

DETECTING FUNDAMENTAL STRUCTURES OF THE
INTERSTELLAR MEDIUM WITH MACHINE LEARNING
ALGORITHMS

NATURWISSENSCHAFTLICHE FAKULTÄT
DEPARTMENT PHYSIK
FRIEDRICH-ALEXANDER-UNIVERSITÄT ERLANGEN-NÜRNBERG

ERLANGUNG DES DOKTORGRADES DR. RER. NAT

VORGELEGT VON
DOMINIC BERNREUTHER

JAHR
2023

Als Dissertation genehmigt
von der Naturwissenschaftlichen Fakultät
der Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung: 06.02.2024

GUTACHTER/IN:

Prof. Dr. Manami Sasaki
Prof. Dr. Jacco Vink

ABSTRACT

The detection and classification of fundamental structures of the interstellar medium (ISM) is tedious but tremendously important work for astronomers. Structures like supernova remnants (SNRs) shape and drive the ISM and have a big impact on evolving galaxies. Those fundamental structures are of bubble-like shape. Finding these bubble-like structures is therefore crucial in the study of galaxy evolution and of the universe in general. Currently, the detection of those bubble-like structures is done mainly by eye, by manually examining astronomical images. In the era of machine learning algorithms, however, this process can be automatized. In this thesis, we present an automated detection and classification pipeline for bubble-like structures based on multi-wavelengths images.

The pipeline is called *BScan* and makes use of two neural networks: a convolutional neural network (CNN) is used for detecting the bubble-like structures in the images and a fully connected neural network (FCNN) is used to classify the detected structures. Both networks are used and combined by classical algorithms. The CNN is part of the nested boxes detection algorithm (NBDA), which scans the available images, and the FCNN then uses the output of the NBDA. The configuration is done via a configuration file where all hyperparameters can be set.

In areas of isolated structures or less crowded regions, *BScan* performs relatively well by localizing already cataloged sources with an average distance of 36 arcsec to the source center given by the catalog. The radius of the detected structure, however, is systematically overestimated. In highly crowded regions *BScan* may detect multiple but separated structures as one single structure. *BScan* also shows some inconsistencies in detecting similar bubble-like structures in images of similar or even the same resolution. The classification has a small estimated error of around 1 per 1000 instances (0.1%) but is heavily dependent on the NBDA.

There are different approaches using different techniques to address the idea of automated structure detection, but yet there has been no attempt to combine the detection and classification aspect in one process. With some further developments and improvements, the idea of *BScan* could set new standards in detecting fundamental structures of the ISM and could be a great tool for astronomy in the future.

ZUSAMMENFASSUNG

Die Erkennung und Klassifikation fundamentaler Strukturen des Interstellaren Mediums (ISM) ist eine mühsame, aber für Astronomen extrem wichtige Aufgabe. Strukturen wie Supernova Überreste (SNRs) formen und beeinflussen das ISM und tragen daher maßgeblich zur Entwicklung von Galaxien bei. Solch fundamentale Strukturen haben meist eine blasenartige Form. Das Finden solcher Strukturen ist entscheidend um die Entwicklung von Galaxien und somit die Evolution des gesamten Universums besser zu verstehen. Diese blasenartigen Strukturen werden derzeit hauptsächlich manuell durch das Scannen astronomischer Bilder entdeckt. Im Zeitalter des maschinellen Lernens kann dieser Prozess jedoch automatisiert werden. In dieser Arbeit entwickeln wir eine automatisierte Erkennungs- und Klassifizierungspipeline für blasenartige Strukturen, basierend auf Multiwellenlängendaten.

Die Pipeline heißt *BScan* und nutzt zwei neuronale Netze, ein Convolutional Neural Network (CNN) zur Erkennung der blasenartigen Strukturen in den Bildern und ein einfaches Fully Connected Neural Network (FCNN), um die erkannten Strukturen anschließend zu klassifizieren. Beide Netzwerke werden durch klassische Algorithmen ergänzt. Das CNN ist beispielsweise Teil des Nested Boxes Detection Algorithmuses (NBDA), welcher die astronomischen Bilder durchmustert. *BScans* Konfiguration erfolgt über eine Konfigurationsdatei, in der alle Hyperparameter festgelegt werden können.

In Gebieten mit isolierten Strukturen oder weniger chaotischen Regionen kann *BScan* bereits katalogisierte Strukturen mit einer mittleren Distanz von 36 Bogensekunden zum tatsächlichen Zentrum lokalisieren. Die Ausdehnung der Struktur wird jedoch systematisch größer geschätzt. In chaotischen Regionen erkennt *BScan* räumlich getrennte Strukturen möglicherweise als eine einzige Struktur. *BScan* zeigt auch einige Schwächen bei der Erkennung ähnlicher blasenartiger Strukturen in Bildern mit ähnlicher oder sogar gleicher Auflösung. Die Klassifizierung weist eine geschätzte Fehlerrate von etwa 1 pro 1000 Klassifizierungen (0.1%) auf, ist jedoch erheblich von den Ergebnissen des NBDA abhängig.

Es gibt unterschiedliche Herangehensweisen, um Strukturen in astronomischen Bildern zu detektieren, aber keine davon kombiniert die Detektion mit der instantanen Klassifikation. Die Idee von *BScan* könnte durch Weiterentwicklung und kontinuierlicher Verbesserung einen neuen Standard für die Detektion von fundamentaler Strukturen des Interstellaren Mediums und zukünftig ein großartiges Werkzeug in der Astronomie werden.

CONTENTS

List of Figures viii

List of Tables xxxi

List of Algorithms xxxiii

I	THEORETICAL FOUNDATION	1
1	THE INTERSTELLAR MEDIUM	3
1.1	Three Phased ISM	3
1.1.1	Cold Phase	4
1.1.2	Warm Phase	7
1.1.3	Hot Phase	9
1.2	Bubble-Like Structures	12
1.2.1	HII Regions	12
1.2.2	Stellar Bubbles	13
1.2.3	Planetary Nebulae	13
1.2.4	Supernova Remnants	14
1.2.5	Emission from bubble-like Structures	18
1.2.6	Bubbles and the inhomogeneous ISM	20
1.3	The Magellanic Clouds	21
2	NEURAL NETWORKS	23
2.1	Fundamental Components and Concepts	23
2.1.1	About Neurons, Weights and Layers	23
2.1.2	Information Processing	24
2.1.3	Forward Propagation	26
2.1.4	Activation Functions	28
2.1.5	Back Propagation and Optimization	29
2.1.6	Loss Functions	31
2.1.7	Deep Neural Networks	32
2.2	Fully Connected Neural Networks	38
2.3	Convolutional Neural Networks	38
2.4	Training	38
2.5	Motivation	43
II	DETECTION AND CLASSIFICATION OF BUBBLE-LIKE STRUCTURES	45
3	DATA	47
3.1	Infrared	47
3.2	Optical	48
3.3	Radio	50
3.4	X-ray	51
4	TRAINING AND TEST DATA	55
4.1	Training Data Set	55
4.2	Kolmogorov-Smirnov Test	58

4.3	Data Set for Pre-Training	59
4.4	Test Data Set	61
4.5	Augmented Data Sets	62
5	DETECTION OF BUBBLE-LIKE STRUCTURES	65
5.1	Architectures	65
5.1.1	VGG16	65
5.1.2	Custom Convolutional Neural Networks	66
5.1.3	Model Naming Convention	67
5.2	Training Procedure	68
5.3	Scanning and Detecting Algorithms	71
5.4	Clustering and Merging	74
5.5	Training and Testing	76
5.5.1	Random Parameter Initialization	77
5.5.2	Pre-Training of the SCNN ₇₁	81
5.5.3	Fine-Tuning	84
5.6	Determine Default Values for the NBDA	87
5.6.1	Stride Factor	88
5.6.2	Number of Nested Boxes	89
5.6.3	Threshold	91
5.6.4	Relative Strength Factor	91
5.7	Summary	93
6	CLASSIFYING DETECTED STRUCTURES	95
6.1	Classifier Architecture	95
6.2	Training and Test Data	95
6.2.1	Creating the Look-Up Table	96
6.2.2	Simulating Data Sets	100
6.3	Training and Evaluation	100
6.4	The Bridge	101
6.5	Summary	102
III APPLYING BSCAN 103		
7	APPLYING BSCAN TO ASTRONOMICAL IMAGES	105
7.1	The Configuration File	105
7.2	Small Magellanic Cloud	109
7.2.1	Configuration	110
7.2.2	Isolated Structures	110
7.2.3	Overlapping Structures	111
7.2.4	Crowded Area	113
7.2.5	Summary	116
7.3	Large Magellanic Cloud	116
7.3.1	Configuration	119
7.3.2	Matching of Catalog Sources	119
7.3.3	Unmatched Detections	123
7.3.4	Alternative Configurations	123
7.3.5	Summary	130

IV	DISCUSSION	137
8	DISCUSSION	139
8.1	Methodology	139
8.1.1	The Training Data Set	139
8.1.2	A Multichannel Neural Network	140
8.1.3	Modification of the Configuration	141
8.2	Detection and Characterization Performance	142
8.3	A More Complete Characterization of SNRs	144
8.3.1	The Newly Detected eROSITA Sources	145
8.3.2	Core Collapse and Thermonuclear Supernovae	146
8.3.3	Morphologies	148
8.4	Beyond the Current BScan Version	149
8.5	BScan as an Astronomical Tool	152
V	APPENDIX	153
A	MATCHED BUBBLE-LIKE STRUCTURES	155
B	UNMATCHED BUBBLE-LIKE STRUCTURES	167
C	NEW EROSITA SNRS AND SNR CANDIDATES	169
	BIBLIOGRAPHY	176

LIST OF FIGURES

- Figure 1.1 Image of the dark cloud Barnard 68 observed in six different wavelengths. The extinction decreases with increasing wavelength. Credit: ESO. 6
- Figure 1.2 Illustration of a molecular cloud with ongoing star formation and subgroups of massive stars compressing the surroundings and trigger new star formation. From Ward-Thompson et al. (2011). 7
- Figure 1.3 This plot shows the temperature dependency of the cooling rate coefficient. Up to temperatures of $T \approx 10^4\text{K}$ cooling is dominated by single ionized carbon. For higher temperatures, the more effective cooling via the Lyman α emission of excited hydrogen atoms sets in. The steep increase of the cooling rate coefficient is referred to as the Lyman α wall. From Williams (2021). 8
- Figure 1.4 Illustration of a shock front propagating with velocity U through the gas in front of the shock (index 1). The medium behind the shock is denoted by the index 2. The pressure p , density ρ , temperature T and velocity is shown. In the rest frame of the shock, the gas in front of the shock moves with $v_1 = |U|$ towards the shock. The shocked gas moves with velocity v_2 away from the shock front. From Longair (2011). 10
- Figure 1.5 A schematic structure of a stellar bubble with its different regions. The stellar wind (a) expands freely till R_1 . The stellar wind starts to interact with the ISM. At a distance R_c from the star, a reverse shock is formed heating up the gas inside (b). A shock wave propagates outwards sweeping up material (c) of the ambient gas and, thus, forming a thin dense shell at R_2 . From Weaver et al. (1977). 14

- Figure 1.6 The Helix Nebula. The white dwarf (small white dot in the center) is surrounded by the ejected material. The ejecta is ionized by the white dwarf and is glowing therefore. Credit: NASA, NOAO, ESA, the Hubble Helix Nebula Team, M. Meixner (STScI), and T.A. Rector (NRAO) 15
- Figure 1.7 The supernova classification scheme from Vink (2020). 16
- Figure 1.8 (a) An illustration of the SNR with the forward and reverse shock heating up the gas (temperature is color coded). The initial ejected mass is indicated by the arrows in the center. (b) Structure of a composite SNR with a pulsar (cross) and the pulsar wind nebula (PWN, blue) embedded in the SNR. From Vink (2020). 17
- Figure 1.9 This plot shows an example of an SNR evolution. The solid line represents the expansion (radius) of the remnant and the dashed line the velocity with which the remnant expands into the ISM. The light gray region marks the time of the Sedov-Taylor phase. From Williams (2021) 18
- Figure 1.10 A schematic image of the fractional ISM within the host galaxy. Molecular clouds and cold gas can be found near the galactic plane, surrounded by the WNM and WIM. Superbubbles are blown into the HIM, which covers a large fraction of the galaxy. From Williams (2021) 20
- Figure 1.11 The Large Magellanic Cloud (a) and the Small Magellanic Cloud (b) in H_α line emission. The structures of different size and brightness are visible in both galaxies Smith C. (2000) 21
- Figure 2.1 A scheme of a neural network. N_n^l denotes the n -th neuron in the layer l . Its basic components are the input ($l = 0$) and output ($l = 3$) layer. Layers between the input and output layer are called hidden layers. The neurons (circles) are connected with each other, indicated by arrows (weights). 24

- Figure 2.2 Illustration of a neural network and its activated neurons. Each neuron represents a gray value. To predict the mean gray value of the three input neurons on the left, the network finds a generalized representation to predicted mean gray value in the output layer on the right. 25
- Figure 2.3 A look into a neural network, showing how the activation of a single neuron is calculated. All weights w_{mn}^{l-1} , activations a_n^{l-1} of every neuron of the previous layer and the bias has an influence on the activation a_m^l of the considered layer. 27
- Figure 2.4 The plots show the ReLU (a) and sigmoid (b) activation function. ReLU activation functions are mainly used for hidden layers, whereas sigmoid activation functions are used in the output layer to interpret the output as a probabilistic value. Plots are taken from Michelucci (2018). 28
- Figure 2.5 This plot shows the principle of gradient descent on a 2D surface in the x_1 - x_2 plane (contours). The black points show the evolution of the loss function after each iteration of the gradient descent. After each iteration the step size in the direction of the steepest descent gets smaller. The gradient descent algorithm converges after 181 iterations and finds a local minimum. 31
- Figure 2.6 Illustration how momentum helps to accelerate the optimization algorithm to converge faster. Without momentum, more iterations are needed because of oscillations (red). Introducing momentum in gradient descent helps to implement bigger steps towards the steepest gradient descent (green) leading to fewer iterations needed. Adopted from Ruder (2016). 32
- Figure 2.7 Comparing the loss of a training batch (solid line) and a test batch, over- and underfitting can be spotted. While a significant higher but steadily decreasing validation loss indicates that the training suffers from underfitting (dashed line), an increasing validation loss after reaching a local minimum hints overfitting (dotted line). 34

- Figure 2.8 Illustration of convolve a 2D input image I with a kernel K . By sliding the kernel over the image and calculating the scalar product (dark gray) the feature map F is created. In order to keep the dimension of the input image, padding (light gray) is applied on I . 34
- Figure 2.9 This Figure demonstrates the maximum pooling operation. A (2×2) kernel is applied on a (4×4) matrix. The kernel hovers over the matrix (gray, left) with a stride of 2. Maximum Pooling returns the maximum value (gray, right) within the kernel. 36
- Figure 2.10 If Dropout is applied during training, neurons are disconnected temporarily by removing connections (light gray) to other neurons of the network. 37
- Figure 2.11 A set of four augmentation operations applied on an original image. With augmentation the original image is transformed in a different image preserving the characteristics of the original. Augmentation operations could be (a) rotations, (b) cropping, (c) vertical and (d) horizontal flipping. In this example of a chess board, (d) does not preserve the characteristics of the original image because of the unnatural scenario of an upside down chess board and should therefore be avoided for augmentation. The image was generated using Midjourney (<https://www.midjourney.com>). 42
- Figure 3.1 This Figure shows the cross-section of the Spitzer Space Telescope with all hardware parts. From Werner et al. (2004). 47
- Figure 3.2 The LMC recorded by the IRAC (a) and MIPS (b) instrument of the Spitzer Space Telescope. Credit: Fazio et al. (2004) and Rieke et al. (2004) 48
- Figure 3.3 The Large (a) and Small (b) Magellanic Cloud observed at the optical emission lines $H\alpha$ (red), [SII] (green) and [OIII] (blue). Credit: Smith C. (2000) 49
- Figure 3.4 This picture shows the Cerro Tololo Inter-American Observatory at sunset. Credit: NOAO/-NSF/AURA 49
- Figure 3.5 This image shows ASKAP telescopes on site. Credit: CSIO. 50

- Figure 3.6 The LMC in radio as seen by ASCAP (a) and ATCA (b). Credit: Pennock et al. (2021) 51
- Figure 3.7 The effective (vignetted) exposure map derived from the eRASS 1 in the 0.6to2.3keV energy band and galactic coordinates. From Predehl et al. (2021). 51
- Figure 3.8 The X-ray telescopes named eROSITA (a) and XMM-Newton (b) are shown schematically, with their X-ray mirrors and detectors, in this Figure. From ESA: XMM-Newton SOC (2022) and Predehl et al. (2021). 52
- Figure 3.9 The images show roughly the same cutout of the LMC recorded by XMM-Newton (a) and eROSITA after eRASS4 (b). Credit: ESA, MPE 52
- Figure 4.1 Each image shows an example of the defined classes *shell* (a), *diffuse* (b) and *none* (c) created from the MCELS2 image showing H α emission. 55
- Figure 4.2 The generation of the dataset represented in a flowchart. Defined regions are used to create images for the training data set by cutting out the regions and standardize the pixel values. Magnetohydrodynamic simulations are added to extend the dataset. 56
- Figure 4.3 These images show the optical (a), soft (b) and hard (c) X-ray simulation of an SNR. Credit: Ustamujic S. et al. at the INAF. 58
- Figure 4.4 KST applied on the entire data set of the Large Magellanic Cloud, Milky Way and magnetohydrodynamic simulations combined in one training data set. Less than 0.3% of the samples within the data set are similar. In fact, similar samples are only found in class *none*. Overall, samples in the training data set are considered as not identical according to the KST. 59
- Figure 4.5 This image shows examples of the Galaxy10 SDSS data set. One image is shown for each class listed in Table 4.2. From <https://astronn.readthedocs.io/en/latest/galaxy10sdss.html>. 60
- Figure 4.6 KST for the test data set. Similar to the training data set, most samples (0.2%) similar to each other are found in the class *none*. 61

- Figure 5.1 Illustration of a possible convolutional blocks. The block consists of two convolutional layers and a maximum pooling layer at the end. Feature map values are normalized by using batch normalization before applying an activation function. The number of convolutional layers, used kernels and the kernel size can vary for each convolutional block. 66
- Figure 5.2 This image shows an illustration of the fully connected top of the SCNN. This part of the network is responsible for classifying. The feature map of the last convolutional block is flattened into a one dimensional feature vector. The length of this vector is equal to the used neurons in the first fully connected layer. For each fully connected layer the number of neurons is reduced. Dropout is used for regularization purpose. 67
- Figure 5.3 The network architecture of a SCNN41. An input image of 50×50 pixel (left) is fed to the network. The network consists of four convolutional blocks (ConvBlock). Each convolutional block i generates m_i feature maps of $n_i \times n_i$ pixels. The first convolutional block generates $m_1 = 8$ feature maps with $n_1 \times n_1 = 50 \times 50$ pixels. On top of the convolutional blocks fully connected layers of 128, 64, 32 and 3 neurons are used. 68
- Figure 5.4 The step (a) and the exponential (b) decay function to control the learning rate during training with an initial learning rate of $LR = 0.001$. The step decay is initialized with a step size of $E' = 25$ epochs and a step decay rate of $S = 0.75$. For the exponential decay a decay rate of $k = 0.01$ is used. 69

- Figure 5.5 This Figure shows how the grid (points) is placed on an image and how the nested boxes (dashed) are defined at a grid position. The grid is centered taking the defined border $\beta^{x,y}$ and the maximum box size Λ into account. Its size is defined by the stride factor γ and the minimum box size $\lambda^{x,y}$ by $\gamma \cdot \beta^{x,y}$. From the first grid position (x_0, y_0) the nested boxes move on the grid as indicated by the arrows till the end position (x_ρ, y_ρ) is reached. The spacing d_s between the nested boxes depends on the number of nested boxes and the minimum and maximum box size. 72
- Figure 5.6 If there are several single detections at different positions (dashed, light grey) detecting the same source (S), the single detections are clustered and merged (M). The merged cluster is then considered as final detection of the source. 75
- Figure 5.7 Training history of the SCNN71-TRS. (a) During training, the precision (red) is increasing and the loss (orange) is decreasing steadily with epochs. The maximum precision is 0.98. There is no sign of overfitting. (b) The validation of the training progress shows similar behavior but bigger fluctuations in precision (green) and loss (blue). For validation the precision reaches a maximum of 0.99. 78
- Figure 5.8 This plot shows the maximum precision for the trained networks of type SCNN ij -TRS (number of convolutional blocks $i \in [3,9]$, number of convolutional layers within a block $j \in [1,3]$). The maximum precision forms a plateau for networks with $i \in [6,8] \wedge j = 1$. For $j = 2$ the maximum precision reaches its peak at $i = 6$, while for $j = 3$ and larger i the tendency is toward lower maximum precision. Just looking at the precision the SCNN6 and SCNN7 perform similar during training. 79

- Figure 5.9 These Figures show the confusion matrices for the trained SCNN₇₁-TRS tested on the test data set (a), the test data set for faint (b) and the test data set for low contrast (c) samples. While the SCNN₇₁-TRS is confident in classifying the class *none*, in case (a), it is confused between the classes *shell* and *diffuse*. Compared to (a), the SCNN₇₁-TRS gets increasingly confused in case (b), not only between *diffuse* and shells, it classifies bubble-like structures into class *none* as well. The result for (c) is highly confused. 80
- Figure 5.10 The Training history of the SCNN₇₁-G6. (a) The training progress with the measured precision (red) and the loss (orange). After 250 epochs the maximum precision is 0.87. The measured loss decreases steadily, hence, no overfitting is observed while training. (b) Validation of the network paint a similar behavior in precision (green) and loss (blue) but with bigger fluctuations which become smaller with epochs. 81
- Figure 5.11 Confusion matrix for the pre-training of the SCNN₇₁ architecture on the G6 data set. The most accurate predictions are for smooth, completely round galaxies (class label B) with 97%, followed by the disk-like galaxies with tight spiral arms observed face on (class label D) with 90%, of true positive predictions. For all other classes the network is confused on a higher level with less than 90% correct predictions. 82
- Figure 5.12 The Training history of the SCNN₇₁-G6-TRS. (a) The precision (red) and loss (orange) over epochs. There is no sign of overfitting while training, since the loss is decreasing steadily. (b) The validation of the training progress shows a slightly different picture. While the precision (green) is also slightly increasing, the loss (blue) seems to have an ascending trend after 200 epochs. However, the loss values over all epochs are small and can be considered as nearly constant. 83

- Figure 5.13 The confusion matrices for the trained SCNN₇₁-G6-TRS tested on the test data set (a), the data set for faint (b) and low contrast (c) samples. Compared to the SCNN₇₁-TRS, which is not pre-trained, the test performance of the SCNN₇₁-G6-TRS most improved on the TES-F but becomes worse on the TES-LC. On the TES the SCNN₇₁-G6-TRS predicts some no-bubble-like structures, most likely filaments, as shells. This is not the case for the SCNN₇₁-TRS. 84
- Figure 5.14 The Training history of the SCNN₇₁-G6-TRS-F recorded for training and validation on the data set for faint samples TRS-F used for fine-tuning. The precision (red) and loss (orange) for training re shown in (a). For validation the precision (green) and loss (blue) are shown in (b). 86
- Figure 5.15 The figure shows the confusion matrices of the fine-tuned network SCNN₇₁-G6-TRS-F applied on the test data sets TES (a), TES-F (b) and TES-LC (c). Compared to the confusion matrices of the pre-trained network SCNN₇₁-G6-TRS the biggest improvement is shown in (c). 87
- Figure 5.16 An image with a bubble-like structure of class *shell* with an image size of 1000×1000 px. The defined grid is using a minimum and maximum box size $\lambda = 50$ px and $\Lambda = 500$ px, respectively, and $\beta = 25$. The stride factor is set to one. Therefore, the bubble-like structure in the center of the image is not centered on any grid position. 88
- Figure 5.17 The 2D histograms to compare the NBDA results on the 30 test images for different stride factors: (a) $\gamma = 0.5$ (b) $\gamma = 0.75$ and (c) $\gamma = 1$ 89
- Figure 5.18 Resulting 2D histograms using different numbers of nested boxes $n = 5$ (a), $n = 10$ (b) and $n = 15$ (c) on the test images. The number of detections per grid position and the distribution in x- and y-direction are similar in all three cases. 90

- Figure 5.19 This Figure shows three meaningful results of the NBDA varying the threshold from $\tau = 0$ up to $\tau = 0.99$. From $\tau = 0$ to $\tau = 0.9$ the detection quality of the central bubble-like structure improves because detections in the outer regions decrease while the detections per grid position stay nearly the same. For $\tau > 0.9$ some detection of the central structure is rejected because of a high threshold. This leads to lower detection per grid position in the central region. The default value is set to $\tau = 0.9$. 92
- Figure 5.20 The 2D histograms of three meaningful results of applying the NBDA with different relative strength factors (RSF) to the test images: (a) $RSF = 10$, (b) $RSF = 10^5$ and (c) $RSF = 10^8$. The default value is set to $RSF = 10^5$. 93
- Figure 6.1 The confusion matrix of the classifier network. It shows now confusion between different labels. All predicted labels match the true label. 101
- Figure 7.1 This Figure shows the pipeline structure for *BScan*. Solid lines indicate fixed parts of the pipeline. Dashed lines are optional paths. While diamonds show the location of the neural networks, boxes show parts of classical algorithms. The circles show the output. 106
- Figure 7.2 Results of *BScan* version V0 (a) and V1 (b) in the region Rg1. The region is located in the SMC at $RA = 14.97^\circ$, $DEC = -71.67^\circ$ with a radius of $R_1 = 0.27^\circ$. This corresponds to a $972px \times 972px$ image. While V1 detects at least some bubble-like structures in this region, V0 cannot detect any structure. 111
- Figure 7.3 Results of *BScan* version V1 in region Rg1 but with different threshold typed parameters: (a) a relative strength factor of $RSF = 0$ and threshold of $\tau = 0.9$, (b) a threshold of $\tau = 0.5$ but the same RSF as in (a). Compared to Figure 7.2 (b) the number of detected bubble-like structures increases. 112

- Figure 7.4 Results of *BScan* version V0 (a) and V1 (b) in the region Rg2. The region is located in the SMC at RA= 21.87°, DEC= -73.33° with a radius of $R_2 = 0.49^\circ$. This corresponds to a 1764px × 1764px image. Both versions detect the brightest bubble-like structures in this region in H α and [OIII]. However, V1 detects the brightest structures in all three bands and detects more structures than V0 in each one. Detected structures in (b) marked with A, B, C and D are fainter compared to other structures and need further investigation. 113
- Figure 7.5 The faint source detected in region A. The detection in H α is clearly visible and has an even fainter counterpart in [SII]. 114
- Figure 7.6 The faint source detected in region B. The detection in [OIII] (blue) is clearly visible. The two shell-like counterparts in H α and [SII] are not detected. This is because of the relatively high threshold typed parameters, the *RSF* and τ . 114
- Figure 7.7 The faint source detected in region C. The detection in [OIII] (blue) is again visible. The extremely faint counterpart in the H α image is not detected. 115
- Figure 7.8 The faint source detected in region D. The detection in [OIII] (blue) is most likely part of a filament with higher local intensity. The filament is part of the large shell-like structure seen in Rg2. 115
- Figure 7.9 This figure shows region B again, but with results of *BScan* version V1 using a relative strength factor of $RSF = 0$ and a threshold of $\tau = 0.5$. With these low thresholds the two shell-like counterparts of the [OIII] in H α only has three detections related to the two structures. 116
- Figure 7.10 This figure shows the results of *BScan* version V0 (a) and V1 (b) in the region Rg3. The region is located in the SMC at RA= 11.76°, DEC= -73.20° with a radius of $R_2 = 0.56^\circ$. This corresponds to a 2016px × 2016px image. The region marked with X is a multiple detection region where more than one bubble-like structures are detected as one single structure. 117

- Figure 7.11 This Figure shows region X marked in Figure 7.10. Multiple structures are detected in one single detection region in $H\alpha$ (red) and [OIII] (blue). 117
- Figure 7.12 In this figure the fraction of detected and characterized bubble-like structures M matching with catalog sources in the LMC $N_c = 92$ is shown. Only 16% of the detected bubble-like structures which are labeled as *SNRc* (green) have counterparts in the catalog. While a small fraction of 4% and 2% is labeled as *likely* (yellow) and candidates (orange), respectively, the majority (17%) of matched bubble-like structures are labeled as *unlikely* (red). Structures which could not be classified as an SNR-like structure but have a counterpart in the catalog are labeled as *other* (magenta) and HII regions (blue). 120
- Figure 7.13 Separation Δ between the center coordinates for each ID. The median separation is $\Delta_m = 36$ arcsec. 121
- Figure 7.14 Ratio between the radius R of the bubble-like structure and the given radius r_c of the matched catalog source. Results are plotted for each ID. 121
- Figure 7.15 Total number of detected bubble-like structures N and the corresponding labels. The structures with matches in the catalog are colored in dark blue and unmatched structures are colored in teal. Most detected structures are labeled as *unlikely* or *other*. For both classes the vast majority is not matched to a catalog source. 124
- Figure 7.16 The five catalog sources which are not detected with the default *BScan* configuration. (a) and (b) show sources listed in (Bozzetto et al., 2017), (c) show a source listed in (Yew et al., 2021) and (d) and (e) show sources listed in (Maggi et al., 2016). The sources are present within the blue circles. 131

- Figure 7.17 NBDA results of the five selected catalog sources using the alternative configuration AC1. The AC1 uses the default configuration but a stride factor of $\gamma = 0.5$. While the sources in (a), (c), (d) and (e) are not detected in any image, the source (blue) in (b) is detected in H α and radio (black). 132
- Figure 7.18 NBDA results of the five selected catalog sources using the alternative configuration AC2. The AC2 uses the default configuration but a stride factor of $\gamma = 0.5$ and a $RSF = 0$. Each source (blue) from (a) to (e) is detected (black) in at least two of the five images. In addition to the detected sources the NBDA detects extremely faint and point-like structures or intensity fluctuations in large scale emission around the central source 133
- Figure 7.19 NBDA results of the five selected catalog sources using the alternative configuration AC3. The AC3 uses the default configuration but a stride factor of $\gamma = 0.5$, a $RSF = 0$ and a primary threshold of $\tau = 0.5$. Each source (blue) from (a) to (e) is detected (black) at least in two of the five images. The AC3 results in even more detections around the central sources. The AC3 is not recommended. 134
- Figure 8.1 The relative frequency of detected features for each classification label evaluated for every bubble-like structure detected in the LMC. 143
- Figure 8.2 Total number of feature vectors responsible for the classification of all detected and labeled bubble-like structures in the LMC. The given ID is the feature vector ID as presented in the look-up table 6.2 in section 6.2.1. 144
- Figure 8.3 The *BScan* detection results for the LMC (a) and zoomed-in region (b). White circles show bubble-like structures as detected by *BScan*. The blue box in (a) shows the location and size of the zoomed-in region. The colored circles in (b) show some examples of X-ray background sources: (cyan) An active galactic nucleus, (red) unlabeled X-ray source of the 1RXS and (green) unlabeled X-ray source of the 2XMM catalog. 145

- Figure 8.4 Different morphology types of SNRs: (a) A shell-type SNR (G299.2-2.9) as seen in X-ray (orange) and infrared (red, green, blue). Credit: NASA, CXC, U. Texas at Arlington, S.Park et al, ROSAT; 2MASS, UMass, IPAC-Caltech, NASA, NSF. (b) The well known mixed-morphology SNR W44 observed in X-ray (cyan) and infrared (red, green, blue). Credit: NASA, CXC, University of Georgia, R.Shelton and NASA, CXC, GSFC, R. Petre; NASA, JPL-Caltech. (c) A composite SNR (G11.2 – 0.3) in the X-ray regime with its central PWN (blue, 2.5 – 8keV) and the SNR (red 0.5 – 1.5keV, green 1.5 – 2.5-keV). Credit: NASA, CXC, Eureka Scientific, M.Roberts et al. (d) The well known Crab Nebula, a pulsar wind nebula as seen in the X-ray (blue), optical (red, yellow) and infrared (purple). Credit: NASA, CXC, SAO, F.Seward; NASA, ESA, ASU, J.Hester and A.Loll; NASA, JPL-Caltech, Univ. Minn., R.Gehrz. 150
- Figure 8.5 This images shows superbubbles inside the nebula N44. (blue) Chandra X-ray data showing hot gas heated by winds and shocks driven by stars and supernovae. (red) Infrared data from the Spitzer Space Telescope which highlight dust and cooler gas. (yellow) Optical data from the 2.2m Max-Planck-ESO telescope showing hot, young stars which are responsible for the nebula to glow due to ultraviolet radiation. Credit: NASA, CXC, U.Mich., S.Oey; NASA, JPL; ESO, WFI 151
- Figure A.1 This figure shows the multi wavelength plot of ID0 at RA = 81.98° , Dec = -65.84° and R = 0.104° . The detected structure is classified as SNRc (green) and is detected in all bands except of X-ray (black). The matched SNR J0527 – 6550 is shown in blue. 155
- Figure A.2 This figure shows the multi wavelength plot of ID1 at RA = 73.93° , Dec = -68.66° and R = 0.077° . The detected structure is classified as SNRc (green) and is detected in all bands except of [OIII] (black). The matched SNR J0455 – 6839 is shown in blue. 155

- Figure A.3 This figure shows the multi wavelength plot of ID2 at $RA = 72.11^\circ$, $Dec = -67.01^\circ$ and $R = 0.074^\circ$. The detected structure is classified as SNRc (green) and is detected in all bands except of X-ray (black). The matched SNR J0448 – 6700 is shown in blue. 155
- Figure A.4 This figure shows the multi wavelength plot of ID3 at $RA = 83.02^\circ$, $Dec = -71.01^\circ$ and $R = 0.056^\circ$. The detected structure is classified as SNRc (green) and is detected in all bands except of $H\alpha$ (black). The matched SNR J0531 – 7100 is shown in blue. 156
- Figure A.5 This figure shows the multi wavelength plot of ID4 at $RA = 84.39^\circ$, $Dec = -66.46^\circ$ and $R = 0.066^\circ$. The detected structure is classified as SNRc (green) and is detected in all bands except of $H\alpha$ and [OIII] (black). The matched SNR J0537 – 6628 is shown in blue. 156
- Figure A.6 This figure shows the multi wavelength plot of ID5 at $RA = 81.50^\circ$, $Dec = -66.08^\circ$ and $R = 0.026^\circ$. The detected structure is classified as SNRc (green) and is detected in all bands except of $H\alpha$ and [OIII] (black). The matched SNR J0526 – 6605 is shown in blue. 156
- Figure A.7 This figure shows the multi wavelength plot of ID6 at $RA = 86.80^\circ$, $Dec = -69.69^\circ$ and $R = 0.054^\circ$. The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0547 – 6941 is shown in blue. 156
- Figure A.8 This figure shows the multi wavelength plot of ID7 at $RA = 81.36^\circ$, $Dec = -65.99^\circ$ and $R = 0.084^\circ$. The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0525 – 6559 is shown in blue. 157
- Figure A.9 This figure shows the multi wavelength plot of ID8 at $RA = 79.94^\circ$, $Dec = -69.44^\circ$ and $R = 0.059^\circ$. The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0519 – 6926 is shown in blue. 157

- Figure A.10 This figure shows the multi wavelength plot of ID9 at RA = 83.50° , Dec = -69.91° and R = 0.069° . The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0534 – 6955 is shown in blue. 157
- Figure A.11 This figure shows the multi wavelength plot of ID10 at RA = 81.25° , Dec = -69.64° and R = 0.058° . The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0525 – 6938 is shown in blue. 157
- Figure A.12 This figure shows the multi wavelength plot of ID11 at RA = 85.05° , Dec = -69.34° and R = 0.039° . The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0540 – 6920 is shown in blue. 158
- Figure A.13 This figure shows the multi wavelength plot of ID12 at RA = 84.56° , Dec = -69.36° and R = 0.039° . The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). Although, the X-ray emission is off. The matched SNR J0538 – 6921 is shown in blue. 158
- Figure A.14 This figure shows the multi wavelength plot of ID13 at RA = 84.02° , Dec = -70.63° and R = 0.039° . The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0536 – 7039 is shown in blue. 158
- Figure A.15 This figure shows the multi wavelength plot of ID14 at RA = 83.51° , Dec = -70.57° and R = 0.039° . The detected structure is classified as SNRc (green) and is detected in H α , [SII] and X-ray (black). The matched SNR J0534 – 7033 is shown in blue. 158
- Figure A.16 This figure shows the multi wavelength plot of ID15 at RA = 82.65° , Dec = -70.11° and R = 0.074° . The detected structure is classified as likely (yellow) and is detected in H α and Radio (black). The matched SNR J0530 – 7008 is shown in blue. 159

- Figure A.17 This figure shows the multi wavelength plot of ID16 at $RA = 78.32^\circ$, $Dec = -69.20^\circ$ and $R = 0.098^\circ$. The detected structure is classified as likely (yellow) and is detected in [SII] and Radio (black). The matched SNR J0513 – 6912 is shown in blue. 159
- Figure A.18 This figure shows the multi wavelength plot of ID17 at $RA = 86.97^\circ$, $Dec = -70.41^\circ$ and $R = 0.039^\circ$. The detected structure is classified as likely (yellow) and is detected in $H\alpha$ and X-ray (black). The matched SNR J0547 – 7025 is shown in blue. 159
- Figure A.19 This figure shows the multi wavelength plot of ID18 at $RA = 74.97^\circ$, $Dec = -70.19^\circ$ and $R = 0.115^\circ$. The detected structure is classified as likely (yellow) and is detected in $H\alpha$ and Radio (black). The matched SNR J0459 – 7008 is shown in blue. 159
- Figure A.20 This figure shows the multi wavelength plot of ID19 at $RA = 72.67^\circ$, $Dec = -70.83^\circ$ and $R = 0.103^\circ$. The detected structure is classified as candidate (orange) and is detected in [OIII] and Radio (black). The matched SNR J0450 – 7050 is shown in blue. 160
- Figure A.21 This figure shows the multi wavelength plot of ID20 at $RA = 72.34^\circ$, $Dec = -69.34^\circ$ and $R = 0.049^\circ$. The detected structure is classified as candidate (orange) and is detected in $H\alpha$, [OIII] and Radio (black). The matched SNR J0449 – 6920 is shown in blue. 160
- Figure A.22 This figure shows the multi wavelength plot of ID21 at $RA = 83.62^\circ$, $Dec = -70.55^\circ$ and $R = 0.059^\circ$. The detected structure is classified as unlikely (red) and is detected in $H\alpha$ and [SII] (black). The matched SNR J0534 – 7033 is shown in blue. 160
- Figure A.23 This figure shows the multi wavelength plot of ID22 at $RA = 76.69^\circ$, $Dec = -70.44^\circ$ and $R = 0.039^\circ$. The detected structure is classified as unlikely (red) and is detected in X-ray only (black). The matched SNR J0506 – 7026 is shown in blue. 160

- Figure A.24 This figure shows the multi wavelength plot of ID23 at $RA = 77.15^\circ$, $Dec = -69.47^\circ$ and $R = 0.059^\circ$. The detected structure is classified as unlikely (red) and is detected in $H\alpha$ and [SII] (black). The matched SNR J0508 – 6928 is shown in blue. 161
- Figure A.25 This figure shows the multi wavelength plot of ID24 at $RA = 79.69^\circ$, $Dec = -69.65^\circ$ and $R = 0.020^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0518 – 6939 is shown in blue. 161
- Figure A.26 This figure shows the multi wavelength plot of ID25 at $RA = 84.45^\circ$, $Dec = -69.17^\circ$ and $R = 0.020^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0537 – 6910 is shown in blue. 161
- Figure A.27 This figure shows the multi wavelength plot of ID26 at $RA = 75.51^\circ$, $Dec = -67.66^\circ$ and $R = 0.049^\circ$. The detected structure is classified as unlikely (red) and is detected in $H\alpha$ and [SII] (black). The matched SNR J0502 – 6739 is shown in blue. 161
- Figure A.28 This figure shows the multi wavelength plot of ID27 at $RA = 78.10^\circ$, $Dec = -67.12^\circ$ and $R = 0.026^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0512 – 6707 is shown in blue. 162
- Figure A.29 This figure shows the multi wavelength plot of ID28 at $RA = 82.46^\circ$, $Dec = -66.91^\circ$ and $R = 0.079^\circ$. The detected structure is classified as unlikely (red) and is detected in X-ray only (black). The matched SNR J0529 – 6653 is shown in blue. 162
- Figure A.30 This figure shows the multi wavelength plot of ID29 at $RA = 81.12^\circ$, $Dec = -66.39^\circ$ and $R = 0.079^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0524 – 6624 is shown in blue. 162

- Figure A.31 This figure shows the multi wavelength plot of ID30 at $RA = 73.69^\circ$, $Dec = -66.43^\circ$ and $R = 0.030^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0454 – 6626 is shown in blue. 162
- Figure A.32 This figure shows the multi wavelength plot of ID31 at $RA = 76.54^\circ$, $Dec = -65.70^\circ$ and $R = 0.098^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0506 – 6541 is shown in blue. 163
- Figure A.33 This figure shows the multi wavelength plot of ID32 at $RA = 76.53^\circ$, $Dec = -68.26^\circ$ and $R = 0.126^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0506 – 6815 is shown in blue. 163
- Figure A.34 This figure shows the multi wavelength plot of ID33 at $RA = 83.40^\circ$, $Dec = -72.05^\circ$ and $R = 0.020^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0533 – 7202 is shown in blue. 163
- Figure A.35 This figure shows the multi wavelength plot of ID34 at $RA = 85.79^\circ$, $Dec = -71.07^\circ$ and $R = 0.020^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0542 – 7104 is shown in blue. However, the detected radio emission is at the edge of J0542 – 7104 and comes most likely from another source. 163
- Figure A.36 This figure shows the multi wavelength plot of ID35 at $RA = 87.69^\circ$, $Dec = -68.39^\circ$ and $R = 0.059^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0550 – 6823 is shown in blue. 164
- Figure A.37 This figure shows the multi wavelength plot of ID36 at $RA = 85.44^\circ$, $Dec = -66.99^\circ$ and $R = 0.079^\circ$. The detected structure is classified as unlikely (red) and is detected in X-ray only (black). The matched SNR J0541 – 6659 is shown in blue. 164

- Figure A.38 This figure shows the multi wavelength plot of ID37 at $RA = 75.225^\circ$, $Dec = -65.20^\circ$ and $R = 0.079^\circ$. The detected structure is classified as other (brown) and is detected in $H\alpha$ only (black). The matched SNR J0500 – 6512 is shown in blue. 164
- Figure A.39 This figure shows the multi wavelength plot of ID38 at $RA = 74.41^\circ$, $Dec = -67.653^\circ$ and $R = 0.039^\circ$. The detected structure is classified as other (brown) and is detected in $H\alpha$ and [SII] (black). The matched SNR J0457 – 6739 is shown in blue. 164
- Figure A.40 This figure shows the multi wavelength plot of ID39 at $RA = 84.04^\circ$, $Dec = -67.56^\circ$ and $R = 0.039^\circ$. The detected structure is classified as other (brown) and is detected in [OIII] (black). The matched SNR J0536 – 6735 is shown in blue. 165
- Figure A.41 This figure shows the multi wavelength plot of ID40 at $RA = 82.06^\circ$, $Dec = -67.44^\circ$ and $R = 0.059^\circ$. The detected structure is classified as other (brown) and is detected in [SII] (black). The matched SNR J0528 – 6727 is shown in blue. 165
- Figure A.42 This figure shows the multi wavelength plot of ID41 at $RA = 80.78^\circ$, $Dec = -67.88^\circ$ and $R = 0.039^\circ$. The detected structure is classified as other (brown) and is detected in [SII] (black). The matched SNR J0523 – 6753 is shown in blue. 165
- Figure A.43 This figure shows the multi wavelength plot of ID42 at $RA = 76.72^\circ$, $Dec = -65.16^\circ$ and $R = 0.108^\circ$. The detected structure is classified as other (brown) and is detected in [SII] (black). The matched SNR J0506 – 6509 is shown in blue. 165
- Figure A.44 This figure shows the multi wavelength plot of ID43 at $RA = 80.40^\circ$, $Dec = -65.70^\circ$ and $R = 0.039^\circ$. The detected structure is classified as other (brown) and is detected in [SII] (black). The matched SNR J0521 – 6543 is shown in blue. 166

- Figure A.45 This figure shows the multi wavelength plot of ID44 at RA = 83.89° , Dec = -66.04° and R = 0.118° . The detected structure is classified as HII (cyan) and is detected in H α and X-ray (black). The matched SNR J0535 – 6602 is shown in blue. 166
- Figure B.1 This figure shows the detected bubble-like structure classified as candidate (orange) at RA = $85.00 \pm 0.01^\circ$, Dec = $-69.79 \pm 0.01^\circ$ and radius of R = 0.020° . It is detected in H α , [OIII] and radio (black). There is a YSO within a radius of 0.01° . 167
- Figure B.2 This figure shows the detected bubble-like structure classified as candidate (orange) at RA = $84.18 \pm 0.01^\circ$, Dec = $-66.41 \pm 0.01^\circ$ and radius of R = 0.020° . It is detected in [OIII] and radio (black). Although, the detection in radio is at a different position and much larger compared to the one in [OIII]. It is likely that in radio another emission is detected. Within a radius of 0.01° there are some infrared sources. 167
- Figure B.3 This figure shows the detected bubble-like structure classified as candidate (orange) at RA = $81.59 \pm 0.01^\circ$, Dec = $-66.25 \pm 0.01^\circ$ and radius of R = 0.043° . It is detected in [OIII] and radio (black). In radio, however, the larger structure is detected. There is an HII region within in the reach of 0.01° named L–253. 168
- Figure B.4 This figure shows the detected bubble-like structure classified as unlikely (red) at RA = $72.91 \pm 0.01^\circ$, Dec = $-67.10 \pm 0.01^\circ$ and radius of R = 0.115° . It is detected in H α and [SII] (black). 168
- Figure B.5 This figure shows the detected bubble-like structure classified as unlikely (red) at RA = $72.82 \pm 0.01^\circ$, Dec = $-67.45 \pm 0.01^\circ$ and radius of R = 0.075° . It is detected in H α and [SII] (black). 168
- Figure B.6 This figure shows the detected bubble-like structure classified as other (brown) at RA = $74.94 \pm 0.01^\circ$, Dec = $-67.92 \pm 0.01^\circ$ and radius of R = 0.177° . It is detected in [OIII] only (black). 168

- Figure B.7 This figure shows the detected bubble-like structure classified as other (brown) at $RA = 73.88 \pm 0.01^\circ$, $Dec = -67.14 \pm 0.01^\circ$ and radius of $R = 0.186^\circ$. It is detected in $H\alpha$ only (black). 168
- Figure C.1 J0451 – 6717: This new eROSITA SNR candidate (blue) is located next to structures visible in the optical images. Indeed, *BScan* detects the bubble-like structure closest to the SNR candidate in $H\alpha$ and [OIII]. However, no clear bubble-like structure can be seen at the location of the candidate. In X-ray there is some emission coming from the considered region, but is not detected by *BScan*, unfortunately. 169
- Figure C.2 J0456 – 6830: For this SNR candidate (blue) there is no clear bubble-like structure visible in any image used for this work. Therefore, *BScan* does not detect this newly detected candidate. 169
- Figure C.3 J0507 – 7143: There is no emission visible in the optical images for this SNR candidate (blue). The point source seen in the non-thermal radio and X-ray image is an active galactic nucleus (Zangrandi et al. 2023, in prep.). In the X-ray image there may be some emission coming from the SNR candidate, but it is too faint to be detected by *BScan*. 169
- Figure C.4 J0510 – 6853: In this region of the new SNR candidate (blue) a shell-like structure is visible in the optical images (except of [OIII]). This structure is partly detected in the $H\alpha$ image (black). In radio and X-ray there is also some faint emission but is not detected by *BScan*. 170
- Figure C.5 J0521 – 6853: This eROSITA candidate (blue) is located within a bigger complex optical structure where some subregions show an enhanced $[SII]/H\alpha$ ratio. There may be some enhanced X-ray emission too, but such faint emission is hardly possible to detect by *BScan*. 170
- Figure C.6 J0521 – 6936: For this SNR candidate no clear structure can be seen in all five images used. However, there could be faint emission within the candidate (blue) in the X-ray image. This emission is not detected by *BScan*. 170

- Figure C.7 J0523 – 6804: This candidate (blue) is extremely faint in images used in this work. Consequently, *BScan* is not able to detect J0523 – 6804 detected by Zangrandi et al. (2023, in prep.) 170
- Figure C.8 J0525 – 6621: This eROSITA SNR candidate- (blue) is located in between some bubble-like structure in the optical and radio band. In the non-thermal radio image *BScan* detects one of those structures in the south. In the X-ray image a faint diffuse emission within the candidate's region is visible indeed, but not detected by *BScan*. 171
- Figure C.9 J0528 – 6719: For this SNR candidate (blue) a faint but visible shell-like emission can be spotted in the X-ray image. Unfortunately, it is not detected by *BScan*. In the the optical and radio band no bubble-like structures are visible. 171
- Figure C.10 J0532 – 6554: For this SNR candidate no optical and radio emission is visible. In X-ray a very faint diffuse emission ca be seen within the blue circle. *BScan* does not detect this candidate. 171
- Figure C.11 J0549 – 7001: This SNR candidate is not detected by *BScan*. It is located next to a bright and complex region in the optical band. No non-thermal radio emission which is correlated with the SNR candidate can be spotted. In X-ray a faint diffuse and irregular shape is visible in the region of the remnant (blue). 171
- Figure C.12 J0456 – 6533: For this SNR no structure is visible in the optical and radio band. In X-ray, however, a faint diffuse emission can be spotted within the blue circle. This source is not detected by *BScan*. 172
- Figure C.13 J0506 – 7009: This confirmed SNR (blue) is detected in the X-ray regime (black). According to Zangrandi et al. (2023, in prep.) the SNR is brightest in the mid X-ray range, which supports a thermonuclear supernova scenario, and faint shell is visible in [SII]. Zangrandi et al. (2023, in prep.) also spotted a non-thermal radio semi-shell. However, *BScan* does not detect any structure in optical or radio. 172

Figure C.14 J0543 – 6624: An eROSITA SNR (blue) with an irregular shape brightest in soft X-rays (Zangrandi et al. 2023, in prep.). A shell-like structure is visible in [SII] but not detected by *BScan*.
172

LIST OF TABLES

Table 1.1	This table summarizes the three phases of the Interstellar Medium and give typical temperatures and structures (McKee et al., 1977). 4
Table 2.1	This table gives an overview of the possible outcome of a prediction. A positive prediction indicates that the corresponding sample belongs to a certain label, where a negative prediction indicates that the sample does not belong to this label. 39
Table 4.1	Number of training samples in each class originating from astronomical images of the LMC and MW as well as from MHDS. In total, there are 1563 samples in the training data set. 57
Table 4.2	Overview of the ten classes of the Galaxy10 SDSS data set. In total 21830 samples are available with this data set. The six crosses indicate if the samples of the corresponding class are used in the reduced data set (G6) for pre-training. In the end there are 19296 samples in the pre-training data set. A short label for each class used in the G6 is introduced. 61
Table 4.3	The number of test samples in each class originating from astronomical images of the SMC. In total, there are 100 samples in the test data set. 62
Table 5.1	This table gives an overview of the possible input parameter for the training routine. 70
Table 5.2	Training parameters used to train the networks SCNN _{ij} -TRS with $i \in [3, 9]$ and $j \in [1, 3]$. 77

Table 5.3	This table shows the top four performing networks on the test data set (TES), the augmented data set for faint (TES-F) and low-contrast (TES-LC) samples considering the weighted mean of true positive predictions over all classes and test sets. For each test set, the mean fraction of true positive predictions over all classes are shown. The weighted mean performance Π is calculated using the given weights in square brackets. 80
Table 5.4	This table shows the SCNN ₇₁ architecture with and without pre-training in comparison. Both versions are tested on the test data set (TES), the augmented data set for faint (TES-F) and low-contrast (TES-LC) samples considering the weighted mean performance as shown in Table 5.3. The pre-trained version has a slightly better Π with $\Delta\Pi = 0.1$. This small increase in performance comes from the better performance on the TES-F. 85
Table 5.5	Parameters used to fine-tune the SCNN ₇₁ -G6-TRS. 85
Table 5.6	Comparison of the fine-tuned versions of the SCNN ₇₁ -G6-TRS network. The best weighted mean performance is reached by fine-tuning the network on faint training samples. For a better comparability the same weights are used as for pre-training. 86
Table 5.7	Summary of the minimum and maximum values to find the default values of the NBDA parameters. Considered parameters are the stride factor (γ), the number of nested boxes (n), the threshold (τ) and the relative strength factor (RSI). 88
Table 5.8	This table summarizes the number of detected bubble-like structures in the test images using a 1×1 grid with different numbers of nested boxes $n \in \{5, 10, 15, 20, 25\}$. The maximum number of detections is reached the first time with $n = 10$. 91
Table 5.9	Table of the determined default values for the NBDA. 93

Table 6.1	The six labels which characterize the detected bubble-like structure and a brief description what the label means. The features needed to label the bubble-like structures accordingly is also given (P/SF: primary/secondary feature, P/SFG primary/secondary feature group) 96
Table 6.2	This table shows all 64 feature vectors and the corresponding labels which are used for simulating the training and test data set. It can also be used as look-up table. 97
Table 6.3	Training parameter and their values used for training the classifier. 100
Table 7.1	A summary of the main <i>BScan</i> V1 results discussed and problems revealed in section 7.2. 118
Table 7.2	This table shows the list of detected bubble-like structures matched to catalog sources. Each detection has an ID, the central coordinates of the bubble-like structure in RA, Dec, the estimated radius R, the label according the classification and the feature vector the classification is based on. The name of the matched source is also given. The error on the estimated coordinates is given as $\Delta_m = 0.01^\circ$. 125
Table 7.3	A compact summary of the main results discussed in section 7.3. For each result one or more reasons (or errors) and a comment is given. 135
Table C.1	This table shows the new eROSITA SNRs and SNR candidates detected by Zangrandi et al. - (2023, in prep.). The coordinates (J2000), the current status and a short description of each source is given. If the corresponding source is detected by <i>BScan</i> the emission and the label is added for each source as well. * as described in Zangrandi et al. (2023, in prep.) 173

LIST OF ALGORITHMS

Algorithm 1	Moving to the next grid position	73
Algorithm 2	Defining Nested Boxes	73
Algorithm 3	Search for bubble-like structures in batch	74
Algorithm 4	Nested Boxes Detection Algorithm	74

ACRONYMS

H	hydrogen
He	helium
ISM	interstellar medium
PNe	planetary nebulae
SN	supernova
SNe	supernovae
SNR	supernova remnant
CC	core collapse
PWN	pulsar wind nebula
PWNe	pulsar wind nebulae
ICS	inverse Compton scattering
CNM	cold neutral medium
WNM	warm neutral medium
WIM	warm ionized medium
HIM	hot intercloud medium
UV	ultraviolet
GMC	giant molecular clouds
MW	Milky Way
MC	Magellanic Cloud
LMC	Large Magellanic Cloud
SMC	Small Magellanic Cloud
ReLU	rectified linear unit
ADAM	adaptive movement estimation algorithm
DNN	deep neural network
CCE	categorical cross-entropy
BCE	binary cross-entropy
FCNN	fully connected neural network
CNN	convolutional neural network
TP	true positive

FP	false positive
TN	true negative
FN	false negative
LR	learning rate
MCELS	UM/CTIO magellanic cloud emission-line survey
FOV	field of view
CTIO	Curtis Schmidt telescope at Cerro Tololo inter-American observatory
eROSITA	extended Röntgen survey imaging telescope array
eRASS	eROSITA all-sky survey
SRG	Spectrum-Röntgen-Gamma
HEW	half-energy width
XMM-Newton	X-ray multi-mirror-Newton
FWHM	full width half maximum
EPIC	european photon imaging camera
ASKAP	Australian square kilometre array pathfinder
ATCA	Australian telescope compact array
GLIMPSE	galactic legacy infrared midplane survey extraordinaire
MIPS	multiband imaging photometer
MIPSGAL	survey of the inner galactic disk with MIPS
SAGE	surveying the agents of a galaxy's evolution
IRAC	infrared array camera
MHDS	magnetohydrodynamic simulation
LoS	line of sight
MWP	Milky Way project
IR	infrared
TES	test data set
TES-F	test data set of faint samples
TES-LC	test data set of low-contrast samples
TRS	training data set
TRS-F	training data set of faint samples
TRS-LC	training data set of low-contrast samples
SDSS	Sloan digital sky survey
G ₁₀	Galaxy ₁₀ data set
G ₆	reduced Galaxy ₁₀ data set
KST	Kolmogorov-Smirnov test

NNA	neural network architecture
ILSVRC ₁₄	<i>ImageNet</i> large scale visual recognition challenge 2014
SCNN	small convolutional neural network
NBDA	nested boxes detection algorithm
EOG	end-of-grid
RSI	relative strength index
RSF	relative strength factor
DS ₉	SAOImageDS ₉
kNN	k-nearest neighbor
CM	confusion matrix
PF	primary feature
PFG	primary feature group
SF	secondary feature
SFG	secondary feature group
WCS	world coordinate system
YSO	young stellar object
AC	alternative configuration
AC ₁	first alternative configuration
AC ₂	second alternative configuration
AC ₃	third alternative configuration
GNN	generative neural network
GAN	generative adversarial network
AGN	active galactic nucleus
GGM	gaussian gradient magnitude
SFH	star formation history

Part I

THEORETICAL FOUNDATION

THE INTERSTELLAR MEDIUM

After the discovery of atomic hydrogen (H) clouds in the 1950s, astronomical observations were performed to study the interstellar medium (ISM) which is all barionic matter filling the space between stars of a galaxy. It consists of a mixture of gas and dust in the atomic, molecular or ionized state. H is the most abundant element in the ISM, contributing around 74% of the total mass followed by helium (He) at roughly 25%. The remaining 1% consists of heavier elements. This chapter is mainly based on Williams (2021) and Dopita et al. (2003) if not stated otherwise.

