Real-time processing of I-LOFAR data using signal and image-based Artificial Intelligence/Machine Learning methods

Jeremiah Scully

Supervisor: Dr. Mark Daly

Dr. Ronan Flynn

Prof. Peter Gallagher

Faculty of Engineering and Informatics

Technological University of the Shannon

This dissertation is submitted for the degree of

Doctor of Philosophy

August 2023

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work which I now submit for assessment on the programme of study leading to the award of PhD and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Jeremiah Scully

August 2023

Acknowledgements

There is a quote that states 'The starting point of all achievement is both desire and having a cohort that helps'. Many people have contributed to my research over the years; enabling my ability to learn, broaden my horizons, and challenge conventional wisdom. That cohort is both my supervisors; Dr. Mark Daly and Dr. Ronan Flynn at TUS. Their unwavering commitment, support, and guidance have made an indelible impact on my thesis journey. I have been privileged to have such enthusiastic support and guidance - for that; I thank you!

I would also like to personally thank my other two supervisors; Dr. Eoin Carley and Prof. Peter Gallagher at DIAS. Their expertise in Solar Physics has helped me to push the technical boundaries which have been pivotal in my personal development for this research project.

Furthermore, I want to acknowledge that the unique (and unfortunate) environment of the Pandemic brought many unintended consequences and hardship to so many people. My Supervisors' advocacy, partnership, and empathetic demeanor helped lighten the load; both in my research and in my ongoing well-being and mindfulness. Words can never express (enough); the immeasurable impact this made, during the difficult times. Thank you!

Abstract

Solar flares discharge up to 10^{25}J of magnetic energy into the solar atmosphere and are often linked with high-intensity radio emissions known as Solar Radio Bursts (SRBs). SRBs are commonly found in dynamic spectra and are classified into five major spectral classes, ranging from Type I to Type V, based on their form and frequency, and time extent. The automatic detection and classification of such radio bursts is a challenge in solar radio physics due to their heterogeneous form. Near real-time detection and classification of SRBs has become a necessity in recent years due to large data rates generated by advanced radio telescopes such as the LOw-FRequency ARray (LOFAR).

This thesis proposes a strategy for developing a very fast image and signal classification system that uses artificial intelligence algorithms to process gigabyte/sec data streams in real-time using the Irish-LOFAR array as its prime data source. Currently, the state-of-the-art systems in this area are falling short of the required performance to process such high-bandwidth data streams. Real-time study of SRBs is crucial for effective solar monitoring because it provides timely information about dynamic solar phenomena, such as flares and coronal mass ejections, allowing us to predict space weather impacts on Earth's technology and infrastructure. This real-time data helps spacecraft operators, scientists, and public safety officials make informed decisions, validates and refines models of solar behavior, and drives advancements in monitoring technologies, ensuring accurate and proactive responses to solar disturbances.

In this research project, artificial intelligent algorithms are implemented for simulating, classifying, detecting and segmenting SRBs. The You Only Look Once v2 (YOLOv2) Convolutional Neural Network (CNN) is introduced, which, for real-time classification of Type III SRBs, achieved an accuracy of 82.63% at a maximum frame rate of 77 fps. The Generative Adversarial Network (GAN) methodology was proposed for generating highly realistic simulations of both Type III and Type II SRBs. By combining GAN-generated Type III simulations with observations from LOFAR, the YOLOv2 model was trained to accurately detect Type III bursts with a mean Average Precision (mAP) of 77.71%. Additionally, the study was extended to Type II bursts which, using a GAN to simulate them and a Mask R-CNN model for detection and segmentation, achieved a mAP score of 78.90%.

These findings demonstrate the effectiveness of deep learning models for SRB classification, detection, and segmentation. With the development of LOFAR for Space Weather (LOFAR4SW), a system upgrade aimed at autonomously monitoring solar radioactivity, the need for automated data pipelines for SRBs has become even more pertinent in recent years. Deep learning has been identified as an indispensable element in the future development of software pipelines for autonomously detecting SRBs. This research highlights the significance of deep learning as the crucial component within such pipelines.

Table of contents

List of figures

List of tables

List of tables

Preface

Some passages in this thesis have been quoted verbatim from the papers listed below, which were written during my PhD research.

Paper I: J. Scully, R. Flynn, E. Carley, P. Gallagher and M. Daly, "Type III solar radio burst detection: A deep learning approach," 2021 32nd Irish Signals and Systems Conference (ISSC), Athlone, Ireland, 2021, pp. 1-6, doi: 10.1109/ISSC52156.2021.9467876.

Paper II: Scully, J., Flynn, R., Carley, E. et al. "Simulating Solar Radio Bursts Using Generative Adversarial Networks", Solar Physics 298, 6 (2023). doi: 10.1007/s11207-022-02099-x.

Paper III: J. Scully, R. Flynn, P. Gallagher and M. Daly, "Mitigation of Radio Frequency Interference in Solar Radio observations using Generative Adversarial Networks", International Machine Learning and Computer Vision in Heliophysics conference, Sofia, Bulgaria, 2023.

Paper IV: J. Scully, R. Flynn, E. Carley,, P. Gallagher and M. Daly, "Improved Type III solar radio burst detection using congruent deep learning models", Astronomy & Astrophysics, 2023.

Paper V: J. Scully, R. Flynn, P. Gallagher and M. Daly, "Type II solar radio burst segmentation and detection using multi-model deep learning networks", 2023 34th Irish Signals and Systems Conference (ISSC), Dublin, Ireland.

Chapter 1

Introduction

1.1 Problem statement

This thesis describes a research project to develop a very fast image and signal classification and detection system that uses artificially intelligent algorithms to process gigabyte per second data streams in real-time using the Irish- LOw-FRequency ARray (I-LOFAR) array as its primary data source. State-of-the-art systems in this area currently fall short of the required performance to process such high-bandwidth data streams. The project builds upon the concepts and methods from an earlier project (Carley et al., 2020c) in which the primary goal was to illustrate the potential of Machine Learning (ML) algorithms for classifying Solar Radio Bursts (SRBs) and introduced the concept of using Deep Learning (DL) for detecting SRBs in dynamic spectra. Thus, leading to the overarching research question:

Research Question: How effectively can deep learning methods, when applied to the data collected by the radio telescope I-LOFAR, contribute to the classification, localisation, simulation, detection, segmentation, and mitigation of various SRB phenomena, including Type II and Type III bursts, while addressing challenges related to comparability with observed data and the effects of radio frequency interference in solar observations?

This research develops the concept of DL further for SRB detection and classification and explores new directions for data sourcing. One of the findings of the earlier research was that classification techniques are capable of classifying Type III and Type II SRBs with classification accuracy rates of 90% on observed I-LOFAR data. However, it failed to provide a precise analysis of the SRB data in terms of frequency and time duration of a single burst within an observation. As a result, the initial focus of the research presented here was to ascertain if Artificial Intelligence (AI), particularly DL, can locally detect and classify multiple SRBs in terms of time and frequency within a given observation time frame. From earlier research, it was noted that Convolutional Neural Networks (CNNs) worked particularly well at classifying highly complex shapes and objects such as SRBs (Carley et al., 2020c). In order to improve this theory and not only gain classification accuracy but also detect multiple SRBs within a given observation frame, the methods of object detection were explored. This led to a framework known as YOLO in which an object detection CNN locally detects multiple objects within a given image or video with the capability of doing so in real-time. Consequently, the first sub research question of this thesis was arrived at:

Sub Research Question 1: How effectively can deep learning methods be employed to classify and locally detect Type III solar radio bursts?

The early stages of this research showed that AI could detect and classify SRBs locally in terms of time but not frequency. However, it was discovered that current data sourcing and simulation methods weren't sufficiently efficient to train large DL algorithms (Scully et al. 2021). This led to the second focus, exploring ways to simulate SRB data comparable to real observed data. After considering both mathematical methods and AI methods it was apparent that the DL algorithm Generative Adversarial Network (GAN), with its proven capability at generating realistic

data was most likely to succeed at simulating SRBs. This led to the following sub research question:

Sub Research Question 2: Is it possible for deep learning methods to generate simulations of Type II and Type III solar radio bursts that exhibit comparability with actual observed data?

Once this question was answered affirmatively using GAN, the focus shifted to simulating the relatively scarce data Type II SRBs and train an AI to detect and classify them. Moreover, with the variety of complex shapes that a Type II can assume, the objective evolved to not only to detect and classify them but also capture the Type II's true shape. This led to the third sub research question:

Sub Research Question 3: How effective are deep learning methods in detecting and segmenting Type II solar radio bursts?

Throughout the entire research project one phenomenon was ever present, Radio Frequency Interference (RFI). RFI occurs when the radiation of radio frequency energy from an electronic device produces noise that interferes with the function of an adjacent device. With LOFAR being a low-frequency radio interferometer, the presence of RFI often hindered some detections of the AI algorithms. RFI mitigation became crucial in the research specifically using DL as a means of RFI mitigation, which, led to the final sub research question:

Sub Research Question 4: How effectively can deep learning techniques be applied to mitigate the effects of radio frequency interference in solar observations?

The aim of the research described in this thesis is to address these four research questions.

1.2 Major contributions

The major contributions of this work are summarised below.

- A proposed approach known as You Only Look Once (YOLO) for locally classifying and detecting Type III solar radio bursts in terms of time and frequency within I-LOFAR observations. The method was evaluated with promising accuracy at detecting solar radio bursts at their time and frequency range, and also tested on a very fast I-LOFAR data stream, proving that it's capable of detecting Type IIIs in real-time.

- A method known as Generative Adversarial Network (GAN) for simulating Type II and Type III radio burst data comparable to real observed data made from I-LOFAR. This method eliminates the tedious task of sifting through gigabytes of data when creating training and validation sets. Using this generating method, realistic Type II and Type III examples were created in volume efficiently.

- A re-design and re-implementation of the first approach for detecting and classifying solar radio bursts. Using a combination of YOLO and GAN the localised detection accuracy of solar radio burst data in connection with time and frequency was significantly improved.

- A method for segmenting the complex phenomena that are Type II solar radio bursts. By using the new and improved simulation method, the infrequent Type II SRB can now be generated in volume. Mask R-CNN can be used to detect and segment the complex shape of Type IIs.

- The RFIR-GAN, a modified Eyeglasses Removal GAN (ERGAN) originally designed for removing eyeglasses from an image, was used for mitigating radio frequency interference from solar observations. By increasing the removal

mask size of ERGANs, the GAN network was then trained to identify and remove embedded RFI from an image.

1.3 Document structure

In this section, the content of each chapter is described and the thought process that connects the chapters is highlighted.

Chapter 2 establishes the theoretical basis for this research project, drawing inspiration from solar physics and computer science. A summary of AI and its subsets ML, DL, and Computer Vision (CV) are presented, and discussed in detail. Current state-of-the-art non-machine learning, ML, and DL techniques for SRB and Fast Radio Burst (FRB) classification and detection, and RFI mitigation are explored. The (CNN) architecture and how it was modified in this research project for SRB classification, detection, and RFI mitigation is discussed. The GAN is introduced, setting the stage for a new simulation technique adapted for this research project. The concept of object detection and how it influences this research project is presented. LOFAR is introduced, along with its new REAL-time Transient Acquisition (REALTA) hardware (Murphy et al. 2021) and a proposed AI software pipeline called LOFAR for Space Weather (LOFAR4SW) (Carley et al. 2020a). Finally, the SRB data used in experiments for this project are presented.

Chapter 3 builds and extends on the classification techniques mentioned in Chapter 2 in an attempt to answer research question 1. This experiment uses parametric modeling to simulate Type III SRBs and create a large training and test set. These simulated datasets are then used to train the object detection algorithm YOLO.

Chapter 4 presents a new way of simulating realistic SRB data, comparable to real observed data, with the goal of answering research question 2. In this investigation, GAN was trained using real observed data to create realistic simulated Type II and Type III SRB data. This improvement in creating simulated SRB data will benefit

future research by generating realistic data in volume and in a short space of time. GANs are used extensively in the following chapters to create train, validate, and test to determine model robustness and accuracy.

In Chapter 5, the GAN and YOLO algorithms are combined to create a congruent DL model to further address research question 1. The introduction of the GAN-generated simulated Type III SRB data add more robustness and accuracy to the YOLO model, so that the mean Average Precision (mAP) could be calculated.

Chapter 6 introduces the Mask R-CNN for detecting and segmenting Type II SRBs (research question 3). Once again, GAN was used to simulate the many complex forms a Type II may assume using a heavily augmented training set. The training set was then used to train Mask R-CNN to get an accurate visualisation of the Type II's heterogeneous shape.

In Chapter 7, the GAN architecture was modified, and, using two training sets, the new Radio Frequency Interference Removal Generative Adversarial Network (RFIRGAN) was trained to attempt to mitigate the effects of RFI in solar observations.

Chapter 8 provides a synopsis of the thesis, presents some conclusions about this research project, and proposes future directions for research continuing on from, or inspired by, this project.

Chapter 2

Literature Review

This research project incorporates concepts from AI, ML, DL, solar physics and radio astronomy. The content in this chapter introduces those concepts, describes the application domains to be focused upon, establishes the theoretical basis for this research, and describes the data source used in the experiments performed as part of this research.

2.1 What is Artificial Intelligence?

The world in which we find ourselves now has many characteristics that are reminiscent of the Wonderland that was portrayed in the works of Lewis Carroll. AI is defined as "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Haenlein and Kaplan 2019). Advances in AI have made possible such technologies as image recognition, smart speakers and self-driving cars, to name a few. Although it had been recognised as a field of study since the 1950s, AI continued to languish in a state of relative scientific obscurity and limited practical relevance for well over half a century due to several reasons such as unrealistic expectations leading to an overestimation of the capabilities of AI, limited computing power and

data storage, lack of data, insufficient funding and the complexity of AI (Mijwel 2015). Today, as a result of the proliferation of Big Data and the advancements in computing power, it has made its way into the cosmic world as well as public conversation (Ekmekci and Arda 2020).

AI refers to the ability of a machine or computer system to perform tasks that would typically require human intelligence, such as learning, problem-solving, and decision-making (Diaz-Ramirez 2021). AI can be divided into two main categories: narrow and general (Khan 2021). Narrow AI refers to systems that are designed to perform a specific task or set of tasks, such as Siri™ or Alexa™. Narrow AI, which includes subsets such as ML, DL and CV, is a field that focuses on creating intelligent systems capable of learning from data, imitating human brain functions, and understanding visual information, see Figure 2.8. General AI refers to systems that are designed to be more adaptable and capable of performing a wider range of tasks. The concept of general AI refers to an AI system that can perform any intellectual task that a human can. Artificial general intelligence is a hypothetical concept, and while many researchers believe it is possible to create such a system, it remains a significant challenge for the field of AI (Fjelland 2020).

ML is a category of AI that involves the use of algorithms to automatically learn and improve from experience without being explicitly programmed (Diaz-Ramirez 2021). In other words, the algorithms do not need to be told exactly how to solve a specific task or problem. Instead, they learn to recognise patterns and relationships in the data through training, and use this learning to make predictions or decisions. ML algorithms are trained on a dataset, which is a collection of data used to teach the algorithm about a particular topic or task. For example, a ML algorithm that is trained on a dataset of images of dogs and cats would be able to recognise and classify new images of dogs and cats based on the patterns it learned from the training dataset (Zhang et al. 2020).

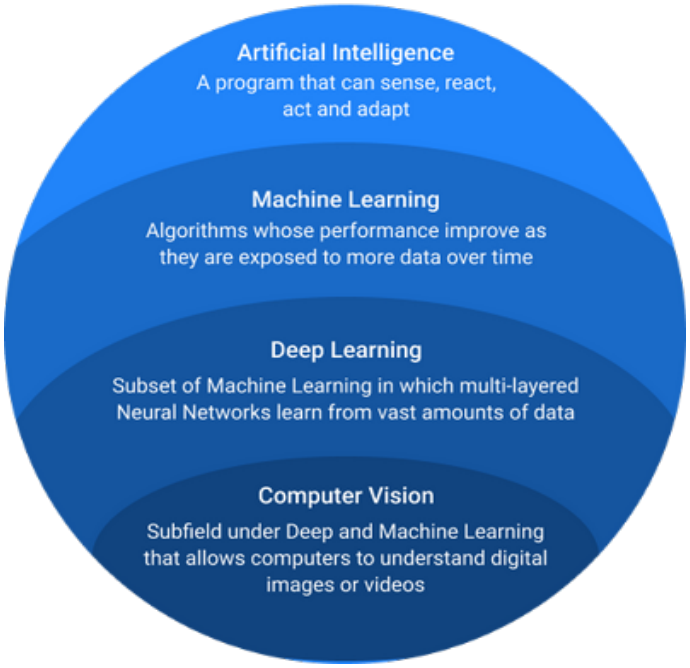


Fig. 2.1 A representation of AI and its subsets. These are employed in narrow AI systems to perform specific tasks efficiently and effectively within their limited scope. The choice of subsets depends on the requirements and objectives of the particular narrow AI application. (Meel 2022).

DL is a subset of ML that involves the use of artificial neural networks (ANNs), which are inspired by the structure and function of the human brain (Diaz-Ramirez 2021). ANNs are made up of layers of interconnected nodes that are used to process and analyse data (Brüggenwirth and Wagner 2021). Each layer of nodes is responsible for a different level of abstraction, with the lower layers handling basic features and the higher layers handling more complex features. DL algorithms are able to learn and improve by adjusting the weights and biases of the nodes in the neural network, which allows them to better recognise and classify new data (Diaz-Ramirez 2021).

One of the key advantages of AI, ML and DL is their ability to perform tasks at a faster and more efficient rate than humans. For example, ML algorithms are able to analyse and process large amounts of data quickly, efficiently and accurately, without the need for extensive computing power, which would be impossible for a human

to do manually. Similarly, DL algorithms are able to recognise and classify objects in images and videos with a high degree of accuracy, which would be difficult for a human to do consistently (Zhang et al. 2020).

Another advantage of these technologies is their ability to evolve over time. As ML algorithms are exposed to more data, they are able to learn and improve their performance, allowing them to solve problems more efficiently and make better decisions.

CV is a field of study that aims to enable computers to interpret and understand visual information, just as humans do (Wiley and Lucas 2018). This includes tasks such as image recognition, object detection and scene understanding. To achieve this, researchers use techniques from a variety of fields, including ML, DL computer graphics, image processing and robotics. They often use large datasets of labelled images to train ML models to recognise objects and then to test these models on new images to see how well they perform on the data object they were trained on, such as cars or pedestrians. There are different ways to approach CV problems, depending on the nature of the task, but the most common is to use DL, where CNNs and other deep architectures are trained and fine-tuned to perform a specific task (Zhang et al. 2020).

AI, ML and DL have a wide range of applications, including image and speech recognition, natural language processing and predictive analytics. These technologies are used in a variety of industries, including healthcare, finance and retail. For example, ML algorithms are used to analyse medical data and predict the likelihood of a patient developing a particular disease, while DL algorithms are used to analyse images and videos to detect abnormalities or objects of interest (Littman et al. 2022). In the research presented here, these technologies are applied to the analysis of solar weather.

2.2 Convolutional Neural Network (CNN)

A CNN is a deep-learning neural network that is specifically designed for visual feature recognition. It is distinguished from a standard neural network by the presence of numerous convolutional layers, see Figure 2.9. These layers do image filtering to produce image intensity gradients and feature activations. Each filter produces a different gradient and is responsive to particular shapes. For example, the first layers of the CNN may contain filters that respond to horizontal or vertical lines, curves and other simple geometric shapes. After the first filters are applied, ReLU (Rectified Linear Activation) is applied a simple mathematical function used in artificial neural networks that helps determine the output of a neuron: if the input is positive, it passes through unchanged, but if it's negative, it becomes zero, see Figure 2.9. The CNN then applies a max-pooling layer, which is a type of summing and downsampling (Wu 2017). Max-pooling works by allowing deeper layers of the network to access larger and larger portions of the image. While simple geometries may be responsive to the early layers, with max-pooling, the subsequent layers become responsive to more complex shapes made from these geometries, such as circles, triangles and complex polygons. The max-pooling and convolving can continue until the deepest levels of the network react to recognisable patterns or artefacts, for example, facial characteristics (Zhang et al. 2020).

In this research, the network's final layers respond to SRBs and their precise structure. Depending on the complexity of the network, the number of convolutional and max pooling layers vary, but the final result is a single vector representation of the image. This vector is then used to classify the image.

Similar to smaller neural networks, a CNN's weights and biases (including those of the convolutional layers) must be trained. Given the size of the networks, fitting tens of thousands, or even millions, of parameters may be necessary for the most complex networks. To avoid under- or overfitting (Ying 2019), the number of training

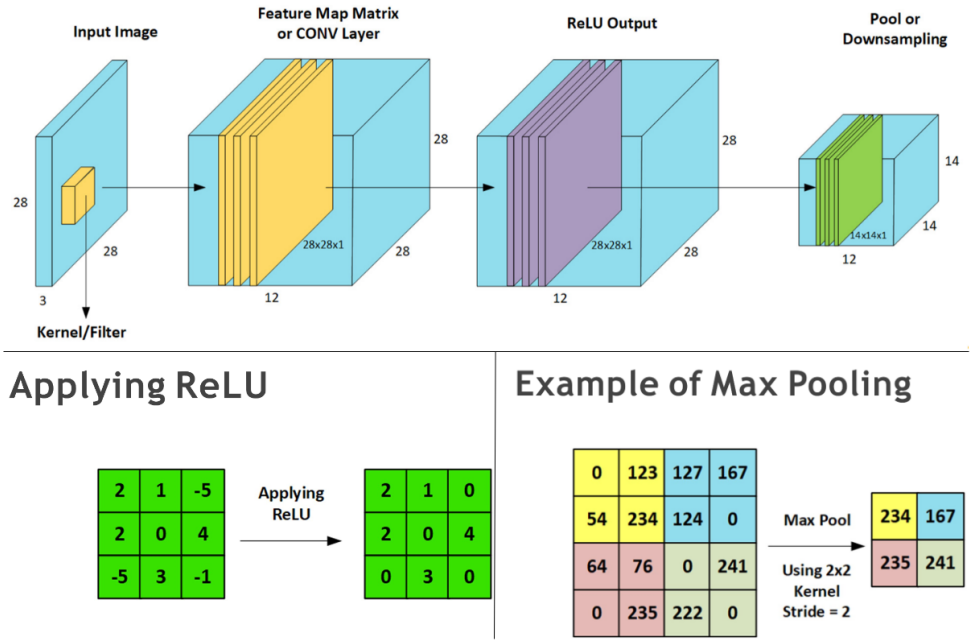


Fig. 2.2 An overview of the CNN architecture, convolutional layers, max pooling and ReLU activation function.

instances should at least be in, or close to, the same order of magnitude as the number of weights and biases in the network. A resulting problem is the lack of accessible databases with a sufficient number of images to train for phenomena like radio bursts. There are two ways to address this problem: (i) simulate a large number of training samples for the network, or (ii) use a method called transfer learning (Zhuang et al. 2021), in which an advanced and powerful CNN that has already been trained on millions of images of generic scenes (containing, among other things, everyday objects like cars and people) and retrain a smaller portion of the network on a specific set of data (radio bursts). Transfer learning is based on the notion that the general forms learned by the CNN from ordinary items may be recycled for new objects, with studies showing that this is possible even for images that are wholly morphologically unlike, such as cars and SRBs (Zhuang et al. 2021).