To get an idea of the scale of the ISM, consider a galaxy containing mainly H, with a particle density of $n_{\text{H}} \approx 1\text{cm}^{-3}$, and an average distance of $\approx 10^{16}\text{m}$ between stars. The estimated time between collisions of two H particles is then $t_{\text{coll}} \approx 10^{10}\text{s} \simeq 10^3\text{yrs}$. In comparison, the average time between collisions for air at sea level can be estimated to $t_{\text{coll}} \approx 10^{-9}\text{s}$. This illustrates how rare collision events are in the ISM. In fact, the probability that an excited element in the ISM is de-excited via collision is so low, it is more likely that the particle emits its excess energy via radiation, making it possible to observe emission lines not observable in earth based laboratories.

The ISM is highly affected by the activity of stars as it hosts dense molecular clouds which are regions of intense star formation. The radiation, e. g., from massive stars ionize their vicinity by photoionization. Stellar winds cause shocks and can ionize the surrounding medium as well. If a dying star explodes in a supernova (SN), it shocks its surroundings and enriches the ISM with new elements. The resulting supernova remnant (SNR) expands into the ISM. In short, stars in a galaxy influence the surrounding ISM which in turn causes the galaxy to evolve.

Thus, the ISM is a unique laboratory for astronomy studying star formation, evolution and their interaction with the ISM itself by study the shocks and remnants of the above-mentioned events. In section 1.1 three phases of the ISM are discussed, while in section 1.2 the formation of some bubble-like structures are explained. The Magellanic Clouds are well suited to study these bubble-like structures. Section 1.3 shows why.

1.1 THREE PHASED ISM

McKee et al. (1977) developed the model of a three-phased ISM consisting of gas and dust. The ISM can be divided into a cold neutral

(section 1.1.1), warm partially ionized (section 1.1.2) and a hot diffuse and ionized gas (section 1.1.3).

The different phases can be understood as a dynamic ISM which undergoes heating and cooling processes. A phase is considered stable if it is in thermal equilibrium, that means

$$N^2\Lambda(T) = N\Gamma(T), \quad (1)$$

where N is the gas number density and Λ and Γ are the cooling and heating rate coefficients, respectively. The heating and cooling mechanisms contributing to Λ or Γ vary with temperature T . Therefore, each phase has a characteristic temperature. Table 1.1 summarizes the typical temperatures and structures that can exist in each phase.

phase	typical T [K]	typical structures
cold	10^2	molecular clouds
warm	$10^2 - 10^4$	HII regions
hot	$> 10^6$	supernova remnants

Table 1.1: This table summarizes the three phases of the Interstellar Medium and give typical temperatures and structures (McKee et al., 1977).

Different phases can coexist if there is pressure equilibrium. The pressure in each phase is given by

$$p = Nk_B T \quad (2)$$

where k_B the Boltzmann constant. Comparing the wide range of typical phase temperatures in Table 1.1 leads to the conclusion of a wide range of densities, assuming thermal and pressure equilibrium.

1.1.1.1 Cold Phase

A small fraction of the ISM volume is made out of cold neutral gas, the cold neutral medium (CNM), and dense molecular clouds. It can be observed within the galactic plane. In total, the CNM occupies just 2 – 4% of the entire ISM volume (McKee et al., 1977) and mostly consists of atomic hydrogen. The cold dense clouds consist of dust and gas with densities from 10^2cm^{-3} up to 10^3cm^{-3} . In these high density environments molecules can be formed on the surface of dust grains.

Cold Neutral Medium

In areas of high densities, the CNM consists mainly of atomic hydrogen (HI). There are a few other elements like carbon, nitrogen and oxygen present in the neutral medium but because of the dominating abundance of hydrogen, the gas is referred to as hydrogen gas

or HI-gas. HI exists mostly in its ground state, with the quantum number $n = 1$, but can still be detected via transitions of the hyper-fine structure. It originates from a change in spin orientation of the electron and proton. The probability of a change in the spin state is so low, that it takes about 11Myr for the transition to occur. Because of the high number of neutral hydrogen atoms in the ISM, a small fraction decay via spontaneous emission and emit photons of a wavelength $\lambda \approx 21\text{cm} = \lambda_{21}$. The λ_{21} emission line cannot be observed on earth since the decay is extremely rare and a large t_{coll} is required. It was thought to be impossible to detect the λ_{21} emission till proper instruments were engineered. That is why it is called a forbidden emission line. Since molecular clouds form predominantly alongside dense atomic hydrogen gas, the λ_{21} emission line is used as tracking tool for molecular clouds.

The temperature of an atomic gas in the ISM depends on the number of photoelectrons and their kinetic energy. Kinetic energy transfer between these electrons and other particles within the gas will increase its temperature. A fraction of the electrons excite atoms, which leads to radiative losses and, therefore, to lower temperatures. If the kinetic energy is high enough to excite the atomic hydrogen gas the cooling process via the Lyman α (transition of quantum number $n = 2 \rightarrow 1$) emission occurs. Because of the high number of hydrogen this cooling mechanism is very effective. Observations revealed temperatures of atomic hydrogen gas from $T \sim 100\text{K}$ to roughly 8000K which is evidence for a two-phased neutral medium: a cold and a warm neutral medium. The warm neutral medium (WNM) is discussed in section 1.1.2.

Dust

The presence of dust grains lead to wavelength dependent extinction $A(\lambda) \sim \lambda^{-1.7}$. Observations show that dust grains in the ISM are a few μm in size. Although there are more small than large dust grains, most of the mass is in larger grains. Extinction is mainly caused by smaller dust grains at shorter wavelengths and are effective absorbers of optical and ultraviolet (UV) photons. Whereas larger grains dominate at longer wavelength. Thus, depending on the number density in the line of sight, dust absorbs most of the visible light from stars in the background (see Figure 1.1). Dust also works like a protection shield against UV emission. Far-ultraviolet photons ($\lambda < 200\text{nm}$), e.g., can destroy molecules by breaking the chemical bonds. This process is called photodissociation. Consequently, the inner regions of cold, dark clouds consist of molecules and are called molecular clouds (Ward-Thompson et al., 2011). While absorbing photons, the temperature of dust grains increases. Since dust grains are solid particles, they emit photons over a continuous spectrum in the infrared

range and, therefore, cool down again. Typical temperatures of dust are around $T \sim 10\text{K}$.

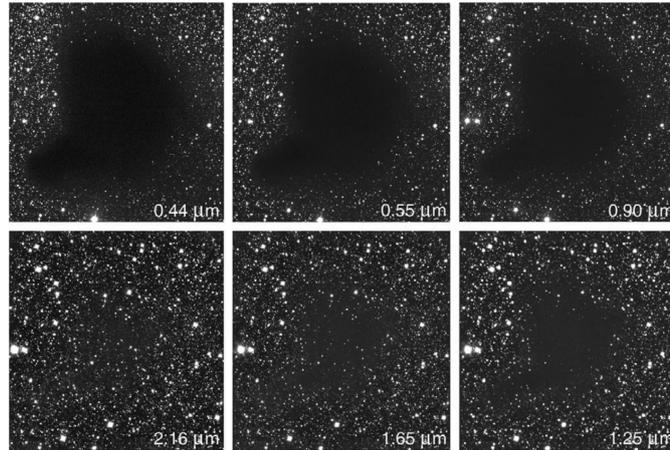


Figure 1.1: Image of the dark cloud Barnard 68 observed in six different wavelengths. The extinction decreases with increasing wavelength. Credit: ESO.

Molecular Clouds

The most abundant molecule in the ISM is molecular hydrogen (H_2). Its symmetry prevents rotational transitions and because of the absence of charge separation, H_2 has no dipole moment. However, it can be excited by collisions. The resulting vibrational modes can lead to detectable emission from molecular hydrogen. The lowest state can only be excited in warm regions with temperatures of $T > 500\text{K}$. In molecular clouds, with temperatures in the range between $10 - 30\text{K}$, molecular hydrogen is in its ground state and, therefore, not observable. Other molecules, like carbon monoxide (CO), are far less common but have an asymmetric charge and mass distribution leading to a dipole moment and a variety of rotational transitions possible. Although these types of molecules are rare compared to H_2 , they emit a detectable amount of photons with a wavelength of a few millimeters. CO is therefore used as a tracer of molecular clouds.

The size of molecular clouds can range from 0.1pc to roughly 100pc , where the largest clouds are called giant molecular clouds (GMC). Within these clouds, the gas density can reach a critical level such that self-gravity forces the gas to collapse and form even denser regions which are known as new star forming regions. Usually, stars are born in a group of a few tens to 10^7 members. Particularly interesting are so-called OB associations, a group of young massive stars with spectral types O and B. The environment of the OB association is dissolved by the OB-stars because of their stellar winds and UV emission creating an expanding HII region (see section 1.2.1). The subgroup

of newborn stars become optically visible. In a cosmological sense, OB-stars are short-lived stars resulting in SN explosions. The remnant expands into the ISM and compresses the surrounded medium, forming new high density regions which can trigger star formation again (see Figure 1.2). Therefore, several subgroups at different evolution times can be observed in relatively proximity to each other (Ward-Thompson et al., 2011). Molecular Clouds are not only sites of star formation, they drive the evolution of the ISM and consequently their host galaxy.

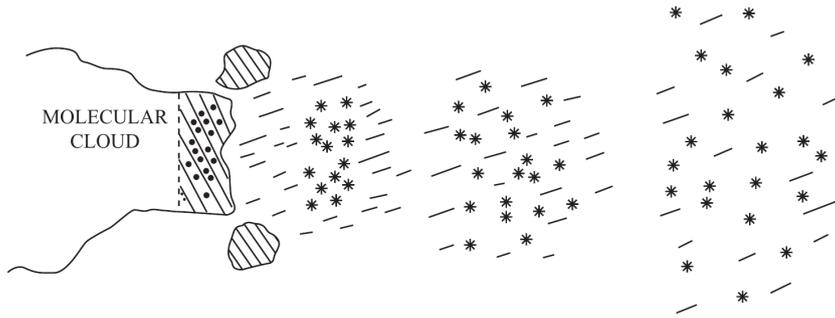


Figure 1.2: Illustration of a molecular cloud with ongoing star formation and subgroups of massive stars compressing the surroundings and trigger new star formation. From Ward-Thompson et al. (2011).

1.1.2 Warm Phase

The warm phase is made of warm, partially photoionized material surrounding the cold dense clouds, also called the „intercloud medium“. It occupies a much larger volume, but includes much less mass (McKee et al., 1977).

Warm Neutral Medium

As stated in section 1.1.1 the temperature of the neutral atomic hydrogen gas depends on the number of photoelectrons and their kinetic energy. In fact, the energy of the first state of hydrogen is relatively high and the kinetic energy of the photoelectrons usually present in the ISM is not sufficient to excite this level. Therefore, photoelectrons transfer some of their kinetic energy to other particles via collisions and heating up the gas. The next largest component of the neutral medium is carbon (C) which can be partially ionized by the electrons. Indeed, the single ionized carbon (CII) with the emission line at $\lambda = 158\mu\text{m}$ in the far-infrared is the dominant cooling mechanism in HI-gas. Because of the small ratio between C- and H-atoms, the heating process mentioned before is more efficient up to temperatures, above which hydrogen atoms can be excited. At this point, the

gas cools via the Lyman α line emission. This point is referred to as the Lyman α wall and sets an upper limit for the temperature of the atomic hydrogen gas (Figure 1.3). During the heating process, the WNM can reach a temperature of $T \sim 8000\text{K}$ with densities of around 1cm^{-3} .

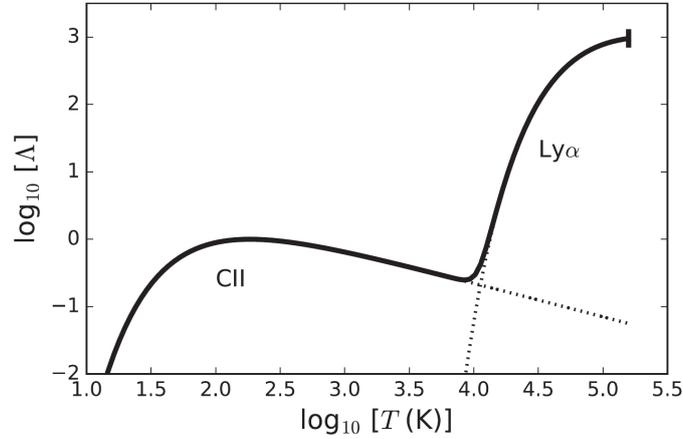


Figure 1.3: This plot shows the temperature dependency of the cooling rate coefficient. Up to temperatures of $T \approx 10^4\text{K}$ cooling is dominated by single ionized carbon. For higher temperatures, the more effective cooling via the Lyman α emission of excited hydrogen atoms sets in. The steep increase of the cooling rate coefficient is referred to as the Lyman α wall. From Williams (2021).

Warm Ionized Medium

Ionization is a single-body process involving photons with sufficient energy to ionize the available atoms. The photoionization rate of a neutral hydrogen gas therefore depends on the number of photons. Massive young stars, like in the aforementioned OB associations, emit photons of sufficient energy isotropically into the ISM. In fact, older and hotter stars emit more ionizing photons than younger ones. As hydrogen is the most abundant element, the ionization process



is quite common. This process is compromised by the recombination of free electrons and protons. Recombination is therefore a two-body process. If a proton captures an electron, a hydrogen atom in an excited state ($n \geq 2$) is created, in the relevant cases. It can then radiate energy. In the case that the electron recombines in the ground state, the resulting photon can ionize other hydrogen atoms again, meaning the electron-proton population in the hydrogen gas remains the same. Recombination depends on the availability of free electrons and protons and, therefore, on the hydrogen density. Since the ISM is not homogeneous, and stars tend to be formed in molecular clouds,

the density decreases rapidly with increasing distance from the star. Thus, the ionization process can be faster than the recombination of electrons and protons. Observations in the radio and optical bands, for example, show that a large fraction of the ISM volume is occupied by ionized gas, the warm ionized medium (WIM).

The heating and cooling of the ionized gas is controlled by photoionization, electrons with high kinetic energy, and radiation losses, which are dominated by forbidden lines of abundant elements like oxygen, e. g. the double ionized oxygen line [OIII]. A small amount of energy is carried away via Bremsstrahlung. For sufficiently low densities where $N\Lambda(T) < \Gamma(T)$ the gas is effectively heated. This is typically the case till the gas reaches a temperature of about $T \approx 8000\text{K}$. With increasing temperature the heating rate coefficient $\Gamma(T)$ becomes smaller and cooling becomes more efficient. Above temperatures of 8000K cooling processes take over leading to temperatures similar to the WNM with similar densities (see above).

1.1.3 Hot Phase

Components of the hot phase fills about 70 – 80% of the ISM volume and is by far the most dominant phase of the ISM. Furthermore, the material in this phase is permanently shocked. Shock fronts originating from different sources influence the ISM (McKee et al., 1977). Stars of OB associations are similar in age evolutionary stages. At the end of their lifetime, these types of stars have strong stellar winds and explode as supernovae (SNe), driving shocks into the ISM. While the volume affected by the strong winds is relatively small, the SNe driven shock fronts expand into a much larger fraction of the ISM ionizing the material with speeds $\geq 200\text{km s}^{-1}$. It is estimated that around two SNe per century occur in our galaxy, leading to $\sim 2 \times 10^4$ SNRs. Most of which are obscured and therefore not detected yet. The progenitor stars of these SNRs are locally grouped as well. Consequently, SNRs expand not only into the ISM but can merge, exchange energy and evolve into superbubbles (see section 1.2.6) maintaining the hot intercloud medium (HIM). Shocks expand until the temperature decreases and the pressure starts to balance with the surrounding medium. At this point the temperature reaches values of 10^6K and a number density around 10^{-2}cm^{-3} . Because of these low densities, two-body processes like collisions and recombination are extremely rare. Consequently, cooling via emission is very inefficient in the HIM and lasts for a comparably long time. Indeed, diffuse low-intensity X-ray emission was found along the galactic plane originating from highly ionized gas and is evidence for an existing hot phase of the ISM. This phase is critical for the evolution of the ISM and galaxies alike and leads to the co-existence of the three characteristic phases of the ISM.

Interstellar Shocks

Energetic events like stellar winds from, e. g., OB-stars or SNe, cause pressure perturbations which propagate with a certain velocity away from their origin and, therefore, expand spherically into the ISM. If the velocity is greater than the speed of sound in the medium, a discontinuity occurs between the regions that have already been affected by the perturbation and those that have not. This discontinuity is referred to as a shock wave (Longair, 2011).

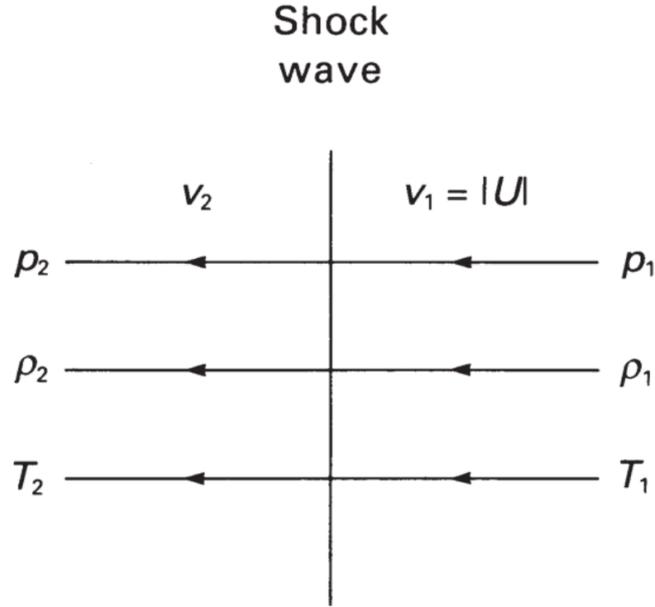


Figure 1.4: Illustration of a shock front propagating with velocity U through the gas in front of the shock (index 1). The medium behind the shock is denoted by the index 2. The pressure p , density ρ , temperature T and velocity is shown. In the rest frame of the shock, the gas in front of the shock moves with $v_1 = |U|$ towards the shock. The shocked gas moves with velocity v_2 away from the shock front. From Longair (2011).

The HIM is created and maintained by strong shock waves propagating through the ISM. Figure 1.4 shows a simplified shock situation in the rest frame of the shock. In this frame, the unshocked gas ahead of the shock moves with velocity v_1 towards the shock, has a pressure p_1 , density ρ_1 and a temperature of T_1 . Behind the shock the shocked gas has velocity v_2 , pressure p_2 , density ρ_2 and temperature T_2 . In the literature the compressible gas is considered as a fluid. While the gas in the unshocked region passes the shock, mass is conserved. Therefore, the mass flux before and after the shock is balanced according to

$$\rho_1 v_1 = \rho_2 v_2. \quad (4)$$

In standard fluid dynamics, the energy flux through a surface with normal vector \mathbf{n} and $\mathbf{v} \parallel \mathbf{n}$ is given by $\rho \mathbf{v} (1/2v^2 + w)$, with w being the enthalpy per unit mass. Considering a plane shock wave, perpendicular to \mathbf{v} , a continuous energy flux and conservation of energy

$$\rho_1 v_1 \left(\frac{1}{2} v_1^2 + w_1 \right) = \rho_2 v_2 \left(\frac{1}{2} v_2^2 + w_2 \right) \quad (5)$$

follows. In the shock scenario described here, the momentum flux $p + \rho v^2$ is considered as continuous and is balanced before and after the shock as follows

$$p_1 + \rho_1 v_1^2 = p_2 + \rho_2 v_2^2. \quad (6)$$

These three conditions, resulting from conservation of mass, energy and momentum, are referred to as shock conditions (Longair, 2011).

To see how the shock influences the ISM, the shock wave is considered in an ideal gas, where the enthalpy can be calculated as

$$w = \frac{\gamma p V}{(\gamma - 1)} \quad (7)$$

where V is the specific volume, $\gamma = 1 + 2/n_f$ the specific heat capacity of the gas components (with n_f the degrees of freedom). In Longair (2011) the pressure, density and temperature ratio between the shocked and unshocked region are derived to be

$$\frac{p_2}{p_1} = \frac{2\gamma M_1^2 - (\gamma - 1)}{\gamma + 1}, \quad (8)$$

$$\frac{\rho_2}{\rho_1} = \frac{v_1}{v_2} = \frac{\gamma + 1}{(\gamma - 1) + \frac{2}{M_1^2}} \quad \text{and} \quad (9)$$

$$\frac{T_2}{T_1} = \frac{(2\gamma M_1^2 - (\gamma - 1)) (2 + (\gamma - 1) M_1^2)}{(\gamma + 1)^2 M_1^2} \quad (10)$$

where $M_1 = v_1/c_1$ is the Mach number and $c_1 = \sqrt{\gamma p_1/\rho_1}$ is the speed of sound in the unshocked region. As previously stated, the HIM is created by strong shocks, which have shock velocities $v_1 \gg c_1$ and, therefore, $M_1 \gg 1$. In this case equation 9 can be written as

$$\frac{\rho_2}{\rho_1} \sim \frac{\gamma + 1}{\gamma - 1}. \quad (11)$$

Consequently, the density ratio is constant. For a monoatomic gas, like hydrogen, where $n_f = 3$ and $\gamma = 5/3$, the density behind the shock is limited to $\rho_2 = 4\rho_1$. In the considered scenario the pressure ratio

$$\frac{p_2}{p_1} \propto M_1^2 \quad (12)$$

and temperature ratio

$$\frac{T_2}{T_1} \propto M_1^2 \quad (13)$$

still depend on the shock velocity via M_1^2 . Theoretically, the gas behind the shock can reach arbitrarily large temperatures (Longair, 2011).

1.2 BUBBLE-LIKE STRUCTURES

Bubble-like structures in the ISM originate from massive stars and are formed by HII regions (section 1.2.1), stellar winds (section 1.2.2), planetary nebulae (PNe) (section 1.2.3) or SNRs (section 1.2.4). These structures emit photons with energies across the electromagnetic spectrum which can be observed and used to identify the type of structure that emitted them. The types of emission coming from those bubble-like structures are described in section 1.2.5. Most models of bubbles in the ISM, e. g., Weaver et al. (1977), are based on some simplifying assumptions, for example that the environment is homogeneous. The reality, however, is quite the opposite. Section 1.2.6 gives a brief overview of a more realistic ISM.

1.2.1 HII Regions

Sources of ionizing photons are embedded in gas clouds called HII regions which are often fully ionized. One example of a photoionizing source is a massive star of spectral type O. There are about 10^4 young O-star clusters in our galaxy and, therefore, at least the same number of HII regions in the ISM.

Since the most abundant element in the ISM is hydrogen, the photoionizing photons produced by an O-type star need at least an energy of $E_{\text{ph}} \geq 13.6\text{eV}$ to fuel the ionization process given by equation 3. Consequently, the number density in the ionized gas will increase by a factor of two, since the ionization of one hydrogen atom results in a proton and electron. The pressure and temperature of the ionized gas increase drastically and the HII region expands supersonically into the neutral medium forming a strong shock and an ionization front. In contrast to the photoionization, electrons can be captured by protons recombining to any excited state of hydrogen while emitting a photon. If the excited hydrogen atom undergoes a cascade of transitions to lower states, photons are emitted for every transition until the hydrogen atom is in its ground state. Therefore, HII regions can be observed mainly via radiative recombination. One dominant transition is the H α emission line which originates from a de-excitation from state $n = 3$ to $n = 2$. Since recombination is inversely propor-

tional to the electron temperature, recombination is a slow process in hot gas (Tielens, 2005). In the case that the degree of ionization is

$$x = \frac{n_p}{n_p + n_H} \sim 1, \quad (14)$$

i. e., the gas is almost fully ionized, the radius of the HII region would be the Strömngren radius

$$R_S = \left(\frac{3\dot{N}_{ph}}{4\pi n_H^2 \beta} \right)^{\frac{1}{3}}. \quad (15)$$

Here n_p is the number density of protons, n_H the number density of hydrogen, \dot{N}_{ph} is the rate of photoionizing photons and β is the recombination rate (Tielens, 2005).

If the ionization rate decreases below the rate of hydrogen atoms crossing the ionization front, the ionization front separates from the shock. The shock propagates further but without ionizing the neutral medium. Consequently, the material which is swept up by the shock forms a dense shell of neutral gas. With larger radii the density decreases and eventually reaches a similar density as the surrounding ISM. At this point the expansion of the HII regions stops.

1.2.2 *Stellar Bubbles*

Hot, massive stars have strong stellar winds that blow stellar bubbles into the ISM by constantly transporting mass and momentum. The photons, produced by the star, are scattered several times within the atmosphere of the star, transferring momentum to the surrounding gas. This momentum transfer drives winds with up to 3000km s^{-1} (Tielens, 2005) into the ISM. The stellar wind expands freely until it starts to interact with the material surrounding the star. This free expansion phase is indicated by R_1 enclosing the area (a) in Figure 1.5. The pressure of the stellar wind decreases with the radius. At the point where the pressure is lower than the pressure of the ambient gas (Figure 1.5, area d), a shock wave is created and propagates into the interstellar gas, sweeping up material. A thin dense shell at R_2 is formed. A reverse shock heats up the gas located within R_c to high temperatures. The heated gas has a lower density and cannot cool down like the dense shell. Therefore, R_c divides the two shocked regions. The resulting bubble consists of a central hot gas and a dense but cooler shell.

1.2.3 *Planetary Nebulae*

Stars produce elements throughout their evolution. Starting from a main sequence star, characterized by hydrogen burning (nuclear fusion of hydrogen atoms), the star produce helium in its core. If there

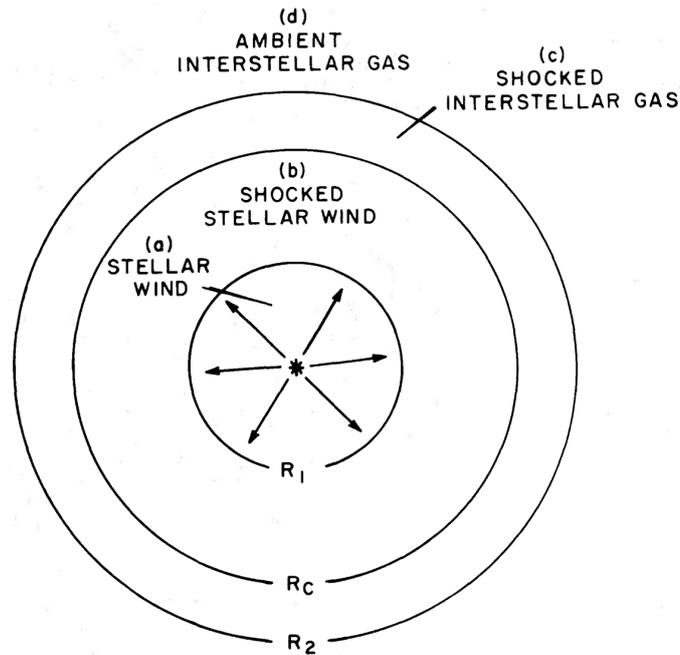


Figure 1.5: A schematic structure of a stellar bubble with its different regions. The stellar wind (a) expands freely till R_1 . The stellar wind starts to interact with the ISM. At a distance R_c from the star, a reverse shock is formed heating up the gas inside (b). A shock wave propagates outwards sweeping up material (c) of the ambient gas and, thus, forming a thin dense shell at R_2 . From Weaver et al. (1977).

is no hydrogen left to fuel hydrogen burning, the star starts burning helium. The following evolution of a star depends heavily on its mass. Stars with a mass less than $3M_{\odot}$ cannot reach core temperatures high enough to reach the next stage of burning, namely carbon burning. Consequently, the core of the star is enriched with carbon. Hydrogen and helium burning continues in shells around the core. At the end of the stars live it expels the hydrogen and helium layers which expands into the ISM (Karttunen et al., 2003). A white dwarf is formed in the center, and it is ionizing the expelled material due to its UV-radiation. The ejected helium and hydrogen shells create a so-called planetary nebula. Figure 1.6 shows the Helix Nebula, a well known planetary nebula with its white dwarf star clearly visible in the center.

1.2.4 *Supernova Remnants*

Supernova remnants (SNRs) originate from stellar explosions, and give rise to the progenitor star and explosion mechanisms. One distinguish between two explosion mechanisms: The core collapse (CC) and thermonuclear supernova (Karttunen et al., 2003). A more detailed classification is based on optical spectra and so-called light curves



Figure 1.6: The Helix Nebula. The white dwarf (small white dot in the center) is surrounded by the ejected material. The ejecta is ionized by the white dwarf and is glowing therefore. Credit: NASA, NOAO, ESA, the Hubble Helix Nebula Team, M. Meixner (STScI), and T.A. Rector (NRAO)

(Vink, 2020). Figure 1.7 shows the classification scheme for SNe. Depending on the hydrogen abundances SNe are divided by Type I and Type II supernovae. Each type has sub-categories as seen in Figure 1.7. The explosion mechanisms are marked with a dashed blue and red box which represent thermonuclear and core collapse SNe, respectively. Note that a thermonuclear SN is exclusively classified as Type Ia while there are many sub-categories of CC supernovae.

Type Ia supernovae are based on the same systems of close binary stars and therefore have the same or at least very similar explosion mechanisms. These systems are a topic of themselves and are not discussed in detail. In a common case, however, one of the stars is a white dwarf (a product of a dying low mass star) reaching the Chandraseka mass, because of mass transfer from its companion, and ends up in a thermonuclear SN. The thermonuclear explosion of the white dwarf sets in if nuclear reactions are triggered within the carbon-oxygen core (Vink, 2012). This trigger mechanism is not yet understood properly and therefore still part of discussion.

Core collapse supernovae have typically progenitor stars with masses of $M > 8M_{\odot}$. Those stars can produce heavier elements up to iron in their cores via nuclear fusion (nuclear burning). Before the explosion, the star consist of different layers, starting with hydrogen and helium at the outer layers to iron in the central region of the star.

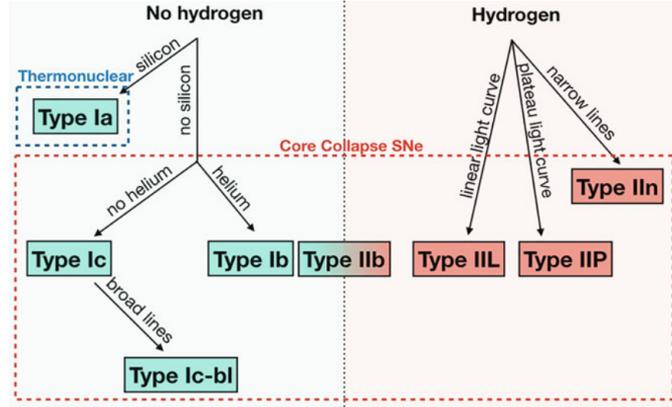


Figure 1.7: The supernova classification scheme from Vink (2020).

Layers between the outer and inner layers consist of layers filled with products from burning processes of elements, for example, carbon and oxygen. The burning process stops with iron in the core, since nuclear fusion of iron does not lead to an energy gain, and the core collapses. As a result either a neutron star or a black hole can be formed depending on the progenitor mass (Vink, 2012). Depending on the initial mass and therefore on the evolution of the massive star, core collapse SNe can have a variety of characteristics leading to the sub-categories shown in Figure 1.7. In case of a fast rotating neutron star as a stellar remnant, a so-called pulsar, a pulsar wind nebula (non-thermal plasma embedded inside the SNR) can be formed. In this case the remnant is called a composite SNR (Figure 1.8, b).

In general, SNe inject a large amount of energy $E_{\text{SN}} \sim 10^{51}$ ergs and mass M_{ej} into the ISM in a short time. They enrich their environment with elements produced inside the progenitor star during its evolution or during the explosion itself. The resulting SNRs expand into the ISM and evolve in three different phases. Assuming a homogeneous atomic gas surrounding the progenitor star, the three phases can be characterized as follows.

In the first phase, the remnant expands freely into the ISM because of the overwhelming mass ejected with high energy (Figure 1.8, a). The expanding remnant is heating and sweeping up surrounded material. Therefore, this phase depends on the initial conditions of the SN explosion and the ISM. As soon as the mass of the swept up material is $M_{\text{sw}} \sim M_{\text{ej}}$, the free expansion stops after a relatively short time and a thin shell of $\Delta R = R_s/12$ has formed. At this point of evolution, a reverse shock (Figure 1.8, a) propagated inwards heating up the gas even further to temperatures of $T \sim 10^6$ K (Tielens, 2005). The gas is almost fully ionized and, therefore, cooling is inefficient.

In the second, so-called adiabatic or Sedov-Taylor phase of the SNR, the pressure of the hot gas is high, driving the forward shock (Figure 1.8, a) into the ISM. Because more and more material is swept up, and no additional energy is injected, the expansion speed $v \propto t^{-3/5}$

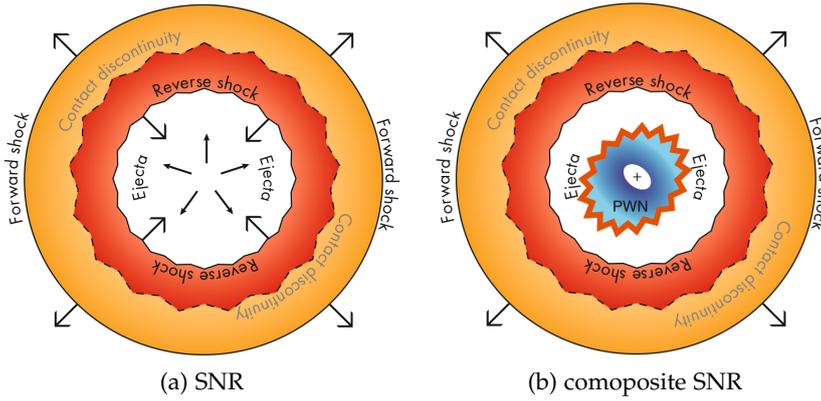


Figure 1.8: (a) An illustration of the SNR with the forward and reverse shock heating up the gas (temperature is color coded). The initial ejected mass is indicated by the arrows in the center. (b) Structure of a composite SNR with a pulsar (cross) and the pulsar wind nebula (PWN, blue) embedded in the SNR. From Vink (2020).

is slowing down with time. This phase can typically last for about 10^4 yr to 10^5 yr and the remnants can reach radii of some ten of parsecs depending on the initial SN parameter (Tielens, 2005; Williams, 2021). If v reaches values of around 200 km s^{-1} and the corresponding temperature falls below $T = 10^6 \text{ K}$ radiative cooling processes become more relevant introducing the third phase.

In the radiative phase, recombination becomes possible radiating away energy and cooling the gas. Assuming a constant density of the not yet affected medium, the remnant slows down further with $v \propto t^{-3/4}$ because of piling up more material at the edge, and is therefore called snowplow phase. The expansion continues due to the higher pressure of the shocked and heated material inside the SNR. At the end, the velocity of the remnant is similar to the thermal velocity of the WNM. As a result, the remnant merge into the ISM. Figure 1.9 shows the three different phases of the SNR by plotting the radius over the time.

The expansion of the remnant into the ISM can lead to complex morphologies, depending on the density of the ISM. In case of a homogeneous ISM the morphology of the SNR is typically shell-like. An inhomogeneous ISM is reflected in a more asymmetric shape. A more specific remnant is the mixed-morphology (or thermal-composite) SNR which has a bright shell-like structure in radio but show center filled X-ray emission. In most cases, mixed-morphology remnants are found in a more inhomogeneous ISM (Ustamujic et al., 2021a) and interact with dense molecular clouds in the vicinity of the SN explosion (Okon et al., 2020). Those clouds survived the forward shock but evaporate inside the hot after-shock medium (White et al., 1991).

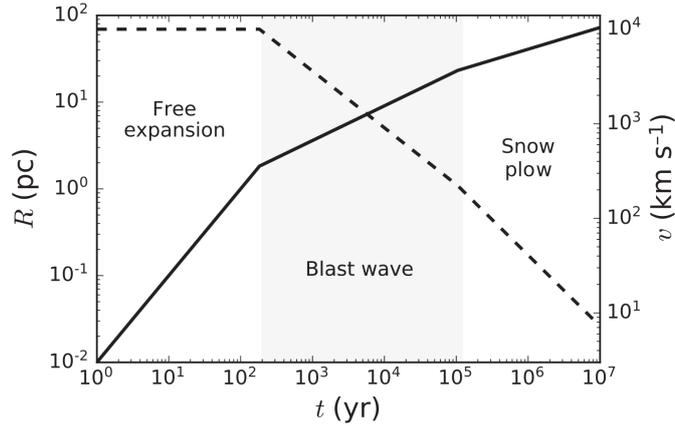


Figure 1.9: This plot shows an example of an SNR evolution. The solid line represents the expansion (radius) of the remnant and the dashed line the velocity with which the remnant expands into the ISM. The light gray region marks the time of the Sedov-Taylor phase. From Williams (2021)

1.2.5 Emission from bubble-like Structures

From lower energy photons, e. g., from dust in the infrared regime, to high energetic X-ray photons, the bubble-like structures described above emit photons in a wide range of the electromagnetic spectrum. The reason for that is the variety of emission processes which typically occur in these structures under different physical conditions like different temperatures.

An important emission process is called Bremsstrahlung or free-free emission. Bremsstrahlung occurs if a charged particle, e. g. a light particle (an electron), is accelerated in the Coulomb field of another charged particle (an ion). Depending on the particle population, the emitted photon energies can range from the infrared to the X-ray regime. Bremsstrahlung is an important emission process in the ionized gas of HII regions. It produces a continuum emission from radio to infrared wavelengths. The heated gas in SNRs, in contrast, can have much higher temperatures and can produce photons in the X-ray regime (Vink, 2020).

Relativistic charged particles produce synchrotron radiation, if they are deflected in strong magnetic fields. Such strong magnetic fields can be found in SNRs or pulsar wind nebulae. Electrons are effectively radiating synchrotron photons due to their low mass. That is why they are responsible for most synchrotron radiation in the aforementioned bubble-like structures. Synchrotron emission can be detected in the radio band but also in X-ray or even γ -ray bands. Structures which emit high energy synchrotron photons need to actively accelerate particles (Vink, 2020).

Another radiation process can be found in each region with a population of electrons and a photon field present. In those regions a

photon can collide with an electron. This process is known as Compton scattering in which both particles can either gain or lose energy. If the photon gains energy, i. e. , it is up-scattered, the process is referred to as inverse Compton scattering (ICS). This process is a common radiation mechanism in SNRs and pulsar wind nebulae (PWNe) to produce γ -ray emission (Vink, 2020).

In contrast to the processes described above, which produce continuum emission, another type of observable radiation leads to distinct photon energies called emission lines. Those emission lines originate from discrete atomic transitions.

There are recombination lines produced by electron capture of an ion. In these processes, electrons of discrete energies can be captured, resulting in multiple possible energy levels and, therefore, a variety of radiative transitions to lower energy levels can occur, e. g. , to the ground state is reached. This cascade of radiative transitions produces photons of different and discrete energies. These transitions follow a nomenclature which indicate the final state $n \in \mathbb{N}$, also labeled with capital letters K ($n = 1$), L ($n = 2$), M ($n = 3$) etc., and the energy of the emitted photon, or the number of levels Δn skipped to reach n . The first two spectral series are the Lyman ($n = 1$) and the Balmer ($n = 2$) series. The skipped levels are denoted by Greek letters α ($\Delta n = 1$), β ($\Delta n = 2$), γ ($\Delta n = 3$) and so on. A prominent example is the H α emission line in HII regions (H stands for the Balmer series).

Collisional lines can also be observed if an excited atom is de-excited by collision. In the environment of bubble-like structures the kinetic energy of the gas is not sufficient to excite hydrogen or helium, the two most abundant elements. However, it is sufficient to excite heavier elements. The de-excitation via collision depends on the number density and the temperature of the gas.

In the hot gas of young SNRs, for example, a high fraction of alpha-elements like oxygen (O) or sulfur (S) are in a hydrogen-like or helium-like ionization states. That means corresponding line emission involves the K-level of the atom. Line emission from those elements emit photons in the 0.5keV – 10keV range (Vink, 2020) and lie therefore in the X-ray regime. Emission lines from OVII and OVIII, oxygen which is six- and seven-times ionized, respectively, have energies around 0.6keV (Winkler et al., 1981). Elements from the iron-group can have emission lines between 0.7keV – 1.5keV or even higher energies between 6keV and 8keV (Vink, 2020). Prominent iron lines around 1keV, the Fe-L complex, are typically observed for Type Ia SNe and is therefore be used to discriminate between the two general types of SNe explosions (Vink, 2020).

In bubble-like structures with ionized gas, but lower temperatures, optical line emission can be produced by elements like oxygen and sulfur with lower ionization states. This is the case in HII regions or in the shell of older SNRs. In environments of low density, optical emis-

sion from forbidden lines like the [OIII] (two-times ionized) or/and [SII] (singly ionized) can be observed. The corresponding emission line positions are at [OIII] $\lambda\lambda$ 495.9nm, 500.7nm and [SII] $\lambda\lambda$ 671.7nm, 673.1nm (Long, 2017).

According to observations, the [SII] lines observed in shells of SNRs are about the same strength as the $H\alpha$ lines. In HII regions, however, [SII] line emission is an order of magnitude less prominent. Mathewson et al. (1973) used the ratio $[SII]/H\alpha$ to distinguish between SNRs and HII regions for the first time. For isolated SNRs the ratio is $[SII]/H\alpha > 0.4$. SNRs embedded in HII regions can be contaminated by photoionizing photons, therefore, the ratio of $[SII]/H\alpha > 0.67$ is used (Fesen et al., 1985; Kavanagh et al., 2016; Mathewson et al., 1973).

1.2.6 Bubbles and the inhomogeneous ISM

The bubble-like structures discussed in section 1.2.1, 1.2.2, 1.2.3 and 1.2.4 are considered as single structures in a homogeneous and, therefore, simplified ISM. However, in reality, the ISM is highly inhomogeneous and influenced by all energetic events, especially SNe and SNRs shape the ISM. A detailed analysis of the ISM is far beyond the scope of this thesis, but the complexity that current research is investigating is worth noting. Figure 1.10 shows a more detailed, edge-on view of a galaxy with the different phases of the ISM. The CNM and GMCs are spread within the galactic plane while the WNM and WIM extends further out and are surrounded by the HIM.

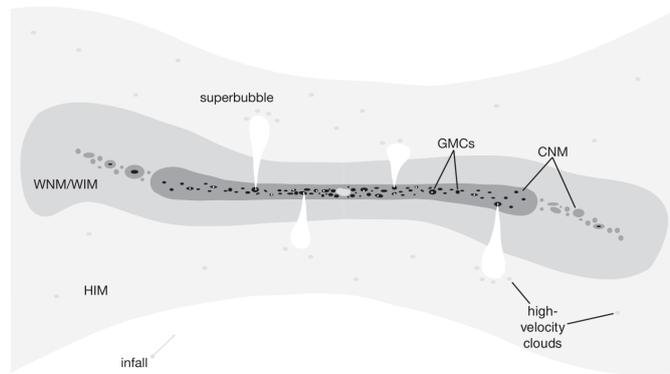


Figure 1.10: A schematic image of the fractional ISM within the host galaxy. Molecular clouds and cold gas can be found near the galactic plane, surrounded by the WNM and WIM. Superbubbles are blown into the HIM, which covers a large fraction of the galaxy. From Williams (2021)

Tielens (2005) analyzed the effect of small dense clouds embedded in the expanding SNR. Evaporation of these clouds leads to mass, energy and momentum transfer and are affecting the evolution of the SNR by slowing down the remnant locally. In general, the biggest impact on the ISM comes from the most rare stars. OB-stars create HII re-

gions and blow strong stellar winds into the ISM. Moreover, OB-stars are usually formed in a group and create SN events. Consequently, SNR can be embedded in HII regions, overlapping with each other and transferring energy. A cluster of OB-stars can, therefore, produce even larger bubble-like structure called superbubbles (Figure 1.10), hot bubbles reaching hundreds of parsecs in size by breaking through regions of lower pressure.

This emphasizes the need of detecting each bubble-like structure in the ISM to grasp a bigger picture and a better understanding of the evolution of the host galaxy.

1.3 THE MAGELLANIC CLOUDS

Among all satellite galaxies, orbiting our host galaxy, the Magellanic Clouds (MCs) are the largest irregular dwarf galaxies. The Large Magellanic Cloud (LMC) is closer at a distance of 50kpc, compared to the Small Magellanic Cloud (SMC) at around 62kpc (Graczyk et al., 2014; van den Bergh, 1999). Thanks to their location on the sky, above the galactic plane, the absorption along the line of sight is minimal. Moreover, the LMC and the SMC can be observed face-on revealing details of their structures (Figure 1.11). These observation properties make the MCs perfect laboratories to study the ISM.

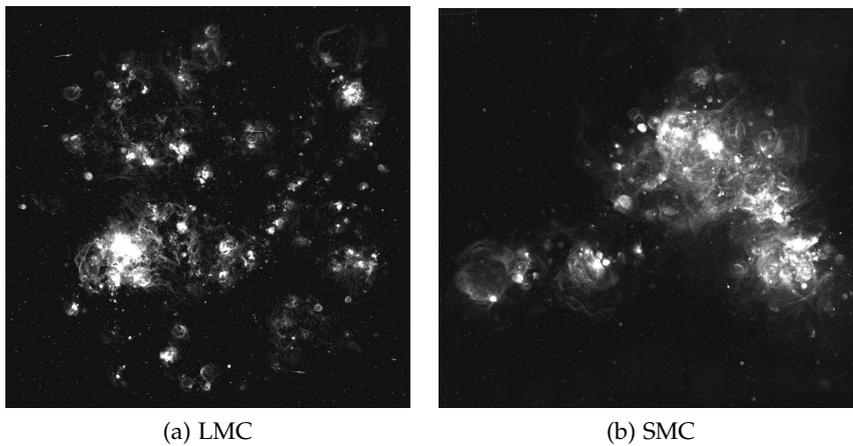


Figure 1.11: The Large Magellanic Cloud (a) and the Small Magellanic Cloud (b) in H_{α} line emission. The structures of different size and brightness are visible in both galaxies Smith C. (2000)

H_{α} emission line surveys show that the MCs host many HII regions and bubble-like structures of different size (Davies et al., 1976; Henize, 1956). Therefore, the MCs are the subject of several studies investigating the properties and the origin of these bubbles, for example in Bozzetto et al. (2023), Kavanagh et al. (2016), Maggi et al. (2019), and Yew et al. (2021). According to observations, star formation is still ongoing in the MCs leading to possible SN explosions. Bozzetto et al.

(2017) estimated a rate of SNe of around $1/\text{yr}$ in the LMC. This is evidence for the presence of both young and older SNRs. Indeed, several recent publications have confirmed the presence of a variety of SNRs, for instance Bozzetto et al. (2023) and Yew et al. (2021) or Zangrandi et al. (2023, in prep.) to name a few. With better telescopes and analysis techniques, faint structures are found. This includes SNRs at the end of their evolution merging into a thin environment, making them difficult to detect (Yew et al., 2021). To date, the MCs are one of the best sites for astronomers to study the ISM.

With every single existing and new telescope around, performing single observations or entire surveys, more and more data are available in astronomy. Therefore, data are growing in size and complexity. As a result, algorithmic based detection methods will become important in the future. Collischon et al. (2021) developed an automated detection algorithm based on Minkowski tensors, to detect and narrow down the origin of bubbles in the MCs. Meanwhile, machine learning algorithms get widely used in astronomy as well including bubble-like structures, e. g., in Liu et al. (2020). Detecting bubble-like structures with machine learning methods is a hard task, because they are of irregular shape.

In this chapter an intuitive approach of understanding neural networks, the general structure and the interplay of its different components is presented. This includes the mathematical formalism to describe all crucial parts of a neural network, starting from a single neuron, its activation and how neurons are connected with each other. Furthermore, the concepts of forward- and back propagation are introduced, together with training procedures for neural networks and methods to improve the training. A motivation, why to use neural networks in this work is given at the end of the chapter.

2.1 FUNDAMENTAL COMPONENTS AND CONCEPTS

This section mainly follows the description and notation given in Michelucci (2018), unless otherwise stated.

2.1.1 About Neurons, Weights and Layers

Every neural network consists of three different types of layers L_l (Figure 2.1), where $l \in \mathbb{N}_0$ denotes the number of the corresponding layer. The input layer L_0 can be considered as the receptor, the eye of the neural network, receiving information. This information is processed in a certain amount of hidden layers L_l , with $0 < l < \Omega$ and $\Omega \in \mathbb{N}$, working on a generalized representation of the given input information. Last but not least, the output layer L_Ω process the pattern provided by the last hidden layer $L_{\Omega-1}$ to give the final result.

Each layer consists of several neurons N_n^l , with $n \in \mathbb{N}_0$ the number of the neurons in layer L_l (Figure 2.1). Neurons N_n^l of each layer L_l are connected with neurons N_n^{l-1} and N_n^{l+1} of the layer L_{l-1} and L_{l+1} , respectively. These connections are called weights $w \in \mathbb{R}$. The

neurons in the first hidden layer L_1 therefore are connected to the neurons of the input layer L_0 and the neurons of the last hidden layer $L_{\Omega-1}$ are connected to neurons of the output layer L_{Ω} . Consequently, a net of neurons is created where neurons depend on the behavior of the previous ones.

At this stage of understanding, neurons can be considered as a storage of value a , where a is called the activation of the neuron. Imagine a neuron fires if it is needed to represent the input information in the generalized representation. The larger a the more important is the neuron for processing the information.

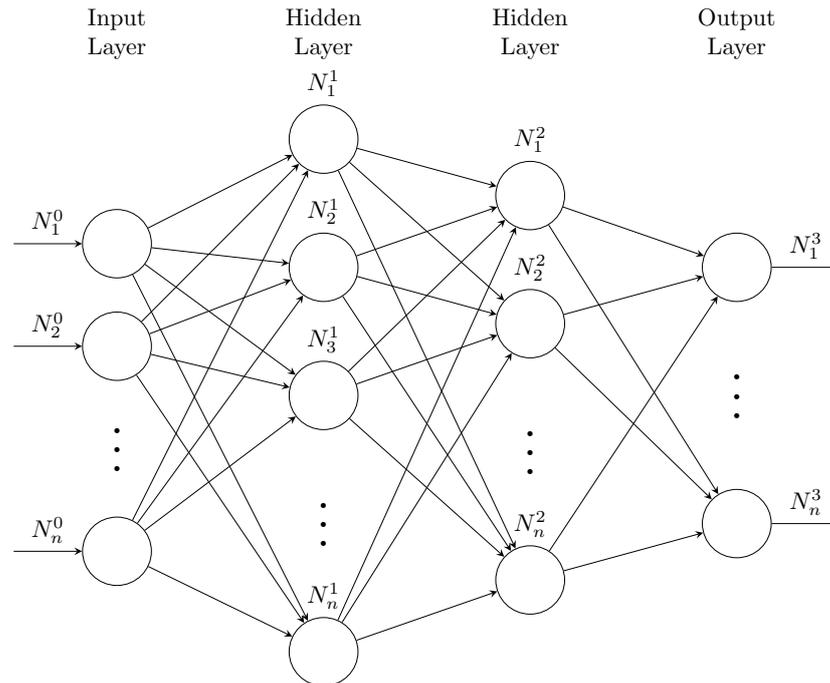


Figure 2.1: A scheme of a neural network. N_n^l denotes the n -th neuron in the layer l . Its basic components are the input ($l = 0$) and output ($l = 3$) layer. Layers between the input and output layer are called hidden layers. The neurons (circles) are connected with each other, indicated by arrows (weights).

2.1.2 Information Processing

Every single information provided at the input layer L_0 passes the network and is processed such that the network can resolve a certain problem. Imagine a neural network which has to figure out the mean gray value of three input values, stored in the three input neurons (Figure 2.2). Depending on the gray value, the network must be able to represent the given information in a generalized way in order to be able to figure out the mean gray value at the output layer L_{Ω} . Consequently, the activation of neurons in the hidden layers have

to be specific for every combination of gray values provided at the input layer. This raises the question of how the network generates the generalized representation of the input information?

The answer lies in the weights w connecting neurons of each layer with each other. These weights can be adjusted, meaning weaken or strengthen the connection between neurons. The combination of weights and activation of neurons lead to a dynamic behavior of the whole network, which is therefore able to react to the given input information accordingly.

But why using several hidden layers for a neural network and connect these layers with each other. With each additional layer, new neurons and weights are added to the neural network. Imagine one hidden layer is responsible for a set of patterns and every neuron is representing one of these patterns. In the case of determine the mean

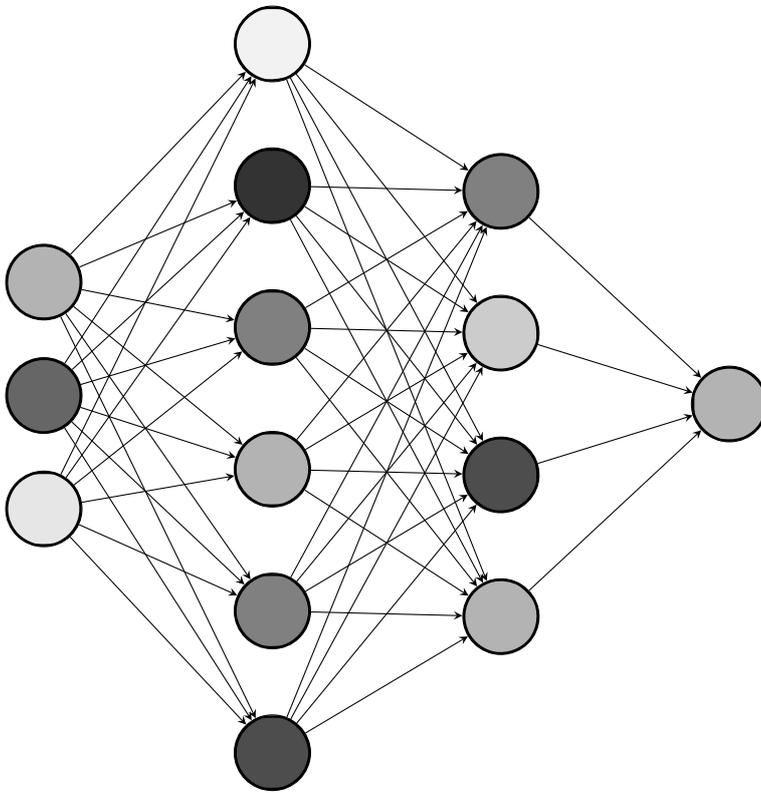


Figure 2.2: Illustration of a neural network and its activated neurons. Each neuron represents a gray value. To predict the mean gray value of the three input neurons on the left, the network finds a generalized representation to predicted mean gray value in the output layer on the right.

gray value in Figure 2.2, these patterns would be different gray values. Because of more possible representations of the gray values at the input layer, the network would be able to map more gray values to the output layer. More layers and neurons results therefore in the ability of the network to solve more complex problems, due to a more versa-

tile neural network. A neural network with more than three hidden layers is usually considered as a deep neural network (DNN).

In general, smaller neural networks might have a few hundred to a few thousand neurons, while larger neural networks can have tens of thousands or even millions of neurons. It is worth mentioning, that having more neurons does not necessarily guarantee better performance of the neural network, because it heavily depends on the structure of the network.

In the end, the activation of neurons in the last hidden layer $L_{\Omega-1}$ has a certain pattern leading to fire a certain neuron in the output layer L_{Ω} . The output neuron in Figure 2.2 simply represents the mean gray value of the input. This entire information processing is called forward propagation. In some cases, the activation of the output neurons can be considered as a probability of how certain the network is in its answer.

Of course, if the network works as painted in a more simplistic picture above, may be doubted. If one analyze each layer and its parameters, one will find that the network has a non-intuitive behavior. This is because of the amount of parameter used in a neural network, which makes it hard to grasp a clue on what actually is going on. This approach provides at least an expectation of how a neural network could work. With this expectation, let's dive into a more formal way to describe neural networks.