The CNN's framework was modified and added to as part of this research project. In the case of GAN, two CNN architectures were created, one of which was the opposite of the other, and put them against each other to simulate real Type III and

Type II SRBs (Scully et al. 2022). Then, this data was combined with real I-LOFAR data to make a large set of data that was used to train the CNN architecture YOLO and several other CNN architectures that make up the Mask R-CNN algorithm.

2.3 Object Detection

Object detection is a task in CV that entails recognising and localising objects within an image or video (Zhao et al. 2019). This can be done either manually or automatically. It is a crucial activity in many different apps, including self-driving vehicles, surveillance systems and robotics, because it enables these systems to perceive and interact with the environment around them. There are two primary methods for detecting objects: the traditional object detection methods built on handcrafted features and shallow trainable architectures, and the DL-based method (Zhao et al. 2019). The identification and localisation of objects are accomplished via traditional object detection methods using hand-crafted features, such as SIFT or SURF (Bilik and Horak 2022), and classifiers, such as SVM (Evgeniou and Pontil 2001) or AdaBoost (Gupta 2015). These methods frequently involve a high level of computing complexity and may not be as accurate as methods that are based on DL (Zhao et al. 2019). Object detection algorithms that are based on DL make use of CNNs to automatically learn features from the input data. These DL approaches have demonstrated performance at the cutting edge across a wide range of object detection tasks, including object classification and localisation.

For the purpose of object identification, some of the most common DL architectures include the Single Shot Multibox Detector (SSD) (Liu et al. 2016), Region-CNN (R-CNN) (Girshick et al. 2016a), Fast R-CNN (Girshick 2015), Faster R-CNN (Ren et al. 2017), Mask R-CNN (He et al. 2020) and YOLO (Redmon et al. 2016). These systems use CNNs to extract features from the input data and then use those features to make predictions about the location and type of objects contained

inside an image. The vast number of alternative positions and sizes for objects within an image presents a challenge to the process of object detection. Many object detection systems make use of a sliding window approach in order to circumvent this issue (Wu 2017). In this method, a CNN is applied to multiple sub-windows of the input image, and the output is then utilised to predict the presence of objects, as well as their location, within each window. Other methods, such as region proposal networks (Ren et al. 2017), utilise a two-step process to first identify potential object regions and then use a second network to classify and localise objects within those regions.

The ability to manage considerable variations in object appearance, such as changes in lighting, position and occlusion, is another problem that comes up while attempting to recognise objects in a scene. Many object detection systems make use of data augmentation techniques, such as image cropping, rotation and scaling, to intentionally modify the appearance of the training data (Zoph et al. 2020). The network is able to acquire the knowledge necessary to recognise objects in a wider variety of settings as a result of this. In general, object detection is an essential step in a wide variety of applications, techniques based on DL have achieved outstanding results on a number of different types of object detection tasks (Zhao et al. 2019).

For the purpose of this research project, YOLO was selected as the preferred detection algorithm for identifying SRBs was grounded due to its unparalleled ability to achieve both remarkable precision and real-time performance across various datasets, including COCO (Lin et al. 2015) and ImageNet (Russakovsky et al. 2015). This choice is especially pertinent to the experimental context due to the intermittent occurrence of Type III SRBs at intervals spanning approximately one to ten seconds. YOLO's standout attribute lies in its exceptional speed, aligning perfectly with the rapid temporal dynamics of Type III SRBs, while maintaining the potential to attain accuracy levels comparable to those exhibited by Faster R-CNN (Redmon and Farhadi 2017). This amalgamation of swift responsiveness and high precision

renders YOLO a compelling option for effectively capturing the transient nature of SRBs within the designated timeframe.

2.4 Generative Adversarial Networks (GANs)

Goodfellow et al. acknowledged that the launch of GANs in 2014 was a significant step forward for the field of CV (Goodfellow et al. 2016). A GAN has an architecture that is made up of two CNNs (generator and discriminator), with one of the CNNs having its connections reversed turning it into an upsampling network. The purpose of this research was to demonstrate that GAN were capable of creating synthetic images that were almost identical to the MNIST handwritten digit dataset (Deng 2012). The concept of creating datasets by utilising neural networks was proposed as a result of this and since then it has undergone significant development.

The subsequent development of Deep Convolutional GAN (DCGAN) (Radford et al. 2016) introduced architecture structure and stability to the two CNNs in the GAN, in particular the upsampling generator network. Their study presented novel approaches, such as batch normalisation, convolutional strides, elimination of fully-connected layers, ReLu/Leaky ReLu activation functions (Figure 2.9) and the Tanh activation function, for the Generator. As a result of these modifications to the architecture of the GAN, the training process was far more efficient, which was a great benefit when training on custom datasets. In the course of this investigation, SRBs are simulated by employing these various modifications.

Salimans et al. demonstrated a considerable gain in resolution that had never been seen before in a generating network (Salimans et al. 2016). The researchers were able to produce images with a resolution of 128 x 128 pixels as a result of improved training stabilisation and by the introduction of an assessment metric. This was crucial since it meant that GAN became more "training friendly" to custom

datasets. They also were capable of producing larger good quality images, despite the increased computing overheads.

Recent research has adopted the GANs concept and enhanced it further (Curto et al. 2017), resulting in the creation of High-resolution Deep Convolutional Generative Adversarial Networks (HDCGANs). GANs now have the capacity to build synthetic datasets, as evidenced by HDCGANs producing images with a resolution of 512 x 512 pixels that depict realistic synthetic human faces.

In this research project, a GAN was implemented to generate both Type III and Type II SRB simulation dynamic spectra. Real SRB data was gathered from the I-LOFAR data archive containing examples of real Type IIIs and real Type IIs respectively. The GAN was then trained on these two separate training sets and produced realistic SRB images.

2.5 The LOw Frequency ARray (LOFAR)

The LOFAR radio interferometer (Van Haarlem et al. 2013), which was constructed in the northern part of the Netherlands and throughout Europe, including Ireland's I-LOFAR station (see Figure 2.1), was the source of the data for this investigation. In order to monitor the radio universe in the relatively unknown low-frequency range of 10-240 MHz, LOFAR provides a unique selection of observational modes that can be used to monitor the radio universe. LOFAR has the capability to observe a large number of radio astronomical objects at the same time; it is capable of functioning as either a multi-station, very long baseline interferometer, or a single station functioning independently as a radio telescope.

The fundamental capabilities offered by LOFAR antenna stations are identical to those offered by conventional interferometric radio dishes, in that these stations have the capacity to position and track their antennae, in addition to having a large collecting area and excellent sensitivity. LOFAR stations do not move like regular

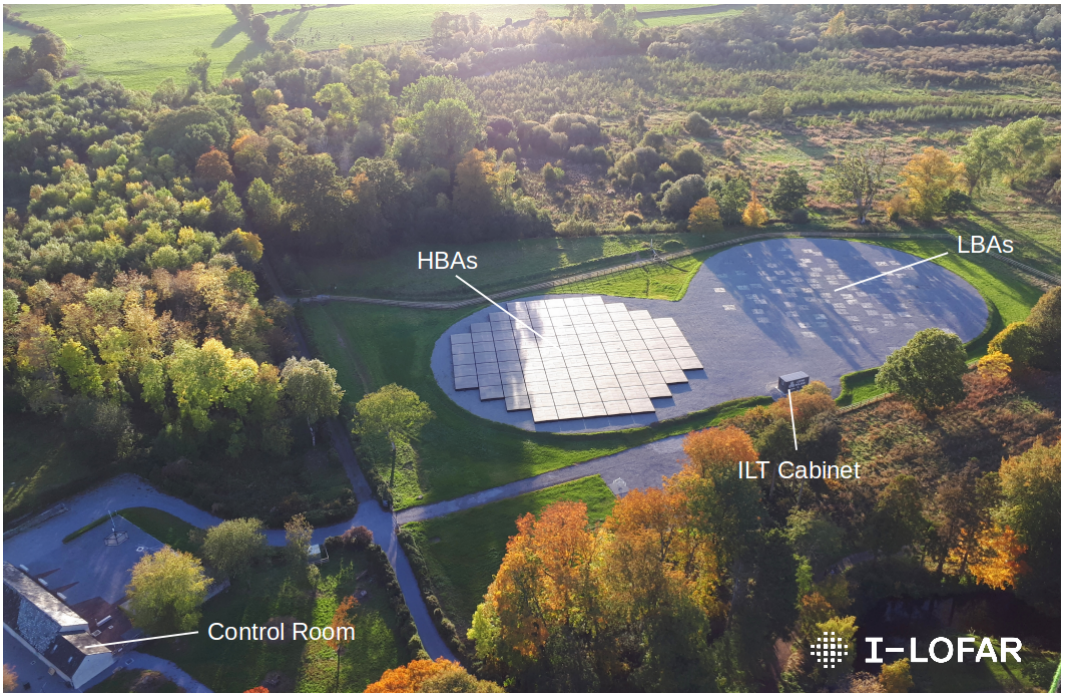


Fig. 2.3 The Irish Low-Frequency Array station IE613 (I-LOFAR) at Birr Castle, County Offaly. Coaxial cables are used to transport data from the Low Band Antennas (LBAs) (10-90 MHz) and High Band Antennas (HBAs) (100-240 MHz) to the ILT Cabinet, where they are amplified, filtered and digitised. Data is sent in international mode at a speed of $\sim$3.2 Gbps to Groningen, Netherlands. In the I-LOFAR Control Room, data is processed using REALTA in local mode (Murphy et al. 2021).

radio dishes; data from individual antennas are mixed to produce a phased array using a combination of analog and digital beam-shaping techniques, which makes the system more efficient. Beamforming at the station level makes it possible to perform rapid re-pointing of the telescope as well as several simultaneous observations from a single station.

The digitised and beam-formed data collected by the stations are then processed by a central processing unit, which correlates the data for imaging and observation analysis (an example image is shown in Figure 2.2). In most cases, the data from a single-station beamformer are collected into a dynamic spectrum with a time resolution of 5 microseconds and a frequency resolution of 195 kHz. The dynamic spectrum, which consists of 488 frequency channels, has the capability of recording data at several terabytes per hour. The I-LOFAR team has recently developed a high-

performance computer system in order to process and store unprocessed beamformed data (Murphy et al. 2021). Due to the sheer volume of data, it is extremely difficult to manually analyse and classify any relevant events, so automated techniques are required. The primary goal of this research is to identify SRBs that have been detected. Using the processed observations that were generated by I-LOFAR, training sets were contructed in this investigation so that GAN, YOLO, and Mask R-CNN could be trained and evaluated using these datasets.

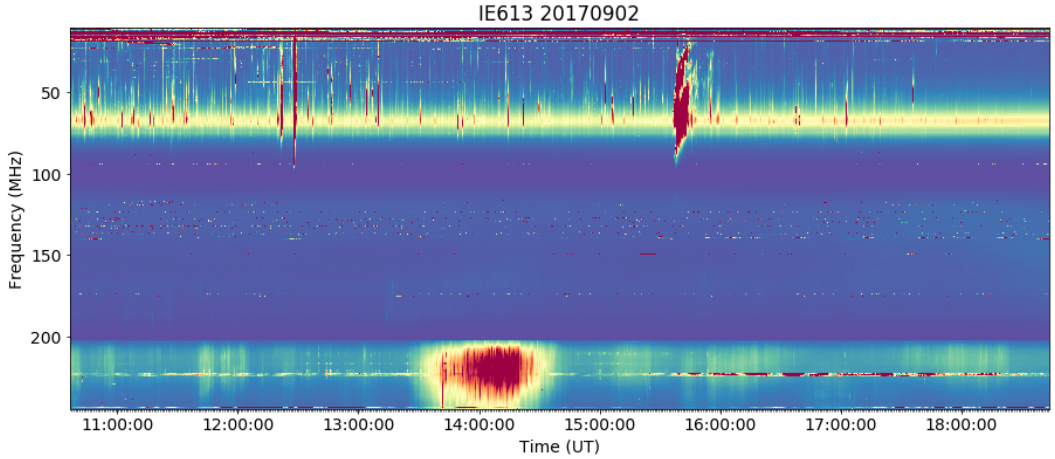


Fig. 2.4 An I-LOFAR observation from September 2, 2017. This I-LOFAR raw data taken from the I-LOFAR archive has been background-subtracted, frequency-scaled and plotted from 20 to 250 MHz.

2.6 Solar Radio Bursts

Solar flares are known to cause the release of up to 10^{25} J of magnetic energy in the solar atmosphere, see Figure 2.3, as well as the acceleration of particles to relativistic energies and the large-scale expulsion of up to 10^{12} kg of plasma (Lin 2011). From X-rays to radio waves, the accelerated electrons produce light across the full electromagnetic spectrum. Radio emission, which is frequently intense, is typically observed as intricate patterns in dynamic spectra known as SRBs (Pick M. 2009). SRBs are often divided into five categories, referred to as Type I to Type V,

based on the shape of their dynamic spectra. Automatically detecting their presence and figuring out their spectral properties is computationally difficult because they can occur thousands of times per day (particularly Type III bursts). SRBs have been

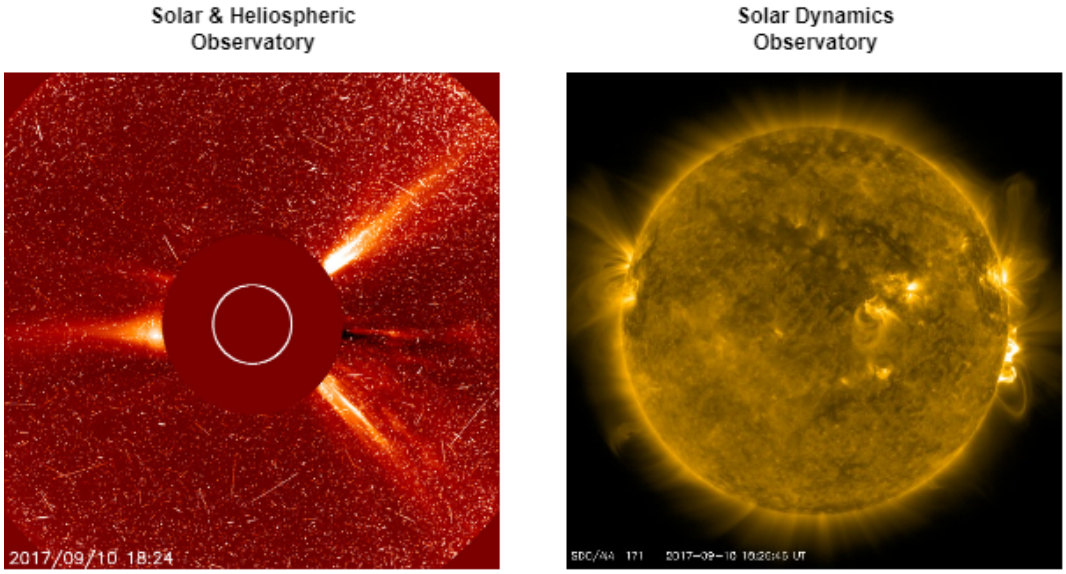


Fig. 2.5 Solar flare and Coronal Mass Ejections (CME) observed by Solar Dynamics Observatory (SDO) and Solar and Heliospheric Observatory (SOHO) (Aeronautics and Administration 2017)

observed since the 1960s and were generally classed according to their shape and extent in time and frequency when observed in dynamic spectra (Wild et al. 1963). It was soon realised that these different spectral shapes pertain to different physics occurring in the solar corona. For example, at metric wavelengths, Type III bursts are often short-lived (between 0.1-3 seconds), have a drift rate of 500 MHz s^{-1} in dynamic spectra, and are a signature of electron acceleration and propagation through the solar atmosphere (Zhan et al. 2014). Type II bursts on the other hand last for up to 10 minutes, have a slower drift rate of 10 MHz s^{-1} and are a signature of shocks traveling through the corona (Aurass 1997). Both of these radio bursts are a signature of space weather phenomena that could have adverse effects on earth, e.g., the particles associated with Type III emissions can cause discharges in spacecraft electronics (Glover et al. 2008), while the shocks responsible for Type II emissions can be a signifier of potential incoming geomagnetic storms that can cause damage

to electricity grid networks on Earth (Warmuth and Mann 2004), see Figure 2.4. Hence, there is a need for automated monitoring and analysis of these radio bursts in a space weather operations capacity. SRBs are most frequently observed as dynamic

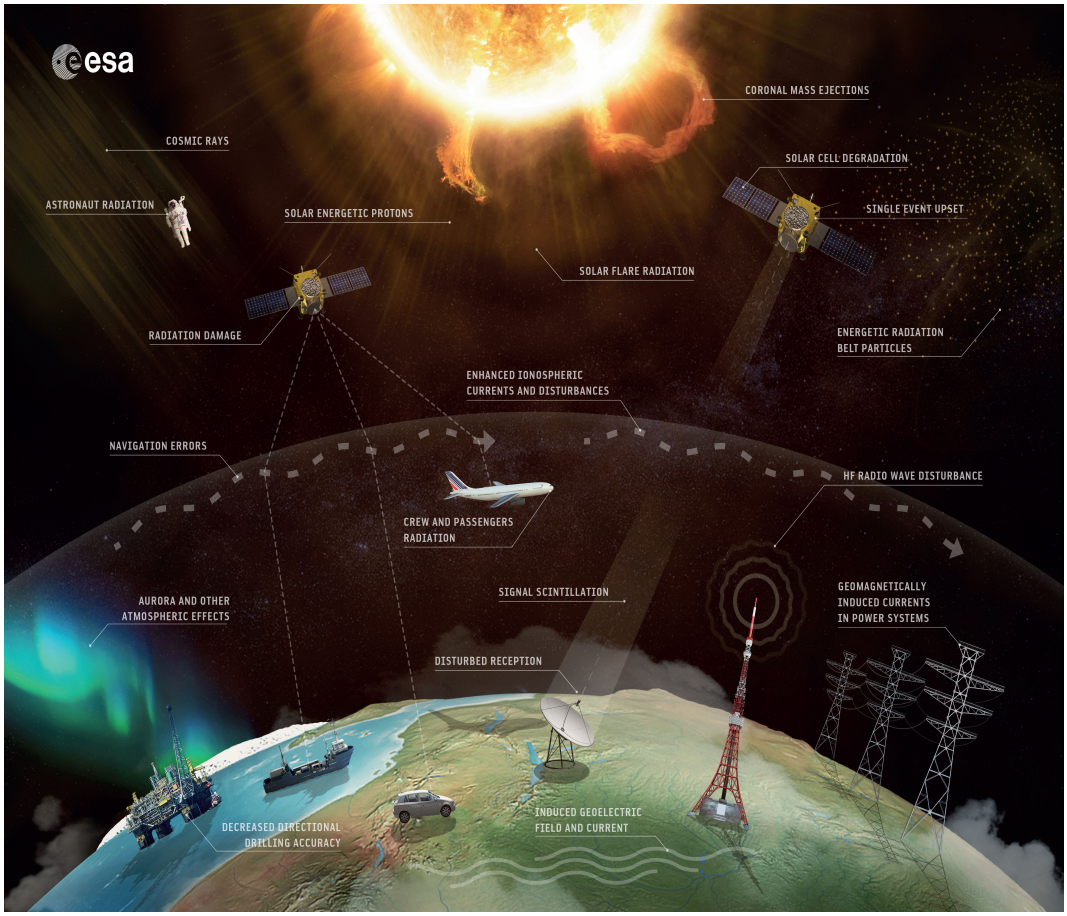


Fig. 2.6 Space weather effects on earth (ESA 2018).

frequency vs time spectra. The most prevalent radio burst is Type III, which is short-lived and physically represents a vertical stripe in dynamic spectra, as depicted in Figure 2.5. The task is made more complicated by the diversity of shapes a Type III might take, such as being smooth or patchy, faint or powerful, superimposed on other radio bursts, freestanding or in groups, or embedded in strong RFI.

The duration of Type II radio bursts can range from a few milliseconds to several minutes, and they can contain multiple lanes of emission. Once more, the range of shapes a Type II can have makes the task to identify them challenging. They can be

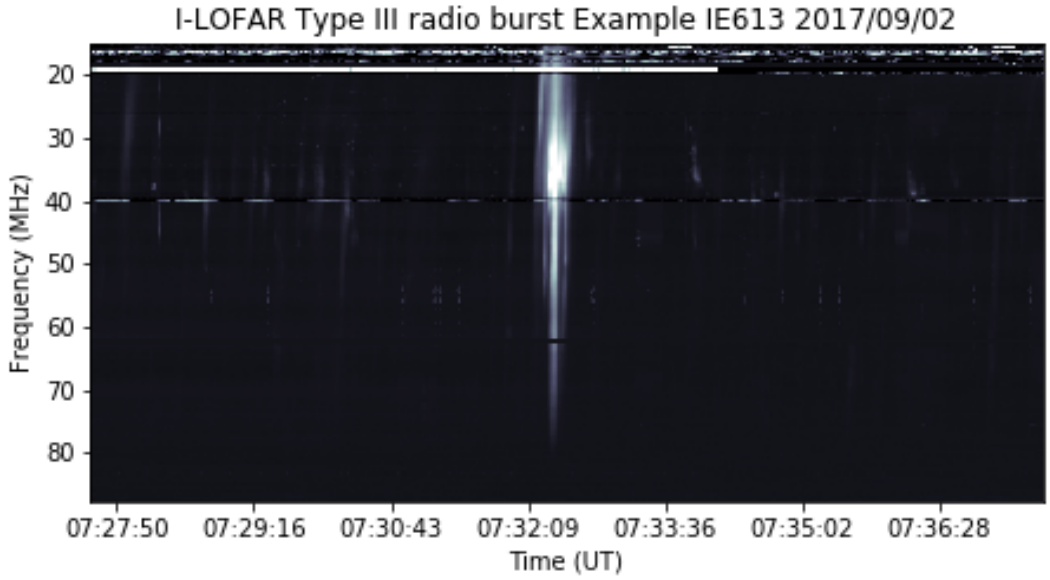


Fig. 2.7 Example of a dynamic spectrum showing a real Type III SRB between 20-90MHz at a time scale of 10 minutes. Notice the Type III's vertical strip-like shape and short duration in time, lasting only a couple of seconds.

inhomogeneous, short-lived, embedded in potent RFI, and overlaid on other bursts, just as Type IIIs. Figure 2.6 illustrates how challenging Type II classification is due to the Type II's complex shape. They often have the same shape from high to low frequency, albeit they tend to drift more quickly at high frequency, unlike Type IIIs where the shape is typically affected by the frequency of observation.

2.7 Traditional SRB Classification Methods

There have been various attempts to automatically detect SRBs in dynamic spectra. The current state-of-the-art algorithms implement a Radon transform to identify preset parametric shapes in images (Lobzin et al. 2014). Automatically identifying Type III radio bursts in data from the High-Frequency Receiver (HFR) of the S/WAVES radio instrument on board a STEREO spacecraft is described as one example. The main goal of the implementation was to apply the Radon transform to find bursts in the high-frequency range of the instrument that are roughly parallel

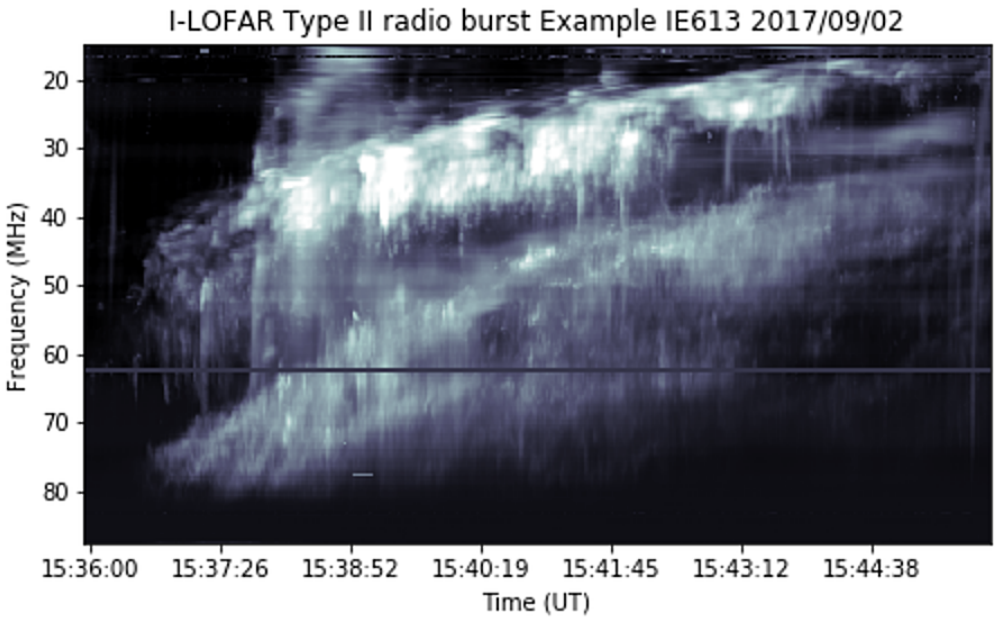


Fig. 2.8 Example of a real Type II SRB as seen in a dynamic spectrum at a frequency range of 20-90MHz at a time scale of 10 minutes. The structure of a Type II burst typically consists of a series of narrow-band emissions, called herringbones, that appear at successively higher frequencies as the shock wave propagates outward. The herringbones are caused by the acceleration of electrons in the shock wave.

to the frequency axis. After compensating for the frequency-dependent delays for the burst maxima observed at various frequencies, the Radon transform was applied in the low-frequency region. For this method, there are two detection methods via thresholding on time and frequency. Only the LFB (low-frequency band) was suitable for the first method which could be used in the onboard software to identify periods of relatively strong solar activity and make preliminary detections of the bursts. The second method allowed for the simultaneous detection of bursts in the LFB and HFB (high-frequency band) spectra. Both the HFB and LFB spectra undergo a Radon transform. Depending on the type of radio burst being categorised, these algorithms could reach an accuracy of up to 82% (Lobzin et al. 2014). However, The method's lack of robustness during model testing becomes apparent as its evaluation was limited to a mere 141 bursts, potentially compromising the generalisability and

reliability of its performance across a broader spectrum of scenarios in terms of burst structure and intensity.

Singh et al. described an algorithm that automatically identifies and classifies radio bursts in the solar corona using a novel statistical method, with the goal of providing clues to forecast space weather hazards (Singh et al. 2019). Using data from the e-CALLISTO radio spectrometer observed at Gauribidanur observatory near Bangalore in India between 2013-2014, the performance of the classifier was studied using receiver operating characteristics. The statistical method could detect radio bursts with an accuracy between 57.86% and 68.38%. The study also found that 75% of the observed Type III bursts were below 200 MHz, and the positions of the flare sites associated with these Type III bursts had an upper-frequency cut-off at 200 MHz. The paper aimed to help identify radio bursts from the solar corona automatically and accurately from large observational data sets.

Detecting radio bursts using de-noising and adaptive thresholding in dynamic spectra was a method described in (Salmane et al. 2018). The elimination of calibration signals and RFI are just two examples of how nature and noise variations are taken into account in this method of adaptive noise reduction. In order to apply automated identification in real-time applications or radio telescope observations, it is necessary to detect SRBs in a dynamic spectrum. The Constant False Alarm Rate (CFAR) was used in radar systems to find SRBs against a backdrop of interference, noise and clutter. It achieved an accuracy of up to 70% and performed well for many different kinds of radio bursts (Lu et al. 2004). However, while the adaptive noise reduction effectively emphasised the bursts in the temporal and frequency domains, it resulted in a notable degradation of the overall quality of the bursts within dynamic spectra.

(Ghurah 2018) reviewed different non-ML technologies that allowed for efficient monitoring of solar radio activity on the ground, with data recorded every second for 24 hours per day. However, the large amount of data produced makes it difficult

to identify solar bursts. The paper discusses various techniques, such as intensity thresholding, denoising, normalisation, spectrum enhancement and image down-sampling, that were successful in automatically identifying these bursts to help solar astronomers find the best technique for pre-processing before moving on to detection.

The most successful algorithms to date for detecting predefined parametric shapes in images have been the Hough and Radon transforms. These algorithms achieved 84% accuracy on Type IIIs and 80% on Type IIs, with a low false positive rates (Lobzin et al. 2009, Lobzin et al. 2010); the technique was applied to data from the Nancay Decametric Array (Lecacheux 2000) analyse statistics on Type III bursts (Zhang et al. 2018a). (Bonnin et al. 2011) used such techniques as a means to monitor Type III bursts on the HELIO feature catalogue, while (Carley et al. 2015) used it to automatically detect the herringbone fine structure in a single Type II radio burst.

The methods discussed in this section have shown significant success in accurately classifying various SRBs, yet their robustness hasn't been thoroughly assessed on extensive datasets encompassing diverse bursts in terms of time, frequency, and intensity. Lobzin et al. as an example, although illustrating 84% detection accuracy on Type III SRBs had a test set size of 35 Type III SRBs and on detected 3 Type III SRBs when tested on a real data stream. The limited size of test sets prevents a comprehensive evaluation of the true robustness and accuracy of these conventional techniques. Furthermore, thresholding methods are constrained by fixed thresholds, potentially missing bursts below such thresholds. Is it possible to develop an approach that identifies SRBs based on pixel values rather than fixed thresholds? The following section delves into the exploration of AI as a potential avenue for SRB identification.

2.8 SRB Classification and Detection using AI

In recent years, ML has found application in the classification of SRBs through the utilization of a Support Vector Machine (SVM) model. This approach leveraged the impact of SRBs on Global Navigation Satellite Systems signals (Luo et al. 2020). By preprocessing various data parameters, such as carrier-to-noise ratio, positioning errors, and geometric dilution of precision, an eigenvector was derived. Subsequently, the SVM classifier was employed to categorize the sample points, yielding an impressive average accuracy exceeding 93%. Notably, this method exhibited continuous operation in diverse weather conditions, demonstrating exceptional efficiency, accuracy, and a straightforward implementation process. Although achieving high classification accuracy, the method fails to provide precise analysis of burst location and structure in terms of time and frequency.

Li et al. discussed how studying SRBs is important for understanding solar activity and providing space weather warnings (Li et al. 2022). However, traditional image classification methods often require large amounts of training data. The paper proposed a self-supervised learning method for solar radio spectrum classification that uses self-masking in natural language processing and pre-training with existing data. This method achieves a classification accuracy of 99.5%, similar to that of other methods discussed using supervised training.

Carley et al. evaluated the capability of a variety of supervised ML methods to automatically classify Type II and Type III SRBs (Carley et al. 2020c). The research evaluated the effectiveness of SVM, random forests (RF), feed-forward neural networks (FNN) and an implementation of transfer learning on the RF CNN to accurately classify anomalies (see Appendix A for an in-depth description). The training data consisted of approximately 50,000 images and was compiled from observations of Type II and Type III radio bursts made by the Radio Solar Telescope Network (RSTN) between the years 1996 and 2018. These observations were

supplemented by computer simulations of Type II and Type III radio bursts. Each method was successful in the classification of radio bursts, with accuracies of 77% for SVM, 70% for RF and 81% for FNN. The CNNs demonstrated the highest level of performance, frequently achieving an accuracy level that was higher than 95% on the validation set and 90% when tested on I-LOFAR data. This demonstrates that ML techniques and DL techniques (CNNs in particular) were able to classify radio bursts on both RSTN and LOFAR data.

In recent years DL neural networks and classification techniques have been applied to SRB detection and classification. Multi-modal DL applied to a spectrogram at millimetric wavelengths has proven to be very effective at detecting radio bursts using deep neural networks (Ma et al. 2017). The technique used autoencoders (Lopez Pinaya et al. 2019) and regularisation to produce a burst detection accuracy of 82%, although it was not applied to metric wavelengths (the LOFAR range) where bursts can have substantially more complex geometries.

When classifying multiple types of SRBs, generative DL models such as GAN (Zhang et al. 2021) have played a key role. The GAN was converted into a classifier by modifying the generative model. Using LOFAR metric wavelengths, this modified GAN classified Type III and Type II SRBs SRBs with an accuracy range of 89-92% and 85-86%, respectively. For the work presented in this thesis the GAN is used as a means of SRB simulation.

Guo et al. discussed the identification of SRBs from large datasets of solar spectrum observations using a neural network model (Guo et al. 2022). The innovation of this study was to enhance the original radio spectrum dataset by using an unbalanced sample distribution and a proposed neural network model that was a hybrid of CNN and Long Short Term Memory (LSTM) networks. This network could extract both frequency structure and time-series features, making it more sensitive to small features in the spectrum image. Using data from the Solar Broadband

Radio Spectrometer (SBRS) in China, the proposed model improved the average classification accuracy of spectrum images to 98.73%.

A proposed automatic radio burst detection system, based on the AlexNet CNN (Gonzalez 2007) on observed solar spectrograms, was presented in (Gordo and Mateo 2022). A methodology for model training, performance evaluation, and feedback to the model generator was developed with a focus on robustness, metrics adapted to the imbalanced structure of solar bursts, tunable parameters for probability threshold optimisation, and burst coincidence cross-matching with other observatories. The neural network configuration was designed to accept data from multiple observatories. The model achieved a classification accuracy of 88% and had typical false negative and false positive scores in the single-observatory mode of 10-16% and 6-8%, respectively, which improved further in cross-match mode, providing new ways to classify SRBs within spectrograms with a high level of confidence.

(Afandi et al. 2016) describes how the sun's natural activities, such as solar flares and CMEs, produce radio signals that could be captured by the radio telescope CALLISTO. These signals were associated with the production of Type II and III SRBs. However, the process of identifying and detecting these bursts was done manually using spectrograms that were produced every 15 minutes. To assist in this process, a new algorithm was proposed that uses the frequency drift value as the principal mechanism for automatically classifying Type II and III SRBs. The algorithm includes pre-processing, identification and classification steps. The results show that the algorithm was able to provide similar results to manual detection, however, there were suggestions for improvement in future detection systems, such as considering specific characteristics of bursts and improved noise elimination.

The Solar and Heliospheric Observatory's (SOHO) Michelson Doppler Imager (MDI) full-disk solar magnetograms were used as inputs in the study by (Ma et al. 2022). Their model makes use of a CNN to automatically extract feature information from the input images. By establishing a connection between the magnetic-field

properties of the full-disk solar magnetograms and the SRBs, the CNN can be used to predict whether or not there will be a SRB on a given day. Using this strategy, they have an accuracy of 87.50% when it comes to predicting Type III and Type II SRBs. The study also showed that Type IIs are more unpredictable than Type IIIs.

Researchers have recently explored object detection algorithms, including Faster R-CNN (Hou et al. 2020), to pinpoint the precise origins of SRBs. This DL neural network demonstrates high accuracy in discerning subtle traits of SRBs, achieving an average precision (AP) of 91%. Average precision, a metric commonly used in evaluating object detection algorithms, measures the accuracy of the model in identifying objects within a range of different precision-recall trade-offs. However, while the model excels in accurate SRB characterisation, it falls short in real-time detection capabilities, achieving a maximum processing speed of 17 frames per second when tested on open sourced like COCO (Girshick et al. 2016b).

A combination of Histogram of Oriented Gradients (HOG) (Socarras Salas et al. 2012) and logistic regression (Park 2013) was used for segmenting and classifying Type II SRBs in (Jenkins et al. 2020). The model was highly effective in detecting and segmenting the shape of a Type II prominence with an accuracy range of 70-73% over a wide variety of solar activity and wavelengths. An experiment described in this thesis builds on this concept by utilising an object detection method called Mask R-CNN in order to identify and classify Type II SRBs.

A summary of the SRB classification and detection accuracies from Type II and Type III bursts is presented in Table 2.1.

2.9 Fast Radio Burst Classification

The Search for Extraterrestrial Intelligence (SETI) and FRBs, is another example of how ML was combined with radio interferometer observations for radio burst detection (Zhang et al. 2018b). Using the Allen telescope array, SETI used ML to

Table 2.1 Summary of published accuracy scores for Type II and Type III SRB classification and detection.

Research	Methods used	Type II	Type III	Technique	Accuracy
(Ma et al. 2017)	Multi modal	✓	✓	classification	82%
(Zhang et al. 2021)	DCGAN	✓	✓	classification	85-92%
(Luo et al. 2020)	SVM	✓	✓	classification	93%
(Guo et al. 2022)	CNN LSTM	✓	✓	classification	98.72%
(Li et al. 2022)	self supervised	✓	✓	classification	99.50%
(Ma et al. 2022)	CNN	✓	✓	classification	87.50%
(Hou et al. 2020)	Faster R-CNN		✓	detection	91%
(Socarras Salas et al. 2012)	HOG	✓		detection	71-73%
(Carley et al. 2020c)	Inception	✓	✓	classification	95%

search for FRBs in planetary systems. In order to remove irregular high-frequency FRB spikes within the dynamic spectrum and highlight regular noise frequencies and RFI , SETI employed a deep CNN known as ResNet (He et al. 2016). The recall score for this model was 95%.

Connor and van Leeuwen demonstrated that deep neural networks are essential for obtaining the properties of FRBs that can be used for categorisation (Connor and van Leeuwen 2018). In order to create an approximator of a dataset for the purpose of training the CNN, the research used both simulated and real single Galactic pulsars. In this particular instance, because there were so few FRBs available, the bulk of the training set was comprised of simulated bursts. They created the majority of their true positives through simulation. For their false positives, they only used results that were generated in actual observations and manually labeled. CNNs were used for this limited data and they achieved an accuracy of greater than 99% when classifying these occurrences. This FRB extraction method illustrates how robust a CNN can be when using a well designed hybrid dataset containing both real and simulated data.

When considering simulation approaches in conjunction with classification and detection methods, this FRB research has provided much insight on the development for this research project and thesis.

2.10 Radio Frequency Interference (RFI) Mitigation

According to (Pankonin and Price 1981) and (Thompson et al. 2017), one of the primary challenges in radio astronomy is RFI, which severely degrades the quality of astronomical observations and consists of any radio signals other than the Signal Of Interest (SOI) picked up by radio inferometers. The presence of RFI in the solar radio spectrum has an impact on the accuracy and efficiency of the extraction of burst parameters, which affects the quality of scientific results and the veracity of conclusions (Pankonin and Price 1981). Therefore, conducting research on RFI recognition and mitigation algorithms for solar radio data is essential (Pankonin and Price 1981).

There are many traditional ways to mitigate the affects of RFI that do not involve AI. They include eliminating RFI by utilising reference antennas (Barnbaum and Bradley 1998), using spatial information provided in interferometers or multi-feed systems to null directions (Smith 1986, Smolders and Hampson 2002), and blanking out unlikely high values at high time resolutions (Leshem et al. 2000, Niamsuwan et al. 2005). Excision of RFI can occur during post-processing and involve identifying the RFI in time, frequency and antenna space, and then ignoring the tainted data during subsequent processing steps. The term data flagging is frequently used to refer to this stage. There has been a significant improvement in both the resolution and bandwidth of observatories, which has resulted in observations totaling tens of terabytes. As a result, manual data flagging is no longer an option. Additionally, in order to automatically detect RFI with a high degree of precision, brand-new algorithms such as Sumthresholding and Varthresholding have been created (Offringa

et al. 2010). The use of these algorithms has revealed, according to the preliminary findings, that only a small percentage of the data is lost as a result of RFI (Offringa et al. 2012).

Akeret et al. proposed a novel approach to mitigate RFI signals in radio data using DL (Akeret et al. 2017a). The proposed method used a CNN called U-Net (Weng and Zhu 2021) to classify clean signals and RFI signatures in 2D time-ordered data from a radio telescope. The performance of the U-Net was assessed using the HIDE and SEEK radio data simulation and processing packages (Akeret et al. 2017b), as well as early science verification data (testing under different observation conditions such as weather, target sources, and observing modes) from a 7m single-dish telescope at the Bleien Observatory. The results showed that the proposed U-Net implementation was accurate and competitive with classical RFI mitigation algorithms such as SEEK's SumThreshold implementation.

Cheng et al. compared the recognition performance of six different deep-learning networks (FCN, Deconvnet, Segnet, U-Net, Dual-Resunet and DSC-Based Dual-Resunet) on the RFI in solar radio spectra observed by the Chinese Solar Broadband Radio Spectrometer (SBRS) (Cheng et al. 2022). The accuracy and convergence speed in the training process, as well as various performance metrics in the test, indicated that the proposed DSC-Based Dual-Resunet in which multiple U-Net architectures are combined to incorporate both high and low level features of RFI, was the most suitable neural network for RFI mitigation. The RFI recognition accuracy of the DSC-Based Dual-Resunet was close to U-Net when there was no burst in the spectrum; when there was a burst, DSC-Based Dual-Resunet was better than U-Net in terms of RFI recognition. Moreover, the model size and number of parameters are approximately 12.5% of those of U-Net, and the amount of computation in terms of parameter computing and time is 38% of that of U-Net, which greatly improves the computation efficiency and is significant for the realisation of the network on mobile hardware. The proposed model identified and removed RFI quite successfully

within solar observations and was promising for the large-scale application of RFI recognition for solar radio telescopes. However, the results also showed a loss in SRB data when this mitigation is applied.

Du et al. proposed using LSTMs, a specific type of recurrent neural network, to predict and remove RFI signals in solar radio research (Du et al. 2021). RFI signals with high intensity can obscure burst event observations. The proposed method was able to process RFI signals of various types, such as broadcast signals, mobile phone signals, and sea vessel signals, by predicting their values and subtracting them from the data containing the burst events and RFI signals to remove the interference. The method also applied a digital mapping method to convert the prediction problem into a time series classification problem. The experimental results indicate that the proposed method was effective in removing radio interference and amplified burst events in the solar spectrum.

2.11 LOFAR for Space Weather (LOFAR4SW)

As stated in Section 2.5, LOFAR is a phased array interferometer that consists of a total of 38 stations in the Netherlands and 13 international sites spread out over Europe. The center of the interferometer is located in the Netherlands. The instrument has a frequency range of 10 to 240 MHz, and it can be utilised for a wide variety of astrophysical science applications. The goal of a new initiative (LOFAR4SW) is to create an update to the LOFAR system that will make it possible for the entire array to study Jupiter, the heliosphere, the ionosphere of Earth, and the Sun for the entirety of its observation window (Carley et al. 2020b). This will enable the telescope to function as a platform for viewing space weather, which will make space weather science and operations easier to carry out. Carley et al. investigated the radio instrumentation and facilities that were in use in space weather science and operations. These radio instrumentation and facilities included instruments that

are currently involved in solar, heliospheric, and ionospheric research. Additionally, they compared LOFAR4SW to the existing radio research infrastructure in space weather and demonstrated how it may advance space weather science and operations in the radio domain in the near future. The research described in this thesis explores the possibility of AI as an important component of LOFAR4SW system.

2.12 The REAL-time Transient Acquisition backend (REALTA)

The REAL-time Transient Acquisition backend, or REALTA, is a computational backend located at the Irish LOFAR station (I-LOFAR) that makes it possible to record data in near real-time and perform post-processing (Murphy et al. 2021). Its purpose is to capture and analyse the raw beamformed data that is being collected from international LOFAR stations in real-time, see Figure 2.7. This is useful, not only to people who study space weather but also to the radio astronomy community as a whole, because it will broaden our understanding of how space weather can affect the propagation of radio waves in the inner heliosphere and disturbances in the ionosphere, and the effect that this has on observing astronomical sources. A backend such as REALTA will be required to capture the data stream and process the raw data in order to make it usable by space weather researchers and forecasters in order to record the data streams from a LOFAR4SW-enabled international station operating in local mode. This is necessary in order to record the data streams. At this time, research is being conducted to determine whether or not ML algorithms are effective at detecting SRBs using REALTA. In this research, DL is explored as a possible crucial part of such a pipeline.