2.1.3 Forward Propagation

As already mentioned, weights do not only connect neurons of different layers, they also influence the activation of the connected neurons and, thus, the information which is fed forward within the network. Let's consider the activation a_m^l of neuron N_m^l (Figure 2.3). That means one has to take into account every weight w_{mn} , connecting the considered neuron N_m^l with the neurons N_n^{l-1} . Remember that the activation is a measure of how important the corresponding neuron is for the generalized representation of the input information. By introducing the so-called bias $b_n^{l-1} \in \mathbb{R}$, a neuron can be more or less sensitive to the previous neuron. The connection between neuron m and all connected neurons n can therefore be written as the weighted sum

$$z_m^l = \sum_n w_{mn}^{l-1} \cdot a_n^{l-1} + b_m^{l-1}. \quad (16)$$

There is just a linear connection between neurons so far. Consequently, the network would not be able to learn non-linear and, therefore, more complex correlations. Applying a non-linear function is essential for neural networks to find a generalized representation and model the input data. These non-linear functions are called activation

functions $\sigma(x)$. Some widely used activation functions, their characteristics, advantages and disadvantages are discussed in section 2.1.4.

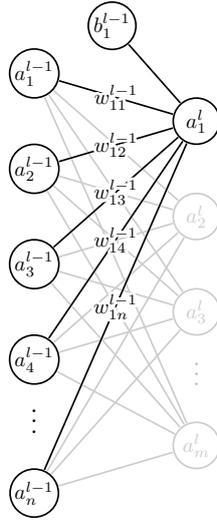


Figure 2.3: A look into a neural network, showing how the activation of a single neuron is calculated. All weights w_{mn}^{l-1} , activations a_n^{l-1} of every neuron of the previous layer and the bias has an influence on the activation a_m^l of the considered layer.

Finally, the activation of a single neuron in the corresponding layer can be calculated as

$$a_m^l = \sigma(z_m^l) = \sigma\left(\sum_n w_{mn}^{l-1} \cdot a_n^{l-1} + b_m^{l-1}\right). \quad (17)$$

In a neural network the activation of every single neuron within the considered layer has to be calculated. Therefore, the activation a^l for every neuron of layer L_l can be written in a more compact form

$$\mathbf{a}^l = \sigma\left(\begin{bmatrix} w_{00}^{l-1} & w_{01}^{l-1} & \dots & w_{0n}^{l-1} \\ w_{10}^{l-1} & w_{11}^{l-1} & \dots & w_{1n}^{l-1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m0}^{l-1} & w_{m1}^{l-1} & \dots & w_{mn}^{l-1} \end{bmatrix} \cdot \begin{bmatrix} a_0^{l-1} \\ a_1^{l-1} \\ \vdots \\ a_n^{l-1} \end{bmatrix} + \begin{bmatrix} b_0^{l-1} \\ b_1^{l-1} \\ \vdots \\ b_m^{l-1} \end{bmatrix}\right) \quad (18)$$

or for short

$$\mathbf{a}^l = \sigma\left(\mathbf{W}^{l-1} \cdot \mathbf{a}^{l-1} + \mathbf{b}^{l-1}\right). \quad (19)$$

Applying equation 19 to each layer is called forward propagation and describes how input data is fed forward through the network. To ensure that the output of the network maps the input data correctly, the parameters need to be adjusted such that the patterns in the input data are represented as good as possible by the network. A measurement is needed of how much off the output of the network is, to map it to the input data. This reverse process is called back propagation, but before diving into back propagation, let's have a look at some activation functions.

2.1.4 Activation Functions

Activation functions determine the activation of a neuron. They are used to introducing non-linearity, allow the network to perform more complex computations and model a wider range of input information. Common examples of activation functions are the sigmoid, rectified linear unit (ReLU) and softmax functions. The choice of the activation function depends on the specific task and architecture of the neural network. More about architectures later in this chapter and in section 5.1.

The ReLU activation function (Figure 2.4 (a)) is defined as

$$\sigma_{\text{ReLU}}(z_m^l) = \begin{cases} 0 & \text{if } z_m^l < 0 \\ z_m^l & \text{if } z_m^l \geq 0 \end{cases} \quad (20)$$

and is often used because its beneficial properties. One of the main advantages of the ReLU is that it is computationally efficient. Using ReLU activation functions allow deep neural networks with many neurons to learn effectively. Due to these main reasons, the ReLU activation function is commonly used in deep neural networks, in particular in the hidden layers.

The sigmoid activation function (Figure 2.4 (b)) is defined as

$$\sigma_{\text{sig}}(z_m^l) = \frac{1}{1 + e^{-z_m^l}}, \quad (21)$$

and maps any input z_m^l to $[0, 1]$. Therefore, the activation can be interpreted as a probability. It is often used in the output layer of a neural network and for binary classification tasks. The sigmoid function has a characteristic s-shape, and it is differentiable.

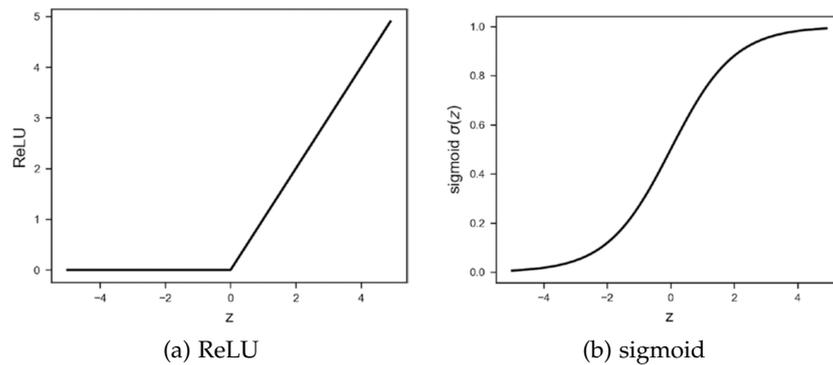


Figure 2.4: The plots show the ReLU (a) and sigmoid (b) activation function. ReLU activation functions are mainly used for hidden layers, whereas sigmoid activation functions are used in the output layer to interpret the output as a probabilistic value. Plots are taken from Michelucci (2018).

The softmax activation function is defined as

$$\sigma_{\text{soft}}(z_m^l) = \frac{e^{z_m^l}}{\sum_m e^{z_m^l}}. \quad (22)$$

Due to the normalization in the softmax function, the output is a probability distribution over all classes. Thus, it is frequently used in the output layer of a multi-class classification problem, where the neuron with the highest probability represents the predicted class.

2.1.5 Back Propagation and Optimization

Till now, its discussed how information passes through a neural network, assuming the network can handle the information given as input. But the weights and biases usually are initialized randomly. Therefore, it is unlikely that the network gives the desired output. The network needs to learn to find the right output for the given input information by finding proper values for the weights and biases. In other words, the neural network needs to model the given input data. In machine learning this is achieved by providing the desired output, so-called labels or classes, to the network. The goal is that the output of the network, its prediction, matches the given label. This is referred to as supervised learning.

Every prediction comes with an error, or loss. This loss can be considered as the discrepancy between the true label, provided to the network, and the label predicted by the network itself. According to the loss, the parameters, the weights and biases, are updated in order to reduce the loss of future predictions. This is achieved by propagating the loss backwards through the network. In literature this process is called back propagation (Goodfellow et al., 2016; Michelucci, 2018). To calculate the new weights and biases, optimization algorithms are implemented.

The learning process, or training process, is often referred to as an optimization problem. In machine learning the performance of a network is measured with some metrics, e.g. the loss. By reducing the loss the performance can be improved. That means the network is fitted to a set of data, called training data. Therefore, in the core of every training process, the so-called loss function L is optimized. Loss functions are discussed in section 2.1.6. The training process is, consequently, an indirect optimization problem (Goodfellow et al., 2016). In the machine learning space the training data is organized in batches. Batches are subsets of the data. An average loss is calculated over an entire batch which is then used to update the parameters.

Training a network combines back propagation with an optimization algorithm to define how to update parameters during training. Mathematically speaking, back propagation means applying the chain

rule to the loss function with respect to the parameters of the network. Consequently, the back propagation for each layer is calculated as

$$\frac{\partial L}{\partial \mathbf{W}^l} = \frac{\partial L}{\partial \mathbf{a}^l} \cdot \frac{\partial \mathbf{a}^l}{\partial \mathbf{z}^l} \cdot \frac{\partial \mathbf{z}^l}{\partial \mathbf{W}^l} \quad (23)$$

for the weights and

$$\frac{\partial L}{\partial \mathbf{b}^l} = \frac{\partial L}{\partial \mathbf{a}^l} \cdot \frac{\partial \mathbf{a}^l}{\partial \mathbf{z}^l} \cdot \frac{\partial \mathbf{z}^l}{\partial \mathbf{b}^l} \quad (24)$$

for the biases (Goodfellow et al., 2016; Kochenderfer et al., 2019).

One widely used optimization algorithm, especially for neural networks, is *Gradient Descent*. It is used to minimize the loss function by updating the weights and biases of the neural network. The gradient descent algorithm leverages the negative gradient calculated in back propagation to reach a local minimum (Ruder, 2016). The step size towards the negative gradient is defined by the learning rate α . The larger α the fewer iterations needed to reach the minimum. If α is too large, the gradient descent algorithm can diverge. Figure 2.5 shows the principle of gradient descent to minimize the loss function. The parameters of a neural network can be updated using gradient descent as follows

$$\mathbf{W}^l = \mathbf{W}^l - \alpha \cdot \frac{\partial L}{\partial \mathbf{W}^l} \quad (25)$$

and

$$\mathbf{b}^l = \mathbf{b}^l - \alpha \cdot \frac{\partial L}{\partial \mathbf{b}^l} \quad (26)$$

for the weights and biases. This is an iterative process of many iterations and, therefore, training the network means finding the best values of the parameters. The goal is to generate a network which is generalized properly. Generalization is the ability of the network to deal with input information never seen before (Goodfellow et al., 2016).

Gradient descent is the umbrella term of various types of optimization algorithms. One of these algorithms is the adaptive movement estimation algorithm (ADAM) introduced by Kingma et al. (2014). It improves efficiency by using an adaptive learning rate and momentum. As the name suggests, it changes the initial learning rate during the process. Momentum helps to accelerate the algorithm along the desired direction (Figure 2.6) of the negative gradient and dampens oscillations (Ruder, 2016). The computational memory usage is also better compared to other optimization algorithms (Kingma et al., 2014). According to Goodfellow et al. (2016), ADAM is considered as robust when it comes to the choice of hyperparameters. ADAM is used for train the neural networks for this work as well.

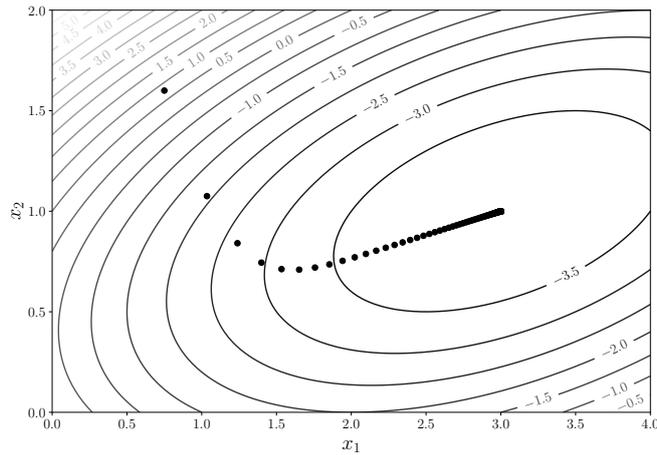


Figure 2.5: This plot shows the principle of gradient descent on a 2D surface in the x_1 - x_2 plane (contours). The black points show the evolution of the loss function after each iteration of the gradient descent. After each iteration the step size in the direction of the steepest descent gets smaller. The gradient descent algorithm converges after 181 iterations and finds a local minimum.

2.1.6 Loss Functions

A loss function is used to measure the loss between the predicted output of the neural network and the expected output. This loss is essential for training a neural network. During training, the goal is to minimize the value of the loss function (Goodfellow et al., 2016), which is typically achieved by adjusting the parameter of the neural network using an optimization algorithm like Gradient Descent. Depending on the task and type of data, different types of loss functions can be used. In this section the focus is on the cross entropy as a loss function.

Given a set of events or random values, cross entropy can measure the difference between the two underlying probability distributions. In machine learning the two distributions are given by the known probability, or label, for a sample belonging to a certain class y . In this work a one-hot encoded vector is used meaning the corresponding vector elements are $y_j \in \{0, 1\}$, and the probability of a sample belonging to a certain class predicted by the network p with $p_j \in [0, 1]$. Therefore, cross entropy is commonly used in deep learning, especially for binary- or multi-classification problems. In Murphy (2012) cross entropy is defined as

$$H(y, p) = - \sum_{j=0}^C y_j \cdot \log(q_j) \quad (27)$$

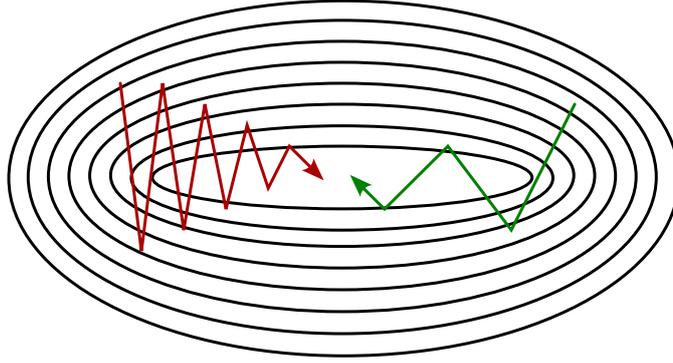


Figure 2.6: Illustration how momentum helps to accelerate the optimization algorithm to converge faster. Without momentum, more iterations are needed because of oscillations (red). Introducing momentum in gradient descent helps to implement bigger steps towards the steepest gradient descent (green) leading to fewer iterations needed. Adopted from Ruder (2016).

with C the number of classes. Considering a single sample with $C \in \{0, 1\}$, i. e., a binary-classification problem, the cross entropy is

$$H(y, p) = y_0 \cdot \log(p_0) + (1 - y_0) \cdot \log(1 - p_0) \quad (28)$$

with $(1 - y_0)$ the label for $C = 1$ and $(1 - p_0)$ the probability that the sample belongs to the class with $C = 1$. In the training process there are N samples in a batch. Taking all samples of the considered binary-classification problem into consideration and calculate the average loss over the entire batch, the so-called binary cross-entropy (BCE) is calculated as follows

$$\text{BCE} = -\frac{1}{N} \sum_{i=0}^N y_{i0} \cdot \log(p_{i0}) + (1 - y_{i0}) \cdot \log(1 - p_{i0}). \quad (29)$$

For a multi-classification problem the categorical cross-entropy (CCE) is calculated

$$\text{CCE} = -\frac{1}{N} \sum_{i=0}^N \sum_{j=0}^C y_{i,j} \cdot \log(p_{i,j}). \quad (30)$$

One disadvantage of the CCE loss function is that it is sensitive to an imbalanced data set.

Both are established loss functions and considered as easy to optimize with a probabilistic interpretation of the output.

2.1.7 Deep Neural Networks

Deep neural networks (DNNs) are neural networks with many hidden layers. These layers are composed of interconnected neurons. Each layer learns a different abstraction of hierarchical representations of

input data as described above. While the first layers may learn some representations of some features, e. g. edges, the deeper layers may learn more complex features, e. g. shapes. This is what makes DNNs powerful tools for various tasks. One of these tasks is image classification.

DNNs can suffer from different effects which reduce the performance of the network. In the following some important effects will be discussed.

Overfitting is a common effect appearing during the training process of a DNN. By learning the underlying features and noise of the training data set in great detail, the network is not able to generalize to data never seen before. Therefore, the performance of the network on the training data set differs significantly from the performance on a test data set. The test data set is a smaller data set organized like the training data set but is not used for training. In practice, a fraction of the training data set ($\sim 10\%$) is used as the test data set. Overfitting can therefore be spotted by evaluating each training step (iteration) on a test data set by plotting the loss for each iteration. Such a plot is shown in Figure 2.7. If overfitting occurs, the network improves on the training data, but the performance on the test data decreases (Goodfellow et al., 2016; Michelucci, 2018). Some common techniques to prevent overfitting include using regularization techniques such as dropout or early stopping of the training process. More about dropout and early stopping below. Since large DNNs trained on a relatively small training data sets tend to overfit, using a simpler model with fewer parameters can also reduce overfitting. Consequently, adding more data to the training data set can help to reduce overfitting large DNNs and vice versa (Goodfellow et al., 2016).

Underfitting, in comparison, refers to a network which is not able to model the training data nor is it able to generalize to new data. This scenario is also shown in Figure 2.7. This said, the complexity of the network and size of the training data set to train on need to be properly balanced in order to match the complexity of the network to the complexity of the considered problem.

A DNN consists of different layer types, each of which offers different functionalities to the network. Some layer types can reduce the risk of overfitting, for example, others can be trained to recognize certain features in the input data. In the following paragraphs, some layer types which are used in this work are introduced.

Convolutional layers are a special type of layers used in convolutional neural networks (CNNs). They are powerful tools for recognizing features present in a two-dimensional input image, e. g. , lines, edges or shapes. Convolutional layers apply filters, or so-called kernels \mathbf{K} , to an input image \mathbf{I} (Figure 2.8). In the context of CNNs the operation

$$\mathbf{F}_{i,j'} = (\mathbf{I} * \mathbf{K})_{i,j} = \sum_m \sum_n \mathbf{I}_{i+m,j+n} \mathbf{K}_{m,n} \quad (31)$$

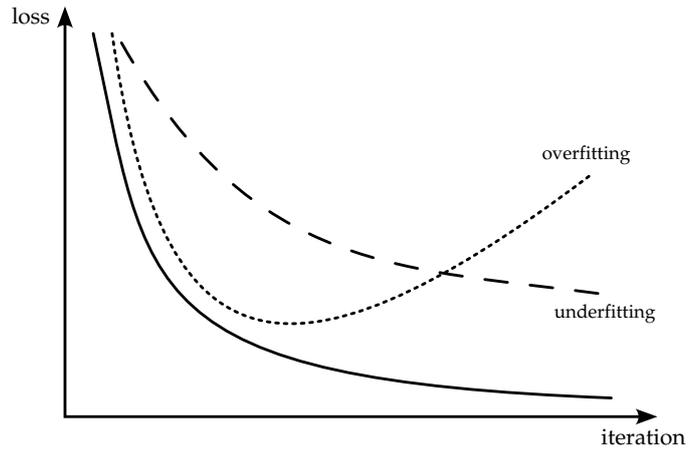


Figure 2.7: Comparing the loss of a training batch (solid line) and a test batch, over- and underfitting can be spotted. While a significant higher but steadily decreasing validation loss indicates that the training suffers from underfitting (dashed line), an increasing validation loss after reaching a local minimum hints overfitting (dotted line).

is called convolution. Where m and n denotes the size of the used kernel with $m, n < i, j$, the size of the input image. $F_{i,j}$ is the resulting feature map, which indicates the locations and the dominance of a certain feature in the given input image. By introducing a stride, the kernel slides over the entire input image and the convolution for each pixel is calculated. In order to maintain the dimension of the input image, the so-called *padding* must be applied to the convolution. Padding is a method in computer vision to extend the input image with zeros, such that $\dim(F_{i',j'}) = \dim(F_{i,j})$ (Figure 2.8).

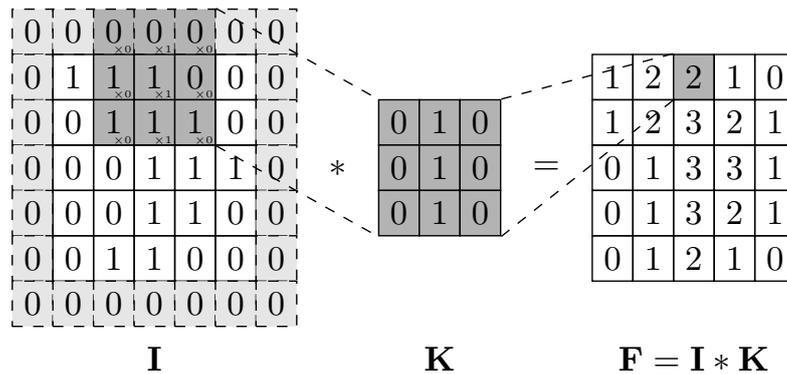


Figure 2.8: Illustration of convolve a 2D input image I with a kernel K . By sliding the kernel over the image and calculating the scalar product (dark gray) the feature map F is created. In order to keep the dimension of the input image, padding (light gray) is applied on I .

Generally, CNNs can be applied on every kind of data, able to define a grid on, e.g., time-series (Goodfellow et al., 2016). The power

of convolutional layers unfolds by the fact, that several kernels can be trained simultaneously. For every kernel, one feature map is generated. Consequently, several features can be detected in an image. The kernel in Figure 2.8 is used for vertical edge detection, for example. Those feature maps can be used as input for other layers. Different level of features can be detected by stacking convolutional layers. While the first convolutional layers extract more low-level features, stacked convolutional layers are on a higher level of abstraction and can extract more complex features. Increasing the depth of a CNN increases the complexity of the network by adding more parameters to train.

Pooling layers are commonly used in CNNs to apply down-sampling to the feature maps. As stated above, these feature maps are not translation invariant. To improve the translation invariance against small translations of the input, pooling is applied to the feature maps and creates a new set of lower resolution feature maps. These maps can then be used as a new input for the following convolutional layer again. How the down-sampling is affecting the input depends on the kernel size and stride of the pooling operation. There are two common pooling operations, the so-called *average* and *maximum* pooling. While the average pooling operation returns the average value

$$AvP = \frac{\sum_{ij} K_{ij}}{\sum_{ij} 1}, \quad (32)$$

maximum pooling returns the maximum value

$$MaP = \max_{ij} (K_{ij}) \quad (33)$$

of the given input within the kernel (Goodfellow et al., 2016; Zhou et al., 1988). Figure 2.9 demonstrate maximum pooling applying a (2×2) kernel and a stride of 2 on a (4×4) matrix. In this example, a kernel size of (2×2) and a stride equal to the kernel size means a down-sampling rate of two.

Dropout is the practice of pretending to train a large neural network of different architectures in parallel by dropping out neurons by cutting their connections for a training iteration temporarily (Srivastava et al., 2014). Figure 2.10 the principle of dropout on a fully connected neural network. Applying Dropout to a neural network can be considered as training an ensemble of networks with a different architecture (Ensemble Learning) but is computational cheaper. It is shown, that Ensemble Learning reduces the risk of overfitting (Goodfellow et al., 2016). Thus, dropout neurons during training is an effective way to tackle overfitting. It is also more effective than other standard regularization. Dropout can be beneficial, especially for problems with a small training data set (Goodfellow et al., 2016).

Normalization layers are transforming the input to be on similar scale. Some features in an image may be more dominant in an image,

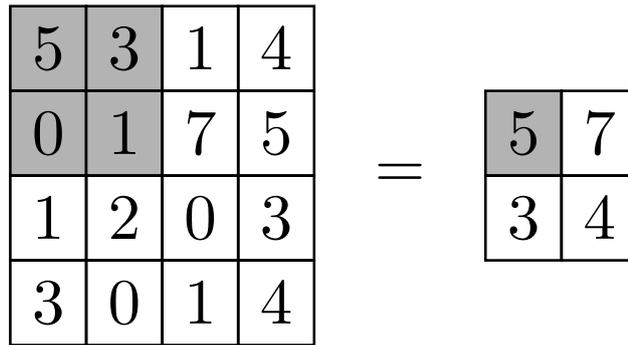


Figure 2.9: This Figure demonstrates the maximum pooling operation. A (2×2) kernel is applied on a (4×4) matrix. The kernel hovers over the matrix (gray, left) with a stride of 2. Maximum Pooling returns the maximum value (gray, right) within the kernel.

e. g., brighter compared to other features. In the machine learning space, it is common use to scale the data that input values are between $[0, 1]$ or $[-1, 1]$. With this approach different features in an image are on similar scale. In practice, normalizing data for training a neural network is not trivial, because input values of each layer within the network changes during training due to the change of parameters after each batch. This is referred to as internal covariate shift (Ioffe et al., 2015). Reducing internal covariate shift stabilizes the training of neural networks by achieving a stable distribution of activation values and speeds up the training (Ioffe et al., 2015). Therefore, Ioffe et al. (2015) introduced *batch normalization*, a technique for normalizing layer inputs for each training batch. It is used before applying non-linearity to increase the likelihood to get stable activation values during training (Ioffe et al., 2015). The batch normalization operation is denoted as

$$\text{BN}_{\gamma, \beta}(x_k) \equiv \gamma \cdot \hat{x}_k + \beta, \quad (34)$$

where γ and β are trainable parameter and \hat{x}_k the normalized input values over a mini-batch k . The details of batch normalization can be found in Ioffe et al. (2015). By implementing batch normalization as part of the model architecture, the benefits range from allowing higher learning rates, less need for careful parameter initialization, to regularization and no need for dropout (Ioffe et al., 2015). Therefore, natch normalization can increase the performance of the trained neural network.

Augmentation layers are processing layers which can help the network to generalize and reduce overfitting. By applying several augmentation transformations to samples within the training data set, it is enlarged artificially. Possible transformations range from translations, rotations, to vertical or horizontal flipping, scaling and reducing or enhancing the contrast and brightness of a sample. A list

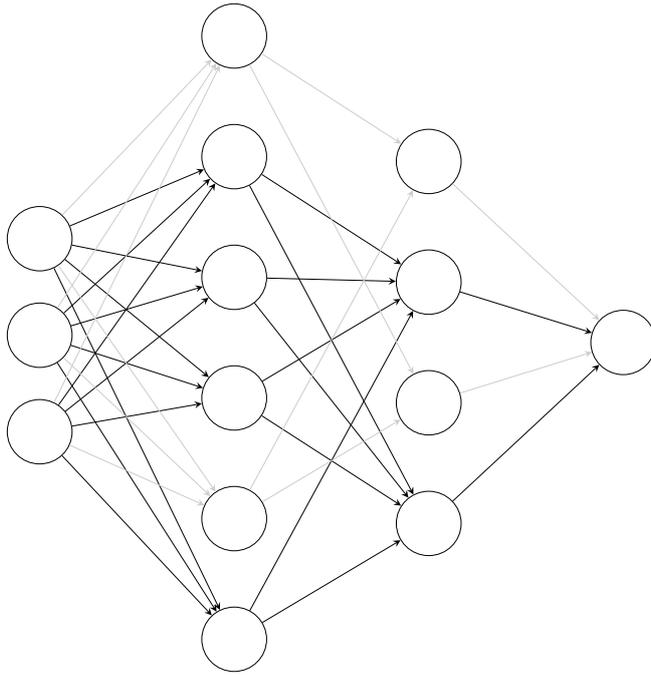


Figure 2.10: If Dropout is applied during training, neurons are disconnected temporarily by removing connections (light gray) to other neurons of the network.

of basic and more advanced augmentation operations can be found in Yang et al. (2022). Figure 2.11 shows four examples of augmentation operations. The choice of which transformation is used in the augmentation depends on the task and data set. For example, using horizontal flipping of a chess board, would result in an unnatural upside down image (Figure 2.11 (d)), while vertical flipping would change the perspective of the image (Figure 2.11 (c)) and add some value to the training data set. Augmentation cannot transform a poor training data set into a rich and diverse data set. It is but a technique to increase diversity within a training data set (Shorten et al., 2019). There is no quantitative standard of how to apply augmentation properly, it is mainly based on experience and experimenting with different augmentation operations (Yang et al., 2022).

Early stopping is a simple technique to prevent overfitting by setting up a schedule which terminates the training process if a certain criterion is fulfilled. Ending the training if the loss increases more than an upper limit, after reaching a minimum, would be such a schedule. However, the calculated loss fluctuates during training. Finding a balanced upper limit is crucial, to not stop training because of those loss fluctuations while the tendency is still towards lower losses.

2.2 FULLY CONNECTED NEURAL NETWORKS

fully connected neural networks (FCNNs) are neural networks where all neurons in layer L_l are connected to all the neurons in layer L_{l+1} (e.g. Figure 2.2). Each neuron in layer L_{l+1} receives an input from every neuron in layer L_l . Therefore, FCNNs require many parameters and are computationally expensive to train. FCNNs tend to overfitting, in particular if the training data set is limited. With an increasing number of parameters, the network becomes too complex and captures the noise in the training data, rather than the underlying pattern. Nevertheless, FCNNs are used as a classifier on top of DNNs, because of their ability to represent a variety of input data.

2.3 CONVOLUTIONAL NEURAL NETWORKS

The main building blocks of CNNs are convolutional layers. Due to their capabilities of detecting features and patterns in 2D images, CNNs are widely used for image classification. Image classification is considered as the prediction of a certain structure in an image and label it correctly. An example of image classification would be matching handwritten digits to the correct number it is representing. In this thesis, CNNs are used to detect bubble-like structures in astronomical images.

2.4 TRAINING

The training of a neural network is a learning process which ends with the ability of the network to predict the correct label of the given input data. In general one distinguishes between supervised and unsupervised learning. Supervised learning provides labels for each given input. The network learns to match these label-input pairs to predict the right label for an unknown input. Unsupervised learning is used to learn structure the data inherently without providing labels. Therefore, unsupervised learning algorithms can be used for clustering or for detecting anomalies, for example. In the following the focus is on supervised learning.

The input for training a CNN can be in the form of a 2D image. Predictions for an image classification problem are then the hopefully correct labels of the given images. Neural networks can be trained on a variety of tasks and depending on the task, a proper training routine has to be developed to achieve a well-trained network able to solve the defined task. In this section methods and parameters for the training routine are discussed.

	predicted positive	predicted negative
positive	True Positive (TP)	False Negative (FN)
negative	False Positive (FP)	True Negative (TN)

Table 2.1: This table gives an overview of the possible outcome of a prediction. A positive prediction indicates that the corresponding sample belongs to a certain label, where a negative prediction indicates that the sample does not belong to this label.

The training progress can be tracked with different type of metrics. One of these metrics is the *accuracy*. It calculates the ratio of the total number of true predictions and the total number of predictions

$$\text{accuracy} = \frac{1}{N} \sum_C (\text{TP}(C) + \text{TN}(C)) \quad (35)$$

with N the total number of predictions. If the network predicts the correct class C $\text{TP}(C) = 1$, a so-called true positive (TP) prediction, otherwise $\text{TP}(C) = 0$. If the network predicts that the sample does not belong to a considered class C correctly $\text{TN}(C) = 1$, this is called a true negative (TN) prediction. The sum of $(\text{TP}(C) + \text{TN}(C))$ is therefore one or zero, depending of the outcome of the prediction and the true class of a sample. Table 2.1 gives an overview of possible outcomes of a prediction including false positive (FP) and false negative (FN) predictions.

For imbalanced data sets, that means different classes are represented by significantly different numbers of samples, the accuracy is heavily biased towards the classes with the largest number of samples. This is because the accuracy takes just the TP and TN predictions into account. Although, the network is not able to predict samples of a minority class correctly, the network could still reach a high accuracy of $\sim 99\%$ (He, 2013). Consequently, the accuracy metric is not well suited to measure the quality of the training process with an imbalanced training data set. The *precision*

$$\text{precision} = \frac{\sum_C \text{TP}(C)}{\sum_C (\text{TP}(C) + \text{FP}(C))} \quad (36)$$

is taking FP predictions into account as well, and is therefore better suited for an imbalanced data set. If the network predicts the wrong class of a sample $\text{FP}(C) = 1$, otherwise $\text{FP}(C) = 0$.

To avoid training on an imbalanced training set, a common method is to calculate class weights for each class in the training data set. These weights ensure that samples of a minority class are of same importance as samples of a majority class. Calculating class weights is therefore recommended.

A second metric to track the quality of the training process is the *loss*. The loss is calculated from the loss function (e.g. Equation 30).

While metrics like accuracy and precision should increase, the loss should steadily decrease during the training. Decreasing losses indicate less parameter changes after each training iteration and, therefore, a converging network. A rising loss function, after forming a global minimum, indicates overfitting. Tracking the loss can help therefore to reduce the risk of overfitting by introducing checkpoints. At these checkpoints the training status of the network is saved. The common used rule of a checkpoint is to save the parameters of the network, if the loss of the current training iteration is smaller than the loss of the previous iteration. Thus, if overfitting occurs the saved network parameters are not affected.

Each training iteration is called a *batch*. The training data set is divided into batches, which are sub-data sets with a certain amount of samples, namely the batch size. After each batch, the entire set of trainable parameters is adjusted. How much the parameters are adjusted depends on the loss and on the so-called learning rate (LR). A smaller LR means a finer adjustment of parameters. The LR is a hyperparameter for the training algorithm used.

The main goal of train a neural network is to avoid over- and underfitting the network and create a well generalized network. Going through the training data set once is not sufficient. Instead, each sample of each class is shown to the network several times. These repetitions are called *epochs*.

Beside the basic training procedure for a neural network, there are different training methods. Some of those methods are sketched and the idea behind them is explained in the following.

Pre-Training

The network can be pre-trained on a related data set to initialize the weights and biases of the network. A related data set show similar features like the training data set. To stay in the example of train a CNN to recognize handwritten digits, a related data set could be a set of circles, arcs, horizontal and vertical lines which are related to different parts of digits. Instead of randomly initializing parameters, train the network on a related data set and task, can help to improve the performance of the network on the task it is actually built for. The quality of the related data set is of similar importance like the actual training data set. Note, that the degree of relationship between these two data set is crucial for achieving high precision networks on the desired task.

Transfer-Learning

Using a pre-trained network as a starting point for training the network on the desired task is generally referred to as transfer-learning.

The idea behind this type of training is to utilize the related information in the related data set for the actual defined problem (Goodfellow et al., 2016). While cats and dogs are two distinct animals, the overall appearance of these two pets are evidently. Therefore, they are considered as related. Whereas, a cat and a fish cannot be considered as related in general. The concept of transfer-learning can therefore be applied to expand the abilities of the network. For example, a CNN trained on recognizing cats and dogs is pre-trained for recognizing wolves. All three classes have some characteristics in common, e. g., four legs and fur, but differ in size and color, for example.

Within the scope of transfer-learning, there are a variety of approaches one can take. Freezing layers within the network is one additional method to avoid parameter changes during training for those layers. How many layers are frozen is an experimental task or an educated guess and therefore a question of experience. Just an amount of certain layers of the pre-trained network are used for transfer-learning sometimes, while a set of new initialized layers are put on top. In this case, the newly initialized layers are trained on the output of the pre-trained layers, whereas the pre-trained layers are not effected during training. There is no strict rule of how to approach transfer-learning.

Fine-Tuning

If the accuracy of the network reaches an appropriate level, the network can be fine-tuned in a niche of the defined problem. Having a network able to recognize cats and dogs, trying to classify a bear with the same network will result in a FP prediction of one of the classes cat or dog. Since the features of bears are similar to the features of cats and dogs, fine-tuning can be used to extend the task of the network to recognize cats, dogs and bears.

Similar to transfer-learning, there is no right or wrong for fine-tuning a network. It is again an experimental task to figure out what works best. The methods used for fine-tuning are also similar to transfer-learning: Not affecting parameters during training by freezing certain layers is one shared method. In addition, a smaller LR can be used, applying fine adjustments to parameter which are not frozen. One more hyperparameter to use for fine-tuning is the batch size. A larger batch size reduces the number of parameter adjustments by averaging over more samples during training. Combining different methods in a way that improves the performance of a network in a niche can be useful and time saving by no need to train a network from scratch.



(a) rotation



(b) crop



(c) vertical flip



(d) horizontal flip

Figure 2.11: A set of four augmentation operations applied on an original image. With augmentation the original image is transformed in a different image preserving the characteristics of the original. Augmentation operations could be (a) rotations, (b) cropping, (c) vertical and (d) horizontal flipping. In this example of a chess board, (d) does not preserve the characteristics of the original image because of the unnatural scenario of an upside down chess board and should therefore be avoided for augmentation. The image was generated using Midjourney (<https://www.midjourney.com>).

2.5 MOTIVATION

The amount and complexity of data is permanently increasing in the future. This is also true for astronomy, where data from different ground and space based instruments are collected and shared. This will lead to the unprecedented opportunity to develop new ways of analyzing data. Machine learning algorithms play an important and increasing role in this regard. Neural networks become more and more popular among astronomers to develop automated routines to extract information from large data sets, for example to identify and classify galaxies, stars and exoplanets (G et al., 2023; Walmsley et al., 2022; Zhao et al., 2023).

In this work we develop a pipeline for automated detection and classification of fundamental structures of the ISM in several wavelength regimes simultaneously. This pipeline is called *BScan*. It leverages machine learning algorithms to detect and characterize bubble-like structures with the focus on SNRs and HII regions. The final goal is to provide *BScan* with a set of images and get a list of detections as a result. *BScan* is written in the programming language Python and consists of parts which will be explained in part ii of this thesis.

Part II

DETECTION AND CLASSIFICATION OF
BUBBLE-LIKE STRUCTURES

DATA

We use multi-wavelength data from different instruments and surveys to detect and characterize bubble-like structures introduced in section 1.2. The data are astronomical images of the MCs and the Milky Way (MW), our own host galaxy. The performance of *BScan* is heavily depending on high quality data. Details of these data are given in the following.

3.1 INFRARED

The Spitzer Space Telescope is an infrared observatory in an earth-trailing solar orbit (Werner et al., 2004). It was decommissioned in 2020 and provided data to study, e. g., dust and dust formation in SNRs (Matsuura et al., 2022; Rho et al., 2009). With its imaging capabilities in the range of $(3 - 180)\mu\text{m}$ (Caltech, 2023), Spitzer images of the Milky Way and the LMC are well suited for training purposes.

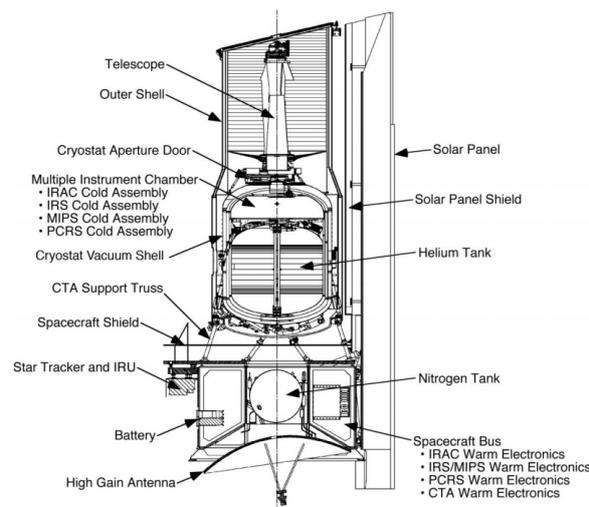


Figure 3.1: This Figure shows the cross-section of the Spitzer Space Telescope with all hardware parts. From Werner et al. (2004).

The multiband imaging photometer (MIPS) is one of three instruments of the Spitzer Space Telescope. It has a high sensitivity and a large imaging field at $24\mu\text{m}$, $70\mu\text{m}$ and $160\mu\text{m}$ with bandwidths of $5\mu\text{m}$, $19\mu\text{m}$, and μm , respectively (Rieke et al., 2004). With the survey of the inner galactic disk with MIPS (MIPSGAL)¹ the inner 248deg^2 of the Milky Way's galactic plane was imaged. The images are taken at

¹ <https://irsa.ipac.caltech.edu/data/SPITZER/MIPSGAL/overview.html>

24 μm and 70 μm with a resolution of 6 arcsec and 18 arcsec, respectively. In total MIPS GAL covers the galactic latitudes of $-1^\circ < b < 1^\circ$ and longitudes of $l < 69^\circ$ and $292^\circ < l$. More information about MIPS GAL can be found in Carey et al. (2009).

The infrared array camera (IRAC) is the second imaging instrument beside the MIPS. It has a four-channel camera able to take broadband images at wavelength of 3.6 μm , 4.5 μm , 5.8 μm and 8.0 μm . The resolution measured as the full width half maximum (FWHM) at these wavelengths is 1.66 arcsec, 1.72 arcsec, 1.88 arcsec and 1.98 arcsec, respectively. With its large field of view (FOV) and high sensitivity, IRAC provides high quality images (Fazio et al., 2004). The galactic legacy infrared midplane survey extraordinaire (GLIMPSE) I² was taken by the IRAC instrument, for example. It covers roughly 220deg² of the Milky Way's galactic plane. The observations range between a latitude of $-1^\circ < b < 1^\circ$ and longitude of $-65^\circ < l < 10^\circ$ and $10^\circ < l < 65^\circ$. In addition to the already mentioned surveys, the surveying the agents of a galaxy's evolution (SAGE) survey was taken by the MIPS and IRAC instruments and include approximately $(7 - 8)^\circ$ of the LMC³ and $\approx 30\text{deg}^2$ of the SMC⁴. Figure 3.2 shows mosaic images of the LMC created from SAGE observations.

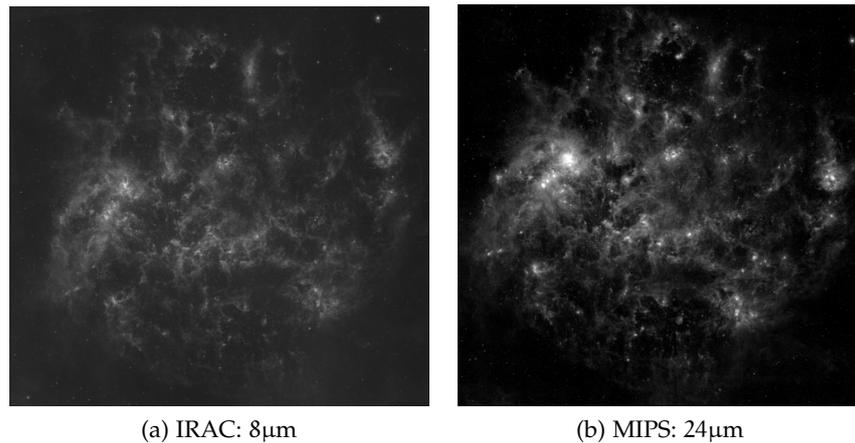


Figure 3.2: The LMC recorded by the IRAC (a) and MIPS (b) instrument of the Spitzer Space Telescope. Credit: Fazio et al. (2004) and Rieke et al. (2004)

3.2 OPTICAL

The UM/CTIO magellanic cloud emission-line survey (MCELS) is a deep imaging multiple emission-line survey of the MCs in the optical regime (Smith C., 2000). As described in chapter 1.3, the MCs provide

² irsa.ipac.caltech.edu/data/SPITZER/GLIMPSE/overview.html

³ irsa.ipac.caltech.edu/data/SPITZER/SAGE/overview.html

⁴ irsa.ipac.caltech.edu/data/SPITZER/SAGE-SMC/overview.html

a unique laboratory to identify and classify structures of the ISM and allow detailed studies of SNRs, for example. The MCELS images cover the central $(8 \times 8)^\circ$ of the LMC (Figure 3.3 a) and $(3.5 \times 3.5)^\circ$ of the SMC (Figure 3.3 b). Therefore, the vast majority of the gaseous extent of these dwarf galaxies is visible (Smith C., 2000).

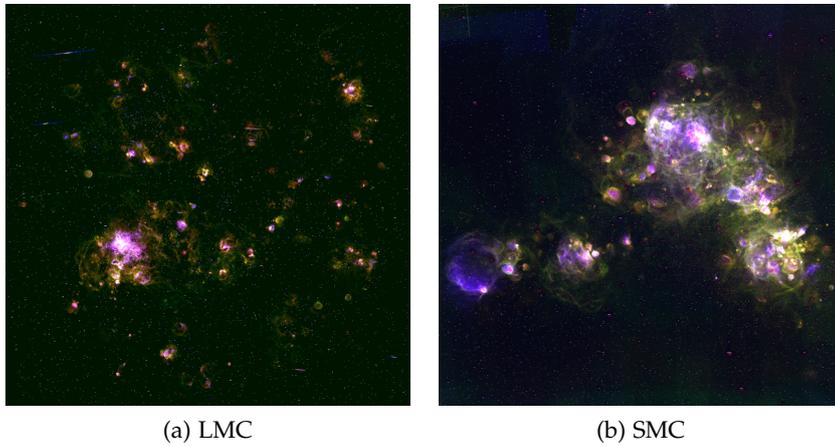


Figure 3.3: The Large (a) and Small (b) Magellanic Cloud observed at the optical emission lines $H\alpha$ (red), [SII] (green) and [OIII] (blue). Credit: Smith C. (2000)

In order to accomplish a resolution of $(3 - 4)''$ and a FOV of around 1° , the survey leveraged the ground based Curtis Schmidt telescope at Cerro Tololo Inter-American observatory (CTIO) in La Serena, Chile (see Figure 3.4). The narrowband images were taken with filters centered on the $H\alpha$ ($\lambda_c = 6563\text{\AA}$, $\Delta\lambda = 30\text{\AA}$), [SII] ($\lambda_c = 6724\text{\AA}$, $\Delta\lambda = 40\text{\AA}$) and [OIII] ($\lambda_c = 5007\text{\AA}$, $\Delta\lambda = 40\text{\AA}$) emission lines. In addition, two continuum band images with filters centered at $\lambda_c = 6850\text{\AA}$ and $\lambda_c = 5130\text{\AA}$ were taken to subtract the stellar background (Smith C., 2000). Since the recording, these data were used by optical and multi-wavelength studies to gain knowledge about the ISM in the MCs, see e. g. Bozzetto et al. (2023), Maggi et al. (2019), and Yew et al. (2021).



Figure 3.4: This picture shows the Cerro Tololo Inter-American Observatory at sunset. Credit: NOAO/NSF/AURA

3.3 RADIO

In the radio regime a 120deg^2 continuum image of the LMC from the Australian square kilometre array pathfinder (ASKAP) is used. ASKAP is a radio telescope, located in Australia, made out of 36 radio dish antennas, each with a diameter of 12m (see Figure 3.5). A single dish has a wide FOV of $13.9 \times 12.1\text{arcsec}^2$. The array of antennas is arranged in two dimensions with a maximum baseline of 6km. With this setup ASKAP is able to perform surveys in the frequency range of 700 – 1800MHz (CSIRO, 2023). The data were taken at 888MHz and a bandwidth of 288MHz with a spatial resolution of $13.9 \times 12.1\text{arcsec}^2$ (Figure 3.6 a). Further information about observing and data processing can be found in Pennock et al. (2021).



Figure 3.5: This image shows ASKAP telescopes on site. Credit: CSIRO.

A second radio mosaic image of the LMC comes from the Australian telescope compact array (ATCA). With its antennas of 22m in diameter, ATCA observes the radio sky from millimeters to a few centimeters (CSIRO, 2023). The scientific objectives are, among others, studying SNe and SNRs. For the purpose of this thesis, archival ATCA data at $1.4\text{GHz} = 21\text{cm}$ are used (Figure 3.6 b).

The ASKAP data in Figure 3.6 (a) is used to create a non-thermal radio image. Therefore, thermal radio emission has to be subtracted. Assuming that thermal radio and $\text{H}\alpha$ emission originate from the same electron population, the thermal component can be subtracted from the non-thermal component by calculating a normalization factor for the $\text{H}\alpha$ emission in regions which show thermal emission only. This procedure is also done in (Bozzetto et al., 2023; Ye et al., 1991) or Zangrandy et al. (2023, in prep.). We use the resulting non-thermal radio image to detect and discriminate SNRs and HII regions.

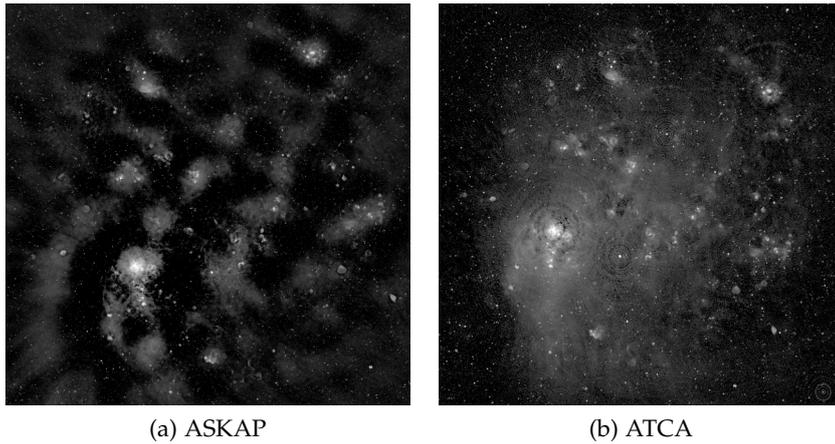


Figure 3.6: The LMC in radio as seen by ASKAP (a) and ATCA (b). Credit: Pennock et al. (2021)

3.4 X-RAY

The extended Röntgen survey imaging telescope array (eROSITA) is an X-ray instrument on the Spectrum-Röntgen-Gamma (SRG) spacecraft. It is a wide-field X-ray telescope with a large FOV compared to competitors like Chandra or the X-ray multi-mirror-Newton (XMM-Newton) (Predehl et al., 2021). From its L2 orbit, it is planned to perform in total eight eROSITA all-sky surveys (eRASSs) by scanning the hole sky several times. At the time of writing, four of these surveys are finished. The scanning geometry is drawing a circle on the sky, with

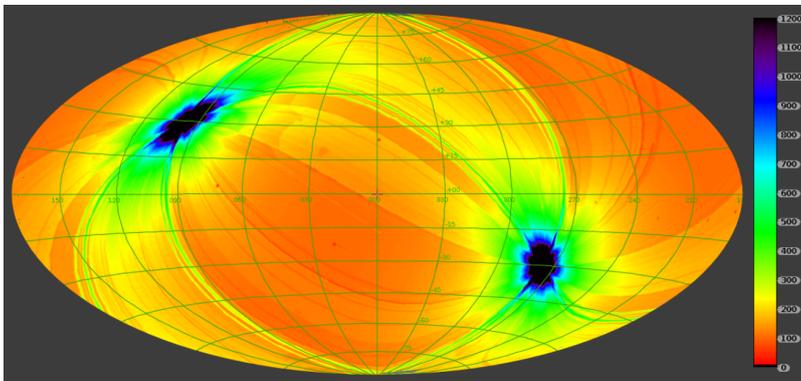


Figure 3.7: The effective (vignetted) exposure map derived from the eRASS₁ in the 0.6to2.3keV energy band and galactic coordinates. From Predehl et al. (2021).

SRG and, thus, eROSITA in its center. Adding up the circles creates a sphere with its poles coinciding with the ecliptic poles. Consequently, the effective exposure is the highest at the northern and southern ecliptic poles (Figure 3.7). The LMC is luckily located near the southern ecliptic pole, which makes eROSITA images the perfect set of data

to study the LMC in the X-ray regime. eROSITA has seven cameras and seven identical X-ray mirror assemblies co-aligned (see Figure 3.8 a), collecting photons in the energy range of 0.2 – 8keV (Sunyaev et al., 2021). The energy and angular resolution of eROSITA is 80eV at 1.5keV and 30 arcsec half-energy width (HEW) (Sunyaev et al., 2021), respectively.



Figure 3.8: The X-ray telescopes named eROSITA (a) and XMM-Newton (b) are shown schematically, with their X-ray mirrors and detectors, in this Figure. From ESA: XMM-Newton SOC (2022) and Predehl et al. (2021).

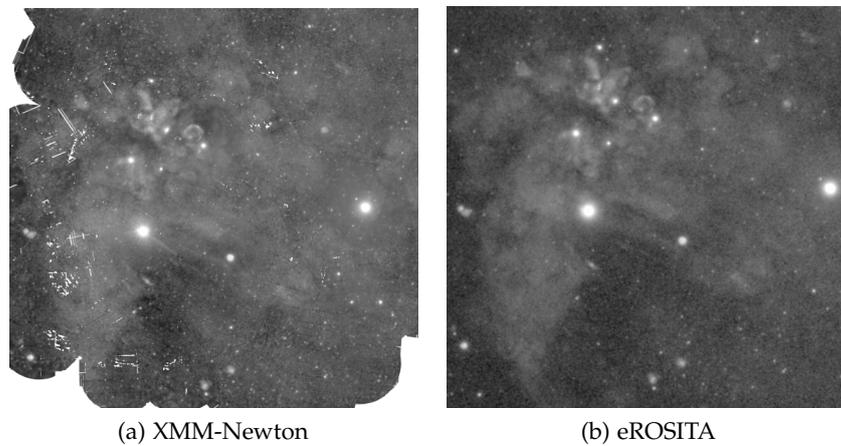


Figure 3.9: The images show roughly the same cutout of the LMC recorded by XMM-Newton (a) and eROSITA after eRASS4 (b). Credit: ESA, MPE

The second X-ray telescope used is the XMM-Newton. Launched in 1999, XMM-Newton observes the universe in the X-ray and optical-UV regime (ESA: XMM-Newton SOC, 2022). It has three telescopes, each equipped with an european photon imaging camera (EPIC) and similar X-ray mirrors, focusing the incoming photons onto the cameras. The observable energies of the XMM-Newton range from (0.2 – 12)keV with an angular and spectral on-axis resolution of 6 arcsec FWHM and

$E/\Delta E \approx 20 - 50$, respectively (ESA: XMM-Newton SOC, 2022). With its FOV of just 30 arcmin, XMM-Newton observes targets in single pointing mode. These single observations can be setup as a mosaic image by combining appropriate observations as done for the LMC. Figure 3.9 shows the LMC in X-ray as seen by XMM-Newton and eROSITA.

TRAINING AND TEST DATA

We use the data in chapter 3 to create training samples to train a neural network on detecting bubble-like structures in astronomical images later on. The samples are grayscale sub-images with an image size of (50×50) pixel and a bubble-like structure in the center (Figure 4.1 a, b). In addition, samples for no-bubble-like structures are added to the set of training samples (Figure 4.1 c). Each sample is classified into one of three classes (or labels), namely:

- *shell*: a ring-like structure which is brighter at the edge than in the center
- *diffuse*: an extended diffuse structure with a gaussian-like, or plateau-like, profile and
- *none*: no bubble-like structure present.

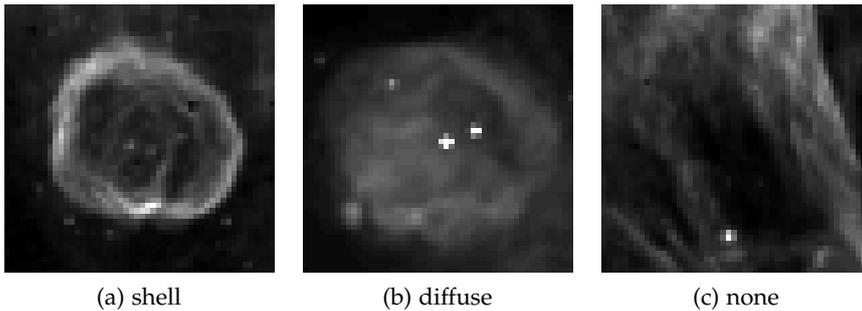


Figure 4.1: Each image shows an example of the defined classes *shell* (a), *diffuse* (b) and *none* (c) created from the MCELS2 image showing $H\alpha$ emission.

4.1 TRAINING DATA SET

The samples in the training data set (TRS) are created by defining regions on each available astronomical image, by manually marking a bubble-like structure. A guideline for the region definition process is that the bubble-like structure covers roughly 70 – 80% of the defined region. The area of the remaining 20% of the image is used for augmentation purpose (e. g. translations and cropping). Since there is no quantitative way to classify the defined regions (samples) into the defined classes at this point, we classify them by the human eye and decide if the bubble-like structure shows one of the characteristics described above or not. Structures of class *none* are created by randomly

defining regions on the image without enclosing bubble-like structures. That means the class *none* includes point-like or filamentous structures. Since there is no quantitative labeling process involved, the classification process is depending on the human eye and is therefore biased. In real data, the class characteristics may not be clearly distinguishable to the human eye. A diffuse bubble-like structure may also have some signs of a shell and vice versa. The labels may therefore be different depending on the observer.

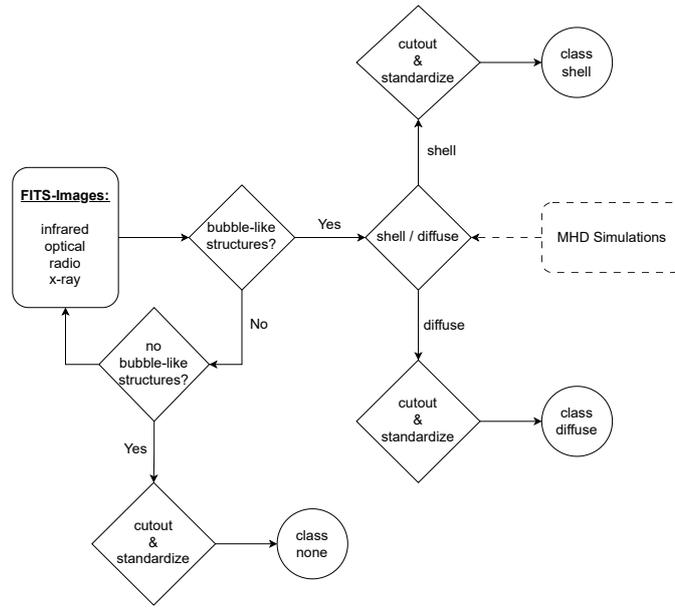


Figure 4.2: The generation of the dataset represented in a flowchart. Defined regions are used to create images for the training data set by cutting out the regions and standardize the pixel values. Magneto-hydrodynamic simulations are added to extend the dataset.