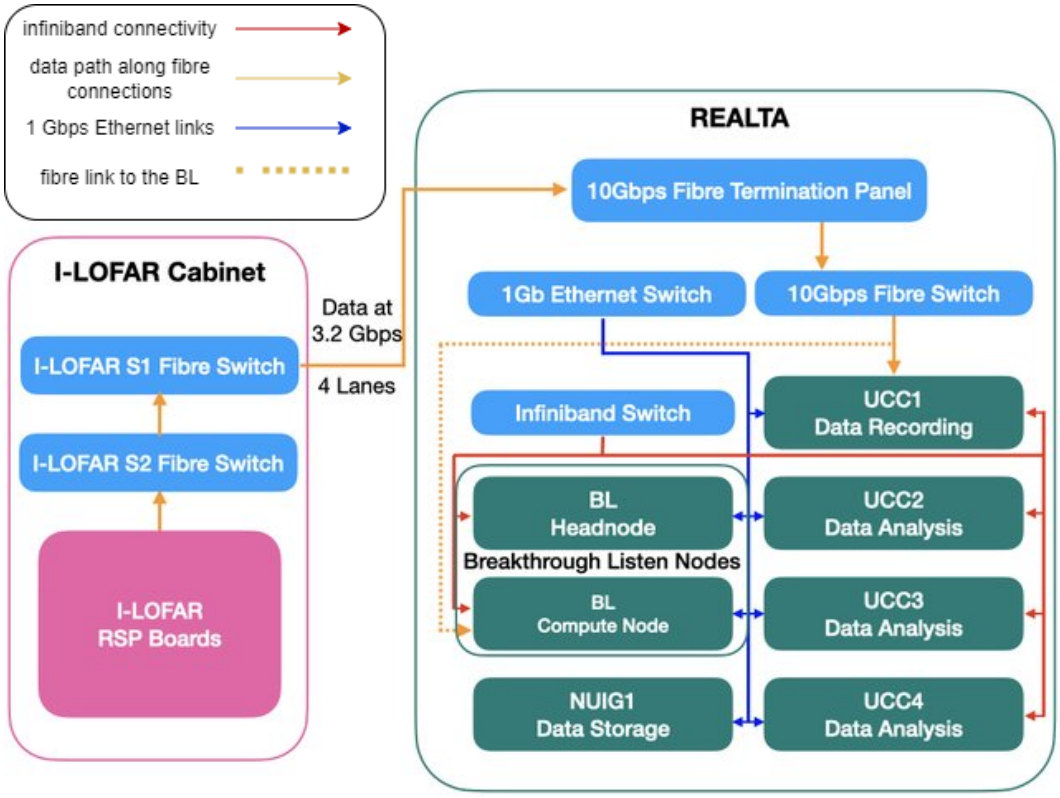


Fig. 2.9 Block diagram for REALTA and I-LOFAR. Data recorded at the Remote Station Processing (RSP) boards are sent to the S1 fibre switch in the I-LOFAR container. Here the data are split into four 'lanes' where each lane contains the data from a maximum of one quarter of the beamlets from the observation. The four lanes of data are then sent over a fibre connection to the I-LOFAR control room where it is recorded by REALTA consisting of 4 V100 GPUs (UCC) for data recording and processing (Murphy et al. 2021).

2.13 Summary

The literature review provides an in-depth exploration of various concepts and methods related to AI, CNNs, object detection, SRBs, GANs, LOFAR and other relevant topics. It covers the fundamentals of AI, including ML and DL, as well as their application in CV for image and video comprehension. Additionally, the review delves into the specifics of CNNs, their layers, and their ability to recognise characteristics and objects within images. It highlights the significance of object detection in achieving computer-based intelligent identification and localisation of objects and is identified as a possible solution to the problem of automatic detection

and classification of SRBs. Furthermore, the review discusses the nature of SRBs, their detection and association with solar events.

The literature review also presents traditional and AI-based methods for classifying SRBs, emphasising the use of ML and DL approaches on different datasets. It explores the significance of applying DL techniques for FRB classification and the utilisation of GAN to generate realistic SRB simulation dynamic spectra. The integration of object detection methods, such as Mask R-CNN, for the identification and classification of Type II SRBs is also discussed. The review recognises the LOFAR radio interferometer as a valuable data source for the research project, particularly in gathering real SRB data. The primary goal of this research project is now apparent, which is the identification of detected SRBs using processed observations from LOFAR, enabling the training and evaluation of AI models like GAN, YOLO, and Mask R-CNN.

In conclusion, the literature review serves as a solid foundation for the research project, providing a comprehensive understanding of the relevant concepts, methodologies and advancements in the field. It establishes the significance of AI and DL in analysing SRBs and explores the integration of object detection and GAN for improved detection, classification and simulation. The review acknowledges the importance of data from LOFAR and emphasises the project's goal of identifying detected SRBs using AI-driven techniques. By building upon the insights gained from the literature review, this research project aims to contribute to the real-time processing of I-LOFAR data using Signal and Image-based AI/DL methods, aiding in the identification and understanding of these phenomena in radio astronomy.

Chapter 3

Type III SRB detection using parametric modelling with YOLO

3.1 Introduction

Recall the first Sub Research Question presented in Chapter 1:

Sub Research Question 1: How effectively can deep learning methods be employed to classify and locally detect Type III solar radio bursts?

This chapter describes the approach used to answer this question using object detection and parametric modelling. Object detection is a problem in CV that refers to the process of determining what entities or objects are present in an image and where those objects are located within the image. Detecting objects in an image is a more challenging task than classifying them because classification can differentiate between different types of objects but not their positions within images or videos. In addition, classification is ineffective when applied to images that contain a significant number of objects (Shen 2022).

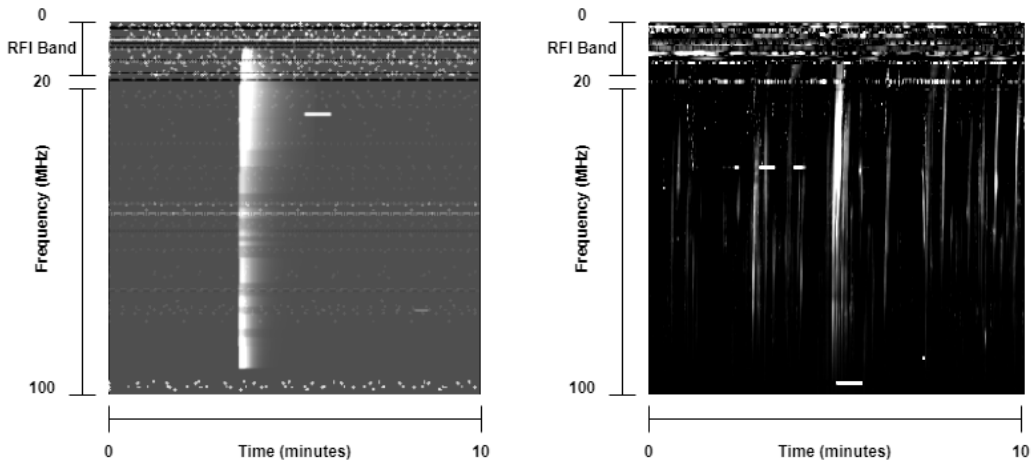
There are many different types of CNNs for object detection such as YOLO (Redmon et al. 2016), Single Shot Detectors (Liu et al. 2016), Region-CNN (R-

CNN) (Girshick et al. 2016a), Fast R-CNN (Girshick 2015), Faster R-CNN (Ren et al. 2017) and Mask R-CNN (He et al. 2020). These approaches have been shown to be very successful for object detection.

YOLO is the only algorithm to deliver high accuracy and real-time detections on datasets such as COCO (Lin et al. 2015) and ImageNet (Russakovsky et al. 2015). For the purpose of this experiment, YOLO was used because of the rate at which Type III SRBs occur, which ranges from about one to ten seconds. YOLO's primary advantage is that it is exceptionally fast and has the potential to produce accuracy that is on a par with that of Faster R-CNN (Redmon and Farhadi 2017).

3.2 Dataset

A sizable training dataset was needed for YOLOv2 in order to develop a usable model. This dataset was created using a parametric modeling method (Kalkan et al. 2018) where polynomials were employed to create the Type III shape in dynamic spectra, while skewed Gaussians were used for their temporal intensity profile at each frequency. The parametric simulated Type III SRBs were random in terms of their number, grouping, intensity, drift rate, and homogeneity. Additionally, there were random variations in the burst duration and frequency range while the bursts themselves were embedded in a background of simulated and random RFI channels as seen in Figure 3.1. Through the utilisation of the parametric modeling method, the bounding box coordinates for the simulated Type III were automatically labeled when generating the simulated Type III images. The simulated Type III was randomly placed within the image. Using these simulated Type IIIs, a training set of 80,000 simulated Type III SRBs was created to train the YOLOv2 model.



(a) Parametric Type III **(b)** Real observed Type III

Fig. 3.1 A comparison between parametrically simulated Type III radio bursts (a) and actual Type III radio bursts (b). The parametric simulated Type III bursts are used to train the DL model YOLOv2.

3.3 YOLOv2 Model Configuration

YOLO is a CNN that, depending on the configuration, can do object detection in near to real-time. The method divides the image into grid regions and predicts bounding boxes and probabilities for each zone using a single neural network applied to the entire image (see Figure 3.2). These bounding boxes get their weights from the probabilities that have been projected. YOLO's popularity can be attributed to the great accuracy and real-time functionality that it provides. This model "just looks once" at the image since it only requires one forward propagation run through the neural network in order to produce predictions. Following non-max suppression, the returned objects are those that were found to have bounding boxes (which ensures that the object detection algorithm only discovers each object once). By utilising YOLO, a single CNN is able to predict several bounding boxes in addition to the class probabilities associated with those boxes. YOLO's detection performance can be improved by using entire images for training purposes.

Once the Type III dataset was created, the YOLOv2 model had to be configured. To do this, the framework known as Darkflow (Trieu 2018), a Tensorflow implemen-

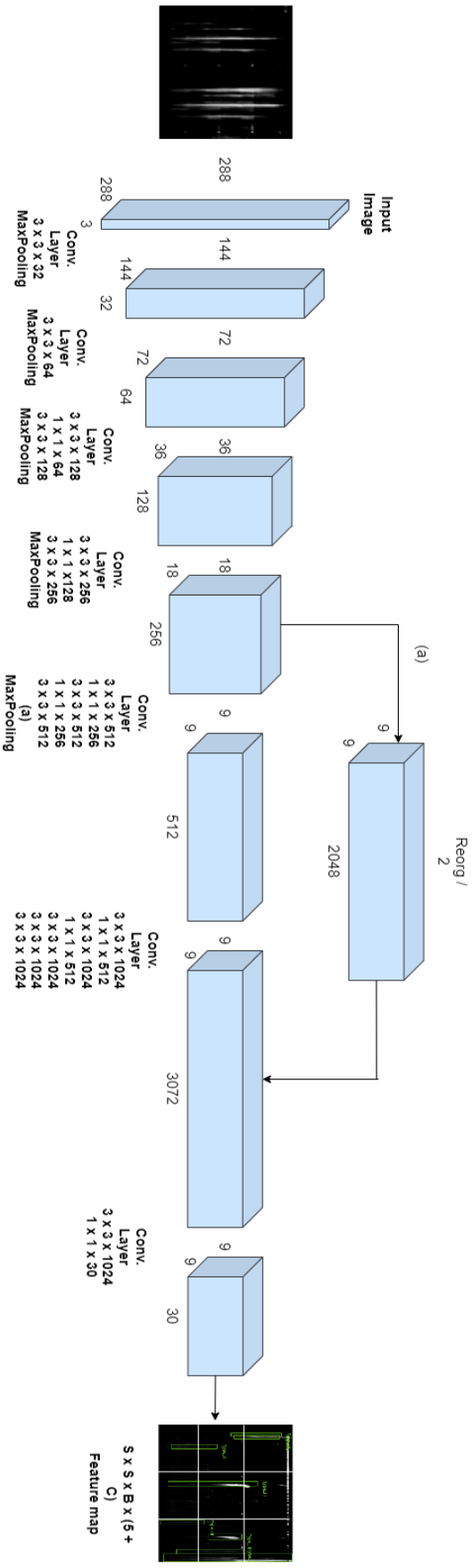


Fig. 3.2 The YOLOv2 Darknet-19 CNN architecture. Five maxpool layers and 19 convolutional layers make up the Darknet-19 architecture. For improved detection accuracy, the Reorg layer includes both high-level and mid-level features. In YOLOv2, the CNN's fully connected layers are changed. K-means classification is used for detection and classification, and the fully connected layers are removed to increase accuracy and speed.

tation of Darknet was used, to develop the YOLOv2 model. In YOLOv2 there are 19 convolutional layers and 5 maxpool layers, as shown in Figure 3.2. The number of filters in the last convolutional layer needed to be changed as only one class was to be detected, namely Type III SRBs. To do this, the following equation was applied,

$$filters = bounding * (classes + coords) \tag{3.1}$$

where $filters$ is the number of filters in the last convolutional layer, *bounding* is the number of bounding boxes per grid cell (YOLOv2 is limited to 5), *classes* is the number of classes being detected, and *coords* is the number of coordinates in each bounding box. Here $bounding = 5$, $classes = 1$ and $coords = 5$. Therefore, the total number $filters$ at the final convolutional layer is 30.

Then the anchor values (predefined bounding box shape and size used for predicting object locations and sizes within an image) were used to change the sizes of the bounding boxes. The bounding box widths were made relatively small because Type IIIs have short time durations, often lasting only a couple of seconds. The height of the bounding box was then set to 10-90MHz to match the size of the simulated Type IIIs, an example of which is shown in Figure 3.1 (a).

In order to achieve real-time frame rates with YOLOv2, the number of bounding boxes predicted on each test image was decreased from 845 to 405, see Figure 3.2. To do this, the input size was changed from 416 x 416 (YOLOv2's default input size) to 288 x 288. As the convolutional layers of YOLOv2 downsample the image by a factor of 32, then, using an input image of 416/32, an output feature map of 13 x 13 is obtained, which effectively sections the image into a 13 x 13 square grid. Multiplying the aforementioned square grid values by the number of bounding boxes per grid cell results in 845 predictions per image, where the number of bounding boxes per grid cell is 5 x 13 x 13. When a 288 x 288 input image is fed, a feature map of 9 x 9 is obtained with 5 bounding boxes per square grid. Consequently, the

number of potential predictions per image is calculated by multiplying 5 x 9 x 9, resulting in 405 possible predictions per image (Figure 3.2). At 288 x 288 it runs at a maximum of 77 fps with an accuracy comparable to Fast R-CNN. This input size complements the dataset input to the YOLOv2 model as only one class in greyscale format is being identified.

3.4 Model Training

The YOLOv2 model was trained to detect and classify Type III SRBs. The training set consisted of 80,000 parametric simulated Type III images and the validation set was a subset of the training set containing 20,000 parametric simulated Type III images. As the training and validation sets were simultaneously labelled as they were created, a high-volume training set along with a list of text files containing bounding box coordinates was built. These text files were then translated into XML to fit the Darkflows training set requirements. Once prerequisites were completed the training and validation sets were then fed into the CNN for training. Darkflow, by default, only uses 1 GPU or CPU for training so Nvidia Scalable Link Interface (SLI) support was added to the framework. The framework was also configured to allow training validation; this gives an insight into whether or not the model was overfitting or underfitting. The model was trained with a learning rate of 0.001, a momentum of 0.9 and a Leaky Rectified Linear Unit (ReLU) activation function. Learning parameters were also updated until convergence using Stochastic Gradient Descent (SGD).

This research was conducted on a machine comprising 2 x SLI inter-connected GPU Nvidia Geforce RTX 2080 Ti, using Ubuntu 20.4.2 LTS on an AMD Ryzen Threadripper 1950x with 32GB of RAM. For the training configuration, 90% of GPU capacity was used for 200 epochs at a batch size of 32. With this configuration, training took 4 days, with both training and validation losses decreasing with every

iteration. These losses rapidly decreased from iteration 0 to 11,000, then stabilised, as seen in Figure 3.3.

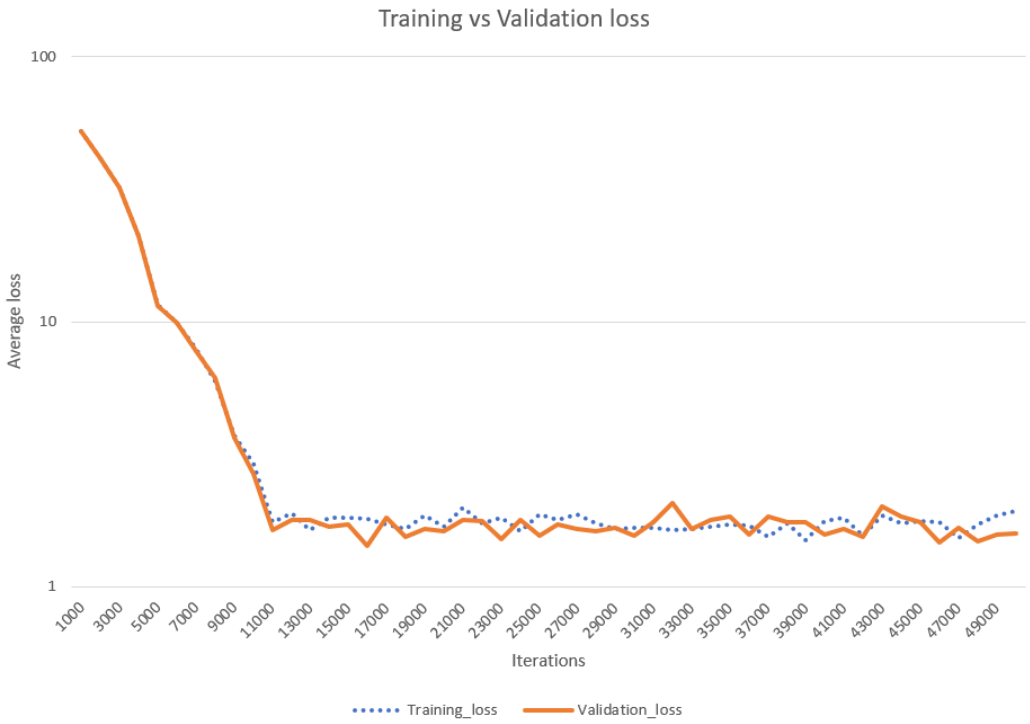


Fig. 3.3 A plot showing both training and validation losses decreasing with every iteration of training. Loss is an excellent way to see if the model is underfitting or overfitting.

3.5 Test Set

The test set for the model was a 12-hour observation made by I-LOFAR on the 10th of September 2017. The raw data was processed and converted into an image and plotted from 0-250MHz over the duration of the observation (time) as seen in Figure 3.4. This image was then converted to greyscale to match the characteristics of the training set so that colour wasn't influenced by the model's predictions (Rafegas and Vanrell 2017). Only the 10-90 MHz range in the observations was considered as this shows the Type III's most prominent attributes. The observation was then cut into 10-minute chunks where the Type IIIs narrow strip like shape becomes most prominent, creating a test set of 1,331 images, containing about 15,000 Type III solar

radio bursts. Once the test set images were created, YOLOs predicted bounding boxes were annotated on the test data as the ground truth bounding box values. This was done by using LabelImg (Tzutalin 2015), an annotating tool used to label objects within an image. Once a Type III was labelled its corresponding bounding box coordinates were stored in an XML file to compare the ground truth coordinates to the model's predicted coordinates.

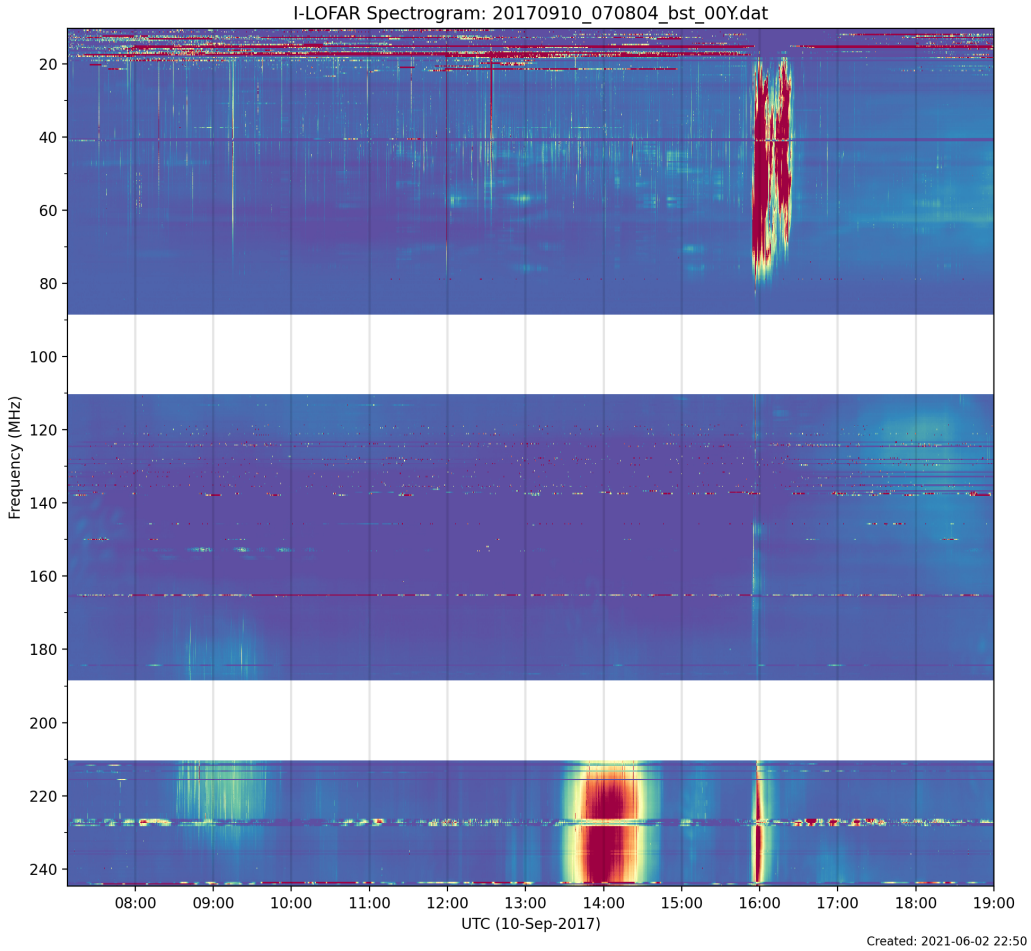


Fig. 3.4 An I-LOFAR observation from September 10, 2017. This I-LOFAR raw data has been background-subtracted, interpolated, frequency-scaled and plotted from 0-250 MHz.

3.6 Model Evaluation

The YOLOv2 model's performance was measured using the F_1-score on the aforementioned test set of 1,331 images. The F_1-score highlights the balance between precision and recall. The precision is how well the model predicted the location of an object and recall is the ratio of the number of true positives to the total number of actual objects. The following equations are used to calculate the F_1-score of the model.

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.3}$$

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.4}$$

$$Intersection over Union = \frac{Area of Overlap}{Area of Union} \tag{3.5}$$

True positive (TP) and false positive (FP) were obtained from intersection over union (IoU) from test data. IoU compares the predicted bounding box to the ground truth bounding box. A prediction was classified as TP if the IoU was greater than 0.5 and FP if it was less than 0.5. A False Negative (FN) was specified for those images where the model failed to detect a known Type III object. One important factor to take into account when evaluating the model performance was the confidence threshold. The confidence threshold measures how confident the model is at predicting a certain object, in this case, Type III SRBs. If the confidence was low, more detections were made on a test image but more false detections were made also. After experimenting with different thresholds it was found that the model was at its optimum classification accuracy with the confidence threshold set to 0.1, see Table 3.1. With this configuration, the resulting F_1-score was 82.63% for detecting Type III solar radio bursts. YOLO's predictions at confidence threshold of 0.1 are illustrated in Figure 3.5 for a typical 10-minute segment.

Table 3.1 YOLOv2's accuracy per confidence threshold.