The defined regions are used to cut out the samples for the training data set. In some defined regions, bright structures dominate the sample and the pixel values span over a wide range. This is the case for low background images with bright point-like structures, for example. The range of pixel values of each sample will be standardized to $[0, 1]$. Compressing the pixel value range improves the training efficiency because of a more controlled back propagation (equations 23 and 24) and, therefore, more efficient parameter updates (equations 25 and 26) while training (see section 2.1.7). The process of generating the training data set is summarized in the flowchart shown in Figure 4.2.

The samples created from the MW data, mostly results from the so called Milky Way project (MWP). It is a citizen science project for visual classification of infrared (IR) bubbles in the MW. With a catalog of 5106 IR bubbles (Simpson et al., 2012) as an outcome, the MWP is used to detect interstellar bubbles with machine learning algorithms like Random Forest as done in Beaumont et al. (2014). The MWP an-

class	shell	diffuse	none	total
LMC	138	209	436	783
MW	120	61	545	726
MHDS	43	11		54
total samples	303	281	981	1563

Table 4.1: Number of training samples in each class originating from astronomical images of the LMC and MW as well as from MHDS. In total, there are 1563 samples in the training data set.

alyzed the GLIMPSE and MIPS GAL observed by Spitzer (see chapter 3). Although, these data are in the IR regime, the MWP is also utilized for this work together with the method described above. We do not use every region marked as a bubble in the MWP for the training set, because of the introduced bias and the fact, that the definition of a bubble-like structure and the scientific objective in the MWP is different compared to this work, see Simpson et al. (2012).

After going through the given data of different instruments and galaxies and counting the number of samples in each class for the LMC and MW, the total amount of training samples are 1509 (Table 4.1). In the context of machine learning, a set of 1509 original training samples is challenging for training a DNN. Note that most of the samples are of class *none*.

We therefore tried to increase the number of training samples by magnetohydrodynamic simulations (MHDSs). Thanks to a team from the INAF – Osservatorio Astronomico di Palermo, Italy, simulated samples of SNRs could be added to the training data set. These simulations use the models from Ustamujic et al. (2021b), Ustamujic et al. (2021c), Orlando et al. (2019), Orlando et al. (2016) and Orlando et al. (2012). With these models the resulting simulations of SNRs have different morphologies and ages. If the line of sight (LoS) and evolution time of the remnant is varied, different images of a SNR can be generated. In addition to the described variations, simulations with different parameters like the ejecta mass or explosion energy can be calculated as well. The morphology are very similar in most cases with these additional parameter changes. There are three images for every simulation: one optical and one X-ray image for the hard ((1 – 2) keV) and soft ((0.3 – 1) keV) band, respectively (Figure 4.3). For the X-ray simulations, the ISM abundances from Maggi et al. (2016) and a column density of $n_{\text{H}} = 3 \cdot 10^{21} \text{ cm}^{-2}$ are taken into account. Abundances from the ejecta is assumed to be 10-times larger than of those from the ISM and an ionization equilibrium is simulated. The resolution of these synthetic images is comparable to the Chandra X-ray Observatory. For the optical images, temperatures between 10^4 K and 10^5 K are considered.

The total number of generated MHDS samples is 54 and is also listed in Table 4.1. Therefore, the total number of training samples are not increased significantly by the MHDSs.

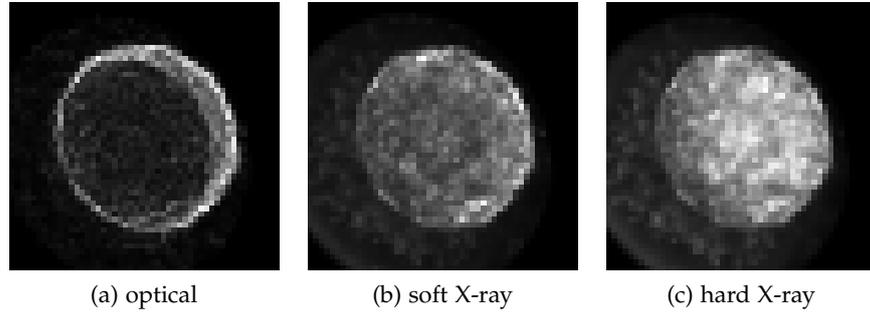


Figure 4.3: These images show the optical (a), soft (b) and hard (c) X-ray simulation of an SNR. Credit: Ustamujic S. et al. at the INAF.

4.2 KOLMOGOROV-SMIRNOV TEST

The presented procedure for the training data set results in three sets of training samples originating from data of the Large Magellanic Cloud, the Milky Way and the magnetohydrodynamic simulations (see Table 4.1). In order to ensure that the created samples are different to each other, we perform a Kolmogorov-Smirnov test (KST). This statistical rank-order test for two independent samples (Hodges, 1958) is named after Andrey Kolmogorov and Nikolai Smirnov and compares the two underlying continuous distributions of the samples which, in this case, are the pixel value distributions of two images within a set of samples. The null hypothesis is that the two distributions are identical. Therefore, the null hypothesis is rejected if the distributions are different. The KST is applied using the `ks_2samp` function, implemented in the `scipy.stats` python package. To verify the test, the `ks_2samp` function returns the p-value. Small p-values lead to a rejection of the null hypothesis. We introduce a statistical significance of $\hat{p} = 0.05$. If $p \leq \hat{p}$ the null hypothesis is rejected and the samples are considered as different. The probability that the tested samples have the same distribution is therefore less than 5%. The value of the statistical significance $\hat{p} = 0.05$ is considered as a compromise between being not too strict to reject, and not too kind to accept the null hypothesis.

Figure 4.4 shows the results of the KST applied to every sample pair in all three data sets listed in Table 4.1 and, therefore, to the entire training set created by combining the single data sets. According to the KST, 0.3% of the samples within the training data set are similar. Similar samples are only found of class *none*. Note that the number of samples in the *none* class is $\sim 55\%$ of the total samples and, thus,

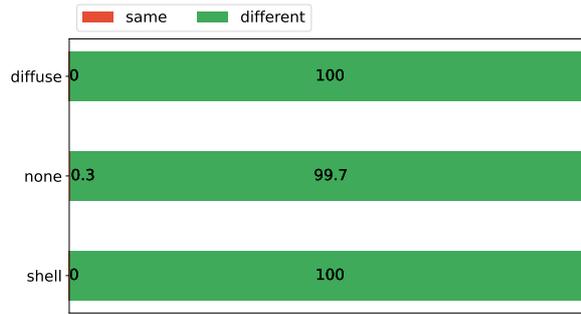


Figure 4.4: KST applied on the entire data set of the Large Magellanic Cloud, Milky Way and magnetohydrodynamic simulations combined in one training data set. Less than 0.3% of the samples within the data set are similar. In fact, similar samples are only found in class *none*. Overall, samples in the training data set are considered as not identical according to the KST.

this class is over-represented. Moreover, samples of class *none* include point-like structures with low background. Consequently, a sample of small pixel values over the entire image except for the point-like structure, could lead to a similar pixel value distribution, if point-like structures are at a similar position.

4.3 DATA SET FOR PRE-TRAINING

The parameters of the network can be initialized randomly to train a network for the first time. Pre-training (see section 2.4) is helpful to adjust the random parameter values to a better starting point for the final training. In this section we introduce two data sets for pre-train the neural networks.

ImageNet

*ImageNet*¹ is a database of more than 14 million hand-labeled RGB images. On average 1000 images visualize a set of cognitive synonyms (synsets) in the lexical *WordNet*² database (*About ImageNet* 2023). In total, *WordNet* contains 117000 such synsets (*About WordNet* 2023). Therefore, *ImageNet* is a useful data set for training large-scale object recognition models to a good level of generalization. There are no astronomical images included in its database. *ImageNet* is used in Geyer-Ramsteck (2020) in form of the pre-trained VGG16 network implemented in *Keras*³, a deep learning framework.

¹ <https://www.image-net.org/about.php>

² <https://wordnet.princeton.edu/>

³ <https://keras.io/api/applications/vgg/>

Galaxy10

The Galaxy10 data set (G_{10}) was created with the Galaxy Zoo (Lintott et al., 2008; Lintott et al., 2011). Similar to the MWP, it used the public to classify around 36000 galaxies. The labels, given by the participants, are not accurate for each galaxy. Therefore, only 21830 galaxies were selected, where 55% of the participants' classifications fall into the same class⁴, and classified into ten distinct classes (see Figure 4.5) for the Galaxy10 Sloan digital sky survey (SDSS) data set (see Table 4.2). The Galaxy10 SDSS set uses images from the Sloan Digital Sky Survey⁵

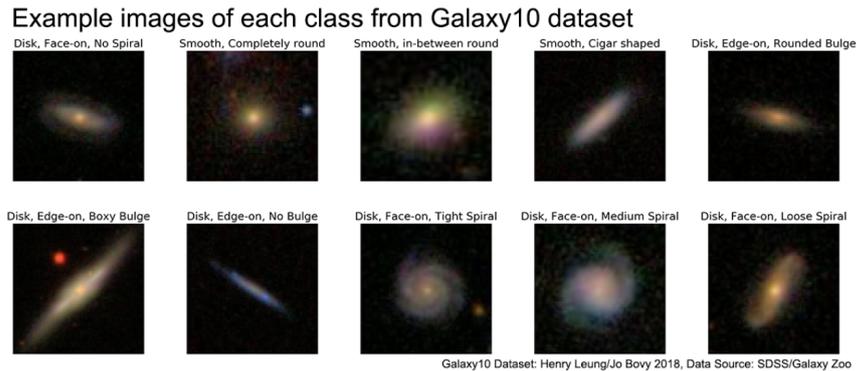


Figure 4.5: This image shows examples of the Galaxy10 SDSS data set. One image is shown for each class listed in Table 4.2. From <https://astronn.readthedocs.io/en/latest/galaxy10sdss.html>.

Galaxies have a different astrophysical appearance compared to bubble-like structures shown in Figure 4.1. Although they have a bright point-like center and may show some spiral arms, they have some characteristics in common, for example the circular shape. In order to initialize the parameters, the network is pre-trained on a subset of the G_{10} . We mark the classes in Table 4.2 which we refer to as the reduced Galaxy10 data set (G_6). It should be noted that G_6 is an unbalanced data set. To train the network on the G_6 , class weights are calculated for each class (see section 2.4).

Images from class number 3,4,5 and 6 are removed from the data set, because of the extremely elongated shape of the galaxies. Galaxies showing spiral arms are used for the G_6 because of their shell-like characteristic. Since G_6 contains just color images, they are transformed into grayscale images merging the red, green and blue channel into one and normalize the pixel values to 255. The idea behind using the G_6 is to pre-train the network on a more astronomical data set compared to *ImageNet*.

⁴ <https://astronn.readthedocs.io/en/latest/galaxy10sdss.html>

⁵ <https://www.sdss.org/>

class nr.	description	total	used	class label
0	Disk, Face-on, No Spiral	3461	X	A
1	Smooth, Completely round	6997	X	B
2	Smooth, in-between round	6292	X	C
3	Smooth, Cigar shaped	394		
4	Disk, Edge-on, Rounded Bulge	1534		
5	Disk, Edge-on, Boxy Bulge	17		
6	Disk, Edge-on, No Bulge	589		
7	Disk, Face-on, Tight Spiral	1121	X	D
8	Disk, Face-on, Medium Spiral	906	X	E
9	Disk, Face-on, Loose Spiral	519	X	F

Table 4.2: Overview of the ten classes of the Galaxy₁₀ SDSS data set. In total 21830 samples are available with this data set. The six crosses indicate if the samples of the corresponding class are used in the reduced data set (G6) for pre-training. In the end there are 19296 samples in the pre-training data set. A short label for each class used in the G6 is introduced.

4.4 TEST DATA SET

In order to test the performance of the neural network, a test data set (TES) is created. We created the test samples with the same methods and guidelines described in section 4.1 for the training samples, but using the SMC data of the MCELS. To ensure the quality of the TES a Kolmogorov-Smirnov test is applied as well. Figure 4.6 shows the results. Similar to the training data set, most samples which are similar to each other are found in class *none* (0.2%) while in the classes *shell* and *diffuse* are no similar samples, according to the KST. It is im-

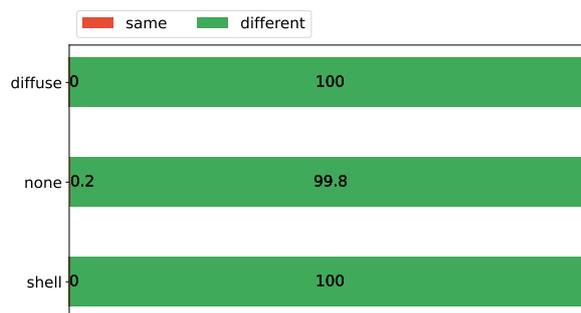


Figure 4.6: KST for the test data set. Similar to the training data set, most samples (0.2%) similar to each other are found in the class *none*.

portant to strictly separate the training samples from the test samples to make sure that applying the network to the test samples is a blind

test for the network. In Table 4.3 the number of test samples are listed.

class	shell	diffuse	none	total
SMC	32	38	30	100

Table 4.3: The number of test samples in each class originating from astronomical images of the SMC. In total, there are 100 samples in the test data set.

4.5 AUGMENTED DATA SETS

In addition to the main data sets for (pre-) training and testing the networks, we generate augmented data sets. Assuming current detection methods, by eye or algorithm based, e. g. , in Collischon et al. (2021), to detect bubble-like structures in the ISM are missing faint or low contrast structures, the idea of the augmented data sets is to fine-tune the trained networks later on, to detect faint bubble-like structures.

The basement of the augmented data sets is the entire training data set (TRS) from section 4.1. Each original sample is augmented once, thus, the size of the TRS is preserved. In the first step, samples of bubble-like structures are augmented such that the brightness is reduced to (10 – 25)% of the original brightness, while keeping the pixel value range the same. In the second step, the pixel values of the augmented samples are added to the pixel values of a random original none-sample. This results, in some cases, in faint bubble-like structures embedded in a background dominated region adding a new characteristic to the augmented sample. Since the none-samples are extracted randomly and some samples of the *none* class show hardly no diffuse emission but bright point-like structures, this method results in faint bubble-like structures with less diffuse background as well. In any case, the bubble-like structure is hardly visible by eye. In addition to the brightness reduction, rotations with a rotation angle of $\varphi < 360^\circ$ and small distortions are applied, to ensure a good augmentation level without misrepresenting the bubble-like structures. The pixel values are normalized as for the original training data set described in section 4.1. These augmented samples are referred to as the training data set of faint samples (TRS-F).

A second augmented data set of low contrast samples is generated from the training data set. Contrary to the TRS-F, the brightness is maintained, but the pixel value range is compressed, resulting in lower contrast. After applying augmentation, the samples have a contrast level of (10 – 25)% compared to the original samples. The idea is to simulate bubble-like structure which are similar to the present background and, therefore, harder to spot in the image. This is more important in crowded regions which are crossed by filaments and

other diffuse background emission. These augmented samples are referred to as the training data set of low-contrast samples (TRS-LC).

The same procedure is done for the test data set. The corresponding data sets are referred to as the test data set of faint samples (TES-F) and the test data set of low-contrast samples (TES-LC). We used Augmentor⁶, a python package specialized on augmentation, for all performed augmentation transformations.

⁶ <https://github.com/mdbloice/Augmentor>, 16. May 2023

DETECTION OF BUBBLE-LIKE STRUCTURES

The *BScan* detection pipeline consists of a CNN, to detect bubble-like structures, and a custom algorithm to scan the input image and estimate the center coordinates as well as the extent of the bubble-like structure. In this chapter we introduce all parts of the detection pipeline.

5.1 ARCHITECTURES

The detector is a CNN trained for an image classification problem. Throughout this thesis, we train and test different neural network architectures (NNAs) to narrow down the best performing architecture for the classification task. The frame of this task is to classify images according to the given classes: *shell*, *diffuse* and *none*. In this section the architectures are described and the ideas behind them are explained.

5.1.1 VGG16

The first architecture used for *BScan* is the well known VGG16, a very deep convolutional network with $134 \cdot 10^6$ parameters used for large-scale image recognition. More details about the architecture of the VGG16 can be found in Simonyan et al. (2014). In short, VGG16 is called after the Visual Geometry Group of the Department of Engineering Science at the University of Oxford, which developed the network for the *ImageNet* large scale visual recognition challenge 2014 (ILSVRC14). It achieved the first place in localization and the second place for classification. The VGG16 consists of 13 convolutional layers with a kernel size of (3×3) pixels, which is the smallest kernel size to understand the concept of left or right, up or down and center (Simonyan et al., 2014). These layers are grouped in five convolutional blocks, each followed by a Maximum Pooling layer. After the last Maximum Pooling layer, a FCNN is applied for the classification. Geyer-Ramsteck (2020) tested and applied a pre-trained VGG16 on LMC data in his master's thesis. The training data set used in this work, however, is the TRS shown in section 4. In fact, the size of the TRS is too small to train a very deep neural network like the VGG16 from scratch. That is why Geyer-Ramsteck (2020) used the VGG16 pre-trained on the *ImageNet* data set (Russakovsky et al., 2015). As already stated, the *ImageNet* data set cannot be considered as an astronomy related data set.

In the following, we introduce a set of smaller CNNs. By keeping the basic idea of the successful VGG16 architecture but reducing the number of parameters, we create customized networks which are better prepared to handle the small TRS. As stated in section 2.1.7, finding the balance between the network complexity and the size of the training data set is important to train a network properly. Therefore, the parameter reduction approach gives more control over the training and performance of the customized networks used for the classification task.

5.1.2 Custom Convolutional Neural Networks

The small convolutional neural networks (SCNNs) are composed of convolutional blocks (Figure 5.1) with a certain number of stacked convolutional layers. As mentioned in section 2.1.7, the level of detectable features can be enhanced by stacking convolutional layers on top of each other. We therefore vary the number of stacked convolutional layers depending on the block position within the SCNNs. The general approach is to implement more convolutional layers per block as deeper the block is located. Therefore, we expect that basic features of an image are detected in the first few layers and more elaborate features are detected in deeper layers. After a set of convolutional layers, a layer for batch normalization and an activation function is implemented.

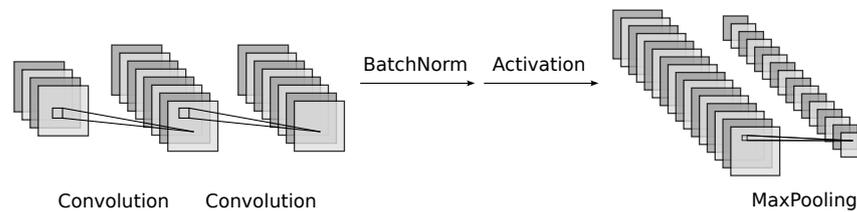


Figure 5.1: Illustration of a possible convolutional blocks. The block consists of two convolutional layers and a maximum pooling layer at the end. Feature map values are normalized by using batch normalization before applying an activation function. The number of convolutional layers, used kernels and the kernel size can vary for each convolutional block.

Batch normalization is stabilizing and speeding up the training procedure. The ReLU activation function (Equation 20) is implemented to calculate the activation of neurons. As an optional layer on top of each convolutional block, a maximum pooling layer can be used to focus on pronounced features. The idea of using maximum pooling is to discriminate between different features present in different classes (see characteristics listed in section 4). Maximum Pooling is therefore predominantly used for deeper convolutional blocks. The training data set in section 4 is set up to detect bubble-like structures

in the center of an image to improve the position estimation. Since maximum pooling leads to translation invariance, using maximum pooling layers contradicts the idea of the training data set to a certain degree. Therefore, maximum pooling is only used on top of some preferable deeper convolutional blocks due to its benefits (see section 2.1.7).

A FCNN is put on top of the last convolutional block for the classification process. The feature map of the last convolutional block is flattened into a one dimensional vector. This vector is used as the input for the FCNN. Each fully connected layer of the FCNN uses the ReLU activation function. The amount of neurons used for a layer decreases by a factor of two with each layer. In this part of the network, dropout is used for regularization purpose. The last fully connected layer (the output layer) of the FCNN has exactly as many neurons as number of classes ($C = 3$). Every neuron represents one class. To interpret the output of each neuron in the output layer as a probability, the softmax activation function (Equation 22) is used. Figure 5.2 illustrates the FCNN of a SCNN.

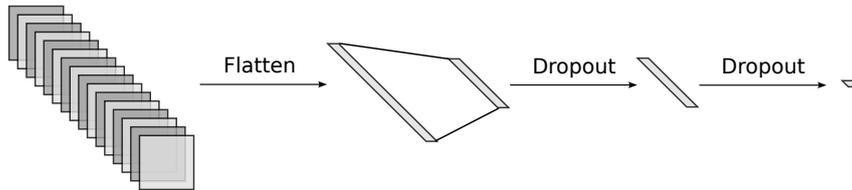


Figure 5.2: This image shows an illustration of the fully connected top of the SCNN. This part of the network is responsible for classifying. The feature map of the last convolutional block is flattened into a one dimensional feature vector. The length of this vector is equal to the used neurons in the first fully connected layer. For each fully connected layer the number of neurons is reduced. Dropout is used for regularization purpose.

5.1.3 Model Naming Convention

Throughout this thesis, different SCNNs are compared to each other to find the best performing network. Therefore, we introduce a naming convention to keep track on the configuration of the considered SCNN. Every network architecture is named after the following guideline

$$\text{SCNN}_{ij},$$

where ij denotes the number of convolutional blocks and the number of convolutional layers within the block, respectively. Figure 5.3 shows an example of a composed SCNN41. In addition, for each network, which can be trained on different data sets, the abbreviation for the data set is added at the end. If the network in the previous

example is pre-trained on the G6, for example, it would be states as SCNN41-G6.

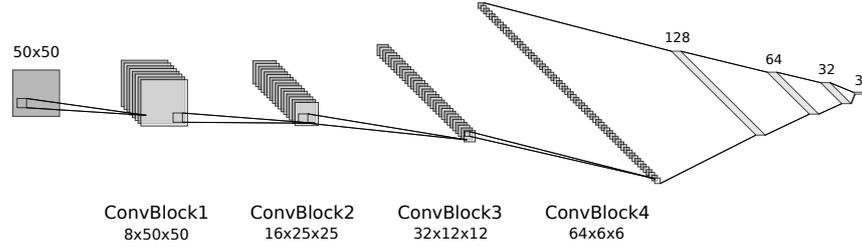


Figure 5.3: The network architecture of a SCNN41. An input image of 50×50 pixel (left) is fed to the network. The network consists of four convolutional blocks (ConvBlock). Each convolutional block i generates m_i feature maps of $n_i \times n_i$ pixels. The first convolutional block generates $m_1 = 8$ feature maps with $n_1 \times n_1 = 50 \times 50$ pixels. On top of the convolutional blocks fully connected layers of 128, 64, 32 and 3 neurons are used.

5.2 TRAINING PROCEDURE

Each SCNN is trained on an image classification task using three classes (see section 4). The training process used throughout this thesis is monitoring the loss for saving the model and avoiding the effect of overfitting. In addition to the loss, the precision is recorded for the entire training process. Due to the imbalanced training data set (see section 4) class weights are calculated using the python package *sklearn*. The implemented function *compute_sample_weights* computes the weights $w(C)$ for each class C

$$w(C) = \frac{N_S}{N_C \cdot N_S(C)}, \quad (37)$$

by taking the total number of samples N_S , the number of classes N_C and the number of samples inside each class $N_S(C)$ into account. The weights are used as an input for the training routine. As an optimizer the ADAM algorithm is used with an initial learning rate $LR = 0.001$ together with the CCE loss function. For more control over the LR during training, e. g., for fine-tuning, we implement custom LR functions (Figure 5.4). The first function is a step decay function, reducing the LR after a certain number of epochs E' . Each new LR is calculated as

$$LR' = LR \cdot S^{\frac{E}{E'}}, \quad (38)$$

where S is the step decay rate $S \in [0, 1]$ and $E = 250$ the number of epochs. A step decay rate of $S = 0.75$ sets the LR after E' epochs to the 75% level of the previous LR. The lower S the steeper the decay of the LR reducing the impact on the parameter updates during transfer-learning, for example.

A second custom LR function is an exponential decay function, where the LR is reduced after each batch according to

$$LR' = LR \cdot \exp(-E \cdot k) \quad (39)$$

with k the decay rate.

Beside these two decay functions, it is also possible to let the LR constant during the entire training process. If no decay function is

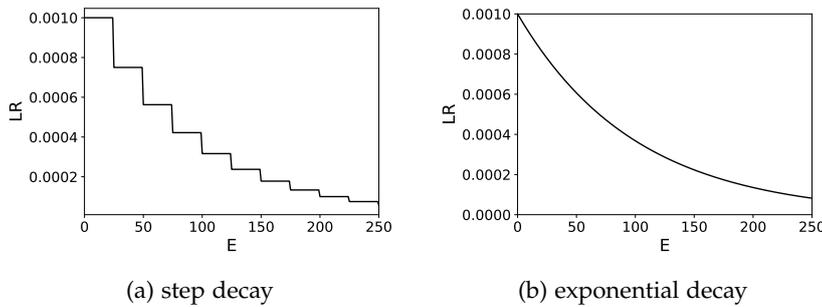


Figure 5.4: The step (a) and the exponential (b) decay function to control the learning rate during training with an initial learning rate of $LR = 0.001$. The step decay is initialized with a step size of $E' = 25$ epochs and a step decay rate of $S = 0.75$. For the exponential decay a decay rate of $k = 0.01$ is used.

specified, the default decay rate is used. In this case the decay rate is determined by the ADAM algorithm.

Furthermore, we arrange the training data set in a certain folder structure. Every sample is placed inside a folder named after the true class. Therefore, the number of folder is the number of classes to train on. This folder structure is accessed by the training routine to import the samples as the training data set. The training routine splits the training data set into the training- and test data set automatically. To control the size of the test data set, the input parameter *split_factor* is used. Its value is in the range of $[0, 1]$. For the data sets used throughout this work, we set a *split_factor* of 0.1. That means 10% of the entire training samples are used for testing and 90% are used for training the network. In addition, the training routine arranges the training data set into batches. The batch size is defined with the input parameter *batch_size*. For the available training data set, the batch size is set to 32. Before training, we rescale the imported samples in the training data set. Since the samples are grayscale images, the pixel values are in the range of $[0, 255]$. However, the pixel values are rescaled such that the values are in $[0, 1]$. Therefore, a rescaling factor of $\frac{1}{255}$ is applied.

Finally, the metrics used for evaluating the training progress are the precision on the one hand, and the loss on the other hand. The progress is saved to disk, if the loss is lower than the loss of previous iterations. This way, overfitting can be avoided (see Figure 2.7). The

Parameter	type	description
<i>data_path</i>	string	path to training data set
<i>model_path</i>	string	path to save model
<i>model_name</i>	string	name of the model to train
<i>nclasses</i>	integer	number of classes
<i>epochs</i>	integer	number of epochs
<i>batch_size</i>	integer	number of samples in a batch
<i>split_factor</i>	[0, 1]	fraction of samples used for testing
<i>optimizer</i>	string	optimizing algorithm
<i>learning_rate</i>	[0, ∞]	initial learning rate
<i>lr_decay_mode</i>	string	learning rate decay function
<i>step_decay_rate</i> (S)	float	if <i>lr_decay_mode</i> is step_decay
<i>step_size</i> (E')	float	if <i>lr_decay_mode</i> is step_decay
<i>decay_rate</i> (k)	float	if <i>lr_decay_mode</i> is exp_decay

Table 5.1: This table gives an overview of the possible input parameter for the training routine.

name of the model, the path to save the trained model and the path to the training data set is given as input. Table 5.1 summarize the input parameter for the training routine.

Transfer-Learning

If a pre-trained SCNN is trained on the training data set, all convolutional blocks will be trained. Since the data set for pre-training (G6) consists of more than three classes and the classes differ from the ones in the training data set, the FCNN on top of the model is exchanged with $nclasses = 3$ and random weights initialization. All parameter listed in Table 5.1 are the same as for pre-training the network on the G6.

Fine-Tuning

For fine-tuning the trained SCNN, all convolutional layers are frozen except of the last one. Therefore, parameters in the frozen layers are untouched, while parameters in the last convolutional block and the FCNN are fine-tuned. The initial learning rate and the number of epochs for fine-tuning are reduced.

5.3 SCANNING AND DETECTING ALGORITHMS

We call the implemented custom scanning algorithm the nested boxes detection algorithm (NBDA). It scans the input image by defining a grid, centered on the image. On every grid position, a certain number of boxes with different sizes are nested into each other (Figure 5.5). Each nested box frames a region to search for bubble-like structures with a SCNN. The NBDA can be described in three parts:

In the first part, it defines a grid on the given input image of size $X \times Y$. The general structure of the grid is in rows, from left-to-right, i. e., in positive x -direction, and columns, from bottom-to-top, i. e., in positive y -direction. Each grid position is numbered, starting from the lower left (x_0, y_0) to the upper right (x_{EOG}, y_{EOG}) corner of the input image, with end-of-grid (EOG) the last grid position $(x_{EOG} < X, y_{EOG} < Y)$, which marks the end of the NBDA (Figure 5.5). This grid structure is referred to as an ϵ -shaped grid throughout this thesis. The first grid position is defined by the maximum box size $\Lambda = (\Lambda^x, \Lambda^y)$ of the nested boxes and the distance from the edge of the input image to the first grid position $\beta = (\beta^x, \beta^y)$ in x - and y -direction, respectively, and is calculated as

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} \beta^x + \frac{\Lambda^x}{2} \\ \beta^y + \frac{\Lambda^y}{2} \end{pmatrix}, \Lambda^{x,y}, \beta^{x,y} \in \mathbb{N}. \quad (40)$$

Therefore, the parameter β defines an area of the image, which is not covered by any box on any grid position. The distance between two neighboring grid positions (x_k, y_k) and (x_{k+1}, y_{k+1}) is called the grid size g . It is determined by the stride factor γ and the minimum box size of the nested boxes $\lambda = (\lambda^x, \lambda^y)$ in x - and y -direction, respectively, and is calculated as

$$g = \begin{pmatrix} g^x \\ g^y \end{pmatrix} = \begin{pmatrix} \gamma \cdot \lambda^x \\ \gamma \cdot \lambda^y \end{pmatrix} \quad (41)$$

with $\gamma > 0$ ($\gamma \in \mathbb{R}$) and $\lambda^{x,y} \in \mathbb{N}$. Starting from (x_0, y_0) , the NBDA moves the boxes on the grid. If (x_{EOG}, y_k) is reached, the algorithm moves one column up (y_{k+1}) and sets the x -coordinate back to the starting point (x_0). Therefore, the term ϵ -shaped grid. This procedure continues till the EOG position in x - and y -direction is reached. The entire algorithm is shown in Algorithm 1. The dimension of the grid can be calculated using the discussed parameters as

$$(G_x, G_y) = \left(\frac{X - (\lambda^x + 2 \cdot \beta^x)}{\gamma \cdot \lambda^x} + 1, \frac{Y - (\lambda^y + 2 \cdot \beta^y)}{\gamma \cdot \lambda^y} + 1 \right). \quad (42)$$

Assuming an image of size $(X, Y) = (100, 100)$, introducing a border around the image of $\beta^{x,y} = 5\text{px}$, a stride factor of $\gamma = 1$, a minimum and maximum box size of $\lambda^{x,y} = 10\text{px}$ and $\Lambda^{x,y} = 50\text{px}$, respectively, results in a 5×5 grid shown in Figure 5.5.

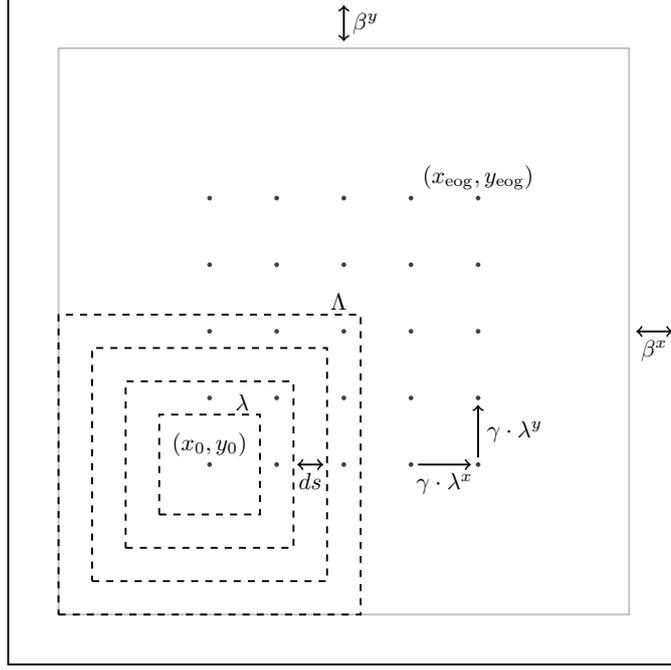


Figure 5.5: This Figure shows how the grid (points) is placed on an image and how the nested boxes (dashed) are defined at a grid position. The grid is centered taking the defined border $\beta^{x,y}$ and the maximum box size Λ into account. Its size is defined by the stride factor γ and the minimum box size $\lambda^{x,y}$ by $\gamma \cdot \beta^{x,y}$. From the first grid position (x_0, y_0) the nested boxes move on the grid as indicated by the arrows till the end position (x_ρ, y_ρ) is reached. The spacing ds between the nested boxes depends on the number of nested boxes and the minimum and maximum box size.

In the second part, the nested boxes are defined. There are n boxes nested into each other on each grid position (Figure 5.5). The box size varies from λ to Λ with a difference in size of

$$ds = \frac{\Lambda - \lambda}{n - 1}. \quad (43)$$

Every box which is used to cutout out the data from the input image at the current grid position with the shape of the defined box. Since the nested boxes are used to search for bubble-like structures with the SCNN on the corresponding grid position, the nested boxes are organized in batches. The batches are then used to predict bubble-like structures inside each batch. Algorithm 2 demonstrate how the nested boxes and, thus, a batch is created.

Part three makes use of the SCNN to predict if there is a bubble-like structure inside a batch or not. The batch element (nested box) with the highest prediction probability is then considered further. For this consideration the threshold (τ), the relative strength factor (RSF) and the prediction probabilities for each class (p_c) is of importance. The threshold is a first order criteria defining the mini-

Algorithm 1 Moving to the next grid position

Require: $(x, y), \gamma > 0 \wedge \gamma \in \mathbb{R}, \lambda \in \mathbb{N}, \Lambda \in \mathbb{N}, \beta \in \mathbb{N}$
 $g^x, g^y \leftarrow \gamma \cdot \lambda^x, \gamma \cdot \lambda^y$
if (x, y) within boundaries \wedge EOG = False **then**
 $nx = x \cdot g^x$
 if nx outside x -boundary **then**
 $nx = \beta^x + \frac{\Lambda^x}{2}$
 $ny = y \cdot g^y$
 if ny outside y -boundary **then return** $(0, 0)$, EOG \leftarrow True
 else return (nx, ny) , EOG \leftarrow False
 end if
 else return (nx, y) , EOG \leftarrow False
 end if
else return $(0, 0)$, EOG \leftarrow True
end if

Algorithm 2 Defining Nested Boxes

Require: $n \in \mathbb{N}, \lambda \in \mathbb{N}, \Lambda \in \mathbb{N}$
 $m \leftarrow 1$
if $\lambda = \Lambda$ **then** $n \leftarrow 1$
else
 $ds^x, ds^y = \frac{\Lambda^x - \lambda^x}{n-1}, \frac{\Lambda^y - \lambda^y}{n-1}$
end if
 $batch \leftarrow (\Lambda^x, \Lambda^y)$ \triangleright initialized as an array of shape $(n, 2)$
while $m < n$ **do**
 $batch \leftarrow (\Lambda^x - m \cdot ds^x, \Lambda^y - m \cdot ds^y)$
 $m \leftarrow m + 1$
end while return $batch$

imum probability needed to consider the prediction as a valid bubble-like structure. With the RSF, a secondary order criteria is introduced. For each batch element the relative strength index (RSI) is calculated, which can be considered as an effective probability that a bubble-like structure $(p_{shell}, p_{diffuse})$ relative to no bubble-like structure $(1 - (p_{shell} + p_{diffuse}))$ is detected. We define the RSI as

$$RSI = \frac{\max\{p_{shell}, p_{diffuse}\} - \min\{p_{shell}, p_{diffuse}\}}{1 - (p_{shell} + p_{diffuse})}. \quad (44)$$

Similar to the threshold, the RSF defines the minimum of the RSI in the considered batch element. If the RSF is set to zero, the secondary order criteria are fulfilled always. The RSF is therefore an optional criterion. The predicted bubble-like structure is considered as detected, if all active criteria are met. In that case, the extent is estimated according to the box size of the considered batch element. The output of Algorithm 3 is the name of the class, the prediction probability and the extent of the detected bubble-like structure.

Algorithm 3 Search for bubble-like structures in batch

Require: $p \in [0, 1] \wedge p \in \mathbb{R}, \tau \in [0, 1] \wedge \tau \in \mathbb{R}, \text{RSF} > 0 \wedge \text{RSF} \in \mathbb{R}, \text{batch}$

$p_{\max} \leftarrow$ calculate maximum probability over entire batch

if $\tau \leq p_{\max}$ **then**

$\text{RSI} \leftarrow$ calculate relative strength index at batch of p_{\max}

if $\text{RSF} \leq \text{RSI}$ **then**

$\text{extent} \leftarrow$ box size of the batch element with p_{\max}

$\text{class} \leftarrow$ predicted class

end if

end if return $\text{detection} \leftarrow \text{extent}, \text{class}, p_{\max}$

Algorithm 4 combines the different parts described above into one algorithm. It is one core element implemented in *BScan* and returns a list of detections. A single detection consists of the information about the grid position (x, y) , the extent (e_x, e_y) , the name of the class C , the detection probability $p_{\max} = p$ and a tag of the input image where the structure is detected. Each input image shows different emission. That is why the tag hints the detected emission coming from the detected bubble-like structure. While the position and the extent define a region framing the detected structure, the class, p and the tag characterize the structure even further. This information can be written to disk in form of a SAOImageDS9 (DS9) region file.

Algorithm 4 Nested Boxes Detection Algorithm

Require: $\text{image}, n \in \mathbb{N}, \gamma > 0 \wedge \gamma \in \mathbb{R}, \lambda \in \mathbb{N}, \Lambda \in \mathbb{N}, \beta \in \mathbb{N}$

$\text{EOG} \leftarrow \text{False}$

$(x, y) = (0, 0)$

while $\text{EOG} = \text{False}$ **do**

$(x, y), \text{EOG} \leftarrow$ Moving to next grid position

if $\text{EOG} = \text{True}$ **then break**

else

$\text{batch} \leftarrow$ Defining Nested Boxes

$\text{detection} \leftarrow$ Search for bubble-like structures in batch

end if

end while return detection

5.4 CLUSTERING AND MERGING

With the NBDA described in 5.3, multiple detections of the same structure is possible. Depending on the input parameters, the grid size g varies. The smaller the grid size the more multiple detections can occur due to overlapping boxes on the grid. To reduce multiple detections and avoid overlapping regions detecting the same structure, we cluster and merge the results after scanning the entire input image.

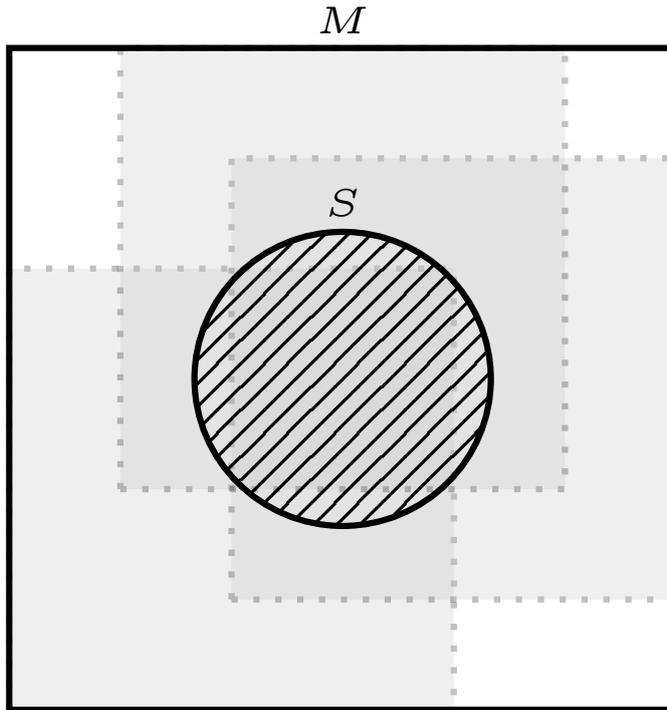


Figure 5.6: If there are several single detections at different positions (dashed, light grey) detecting the same source (S), the single detections are clustered and merged (M). The merged cluster is then considered as final detection of the source.

For defining a cluster, the k -nearest neighbor (kNN) algorithm by Maneewongvatana et al. (1999) is used. This algorithm is implemented into the Python package SciPy as the `KDTree` class¹ and provides a searchable k -dimensional tree (Bentley, 1975) to find members m of a cluster efficiently.

Assigning a list of coordinates (x, y) and radii (r) of detected bubble-like structures, the cluster algorithm (Algorithm 5) groups all structures within a maximum distance of $R = f \cdot r$ to one cluster, but only if the corresponding bubble-like structures are predicted as the same class. This way, detected structures of different classes can still be detected at nearby grid positions. The factor f controls the maximum distance R between cluster members and, therefore, the level of overlap of detected bubble-like structures. As a default value we use $f = 1$ and, thus, all bubble-like structures with a maximum distance of $R = r$ are clustered by default.

After finding the members of a cluster, each cluster is merged into one region (Figure 5.6). The merged region is then considered as the final detection of a bubble-like structure. The merging algorithm takes

¹ <https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.KDTree.html>

Algorithm 5 Cluster Detections Algorithm

Require: list of detections (coordinates, extend, class), $f > 0 \wedge f \in \mathbb{R}$
 $K \leftarrow$ apply kNN-algorithm with $f \cdot \frac{\text{extend}}{2}$
for cluster in K **do**
 for member in cluster **do**
 if class (member) = shell **then**
 $S \leftarrow$ member, coordinates, extend, class, $p(\text{member})$
 else if class (member) = diffuse **then**
 $D \leftarrow$ member, coordinates, extend, class, $p(\text{member})$
 else
 continue
 end if
 end for
end for return concatenated S and D

the output of Algorithm 5 and calculates the mean probability over all cluster members

$$\bar{p} = \frac{1}{N_m} \sum_m p_m \quad (45)$$

where N_m is the number of members inside the cluster and p_m is the probability of the considered member to be of class *shell* or *diffuse*. We estimate the final coordinates by the weighted mean of the member coordinates (x^m, y^m) with p_m as the weights

$$(\bar{x}, \bar{y}) = \left(\frac{\sum_m p_m \cdot x^m}{\sum_m p_m}, \frac{\sum_m p_m \cdot y^m}{\sum_m p_m} \right). \quad (46)$$

Similar to the coordinates, we estimate the final extent of the bubble-like structure by the weighted mean of the extent of the cluster members

$$(\bar{e}_x, \bar{e}_y) = \left(\frac{\sum_m p_m \cdot e_x^m}{\sum_m p_m}, \frac{\sum_m p_m \cdot e_y^m}{\sum_m p_m} \right) \quad (47)$$

with $e_{x,y}^m$ the extent of the corresponding member of the cluster in x - and y -direction, respectively. These characteristics are then written to a DS9 region file.

5.5 TRAINING AND TESTING

In this section, we train networks of different depths on the training data sets. To find the best performing network on detecting bubble-like structures in astronomical images, the trained networks are tested (validated) on the test data sets as well. In addition, the best performing network is used together with the NBDA to test the network on 30 test images to determine the NBDA default values.

5.5.1 *Random Parameter Initialization*

We start with randomly initialized networks of type SCNN $_{ij}$ (see section 5.1.3) with different convolutional block numbers $i \in [3, 9]$ and convolutional layers per block $j \in [1, 3]$. Those SCNNs have a different amount of trainable parameters, which are trained on the training data set (TRS). The number of trainable parameters of these networks range from around $50 \cdot 10^3$ to more than $15 \cdot 10^6$. In comparison, the VGG16 has $\sim 138 \cdot 10^6$ trainable parameters (Simonyan et al., 2014).

The training parameters are the same for each network. This leads to a better comparability. Table 5.2 summarize the parameters needed for the training procedure and the corresponding value.

Figure 5.7 (a) and (b) show the history of the training progress for the network SCNN $_{71}$ -TRS. The SCNN $_{71}$ -TRS is the best performing network in the set of trained SCNNs with randomly initialized parameters (see Table 5.3). The precision of the SCNN $_{71}$ -TRS is steadily

Parameter	value
<i>nclasses</i>	3
<i>epochs</i>	250
<i>batch_size</i>	32
<i>split_factor</i>	0.1
<i>optimizer</i>	Adam
<i>learning_rate</i>	0.001
<i>lr_decay_mode</i>	none

Table 5.2: Training parameters used to train the networks SCNN $_{ij}$ -TRS with $i \in [3, 9]$ and $j \in [1, 3]$.

increasing during training (Figure 5.7, a), whereas the loss decreases with each epoch. Therefore, no overfitting is observed. A similar behavior is shown for validation (Figure 5.7, b), but larger fluctuations are evident. These fluctuations become smaller with epochs. The maximum precision achieved for the SCNN $_{71}$ -TRS is 0.98 and 0.99 for training and validation, respectively. Due to the larger precision fluctuations during validation, the slightly higher maximum precision for validation can be explained. The training history shown in Figure 5.7 is recorded for all trained SCNNs. We compare the maximum precision reached during training to decide which SCNN is the best performing network.

This comparison is done in Figure 5.8. Networks with one convolutional layer per block ($j = 1$) reach a higher maximum performance, except for the SCNN $_{31}$ -TRS, which consists of only three blocks ($i = 3$). The maximum precision reaches the highest values for networks with $i \in \{6, 7, 8\}$. Those perform similar during train-

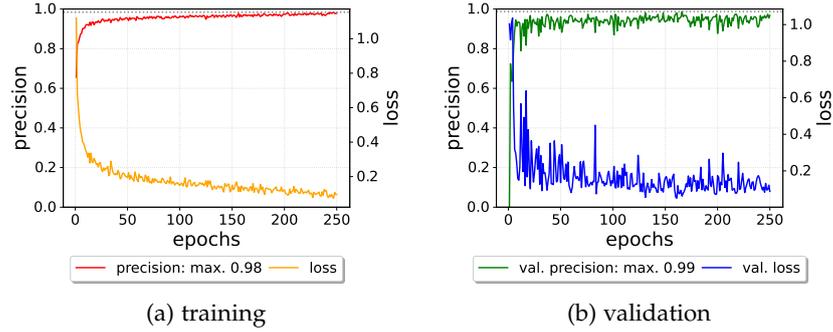


Figure 5.7: Training history of the SCNN71-TRS. (a) During training, the precision (red) is increasing and the loss (orange) is decreasing steadily with epochs. The maximum precision is 0.98. There is no sign of overfitting. (b) The validation of the training progress shows similar behavior but bigger fluctuations in precision (green) and loss (blue). For validation the precision reaches a maximum of 0.99.

ing. In contrast, networks with $j = 3$ have a tendency towards lower maximum precision.

To test the performance after 250 epochs of training, the SCNNs are applied to the test data set (TES) and the augmented data sets of faint (TES-F) and low-contrast (TES-LC) samples. A confusion matrix (CM) is calculated to visualize the performance on these test sets. The CM correlates the true labels, given to the network, with the labels predicted by the network (Figure 5.9). The diagonal shows the true positive (TP) and true negative (TN) predictions. Each row of the CM sums up to 100%. If the diagonals display 100%, it means that all samples in the test set are predicted properly with the corresponding label. On the off diagonals the false positive (FP) and false negative (FN) predictions are represented. If off elements are larger than zero, the network is confused and not able to predict the test samples completely. The FN prediction is especially problematic, because a bubble-like structure would be categorized as a no-bubble-like structure. This would certainly lead to missing out of a bubble-like structure for further analysis. In contrast, with a FP prediction, the network would still recognize the bubble-like structure, but instead of categorizing the diffuse bubble-like structure into the correct class *diffuse*, for example, it would be categorized as a *shell*. At this stage, *BScan* focuses on detecting bubble-like structures, the classes *diffuse* and *shell* are introduced for further applications in the future. Therefore, FP predictions are not as significant as FN predictions for this thesis. Testing each network as described above is resulting in several CMs. To quantitatively describe the best performing SCNN, we calculate the weighted

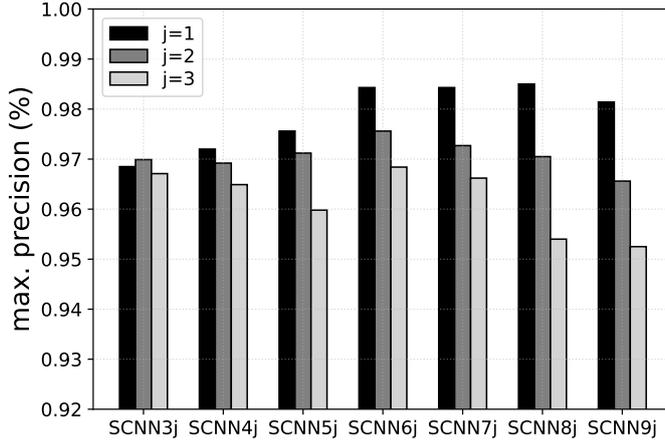


Figure 5.8: This plot shows the maximum precision for the trained networks of type SCNN i j-TRS (number of convolutional blocks $i \in [3, 9]$, number of convolutional layers within a block $j \in [1, 3]$). The maximum precision forms a plateau for networks with $i \in [6, 8] \wedge j = 1$. For $j = 2$ the maximum precision reaches its peak at $i = 6$, while for $j = 3$ and larger i the tendency is toward lower maximum precision. Just looking at the precision the SCNN6 and SCNN7 perform similar during training.

mean of TP predictions over all classes and test sets. The weighted mean performance Π is defined as

$$\Pi = \frac{\sum_t \omega_t \sum_m^C CM_{mm}^t}{C \sum_t \omega_t}, \quad (48)$$

where ω_t is the weight for the corresponding test set t , $C = 3$ is the number of considered classes and CM is the confusion matrix of each test set. Results are shown in Table 5.3 for the four best performing SCNNs. We choose the weights $\omega_{\text{TES}} = 1.00$, $\omega_{\text{TES-F}} = 0.75$ and $\omega_{\text{TES-LC}} = 0.50$ to take into account that the networks are trained on the training data set TRS only and a lower performance on the test sets TES-F and TES-LC is therefore expected. According to this performance measure, the SCNN 7_1 -TRS is the best performing network with a weighted mean performance of

$$\Pi(\text{SCNN}_{7_1}\text{-TRS}) = 0.8. \quad (49)$$

As demonstrated in Figure 5.9, the SCNN 7_1 -TRS performs relatively well on the TES (a). While the network is confused between the classes *diffuse* and *shell*, there are no FN predictions. In numbers, this translates to 10% of samples labeled as *diffuse* are predicted as shells and 13% of shells are predicted as *diffuse*. The test results getting worse if the SCNN 7_1 -TRS is tested on the TES-F (Figure 5.9 b). While the confusion between the two classes *diffuse* and *shell* is nearly the

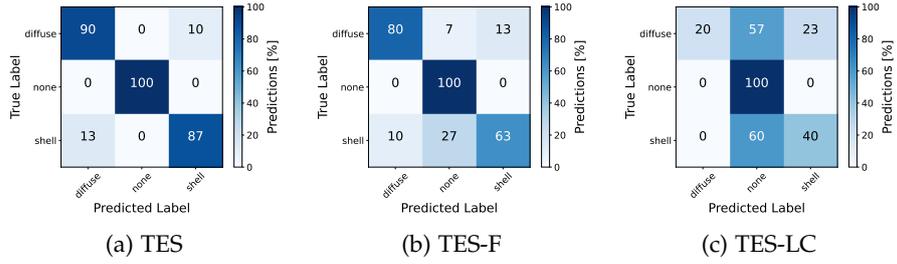


Figure 5.9: These Figures show the confusion matrices for the trained SCNN71-TRS tested on the test data set (a), the test data set for faint (b) and the test data set for low contrast (c) samples. While the SCNN71-TRS is confident in classifying the class *none*, in case (a), it is confused between the classes *shell* and *diffuse*. Compared to (a), the SCNN71-TRS gets increasingly confused in case (b), not only between *diffuse* and shells, it classifies bubble-like structures into class *none* as well. The result for (c) is highly confused.

Nr.	network	TES [1.00]	TES-F [0.75]	TES-LC [0.50]	Π
1	SCNN71-TRS	0.92	0.81	0.53	0.80
2	SCNN51-TRS	0.93	0.85	0.42	0.79
3	SCNN81-TRS	0.92	0.81	0.48	0.79
4	SCNN92-TRS	0.88	0.80	0.56	0.78

Table 5.3: This table shows the top four performing networks on the test data set (TES), the augmented data set for faint (TES-F) and low-contrast (TES-LC) samples considering the weighted mean of true positive predictions over all classes and test sets. For each test set, the mean fraction of true positive predictions over all classes are shown. The weighted mean performance Π is calculated using the given weights in square brackets.

same compared to Figure 5.9 (a), the confusion between no-bubble-like structures (*none*) and bubble-like structures (*diffuse*, *shell*) is increasing. Around 11% of all predictions are FN predictions. These FN predictions originate from predicting 7% of *diffuse* and 27% of shells as class *none*. In fact, the confusion on the TES-LC (Figure 5.9 c) especially the FN predictions getting worse. In total, 39% of all bubble-like samples in the test data set are predicted as no-bubble-like structures (57% of class *diffuse* and 60% of class *shell*).

According to this test results, the SCNN71-TRS is able to predict well-prepared samples present in the TES but cannot handle faint or low-contrast samples as present in the TES-F or TES-LC, respectively. Therefore, the SCNN71-TRS will be fine-tuned on the training data sets for faint (TRS-F) and low contrast (TRS-LC) samples later on.

5.5.2 Pre-Training of the SCNN71

The best performing network architecture in the previous section 5.5.1, the SCNN71, is pre-trained on the reduced Galaxy10 data set (G6) to improve the training on the TRS further. The resulting network SCNN71-G6-TRS will then be compared to the SCNN71-TRS. Both networks will undergo the same testing procedure illustrated in section 5.5.1. We summarize the results in Table 5.4.

The history of the pre-training is shown in Figure 5.10. The training (Figure 5.10, a) and validation (Figure 5.10, b) precision reach similar maxima of 0.87 and 0.88, respectively. But both precision values are relatively small. One possible reason for the low precision could be found in the fact that the architecture focuses on a tiny data set, therefore, it has only a few parameters and therefore lacks of the capability to perform well on a bigger data set like the G6 (see section 2.1.7). Although the precision is increasing steadily, the slope becomes significantly smaller after around 100 epochs, making it computationally expensive to improve and inefficient to train for longer. The loss show again no sign of overfitting during training.

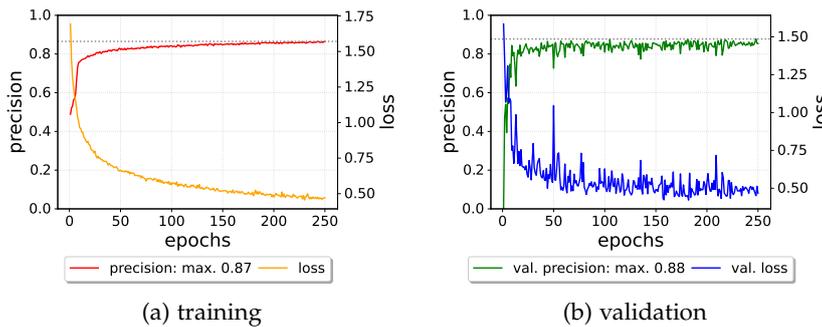


Figure 5.10: The Training history of the SCNN71-G6. (a) The training progress with the measured precision (red) and the loss (orange). After 250 epochs the maximum precision is 0.87. The measured loss decreases steadily, hence, no overfitting is observed while training. (b) Validation of the network paint a similar behavior in precision (green) and loss (blue) but with bigger fluctuations which become smaller with epochs.

Testing the SCNN71-G6 on a sub data set of G6 leads to the confusion matrix (CM) shown in Figure 5.11. It reflects the low precision reached during training in the low fractions of TP predictions on the diagonal. The most accurate predictions are made for smooth, completely round galaxies (see Table 4.2, class label B) and disk-like galaxies with tight spiral arms observed face on (see Table 4.2, class label D) with 97% and 90% of TP predictions, respectively. For all other classes the SCNN71-G6 is more confused with less than 90% of TP predictions. Especially for class A and E, the face-on disk galaxies without and medium spiral arms, the network is heavily confused. A look into

the samples of these classes reveal possible reason related to the data set. In case of class A, samples have a characteristic bright center surrounded by a diffuse disk-like emission of the galaxy. This features can also be found in some, but not all, samples of class B and C. The bright center feature is also present in galaxies with spiral arms. This could be an explanation why the SCNN₇₁-G6 predicts class A for samples over all classes. In case of samples with medium spiral arms (class label E), the network is mainly confused by classes containing spiral arms samples like in D and F. The purpose of pre-training the

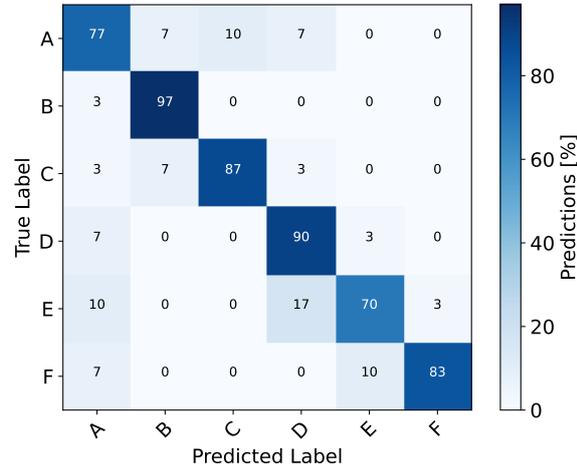


Figure 5.11: Confusion matrix for the pre-training of the SCNN₇₁ architecture on the G6 data set. The most accurate predictions are for smooth, completely round galaxies (class label B) with 97%, followed by the disk-like galaxies with tight spiral arms observed face on (class label D) with 90%, of true positive predictions. For all other classes the network is confused on a higher level with less than 90% correct predictions.

SCNN₇₁ architecture on the G6 is to initialize the trainable parameter to avoid random initialization like in section 5.5.1. Although the pre-training is not performing well considering the achieved maximum precision and the resulting CM, we consider the initial goal as fulfilled. In the next step the network is trained on the TRS.

Comparing the training history of SCNN₇₁-TRS (Figure 5.7) with the one of SCNN₇₁-G6-TRS in Figure 5.12 (a) progress is similar. The precision slope is small after 100 epochs but reaches already values close to one. A steadily decreasing training loss shows no overfitting. For the validation of the SCNN₇₁-G6-TRS (Figure 5.12, b) the precision and the loss is almost constant over all epochs. The loss shows an ascending trend after 200 epochs, which could be a hint for overfitting, but the loss values are relatively small, however. In addition, the training routine saves the network to disk, only if the loss of the current valuation is lower compared to previous losses preventing the trained network from overfitting. Consequently, the status of the

pre-trained SCNN71-G6-TRS, which is used in the following, is saved at epoch 195 with a validation loss of ~ 0.04 and a precision of ~ 0.98 .

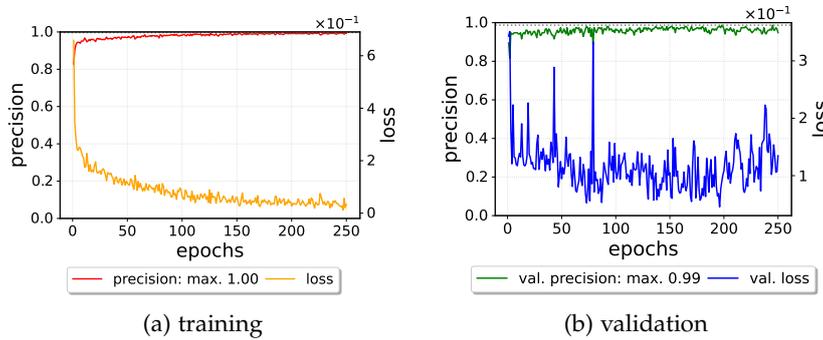


Figure 5.12: The Training history of the SCNN71-G6-TRS. (a) The precision (red) and loss (orange) over epochs. There is no sign of overfitting while training, since the loss is decreasing steadily. (b) The validation of the training progress shows a slightly different picture. While the precision (green) is also slightly increasing, the loss (blue) seems to have an ascending trend after 200 epochs. However, the loss values over all epochs are small and can be considered as nearly constant.

Applying the SCNN71-G6-TRS to the TES, TES-F and TES-LC results in the CMs in Figure 5.12. Although the mean ratio of true positive predictions on the TES is the same for both networks, the SCNN71-G6-TRS and the SCNN71-TRS (see Table 5.4), it is worth to have a closer look to the predictions made by the networks in comparison. The TP predictions for the class *shell* increase from 87% predicted by the SCNN71-TRS, to 97% predicted by the SCNN71-G6-TRS. The ratio of TP predictions, however, decreases from 90% to 87% for the SCNN71-TRS to the SCNN71-G6-TRS. Overall the gain of TP predictions is larger than the loss. This test result is a significant improvement. The pre-trained network SCNN-G6-TRS predicts slightly more bubble-like structures compared to the SCNN71-TRS. This could be considered as a positive test results because more detections of bubble-like structures could lead to more SNR candidates, for example. But this comes with a cost, namely the smaller fraction of TN predictions which has decreased from 100% to 93% of no-bubble-like structures (*none*) predicted correctly. Additional bubble-like structures predicted by the SCNN71-G6-TRS come most likely from filaments represented in class *none* and fall into the category FP predictions. Testing the SCNN71-G6-TRS on the TES-F (Figure 5.12 b) reveals a significant improvement of predicting bubble-like structures of class *shell*. The ratio of true positives increases from 63% to 80% of correctly predicted shells in the TES-F. The ratios for the class *diffuse* and *none* are the same for both networks. Consequently, the FP predictions are less present if applying the SCNN71-G6-TRS to faint bubble-like structures com-

pared to applying the SCNN₇₁-TRS. The opposite is true for applying the SCNN₇₁-G6-TRS to the TES-LC samples (Figure 5.12 c). While the ratio of true positives and true negatives stay the same for the class *diffuse* and *none*, the true positives for the class *shell* decreases, leading to more FP predictions.

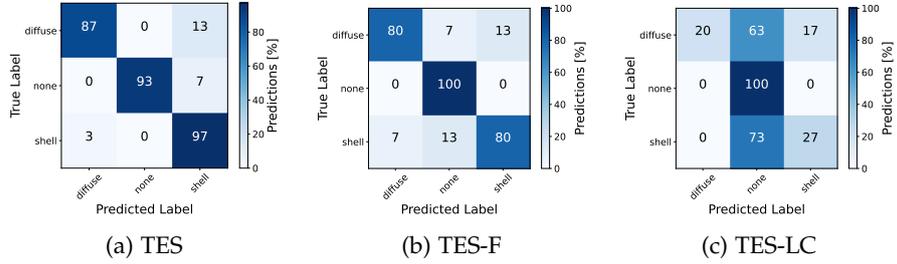


Figure 5.13: The confusion matrices for the trained SCNN₇₁-G6-TRS tested on the test data set (a), the data set for faint (b) and low contrast (c) samples. Compared to the SCNN₇₁-TRS, which is not pre-trained, the test performance of the SCNN₇₁-G6-TRS most improved on the TES-F but becomes worse on the TES-LC. On the TES the SCNN₇₁-G6-TRS predicts some no-bubble-like structures, most likely filaments, as shells. This is not the case for the SCNN₇₁-TRS.

We summarize the discussion above in Table 5.4, where the mean ratio of the TP predictions for each test set and the weighted mean performance Π is calculated for each network. Applying the same weights as in section 5.5.1 results in

$$\Pi(\text{SCNN}_{71}\text{-G6-TRS}) = 0.81 \quad (50)$$

for the SCNN₇₁-G6-TRS. This means an improvement of the weighted mean performance of $\Delta\Pi = 0.1$ (compared to equation 49). A closer look to the mean ratios of TP predictions reveal a larger gain on faint than the loss on low contrast samples. Having this in mind, we consider the pre-trained network SCNN₇₁-G6-TRS as the better performing network.

5.5.3 Fine-Tuning

The pre-trained network SCNN₇₁-G6-TRS is fine-tuned on the corresponding TRS-F and/or TRS-LC in the following. If the network is fine-tuned on both augmented data sets, the name of the SCNN has the label FLC at the end. The fine-tuned networks are tested on the test data sets, and the weighted mean performance is calculated as before.

We freeze all convolutional blocks for the fine-tuning process. Consequently, the only part of the network which is affected while fine-

Nr.	network	TES [1.00]	TES-F [0.75]	TES-LC [0.50]	Π
1	SCNN ₇₁ -G6-TRS	0.92	0.87	0.49	0.81
2	SCNN ₇₁ -TRS	0.92	0.81	0.53	0.80

Table 5.4: This table shows the SCNN₇₁ architecture with and without pre-training in comparison. Both versions are tested on the test data set (TES), the augmented data set for faint (TES-F) and low-contrast (TES-LC) samples considering the weighted mean performance as shown in Table 5.3. The pre-trained version has a slightly better Π with $\Delta\Pi = 0.1$. This small increase in performance comes from the better performance on the TES-F.

tuning is the FCNN at the end of the SCNN. Table 5.5 summarizes the parameters used for fine-tuning. Compared to the training parameters in Table 5.2, changes have been made in order to preserve the training progress of the previous sections: To avoid large changes to the parameters of the FCNN we half the learning rate to $LR = 0.0005$. In addition, we double the batch size to 64 reducing the amount of parameter updates during fine-tuning. Since it is shown that a number of between 100 and 125 epochs are sufficient for training and pre-training, we reduce the number of epochs for fine-tuning to $E = 125$. This saves computational time.

Parameter	value
<i>nclasses</i>	3
<i>epochs</i>	125
<i>batch_size</i>	64
<i>split_factor</i>	0.1
<i>optimizer</i>	Adam
<i>learning_rate</i>	0.0005
<i>lr_decay_mode</i>	none

Table 5.5: Parameters used to fine-tune the SCNN₇₁-G6-TRS.

Calculating the weighted mean performance Π from the calculated confusion matrices results in the values given in Table 5.6 for the best three fine-tuned networks. For better comparison the same weights are used as in section 5.5.1 and 5.5.2. The network SCNN₇₁-G6-TRS-F has improved ratios of TP predictions on all test data sets except on the TES-F compared to the SCNN₇₁-G6-TRS. This is surprising given the fact that it is fine-tuned on the TRS-F. The largest improvement is seen on the test data set for low contrast samples TES-LC. On this test set the ratio of true positives has improved by about 39% (compare Table 5.4). Measuring the weighted mean performance according to

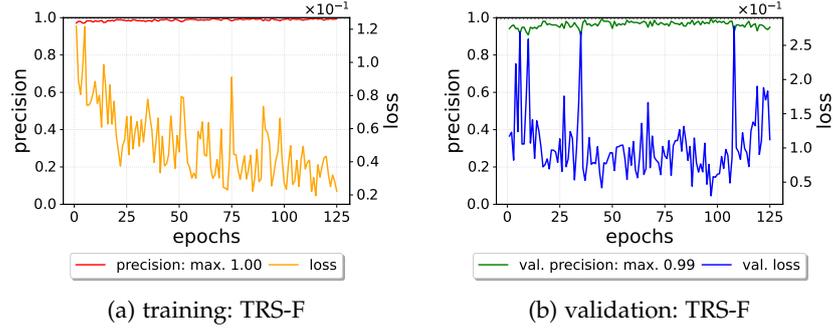


Figure 5.14: The Training history of the SCNN₇₁-G6-TRS-F recorded for training and validation on the data set for faint samples TRS-F used for fine-tuning. The precision (red) and loss (orange) for training re shown in (a). For validation the precision (green) and loss (blue) are shown in (b).

equation 48 the SCNN₇₁-G6-TRS-F is the best performing network with

$$\Pi(\text{SCNN-71-G6-TRS-F}) = 0.85. \quad (51)$$

Its weighted mean performance improved by about 5% compared to the SCNN₇₁-G6-TRS (equation 49). The history of the fine-tuning progress is shown in Figure 5.14.

Nr.	network	TES [1.00]	TES-F [0.75]	TES-LC [0.50]	Π
1	SCNN ₇₁ -G6-TRS-F	0.93	0.84	0.68	0.85
2	SCNN ₇₁ -G6-TRS-LC	0.91	0.83	0.53	0.80
3	SCNN ₇₁ -G6-TRS-FLC	0.88	0.81	0.58	0.79

Table 5.6: Comparison of the fine-tuned versions of the SCNN₇₁-G6-TRS network. The best weighted mean performance is reached by fine-tuning the network on faint training samples. For a better comparability the same weights are used as for pre-training.

Figure 5.15 shows the CMs for the SCNN₇₁-G6-TRS-F applied to all three test data sets as done before. The CM for the TES reveals some confusion between the class *diffuse* and *shell* and, therefore, some FP predictions but no FN predictions. The test performance did not improve on the TES-F which is also seen in the CM in Figure 5.15 (b) compared to Figure 5.12 (b). The ratio of false negatives stayed the same but predictions of type true positives decreased significantly for class *shell*. A different situation is shown in the CM of Figure 5.15 (c), which is the result of applying the SCNN₇₁-G6-TRS-F on the TES-LC. In comparison to Figure 5.12 (c), the CM of the SCNN₇₁-G6-TRS applied on

the TES-LC, false negative predictions are reduced significantly from about 45% to only 19% of all samples predicted as a no-bubble-like structure, while the sample show actually a bubble-like structure.

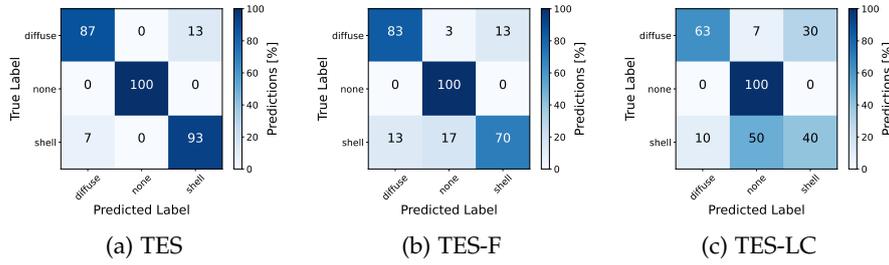


Figure 5.15: The figure shows the confusion matrices of the fine-tuned network SCNN71-G6-TRS-F applied on the test data sets TES (a), TES-F (b) and TES-LC (c). Compared to the confusion matrices of the pre-trained network SCNN71-G6-TRS the biggest improvement is shown in (c).

5.6 DETERMINE DEFAULT VALUES FOR THE NBDA

In this section we apply the best performing network SCNN71-G6-TRS-F with the NBDA on a set of 30 astronomical images. For each wavelength regime (radio, optical and X-ray) ten images are used showing different bubble-like structures of class *diffuse* and *shell* near the center. The central structure can be surrounded by filaments but no additional bubble-like structure for this test. The images are of the size (1000×1000) px. A grid is defined on top of each image using the NBDA. It is defined in a way that the bubble-like structure is not centered on any grid position. An example is presented in Figure 5.16.

Applying the NBDA on each of the 30 images with the same grid definition results in 30 possible detections per grid position. Since the bubble-like structures are located near the center of each image, most detections are expected for grid positions in the central region. Counting and visualizing the detections per grid position in a 2D histogram helps to determine default values of the NBDA parameters by focusing on the central bubble-like structure and assuming detections in the outer regions are false positives. Summing up the detections per grid position in x- and y-direction results in a distribution for each axis with an expected peak in the center of the grid. A sharp peak with low tail is preferred in order to estimate the location and the extent of the source later on. This approach is considered as a more realistic test of the network together with the NBDA.

By changing a single parameter of the NBDA (see section 5.3) while keeping all other parameters the same and apply it to the set of 30 images, the 2D histograms can be used to estimate the default values of

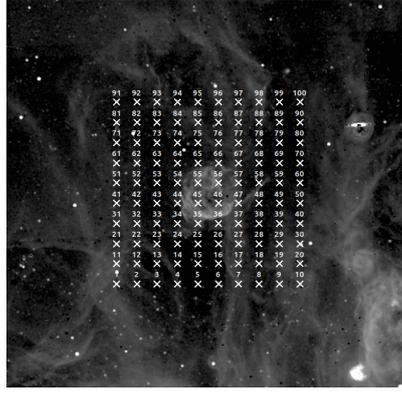


Figure 5.16: An image with a bubble-like structure of class *shell* with an image size of 1000×1000 px. The defined grid is using a minimum and maximum box size $\lambda = 50$ px and $\Lambda = 500$ px, respectively, and $\beta = 25$. The stride factor is set to one. Therefore, the bubble-like structure in the center of the image is not centered on any grid position.

the NBDA parameters. The minimum and maximum values of the parameters used in this section are summarized in Table 5.7. Note that the minimum and maximum box size $\lambda = 50$ px and $\Lambda = 500$ px, respectively, is fixed and depends on the expected extents of the bubble-like structures in the input image. The border $\beta = 25$ px is also fixed and exclude certain areas at the edge of the input image (see Figure 5.5).

Description	parameter	min.	max.
stride factor	γ	0.5	1
number of nested boxes	n	5	15
threshold	τ	0	1
Relative Strength Factor	RSF	0	10^8

Table 5.7: Summary of the minimum and maximum values to find the default values of the NBDA parameters. Considered parameters are the stride factor (γ), the number of nested boxes (n), the threshold (τ) and the relative strength factor (RSI).

5.6.1 Stride Factor

The stride factor γ is the first parameter which we investigate. It is varied between values of $\gamma = 0.5$ and $\gamma = 1.0$. Values $\gamma < 0.5$ are excluded because γ determines the grid size. A smaller grid size means a higher density of grid positions. Thus, with a smaller grid size the likelihood of detecting the same bubble-like structure multiple times increases. For stride factors $\gamma > 1$ the grid size becomes larger than the minimum nested box λ and, therefore, smaller bubble-like struc-

tures may not be detected. This is especially true for those of similar extent as λ . Also bubble-like structures located between neighboring grid positions could be missing with stride factor $\gamma > 1$.

Figure 5.17 shows the result of varying the stride factor with $\Delta\gamma = 0.25$ but keep the number of nested boxes $n = 10$, the threshold $\tau = 0$ and the RSF = 0. For smaller stride factors multiple detections are evident. Comparing Figure 5.17 (a) and (c), for example, the distribution maxima in x - and y -direction increases from around 100 for a stride factor of $\gamma = 1$ to around 200 for a stride factor of $\gamma = 0.5$, while the maximum number of detections per grid position stays nearly the same at around 25. This means the same bubble-like structure is detected at neighboring grid positions. Due to the clustering and merging algorithm the multiple detections are not as problematic. The choice of the right stride factor is therefore not as important as other parameters, as shown later, but a question of computational resources. A smaller stride factor means more computational time. We set the default value therefore to

$$\gamma_D = 1.0. \quad (52)$$

In the following the default value is used, which results in a 10 grid on the test images.

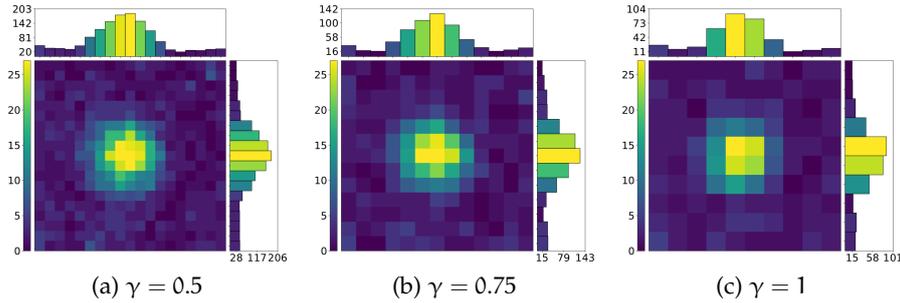


Figure 5.17: The 2D histograms to compare the NBDA results on the 30 test images for different stride factors: (a) $\gamma = 0.5$ (b) $\gamma = 0.75$ and (c) $\gamma = 1$

5.6.2 Number of Nested Boxes

The second parameter we study is the number of nested boxes n . Applying the NBDA on the test images with the stated input parameters in section 5.6.1 but changing n with $\Delta n = 5$ results in the 2D histograms in Figure 5.18. Except slightly smaller numbers in Figure 5.18 (a) compared to (b) and (c), no significant changes in the x - and y -distributions and the number of detections per grid position can be found. This is expected, since the NBDA allows exactly one detection per grid position. A larger number of nested boxes means a smaller difference of the box sizes (see Δs in Figure 5.5). This may play a role

for estimating the extent of the detected bubble-like structure (see section 5.4).

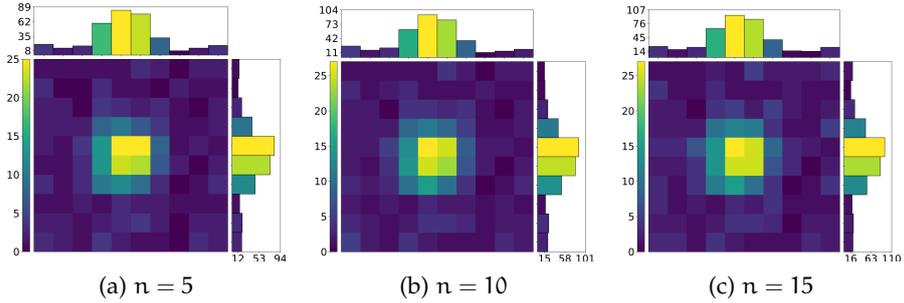


Figure 5.18: Resulting 2D histograms using different numbers of nested boxes $n = 5$ (a), $n = 10$ (b) and $n = 15$ (c) on the test images. The number of detections per grid position and the distribution in x- and y-direction are similar in all three cases.

In order to determine the default value for the number of nested boxes n , we apply the NBDA on the 30 test images again but with a 1×1 grid centered on the bubble-like structure. This way, the detection of the bubble-like structure depends only on the box size. In addition, the tested n is extended to $n \in \{5, 10, 15, 20, 25\}$. The resulting detections on the single grid position are shown in Table 5.8. In total 30 detections are possible, since there are thirty test images with a defined 1×1 grid. For $n \in \{10, 15, 20, 25\}$ there are 27 bubble-like structures detected. The three missing structures are of class *shell* in the X-ray wavelength regime. Note that the training samples for X-ray are far less compared to other wavelengths and mostly of class *diffuse*. Therefore, samples of class *shell* in X-rays are underrepresented in the training data set. This could explain the three missing detections. For $n = 5$ additional bubble-like structures in the optical regime are not detected, namely one of class *diffuse* and one of class *shell*. In this case the size of the nested boxes may not match the size of the bubble-like structure. Consequently, the structure size may not fulfill the criteria of being roughly 80% of the box size used in chapter 4. This could explain the additional missing detections. Because of the results in Figure 5.18 and Table 5.8, we use

$$n_D = 10 \quad (53)$$

as the default value. It is the lowest number of nested boxes which detects the highest number of central sources in the test images. The determined default value is used from now on.

After testing the NBDA with different n , it has to be stated, that the default value $n = 10$ cannot be considered as the best value for every image. It highly depends on the size distribution of bubble-like structures. The larger the difference in size the finer the division of

n	detections
5	25
10	27
15	27
20	27
25	27

Table 5.8: This table summarizes the number of detected bubble-like structures in the test images using a 1×1 grid with different numbers of nested boxes $n \in \{5, 10, 15, 20, 25\}$. The maximum number of detections is reached the first time with $n = 10$.

the box sizes of the nested boxes. Keep in mind that a larger n comes with a higher computational cost.

5.6.3 *Threshold*

The threshold is a primary parameter of the NBDA, since it sets the minimum of the output of the network and, therefore, a lower limit of a prediction of a bubble-like structure that can be considered as detected. It is of importance to find a proper default value which is not too low, to avoid low confidence detections, and not too high, to miss detections which have a lower probability than one but are still detections of high confidence. We show three meaningful results in Figure 5.19, chosen from applying the NBDA to the 30 test images with threshold values of $\tau \in \{0, 0.5, 0.75, 0.8, 0.9, 0.95, 0.99\}$. For the thresholds from $\tau = 0$ to $\tau = 0.9$ the maximum detections per grid position in the center are nearly the same at ~ 25 , while the number of detections in the outer regions are decreasing. This means the tails of the distribution in x - and y -direction show lower numbers, resulting in a sharper peak near the grid center. Consequently, the detection of the central bubble-like structure improves. For thresholds $\tau > 0.9$ this improvement is not as significant, but the maximum detections per grid position in the central region decreases to around 18, which indicates that detections of the central bubble-like structure are rejected because of a high threshold. According to this analysis we set the default threshold to

$$\tau_D = 0.9. \quad (54)$$

5.6.4 *Relative Strength Factor*

The relative strength factor (RSF) is an optional, and therefore secondary parameter which is similar to a threshold (see section 5.3). It

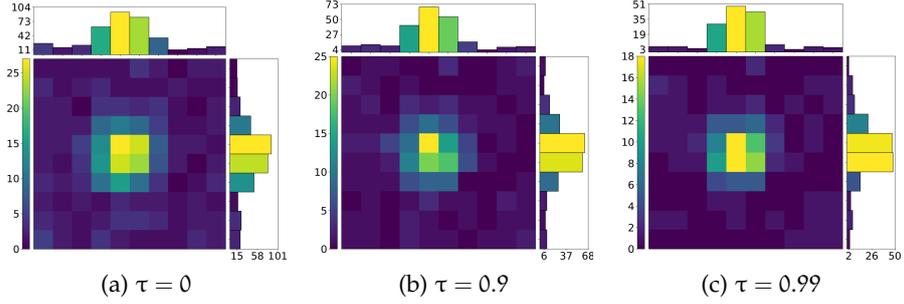


Figure 5.19: This Figure shows three meaningful results of the NBDA varying the threshold from $\tau = 0$ up to $\tau = 0.99$. From $\tau = 0$ to $\tau = 0.9$ the detection quality of the central bubble-like structure improves because detections in the outer regions decrease while the detections per grid position stay nearly the same. For $\tau > 0.9$ some detection of the central structure is rejected because of a high threshold. This leads to lower detection per grid position in the central region. The default value is set to $\tau = 0.9$.

sets the lower limit of the RSI calculated in equation 44. The goal of this part of the analysis is to improve the detection quality of the central bubble-like structure further. That means reducing the detections in the outer regions while keeping the peak of the detection distribution in x - and y -direction nearly at the maximum of 27 detections (from Table 5.8). This should improve the precision of the estimated coordinates as well. The NBDA is applied with the default values determined so far and RSF values of $\text{RSF} = 10^k$ with $k \in [1, 8]$. The values for the RSF are arbitrary but fulfill the goal of this analysis. Figure 5.20 shows the 2D histograms for $\text{RSF} = 10$ (a), $\text{RSF} = 10^5$ (b) and $\text{RSF} = 10^8$ (c). The results shown in Figure 5.20 (a) show similar results compared to Figure 5.19 (b) meaning low values for the RSF have a small impact on the detection quality. With much higher values (Figure 5.20 b) the detections in the outer regions become less while the total number of detections near the center are about 27. Increasing the RSF even further reduces the number of detections in the outer regions but rejects detections from the central region and reduces the number of bubble-like structures in the center to around 15. Consequently, we set the default value for the RSF to

$$\text{RSF}_D = 10^5. \quad (55)$$

Because of arbitrary large numbers for the RSF, it is hard to find the proper value for a specific image. Small changes to the RSF have hardly no impact on the results. This circumstance make it even harder to set the right value for the RSF according to the used input image. It is therefore recommended, to use the RSF carefully.

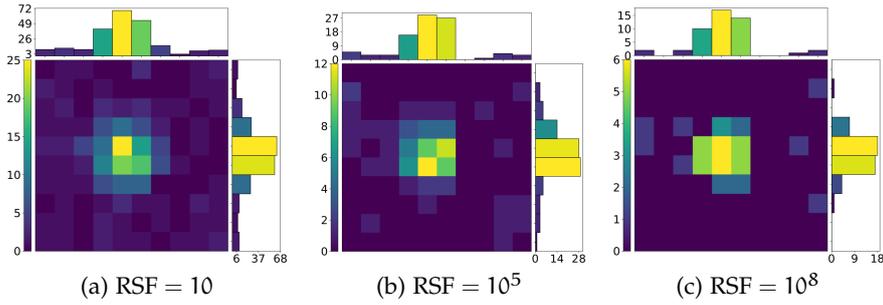


Figure 5.20: The 2D histograms of three meaningful results of applying the NBDA with different relative strength factors (RSF) to the test images: (a) $\text{RSF} = 10$, (b) $\text{RSF} = 10^5$ and (c) $\text{RSF} = 10^8$. The default value is set to $\text{RSF} = 10^5$.

5.7 SUMMARY

In this chapter, we introduced the network architectures of the SCNNs and we trained networks of different depths on the training data sets. The trained networks were tested to find the best performing SCNN using the weighted mean performance Π , a metric we have introduced in this thesis. According to this metric the $\text{SCNN}_{71}\text{-G6-TRS-F}$, a pre-trained and fine-tuned network, has a performance of $\Pi = 0.85$ taking faint and low-contrast samples into account but using lower weights for those samples to calculate the weighted mean performance. To put this performance in perspective, established networks, e.g. the VGG16, mentioned in section 5.1.1, has a valuation error of around 9% (Simonyan et al., 2014) on its trained task. Considering the mean true positive and negative predictions given in Table 5.6, 7% of predictions for the TES test set fall into the false positive and false negative category. These false predictions can be considered as errors. For the TES-F and TES-LC these errors are significantly worse with around 16% and 32%, respectively. Depending on the considered test set the errors of the $\text{SCNN}_{71}\text{-G6-TRS-F}$ can be considerably higher compared to the VGG16.

Description	parameter	default value
stride factor	γ_D	1
number of nested boxes	n_D	10
threshold	τ_D	0.9
Relative Strength Factor	RSF_D	10^5

Table 5.9: Table of the determined default values for the NBDA.

In addition to the training of the SCNNs, we introduced the nested boxes detection algorithm (NBDA). This algorithm calculates a grid on the input image and a batch of nested boxes on each grid position.

In combination with the SCNN_{71-G6-TRS-F}, it can be used to detect bubble-like structures in a 2D input image by return a prediction of a bubble-like structure for each nested box. Table 5.9 summarizes the default values for the NBDA determined in section 5.6. These default values cannot be considered as the perfect values for applying the NBDA on every image, but it should be a good starting point to apply the NBDA to an astronomical image.

CLASSIFYING DETECTED STRUCTURES

Applying the NBDA on all available images in the optical, radio and X-ray regime, results in detections of bubble-like structures for each input image. We combine these results and compare them across the considered images to make it possible to classify the detected structures into fundamental structures of the ISM. Our focus, however, is on SNRs and HII regions.

In the following chapter, we train a second neural network on a different classification task and refer it to the classifier of bubble-like structures, or classifier for short. The architecture of the classifier is shown in section 6.1. While the data sets for training and testing the network are discussed in section 6.2, the training procedure is presented in section 6.3. In the last section 6.4 of this chapter, the bridge between the NBDA and the classifier is explained.

6.1 CLASSIFIER ARCHITECTURE

The network used as classifier is a fully connected neural network (FCNN) discussed in section 2.2 and the principle of the architecture is shown in Figure 2.2, for instance. It expects a one-hot encoded input feature vector of length six. Each feature is therefore represented by a one or a zero meaning the feature is fulfilled or not. The classifier is limited to exactly six features which will be discussed in the next section. Each of the five layers of the FCNN uses the ReLU except of the output layer, which uses the Softmax activation function. The output of the classifier can therefore be interpreted as a probability and each output neuron represents one type of bubble-like structure, similar to the FCNN used for the SCNNs (see section 5.1.2). The entire network has 17100 trainable parameters and is therefore a small FCNN.

6.2 TRAINING AND TEST DATA

The training and test data set for the classifier is a set of feature vectors. Each vector characterizes a detected bubble-like structure according to the fulfilled features. There are six features and six labels. The features are shown in Table 6.2, which is used as a look-up table for labeling the feature vectors. The corresponding labels are summarized in Table 6.1.

The first feature tracks if there is a bubble-like structure detected at a certain position in the $H\alpha$ line emission image. If this is the case, the first element of the feature vector is set to one. Consequently, this

Description	Features	Label
considered as a SNR	PFG	SNRc
likely a SNRc	PF + SFG	likely
candidate of a SNRc	PF + SF	candidate
unlikely a SNRc	PF	unlikely
considered as HII region	H α + X-ray	HII
other bubble-like structures	SF/SFG	other

Table 6.1: The six labels which characterize the detected bubble-like structure and a brief description what the label means. The features needed to label the bubble-like structures accordingly is also given (P/SF: primary/secondary feature, P/SFG primary/secondary feature group)

feature can only be fulfilled if the NBDA is applied to an image showing H α line emission. This procedure is done in the correct order for [SII] and [OIII] line emission, non-thermal radio as well as for X-ray emission (Table 6.2). Thus, the first five elements of a feature vector are directly related to emission in a certain band. It should be highlighted, that the radio feature is tracking the non-thermal radio emission coming from the bubble-like structure. It is therefore required to apply the NBDA to a non-thermal radio image. The sixth feature tracks the ratio between [SII] and H α emission, which is enhanced for SNRs (see section 1.2.5). As a default value we use $^{[SII]}/H\alpha > 0.67$, although 0.4 would be sufficient for isolated SNRs (Kavanagh et al., 2016). The ratio is estimated by dividing the available images for [SII] and H α line emission and calculating the median ratio in the region of the detected structure. Therefore, the $^{[SII]}/H\alpha$ ratio can be underestimated or overestimated depending on the ratio between the estimated and the real extent of the detected structure.

6.2.1 Creating the Look-Up Table

Each feature vector has in total six features and is one-hot encoded, that means each feature value is either one or zero. Consequently, there are $2^6 = 64$ possible feature combinations which we show in Table 6.2. Every combination is labeled according to Table 6.1. The classifier is focused on SNRs, at the moment and, thus, the labels *likely*, *candidate* and *unlikely* are considered as a grading of how likely it is that the detected bubble-like structure could be a SNR indeed.

We divide the features into primary features (PFs) and secondary features (SFs) which form primary feature groups (PFGs) and secondary feature groups (SFGs), respectively. A PFG is necessary to label a feature vector as a *SNRc*. SFGs, however, are not absolutely necessary to classify a feature vector as *SNRc*, but it adds value to the detection.

The non-thermal radio, X-ray and the $[\text{SII}]/\text{H}\alpha$ ratio, together with the $\text{H}\alpha$ and $[\text{SII}]$ emission line features, are primary features. Since SNRs show at least two of these features, a PFG consist of at least two primary features. Secondary features are the single $\text{H}\alpha$, $[\text{SII}]$ and the $[\text{OIII}]$ emission line features. Any combination of the secondary features form a SFG. If the $[\text{SII}]/\text{H}\alpha$ ratio feature is fulfilled (1) but one of the emission line features $\text{H}\alpha$ or $[\text{SII}]$ is missing, it is referred to as contradictory prime feature and is considered as a SFG. If both emission line features or the $[\text{SII}]/\text{H}\alpha$ ratio feature itself is missing, it is referred to as a contradictory secondary feature and cannot be part of a SFG. Which features, or feature combinations, are required for which label is summarized in Table 6.1.

An example of a *SNRc* is the feature vector with the ID 25 in Table 6.2. All features are zero except of two primary features, namely the PFG of the non-thermal radio and X-ray feature. A feature vector with one primary feature only and at least one secondary feature is labeled as *likely*. The ID 42 in Table 6.2 is a feature vector with exactly one primary (non-thermal radio) and a contradictory prime feature with the $[\text{SII}]$ emission line feature missing, a SFG. Therefore, it is labeled as *likely*. As stated before, the missing $[\text{SII}]$ line emission feature does not necessarily mean that there is no $[\text{SII}]$ emission coming from this region, but was simply not detected by the NBDA. With a more detailed analysis of the considered region, e. g. as shown in Zangrandi et al. (2023, in prep.), there could be $[\text{SII}]$ emission spotted. Therefore, the label *likely* is introduced. A vector with the label *candidate* has again a primary feature and only one secondary feature, that means the feature vector has no SFG. An example would be the feature vector with the ID 13 in Table 6.2. Feature vectors which show just a primary feature are labeled as *unlikely* (e. g. ID 9). All other bubble-like structures showing no primary features are labeled as class *other*. There is one exception, namely the scenario of a feature vector with an $\text{H}\alpha$ and X-ray feature, which is labeled as a HII region. This feature combination appears only one time in Table 6.2 and has the ID 18.

Table 6.2: This table shows all 64 feature vectors and the corresponding labels which are used for simulating the training and test data set. It can also be used as look-up table.

ID	$\text{H}\alpha$	$[\text{SII}]$	$[\text{OIII}]$	Radio	Xray	$[\text{SII}]/\text{H}\alpha$	LABLE
1	0	0	0	0	0	0	other
2	1	0	0	0	0	0	other
3	0	1	0	0	0	0	other

Continued on next page

Table 6.2: This table shows all 64 feature vectors and the corresponding labels which are used for simulating the training and test data set. It can also be used as look-up table. (Continued)

ID	H α	[SII]	[OIII]	Radio	Xray	[SII]/H α	LABLE
4	1	1	0	0	0	0	other
5	0	0	1	0	0	0	other
6	1	0	1	0	0	0	other
7	0	1	1	0	0	0	other
8	1	1	1	0	0	0	other
9	0	0	0	1	0	0	unlikely
10	1	0	0	1	0	0	unlikely
11	0	1	0	1	0	0	unlikely
12	1	1	0	1	0	0	unlikely
13	0	0	1	1	0	0	candidate
14	1	0	1	1	0	0	candidate
15	0	1	1	1	0	0	candidate
16	1	1	1	1	0	0	candidate
17	0	0	0	0	1	0	unlikely
18	1	0	0	0	1	0	HII
19	0	1	0	0	1	0	unlikely
20	1	1	0	0	1	0	unlikely
21	0	0	1	0	1	0	candidate
22	1	0	1	0	1	0	candidate
23	0	1	1	0	1	0	candidate
24	1	1	1	0	1	0	candidate
25	0	0	0	1	1	0	SNRc
26	1	0	0	1	1	0	SNRc
27	0	1	0	1	1	0	SNRc
28	1	1	0	1	1	0	SNRc
29	0	0	1	1	1	0	SNRc
30	1	0	1	1	1	0	SNRc
31	0	1	1	1	1	0	SNRc
32	1	1	1	1	1	0	SNRc
33	0	0	0	0	0	1	other
34	1	0	0	0	0	1	other

Continued on next page

Table 6.2: This table shows all 64 feature vectors and the corresponding labels which are used for simulating the training and test data set. It can also be used as look-up table. (Continued)

ID	H α	[SII]	[OIII]	Radio	Xray	[SII]/H α	LABLE
35	0	1	0	0	0	1	other
36	1	1	0	0	0	1	unlikely
37	0	0	1	0	0	1	other
38	1	0	1	0	0	1	other
39	0	1	1	0	0	1	other
40	1	1	1	0	0	1	unlikely
41	0	0	0	1	0	1	unlikely
42	1	0	0	1	0	1	likely
43	0	1	0	1	0	1	likely
44	1	1	0	1	0	1	SNRc
45	0	0	1	1	0	1	candidate
46	1	0	1	1	0	1	likely
47	0	1	1	1	0	1	likely
48	1	1	1	1	0	1	SNRc
49	0	0	0	0	1	1	unlikely
50	1	0	0	0	1	1	likely
51	0	1	0	0	1	1	likely
52	1	1	0	0	1	1	SNRc
53	0	0	1	0	1	1	candidate
54	1	0	1	0	1	1	likely
55	0	1	1	0	1	1	likely
56	1	1	1	0	1	1	SNRc
57	0	0	0	1	1	1	SNRc
58	1	0	0	1	1	1	SNRc
59	0	1	0	1	1	1	SNRc
60	1	1	0	1	1	1	SNRc
61	0	0	1	1	1	1	SNRc
62	1	0	1	1	1	1	SNRc
63	0	1	1	1	1	1	SNRc
64	1	1	1	1	1	1	SNRc

6.2.2 Simulating Data Sets

We use the look-up table 6.2 to simulate a training and test data set for the classifier. There is no overlap between the 64 feature vectors and there is no bias because of a subjective decision-making of an observer. This is in contrast to the training and test data set of the SCNNs in section 4. Usually, it is difficult to simulate training samples to train a neural network, because the simulated samples need to be as realistic as possible to not train the network on features which do not exist in reality. This would lead to a high training but a poor real life performance of the network. If the simulated and the real feature vectors are identical, the network can be trained effectively for the real application. Simulating realistic feature vectors as shown in Table 6.2 is indeed possible. The classifier can therefore be trained effectively with the simulated training data set.

For each data set a *numpy* array, containing a certain number of feature vectors, is simulated by randomly creating features of values $\{0, 1\}$. These simulated feature vectors are labeled according to the look-up table. For the FCNN used as classifier a set of 100000 training samples is sufficient to train the network. This means 10000 validation samples are simulated, which corresponds to the usual fraction of 10% of the training samples. Furthermore, a test set of 1000 feature vectors is simulated to test the network after the training.

6.3 TRAINING AND EVALUATION

The training procedure for the FCNN is much simpler than the one for the SCNNs, since the training data set is large compared to the small network. Furthermore, the training data set is of better quality, because of the data structure discussed in section 6.2.2. These factors make it relatively easy to train the classifier.

Parameter	value
<i>nClasses</i>	6
<i>Epochs</i>	10
<i>BatchSize</i>	64
<i>Optimizer</i>	Adam
<i>LearningRate</i>	0.001
<i>EarlyStopping</i>	yes

Table 6.3: Training parameter and their values used for training the classifier.

Table 6.3 give an overview of the used training parameters. Due to the large training data set, the FCNN is trained with fewer epochs, but a larger batch size compared to the SCNNs. In addition, we use early

stopping (see section 2.1.7) to avoid overfitting and reduce computation time. After the network reaches a precision higher than 99.9% the model is saved to disk and the training is finished. This is the case after two epochs only with a loss of $3.7 \cdot 10^{-4}$, which shows the effective training of the FCNN.

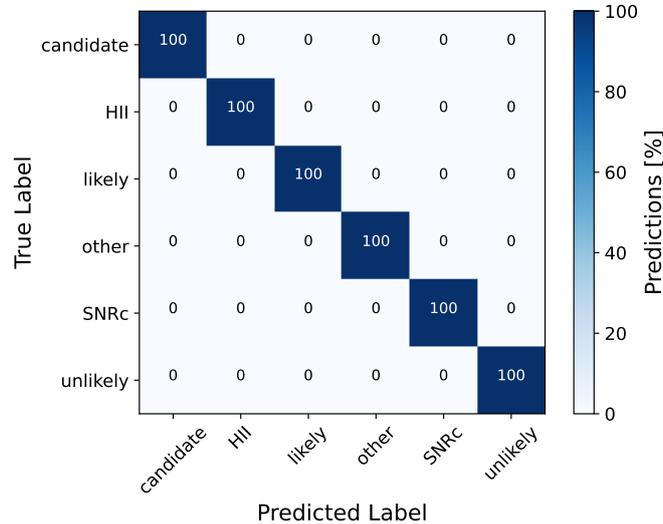


Figure 6.1: The confusion matrix of the classifier network. It shows now confusion between different labels. All predicted labels match the true label.

The network is tested on the simulated test data set after training and a confusion matrix (CM) is calculated. The CM is shown in Figure 6.1. In contrast to the CMs shown in section 5.5, the classifier is not confused between different classes (labels) at all. Consequently, all predicted labels match the true labels. If a feature vector characterize the detected bubble-like structures correctly, the classification is correct with a high probability of $> 99.9\%$. The input feature vectors for the classifier depend on the NBDA results and have to be created before. Feature vectors are set up by comparing all NBDA results and combine them into a set of features. This process is explained in section 6.4 and is called the bridge.

6.4 THE BRIDGE

The NBDA detects bubble-like structures in different images. The bridge loops over all available images and detections for each of these images, matching the coordinates to generate the corresponding feature vector for this specific bubble-like structure.

Results of the NBDA are in the form of a DS9 region files. On each image a bubble-like structure is detected at a certain grid position with a certain box framing the region of the structure. The grid positions and therefore the results of the NBDA are in pixel coordinates. We use

these results to create a mask of detected structures. For the matching of coordinates of different images, which may have different pixel sizes, a world coordinate system (WCS) is required to transform pixel coordinates to world coordinates. For images without such a WCS the bridge and, therefore, the classification of bubble-like structures does not work. While looping over the detected bubble-like structures of an image, the masks of all the other images are checked if there is a bubble-like structure detected as well. If that is the case we set the corresponding feature to one, otherwise it is set to zero. The region of the corresponding structure is flagged as part of the same detection. This avoids multiple feature vectors characterizing the same bubble-like structure. Once the bridge looped over all NBDA results an array of feature vectors is created. This array is then used as input for the classifier.

6.5 SUMMARY

In this chapter we introduced a second neural network, the classifier, to classify the detected bubble-like structures. For that purpose a look-up table with focus on SNRs was created and used as the base-ment of simulations of feature vectors. The resulting data were used for training and testing the classifier. It was shown that the precision is higher than 99.9%. Nevertheless, the classification depends on the performance of the NBDA, which by far does not show a precision as high as the classifier. The bridge combines the detection with the classification pipeline of *BScan*. The overall *BScan* performance is tested in the following part iii of this thesis.

Part III

APPLYING BSCAN

We apply *BScan* to astronomical images in this chapter. *BScan* uses a configuration file to set all necessary parameters. The configuration of *BScan* is presented in section 7.1. *BScan* is then applied to the MCELS data of the SMC (section 7.2) and on radio, optical and X-ray data for the LMC (section 7.3). Results of the SMC are used to compare *BScan*, as we presented in this thesis, to the previous version described in Geyer-Ramsteck (2020). Since the previous version is not able to classify detected structures, the NBDA and the previous *BScan* version is applied to the SMC with comparable parameters. Data from the SMC were not used to train both versions. Thus, applying *BScan* to the optical images of the SMC is a blind test for both. Results of the LMC, however, are used to test the *BScan* performance including the classification. Many samples in the training data set originate from the LMC. Applying *BScan* on images of the entire LMC is therefore not a blind test. Nevertheless, it is the only galaxy with images available for all wavelengths needed for classification.

7.1 THE CONFIGURATION FILE

The configuration of *BScan* is done via a configuration file named *config.ini*. *BScan* is called via command line prompt. Part of this prompt is the relative path to the mentioned configuration file. This way, the *BScan* configurations are done all in one place. The *config.ini* file is divided into four sections. Each section provides a set of certain parameters, some of which are mandatory and some are optional. The sections are

- PARAMETER: sets parameter for the entire pipeline
- PATHS: configures input and output paths
- DATA: sets paths to the 2D images required for classification and
- aDATA: provides additional 2D FITS images.

The given parameters are written into a log-file (.log). Therefore, the used parameters for a certain run are tracked. Furthermore, the configuration file can be used to modify the pipeline. It is therefore possible to use the NBDA or the classifier only. Figure 7.1 shows the optional paths through the pipeline in dashed lines. The diamond shapes mark the parts of *BScan* where neural networks are involved. Boxes, however, show parts of classical algorithms. The circles indicate output of *BScan*.

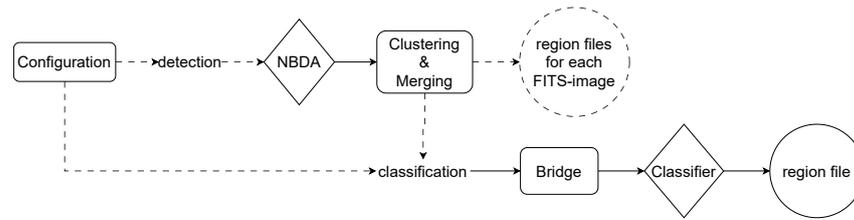


Figure 7.1: This Figure shows the pipeline structure for *BScan*. Solid lines indicate fixed parts of the pipeline. Dashed lines are optional paths. While diamonds show the location of the neural networks, boxes show parts of classical algorithms. The circles show the output.

The `PARAMETER` section sets all the parameters needed for the detection and classification. Mandatory parameters in this section are the *DetectorName*, the name of the CNN trained on image recognition and used for the NBDA, and the *ClassifierName*, the name of the FCNN trained on the feature vectors discussed in section 6. Both are defined without the file ending „h5“, which is the data format used to save the trained parameters of the network to disk and is therefore assumed by *BScan*. The remaining parameters are optional and are set to default values. Changing these default values changes the detection and classification behavior. We summarize the optional parameters which can be set in the configuration file in the following:

Detection

A boolean parameter to configure *BScan* for detecting bubble-like structures for the given 2D FITS images in section *Data* and *aData*. The default value is *True*. If *False*, *BScan* expects results for each given image of the same format as given by the bridge.

nBoxes

Number of nested boxes used by the NBDA ($n \in \mathbb{N}$). The default value is 10.

minBXS

Minimum box size in pixels used by the NBDA ($\lambda \in \mathbb{N}$). The default value is 50px.

maxBXS

Maximum box size in pixels used by the NBDA ($\Lambda \in \mathbb{N}$). It is larger than, or equal to *minBXS*. If Λ is equal to λ , the NBDA uses a single box size. The default value is 500px.

StrideFactor

A factor to define the grid size calculated by the NBDA ($\gamma > 0$). The default value is 1.

Threshold

The minimum probability to consider a prediction of the network as detected. It is used by the NBDA ($\tau > 0$). The default value is 0.9.

OverlapFactor

Defines the possible overlap (> 0) considered by the clustering and merging algorithms. It defines the distance between two neighboring regions by measuring the separation of the two centers. This distance is defined as *OverlapFactor* $\cdot r_0$ with r_0 the radius of the considered region. A neighbor region with distance smaller than *OverlapFactor* $\cdot r_0$ is considered as overlapping. An *OverlapFactor* of 1 means two regions are overlapping, if the center of the second region is within the radius of the first region. The default value is 0.75.

RAW

A boolean parameter. If *True* all sub-results are saved to disk. The default value is *False*.

siizha

Literature value of the $[\text{SII}]/\text{H}\alpha$ ratio (> 0). The default value is 0.67 (see section 1.2.5).

PXS

Pixel size in $\text{arcsec}/\text{pixel}$ of the given $[\text{SII}]$ and $\text{H}\alpha$ image. This value is needed to estimate the $[\text{SII}]/\text{H}\alpha$ ratio and is mandatory for the classification. For the MCELS images the pixel size is $2 \text{ arcsec}/\text{pixel}$.

Classification

A boolean parameter to enable classification. If *False* the NBDA results will not be classified. The default value is *True*.

The `PATHS` section sets the absolute input and output paths. All paths are required parameters:

PathToDetector

Location of the saved CNN.

PathToClassifier

Location of the saved FCNN.

PathToResults

Path to save the output of *BScan*. If this directory does not exist, it will be created by *BScan*.

The last two sections `DATA` and `aDATA` define the absolute paths to the available FITS images to detect bubble-like structures and/or classify the detected structures.

The keys for the `DATA` section

- *radio*
- *ha*
- *sii*
- *oiii*
- *xray*

represent the different wavelengths and should not be changed, because they are hard coded. The data given in this section are also needed for classification. If one of the required input images is missing, the classification does not work properly.

In contrast to `DATA`, keys of the `aData` section can be chosen by the user. Note that no special characters are allowed. Consequently, only the NBDA can be used for images given in the `aDATA` section. *BScan* sets *Classification* to *False* automatically if additional images are given. If *NONE* is given for a certain key, it will be ignored by *BScan*. This is also the case for the `DATA` section.

7.2 SMALL MAGELLANIC CLOUD

In this section we compare the current version of *BScan* and the version presented in Geyer-Ramsteck (2020) qualitatively. The version described in Geyer-Ramsteck (2020) is referred to as the version zero (V0) and the version presented in this thesis is referred to as version one (V1) to clearly distinguish between the two. Before diving into this comparison, we give a short recap of the different versions in the following paragraph.

While both versions leverage a CNN, V0 uses the pre-trained VGG16. It is pre-trained on everyday life color images from the *ImageNet* data set (see 5.1.1). Version V1, however, uses a much smaller CNN with fewer parameters compared to the VGG16 and is pre-trained on the G6 data set (see section 4), which contains grayscale images of different types of galaxies. Therefore, the G6 is considered as a more task related data set and is better suited for pre-training the CNN used for *BScan*. In addition, V1 has the NBDA implemented and is able to detect bubble-like structures of different sizes at the same grid position. This is due to the nested boxes with different box sizes where each box is used to predict if there is a bubble-like structure or not. Version V0 uses a single box size. Nevertheless, both versions detect one structure per grid position and merge overlapping boxes, that means multiple detections of a considered structure at different grid positions are combined to a final region. The final result, represented as a DS9 region, is then written to a .reg file. Version V0 uses the coordinates of the merged box center as location of the detected structure and half of the merged box size as an estimate of the radius. V1 estimates the center coordinates and extent of the detected structure according to equation 46 and equation 47 in section 5.4, respectively. The classifier is just implemented in version V1 and is therefore not part of this comparison.

We apply both versions to the MCELS data of the SMC. Therefore, the detection of bubble-like structures is tested on an $H\alpha$, [SII] and [OIII] image. The pixel size of these images is $2 \text{ arcsec}/\text{pixel}$. Three regions of the SMC are considered for this comparison: Rg1 at RA= 14.97° , Dec= -71.67° with radius $R_1 = 0.27^\circ$ ($\equiv 972\text{px} \times 972\text{px}$ image), Rg2 at RA= 21.87° , Dec= -73.33° with radius $R_2 = 0.49^\circ$ ($\equiv 1764\text{px} \times 1764\text{px}$ image) and Rg3 at RA= 11.76° , Dec= -73.20° with radius $R_3 = 0.56^\circ$ ($\equiv 2016\text{px} \times 2016\text{px}$ image). The first region Rg1 (Figure 7.2) shows mostly isolated bubble-like structures of different extents, region Rg2 (Figure 7.4) shows a big bubble-like structure surrounded by smaller bubble-like structures overlapping with the bigger one and region Rg3 (Figure 7.10) is a crowded area with a high density of bubble-like structures. Therefore, each region is challenging both versions in a different way.

7.2.1 Configuration

We use the networks *model4* for the version V0 (Geyer-Ramsteck, 2020) and the SCNN71-G6-TRS-F for the V1. For both versions the configuration is mainly set to the default values (see Table 5.9 for V1). Parameters which both versions have in common are the

- *StrideFactor* = $\gamma = 1$
- *Threshold* = $\tau = 0.9$ and
- *OverlapFactor* = 0.75

and they are configured with the same values. The maximum box size Λ in version V1 is set to the single box size of 250px used in version V0. This ensures that both versions could detect structures with the same upper limit in size. In comparison to V0, however, the minimum box size λ of V1 is set to 50px with $n = 10$ nested boxes, the only direct advantage for V1 compared to V0 in this configuration. The configuration is chosen to test if version V1 has significantly improved compared to V0 or not. Therefore, the configuration has to include the possibilities given by the NBDA. The relative strength factor is also used and is set to the default value of $RSF = 10^5$. Since the RSF can be considered as an additional threshold which has to be met to detect a bubble-like structure, it may compensate the advantage of the nested boxes.

7.2.2 Isolated Structures

Results of the two *BScan* versions in region Rg1 show a clear picture. While V0 could not detect any structure in any of the three images $H\alpha$, [SII] or [OIII] (Figure 7.2 a), version V1 detects some bubble-like structures in the $H\alpha$ (Figure 7.2 b, red) and [OIII] (Figure 7.2 b, blue) images.

A first look at the results in Figure 7.2 (b) shows that V1 overestimates the extent of the detected structures in $H\alpha$ and [OIII]. The results of the $H\alpha$ image shows that V1 is missing some structures which would be detected visually, for example the bigger shell-like structure in the north at around $RA \approx 15.25^\circ$, $Dec \approx -71.5^\circ$. This may be due to the size of the structure. Since the maximum box size Λ is set to 250px and the structure is larger than Λ , it may not be detected. Even if the structure could be partly detected at some grid positions, the threshold or the *RSF* may be too high to accept the predicted bubble-like structure by the NBDA. Figure 7.3 shows the same region of the SMC but with an applied relative strength factor of $RSF = 0$ and a threshold of $\tau = 0.9$ (a) and $\tau = 0.5$ (b). The considered structure is still not detected in both cases. One possible interpretation is that the structure is indeed too large for the given Λ . Another possibility

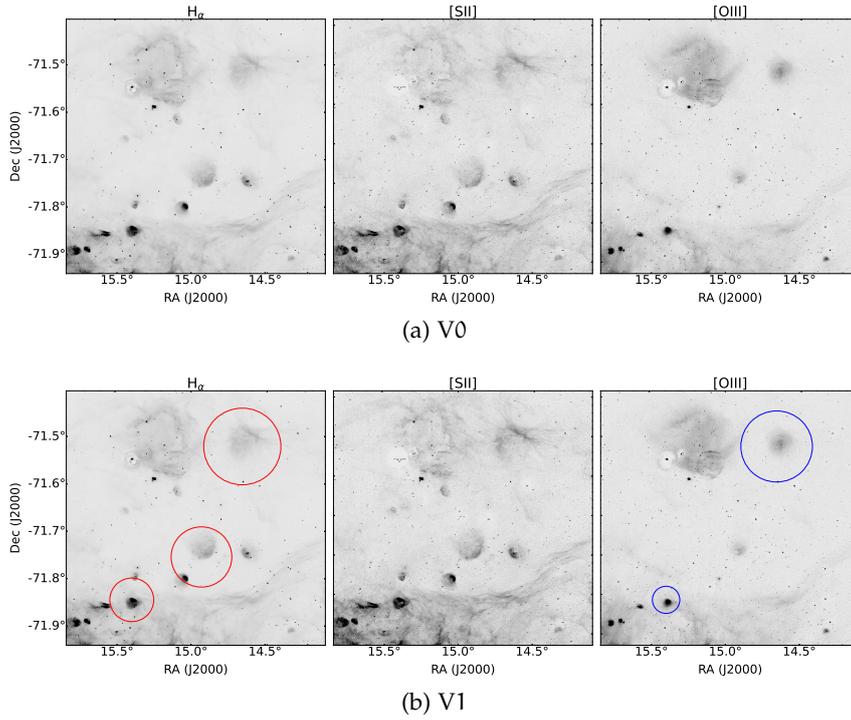


Figure 7.2: Results of *BScan* version V0 (a) and V1 (b) in the region Rg1. The region is located in the SMC at RA= 14.97°, DEC= -71.67° with a radius of $R_1 = 0.27^\circ$. This corresponds to a 972px \times 972px image. While V1 detects at least some bubble-like structures in this region, V0 cannot detect any structure.

is that the network is not generalized properly to detect the structure in the north of the image.