Model accuracy per confidence threshold					
Conf. thresh	Precision	Recall	f1-score	True Positive	False Positive
0.5	10.79%	98.35%	19.44%	1,310	22
0.4	26.22%	98.36%	41.41%	3,185	53
0.3	46.33%	98.13%	63.04%	5,628	107
0.25	56.81%	97.57%	63.04%	6,910	172
0.2	68.20%	96.75%	80.01%	8,313	279
0.15	75.13%	90.73%	82.30%	9,220	942
0.1	83.42%	81.85%	82.63%	9,796	1,947

3.7 Summary

A parametric modelled simulated dataset of Type III SRBs embedded in RFI channels was used to train the YOLOv2 model. This model efficiently detected and classified Type III SRBs with a high F_1 accuracy of 82.63% for real-time frame rates. The quality of the simulated training set was key to achieving high accuracy. However, the parametric modelling method showed some issues with SRB intensity fluctuations, when simulated, and detection robustness and localised accuracy, when applied to real data. The following chapters will address these issues and constraints.

Chapter 4 explains how GAN was applied to the SRB simulation problem to create realistic SRB examples. Using this new simulation method, Chapter 5 shows how a new training set with simulated and real Type III SRBs was created to reevaluate the robustness and accuracy of the YOLO model.

The full source code for SRB detection with YOLOv2 is available on GitHub (Scully 2022a).

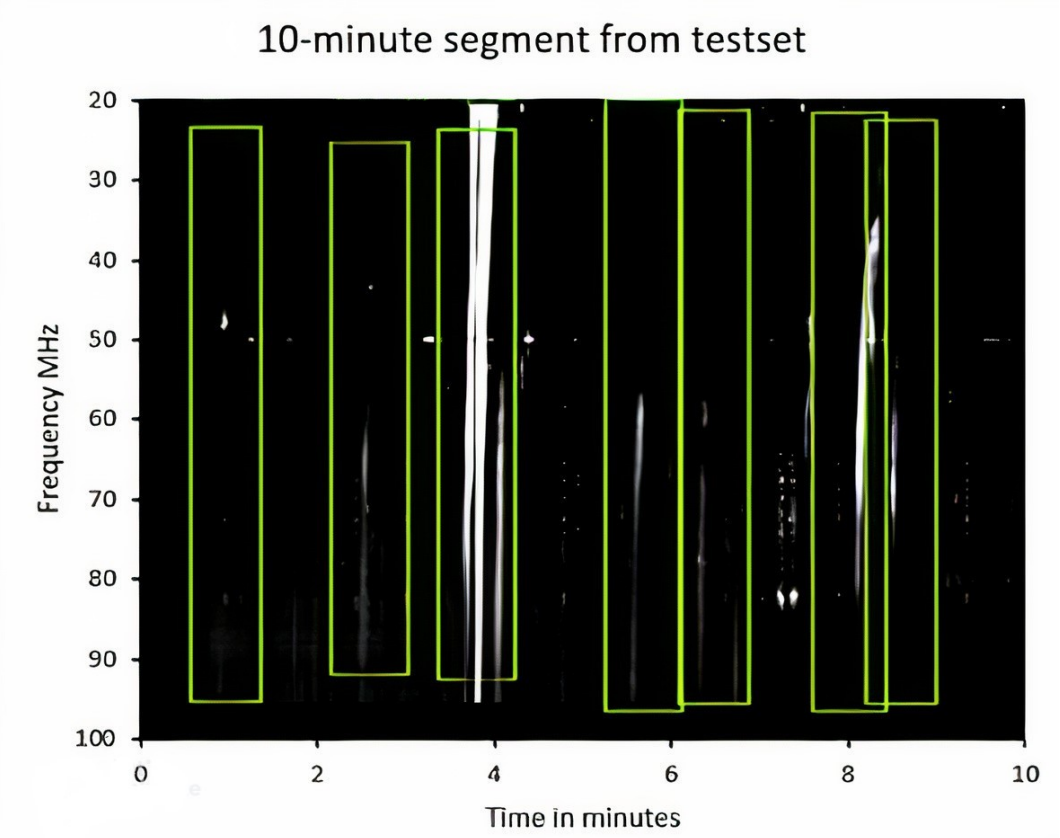


Fig. 3.5 Detections of Type III SRBs on a 10-minute segment at its most optimised classification confidence threshold of 0.1 on the testset.

Chapter 4

GANs for SRB simulation

4.1 Introduction

The development of YOLO and the constraints with the simulated parametric modelled training data led to the second research question:

Sub Research Question 2: Is it possible for deep learning methods to generate simulations of Type II and Type III solar radio bursts that exhibit comparability with actual observed data?

Traditional methods of creating training sets for SRB classification and detection often involve the tedious task of sifting through gigabytes of data archives and collecting relevant training examples of the desired radio burst. The idea of generating simulated data eliminates this tedious task as it allowed us to generate data that not only looks like SRBs but also produces it in volume and in a short space of time. In Chapter 3, parametric modelling created SRB-like images in huge quantities and in a short space of time. However, simulated data produced from simple parametric models do not produce simulated SRB data that are comparable to real observed data in terms of features like time and frequency structure and intensity consistency. This chapter presents the GAN as an approach used to answer Research Question 2.

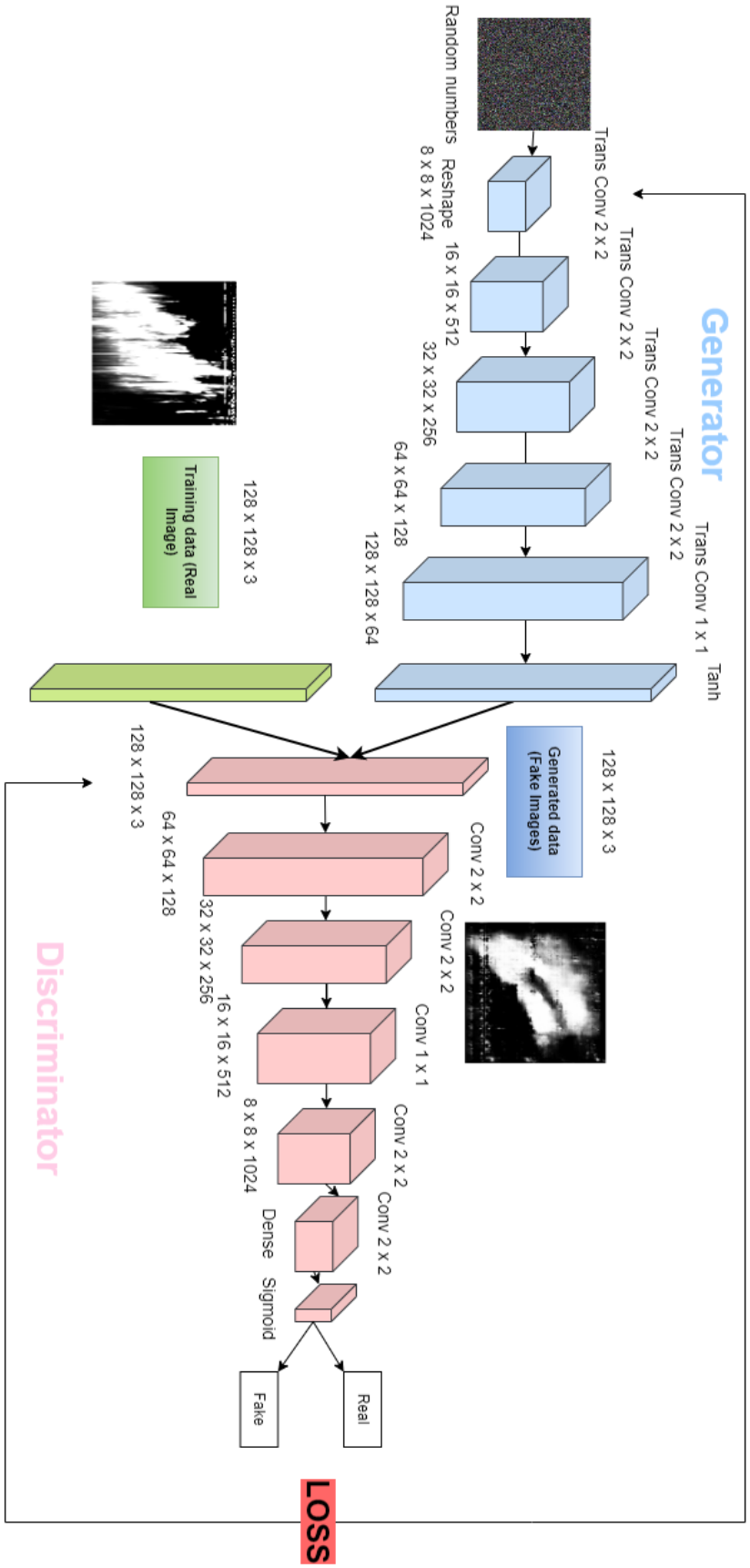


Fig. 4.1 Example of a GAN architecture for generating simulated Type II and Type III SRB data.

4.2 Generative Adversarial Network

As discussed in Section 2.4, GANs are a type of artificial intelligence algorithm used for generating new, previously unseen data samples that are similar to a training dataset. They consist of two deep neural networks: a generator network and a discriminator network, see Figure 4.1. The two networks are trained simultaneously, with the generator trying to produce samples that can fool the discriminator, and the discriminator trying to correctly identify real and fake samples. As the training process continues, the generator improves at generating realistic samples, which look like they come from the training dataset, and the discriminator improves at distinguishing real from fake samples. In effect, GANs generate realistic synthetic data from existing available data.

4.2.1 Discriminator

When describing the GAN process as a whole, the discriminator receives a batch of real or fake images based on the training stage, as shown in Figure 4.1. It passes through five convolutional layers with (2x2) and (1x1) strides to downsample the data. Batch normalisation is applied after each convolutional layer, aiding deep learning by ensuring consistent scale and mean of inputs, leading to smoother training; this is followed by ReLU activation, which sets negative values to zero. The sigmoid activation function is used in the final layer, normalising the output within the range [0,1].

Discriminators are simple binary image classifiers that accept an image as input and output whether the picture is real (output = 1) or fake (output = 0), as seen in Figure 4.1. The goal of the discriminator is to distinguish fake images from real ones. In order to achieve this goal, two possibilities must be considered while determining the amount of error that the discriminator makes during the training phase:

- When a discriminator is fed real images, the error produced is called a real error or positive error.
- When a discriminator is fed fake images created by the generator, the error produced is called a fake error or negative error.

During the training phase, the objective function $\nabla_{\theta d}$ of the discriminator is the sum of positive and negative errors to be optimised (Goodfellow et al. 2016):

$$\nabla_{\theta d} = \frac{1}{m}\sum_{i=1}^{m}[\log D(x^i) + (\log(1 - D(G(z^i)))] \tag{4.1}$$

where $\log(D(x^i))$ refers to the probability that the generator is correctly classifying the i^{th} mini–batch of size m of real images x and $\log(1 - D(G(z^i)))$ is the probability of correctly labeling the i^{th} mini-batch of fake images z that comes from the generator. The GAN is updated by inputting a batch of (128x128x3) real images to the discriminator and using the output to update the discriminator by changing the weights of the discriminator with each epoch to minimise the loss. The resultant output vector will contain values ranging from 0 to 1. These values are then compared to their true labels, i.e., real images as 1 and fake images as 0. To determine the gradient, the derivative of the loss function with respect to the weights in the model is calculated, after calculating the discriminator's loss across the real images. The generator is then used to generate a batch of fake images using some random noise input. These images are then fed into the discriminator, which gives predictions for these fakes (values between 0 and 1). The loss is computed by comparing these predicted values to their true labels, which is 0 in this case. After calculating the loss across fake images, the derivative of the loss function is used to calculate gradients, much like with the real images. Finally, to minimise the overall discriminator loss, the weights W_n are adjusted, where n is the iteration index:

$$W_{n+1} = W_n\,(1 + learningrate * gradient) \tag{4.2}$$

The *learningrate* determines how the weights of the network are adjusted in relation to the loss gradient descent, i.e. how quickly or slowly the model will approach the optimal weights. The gradient descent algorithm minimises a cost function at each step to estimate the model weights over a number of iterations.

4.2.2 Generator

On the other side, the generator takes random numbers as input and utilizes a deconvolutional neural network, as illustrated in Figure 4.1. It employs five transpose convolutional layers with (2x2) and (1x1) strides to upsample the data. After each transpose layer, batch normalization and Leaky ReLU activation are applied, allowing small negative values for inputs below zero. The final layer uses the Tanh activation function, suitable for generating images within the range [0,1] or [-1,1].

The task for a generator is to create an image that is realistic enough to mislead the discriminator. Using the discriminator, the amount of error from the generator during the training phase can be readily calculated (Goodfellow et al. 2016). The objective function of the generator $\nabla_{\theta g}$ is given by:

$$\nabla_{\theta g} = \frac{1}{m}\sum_{i=1}^{m}[\log(1 - D(G(z^i)))] \tag{4.3}$$

Generators are more complicated than discriminators. For each image, they take an array of random numbers, identical in size to the discriminator (128 x 128 x 3) to simulate a noisy input, as seen Figure 4.1. It should be noted here that these random numbers are initially influenced by the output of the discriminator. This input is then fed into the CNN of the generator, the output of which is an image that is further refined by multiple passes through the CNN. When updating the generator, a similar procedure to the discriminator is followed. The model starts by passing a batch of fake images to the discriminator (produced during discriminator training). This is because the discriminator is updated before the generator, necessitating a forward

run of the fake batch. The loss is then determined from the discriminator's output and the real label of the images. It is worth noting that, although these are fake images, their true labels are set to 1 throughout the loss calculations because the generator wants to fool the discriminator into thinking it is producing real images. Ideally, when the fake images are given as input, the generator wants the discriminator to output 1 or as close to 1 as possible. In this case, the loss function is used to reduce the difference between the output of the discriminator for fake images and the output of the discriminator for true images. Finally, based on the gradient, the weights are modified to minimise the total generator loss, and the process repeats until the specified epoch counter is met.

4.2.3 Convergence Failure and Mode Collapse

Due to its volatile training process (Goodfellow 2016), there is no way of knowing when the training process should end and when a final model should be saved for subsequent use because there is no objective assessment of model performance (Borji 2019). As a result, it is common practice to keep the current states of the generator and the model which results in a large number of fake images generated during training. One indicator that the GAN is producing realistic fake generated images is the training loss of the network. When the training loss of the generator spikes, it is a good indication that the generator is producing noisy or bad images during that step; this is known as convergence failure (Goodfellow 2016). It manifests itself as a significant divergence between the training losses of the discriminator and the generator that typically lasts for more than 10 iterations.

Another key feature in training a GAN is mode collapse (Goodfellow 2016), in which the generator attains a state where it always produces the same output, because the main aim of the generator is to fool the discriminator not generate diverse output. When mode collapse happens, the discriminator gets stuck in a local minimum and

is unable to find the best strategy for rejecting the diverse outputs of the generator and thus it keeps producing that one type of image. One method of identifying mode collapse is to plot the training loss of both the generator and discriminator vs epoch. This plot will show oscillations in the loss over time, most notably in the model loss of the generator. Another method to identify mode collapse is to observe the images post-training produced by GAN. As mentioned before during mode collapse, GAN will produce roughly the same type of image over and over again, see Figure 4.2. This enables post-epoch evaluation of each saved generator model based on its generated images. Models can be saved in a systematic manner across training epochs, in this case, the model and images generated by the model are saved after every epoch.

4.3 Training GANs for SRB Generation

4.3.1 Type III Generation

The GAN network in this research was first trained to generate simulated or synthetic Type III SRBs. The training set consisted of 2,763 real Type III images containing just over 33,000 (real) observed Type III examples that were gathered by combining different observation days made by I-LOFAR, focusing on the 20-100 MHz range as this is where a Type III's vertical strip shape is observed. These observation days varied in terms of solar activity and also contained a variety of different Type IIIs, such as inverted-U bursts and Type N bursts (Reid and Ratcliffe 2014). This data was then cleaned by simple image inspection and deletion as the goal was for the GAN to generate images that did not contain interference usually seen in observations, such as embedded RFI. As a GAN is a semi-unsupervised algorithm, there is no need to label any individual image data but to only classify it as a whole, meaning the collection of training images falls under the same Type III label.

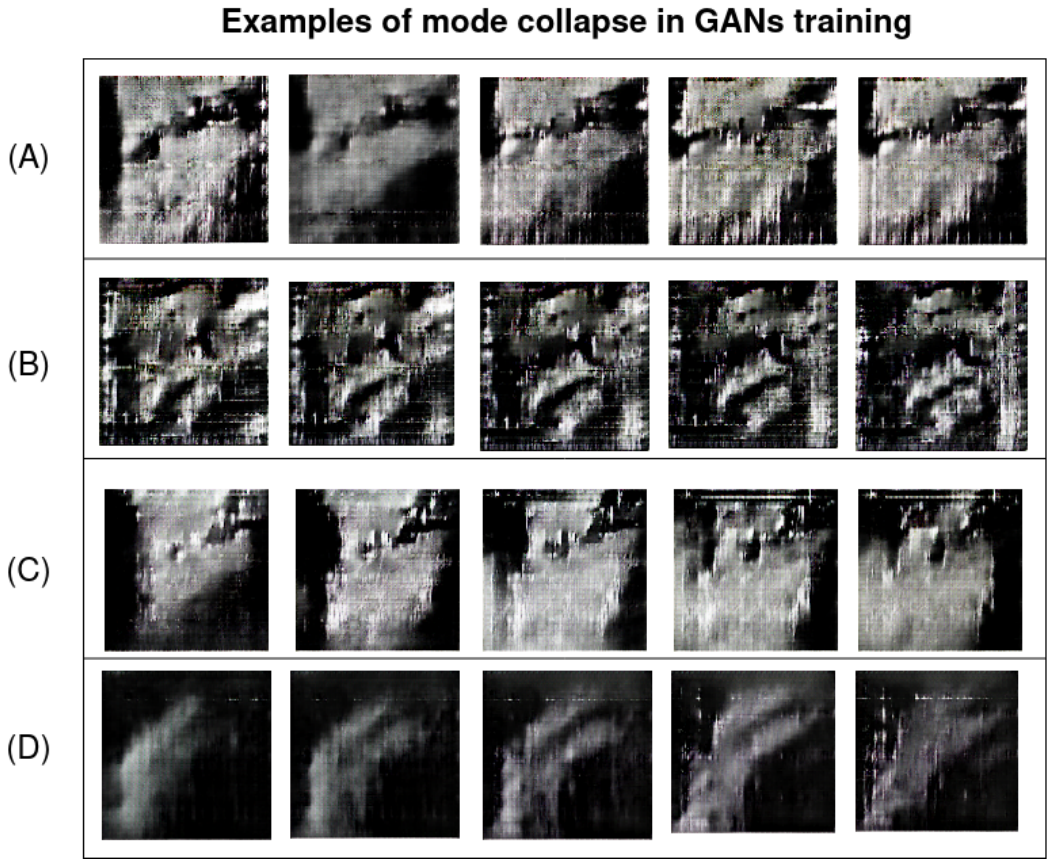


Fig. 4.2 Four different Type II cases of mode collapse during a GAN training instance. As it's quite difficult to identify oscillations in loss plots, post-training evaluation is the best way to detect mode collapse. Cases (A), (C) and (D) are examples of mode collapse in effect where GAN is generating the same Type II image with little variation. Case (B) is an example of a combined convergence failure and mode collapse where the repeated image produced is of low quality and very noisy.

This research was performed on a machine comprising 2 x SLI inter-connected GPU Nvidia Geforce RTX 2080 Ti, using Ubuntu 20.4.2 LTS on an AMD Ryzen Threadripper 1950x with 32GB of RAM. For the training configuration, 90% of GPU capacity was used for a variety of different epochs at a batch size of 32 as seen in Figure 4.3. GANs were trained numerous times, which built a collection of over 5,000 generated simulated Type III SRBs. However, the generated images were small (128x128 pixels), so the images were then bulk-rescaled from (128x128) to (256x256). Once this was done, a new synthetic training set was created that can be used with classification and object detection algorithms such as YOLO.

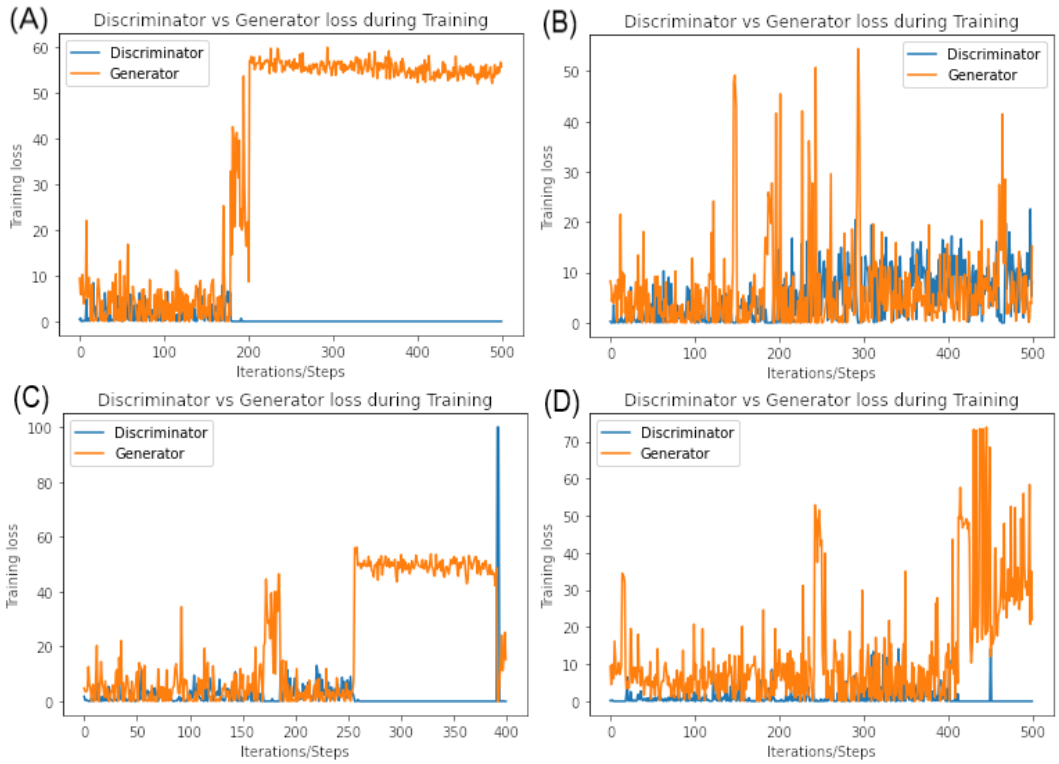


Fig. 4.3 The loss error battle between the discriminator and the generator when generating Type IIIs. This illustrates the GAN's learning pattern. Convergence failure can be seen in three training instances, plots (A), (C) and (D), where the generator's training loss increases for prolonged periods. Case (B) is a relatively good example of GAN training on Type IIIs.