Comparing the results for [SII] and H α in Figure 7.2 (b), especially the detection in the H α image at RA $\approx 15^\circ$ and Dec $\approx -71.75^\circ$, one could expect that V1 should detect the same structure in [SII] as well. The [SII] structure is fainter and shows a slightly different intensity profile but is very similar to the structure visible in H α nevertheless. Figure 7.3 (a), however, shows that the structure is detected in the [SII] image with a $RSF = 0$ and $\tau = 0.9$. Not only the above-mentioned structure is detected but also the south-western structure, detected in H α (Figure 7.2, b), is detected in the [SII] image (Figure 7.3, a). Overall, Figure 7.3 shows a trend of more detected bubble-like structures with lower values of threshold typed parameters. Consequently, *BScan* version V1 can handle the first challenge of mostly isolated structures better than V0.

7.2.3 Overlapping Structures

Comparing the results for region Rg2 of both *BScan* versions show that V0 can detect the brightest bubble-like structures in the H α and [OIII] images (Figure 7.4, a), but the same structures could not be de-

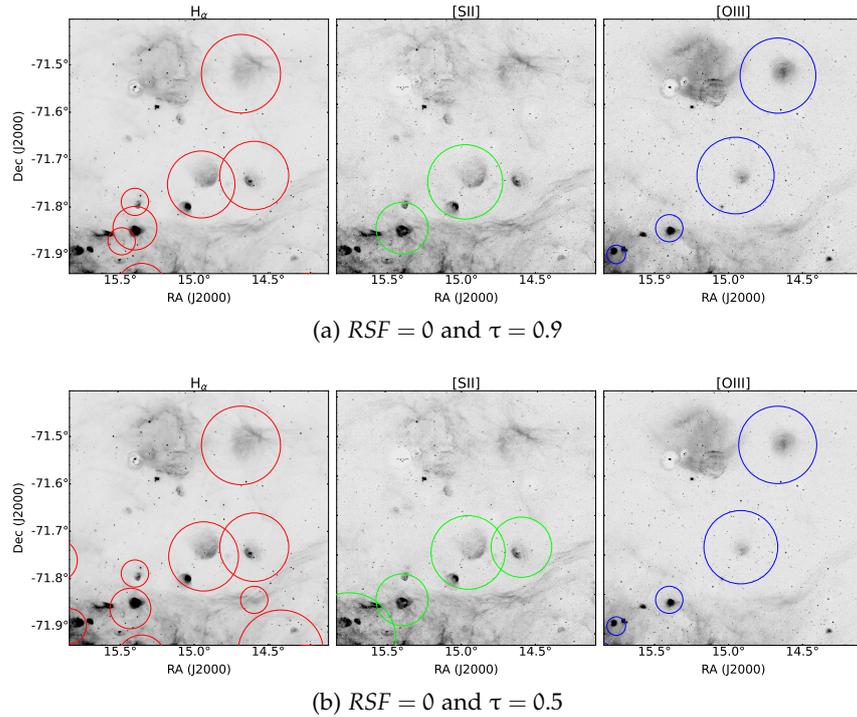


Figure 7.3: Results of *BScan* version V1 in region Rg1 but with different threshold typed parameters: (a) a relative strength factor of $RSF = 0$ and threshold of $\tau = 0.9$, (b) a threshold of $\tau = 0.5$ but the same RSF as in (a). Compared to Figure 7.2 (b) the number of detected bubble-like structures increases.

tected in the [SII] image. Furthermore, V0 underestimates the extent of the detected structures. Version V1 detects the same bubble-like structures as V0 in these images, but in all three images (Figure 7.4, b). The estimated extent is larger than the actual structure and is therefore overestimated again. Moreover, V1 detects more structures in $H\alpha$ and [OIII] compared to V0. Overall, V1 performs better in the second challenge as well.

In some cases it looks like V1 detects bubble-like structures where no structure is present. These detections are marked with A, B, C and D in Figure 7.4 (b). A closer look at these regions, however, reveals that there are structures which fall into the class diffuse or shell. Figures 7.5 to 7.8 zoom into A, B, C and D, respectively. The detected bubble-like structures in A (Figure 7.5), B (Figure 7.6) and D (Figure 7.8) would be characterized as diffuse, while the structure in C (Figure 7.7) would be of class shell. Regions A to C show clearly a bubble-like structure as represented in the training data set (see chapter 4). The structure in region D is most likely a region within a filament. This region is part of the large shell-like structure seen in Figure 7.4, with an increased intensity and seems like a structure of class diffuse. Without a doubt these detections could be interesting as well, but fur-

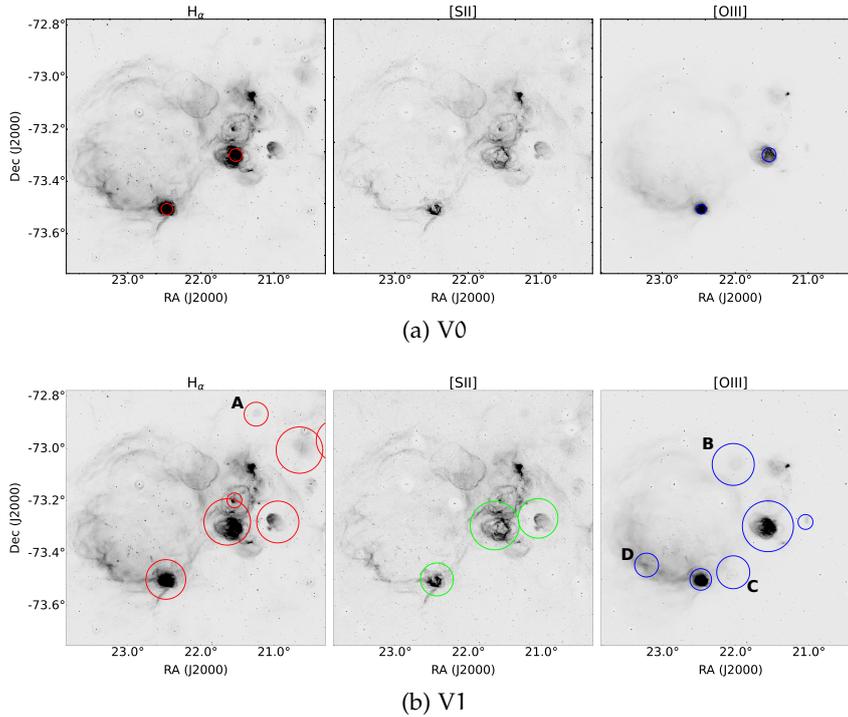


Figure 7.4: Results of *BScan* version V0 (a) and V1 (b) in the region Rg2. The region is located in the SMC at RA= 21.87°, DEC= -73.33° with a radius of $R_2 = 0.49^\circ$. This corresponds to a 1764px \times 1764px image. Both versions detect the brightest bubble-like structures in this region in H α and [OIII]. However, V1 detects the brightest structures in all three bands and detects more structures than V0 in each one. Detected structures in (b) marked with A, B, C and D are fainter compared to other structures and need further investigation.

ther analysis is needed. For this test, this detection is considered as a false positive detection of class diffuse.

If the threshold type parameters are reduced like in section 7.2.2, the *BScan* version V1 is again able to detect additional bubble-like structures. For example in region B, two shell-like counterparts are visible in H α and [SII]. Figure 7.9 shows results for region B using a relative strength factor of $RSF = 0$ and a threshold of $\tau = 0.5$. Note that although the threshold type parameters are extremely low, the shell-like counterparts in [SII] are not detected while there are two related detections in H α . Furthermore, there are more of the diffuse-like filament detections in H α and [OIII].

7.2.4 Crowded Area

The last region to consider is the region Rg3. It encloses an area of many nearby or even overlapping bubble-like structures of different size. Therefore, region Rg3 is the hardest detection challenge for both

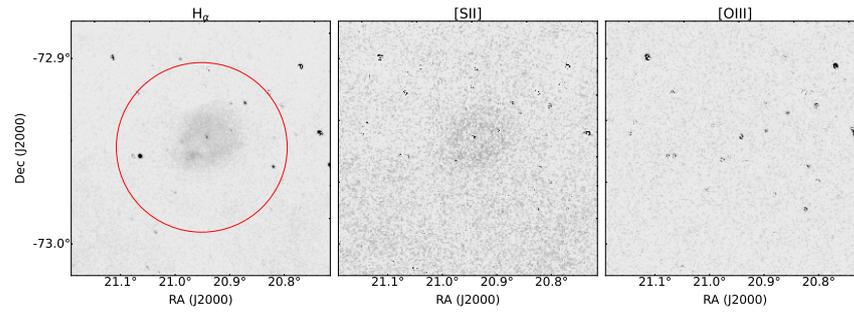


Figure 7.5: The faint source detected in region A. The detection in $H\alpha$ is clearly visible and has an even fainter counterpart in [SII].

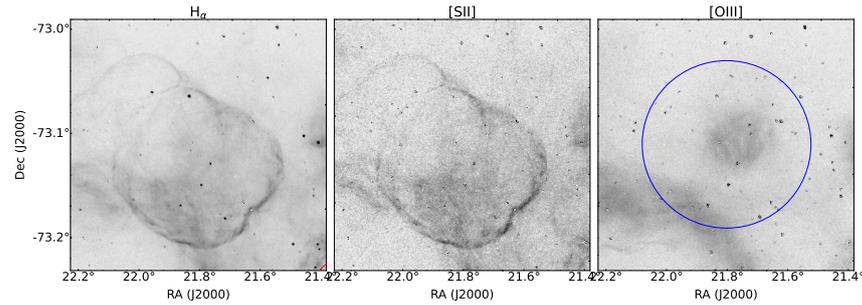


Figure 7.6: The faint source detected in region B. The detection in [OIII] (blue) is clearly visible. The two shell-like counterparts in $H\alpha$ and [SII] are not detected. This is because of the relatively high threshold typed parameters, the RSF and τ .

BScan versions. Figure 7.10 shows the results for *BScan* version V0 (a) and V1 (b). In general, the results are similar to the ones in section 7.2.2 and section 7.2.3 with more bubble-like structures detected by V1 compared to V0 in all three images $H\alpha$, [SII] and [OIII]. V1 shows a significant improvement compared to V0 again.

Especially in $H\alpha$ and [OIII] the results of V1 are surprisingly good, taking into account that the shown results are achieved with the default configuration. Structures of different size and brightness are detected. Most of the structures, which are located close to each other, are detected as separate structures. There is, however, a detection region which encloses multiple structures in the center of the $H\alpha$ and [OIII] images. This region is marked with an X and is shown in Figure 7.11 in more detail. On the one hand, assuming the more point-like structures are detected separately in the first place, would lead to the conclusion that the NBDA results have a larger overlap than it is accepted according to the *OverlapFactor* of 0.75 and consequently are clustered and merged together. On the other hand, the minimum box size λ could be too large to detect the small structures shown in region X. In this case, a too large λ would mean that the extent of the point-like structures is smaller than $0.8 \cdot \lambda$, which is roughly the structure size relative to the box size used as a guideline for the training samples (see section 4.1). Considering that the largest and most dom-

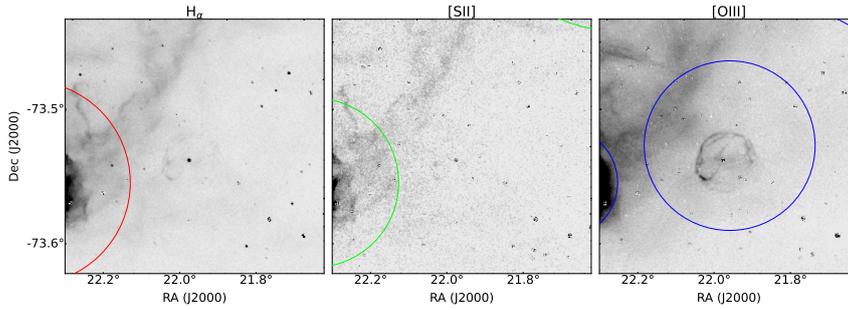


Figure 7.7: The faint source detected in region C. The detection in [OIII] (blue) is again visible. The extremely faint counterpart in the $H\alpha$ image is not detected.

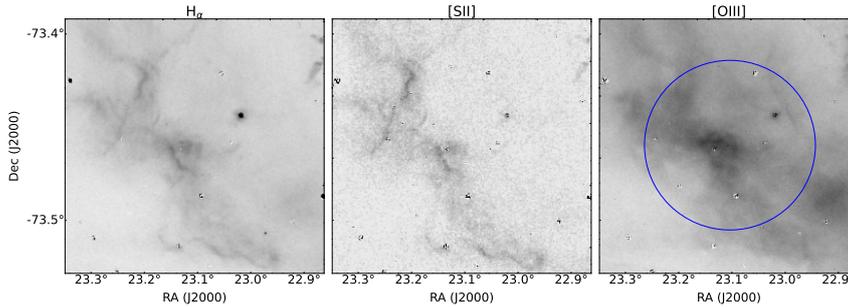


Figure 7.8: The faint source detected in region D. The detection in [OIII] (blue) is most likely part of a filament with higher local intensity. The filament is part of the large shell-like structure seen in Rg2.

inant diffuse structure in Figure 7.11 is off the center of the estimated center coordinates of the detection region, the first explanation seems more likely.

Comparing the results for the $H\alpha$ and [OIII] images with the [SII] image (Figure 7.10) reveals an underperformance of both versions in [SII], although bubble-like structures are clearly visible in the area around $RA \approx 11.7^\circ$, $Dec \approx -73.3^\circ$. Of course the value for the threshold τ can be reduced to accept more predicted bubble-like structures in the NBDA as demonstrated in the previous sections, but this would also lead to more structures detected in $H\alpha$ and [OIII] and consequently increase the potentially false positive detections overall. Since the MCELS images have the same dimension, resolution and pixel size, this inconsistency in detecting bubble-like structures in these images cannot be explained by the used configuration and therefore the grid size, the size of the nested boxes or the used thresholds. Similar structures compared to the ones which are not detected in some images, are detected in $H\alpha$, [OIII] and [SII] itself at other positions. The only plausible explanation at this point is the lack of generalization of the networks. For version V1 this falls back on the tiny training data set. It has to be stated that one set of parameters is used to apply *BScan* to all wavelengths and the corresponding NBDA results are used to generate the input for the classifier later on. If there are bubble-like

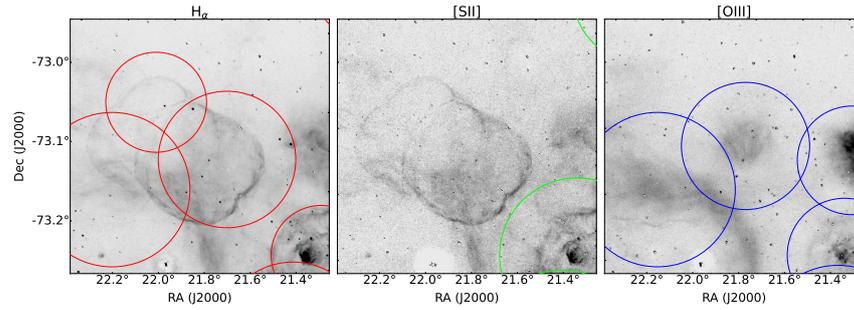


Figure 7.9: This figure shows region B again, but with results of *BScan* version V1 using a relative strength factor of $RSF = 0$ and a threshold of $\tau = 0.5$. With these low thresholds the two shell-like counterparts of the [OIII] in $H\alpha$ only has three detections related to the two structures.

structures detected in some images, but the counterparts present in others are not, the classification cannot work properly.

7.2.5 Summary

The *BScan* version V1, presented in this thesis, performs significantly better than *BScan* version V0, developed in Geyer-Ramsteck (2020), in all three scenarios discussed in section 7.2.2, 7.2.3 and 7.2.4. This means V1 detects more bubble-like structures than V0, including fainter structures and discriminating those which are partly overlapping or which are close to each other. The estimated extent of detected structures are underestimated by V0 and overestimated by V1.

Moreover, this blind test on the SMC data reveals the weaknesses of V1 as well. The inconsistency of detecting some bubble-like structures, but do not detect similar structures in the same or similar images is one example. This can partly be explained by the chosen parameter values like the threshold and the relative strength factor. In some cases, however, the low generalization level of the trained network is the most plausible explanation. In addition, V1 detects multiple bubble-like structures as one single structure in some cases. This results in detection regions with an overlap larger than accepted by the clustering and merging algorithm.

We summarize the results of the entire chapter with focus on the *BScan* version V1 in Table 7.1.

7.3 LARGE MAGELLANIC CLOUD

In this section we perform a more quantitative analysis of *BScan* (V1) results, including the classification, by applying the whole pipeline to the available LMC images and comparing them to a catalog of 92 sources from Bozzetto et al. (2017), Maggi et al. (2016), and Yew et al.

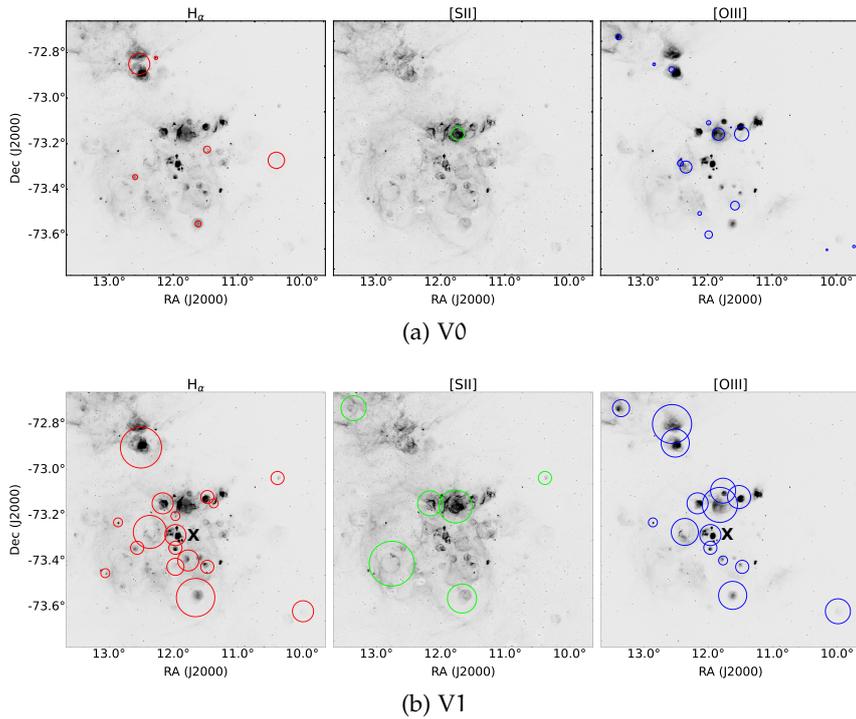


Figure 7.10: This figure shows the results of *BScan* version V0 (a) and V1 (b) in the region Rg3. The region is located in the SMC at RA= 11.76°, DEC= -73.20° with a radius of $R_2 = 0.56^\circ$. This corresponds to a 2016px \times 2016px image. The region marked with X is a multiple detection region where more than one bubble-like structures are detected as one single structure.

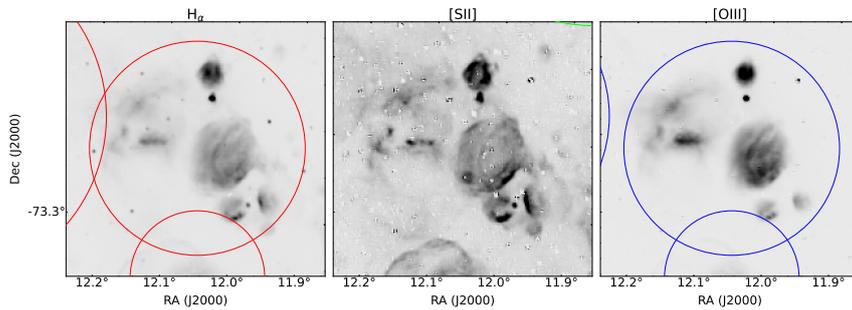


Figure 7.11: This Figure shows region X marked in Figure 7.10. Multiple structures are detected in one single detection region in H α (red) and [OIII] (blue).

Region	Challenge	Configuration	Results	Problems
Rg1	isolated structures	default	inconsistent detection results in different images, lower relative strength factor (RSF) ⇒ more consistent detections in different images	overestimated structure extents
Rg2	overlapping structures	default	detection of diffuse-like filament regions, not all structures could be detected	lower threshold typed parameters (τ , RSF) ⇒ more diffuse-like filament detections, inconsistent detection results in different images even with varying thresholds
Rg3	crowded area	default	structures of different size are detected, overlapping structures are mostly detected separately	multiple structures are detected as one

Table 7.1: A summary of the main *BScan* V1 results discussed and problems revealed in section 7.2.

(2021). These sources are confirmed SNRs or SNR candidates. In the following the sources are referred to as catalog sources or just sources for simplicity. Images are available for $H\alpha$, [SII] and [OIII] line emission in the optical (section 3.2), a radio image showing non-thermal emission (section 3.3) and an X-ray image (section 3.4). Although SNRs of the LMC are used for the training data set and the LMC is, therefore, not completely unknown like the SMC, it is the only available data set which can be used to detect and classify bubble-like structures.

7.3.1 Configuration

The network used to apply *BScan* to the LMC data is again the SCNN71-G6-TRS-F. Parameters of the NBDA are set to default values (see Table 5.9 and section 7.1). The minimum box size of $\lambda = 50\text{px}$ and the maximum box size of $\Lambda = 500\text{px}$ translates to structure sizes of 100 arcsec to 1000 arcsec and, therefore, a wide range of detectable bubble-like structures. The *OverlapFactor* = 1 is used to reduce the merging of nearby structures. For the classification the ratio of $^{[SII]}/H\alpha = 0.67$ is set in configuration file.

In the following images, detected bubble-like structures are represented by black circles (NBDA results), whereas catalog sources are shown in blue circles. Each detected structure is classified according to the labels given in Table 6.1. Each label has a specific color: *SNRc* (green), *likely* (yellow), *candidate* (orange), *unlikely* (red), *other* (brown) and HII regions (cyan).

7.3.2 Matching of Catalog Sources

Figure 7.12 shows the number of detected bubble-like structures which are classified and matched to a catalog source (M), relative to the number of sources in the entire catalog ($N_c = 92$). Only 16% of these structures are labeled as *SNRc* (green). This translates to only 15 confirmed SNRs or SNR candidates which are detected and characterized as such by *BScan*. In addition to the 16% of *SNRc*-labeled structures, roughly 4% and 2% got the label *likely* (yellow) and *candidate* (orange), respectively. In numbers this means 4 bubble-like structures are labeled as *likely* and 2 are labeled as *candidate*. The majority of around 17% (≈ 16 structures), however, are labeled as *unlikely* (red). All detected structures labeled as *likely*, *candidate* or *unlikely* are missing one primary feature (PF). This can be the case due to one of the reasons summarized in section 7.2.5. Also bubble-like structures which are labeled as *other* (brown, $\approx 8\%$) or as HII region (cyan, $\approx 1\%$) can be matched to catalog sources. The fact that each of the contributing structures in Figure 7.12 has matches with the catalog shows that the results of the current *BScan* version (V1) are not yet trustworthy

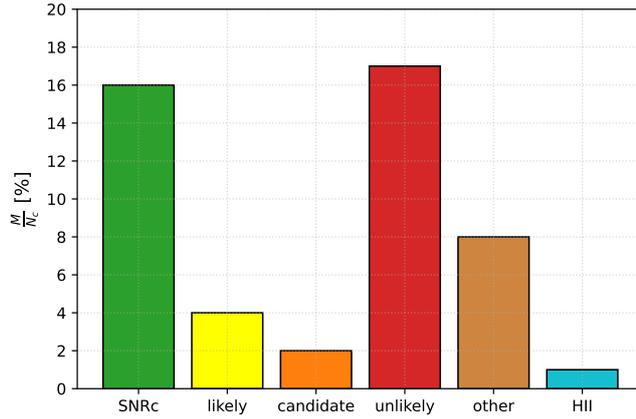


Figure 7.12: In this figure the fraction of detected and characterized bubble-like structures M matching with catalog sources in the LMC $N_c = 92$ is shown. Only 16% of the detected bubble-like structures which are labeled as *SNRc* (green) have counterparts in the catalog. While a small fraction of 4% and 2% is labeled as *likely* (yellow) and *candidate* (orange), respectively, the majority (17%) of matched bubble-like structures are labeled as *unlikely* (red). Structures which could not be classified as an SNR-like structure but have a counterpart in the catalog are labeled as *other* (magenta) and HII regions (blue).

enough to use it as a standard tool for detection and classification of bubble-like structures in astronomy.

In total, we matched 45 detected bubble-like structures to catalog sources, that means only around 49% of the catalog sources are detected. Each detected and matched structure has an ID and is listed in Table 7.2 together with its coordinates in RA, Dec, its estimated radius R , the label and the corresponding feature vector.

One major characteristic of the detections, beside the classification, is the estimated coordinates of the bubble-like structure. We therefore compare the coordinates of the detected structures to the matched sources by measuring the separation between their center coordinates Δ . Figure 7.13 shows the results for each ID. About 96% of the structures have a $\Delta < \lambda$. Only two bubble-like structures show a separation of $\Delta > \lambda$. Those structures have the IDs 35 and 36. The median separation of

$$\Delta_m = 36 \text{ arcsec} = 0.01^\circ \quad (56)$$

can be used to estimate the error of the estimated center coordinates.

In the next step we compare the estimated radius of the detected structure R to the radius r_c given for each catalog source (Figure 7.14). *BScan* overestimates the radius of the bubble-like structures in most cases. Moreover, in around 45% of all detections the estimated radius

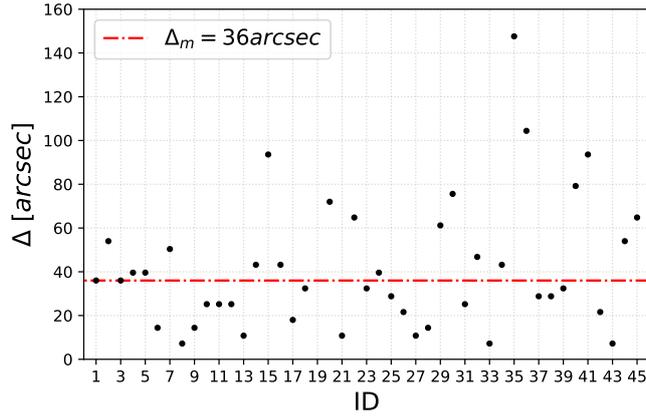


Figure 7.13: Separation Δ between the center coordinates for each ID. The median separation is $\Delta_m = 36$ arcsec.

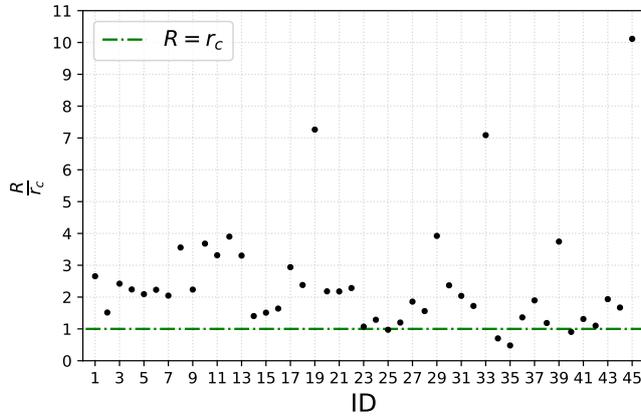


Figure 7.14: Ratio between the radius R of the bubble-like structure and the given radius r_c of the matched catalog source. Results are plotted for each ID.

ranges from $2r_c \leq R \leq 4r_c$. Only in $\approx 9\%$ of the detections the estimated radius is $R/r_c < 1$. There are three outliers, namely for the detection with the ID = 19 and ID = 33 with $R/r_c \approx 7$, which are classified as *likely* and *unlikely*, respectively. ID = 45 with $R/r_c \approx 10$, is the only source detected and classified as an HII region. The latter is shown in the Appendix (Figure A.45) in the different images of $H\alpha$, [SII], [OIII], radio and X-ray. For each image the detection regions (black), the matched catalog source (blue) and the final detected bubble-like structure (cyan) with its estimated radius are shown. The detected diffuse emission in $H\alpha$ coming from this area is larger than the size of the catalog source. The coordinates of the detected HII region are $RA = 83.89^\circ \pm 0.01^\circ$, $Dec = -66.04^\circ \pm 0.01^\circ$ (see Table 7.2 ID45). In fact, the MCELS HII region L310 (or DEM L234) with $RA = 83.890000^\circ$ and $Dec = -66.040556^\circ$ (Pellegrini et al., 2012) match the detection nicely. This shows that *BScan* does not detect

the SNR from the catalog, it detects the HII region in which the SNR is embedded. A ten times larger estimated radius of the bubble-like structure detected by *BScan* can therefore be explained.

Table 7.2 shows not only the ID and coordinates of the detected bubble-like structures which have a match in the catalog, it also shows the feature vector on which the classification is based. Among our 45 matched structures, there is no single mismatch between the predicted label and the given feature vector. Each bubble-like structure in Table 7.2 is shown in the Figures A.1 to A.45 in the Appendix. For each detection the corresponding region of the H α , [SII], [OIII], Radio and X-ray image is plotted. The DS₉ regions are overlaid. Even in the entire set of 673 detected and classified structures there is only one mismatch between the input feature vector and the output label. This means in around 0.1% of classified structures, the classification could be wrong.

A closer look at the Table 7.2 reveals, if the detected bubble-like structure is classified as *likely*, *candidate* or *unlikely* and the feature vector is therefore missing one PF to provide it with the label *SNRc*, for most of the 22 detections (see Table 7.2), the $^{[SII]}/H_{\alpha}$ or the X-ray feature is not fulfilled. For 19 detections the bubble-like structures would be classified as *SNRc* if the PF of $^{[SII]}/H_{\alpha}$ was fulfilled. This translates to $\approx 86\%$ of the detected structures with the corresponding labels. Reasons for the missing $^{[SII]}/H_{\alpha}$ feature could be the overestimation of the extent of the structure and the way the ratio is calculated (see section 6.2). If the estimated extent in equation 47 is significantly larger than the real source, the ratio of [SII] to H α can be underestimated due to the larger area of background emission. Consequently, the PF $^{[SII]}/H_{\alpha}$ may not be fulfilled. As discussed before, *BScan* systematically overestimates the extent and therefore underestimates the ratio of $^{[SII]}/H_{\alpha}$. An extremely faint emission from H α or [SII] may be the reason that the emission from these two bands cannot be detected, on the one hand. On the other hand, if the emission is detected by the network but with a detection probability lower than the threshold used in the configuration, the detection of the H α or [SII] emission can be missing as well. In both cases the PF of $^{[SII]}/H_{\alpha}$ is not fulfilled. Note that the result in section 7.2 showed an underperformance of detecting [SII] structures. Therefore, it is important to improve the estimated extent of the detected structures and the performance of the network in general.

In addition, in 18 ($\approx 82\%$) of the cases the bubble-like structures could be classified as *SNRc* if the primary X-ray feature was fulfilled. In contrast to the $^{[SII]}/H_{\alpha}$ ratio, this feature is already set if X-ray emission is detected in the corresponding image. Since the resolution in this regime is significantly lower and the number of X-ray training samples is low compared to the optical training samples, for example, the detection in the X-ray is limited to bright emission with low back-

ground. If the X-ray emission from the catalog source is faint, or it is surrounded by a higher background, the SCNN_{71-G6-TRS-F} struggles to detect a bubble-like structure. The network is trained on the pixel values of images and is therefore sensitive to changes in pixel values. Although the detection performance is higher for brighter X-ray emission, if pixels are saturated, the X-ray emission is most likely not detected as well.

7.3.3 *Unmatched Detections*

The detected bubble-like structures which have a match in the used catalog are a fraction of all detected bubble-like structures. *BScan* detects many uncatalogued structures. However, the used catalog is not a complete catalog of all bubble-like structures. Figure 7.15 shows the total number of detected and classified bubble-like structures in the LMC. The matched structures in section 7.3.2 are shown in dark blue, whereas unmatched structures are shown in teal. There are no unmatched structures classified as *SNRc* or as *likely*. Therefore, *BScan* did not detect a new SNR candidate in the LMC with the used configuration. For structures labeled as *candidate* there are only three detections without a match. These three detections are shown in the Appendix (Figures B.1 to B.3). To find possible counterparts for the detected structures, a *Simbad* coordinate query with a radius of $\Delta_m = 0.01^\circ$ is executed for each unmatched structure. The detection shown in Figure B.3 at $RA = 81.59 \pm 0.01^\circ$, $Dec = -66.25 \pm 0.01^\circ$ is most likely the HII region called MCELS L-253 (Pellegrini et al., 2012) with the center coordinates $RA = 81.5825$, $Dec = -66.2828$. For the one shown in Figure B.1 and B.2 infrared sources and a young stellar object (YSO) could be found within a radius of 0.01° , respectively. This kind of sources are of course interesting to find and to study, however, *BScan* is not able to classify them in the current version. Most unmatched structures are classified as *unlikely* and *other* as shown in Figure 7.15. In Appendix B some examples are shown in multi wavelength plots. It is beyond the scope of this thesis to analyze each detected structure. Because of the large number of detected bubble-like structures of class *unlikely* and *other* only two of each class are shown.

7.3.4 *Alternative Configurations*

Since 45 detected bubble-like structures could be matched to all 92 catalog sources, 47 sources could not be matched. Some of these 47 catalog sources are not even detected. In this part, we change the configuration of *BScan* and apply it again to the same data. We select five catalog sources which were not detected before and try to detect them with different configurations. These configurations are referred

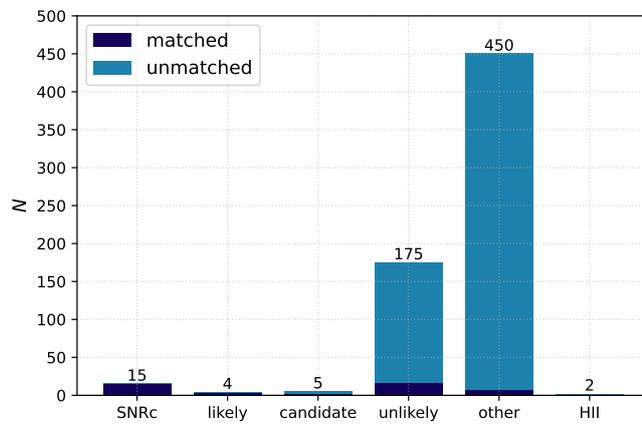


Figure 7.15: Total number of detected bubble-like structures N and the corresponding labels. The structures with matches in the catalog are colored in dark blue and unmatched structures are colored in teal. Most detected structures are labeled as *unlikely* or *other*. For both classes the vast majority is not matched to a catalog source.

Table 7.2: This table shows the list of detected bubble-like structures matched to catalog sources. Each detection has an ID, the central coordinates of the bubble-like structure in RA, Dec, the estimated radius R, the label according the classification and the feature vector the classification is based on. The name of the matched source is also given. The error on the estimated coordinates is given as $\Delta_m = 0.01^\circ$.

ID	RA $^{\pm 0.01}$ [deg]	DEC $^{\pm 0.01}$ [deg]	R [deg]	Label	H α	[SII]	[OIII]	Radio	X-ray	[SII]/H α	Catalog
1	81.98	-65.84	0.104	SNRc	1	1	1	1	0	1	J0527 – 6550
2	73.93	-68.66	0.077	SNRc	1	1	0	1	1	1	J0455 – 6839
3	72.11	-67.01	0.074	SNRc	1	1	1	1	0	1	J0448 – 6700
4	83.02	-71.01	0.056	SNRc	0	1	1	1	1	1	J0531 – 7100
5	84.39	-66.46	0.066	SNRc	0	1	0	1	1	1	J0537 – 6628
6	81.50	-66.08	0.026	SNRc	0	1	0	1	1	1	J0526 – 6605
7	86.80	-69.69	0.054	SNRc	0	0	0	1	1	1	J0547 – 6941
8	81.36	-65.99	0.084	SNRc	0	0	0	1	1	0	J0525 – 6559
9	79.94	-69.44	0.059	SNRc	0	0	0	1	1	1	J0519 – 6926
10	83.50	-69.91	0.069	SNRc	0	0	0	1	1	0	J0534 – 6955
11	81.25	-69.64	0.058	SNRc	0	0	0	1	1	0	J0525 – 6938
12	85.05	-69.34	0.039	SNRc	0	0	0	1	1	1	J0540 – 6920

Continued on next page

Table 7.2: This table shows the list of detected bubble-like structures matched to catalog sources. Each detection has an ID, the central coordinates of the bubble-like structure in RA, Dec, the estimated radius R, the label according the classification and the feature vector the classification is based on. The name of the matched source is also given. The error on the estimated coordinates is given as $\Delta_m = 0.01^\circ$. (Continued)

ID	RA $^{\pm 0.01}$ [deg]	DEC $^{\pm 0.01}$ [deg]	R [deg]	Label	H α	[SII]	[OIII]	Radio	X-ray	[SII]/H α	Catalog
13	84.56	-69.36	0.039	SNRc	0	0	0	1	1	1	J0538 – 6921
14	84.02	-70.63	0.039	SNRc	0	0	0	1	1	1	J0536 – 7039
15	83.51	-70.57	0.039	SNRc	1	1	0	0	1	1	J0534 – 7033
16	82.65	-70.11	0.074	likely	1	0	0	1	0	1	J0530 – 7008
17	78.32	-69.20	0.098	likely	0	1	0	1	0	1	J0513 – 6912
18	86.97	-70.41	0.039	likely	1	0	0	0	1	1	J0547 – 7025
19	74.97	-70.19	0.115	likely	1	0	0	1	0	1	J0459 – 7008
20	72.67	-70.83	0.103	candidate	0	0	1	1	0	0	J0450 – 7050
21	72.34	-69.34	0.049	candidate	1	0	1	1	0	0	J0449 – 6920
22	83.62	-70.55	0.059	unlikely	1	1	0	0	0	1	J0534 – 7033
23	76.69	-70.44	0.039	unlikely	0	0	0	0	1	0	J0506 – 7026
24	77.15	-69.47	0.059	unlikely	1	1	0	0	0	1	J0508 – 6928

Continued on next page

Table 7.2: This table shows the list of detected bubble-like structures matched to catalog sources. Each detection has an ID, the central coordinates of the bubble-like structure in RA, Dec, the estimated radius R, the label according the classification and the feature vector the classification is based on. The name of the matched source is also given. The error on the estimated coordinates is given as $\Delta_m = 0.01^\circ$. (Continued)

ID	RA $^{\pm 0.01}$ [deg]	DEC $^{\pm 0.01}$ [deg]	R [deg]	Label	H α	[SII]	[OIII]	Radio	X-ray	[SII]/H α	Catalog
25	79.69	-69.65	0.020	unlikely	0	0	0	1	0	0	J0518 – 6939
26	84.45	-69.17	0.020	unlikely	0	0	0	1	0	0	J0537 – 6910
27	75.51	-67.66	0.049	unlikely	1	1	0	0	0	1	J0502 – 6739
28	78.10	-67.12	0.026	unlikely	0	0	0	1	0	0	J0512 – 6707
29	82.46	-66.91	0.079	unlikely	0	0	0	0	1	0	J0529 – 6653
30	81.12	-66.39	0.079	unlikely	0	0	0	1	0	0	J0524 – 6624
31	73.69	-66.43	0.030	unlikely	0	0	0	1	0	0	J0454 – 6626
32	76.54	-65.70	0.098	unlikely	0	0	0	1	0	0	J0506 – 6541
33	76.53	-68.26	0.126	unlikely	0	0	0	1	0	0	J0506 – 6815
34	83.40	-72.05	0.020	unlikely	0	0	0	1	0	0	J0533 – 7202
35	85.79	-71.07	0.020	unlikely	0	0	0	1	0	0	J0542 – 7104
36	87.69	-68.39	0.059	unlikely	0	0	0	1	0	0	J0550 – 6823

Continued on next page

Table 7.2: This table shows the list of detected bubble-like structures matched to catalog sources. Each detection has an ID, the central coordinates of the bubble-like structure in RA, Dec, the estimated radius R, the label according the classification and the feature vector the classification is based on. The name of the matched source is also given. The error on the estimated coordinates is given as $\Delta_m = 0.01^\circ$. (Continued)

ID	RA $^{\pm 0.01}$ [deg]	DEC $^{\pm 0.01}$ [deg]	R [deg]	Label	H α	[SII]	[OIII]	Radio	X-ray	[SII]/H α	Catalog
37	85.44	-66.99	0.079	unlikely	0	0	0	0	1	0	J0541 – 6659
38	75.23	-65.20	0.079	other	1	0	0	0	0	1	J0500 – 6512
39	74.41	-67.65	0.039	other	1	1	0	0	0	0	J0457 – 6739
40	84.04	-67.56	0.039	other	0	0	1	0	0	0	J0536 – 6735
41	82.06	-67.44	0.059	other	0	1	0	0	0	0	J0528 – 6727
42	80.78	-67.88	0.039	other	0	1	0	0	0	1	J0523 – 6753
43	76.72	-65.16	0.108	other	0	1	0	0	0	0	J0506 – 6509
44	80.40	-65.70	0.039	other	0	1	0	0	0	1	J0521 – 6543
45	83.89	-66.04	0.118	HII	1	0	0	0	1	0	J0535 – 6602

to as alternative configurations (ACs). The default configuration as discussed in section 7.1 which is the basement of each configuration. For each alternative configuration the stride factor γ , the threshold τ and the *RSF* are altered. The considered sources are: J0507 – 7110 and J0527 – 7134 listed in (Bozzetto et al., 2017) (Figure 7.16 a and b), J0529 – 7004 listed in (Yew et al., 2021) (Figure 7.16 c) as well as J0508 – 6902 and J0527 – 7104 listed in (Maggi et al., 2016) (Figure 7.16 d and e).

The first alternative configuration (AC₁) uses a stride factor of $\gamma = 0.5$ but set the remaining parameters to the default values. Results of the NBDA for the five sources are shown in the multi wavelength plots in Figure 7.17. The source J0527 – 7134 (Figure 7.17 b) is detected in $H\alpha$ and radio, and the source J0508 – 6902 (Figure 7.17 d) is detected in [SII] and radio, whereas the other three sources are still not detected in any image. This demonstrates that the stride factor and the grid size has indeed an influence on the outcome of the NBDA. A smaller stride factor means a smaller grid size which results in more possible detections. This can be explained by the fact that more grid positions are distributed over a possible structure. Consequently, it increases the possibility that the network detects the structure at the proper position as expected according to the training samples. In these cases the detection probability increases and is more likely to exceed the threshold τ and the *RSF*.

For the second alternative configuration (AC₂) we set the relative strength factor to zero compared to AC₁. Comparing Figure 7.17 with Figure 7.18 there are significantly more detections in nearly every image. Each selected catalog source is detected in at least two of the five images. In addition to the bubble-like structures, the NBDA detects more point-like sources, or structures with a smaller extent. This can be seen in Figure 7.18 (b) in radio, (d) in [SII] and [OIII] and (e) in all three optical bands. The latter point-like sources are partly detected with the AC₁ as well. Furthermore, the number of detected faint structures (Figure 7.18 e, $H\alpha$), diffuse large scale emission (Figure 7.18 e, radio), small intensity fluctuations or even noise (Figure 7.18 a and b, X-ray) is increased. This comparison shows that the *RSF* effectively cleans up the NBDA results but reduces the probability to detect bubble-like structures which would be detected without using the *RSF*. The *RSF* was introduced to enhance the quality of the detection and this still holds, because most detections using the *RSF* show a bubble-like structure whereas without using the *RSF* intensity fluctuations or noise is also detected.

Compared to the AC₂ the threshold is set to $\tau = 0.5$ in the third alternative configuration (AC₃). The results for the five selected sources are shown in Figure 7.19. Similar to Figure 7.18 each source is detected in at least two bands. However, the number of detected point-like sources, diffuse large scale emission, intensity fluctuations or noise is

increased even further. The radio detections around the central source in Figure 7.19 (e), for example, cover the entire diffuse large scale emission in the west compared to Figure 7.18 (e) using the AC₂. The central structure in Figure 7.19 (b) in [OIII] is detected twice but with smaller detection regions. This is also true for X-ray in Figure 7.19 (d) and (e). This means that the classifier treats these as two distinct sources which may lead to different classification results. Although it could be of interest to detect bubble-like structures within a larger bubble-like structure, or parts of a large bubble-like structures in general, this is not an intended feature of *BScan* right now. The configuration AC₃ should be treated with caution and a low threshold τ is not recommended. In fact, the threshold τ should be as high as possible.

7.3.5 Summary

BScan is not able to detect every confirmed SNR or SNR candidate in a catalog of 92 sources with the default configuration. The classifications of the detected sources are not precise for most detections. This is mostly because of a missing counterpart detection in other wavelengths. In rare cases the classification can be misleading because of a false positive prediction of the classifier. It could be shown that the estimated center coordinates is within a radius which is smaller than the given minimum box size λ . However, this must not be true for arbitrary λ . The median separation $\Delta = 0.01^\circ$ between the detected and catalog source is used as an estimated error on the center coordinates. The radius is systematically overestimated. This leads to an underestimation of the $^{[SII]}/H\alpha$ ratio and is contributing to the misclassification of bubble-like structures. It is important to note that although there are some catalog sources missing with the default parameters, the parameters need to be chosen carefully in order to not detect every single enhanced emission coming from filaments or even noise in the considered images or region. It is worth to consider different parameter values for the stride factor γ , the threshold τ and the *RSF* depending on the input image. The threshold τ , however, should be as high as possible.

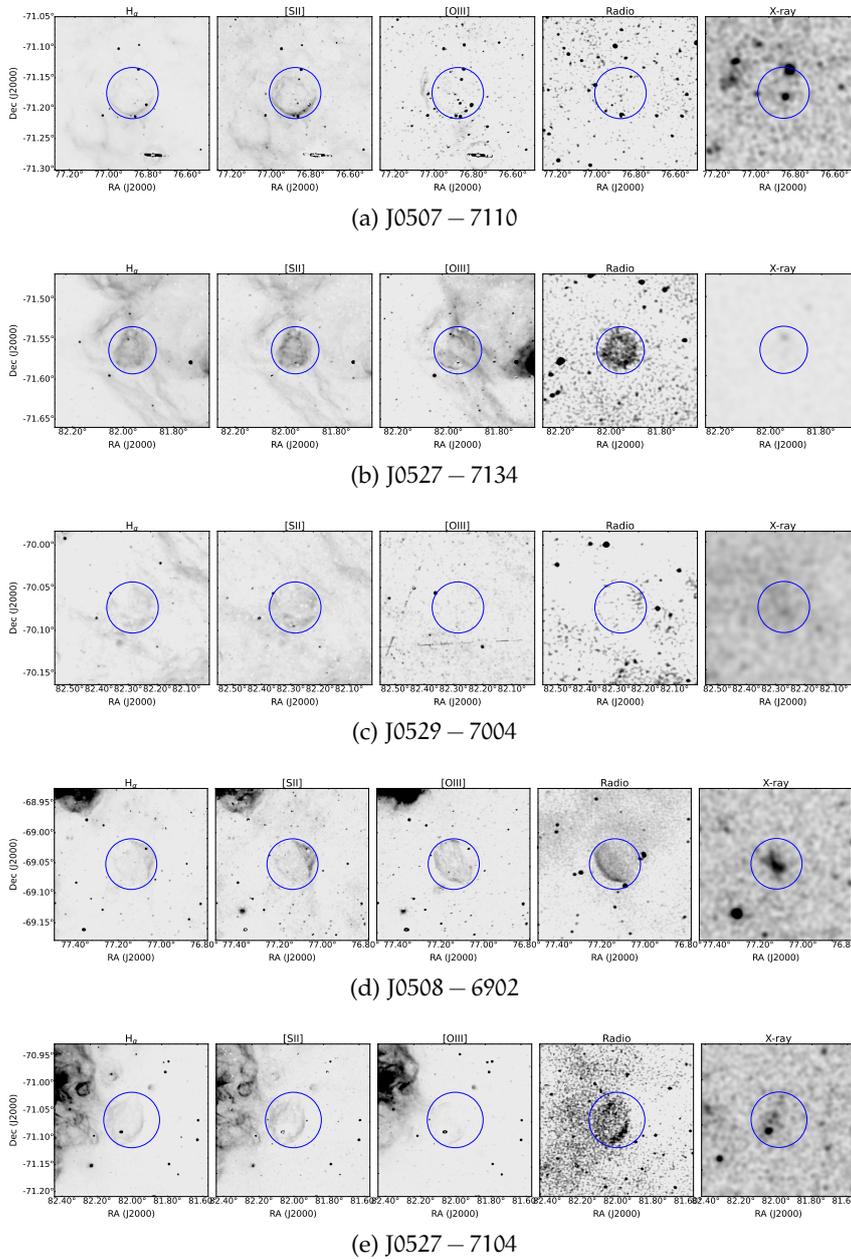
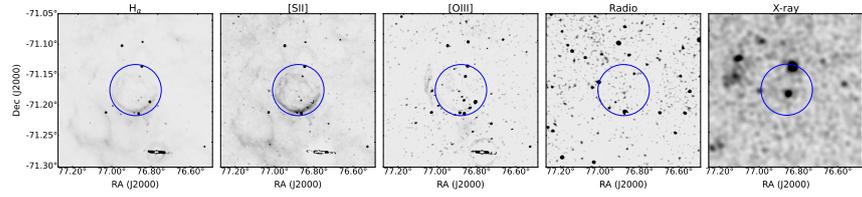
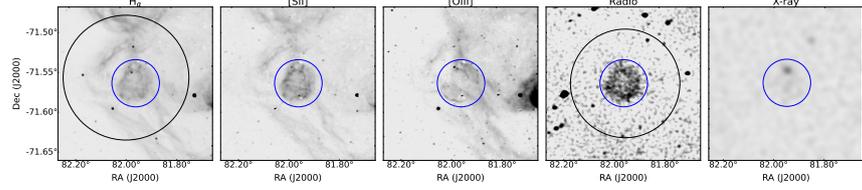


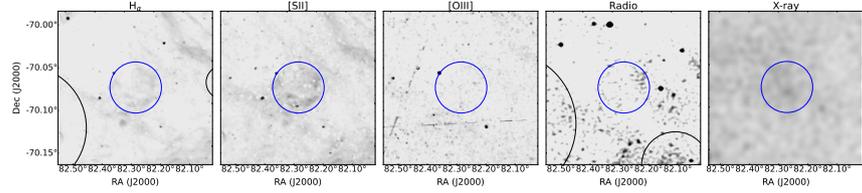
Figure 7.16: The five catalog sources which are not detected with the default *BScan* configuration. (a) and (b) show sources listed in (Bozzetto et al., 2017), (c) show a source listed in (Yew et al., 2021) and (d) and (e) show sources listed in (Maggi et al., 2016). The sources are present within the blue circles.



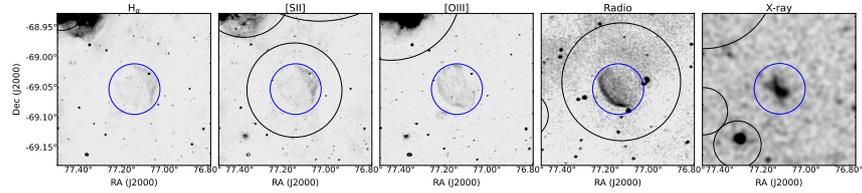
(a) J0507 – 7110



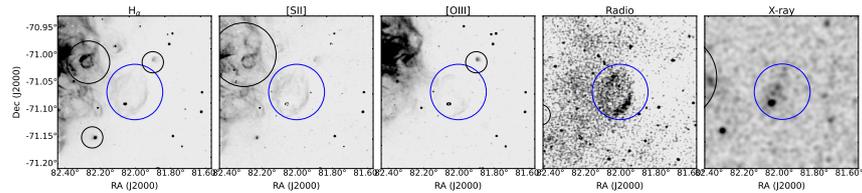
(b) J0527 – 7134



(c) J0529 – 7004



(d) J0508 – 6902



(e) J0527 – 7104

Figure 7.17: NBDA results of the five selected catalog sources using the alternative configuration AC1. The AC1 uses the default configuration but a stride factor of $\gamma = 0.5$. While the sources in (a), (c), (d) and (e) are not detected in any image, the source (blue) in (b) is detected in H α and radio (black).

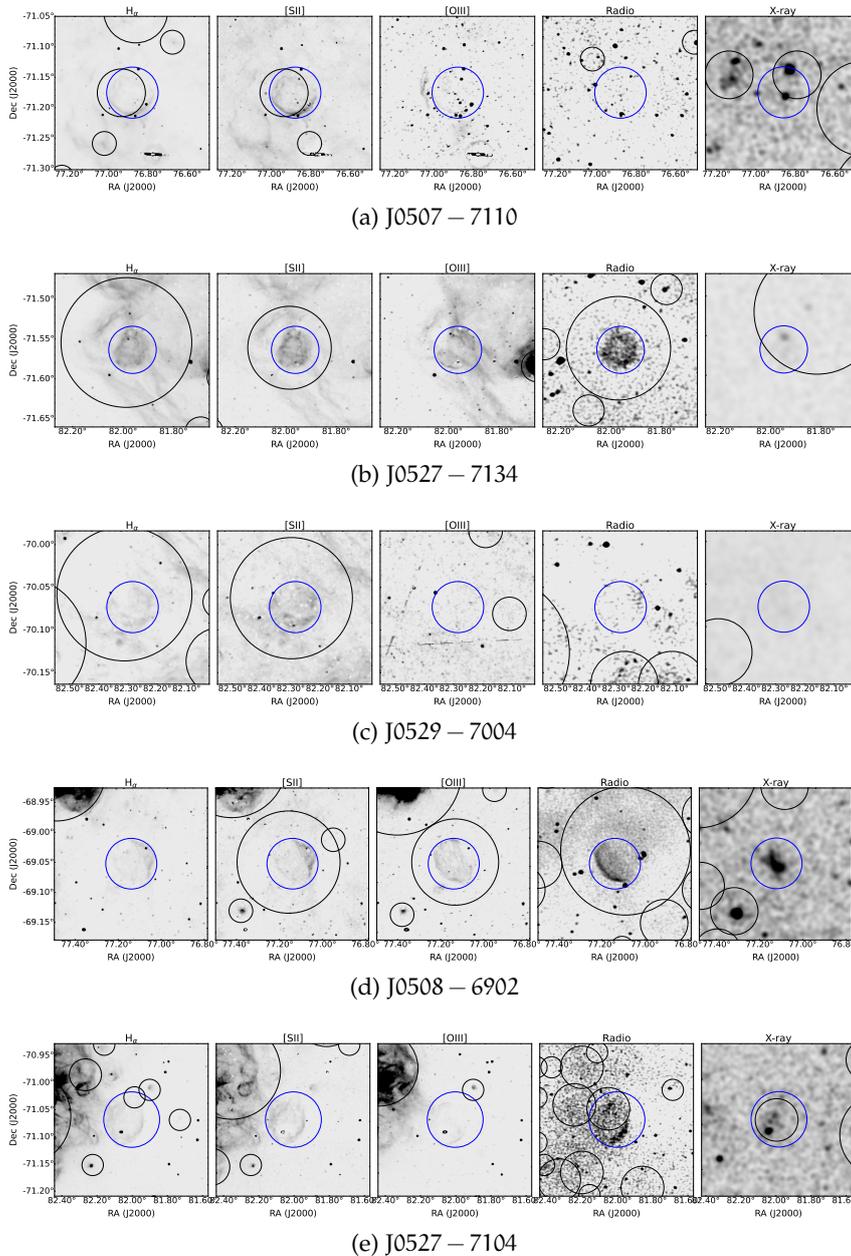
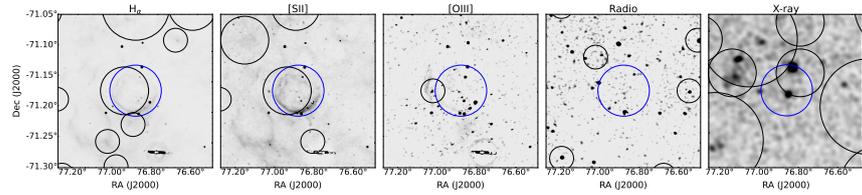
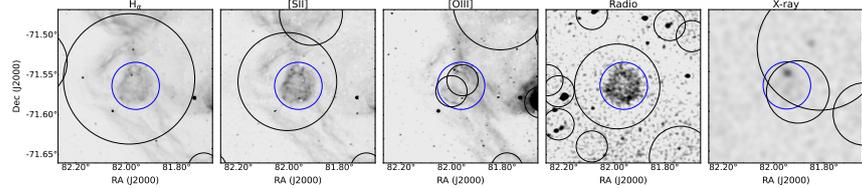


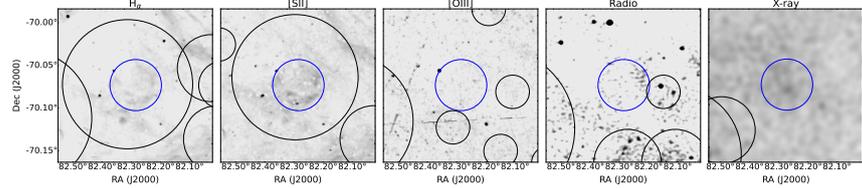
Figure 7.18: NBDA results of the five selected catalog sources using the alternative configuration AC2. The AC2 uses the default configuration but a stride factor of $\gamma = 0.5$ and a $RSF = 0$. Each source (blue) from (a) to (e) is detected (black) in at least two of the five images. In addition to the detected sources the NBDA detects extremely faint and point-like structures or intensity fluctuations in large scale emission around the central source



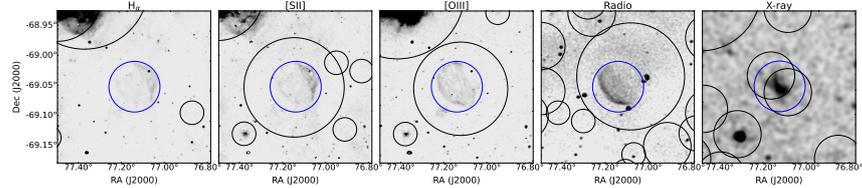
(a) J0507 – 7110



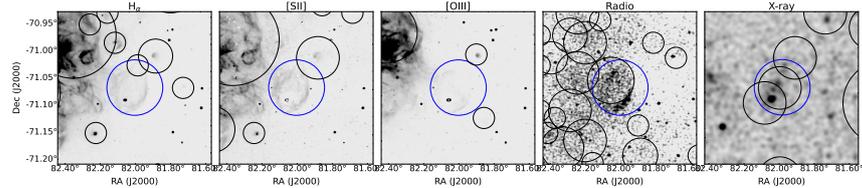
(b) J0527 – 7134



(c) J0529 – 7004



(d) J0508 – 6902



(e) J0527 – 7104

Figure 7.19: NBDA results of the five selected catalog sources using the alternative configuration AC3. The AC3 uses the default configuration but a stride factor of $\gamma = 0.5$, a $RSF = 0$ and a primary threshold of $\tau = 0.5$. Each source (blue) from (a) to (e) is detected (black) at least in two of the five images. The AC3 results in even more detections around the central sources. The AC3 is not recommended.

Description	Result	Reason/Error	Comment
structure localization	within λ	$\Delta_m = 36$ arcsec	in rare cases larger than λ
extent estimation (R)	systematically overestimated	$R \leq 4r_c$	in most cases
stride factor (γ)	impact on NBDA results	the more grid positions at a possible structure the more likely the detection	smaller γ (smaller grid size) \Rightarrow more possible detections
threshold (τ)	should be as high as possible	set threshold for detection probability given by the neural network	low threshold τ \Rightarrow noise can also be detected
Relative Strength Factor (RSF)	cleaner NBDA results but lowering chance for detecting structures with lower detection probability ($\approx \tau$)	additional threshold typed parameter to met	can be adjusted depending on the input image, lower RSF increases false positive detections
[SII]/H $_{\alpha}$ feature	most missing feature for classification	overestimation of R, missing detection	noticeable underperformance of the NBDA on [SII]
X-ray feature	second most missing feature for classification	low signal-to-background ratio, pixel saturation	less training samples available

Table 7.3: A compact summary of the main results discussed in section 7.3. For each result one or more reasons (or errors) and a comment is given.

Part IV
DISCUSSION

DISCUSSION

In part ii and iii we presented a python routine named *BScan*. It was tested on its abilities to find bubble-like structures in astronomical images. Those structures were described in part i chapter 1. This routine uses two deep neural networks, one for detecting and one for classifying bubble-like structures. At the time of writing, *BScan* focused on SNRs for the detection and classification.

In the following, we discuss methodical challenges in section 8.1, while in section 8.2 we discuss the detection and classification performance with respect to the used images. Section 8.3 deals with a more complete characterization of SNR. Here, newly discovered eROSITA SNRs and SNR candidates are discussed. They are partly used to highlight what kind of information could be used to characterize detected SNRs further (section 8.3.2 and 8.3.3). Finally, we give an outlook beyond the status quo of *BScan* in section 8.4 and give a final statement about *BScan* as an astronomical tool (section 8.5).

8.1 METHODOLOGY

In this section methodological problems and improvements will be discussed. Firstly, the biggest problem of this thesis is discussed in section 8.1.1, namely the training data set and how it could be improved in the future, possibly. In section 8.1.2 a multichannel architecture is considered to improve the entire detection and classification process. Last but not least, the role of the configuration is revisited briefly in section 8.1.3.

8.1.1 *The Training Data Set*

One of the main problems, if not the main problem, is the tiny training data set of around 1500 training samples. Furthermore, the way the training samples are generated is highly biased. Since there is no quantitative description of a shell-like or diffuse-like structure. As stated in section 4, machine learning algorithms require a high quality training data set. That means numerous samples showing clear characteristics of the corresponding class are required. Therefore, a larger training data set is necessary to improve *BScan* in the first place.

Augmentation is one possible approach to improve the training set, but can only remedy this problem to a limited extent. The reason lies in the limited transformations and therefore in the limited information that is provided to the network through augmentation.

Adding magnetohydrodynamic simulations of SNRs to the training data was proven extremely difficult, in contrast. Although it is not recommended to train a neural network on simulated data only, a number of simulated bubble-like structures could help to improve pre-training the network in the future. If the simulated data mirror the reality as much as possible, a larger set of magnetohydrodynamic simulations of SNRs could substitute the G6 or could be added to the G6 for pre-training. The pre-training of the SCNNs could be improved this way. A second approach to increase the size of the training data set could be to leverage so-called generative neural networks (GNNs). An example would be the generative adversarial network (GAN) (Goodfellow et al., 2014; Gui et al., 2020), which basically consists of two neural networks. The generator is trained on creating artificial data based on real world examples and the discriminator is trained on classifying the artificial data and real data. If the generated data are realistic enough, that the discriminator cannot distinguish between artificial and real data, the GAN could be used to set up a proper training data set of realistic but artificial training samples based on real world samples. One major problem is the discriminator. If it is not possible to train the discriminator properly to classify artificial and real data with high precision, the generator cannot be trained properly to generate realistic data, which can then be used as training samples. Therefore, this approach would be interesting but challenging at the same time. A high quality training data set should therefore be the focus of future projects to improve *BScan* further.

8.1.2 A Multichannel Neural Network

The detection of bubble-like structures rely on the neural network trained on image recognition on the one hand, and the NBDA on the other hand. The (pre-)trained and fine-tuned neural network SCNN71-G6-TRS-F suffers generalization because of the training data set discussed above. The achieved test performance is better than expected but low compared to high standard neural networks known in the machine learning space. In combination with the NBDA the neural network is the bottleneck of detecting and classifying bubble-like structures in astronomical images. As shown in chapter iii the configuration of the NBDA has a large impact on the detection results. It could be shown that structures which were not detected with the default configuration of the NBDA, could be detected by using alternative configurations, i. e. changing NBDA parameters. Parameters like the *RSF* were introduced to improve the localization of detected bubble-like structures. They are useful in some cases but contributing to the interdependencies between the detection and classification. This is an issue, since the classification depends on the NBDA. Reducing the in-

terdependencies would not only improve the classification process, it would also reduce computing time.

We suggest to introducing a single neural network with a five channel input layer (multichannel network), where each channel is carrying information of a certain wavelength or different information needed for classification. Such a network could solve dependencies within *BScan* by taking care of all available data simultaneously. This would make the bridge and the classifier obsolete and would improve *BScan* drastically, because it would unify the detection and classification process. However, this was one of the first *BScan* architecture ideas which did not work out. The reason can be found in the training data set. Currently, the same bubble-like structure visible in different wavelengths are treated as separate training samples. For a five-channel architecture a proper training data set for each channel and, therefore, for each wavelength is needed. The training data sets used throughout this work is far from the standard high quality data set needed for training such a multichannel architecture. A high quality training data set would not only improve the training of the neural network, it would also allow thinking about alternative network architectures which would reduce the interdependencies of *BScan*. Even more input information like spectral characteristics (see section 8.3.2) could be added to improve classification even further.

The classifier, however, is trained on a high quality training data set of simulated feature vectors. As shown in section 6.3 and section 7.3 the precision of the classification is considerably higher than the detection of bubble-like structures. The classifier as stand-alone tool for available feature vectors, as described in the look-up Table 6.2, is trustworthy up to an estimated false positive rate of 0.1%. That means only one classification in 1000 incidents are expected to be wrong. Nevertheless, because of the already discussed interdependencies of the classifier, the focus should be on removing the classifier from the pipeline in the future (e. g. with a multichannel network).

8.1.3 *Modification of the Configuration*

The default configuration is a good starting point to play around with the parameters for the NBDA and the classifier, but the default values may change depending on the input images. It is therefore needed to examine the input images, considering pixel size and resolution, and the parameter values given in the configuration file before applying *BScan*. In some cases it may be useful, or intended, to detect smaller bubbles or sub-bubbles in larger bubble-like structures. In these situations it could be worth to consider lower values for the RSF or the stride factor and apply *BScan* to the desired subregion only.

8.2 DETECTION AND CHARACTERIZATION PERFORMANCE

In Figure 8.1 we show the relative frequency of detected features for each classification label evaluated for every bubble-like structure in the LMC. It gives rise to which features (wavelengths) are responsible for the resulting label (see table 6.2) of the detected structure. It could be shown that *BScan* struggles to detect structures showing faint emission. Consequently, the more defined the structure appears in the available image (wavelength) the more likely it is that *BScan* detects the corresponding feature.

If a structure is labeled as *SNRc* (Figure 8.1, green) it is most likely detected in X-ray (XR) and radio (RA). The $[\text{SII}]/\text{H}\alpha$ ratio, however, is detected for less than 50% of the *SNRc* labeled structures. If the ratio is considered as a secondary feature group (SFG) (i. e. $[\text{SII}]/\text{H}\alpha > 0.67$ but only one related detection in $\text{H}\alpha$ or [SII]), $\text{H}\alpha$ is most likely the missing feature. This is also supported by the total number of feature vectors of all detected bubble-like structures in the LMC, which are labeled as *SNRc* in Figure 8.2 (green). More than half of the considered feature vectors show no $\text{H}\alpha$ feature, e. g., the vectors with the IDs 25 and 57 (see Table 6.2), which are two of the most common detected feature vectors for the label *SNRc*.

For bubble-like structures labeled as *likely* (Figure 8.1, yellow) the relative frequency of X-ray detections drop significantly while nearly all show an enhanced $[\text{SII}]/\text{H}\alpha$ ratio, but mostly as a SFG with a missing [SII] detection. Only about 25% of all detected structures labeled as *likely* show the [SII] feature. The main reason why these structures are labeled as *likely* is because they are detected in radio. The same can be stated about the detected structures with the label *candidate* (Figure 8.1, orange). In addition, the *candidates* seem to be relatively bright in the optical bands. Especially in [OIII] and $\text{H}\alpha$ *candidates* are frequently detected by *BScan*. Only a few bubble-like structures are labeled as *likely* or *candidate*. This can be seen in Figure 8.2 (yellow or orange) or in section 7.3.2 Figure 7.12.

In contrast, there are many detected structures labeled as *unlikely* (Figure 8.1, red). Comparing the relative frequencies for each feature of the *SNRc* labeled structures (Figure 8.1, green) with the ones labeled as *unlikely*, the relative frequencies of *unlikely* follow a similar pattern but have significant lower values (< 0.5). That means it is unusual to detect primary features, like radio or X-ray, together with a secondary feature, like $\text{H}\alpha$ or [SII] for example. A closer look at Figure 8.2 reveals that the three main feature vectors involved in classifying detected bubble-like structures as *unlikely*, are the vectors with the IDs 17, 41 and 49 (see Table 6.2). What is striking is that all three feature vectors show no features in the optical. Two of them are only showing the X-ray primary feature (IDs 17 and 49). *BScan* does not

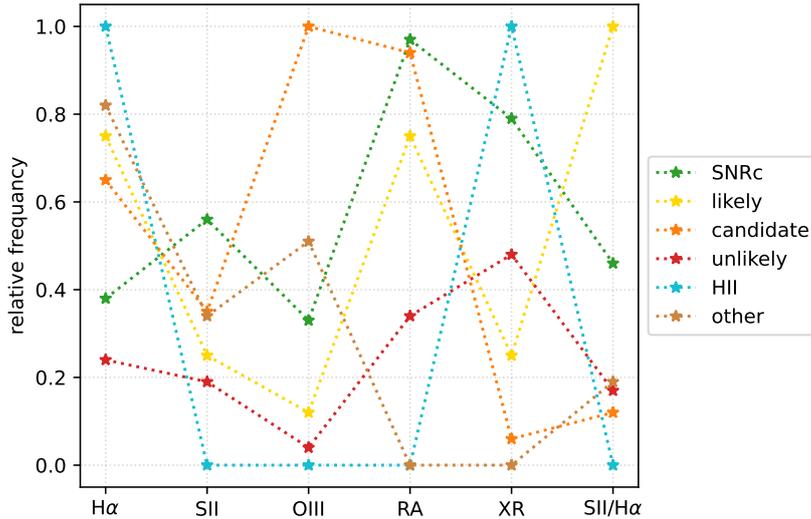


Figure 8.1: The relative frequency of detected features for each classification label evaluated for every bubble-like structure detected in the LMC.

detect a counterpart in other wavelengths for bubble-like structures detected in the X-ray image most likely.

We find the reason for that in X-ray background sources detected by *BScan*. Figure 8.3 shows the *BScan* detection results for the eROSITA image (a) and a zoomed-in region at the edge of the image where no bubble-like structures of the LMC is visible (b). There are many more detections towards the upper left corner than in the center of the image. Most X-ray emission from the LMC is coming from the central region of the image. In addition, most of those detections have the minimum size which is detectable. This leads to the conclusion that the structures detected by *BScan* are relatively small too. Indeed, many structures detected at the edge of the image have a point-like source in its center. A closer look to the zoomed-in region in Figure 8.3 (b) shows that *BScan* detects a bright point-like source (cyan), which can be identified as the active galactic nucleus (AGN) MRC0611 – 663. The detections marked as red and green circles are listed as unlabeled X-ray sources in the 1RXS (First ROSAT X-ray Survey) and 2XMM (Second XMM-Newton Serendipitous Source Catalog). Such X-ray background sources lead to a background emission which, in some cases, is recognized as a bubble-like structure. If an X-ray background sources is too far away or simply is obscured by dense clouds, there could be no optical or radio counterparts because of absorption, for example.

Since there is only one feature vector representing a HII region in Table 6.2.1 (ID 18), it is no surprise that all structures labeled as *HII* are detected in the H α and X-ray image (Figure 8.1, cyan).

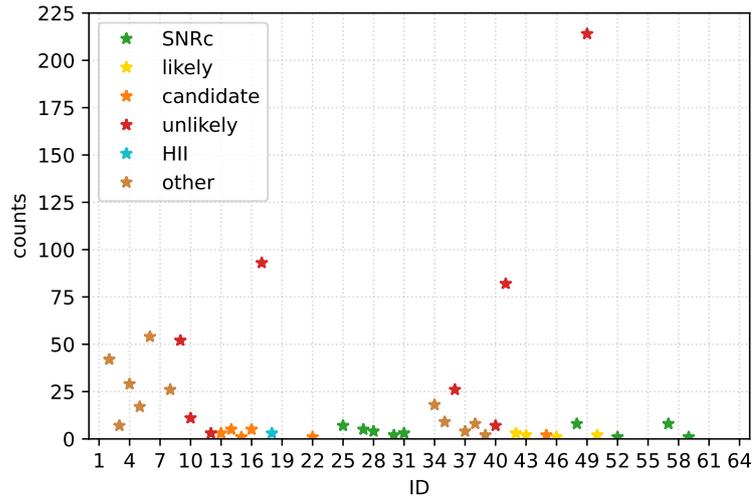


Figure 8.2: Total number of feature vectors responsible for the classification of all detected and labeled bubble-like structures in the LMC. The given ID is the feature vector ID as presented in the look-up table 6.2 in section 6.2.1.

All structures labeled as *other* (Figure 8.1, brown) show no primary feature and thus are only detected in optical wavelengths. Most of them are detected in H α .

To sum up the discussion above, the quality of the used images is of importance. The more defined the structures in the images the better the detection and classification results. Speaking about quality of the images, one crucial criteria is a well subtracted background. That would mean to remove point-like sources and reduce background emission which does not come from the considered region as good as possible. This is a task on its own for an astronomical image. Since the current version of *BScan* is focused on detecting and classifying SNRs, the detection of X-ray background emission or missing X-ray emission from SNRs lead to false positive detections and misclassifications. As shown in Figure 8.3, in some cases there are real X-ray sources like AGNs detected, but the information needed to classify these sources is simply not available to *BScan* right now. Therefore, clean images are of importance for *BScan*. This is true for the X-ray image especially.

8.3 A MORE COMPLETE CHARACTERIZATION OF SNRS

Throughout this thesis the detected bubble-like structures are classified with focus on SNRs. This is achieved by detecting bubble-like structures in the optical, radio and X-ray regime and matching the detected structures according to their coordinates with each other (see chapter 6). With this approach, it can be estimated if the detected structure is a SNR or not. However, a detailed classification of SNRs is

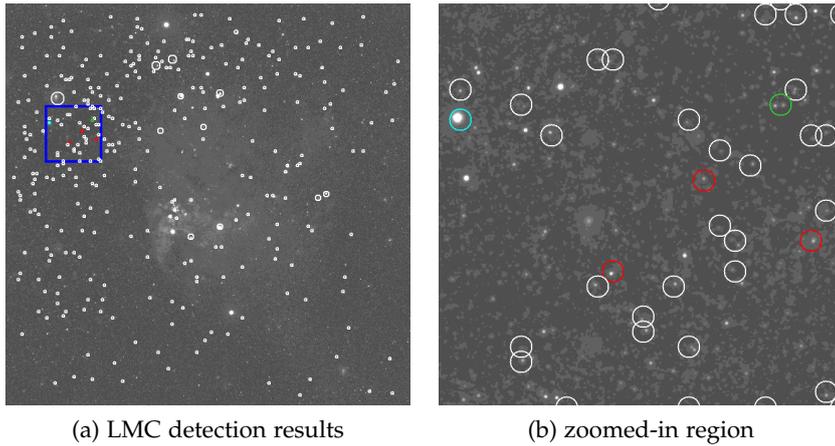


Figure 8.3: The *BScan* detection results for the LMC (a) and zoomed-in region (b). White circles show bubble-like structures as detected by *BScan*. The blue box in (a) shows the location and size of the zoomed-in region. The colored circles in (b) show some examples of X-ray background sources: (cyan) An active galactic nucleus, (red) unlabeled X-ray source of the 1RXS and (green) unlabeled X-ray source of the 2XMM catalog.

far more complicated (see section 1.2.4) and, thus, more complicated than *BScan* can handle right now. If a detected bubble-like structure is classified as a SNR, the questions arise what are the characteristics of the detected SNR and what does the remnant tell us about the exploded progenitor star, for example.

With new telescopes, which have better resolution and better sensitivity, the resulting data improve as well. With these new data, new SNRs become visible and get discovered. In Zangrandi et al. (2023, in prep.), for example, 15 new SNRs and SNR candidates in the LMC were detected with *eROSITA* data for the first time. In section 8.3.1 these sources are discussed briefly in the context of the already presented *BScan* results of the LMC (see section 7.3). Information used to classify SNRs with *BScan* are just the tip of the iceberg. The following sections 8.3.2 and 8.3.3 will discuss additional astrophysical aspects of SNRs, which were also partly used in Zangrandi et al. (2023, in prep.), and therefore would improve the characterization of SNRs if implemented into *BScan*.

8.3.1 The Newly Detected *eROSITA* Sources

Among the 15 newly detected sources there are twelve SNR candidates and three confirmed SNRs, most of which are described as faint X-ray sources. They were detected by eye using gaussian gradient magnitude (GGM) filters to highlight the shocked gas (Zangrandi et al. 2023, in prep.). *BScan* detects just one of the confirmed SNRs which is rela-

tively bright in X-ray, namely the remnant J0506 – 7009 (Figure C.13). It is labeled as *unlikely*. The only SNR candidate detected by *BScan* is the source J0510 – 6853. It is detected in the H α image (Figure C.4) only and is therefore labeled as *other*.

Because of the location of the SNR candidate J0614 – 7251, which is outside the MCELS images, only the remaining eleven candidates are presented in this thesis. All remaining 14 sources are shown in Appendix C. A short description of each new eROSITA source and an overview of the *BScan* results is also given in the Appendix C (see Table C.1).

The low *BScan* detection performance is most likely because of the faint nature of the new eROSITA sources, especially in the X-ray image used in this work. In addition, the shapes of some SNRs or SNR candidates are described as irregular or as a semi-shell in Zangrandi et al. (2023, in prep.). *BScan*, however, is only trained on shell-like and diffuse-like structures which have closed and nearly spherical shapes. According to Zangrandi et al. (2023, in prep.) some new eROSITA sources are embedded in a complex environment or in bigger structures like HII regions (see table C.1, e. g. J0543 – 6624, J0525 – 6621 or J0549 – 7001). It could be shown that *BScan* struggles to detect and clearly separate bubble-like structures in such environments (see section 7.2.4). Furthermore, the *BScan* results are achieved with the default configuration and could be improved by finding the best parameters.

8.3.2 Core Collapse and Thermonuclear Supernovae

As briefly discussed in section 1.2.4, the origin of SNRs is either a core collapse (CC) or a thermonuclear, a Type Ia, supernova. Both are naturally different and lead to different characteristics of the resulting SNR.

One major difference is the stellar remnants left behind, or their progenitor stars. While CC supernovae generally leave a neutron star or a black hole within the remnant, Type Ia supernovae do not show stellar remnants. Therefore, the detection of point-like sources within the SNR can be used to hint the origin of the remnant. Black holes are extremely hard to detect and will not be considered in this thesis. Neutron stars, however, emit a wide range of electromagnetic radiation, usually from radio to X-rays or even γ -rays (Vink, 2020). This emission can be detected by radio and X-ray telescopes and can prove the existence of the neutron star. Consequently, if a neutron star is detected in the center of a SNR, it is evident that the origin of the remnant is indeed a CC supernova.

Since core collapse supernovae have the most massive progenitor stars (Karttunen et al., 2003), the presence of massive stars nearby the remnant suggests a CC event. As mentioned in section 1.1.1 and 1.2.1,

OB-type stars create HII regions. The location of a SNR inside a HII region can therefore be considered as an indicator for a CC origin. However, the mere existence of massive stars in the vicinity of a SNR does not prove a CC origin. The star formation history (SFH) need to be considered as well. The time between the formation of a massive star and its core collapse need to be long enough to evolve into the end state. Since the lifetime of massive stars is a few million years and depends on their mass, the time span should be greater than 10^6 yr. If the SFH does not support the evolution of a possible progenitor star, a CC origin of the corresponding SNR cannot be concluded. Zangrandi et al. (2023, in prep.) estimated the number of OB stars within a radius of 100pc around each SNR and used the SFH to conclude a possible CC origin.

An example of a newly detected SNR with a possible CC origin is the remnant J0543 – 6624 (Figure C.14). It has an irregular rectangular shape and is embedded inside a HII region with a peak in the SFH at 10^7 yr ago (Zangrandi et al. 2023, in prep.).

In addition to the stellar remnants and progenitor stars as indicators of a CC origin, the two explosion mechanisms differ also in the expected emission coming from the SNR. As mentioned in section 1.2.5, Type Ia supernovae are associated with the observation of iron lines around 1keV, while CC supernovae are associated with oxygen-rich remnants. Oxygen can be observed in the optical, in [OIII], and in X-ray around 0.6keV. A SNR which shows bright emission from oxygen suggests therefore a prior CC supernova.

The reason of the different emission lies in the nature of the exploding progenitor star. The most massive stars are responsible for CC supernovae and produce heavy elements up to iron in their cores. This is possible because of their high temperatures. The iron core is surrounded by layers of heavy elements, for example oxygen (see section 1.2.4). While the star collapses, the iron core is contracting into a neutron star. The outer layers, however, will be driven into the ISM and form the SNR (Karttunen et al., 2003). Since the iron core is consumed in the neutron star, most emission of the remnant comes from the elements which surrounded the iron core, e. g., oxygen. Thermonuclear supernovae on the other hand, have their origin most likely in a thermonuclear explosion of a white dwarf (see section 1.2.4). The nuclear reactions taking place during the explosion form also heavy elements like iron. Indeed, it is evident that the amount of produced iron-group elements, e. g. iron (Fe), is about a factor of ten times larger in Type Ia than in core collapse SNe. If iron is in a proper ionization state, the remnant usually show bright iron lines around 1keV. In some cases the iron inside the remnant may not be in a high enough ionization state and is therefore not showing emission coming from iron. A reason for that could be that the iron is not completely shocked by the reverse shock (Vink, 2012). Although not every SNR with Type

Ia origin show bright iron lines, it is a common measurement to discriminate between a CC and a thermonuclear explosion scenario as done in Zangrandi et al. (2023, in prep.) for the newly confirmed SNR J056 – 7009 (Figure C.13).

According to Zangrandi et al. (2023, in prep.) the remnant show X-ray emission between 0.7keV and 1.1keV, which indicates a Type Ia origin. A faint optical shell and a semi-shell structure in the non-thermal radio image is also visible according to Zangrandi et al. (2023, in prep.). J056 – 7009 is detected by *BScan* as a diffuse bubble-like structure in the eROSITA image, but the faint shell (or semi-shell) structures in the optical (or non-thermal radio) images is not detected.

Implementing the described physics into *BScan* would upgrade the classification of SNRs. For example, provide a set of X-ray images to *BScan* which represent a soft and a medium X-ray band (see e.g. Zangrandi et al. 2023, in prep.) to discriminate between oxygen-rich and iron-rich remnants and, thus, between a possible core collapse and thermonuclear explosion scenario. This may not be straight forward for the current *BScan* version but with a multichannel network, as sketched in section 8.1.2, it could be possible in the future. In principle, such a network should also be able to detect a radio and/or an X-ray point source and narrow down the possible scenario of the SN explosion mechanism.

8.3.3 Morphologies

The history of SNRs, e.g., their origin or age, together with the overall structure of the ISM, the density distribution or the presence of molecular clouds, lead to a variety of morphologies. Each of these morphologies provide different information about the remnant. That is why the various morphologies of SNRs is a classification scheme of its own. In general, SNRs can be divided into four classes of morphologies:

The shell-type morphology is one of the most common morphologies for isolated SNRs expanding into a relatively homogeneous ISM. Shell-type SNRs are not necessarily closed shells with a circular shape but also show more irregular shapes with partly interrupted shell-like structure. The shape of a shell depends heavily on the surrounded ISM. It should be noted that SNRs of Type Ia origin tend to be more spherical than SNRs from core collapse events (Lopez et al., 2009; Lopez et al., 2011). Figure 8.4 (a) shows a shell-type SNR of a Type Ia supernova in our Milky Way.

A more complex structure of a remnant is the mixed-morphology SNR. This type is characterized by bright central X-ray emission with shell-type radio morphology. The X-ray emission coming from the central region is of thermal nature. It is believed that this kind of morphology is created by the presence of molecular clouds inside

the remnant and/or an inhomogeneous ISM where the SNR expands into (see section 1.2.4). SNRs with a mixed-morphology are usually remnants older than 10000yr evolving in denser regions of the ISM (Vink, 2020). The connection to older remnants suggests that they already merge into their surroundings. Figure 8.4 (b) shows the known mixed-morphology SNR W44 which is interacting with a dense ISM (Shelton et al., 2004).

In some cases of a CC supernova, a pulsar wind nebula (PWN) can be formed inside a SNR as described in section 1.2.4. This constellation is called a composite SNR (Vink, 2020). In contrast to mixed-morphology remnants, the PWN is emitting non-thermal X-rays and is strictly speaking not part of the SNR itself. It is fueled by the pulsar. In Figure 8.4 (c) the composite SNR G11.2 – 0.3 (Roberts et al., 2002) is shown with a clearly visible shell-type SNR.

If there is no remnant observed around the PWN it is rather referred to as a plerion. The Crab Nebula (Figure 8.4, d), possibly the best known PWN, does not have a distinct SNR but show optical line emission around its PWN which can be related to the ejecta of the SN explosion (Vink, 2012). It can therefore be considered as a SNR and a PWN (plerion) alike.

At the moment of writing, *BScan* does not have any information of SNR morphologies. In fact, for the current version it is hardly possible to implement the described morphologies into the classification process, but it would be once more a step towards a more detailed characterization of SNRs. If implemented, the output of *BScan* could give rise to the age (e. g. mixed-morphology SNRs) or the origin (spherical shape, iron-rich remnant - possibly a Type Ia SN - or detecting a radio and/or an X-ray point-source within an oxygen-rich remnant - possible CC origin) of the SNR. Again a multichannel network could recognize the different morphologies or may find even more details, hidden from the human eye. This could lead to an even better characterization of SNRs. Assuming a proper data set of SNRs showing different morphologies, a multichannel network could be a way to provide *BScan* with more astrophysical information about SNRs and improve the performance further.

8.4 BEYOND THE CURRENT BSCAN VERSION

As introduced in section 1.2 there are more bubble-like structures than just SNRs and HII regions in the ISM. In the future, *BScan* should be upgraded to be able to not only detect bubble-like structures, but also classify all of them accordingly. This is important, because each structure has its own astrophysical processes which characterize it. A deeper understanding of such astrophysical processes and the interplay of the bubble-like structures within the ISM requires the detection and characterization of many sources. The interaction of bubble-like

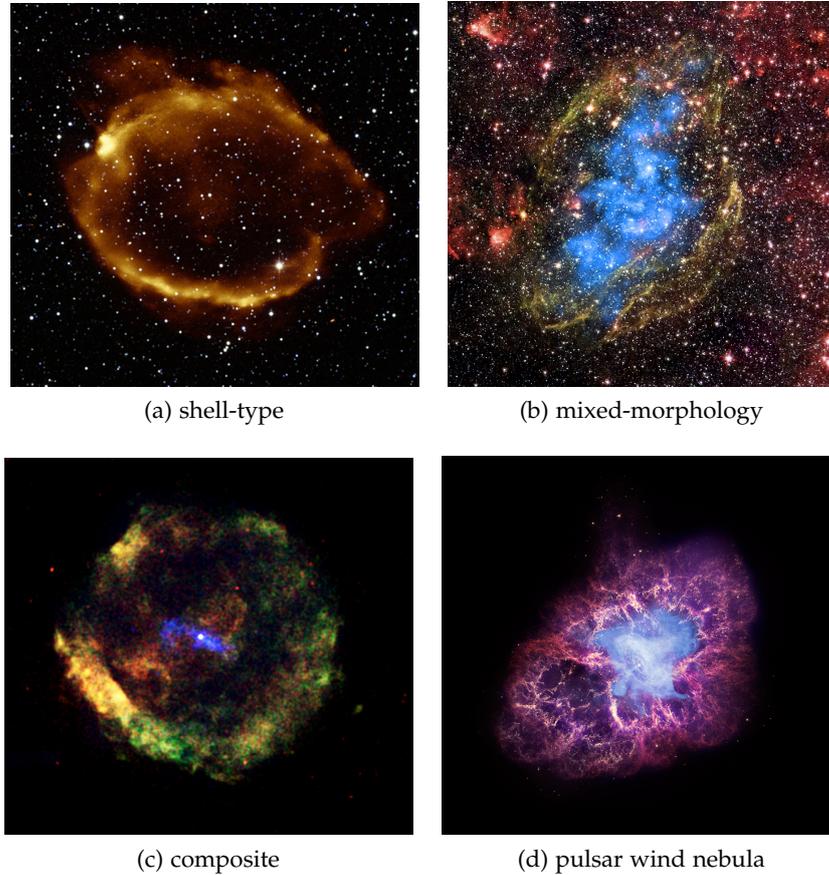


Figure 8.4: Different morphology types of SNRs: (a) A shell-type SNR (G299.2-2.9) as seen in X-ray (orange) and infrared (red, green, blue). Credit: NASA, CXC, U. Texas at Arlington, S.Park et al, ROSAT; 2MASS, UMass, IPAC-Caltech, NASA, NSF. (b) The well known mixed-morphology SNR W44 observed in X-ray (cyan) and infrared (red, green, blue). Credit: NASA, CXC, University of Georgia, R.Shelton and NASA, CXC, GSFC, R. Petre; NASA, JPL-Caltech. (c) A composite SNR (G11.2 – 0.3) in the X-ray regime with its central PWN (blue, 2.5 – 8keV) and the SNR (red 0.5 – 1.5keV, green 1.5 – 2.5keV). Credit: NASA, CXC, Eureka Scientific, M.Roberts et al. (d) The well known Crab Nebula, a pulsar wind nebula as seen in the X-ray (blue), optical (red, yellow) and infrared (purple). Credit: NASA, CXC, SAO, F.Seward; NASA, ESA, ASU, J.Hester and A.Loll; NASA, JPL-Caltech, Univ. Minn., R.Gehrz.



Figure 8.5: This images shows superbubbles inside the nebula N44. (blue) Chandra X-ray data showing hot gas heated by winds and shocks driven by stars and supernovae. (red) Infrared data from the Spitzer Space Telescope which highlight dust and cooler gas. (yellow) Optical data from the 2.2m Max-Planck-ESO telescope showing hot, young stars which are responsible for the nebula to glow due to ultraviolet radiation. Credit: NASA, CXC, U.Mich., S.Oey; NASA, JPL; ESO, WFI

structures can trigger star formation, if two shock fronts collide forming denser regions, or creating so-called superbubbles, if strong stellar winds of stellar clusters and supernovae inside the clusters creating huge cavities.

Figure 8.5 shows the nebula N44 in the LMC with the star cluster NGC1929 in its core. The stellar winds and SNe create superbubbles filled with shocked hot gas (shown in blue) surrounded by dust and cooler gas (shown in red). Superbubbles can extend up to a few hundred parsecs (Kavanagh, 2020) and emit a wide range of emission due to the involved events leading to their formation. Furthermore, Chandra observations have revealed more X-ray emission than expected from pressure-driven superbubbles (Kavanagh, 2020). That is why superbubbles are heavily studied to complete the pressure-driven model (e. g. Weaver et al. (1977)) of a superbubble formation. Indeed, new instruments will provide new data to study those huge bubble-like structures of the ISM. With the first X-ray all sky survey provided by eROSITA, superbubble population studies should be possible in the LMC (Kavanagh, 2020). An addition future X-ray telescope, probably launching in 2031, is the Athena X-ray observatory. It should be able

to reveal faint X-ray emission coming from superbubbles. Since superbubbles are created by single bubble-like structures, being able to detect all those structures forming the superbubble would result in a better understanding of the formation process and, thus, of the ISM.

New observatories in different wavelength regimes will have better resolution and a higher sensitivity to reveal probably even fainter emission coming from fundamental structures of the ISM. In case of the SNRs, this would mean they detect rather old and faint or young and too small remnants (Filipovic, 2021) in greater detail than in current data. With new high quality data new bubble-like structures will be detected. Being able to detect every single bubble-like structure would be a powerful tool for astronomy.

8.5 BSCAN AS AN ASTRONOMICAL TOOL

There are other approaches using different techniques to detect fundamental bubble-like structures in the ISM. One of these techniques is the Minkowski tensor analysis used by Collischon et al. (2021), for example. They are focusing on detecting and localizing these structures in astronomical images. However, the idea of *BScan* is to combine the detection and classification of those fundamental structures in one tool. *BScan* should be able to detect all bubble-like structures present in the ISM, including faint structures like old SNRs, for example. This is a huge challenge, especially for a machine learning approach presented in this work considering the limited training samples available right now. Therefore, we do not consider *BScan* as a reliable tool to fulfill the need of a standard tool in astronomy but with further developments, like the technical and astrophysical aspects discussed in section 8.1 and 8.3, respectively, the idea of *BScan* will become reality in the future.

Part V

APPENDIX

MATCHED BUBBLE-LIKE STRUCTURES

Here every detected bubble-like structure which has a match in the used SNR catalog is plotted in a multi wavelength plot showing the corresponding region in $H\alpha$, [SII], [OIII], radio and X-ray. The structures are classified into six different Classes: SNRc (green), likely (yellow), candidate (orange), unlikely (red), HII (cyan) and other (brown).

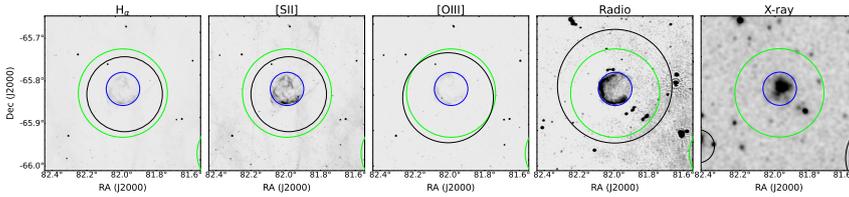


Figure A.1: This figure shows the multi wavelength plot of ID0 at RA = 81.98° , Dec = -65.84° and $R = 0.104^\circ$. The detected structure is classified as SNRc (green) and is detected in all bands except of X-ray (black). The matched SNR J0527 – 6550 is shown in blue.

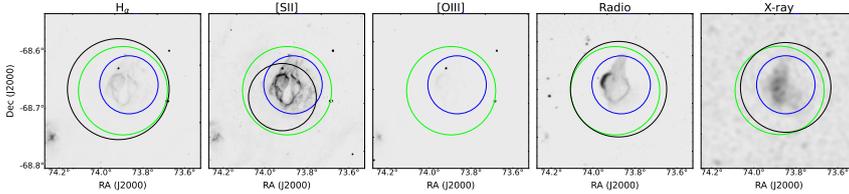


Figure A.2: This figure shows the multi wavelength plot of ID1 at RA = 73.93° , Dec = -68.66° and $R = 0.077^\circ$. The detected structure is classified as SNRc (green) and is detected in all bands except of [OIII] (black). The matched SNR J0455 – 6839 is shown in blue.

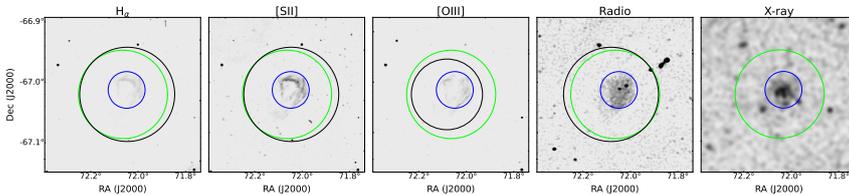


Figure A.3: This figure shows the multi wavelength plot of ID2 at RA = 72.11° , Dec = -67.01° and $R = 0.074^\circ$. The detected structure is classified as SNRc (green) and is detected in all bands except of X-ray (black). The matched SNR J0448 – 6700 is shown in blue.

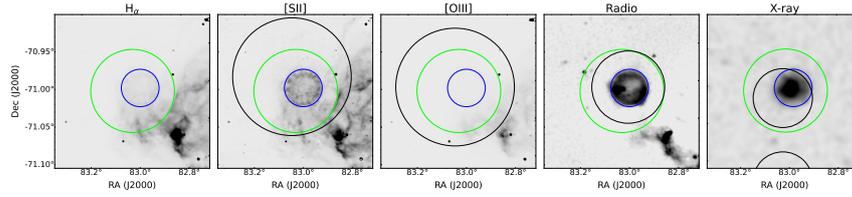


Figure A.4: This figure shows the multi wavelength plot of ID3 at RA = 83.02° , Dec = -71.01° and $R = 0.056^\circ$. The detected structure is classified as SNRc (green) and is detected in all bands except of $H\alpha$ (black). The matched SNR J0531 – 7100 is shown in blue.

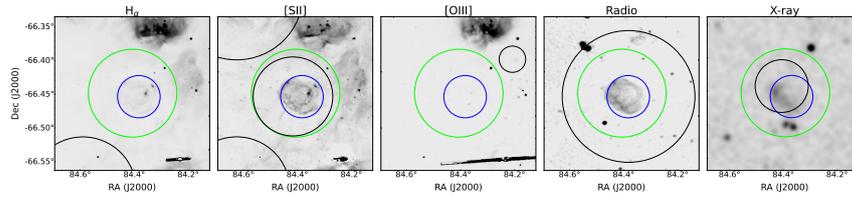


Figure A.5: This figure shows the multi wavelength plot of ID4 at RA = 84.39° , Dec = -66.46° and $R = 0.066^\circ$. The detected structure is classified as SNRc (green) and is detected in all bands except of $H\alpha$ and [OIII] (black). The matched SNR J0537 – 6628 is shown in blue.

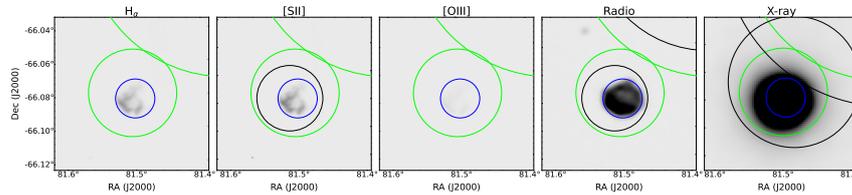


Figure A.6: This figure shows the multi wavelength plot of ID5 at RA = 81.50° , Dec = -66.08° and $R = 0.026^\circ$. The detected structure is classified as SNRc (green) and is detected in all bands except of $H\alpha$ and [OIII] (black). The matched SNR J0526 – 6605 is shown in blue.

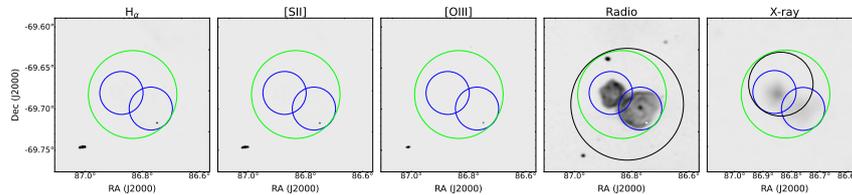


Figure A.7: This figure shows the multi wavelength plot of ID6 at RA = 86.80° , Dec = -69.69° and $R = 0.054^\circ$. The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0547 – 6941 is shown in blue.

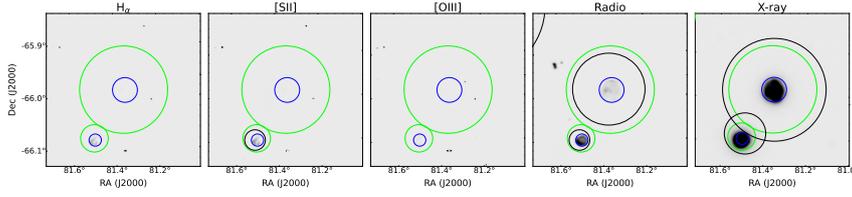


Figure A.8: This figure shows the multi wavelength plot of ID7 at RA = 81.36° , Dec = -65.99° and $R = 0.084^\circ$. The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0525 – 6559 is shown in blue.

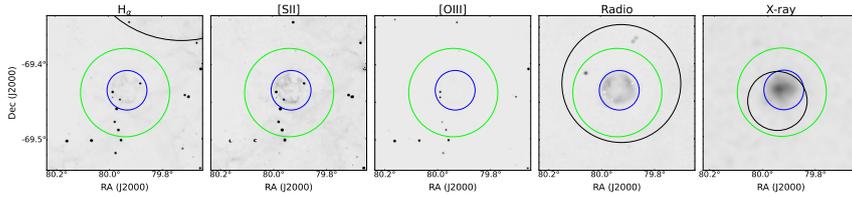


Figure A.9: This figure shows the multi wavelength plot of ID8 at RA = 79.94° , Dec = -69.44° and $R = 0.059^\circ$. The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0519 – 6926 is shown in blue.

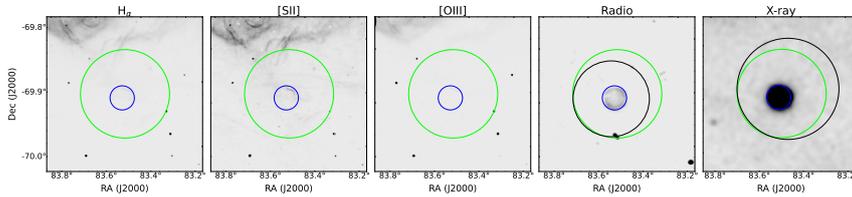


Figure A.10: This figure shows the multi wavelength plot of ID9 at RA = 83.50° , Dec = -69.91° and $R = 0.069^\circ$. The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0534 – 6955 is shown in blue.

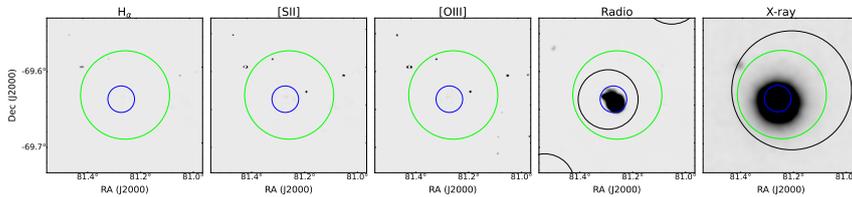


Figure A.11: This figure shows the multi wavelength plot of ID10 at RA = 81.25° , Dec = -69.64° and $R = 0.058^\circ$. The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0525 – 6938 is shown in blue.

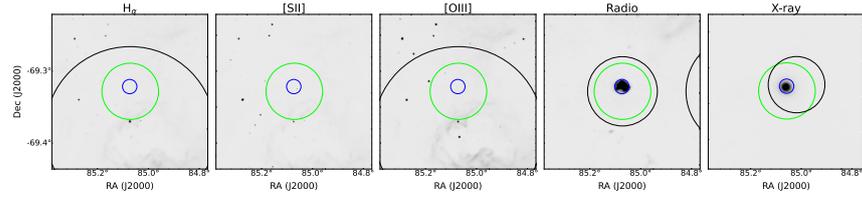


Figure A.12: This figure shows the multi wavelength plot of ID11 at RA = 85.05° , Dec = -69.34° and $R = 0.039^\circ$. The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0540 – 6920 is shown in blue.

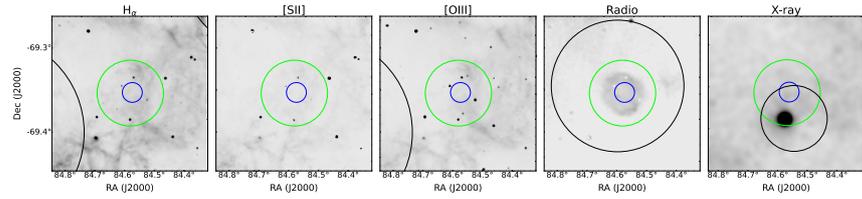


Figure A.13: This figure shows the multi wavelength plot of ID12 at RA = 84.56° , Dec = -69.36° and $R = 0.039^\circ$. The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). Although, the X-ray emission is off. The matched SNR J0538 – 6921 is shown in blue.

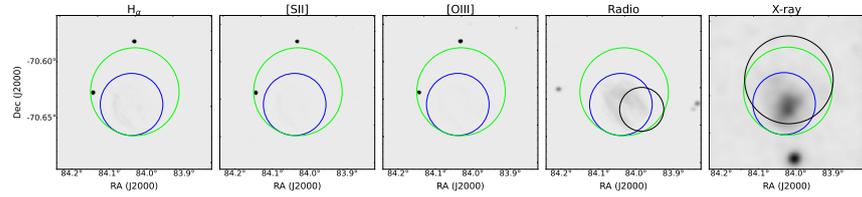


Figure A.14: This figure shows the multi wavelength plot of ID13 at RA = 84.02° , Dec = -70.63° and $R = 0.039^\circ$. The detected structure is classified as SNRc (green) and is detected in Radio and X-ray only (black). The matched SNR J0536 – 7039 is shown in blue.

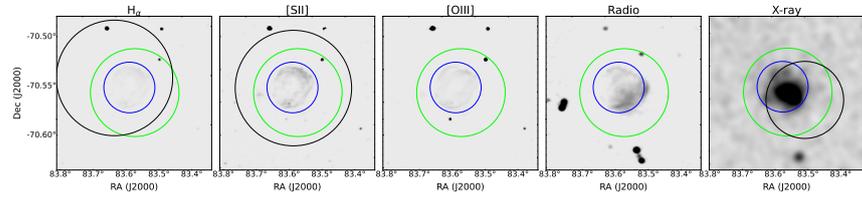


Figure A.15: This figure shows the multi wavelength plot of ID14 at RA = 83.51° , Dec = -70.57° and $R = 0.039^\circ$. The detected structure is classified as SNRc (green) and is detected in H α , [SII] and X-ray (black). The matched SNR J0534 – 7033 is shown in blue.

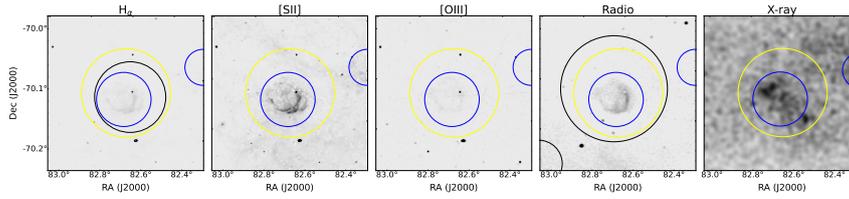


Figure A.16: This figure shows the multi wavelength plot of ID15 at RA = 82.65° , Dec = -70.11° and $R = 0.074^\circ$. The detected structure is classified as likely (yellow) and is detected in H α and Radio (black). The matched SNR J0530 – 7008 is shown in blue.

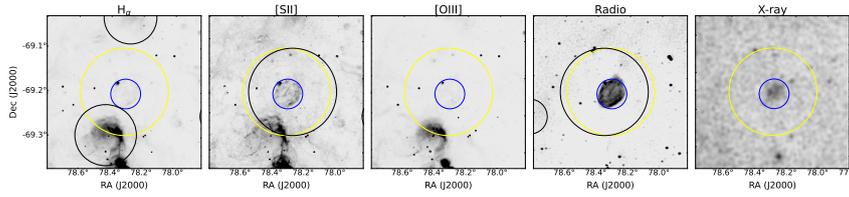


Figure A.17: This figure shows the multi wavelength plot of ID16 at RA = 78.32° , Dec = -69.20° and $R = 0.098^\circ$. The detected structure is classified as likely (yellow) and is detected in [SII] and Radio (black). The matched SNR J0513 – 6912 is shown in blue.

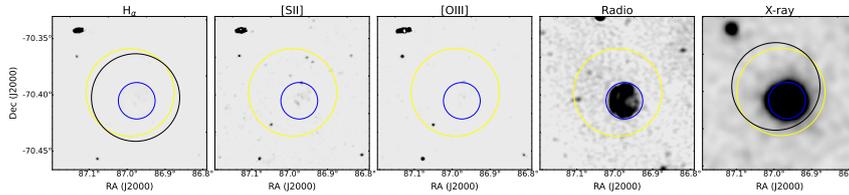


Figure A.18: This figure shows the multi wavelength plot of ID17 at RA = 86.97° , Dec = -70.41° and $R = 0.039^\circ$. The detected structure is classified as likely (yellow) and is detected in H α and X-ray (black). The matched SNR J0547 – 7025 is shown in blue.

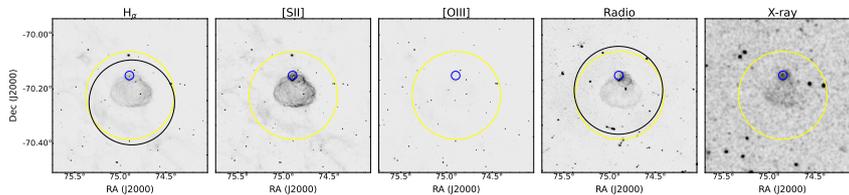


Figure A.19: This figure shows the multi wavelength plot of ID18 at RA = 74.97° , Dec = -70.19° and $R = 0.115^\circ$. The detected structure is classified as likely (yellow) and is detected in H α and Radio (black). The matched SNR J0459 – 7008 is shown in blue.

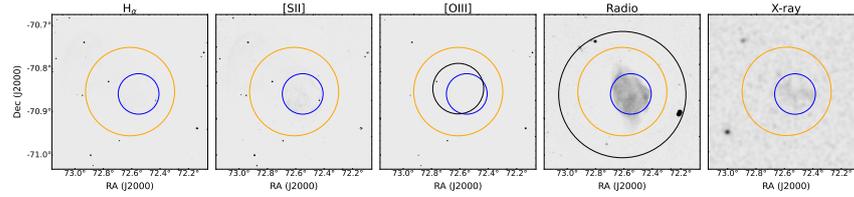


Figure A.20: This figure shows the multi wavelength plot of ID19 at RA = 72.67° , Dec = -70.83° and $R = 0.103^\circ$. The detected structure is classified as candidate (orange) and is detected in [OIII] and Radio (black). The matched SNR J0450 – 7050 is shown in blue.

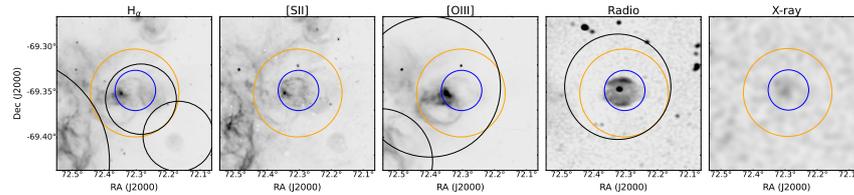


Figure A.21: This figure shows the multi wavelength plot of ID20 at RA = 72.34° , Dec = -69.34° and $R = 0.049^\circ$. The detected structure is classified as candidate (orange) and is detected in $H\alpha$, [OIII] and Radio (black). The matched SNR J0449 – 6920 is shown in blue.

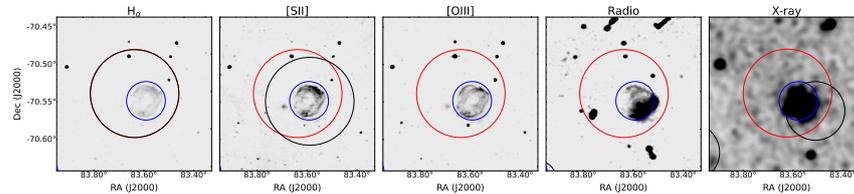


Figure A.22: This figure shows the multi wavelength plot of ID21 at RA = 83.62° , Dec = -70.55° and $R = 0.059^\circ$. The detected structure is classified as unlikely (red) and is detected in $H\alpha$ and [SII] (black). The matched SNR J0534 – 7033 is shown in blue.

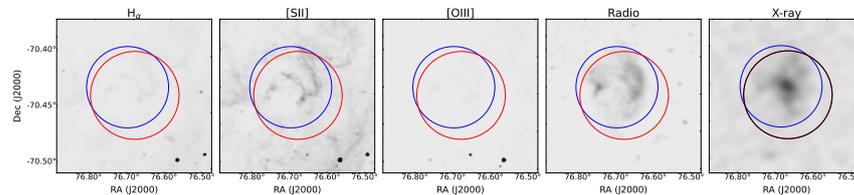


Figure A.23: This figure shows the multi wavelength plot of ID22 at RA = 76.69° , Dec = -70.44° and $R = 0.039^\circ$. The detected structure is classified as unlikely (red) and is detected in X-ray only (black). The matched SNR J0506 – 7026 is shown in blue.

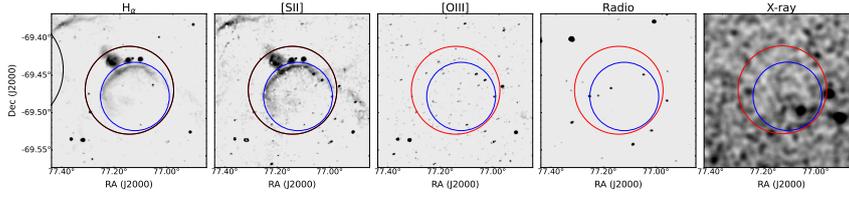


Figure A.24: This figure shows the multi wavelength plot of ID23 at RA = 77.15° , Dec = -69.47° and R = 0.059° . The detected structure is classified as unlikely (red) and is detected in H α and [SII] (black). The matched SNR J0508 – 6928 is shown in blue.