4.3.2 Type II Generation

Training a GAN to generate Type II images is a more complex task than that for Type III SRBs. One constraint is the lack of observed data from the I-LOFAR data archive. This was due to the Sun's 11-year solar cycle. Since 2017, the date I-LOFAR was commissioned, the Sun has been in a solar minimum so the activity has been much lower than that prior to 2017. Consequently, only five observed Type II SRBs were gathered. The five observed Type IIs spanned a couple of hours, and to focus on their prominent shapes, they were divided into 10-minute segments, resulting in a training set of 214 images for the Generative Adversarial Network to learn from. This amount of data is not enough to train the GAN as it creates issues such as over-fitting and mode collapse. To combat this, data augmentation (Perez and Wang 2017) was implemented by taking copies of the images and editing the images slightly using

techniques such as blurring, cropping and feature removal to increase the size of the dataset to 1,527 images. Another issue with Type II SRBs is their shape as their structure has a degree of randomness. All images were considered even if there was radio interference within the data as the number of images needed to be maximised for training the GAN.

The Type IIs were trained on the same machine and training setup as the Type IIIs; the only difference was the increased number of epochs (see Figure 4.4) to allow the GAN to work with this complex yet very small amount of data. The GAN was once again trained numerous times to create a collection of over 5,000 Type II generated simulated examples. The generated data from the GAN was resized to 256 x 256. However, due to the constraints discussed previously, the generated Type IIs contained a lot of noise. This noise contained RGB colouring that distorted the image and in some cases disrupted the shape of the Type II. To combat this, the generated simulated Type IIs were converted to greyscale thus removing the RGB noise produced by the GAN in training.

4.4 GAN: Evaluation and Results

The GAN was trained multiple times on Type III and Type II datasets for varying numbers of epochs. During training, eight images were generated after each epoch. Then the noisy generated images corresponding to significant spikes in loss error (convergence failure) were manually filtered out along with repeated images produced by mode collapse. The collection of images was then bulk resized to 256x256 and converted to greyscale. Two methods were used to evaluate the GAN, namely, human perception and Fréchet Inception Distance (FID).

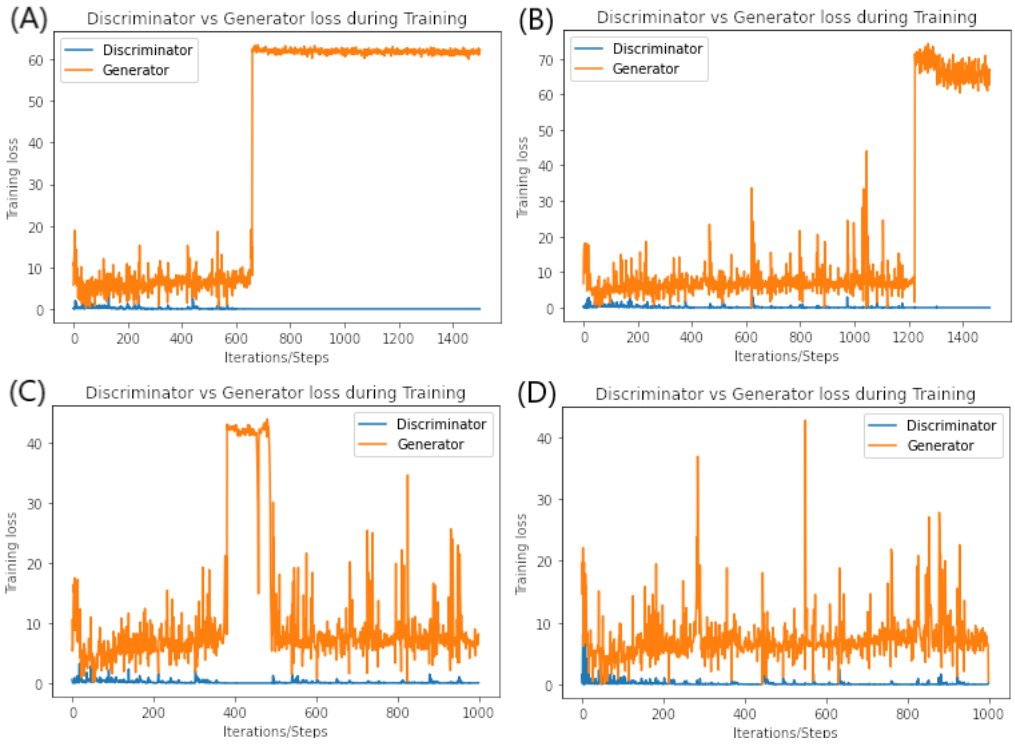


Fig. 4.4 The loss error battle between the discriminator and the generator when generating Type IIs, showing the learning pattern of GANs on Type II data. The number of epochs has increased relative to Type III training, due to the lack of data in the training set but the algorithm seems to train well up to the 1,500 epoch mark. Again, convergence failure is seen in (A), (B), and (C) where the modelling actually corrects itself during training. Case (D) is an example of a relatively healthy GAN learning pattern with no major generator loss spikes.

4.4.1 Human Perception

Manual assessment or human interpretation of images produced by a generator model is the common way to evaluate a GAN (Borji 2019). The generator is used to create a batch of fake generated images, then an evaluation of the quality and diversity of the images in relation to the target domain (in this case, SRBs) is performed. To accomplish this, the GAN was trained over a number of epochs several times. Once a training instance was complete, the generated images were compared to both parametric and real observed data.

When visually comparing parametric-produced, GAN-produced, and real Type III data in Figure 4.5, notice how the parametric modelling method fails to replicate the consistent brightness intensity seen in the observed data. The parametric model

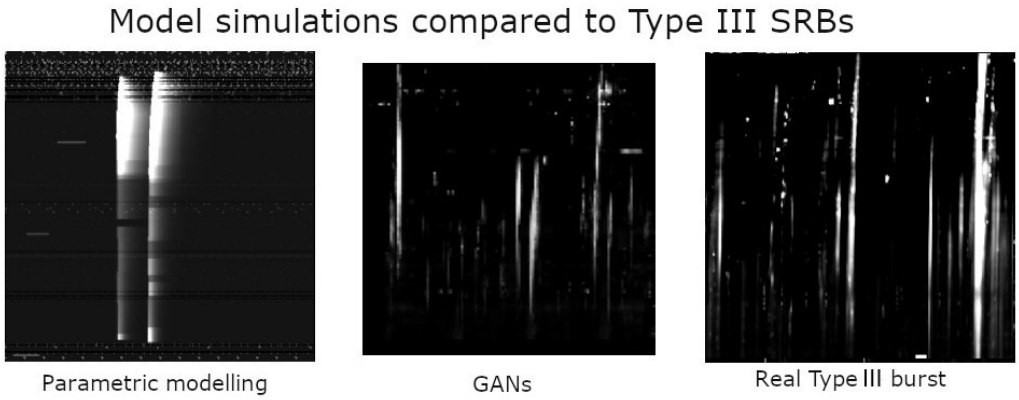


Fig. 4.5 Comparison of Type III SRBs (10-minute segments in the frequency range of 20-90 MHz) generated by parametric models and the GAN to real Type IIIs generated by I-LOFAR.

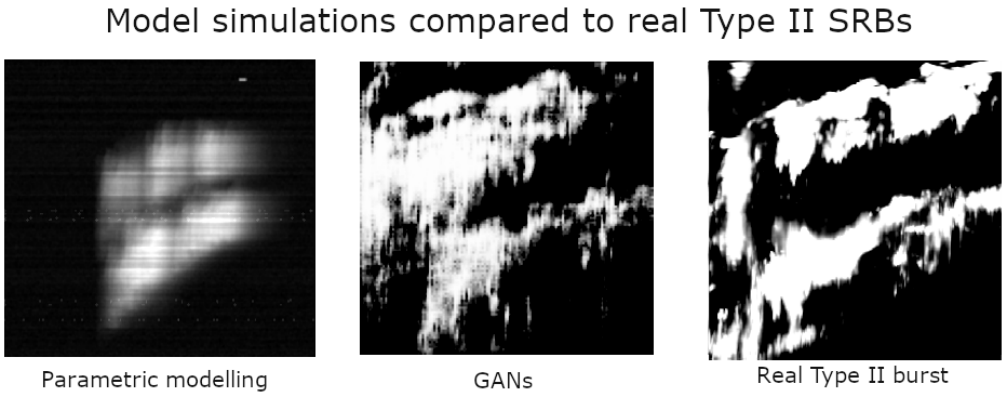


Fig. 4.6 Comparison of Type II SRBs (10-minute segments in the frequency range of 20-90 MHz) generated by parametric models and the GAN to a real Type II generated by I-LOFAR

seems to have an abrupt change in intensity about halfway through the burst. The GAN on the other hand has a consistent intensity drift from start to end with no abrupt changes in shape and intensity. While parametric modelling has the tadpole-like shape of Type IIIs it does not model the top and bottom of the SRBs as seen in the GAN and observed data, i.e., parametric models have flat tops and bottoms not seen in real data. This can be seen in the comparison between parametric, GAN and real Type III data in Figure 4.5. When comparing Type IIs in Figure 4.6, parametric models have the general shape of observed data but, in appearance, look like a real Type II that has been blurred. This blurring obscures crucial structure within the Type II. Furthermore, parametric Type IIs do not have the inherent RFI interference

or embedded Type III SRBs seen in real observations. A comparison between parametric-produced, GAN-produced and real Type II data is shown in Figure 4.6. The Generated images were also evaluated by experts in the field for more clarificaion on the resemblence between the real and simulated datasets. The experts assessed the GAN-generated SRB images by closely examining their visual fidelity, evaluating their resemblance to real data in terms of details, textures, and overall coherence. They analysed the images for realism, diversity, and finer elements, considering factors such as time duration, frequency, and intensity drift rate. Additionally, they compared the generated images to authentic examples. This subjective analysis, guided by the experts' domain-specific knowledge, provides insights into the GAN's effectiveness in producing accurate and compelling visual content for the intended application. Human perception can be used to evaluate a GAN, however, it is especially subject to observer bias; SRBs are not like cars, people, animals, etc. but are abstract objects that can be easily misclassified. The issue of human biases is addressed metrically by the FID which calculates the similarities between datasets produced by parametric modelling, GANs and real observed data.

4.4.2 Fréchet Inception Distance

The FID is a metric for assessing the quality of generated images and was created primarily to assess the GAN performance without human perception biases (Heusel et al. 2017). The FID score was designed with the aim of comparing the feature vectors (a feature vector is a list of numbers used to represent an image) of a group of fake-generated images to the feature vectors of a collection of real images from the target domain, in this case, SRBs, to evaluate how good the fake-generated images are compared to the real observed data. A lower FID suggests higher-quality images, a higher score indicates lower-quality images, as shown in Figure 4.7.

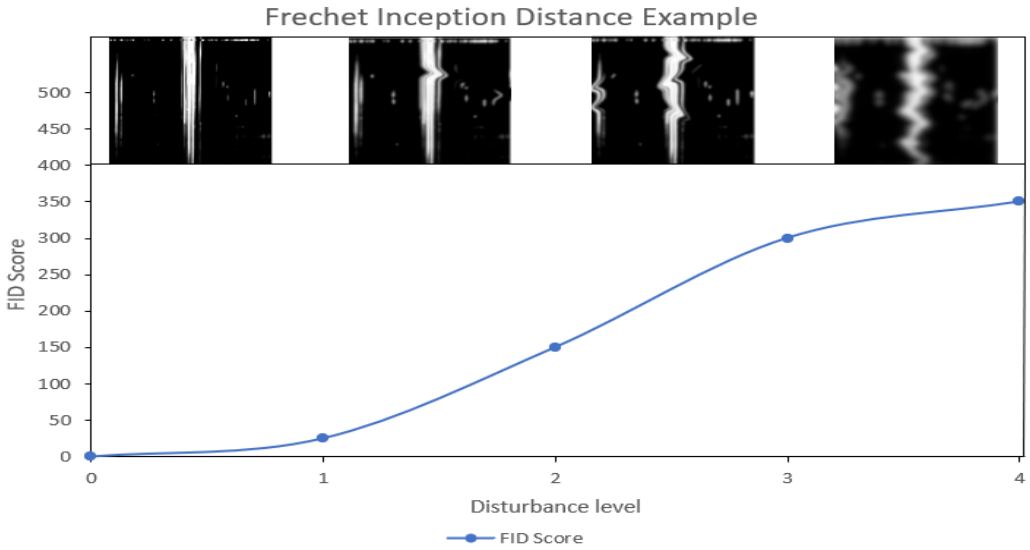


Fig. 4.7 An example of FID comparing the disturbance level of the same image. Notice how when the image is distorted the FID score rises. The lower the FID score the better the quality of the generated SRB image compared to a real SRB image.

The FID score utilises Google's pre-trained RF CNN model (Szegedy et al. 2015). Specifically, the last pooling layer prior to the output classification of images is used to capture features of an input image, without the classification scores. The model's output layer is removed, and the feature activations from the last pooling layer, a global spatial pooling layer, are used instead. Each image is predicted as having 2,048 activation features because this output layer includes 2,048 activations. To offer a reference for how real images are represented, a 2,048 feature vector is predicted for a set of real images from the problem domain (SRBs). For fake-generated images, feature vectors are then computed. Two collections of 2,048 feature vectors for real and generated images will result. These feature vectors are calculated for a collection of real and generated images. By computing the mean and covariance of the images, the feature activations are summarised as a multivariate Gaussian. The activations across the collection of real and generated images are then calculated using these statistics. The Fréchet distance is used to calculate the similarity distance between these two feature distributions (real and fake). The following equation is then used

to determine the FID score:

$$d^2 = ||mu_1 - mu_2||^2 + Tr(C_1 + C_2 - 2 * \sqrt{C_1 * C_2}) \tag{4.4}$$

where the FID score is referred to as d^2, showing that it is a distance and has squared units. The mu_1 and mu_2 are the feature-wise mean of the real and generated images, respectively. Tr represents the trace of a matrix. The C_1 and C_2 are the covariance matrices for the real and generated feature vectors respectively (known as sigma). The $||mu_1 - mu_2||^2$ is the sum squared difference between the two mean vectors. Using this method, the FID compared scores of both parametric and GAN-generated Type IIIs and Type IIs to real observed Type IIIs and Type IIs as seen in Table 4.1.

SRB Type	Comparison	FID Score
Real Type III	Real Type III	0
	Parametric Type III	34.08
	GAN Type III	9.23
Real Type II	Real Type II	0
	Parametric Type II	78.15
	GAN Type II	34.01

Table 4.1 A comparison of FID scores between simulated methods compared to real SRBs. Notice when comparing real observed training sets to themselves the result is 0. The objective is to get a FID score as close to 0 as possible, indicating a good similarity between simulated and real data.

4.5 Summary

A GAN was utilised to produce simulated Type III and Type II SRBs from real observations in the LOw Frequency Array archive. The evaluation of the generated images using human perception and the FID score showed the effectiveness of GANs in creating realistic models for classification and detection algorithms. The combination of simulated and observed data was also discussed as a means of enhancing detection accuracy.

In Chapter 5, the focus is on creating a more diverse training set by combining simulated and observed data for Type III SRBs using GANs. The ultimate goal is to improve the accuracy and robustness of the YOLOv2 model for localised Type III SRB detection.

The full source code for the GAN SRB simulation method is available on GitHub (Scully 2022b).

Chapter 5

Improved Type III SRB detection using congruent deep learning models

5.1 Introduction

Once again, recall the first sub research question presented in Chapter 1:

Sub Research Question 1: How effectively can deep learning methods be employed to classify and locally detect Type III solar radio bursts?

This investigation combines the approaches described in Chapters 3 and 4 to answer this research question. In Chapter 3, the YOLO algorithm was adapted to detect Type III SRBs. Using this configuration, a classification accuracy score of 82.63% was obtained. However, one score that couldn't be obtained was mean Average Precision (mAP), which is the evaluation metric that determines how precise the algorithm can locate a certain object in an image, in this case, Type III SRBs. Some key areas were noted for improvement from the investigation in Chapter 3, the main one being the simulated training set. In this experiment, congruent deep-learning methods are applied to the problem of SRB simulation, detection, and classification. In deep learning, congruence refers to the similarity or consistency between different parts

of a model or different models. It also refers to the similarity between the weights of different models trained on the same task, or the consistency of the outputs of different models when presented with the same input. The GAN described in Chapter 4 is used to simulate Type III SRBs comparable to real observations, and create a new training set for input to YOLO to detect and classify the locality of Type III SRBs.

5.2 SRB simulation

Creating big datasets for training classifier systems has necessitated simulating SRBs. It essentially eliminates the time-consuming task of searching for data within data archives and sifting through it to find good, clean images suitable for a training set. Simulating data also gets rid of the need for cleaning archival data of any interferences such as RFI. Parametric modelling was used to simulate SRBs in Chapter 3, in which polynomials were used to create the Type III shape in dynamic spectra, while skewed Gaussians were used for their intensity profile across time at each frequency. The bursts were then placed in front of a background of synthetic and random RFI channels.

When training YOLO, using simulated data has a number of advantages, including the ability to generate enormous amounts of data in a short amount of time. Using parametric modelling, a dataset of 80,000 images of Type III SRBs was constructed for the previous YOLO model. On paper, having 80,000 images and a class of one phenomenon sounds great, but having a large amount of data is only one element of the equation, and the training dataset in this case had some drawbacks. The first difficulty is that this data must be accurately labelled for training; YOLO needs to see what it's looking at because it's a supervised learning technique that requires it to train on a labelled dataset in order to identify a class. The automatic labelling that is characteristic of the parametric modelling saved a lot of time, but there was no

variation heights in terms of frequency of these labelled bounding boxes. In other words, the height variable became saturated with static values in the training set, therefore the final YOLO detections had the same height variation as the training set, as seen in Figure 5.1. As a result, the localised precision of the model could not be calculated. The second difficulty experienced with the parametric modelling approach was the model's lack of realism. While it provided many possibilities in terms of position, grouping, and overall shape and intensity, it did not provide the exact shape and intensity variation seen in a real observation.

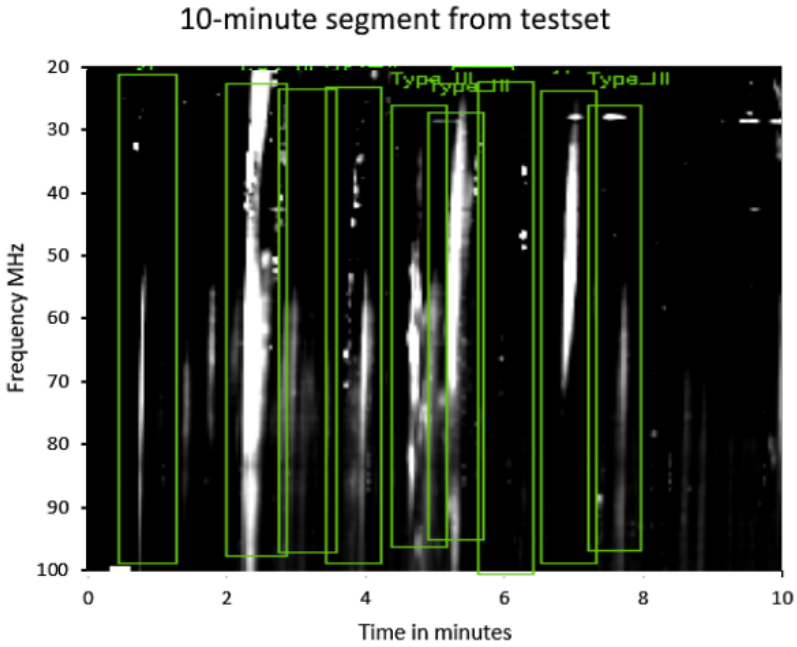


Fig. 5.1 A 10-minute segment from the test set of the previous model attempt described in Section 3.5. The lack of variation and over-saturation in the Y-axis or height variable in the training set meant that when the previous YOLO model was evaluated, the predictions in the test set had no variation in the Y-axis or height variable no matter the size of the Type III.

For Type III SRB simulations, GAN was used as described in Chapter 4. The Type III SRB simulations generated by the GAN looked realistic. GAN created SRBs that were similar to those seen in actual I-LOFAR observations obviating the need to trawl through data archives for appropriate images for the training set. These

new simulated Type III SRBs also offered the Y-axis variation that YOLO needed to make localised Type III SRB detections.

5.3 Type III simulation with GANs

The same method for Type III simulation was used as described in Section 4.3.1, where the GAN network was employed to create simulated Type III SRBs using real images obtained by I-LOFAR that were cleaned using colour augmentation to remove interference. The training set was made up of 2,763 images and the algorithm was run on a machine with two Nvidia Geforce RTX 2080 Ti GPUs. The training configuration used 90% of GPU capacity at a batch size of 32 and generated images were output after each epoch to create a dataset of over 4,500 simulated Type III SRBs. The generated images were initially 128x128 pixels but were then bulk-rescaled to 256x256 pixels.

5.4 Dataset

The key feature of the updated YOLOv2 model, compared to that in Chapter 3 was the dataset. Instead of using only parametric modelling generated data, a hybrid dataset was used. Which offered more realistic data for YOLO to train on compared to the parametric modelling. The hybrid dataset consisted of data generated by the GAN and real observed data from I-LOFAR. This improved dataset (6,732 images) was considerably smaller compared to the parametric modelling training set (80,000 images). The memory requirement of the training set is smaller so it can be easily transferred without the risk of data corruption and there are also fewer data to be labelled. The only constraint with this approach was that this dataset needed to be manually labelled. Once the dataset was labelled, the training set consisted of 6,732 images with just over 60,000 examples of Type III SRBs.

5.5 Model Configuration

The YOLO model used for this experiment was identical to that described in Section 3.3, i.e., it has 19 convolutional layers, 5 maxpool layers, and the number of filters in the last convolutional layer is changed to detect only one class, Type IIIs. Using the anchor values, the sizes of the bounding boxes were changed in width and height. The bounding boxes were made relatively narrow since Type IIIs are quite short-lived temporally, usually lasting only a few seconds. As mentioned in Chapter 3, the original height of the bounding box was static due to the lack of Y-axis value variation in the parametric modelling approach. With the new and improved training set, the anchor values or the height range were adjusted to 30MHz, 40Mhz, 50MHz, 60MHz and 80MHz. This captured most Type III sizes in terms of width and height detected by YOLOv2. As before, to improve real-time frame rates, the number of bounding boxes predicted on each test image was decreased from 845 to 405 and the input size was changed from 416x416 to 288x288. At this input size, the model ran at a maximum of 77 fps with an accuracy comparable to Fast R-CNN (Ren et al. 2017).

5.6 Training and Validation

The YOLOv2 model was trained to detect and classify Type III SRBs. The training set implemented a hybrid system in that a portion of the training set contained 3,000 simulated GAN images and 3,732 real images. A subset of the training set, the validation set, contained 1,500 Type III images produced by the GAN. The training and validation sets were both manually labelled using LabelImg (Tzutalin 2015), which labels objects within an image in Pascal VOC format or XML. Although manual labelling was a tedious task it provided YOLO with precise instructions on what to train on within the specified image. To fulfill the requirements of the

Darkflow framework, these manually labeled training set images were converted into XML. The training and validation set were then fed into YOLO for training once prerequisites (dataset and model configuration) had been completed.

To determine whether the model was overfitting or underfitting, the Darkflow framework was constructed to allow training validation. Leaky ReLU was used as the activation function during training and the model was trained using a learning rate based on the learning pattern of the model. If the model learned too quickly, the learning rate was updated, resulting in a smoother learning pattern. Stochastic Gradient Descent (SGD) was used to continuously update the learning parameters until convergence. This study was conducted on a system with an Intel Core i7-10700K processor, 32GB of RAM, and an Nvidia Geforce RTX 2080 Super GPU running Ubuntu 20.4.2 LTS. For the training configuration, 70% of GPU capacity was employed for 100 epochs at a batch size of 16. With this configuration, training took 7 days with both training and validation losses decreasing with every iteration, see Figure 5.2.

5.7 Test Set

When testing the previous model, a specific observation date was chosen, the 10th of September 2017. This specific date was chosen because it contained clear Type III examples with barely any RFI or smaller Type III examples that could properly test the YOLO model. Although the resulting data provided good benchmark results, the particular observation day never tested the robustness of the model as this date was relatively uneventful. To test how robust the model was, a test set containing a variety of Type III examples was required. So, instead of having just one observation, multiple observations from different dates in the I-LOFAR archive were gathered to build a test set that contained examples of busy and quiet periods of solar activity. The archive's raw data was then transformed into an image and plotted from 0-250

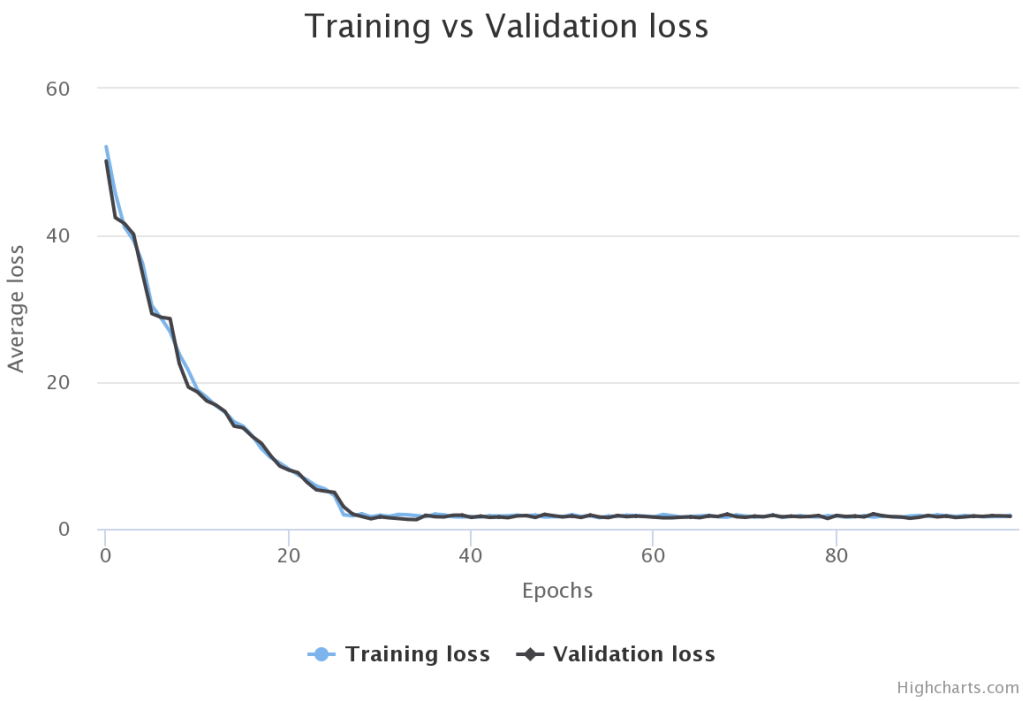


Fig. 5.2 Comparing training loss with validation loss illustrates how good YOLO learned on the training set. It is also used to prevent algorithm overfitting. Each epoch represents 422 iterations or when the dataset was passed forward and backward through YOLO once.

MHz over the course of the observation (time). To remove the affect of colour on the model's predictions, this image was then converted to greyscale to match the traits of the training set. As before, the observations were restricted to the 10-90 MHz range because this was where the Type III's vertical strip shape was most prominent. The observations were then divided into 10-minute intervals to provide a test set of 2,763 images containing around 35,000 Type III solar radio bursts. Using LabelImg, the ground truth bounding box values were precisely annotated with appropriate bounding box coordinates saved in an XML file. These box values were used to compare the ground truth coordinates to the predicted coordinates of the models.

5.8 Mean Average Precision

The YOLOv2 model's performance was measured using the test set of 2,763 images from I-LOFAR described previously in Section 3.6. In Chapter 3, the model's

performance was evaluated as a detection classification problem. This was done by prediction on the test set using the model, followed by a search through the entire test set for correctly identified Type III SRBs. Once a bounding box encompassed a Type III, it was categorised as correctly identified and then annotated a bounding box around the predicted bounding box. This meant that the previous model's performance (from an accuracy perspective) could only be represented by the F_1-score. Although it was possible to determine the accuracy of the model using the

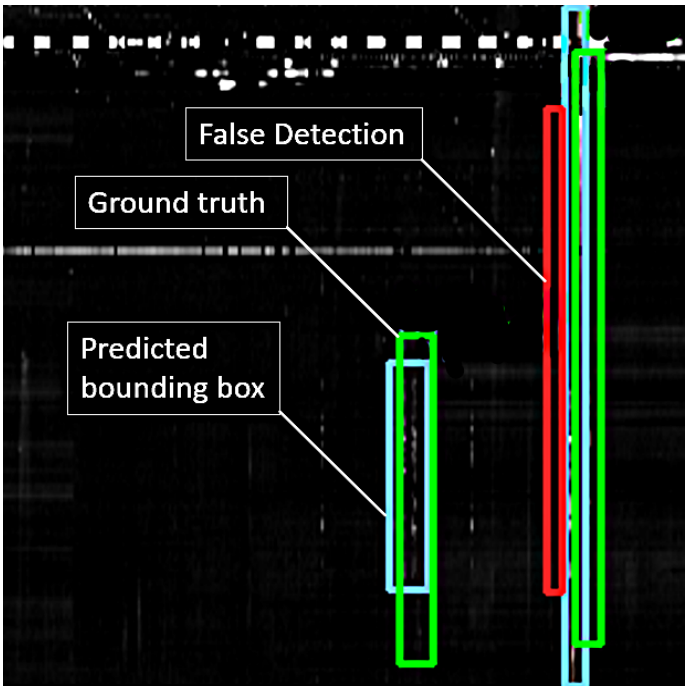


Fig. 5.3 Visual representation of IoU thresholding: the green bounding box indicates the ground-truth or actual Type III, blue indicates the predicted bounding box by YOLO and red indicates a false detection (FN) or IOU$<$0.5.

F_1-score in Chapter 3, it was not possible to calculate mAP, a metric used in computer vision to evaluate the performance of object detection and image classification algorithms. It calculates the average precision across all classes and is expressed as a fraction between 0 and 1. It takes into account both the number of true positive (TP) detections and false positive (FP) detections. A higher mAP score indicates a better performance of the algorithm. The TP and FP variables in mAP are determined by

comparing the ground truth bounding box to the model's predicted bounding box, the IoU.

The IoU values from tested data was used to determine the TPs and FPs. YOLO's predicted bounding box were compared to the actual ground truth bounding box to obtain the IoU. If the IoU was greater than 0.5, a prediction was categorised as a TP, and if it was less than 0.5, as a FP. An instance of a false detection (FN) where the model missed an identified Type III SRB is shown in Figure 5.3. The confidence threshold was an important consideration when assessing the model's performance. Figure 5.4 shows that the lower the confidence, the more detections made on a test image, but also the more false detections made. The model was found to be optimal with the confidence threshold set to 0.35, as there was a balance between the TP and FP rate in terms of the models predictions (see Table 5.1 and Figure 5.4). Figure 5.5 shows the bounding box predictions when model configuration was set to threshold 0.35, the resulting mAP for detecting Type III SRBs is 77.71%. This result is quite promising, demonstrating a high detection accuracy even on abstract objects. For reference, YOLOv2 achieves a mAP accuracy of 69% on the COCO dataset. The localised YOLO detections are plotted into a dynamic spectra observation made on the 10th of September 2017, see Figure 5.6.

5.9 Summary

A combination of congruent deep learning models was applied to generating, detecting and classifying SRBs in real-time to produce realistic Type III bursts based on real observations. The GAN-generated data was then combined with real observed data to produce a training set that was used to evaluate the YOLOv2 model for accurate detection. Using this method, the model achieved a mAP accuracy of 77.71% at detecting Type III SRBs. The model also demonstrated significant robustness when tested on a larger test set of over 35,000 Type III examples. The successful

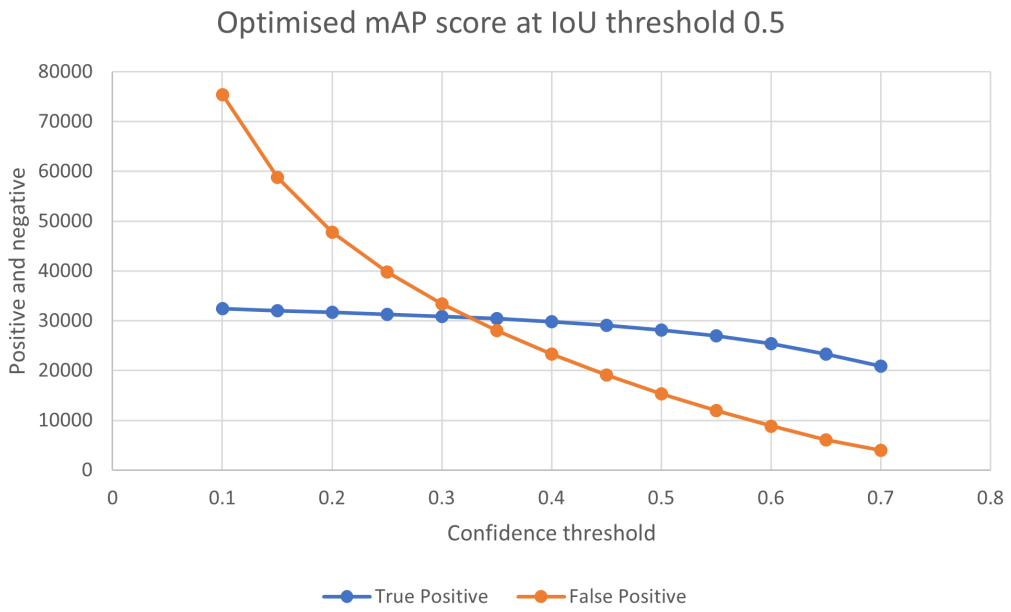


Fig. 5.4 A visual representation of Table 5.1. As the confidence threshold decreases, the true positive and negative values increase. YOLO performs at its most optimised confidence threshold at 0.35. The key is to find the balance between both true positive and true negative values. The model has been evaluated using IoU threshold at 0.5 as it tests the models robustness and localised accuracy at detecting Type III SRBs.

integration of simulated SRBs, using a GAN, into object detection algorithms raised the possibility of expanding the detection to Type II SRBs.

Chapter 6 introduces Mask R-CNN, an object detector that not only detects but segments the objects it's trained on. The objective is to not only detect Type II SRBs but to extract their true shape and intensity within dynamic spectra.

Table 5.1 The mAP scores associated with different confidence thresholds set in YOLO at different IoU thresholds. Notice how when the confidence threshold decreases the mAP increases but so too does the true negative and positive rate. The challenge is to find a balance between metrics for optimised performance in terms of accuracy.

Model accuracy per confidence threshold					
IOU Threshold	Confidence Threshold	Recall	True Positive	False Negative	mAP
IOU Threshold @ 0.5	0.7	84.03%	20838	3959	57.56%
	0.65	79.30%	23347	6093	63.57%
	0.6	74.14%	25398	8858	68.19%
	0.55	69.34%	26942	11911	71.44%
	0.5	64.74%	28140	15321	73.79%
	0.45	60.36%	29094	19099	75.55%
	0.4	56.18%	29828	23261	76.80%
	0.35	52.03%	30404	28026	77.71%
	0.33	50.42%	30592	30073	78.00%
	0.3	48.02%	30885	33422	78.42%
	0.25	44.00%	31295	39827	78.97%
	0.2	39.86%	31685	47801	79.45%
	0.15	35.28%	32060	58811	79.87%
	0.1	30.08%	32422	75336	80.21%
IOU Threshold @ 0.1	0.7	84.74%	21014	3783	58.32%
	0.65	79.99%	23549	5891	64.45%
	0.6	74.75%	25604	8648	69.12%
	0.55	69.90%	27162	11691	72.42%
	0.5	65.23%	28353	15108	74.78%
	0.45	60.79%	29298	18895	76.53%
	0.4	56.58%	30039	23050	77.80%
	0.35	52.38%	30604	27822	78.71%
	0.33	50.75%	30789	29876	78.99%
	0.3	48.33%	31082	33225	79.41%
	0.25	44.27%	31387	39635	79.96%
	0.2	40.11%	31880	47606	80.45%
	0.15	35.49%	32255	58616	80.87%
	0.1	30.27%	32625	75133	81.22%

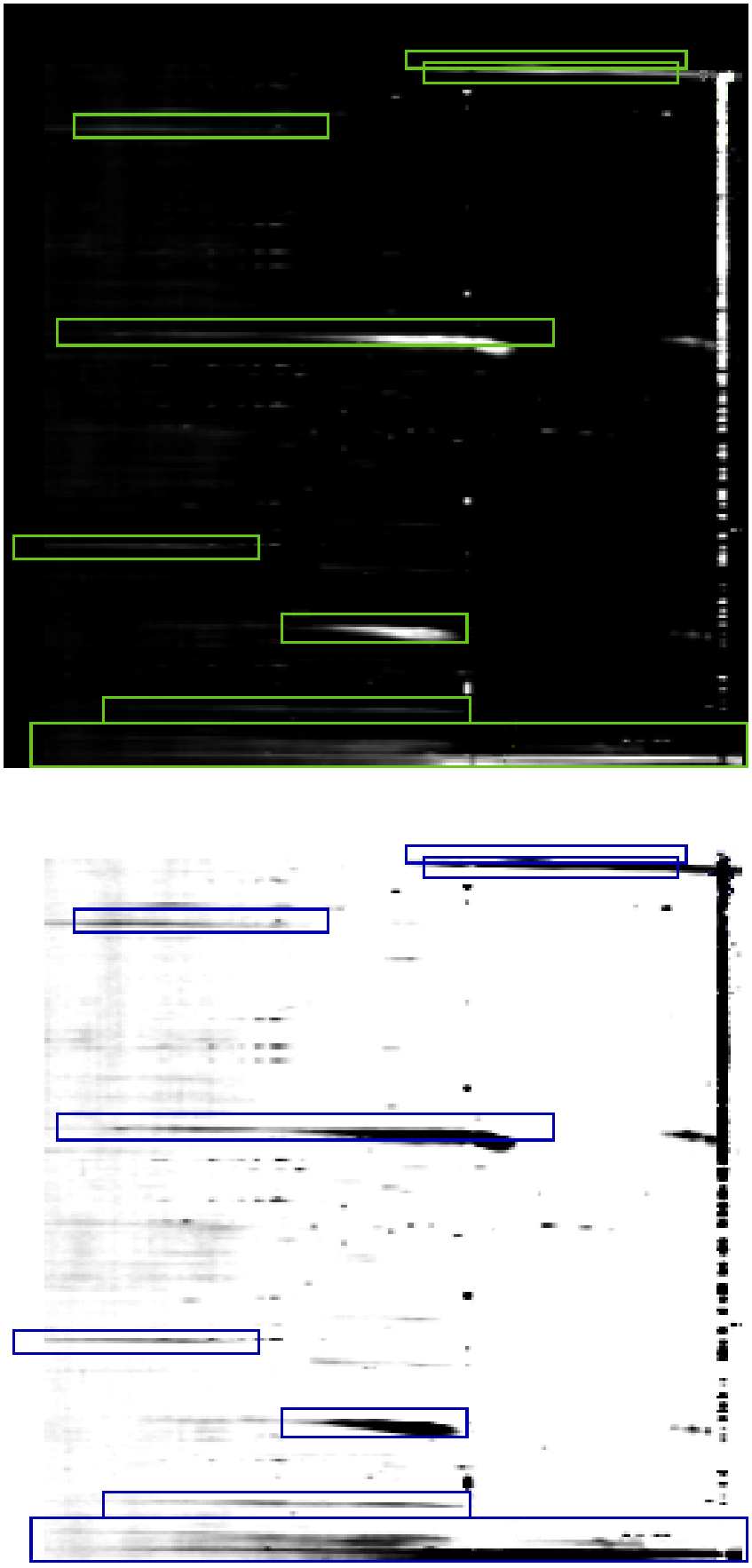


Fig. 5.5 YOLO making localised detections on a 10-minute segment at its most optimised confidence threshold of 0.35 on the testset (a). When the image is colour inverted (b), see how the faint Type IIIs are picked up by YOLO. Notice how YOLO picks up most Type IIIs in the image and ignores most RFI.

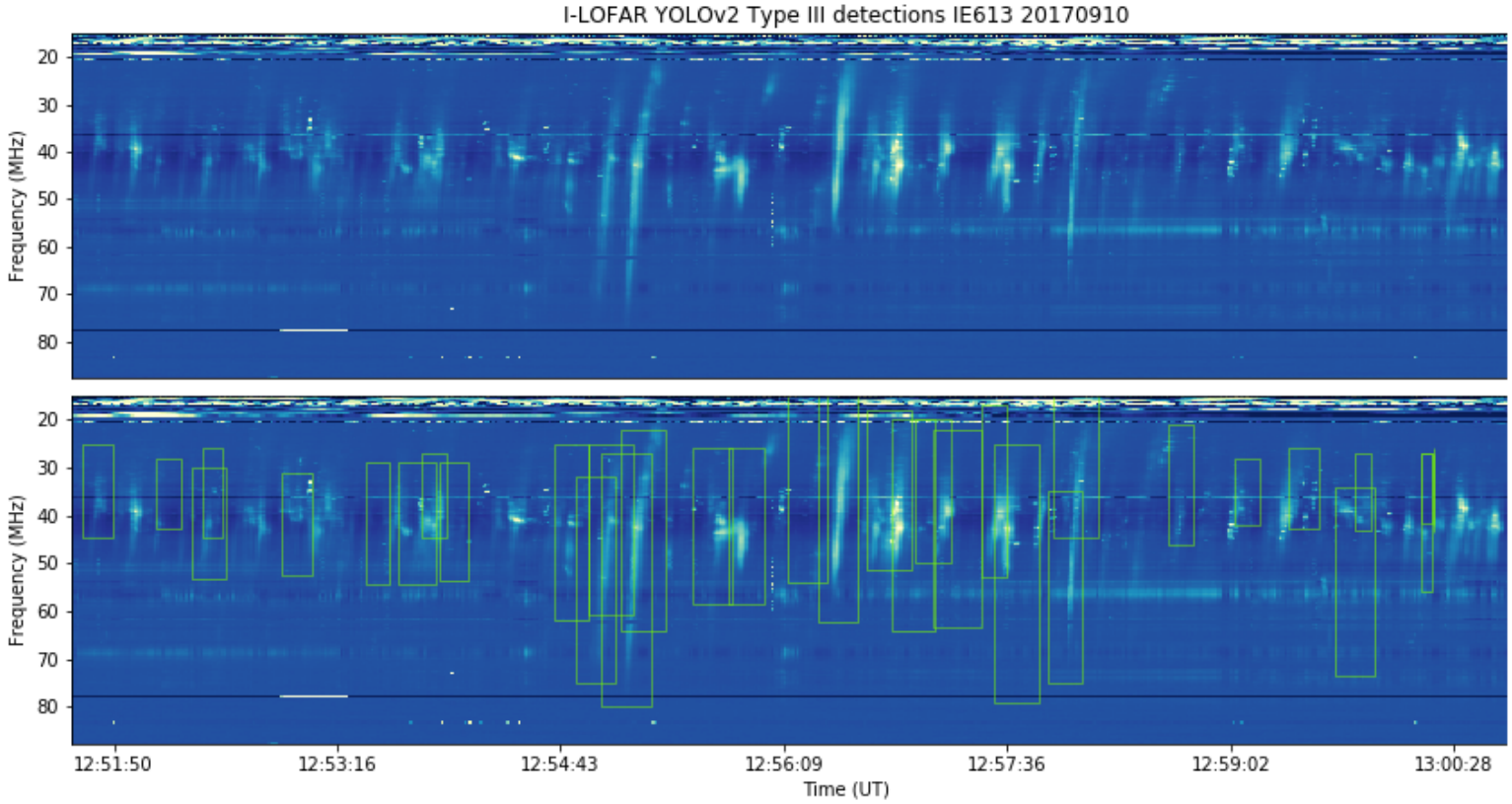


Fig. 5.6 YOLOv2 applied to an I-LOFAR observation made on the 10th of September 2017. The model's detections capture the Type IIIs' frequency range and length in time. The model predicts the most intense Type IIIs correctly and ignores somewhat low intensities as they are quite difficult to distinguish between Type III and RFI, even to the human eye.

Chapter 6

Using Mask R-CNN to automatically detect and segment Type II SRBs

6.1 Introduction

Sub Research Question 3: How effective are deep learning methods in detecting and segmenting Type II solar radio bursts?

Type II SRBs are the result of particle acceleration by shock waves in the solar corona and interplanetary medium. The shocks are created by solar eruptions involving Coronal Mass Ejections (CMEs) traveling at super-Alfvenic speeds (Jebaraj et al. 2021), see Figure 6.1. The automatic detection, classification and segmentation of such radio bursts is a challenge in solar radio physics due to their heterogeneous form. In this Chapter, a GAN is used to simulate Type II SRBs and this simulated data is used as a training set for Mask R-CNN to detect and segment Type II SRBs automatically.