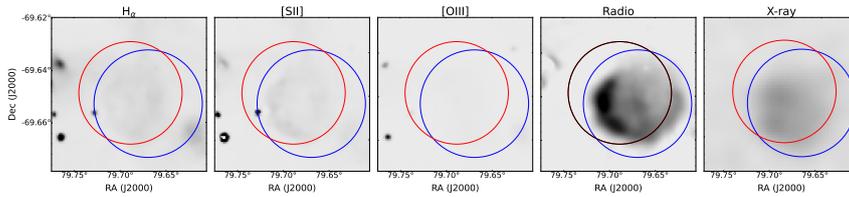


Figure A.25: This figure shows the multi wavelength plot of ID24 at RA = 79.69° , Dec = -69.65° and R = 0.020° . The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0518 – 6939 is shown in blue.

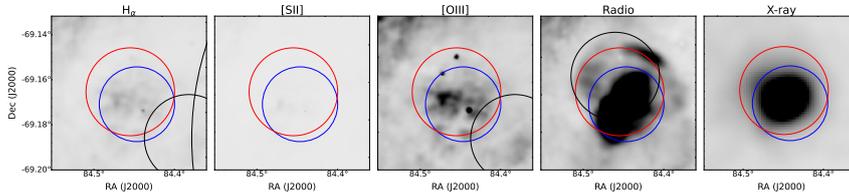


Figure A.26: This figure shows the multi wavelength plot of ID25 at RA = 84.45° , Dec = -69.17° and R = 0.020° . The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0537 – 6910 is shown in blue.

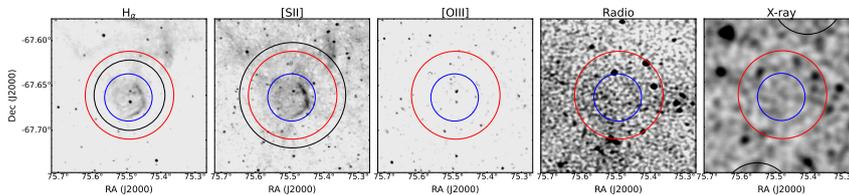


Figure A.27: This figure shows the multi wavelength plot of ID26 at RA = 75.51° , Dec = -67.66° and R = 0.049° . The detected structure is classified as unlikely (red) and is detected in H α and [SII] (black). The matched SNR J0502 – 6739 is shown in blue.

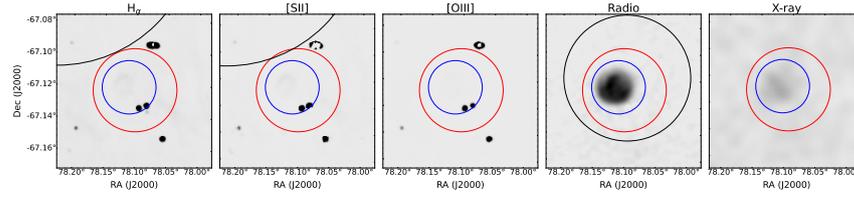


Figure A.28: This figure shows the multi wavelength plot of ID27 at RA = 78.10° , Dec = -67.12° and $R = 0.026^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0512 – 6707 is shown in blue.

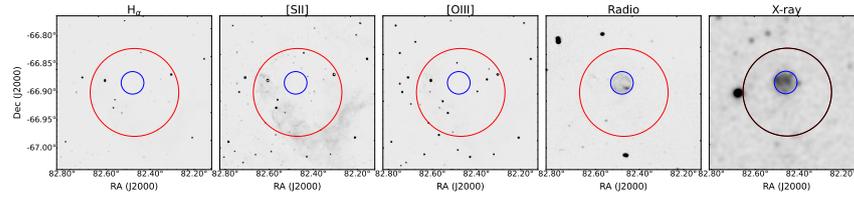


Figure A.29: This figure shows the multi wavelength plot of ID28 at RA = 82.46° , Dec = -66.91° and $R = 0.079^\circ$. The detected structure is classified as unlikely (red) and is detected in X-ray only (black). The matched SNR J0529 – 6653 is shown in blue.

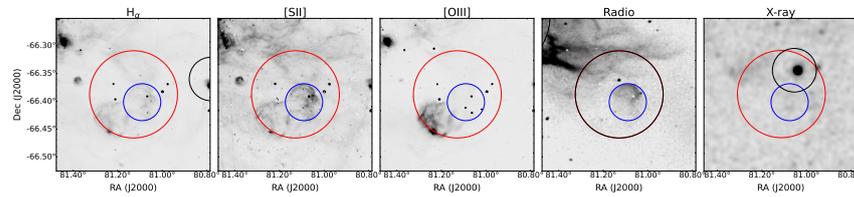


Figure A.30: This figure shows the multi wavelength plot of ID29 at RA = 81.12° , Dec = -66.39° and $R = 0.079^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0524 – 6624 is shown in blue.

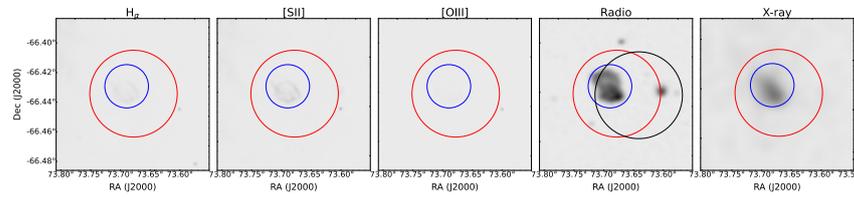


Figure A.31: This figure shows the multi wavelength plot of ID30 at RA = 73.69° , Dec = -66.43° and $R = 0.030^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0454 – 6626 is shown in blue.

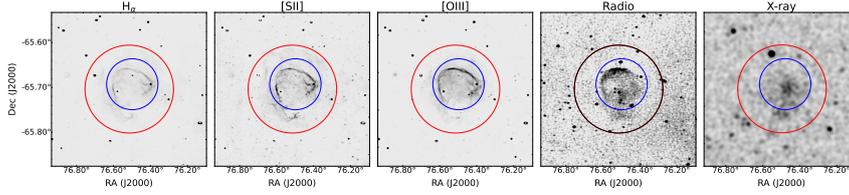


Figure A.32: This figure shows the multi wavelength plot of ID31 at RA = 76.54° , Dec = -65.70° and $R = 0.098^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0506 – 6541 is shown in blue.

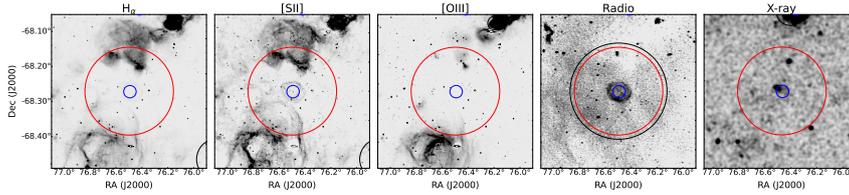


Figure A.33: This figure shows the multi wavelength plot of ID32 at RA = 76.53° , Dec = -68.26° and $R = 0.126^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0506 – 6815 is shown in blue.

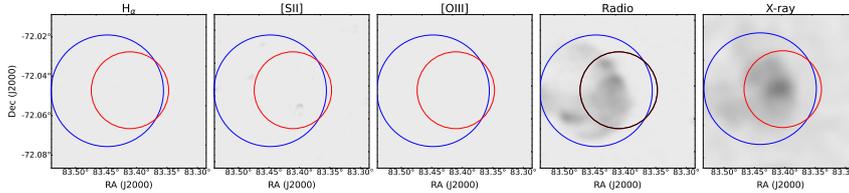


Figure A.34: This figure shows the multi wavelength plot of ID33 at RA = 83.40° , Dec = -72.05° and $R = 0.020^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0533 – 7202 is shown in blue.

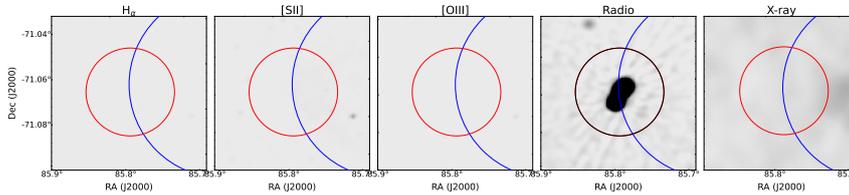


Figure A.35: This figure shows the multi wavelength plot of ID34 at RA = 85.79° , Dec = -71.07° and $R = 0.020^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0542 – 7104 is shown in blue. However, the detected radio emission is at the edge of J0542 – 7104 and comes most likely from another source.

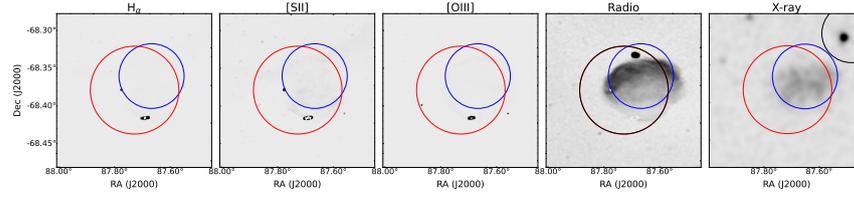


Figure A.36: This figure shows the multi wavelength plot of ID35 at RA = 87.69° , Dec = -68.39° and $R = 0.059^\circ$. The detected structure is classified as unlikely (red) and is detected in Radio only (black). The matched SNR J0550 – 6823 is shown in blue.

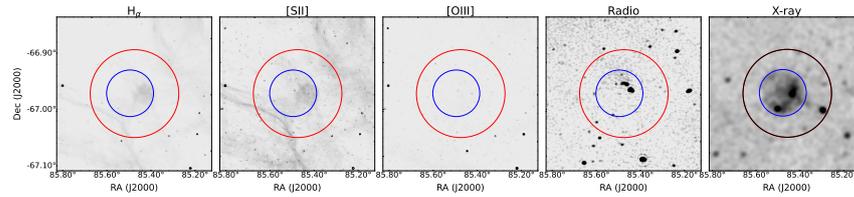


Figure A.37: This figure shows the multi wavelength plot of ID36 at RA = 85.44° , Dec = -66.99° and $R = 0.079^\circ$. The detected structure is classified as unlikely (red) and is detected in X-ray only (black). The matched SNR J0541 – 6659 is shown in blue.

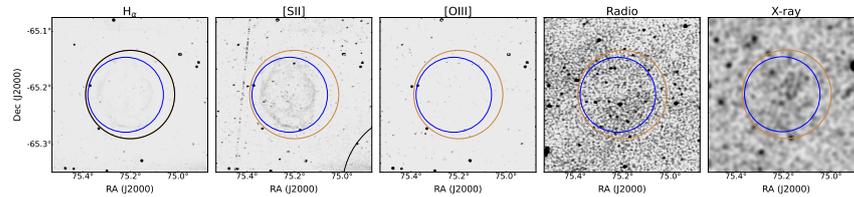


Figure A.38: This figure shows the multi wavelength plot of ID37 at RA = 75.225° , Dec = -65.20° and $R = 0.079^\circ$. The detected structure is classified as other (brown) and is detected in $H\alpha$ only (black). The matched SNR J0500 – 6512 is shown in blue.

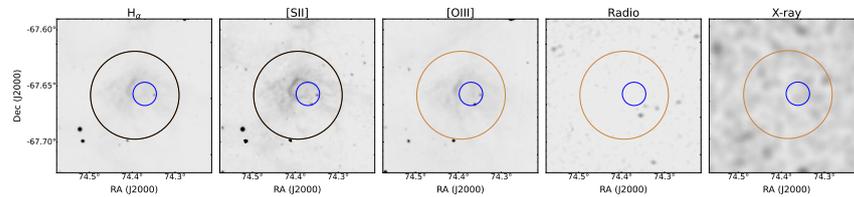


Figure A.39: This figure shows the multi wavelength plot of ID38 at RA = 74.41° , Dec = -67.653° and $R = 0.039^\circ$. The detected structure is classified as other (brown) and is detected in $H\alpha$ and [SII] (black). The matched SNR J0457 – 6739 is shown in blue.

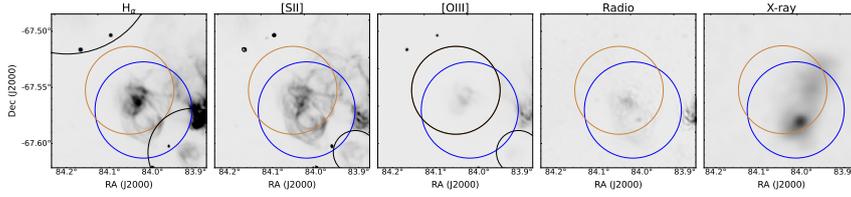


Figure A.40: This figure shows the multi wavelength plot of ID39 at RA = 84.04° , Dec = -67.56° and $R = 0.039^\circ$. The detected structure is classified as other (brown) and is detected in [OIII] (black). The matched SNR J0536 – 6735 is shown in blue.

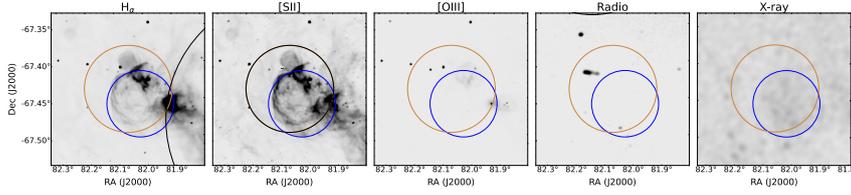


Figure A.41: This figure shows the multi wavelength plot of ID40 at RA = 82.06° , Dec = -67.44° and $R = 0.059^\circ$. The detected structure is classified as other (brown) and is detected in [SII] (black). The matched SNR J0528 – 6727 is shown in blue.

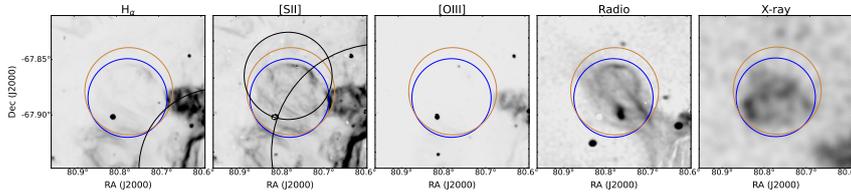


Figure A.42: This figure shows the multi wavelength plot of ID41 at RA = 80.78° , Dec = -67.88° and $R = 0.039^\circ$. The detected structure is classified as other (brown) and is detected in [SII] (black). The matched SNR J0523 – 6753 is shown in blue.

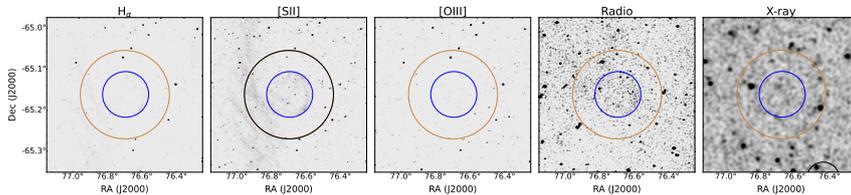


Figure A.43: This figure shows the multi wavelength plot of ID42 at RA = 76.72° , Dec = -65.16° and $R = 0.108^\circ$. The detected structure is classified as other (brown) and is detected in [SII] (black). The matched SNR J0506 – 6509 is shown in blue.

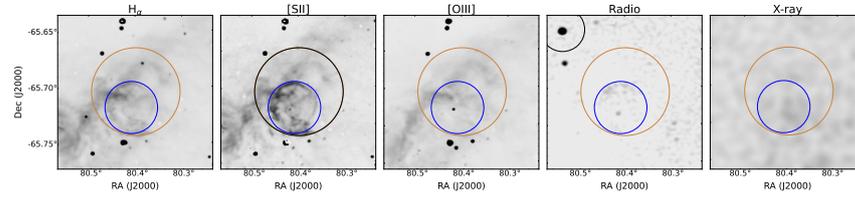


Figure A.44: This figure shows the multi wavelength plot of ID43 at RA = 80.40° , Dec = -65.70° and $R = 0.039^\circ$. The detected structure is classified as other (brown) and is detected in [SII] (black). The matched SNR J0521 – 6543 is shown in blue.

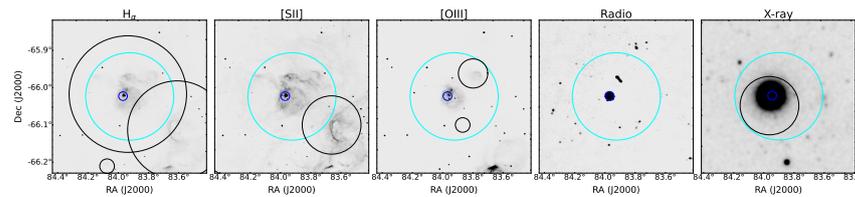


Figure A.45: This figure shows the multi wavelength plot of ID44 at RA = 83.89° , Dec = -66.04° and $R = 0.118^\circ$. The detected structure is classified as HII (cyan) and is detected in H α and X-ray (black). The matched SNR J0535 – 6602 is shown in blue.

UNMATCHED BUBBLE-LIKE STRUCTURES

CANDIDATE

In this section all unmatched bubble-like structures classified as candidate (Figures B.1 to B.3) are shown in multi wavelength plots. Within a search radius of 0.01° there are some detectable structures like HII regions, YSOs or infrared sources.

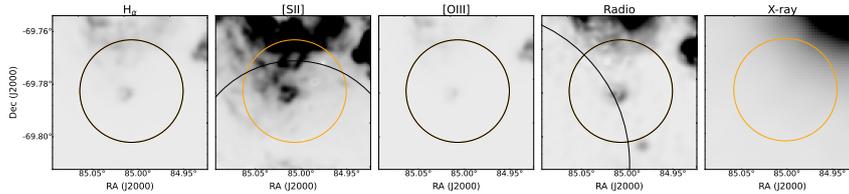


Figure B.1: This figure shows the detected bubble-like structure classified as candidate (orange) at $RA = 85.00 \pm 0.01^\circ$, $Dec = -69.79 \pm 0.01^\circ$ and radius of $R = 0.020^\circ$. It is detected in $H\alpha$, [OIII] and radio (black). There is a YSO within a radius of 0.01° .

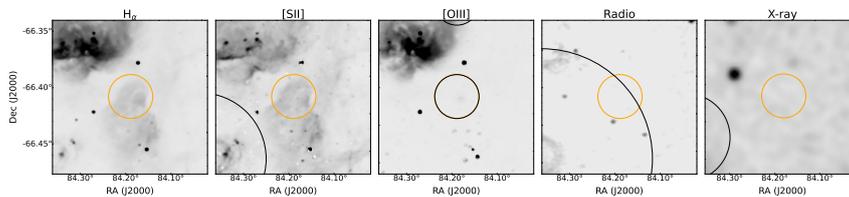


Figure B.2: This figure shows the detected bubble-like structure classified as candidate (orange) at $RA = 84.18 \pm 0.01^\circ$, $Dec = -66.41 \pm 0.01^\circ$ and radius of $R = 0.020^\circ$. It is detected in [OIII] and radio (black). Although, the detection in radio is at a different position and much larger compared to the one in [OIII]. It is likely that in radio another emission is detected. Within a radius of 0.01° there are some infrared sources.

UNLIKELY AND OTHER

In this section unmatched bubble-like structures classified as unlikely and other are shown in a multi wavelength plot. Because of the large number of detected and classified structures into these classes, only two of each class are shown. Figures B.4 and B.5 show examples of class unlikely, while Figures B.6 and B.7 show examples of class other.

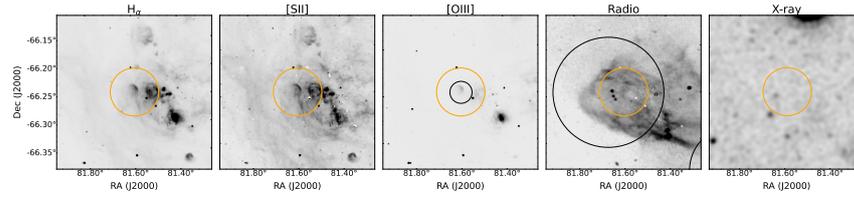


Figure B.3: This figure shows the detected bubble-like structure classified as candidate (orange) at $RA = 81.59 \pm 0.01^\circ$, $Dec = -66.25 \pm 0.01^\circ$ and radius of $R = 0.043^\circ$. It is detected in [OIII] and radio (black). In radio, however, the larger structure is detected. There is an HII region within in the reach of 0.01° named L-253.

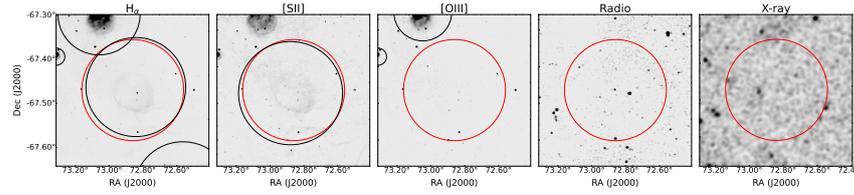


Figure B.4: This figure shows the detected bubble-like structure classified as unlikely (red) at $RA = 72.91 \pm 0.01^\circ$, $Dec = -67.10 \pm 0.01^\circ$ and radius of $R = 0.115^\circ$. It is detected in H α and [SII] (black).

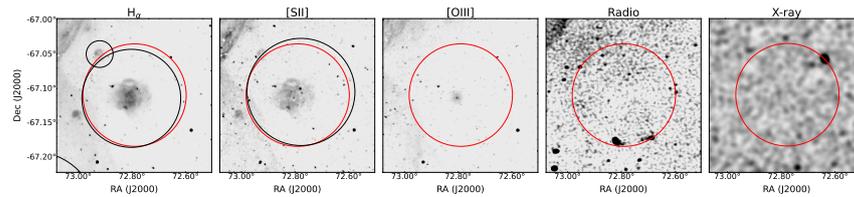


Figure B.5: This figure shows the detected bubble-like structure classified as unlikely (red) at $RA = 72.82 \pm 0.01^\circ$, $Dec = -67.45 \pm 0.01^\circ$ and radius of $R = 0.075^\circ$. It is detected in H α and [SII] (black).

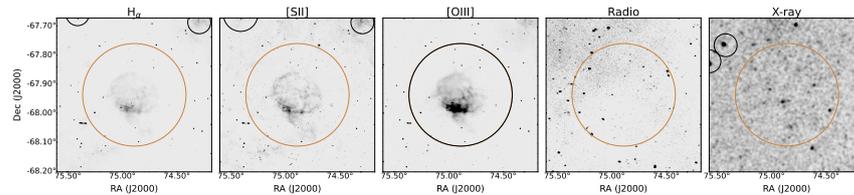


Figure B.6: This figure shows the detected bubble-like structure classified as other (brown) at $RA = 74.94 \pm 0.01^\circ$, $Dec = -67.92 \pm 0.01^\circ$ and radius of $R = 0.177^\circ$. It is detected in [OIII] only (black).

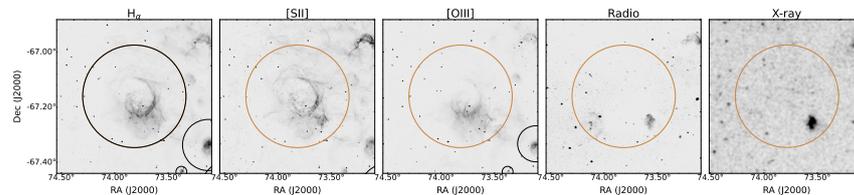


Figure B.7: This figure shows the detected bubble-like structure classified as other (brown) at $RA = 73.88 \pm 0.01^\circ$, $Dec = -67.14 \pm 0.01^\circ$ and radius of $R = 0.186^\circ$. It is detected in H α only (black).

NEW EROSITA SNRS AND SNR CANDIDATES

This section show new confirmed SNRs and SNR candidates in the LMC detected with eROSITA data for the first time (Zangrandi et al. 2023, in prep.). All descriptions of the remnants from Figure C.1 to C.14 are based on Zangrandi et al. (2023, in prep.).

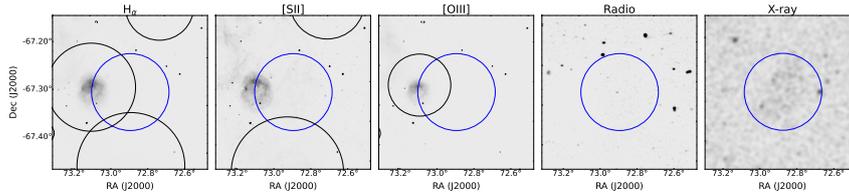


Figure C.1: J0451 – 6717: This new eROSITA SNR candidate (blue) is located next to structures visible in the optical images. Indeed, *BScan* detects the bubble-like structure closest to the SNR candidate in $H\alpha$ and $[OIII]$. However, no clear bubble-like structure can be seen at the location of the candidate. In X-ray there is some emission coming from the considered region, but is not detected by *BScan*, unfortunately.

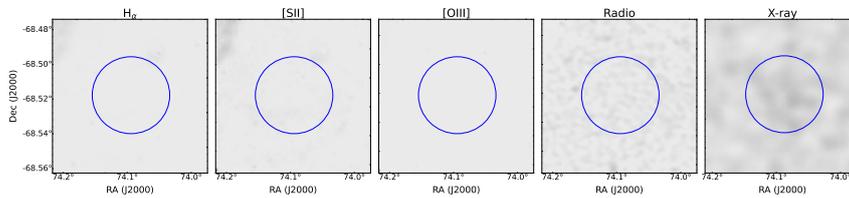


Figure C.2: J0456 – 6830: For this SNR candidate (blue) there is no clear bubble-like structure visible in any image used for this work. Therefore, *BScan* does not detect this newly detected candidate.

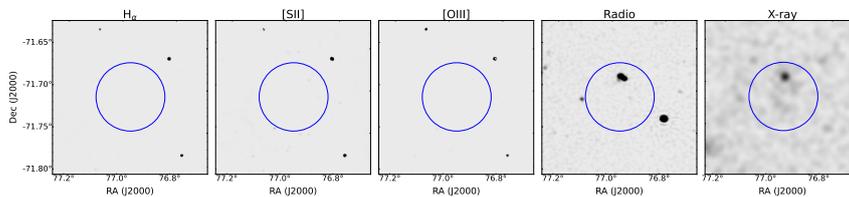


Figure C.3: J0507 – 7143: There is no emission visible in the optical images for this SNR candidate (blue). The point source seen in the non-thermal radio and X-ray image is an active galactic nucleus (Zangrandi et al. 2023, in prep.). In the X-ray image there may be some emission coming from the SNR candidate, but it is too faint to be detected by *BScan*.

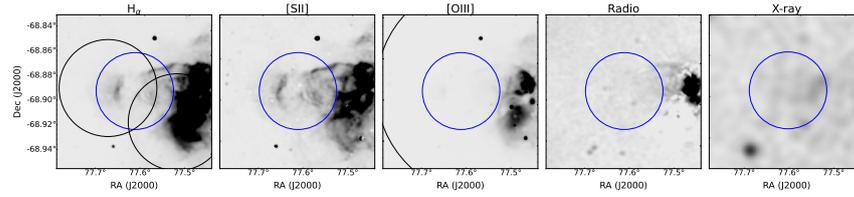


Figure C.4: J0510 – 6853: In this region of the new SNR candidate (blue) a shell-like structure is visible in the optical images (except for [OIII]). This structure is partly detected in the $H\alpha$ image (black). In radio and X-ray there is also some faint emission but is not detected by *BScan*.

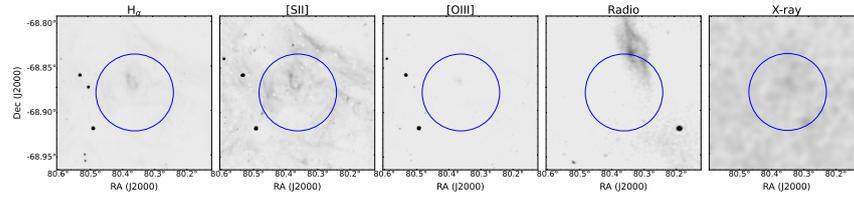


Figure C.5: J0521 – 6853: This eROSITA candidate (blue) is located within a bigger complex optical structure where some subregions show an enhanced $[SII]/H\alpha$ ratio. There may be some enhanced X-ray emission too, but such faint emission is hardly possible to detect by *BScan*.

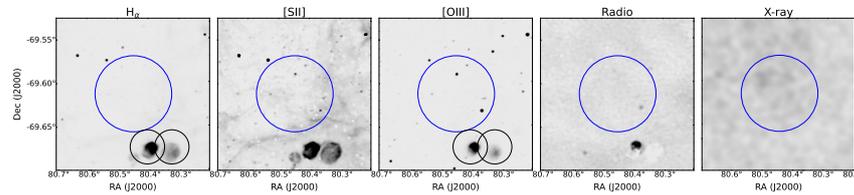


Figure C.6: J0521 – 6936: For this SNR candidate no clear structure can be seen in all five images used. However, there could be faint emission within the candidate (blue) in the X-ray image. This emission is not detected by *BScan*.

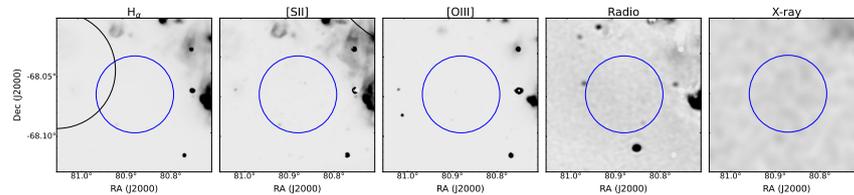


Figure C.7: J0523 – 6804: This candidate (blue) is extremely faint in images used in this work. Consequently, *BScan* is not able to detect J0523 – 6804 detected by Zangrandi et al. (2023, in prep.)

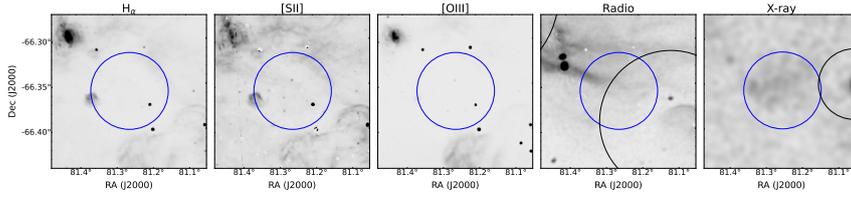


Figure C.8: J0525 – 6621: This eROSITA SNR candidate (blue) is located in between some bubble-like structure in the optical and radio band. In the non-thermal radio image *BScan* detects one of those structures in the south. In the X-ray image a faint diffuse emission within the candidate's region is visible indeed, but not detected by *BScan*.

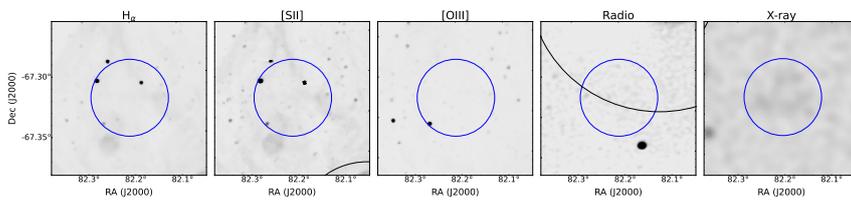


Figure C.9: J0528 – 6719: For this SNR candidate (blue) a faint but visible shell-like emission can be spotted in the X-ray image. Unfortunately, it is not detected by *BScan*. In the the optical and radio band no bubble-like structures are visible.

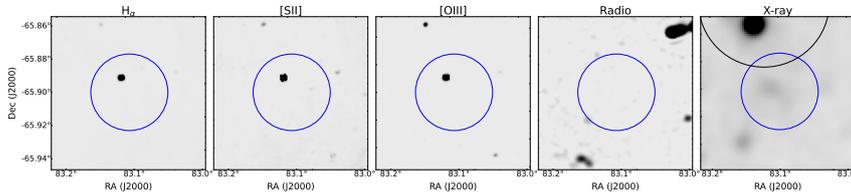


Figure C.10: J0532 – 6554: For this SNR candidate no optical and radio emission is visible. In X-ray a very faint diffuse emission can be seen within the blue circle. *BScan* does not detect this candidate.

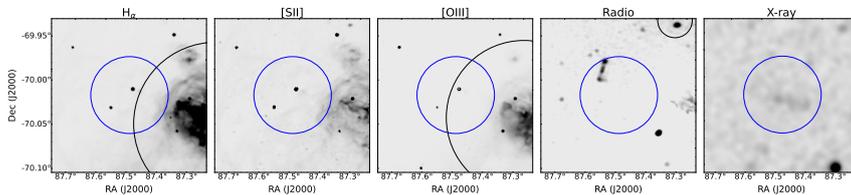


Figure C.11: J0549 – 7001: This SNR candidate is not detected by *BScan*. It is located next to a bright and complex region in the optical band. No non-thermal radio emission which is correlated with the SNR candidate can be spotted. In X-ray a faint diffuse and irregular shape is visible in the region of the remnant (blue).

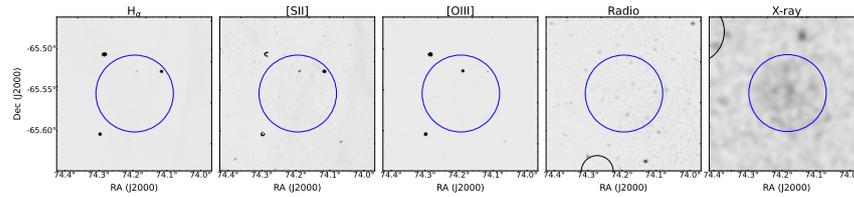


Figure C.12: J0456 – 6533: For this SNR no structure is visible in the optical and radio band. In X-ray, however, a faint diffuse emission can be spotted within the blue circle. This source is not detected by *BScan*.

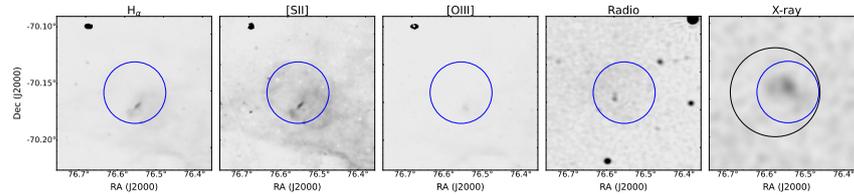


Figure C.13: J0506 – 7009: This confirmed SNR (blue) is detected in the X-ray regime (black). According to Zangrandi et al. (2023, in prep.) the SNR is brightest in the mid X-ray range, which supports a thermonuclear supernova scenario, and faint shell is visible in [SII]. Zangrandi et al. (2023, in prep.) also spotted a non-thermal radio semi-shell. However, *BScan* does not detect any structure in optical or radio.

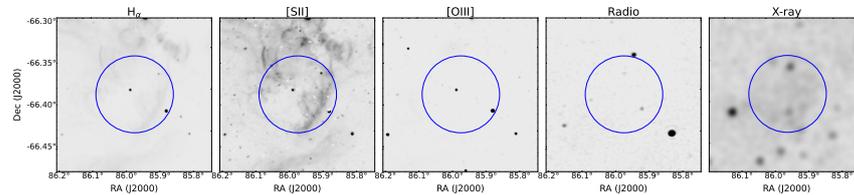


Figure C.14: J0543 – 6624: An eROSITA SNR (blue) with an irregular shape brightest in soft X-rays (Zangrandi et al. 2023, in prep.). A shell-like structure is visible in [SII] but not detected by *BScan*.

Table C.1: This table shows the new eROSITA SNRs and SNR candidates detected by Zangrandi et al. (2023, in prep.). The coordinates (J2000), the current status and a short description of each source is given. If the corresponding source is detected by *BScan* the emission and the label is added for each source as well. * as described in Zangrandi et al. (2023, in prep.)

Source	RA [deg]	DEC [deg]	R [deg]	status*	description*	detected by <i>BScan</i>	label
J0451 – 6717	72.9405	–67.2900	0.0812	candidate	diffuse X-ray elongated shell in H α	/	/
J0456 – 6830	74.1063	–68.5145	0.0222	candidate	faint diffuse X-ray possible Type Ia SN	/	/
J0507 – 7143	76.9615	–71.7115	0.0406	candidate	faint diffuse X-ray possible Type Ia SN	/	/
J0510 – 6853	77.6180	–68.8915	0.0312	candidate	faint X-ray optical shell partly shell-like non-thermal radio possible CC origin	H α	other
J0521 – 6853	80.3604	–68.8793	0.0432	candidate	bright X-ray env. elongated non-thermal radio diffuse optical	/	/
J0521 – 6936	80.4531	–69.6136	0.0436	candidate	bright X-ray env. shell-like optical	/	/

Continued on next page

Table C.1: This table shows the new eROSITA SNRs and SNR candidates detected by Zangrandi et al. (2023, in prep.). The coordinates (J2000), the current status and a short description of each source is given. If the corresponding source is detected by *BScan* the emission and the label is added for each source as well. * as described in Zangrandi et al. (2023, in prep.) (Continued)

Source	RA [deg]	DEC [deg]	R [deg]	status*	description*	detected by <i>BScan</i>	label
J0523 – 6804	80.8781	–68.0668	0.0323	candidate	circular X-ray non-thermal radio possible Type Ia SN	/	/
J0525 – 6621	81.2614	–66.3569	0.0428	candidate	irregular X-ray complex non-thermal radio env. possible Type Ia SN	/	/
J0528 – 6719	82.2073	–67.3204	0.0318	candidate	half-shell X-ray possible CC origin	/	/
J0532 – 6554	83.1020	–65.9032	0.0231	candidate	diffuse X-ray	/	/
J0549 – 7001	87.4572	–70.0291	0.0440	candidate	elongated X-ray filamentous optical possible Type Ia SN	/	/

Continued on next page

Table C.1: This table shows the new eROSITA SNRs and SNR candidates detected by Zangrandi et al. (2023, in prep.). The coordinates (J2000), the current status and a short description of each source is given. If the corresponding source is detected by *BScan* the emission and the label is added for each source as well. * as described in Zangrandi et al. (2023, in prep.) (Continued)

Source	RA [deg]	DEC [deg]	R [deg]	status*	description*	detected by <i>BScan</i>	label
J0456 – 6533	74.2153	–65.5472	0.0473	confirmed	circular X-ray optical shell faint radio shell possible Type Ia SN	/	/
J0506 – 7009	76.5661	–70.1555	0.0272	confirmed	diffuse X-ray faint optical shell non-thermal radio semi-shell possible Type Ia SN	X-ray	unlikely
J0543 – 6624	85.9524	–66.3974	0.0464	confirmed	irregular X-ray similar optical shape embedded in HII region possible CC origin	/	/

BIBLIOGRAPHY

- About ImageNet* (2023). <https://www.image-net.org/about.php>. Accessed: 2023-03-29.
- About WordNet* (2023). <https://wordnet.princeton.edu/>. Accessed: 2023-03-29.
- Beaumont, Christopher N., Alyssa A. Goodman, Sarah Kendrew, J. P. Williams, and Robert Simpson (2014). "The Milky Way Project: Leveraging Citizen Science And Machine Learning To Detect Interstellar Bubbles." In: *The Astrophysical Journal Supplement Series* 214.1, p. 3. DOI: 10.1088/0067-0049/214/1/3.
- Bentley, J.L. (1975). "Multi Dimensional Binary Search Trees Used Associative Searching." In: *Communication of the ACM* 18, pp. 509–517.
- Bozzetto, Luke M. et al. (2017). "Statistical Analysis of Supernova Remnants in the Large Magellanic Cloud." In: *The Astrophysical Journal Supplement Series* 230.1, p. 2. DOI: 10.3847/1538-4365/aa653c.
- Bozzetto, Luke M. et al. (2023). "New ASKAP radio supernova remnants and candidates in the Large Magellanic Cloud." In: *Monthly Notices of the Royal Astronomical Society*, 518.2. DOI: 10.1093/mnras/stac2922.
- CSIRO (2023). *The ASKAP Radio Telescope*. <https://www.atnf.csiro.au/projects/askap/index.html>. Accessed: 2023-13-04.
- Caltech (2023). *Spitzer Space Telescope*. <https://www.spitzer.caltech.edu/mission/fast-facts>. Accessed: 2023-13-04.
- Carey, S. J. et al. (2009). "MIPSGAL: A Survey of the Inner Galactic Plane at 24 and 70 μm ." In: *Publications of the Astronomical Society of the Pacific* 121.875, p. 76. DOI: 10.1086/596581.
- Collischon, Caroline, Manami Sasaki, Klaus Mecke, Sean D. Points, and Michael A. Klatt (2021). "Tracking down the origin of superbubbles and supergiant shells in the Magellanic Clouds with Minkowski tensor analysis." In: *Astronomy & Astrophysics* 653, A16. DOI: 10.1051/0004-6361/202040153.
- Davies, R. D., K. H. Elliott, and J. Meaburn (1976). "The nebular complexes of the Large and Small Magellanic Clouds." In: *Memoirs of the Royal Astronomical Society* 81, p. 89.
- Dopita, Michael A. and Ralph S. Sutherland (2003). *Astrophysics of the diffuse universe*. DOI: 10.1007/978-3-662-05866-4.
- ESA: XMM-Newton SOC (2022). *XMM-Newton Users Handbook*. https://xmm-tools.cosmos.esa.int/external/xmm_user_support/documentation/uhb/. Issue 2.20. Accessed: 2023-13-04.

- Fazio, G. G. et al. (2004). "The Infrared Array Camera (IRAC) for the Spitzer Space Telescope." In: *The Astrophysical Journal Supplement Series* 154.1, p. 10. DOI: 10.1086/422843. URL: <https://dx.doi.org/10.1086/422843>.
- Fesen, R. A., W. P. Blair, and R. P. Kirshner (1985). "Optical emission-line properties of evolved galactic supernova remnants." In: *Astrophysical Journal* 292, pp. 29–48. DOI: 10.1086/163130.
- Filipovic, Miroslav (2021). "The Future is Certain: Supernova Remnants in the Multi-Messenger Era - CTA-Radio Link." In: *43rd COSPAR Scientific Assembly. Held 28 January - 4 February*. Vol. 43, 1287, p. 1287.
- G, Prithivraj and Alka Kumari (2023). "Identification and Classification of Exoplanets Using Machine Learning Techniques." In: *arXiv e-prints*. DOI: 10.48550/arXiv.2305.09596. eprint: 2305.09596.
- Geyer-Ramsteck, Jonas (2020). "The predictive power of convolutional neural networks in Astrophysics as a discovery tool." In: *arXiv e-prints*, arXiv:1406.2661. DOI: 10.48550/arXiv.1406.2661. eprint: 1406.2661.
- Goodfellow, Ian J. et al. (2014). "Generative Adversarial Networks." In: *arXiv e-prints*, arXiv:1406.2661. DOI: 10.48550/arXiv.1406.2661. eprint: 1406.2661.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- Graczyk, Dariusz et al. (2014). "The Araucaria Project. The Distance to the Small Magellanic Cloud from Late-type Eclipsing Binaries." In: *The Astrophysical Journal* 780.1, p. 59. DOI: 10.1088/0004-637X/780/1/59.
- Gui, Jie, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye (2020). "A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications." In: *arXiv e-prints*. DOI: 10.48550/arXiv.2001.06937. eprint: 2001.06937.
- He, Haibo (2013). *Imbalanced Learning. Foundations, Algorithms, and Applications*. John Wiley & Sons, Ltd. ISBN: 9781118646106. DOI: <https://doi.org/10.1002/9781118646106.ch1>.
- Henize, Karl G. (1956). "Catalogues of H α -emission Stars and Nebulae in the Magellanic Clouds." In: *Astrophysical Journal Supplement* 2, p. 315. DOI: 10.1086/190025.
- Hodges, J. L. (1958). "The significance probability of the smirnov two-sample test." In: *Arkiv for Matematik* 3.5, pp. 469–486. DOI: 10.1007/BF02589501.
- Ioffe, Sergey and Christian Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37. ICML'15*. Lille, France: JMLR.org, 448–456.

- Karttunen, Hannu, Pekka Kröger, Heikki Oja, Markku Poutanen, and Karl Johan Donner (2003). *Fundamental Astronomy; 4th ed.* Berlin: Springer. DOI: 10.1007/978-3-540-34144-4.
- Kavanagh, P. J. et al. (2016). "Two evolved supernova remnants with newly identified Fe-rich cores in the Large Magellanic Cloud." In: *Astronomy & Astrophysics* 586, A4. DOI: 10.1051/0004-6361/201527414.
- Kavanagh, Patrick J. (2020). "Thermal and non-thermal X-ray emission from stellar clusters and superbubbles." In: *Astrophysics and Space Science* 365.1, 6, p. 6. DOI: 10.1007/s10509-019-3719-5.
- Kingma, Diederik P. and Jimmy Ba (2014). "Adam: A Method for Stochastic Optimization." In: *arXiv e-prints*, arXiv:1412.6980. DOI: 10.48550/arXiv.1412.6980.
- Kochenderfer, Mykel J. and Tim A. Wheeler (2019). *Algorithms for Optimization*. MIT Press.
- Lintott, Chris J. et al. (2008). "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey." In: *Monthly Notices of the Royal Astronomical Society* 389.3, pp. 1179–1189. DOI: 10.1111/j.1365-2966.2008.13689.x.
- Lintott, Chris et al. (2011). "Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies." In: *Monthly Notices of the Royal Astronomical Society* 410.1, pp. 166–178. DOI: 10.1111/j.1365-2966.2010.17432.x.
- Liu, W., X. C. Yu, and B. Y. Wang (2020). "Deep Learning for Supernova Remnants Detection and Localization." In: *Astronomical Data Analysis Software and Systems XXVII*. Ed. by Pascal Ballester, Jorge Ibsen, Mauricio Solar, and Keith Shortridge. Vol. 522. Astronomical Society of the Pacific Conference Series, p. 451.
- Long, Knox S. (2017). "Galactic and Extragalactic Samples of Supernova Remnants: How They Are Identified and What They Tell Us." In: *Handbook of Supernovae*. Ed. by Athem W. Alsabti and Paul Murdin, p. 2005. DOI: 10.1007/978-3-319-21846-5_90.
- Longair, Malcom S. (2011). *High Energy Astrophysics*. Cambridge University Press.
- Lopez, L. A. et al. (2009). "Typing Supernova Remnants Using X-ray Line Emission Morphologies." In: *The Astrophysical Journal* 706.1, p. L106. DOI: 10.1088/0004-637X/706/1/L106.
- Lopez, Laura A., Enrico Ramirez-Ruiz, Daniela Huppenkothen, Carles Badenes, and David A. Pooley (2011). "Using The X-ray Morphology Of Young Supernova Remnants To Constrain Explosion Type, Ejecta Distribution, And Chemical Mixing." In: *The Astrophysical Journal* 732.2, p. 114. DOI: 10.1088/0004-637X/732/2/114.
- Maggi, P. et al. (2016). "The population of X-ray supernova remnants in the Large Magellanic Cloud." In: *Astronomy & Astrophysics* 585, A162. DOI: 10.1051/0004-6361/201526932.

- Maggi, Pierre et al. (2019). "The supernova remnant population of the Small Magellanic Cloud." In: *Astronomy & Astrophysics* 631. DOI: 10.1051/0004-6361/201936583.
- Maneewongvatana, Songrit and David M. Mount (1999). "Analysis of approximate nearest neighbor searching with clustered point sets." In: *arXiv e-prints*. DOI: 10.48550/arXiv.cs/9901013.
- Mathewson, D. S. and J. N. Clarke (1973). "Supernova remnants in the Large Magellanic Cloud." In: *Astrophysical Journal* 180, pp. 725–738. DOI: 10.1086/152002.
- Matsuura, Mikako et al. (2022). "Spitzer and Herschel studies of dust in supernova remnants in the Small Magellanic Cloud." In: *Monthly Notices of the Royal Astronomical Society* 513.1, pp. 1154–1174. DOI: 10.1093/mnras/stac583.
- McKee, C. F. and J. P. Ostriker (1977). "A theory of the interstellar medium: three components regulated by supernova explosions in an inhomogeneous substrate." In: *Astrophysical Journal* 218, pp. 148–169. DOI: 10.1086/155667.
- Michelucci, Umberto (2018). *Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks*. Apress Berkeley. DOI: <https://doi.org/10.1007/978-1-4842-3790-8>.
- Murphy, Kevin P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Okon, Hiromichi et al. (2020). "Deep XMM-Newton Observations Reveal the Origin of Recombining Plasma in the Supernova Remnant W44." In: *The Astrophysical Journal* 890.1, p. 62. DOI: 10.3847/1538-4357/ab6987.
- Orlando, S., F. Bocchino, M. Miceli, O. Petruk, and M. L. Pumo (2012). "Role of Ejecta Clumping and Back-reaction of Accelerated Cosmic Rays in the Evolution of Type Ia Supernova Remnants." In: *The Astrophysical Journal* 749.2, p. 156. DOI: 10.1088/0004-637X/749/2/156.
- Orlando, S., M. Miceli, M. L. Pumo, and F. Bocchino (2016). "Modeling SNR Cassiopeia A from the Supernova Explosion to its Current Age: The Role of Post-explosion Anisotropies of Ejecta." In: *The Astrophysical Journal* 822.1, p. 22. DOI: 10.3847/0004-637X/822/1/22.
- Orlando, S. et al. (2019). "3D MHD modeling of the expanding remnant of SN 1987A. Role of magnetic field and non-thermal radio emission." In: *Astronomy & Astrophysics* 622, A73. DOI: 10.1051/0004-6361/201834487.
- Pellegrini, E. W. et al. (2012). "The Optical Depth of H II Regions in the Magellanic Clouds." In: *The Astrophysical Journal* 755.1, p. 40. DOI: 10.1088/0004-637X/755/1/40. eprint: 1202.3334.
- Pennock, Clara M et al. (2021). "The ASKAP-EMU Early Science Project: 888 MHz radio continuum survey of the Large Magellanic

- Cloud." In: *Monthly Notices of the Royal Astronomical Society* 506.3, pp. 3540–3559. DOI: 10.1093/mnras/stab1858.
- Predehl, P. et al. (2021). "The eROSITA X-ray telescope on SRG." In: *Astronomy & Astrophysics* 647, A1. DOI: 10.1051/0004-6361/202039313.
- Rho, J. et al. (2009). "Dust Formation Observed in Young Supernova Remnants with Spitzer." In: *Cosmic Dust - Near and Far*. Ed. by T. Henning, E. Grün, and J. Steinacker. Vol. 414. Astronomical Society of the Pacific Conference Series, p. 22. DOI: 10.48550/arXiv.0901.1699.
- Rieke, G. H. et al. (2004). "The Multiband Imaging Photometer for Spitzer (MIPS)." In: *The Astrophysical Journal Supplement Series* 154.1, p. 25. DOI: 10.1086/422717. URL: <https://dx.doi.org/10.1086/422717>.
- Roberts, M. S. E. et al. (2002). "The Pulsar Wind Nebula of G11.2-0.3." In: *Neutron Stars in Supernova Remnants*. Ed. by Patrick O. Slane and Bryan M. Gaensler. Vol. 271. Astronomical Society of the Pacific Conference Series, p. 229. DOI: 10.48550/arXiv.astro-ph/0112053.
- Ruder, Sebastian (2016). "An overview of gradient descent optimization algorithms." In: *arXiv e-prints*. DOI: 10.48550/arXiv.1609.04747. eprint: 1609.04747.
- Russakovsky, Olga et al. (2015). "ImageNet Large Scale Visual Recognition Challenge." In: *International Journal of Computer Vision (IJCV)* 115.3, pp. 211–252. DOI: 10.1007/s11263-015-0816-y.
- Shelton, R. L., K. D. Kuntz, and R. Petre (2004). "Chandra Observations and Models of the Mixed-Morphology Supernova Remnant W44: Global Trends." In: *The Astrophysical Journal* 611.2, pp. 906–918. DOI: 10.1086/422352.
- Shorten, Connor and Taghi M. Khoshgoftaar (2019). "A survey on Image Data Augmentation for Deep Learning." In: *Journal of Big Data*. DOI: <https://doi.org/10.1186/s40537-019-0197-0>.
- Simonyan, Karen and Andrew Zisserman (Sept. 2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition." In: *arXiv e-prints*. DOI: 10.48550/arXiv.1409.1556. eprint: 1409.1556.
- Simpson, R. J. et al. (2012). "The Milky Way Project First Data Release: a bubblier Galactic disc." In: *Monthly Notices of the Royal Astronomical Society* 424.4, pp. 2442–2460. DOI: 10.1111/j.1365-2966.2012.20770.x.
- The UM/CTIO Magellanic Cloud Emission Line Survey (MCELS)* (2000). Vol. 221. Stars, Gas and Dust in Galaxies: Exploring the Links. Astronomical Society of the Pacific.
- Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov (2014). "Dropout: A Simple Way to

- Prevent Neural Networks from Overfitting." In: *Journal of Machine Learning Research* 15,56, pp. 1929–1958.
- Sunyaev, R. et al. (2021). "SRG X-ray orbital observatory. Its telescopes and first scientific results." In: *Astronomy & Astrophysics* 656, A132. DOI: 10.1051/0004-6361/202141179.
- Tielens, A. G. G. M. (2005). *The Physics and Chemistry of the Interstellar Medium*. Cambridge University Press.
- Ustamujic, S. et al. (2021a). "Modeling the mixed-morphology supernova remnant IC 443. Origins of its complex morphology and X-ray emission." In: *Astronomy & Astrophysics* 649, A14. DOI: 10.1051/0004-6361/202039940.
- (2021b). "Modeling the mixed-morphology supernova remnant IC 443. Origins of its complex morphology and X-ray emission." In: *Astronomy & Astrophysics* 649, A14. DOI: 10.1051/0004-6361/202039940.
- Ustamujic, S. et al. (2021c). "Modeling the remnants of core-collapse supernovae from luminous blue variable stars." In: *Astronomy & Astrophysics* 654, A167. DOI: 10.1051/0004-6361/202141569.
- Vink, Jacco (2012). "Supernova remnants: the X-ray perspective." In: *The Astronomy and Astrophysics Review* 20, p. 49. DOI: 10.1007/s00159-011-0049-1.
- Vink, Jacco (2020). *Physics and Evolution of Supernova Remnants*. Springer Cham. DOI: 10.1007/978-3-030-55231-2.
- Walmsley, Mike et al. (2022). "Galaxy Zoo DECaLS: Detailed visual morphology measurements from volunteers and deep learning for 314 000 galaxies." In: *Monthly Notices of the Royal Astronomical Society* 509.3, pp. 3966–3988. DOI: 10.1093/mnras/stab2093.
- Ward-Thompson, Derek and Anthony P. Whitworth (2011). *An Introduction to Star Formation*. Cambridge University Press. DOI: 10.1017/CB09780511974021.
- Weaver, R., R. McCray, J. Castor, P. Shapiro, and R. Moore (1977). "Interstellar bubbles. II. Structure and evolution." In: *Astrophysical Journal* 218, pp. 377–395. DOI: 10.1086/155692.
- Werner, M. W. et al. (2004). "The Spitzer Space Telescope Mission." In: *The Astrophysical Journal Supplement Series* 154.1, p. 1. DOI: 10.1086/422992. URL: <https://dx.doi.org/10.1086/422992>.
- White, Richard L. and Knox S. Long (1991). "Supernova Remnant Evolution in an Interstellar Medium with Evaporating Clouds." In: *Astrophysical Journal* 373, p. 543. DOI: 10.1086/170073.
- Williams, Jonathan P. (2021). *Introduction to the Interstellar Medium*. Cambridge University Press. DOI: 10.1017/9781108691178.
- Winkler, P. F., G. W. Clark, T. H. Markert, R. Petre, and C. R. Canizares (1981). "X-ray line emission from the Puppis A supernova remnant - Oxygen lines." In: *Astrophysical Journal* 245, pp. 574–580. DOI: 10.1086/158832.

- Yang, Suorong et al. (2022). "Image Data Augmentation for Deep Learning: A Survey." In: *arXiv e-prints*. DOI: 10.48550/arXiv.2204.08610.
- Ye, Taisheng, A. J. Turtle, and Jr Kennicutt R. C. (1991). "A new study of N66 in the Small Magellanic Cloud." In: *Monthly Notices of the Royal Astronomical Society* 249.4, pp. 722–728. ISSN: 0035-8711. DOI: 10.1093/mnras/249.4.722.
- Yew, Miranda et al. (2021). "New optically identified supernova remnants in the Large Magellanic Cloud." In: *Monthly Notices of the Royal Astronomical Society* 500.2. DOI: 10.1093/mnras/staa3382.
- Zhao, Guiyu et al. (2023). "Deep Learning Applications Based on WISE Infrared Data: Classification of Stars, Galaxies and Quasars." In: *arXiv e-prints*, arXiv:2305.10217. DOI: 10.48550/arXiv.2305.10217.
- Zhou and Chellappa (1988). "Computation of optical flow using a neural network." In: *IEEE 1988 International Conference on Neural Networks*. Vol. 2, pp. 71–78. DOI: 10.1109/ICNN.1988.23914.
- van den Bergh, S. (1999). "The Magellanic Clouds, Past, Present and Future - A Summary of IAU Symposium No. 190." In: *New Views of the Magellanic Clouds*. Vol. 190, p. 569. DOI: 10.48550/arXiv.astro-ph/9810045.