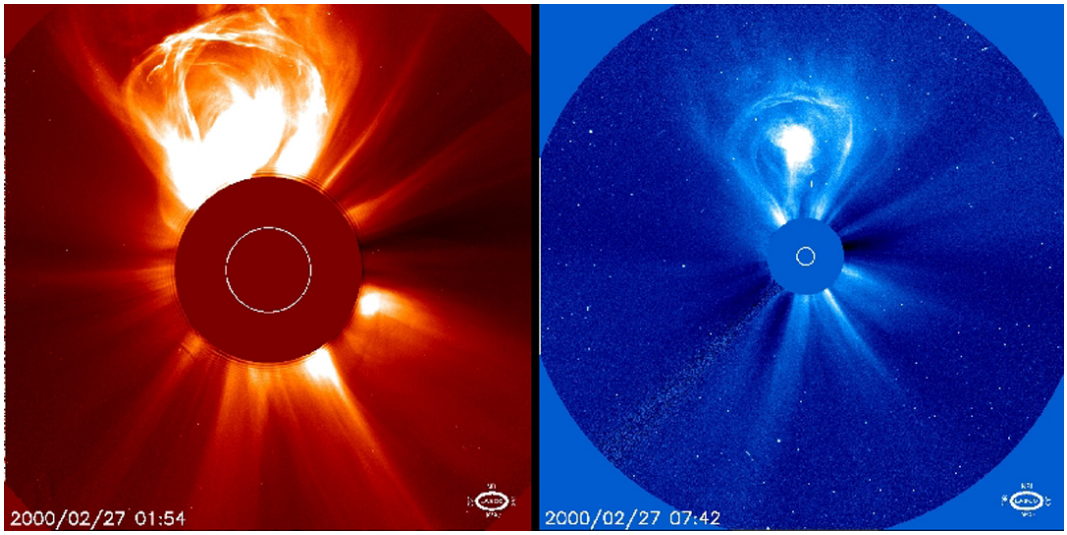


Fig. 6.1 Strong magnetic fields shape the corona, the outer solar atmosphere. The limited solar atmosphere can suddenly and violently expel bubbles of gas and magnetic fields known as CMEs where these fields are closed, frequently over sunspot groupings. One billion tons of matter, which can be accelerated to several million miles per hour in a spectacular explosion, can be found in a massive CME. Through the interplanetary medium, solar material flows out, striking any planets or spacecraft in its path. CMEs can happen independently but are occasionally linked to flares (Linker et al. 2003).

6.2 Dataset

The method used for Type II simulation is the same as that described in Section 4.3.2, where the GAN was used to generate simulated Type II SRBs using real images obtained by I-LOFAR that were cleaned using colour augmentation to remove interference. As discussed in Section 4.5.2, it was difficult to train the GAN to produce Type II images due to a lack of observed data from the I-LOFAR archive, caused by the Sun's 11-year solar cycle. This time, 9 observed Type II SRBs could only be sourced from the I-LOFAR repository and by breaking these into smaller chunks and using data augmentation, the training set was expanded to 1,835 images. Another issue with Type II SRBs is their arbitrary shape, so images with radio interference were once again included to provide more images for the GAN to train on.

6.3 Mask R-CNN architecture

Mask R-CNN is a state-of-the-art model for instance segmentation, developed on top of Faster R-CNN (He et al. 2020). Not only does Mask R-CNN return a classification and bounding box on a particular object, it also applies instance segmentation to the target object where a class label is assigned to each pixel of the image. The benefit of this is that, in addition to getting detection and classification, the true shape of the object (Type II SRB) is also captured. The Mask R-CNN can be split into two stages, see Figure 6.2. Two networks, ResNet 101 CNN and the Region Proposal Network (RPN), make up the first stage (He et al. 2016). These networks execute once for each image to provide a list of region proposals, which are feature map regions that contain the object. For each proposed region in this list, the network predicts bounding boxes and object classes in the second stage. While fully-connected layers in networks always require a constant-size vector to generate predictions, each proposed region can be of varied size. These proposed region sizes are set using either the RoIAlign technique or the Region of Interest (RoI) pool, which is quite similar to MaxPooling. The RoIAlign algorithm, which helps to maintain spatial information that becomes misaligned (as can happen with RoI pool), is used in the second stage of Mask R-CNN. Using binary interpolation, RoIAlign creates a fixed-size feature map, such as a 7 x 7 map (Figure 6.2). Then, using a regression model to further enhance border box prediction, warped features are input into fully-connected layers to perform classification using softmax activation. Additionally, the output of the RoIAlign layer is passed into the Mask classifier head, which has two convolutional layers. It creates a mask for each RoI, pixel-by-pixel segmenting an image. Then the bounding box with the greatest value for IoU is chosen and the other bounding boxes that are used to identify the same item are suppressed; this is done using non-Max Suppression. Finally, any bounding boxes with an IoU less than or equal to 0.5 are deleted.

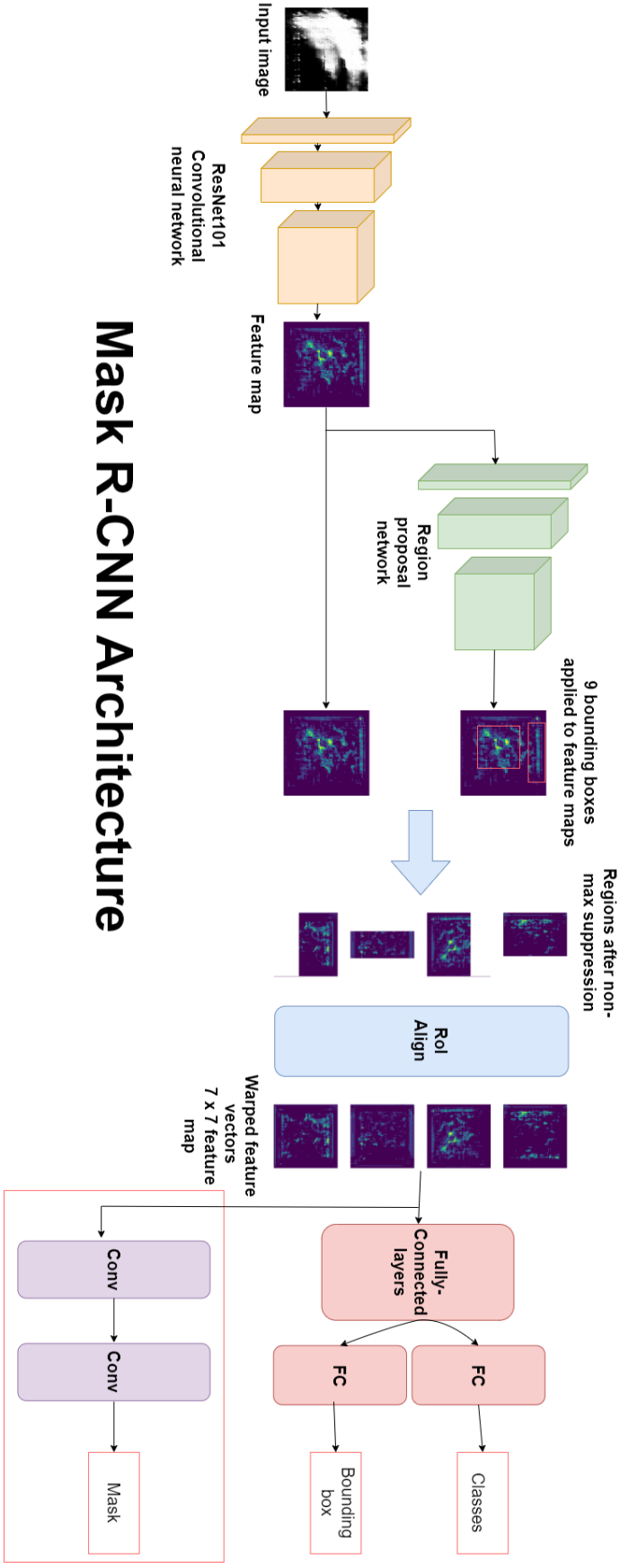


Fig. 6.2 The Mask R-CNN architecture is constructed using the ResNet 101 CNN. The model can be divided into two parts where the RPN proposes candidate object bounding boxes and then the binary mask classifier generates a mask for every class, in this case, Type II SRBs.

The architecture used in this research was built upon a Matterports implementation of Mask R-CNN (Abdulla 2017). However, a few adjustments had to be made to the network as the simulated data generated by the GAN were in greyscale and the Mask R-CNN network expects RGB image input. To do this, the *ImageChannelCount* and *MeanPixel* was changed from 3 (RGB) to 1 (greyscale) and then the shape of the image was altered when inputted. Once the shape of the input was altered, the weights loaded into the first convolutional layer were excluded. This allowed the use of pre-trained weights for all convolutional layers except the first, which was initialised to random inputs. Therefore, transfer learning was used in the Type II detection problem and training a model from scratch was not necessary.

6.4 Training and Validation

The Mask R-CNN training set uses a GAN-generated data system containing 5,000 Type II simulated GAN images. Random intensity, drift rate, inhomogeneity, start-end frequency and start-end time are some of the instance features provided by the data set. Five hundred Type II images produced by the GAN were included in the validation set. The training and validation sets were both manually labelled using an application called VGG Annotator (Dutta and Zisserman 2019), an annotating tool that labels objects in polygon JSON format within an image, see Figure 6.3. Although manual labelling was a tedious task it provided Mask R-CNN with precise polygon coordinates on what to train within the specified image.

The training and validation set were then fed into Mask R-CNN after the model configuration prerequisites was satisfied. This study was conducted on a system with an Intel Core i7-10700K processor, 32GB of RAM, and an Nvidia Geforce RTX 2080 Super GPU running Ubuntu 20.4.2 LTS. Eighty percent of the GPU's capability was used for 1,000 epochs of training using an 8-batch configuration. The training

of the Mask R-CNN was then assessed by plotting the training and validation loss to determine if the model was over or under-fitted.

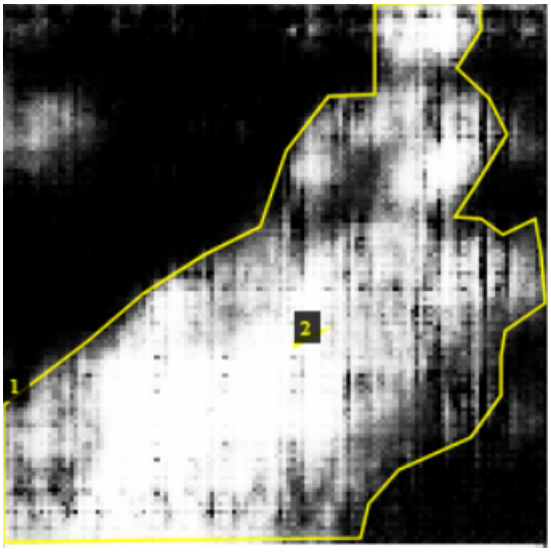


Fig. 6.3 An example of the polygon annotation tool VGG Annotator applied to a Type II SRB.

6.5 Test set

To test how robust the model was, a test set that had a variety of Type II examples was needed. One major constraint, was, as in Section 4.3.2, creating the test set as there were only 9 observed Type II examples in the I-LOFAR archive, totalling about 298 images when divided into 10-minute segments. To increase the size of the test set, 300 simulated examples of Type II were generated using the GAN to test Mask R-CNNs robustness. The real observation images were then cut into a range between 10-30 minute chunks, as they can take different shapes at different time segments. By combining real and simulated Type II data, a test set of 598 images was created. Then the ground truth polygon boxes were precisely annotated using VGG Annotator. Once a Type II was labelled, its corresponding bounding box coordinates were stored in a JSON file that was used to compare the ground truth coordinates to the model's predicted coordinates.

6.6 Results

When evaluating the Mask R-CNN model, the mAP described in Section 5.8 was again used. Using this metric, the ground truth test set values were compared to the predicted test set values made by the Mask R-CNN model. The TPs and FPs were determined by the IoU score, which was set to 0.5. Using this configuration, a mAP score of 78.90% was obtained on the test set.

In order to evaluate Mask R-CNN visually, a combination of GAN simulated and real Type II observations made by I-LOFAR on several observation dates were used. The time parameter was changed from 10-minute segments used for the training set to 30-minute segments so that the Type II's whole structure would be captured in a single image. The detections and instance segmentations of the Mask R-CNN are shown in Figure 6.4.

The feature activation's from the RPN, the offesets from the anchors, the bounding box distribution, and the segmented mask applied to the input image is illustrated in Figure 6.5. The RPN analyses feature maps to generate potential object regions, producing region proposals and corresponding objectness scores. These proposals are refined by adjusting anchor boxes using anchor offsets, which account for translations and scales. The bounding box distribution further enhances accuracy by adjusting the proposed boxes to better align with the ground truth bounding boxes. Finally, Mask R-CNN generates segmented masks by assigning class labels to each proposal and predicting pixel-level masks for object segmentation within the proposed regions. By combining these components, Mask R-CNN effectively detects objects, refines bounding box predictions, and produces high-quality segmentation masks for accurate object delineation.

6.7 Summary

The investigation in Chapter 6 utilised a dataset of simulated Type II images generated using a GAN. Since there was a lack of Type II data during solar minimum, it was more challenging to train the GAN to produce Type II images. The training set used in this investigation contained nine Type II SRBs, which were divided into 10-minute segments, and 298 training images were generated from them. To enlarge the training set, techniques such as blurring, cropping, and removing features were used to produce a total of 1,835 images, including the presence of RFI.

Using this training set, the results demonstrated that a combination of deep learning models was an effective method for detecting, classifying and segmenting Type II SRBs. In particular, the Mask R-CNN configuration achieved a mAP accuracy of 78.90% on Type II data, consisting of 500 Type II examples. By extending the SRB detection analysis into both detection and object segmentation, the shape and intensity of the Type II bursts were captured.

However, due to the Type II's erratic shape it is often susceptible to large amounts of RFI. This was not only evident from this investigation but also from the results in Chapters 3 and 5. Chapter 7 explores a new way of mitigating the effects of RFI using a modified GAN architecture.

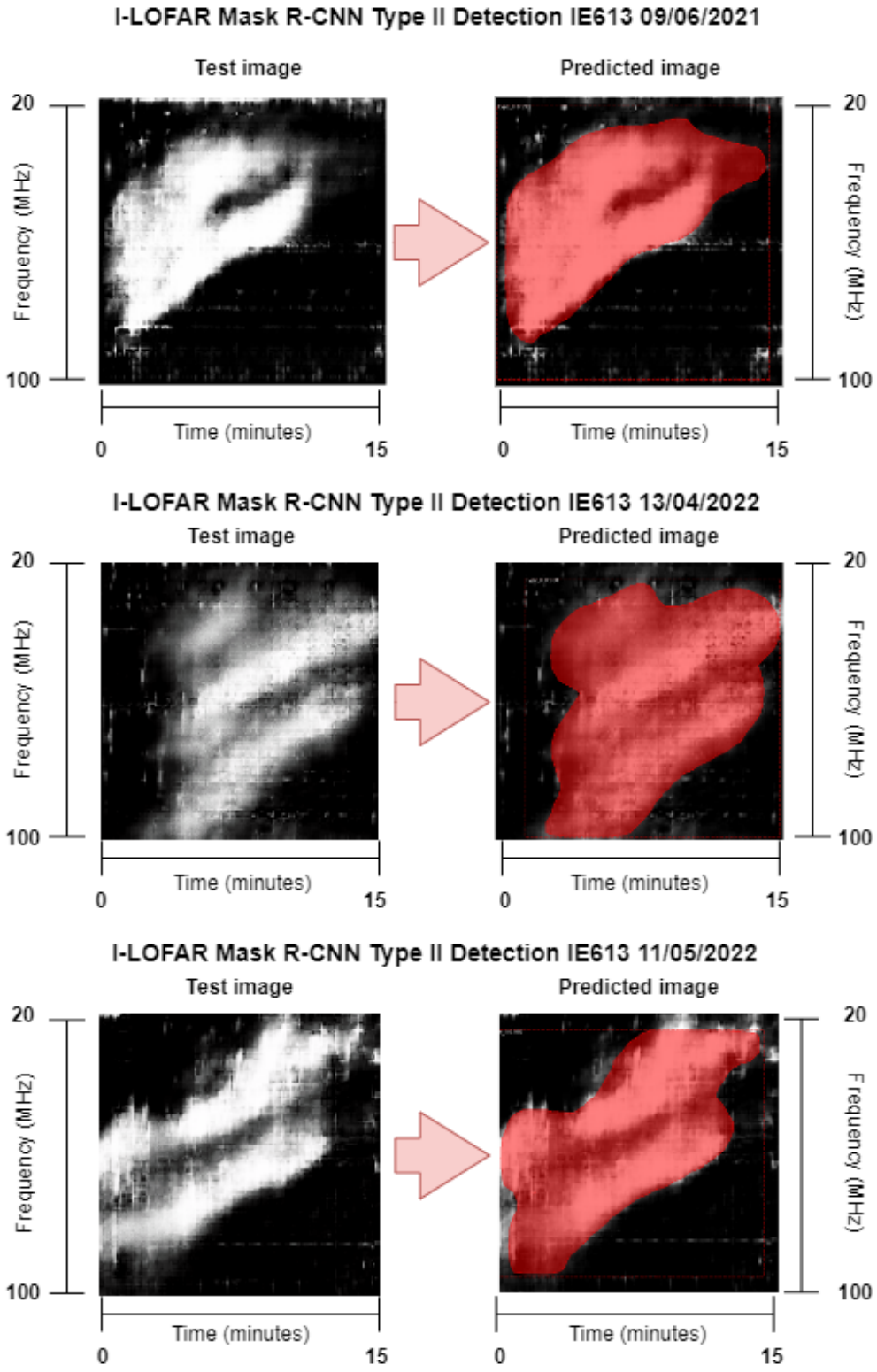


Fig. 6.4 Mask R-CNN applied to different Type II observations made by I-LOFAR. Each Type II is is plotted at a time range of 0-15 minutes and at a frequency range of 20-100MHz

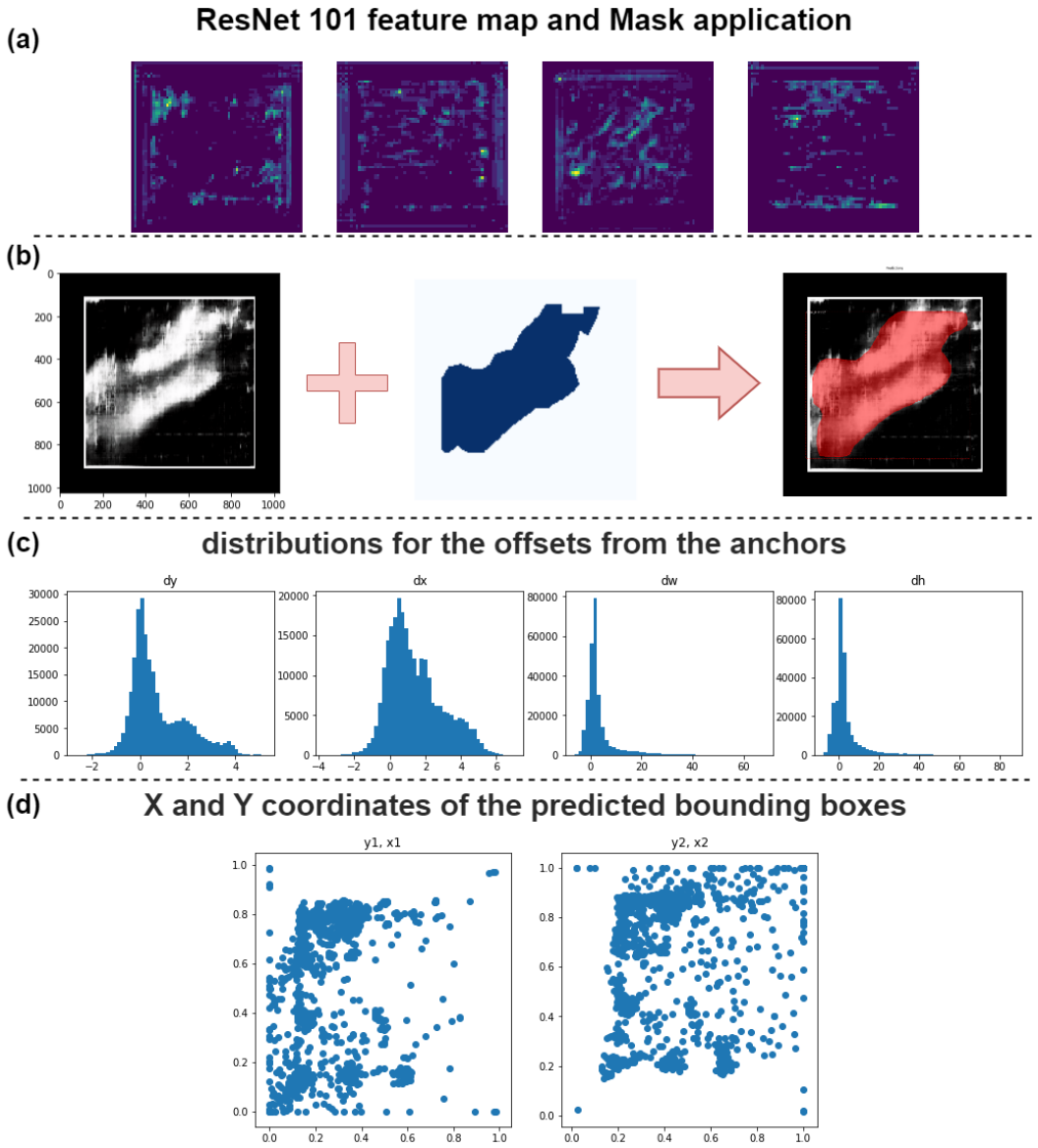


Fig. 6.5 The Mask R-CNN process visualised on a Type II example. The resulting feature maps from the RPN are illustrated in (a). The segmented mask is applied to the Type II along with the detection in (b). The offsets from the anchors (c) and bounding box distribution (d) represent the difference between the coordinates of a bounding box anchor and the coordinates of the corresponding object instance.

Chapter 7

Mitigation of Radio Frequency Interference in Solar Radio Observations using Generative Adversarial Networks

7.1 Introduction

Recall from chapter 1:

Research Question 4: How effectively can deep learning techniques be applied to mitigate the effects of radio frequency interference in solar observations?

Radio astronomy is an essential instrument for astronomers in their research of the cosmos, and it has recently been the focus of a resurgence in both interest and technological development. When compared to previous, more conventional radio telescopes, next-generation radio telescope arrays like LOFAR are being developed to have a substantially higher sensitivity to low radio frequencies. However, this increased sensitivity makes them more vulnerable to RFI, which can potentially

corrupt or distort the signal received by radio telescopes. RFI occurs when an electronic piece of equipment emits radio waves, disrupting the operation of another nearby electronic device, in this case, the radio inferometer LOFAR. Being a low-frequency radio interferometer, LOFAR is particularly susceptible to RFI, which often comprimised the AI algorithms' detections. During the training of algorithms, efforts were made to ignore this interference. Although this approach was effective in some instances, the presence of interferences in solar observations continued to impede the detection of YOLO, as illustrated in Figure 7.1. This emphasises the importance of RFI mitigation techniques that are both efficient and effective in the field of radio astronomy. An approach for mitigating RFI is presented, which is based on DL, that identifies and removes RFI within solar RFI-corrupted spectrogram observations in an unsupervised manner, utilising a modified RFI Removal GAN (RFIRGAN) framework.

7.2 Dataset

In order to train a usable model with RFIRGANs, two datasets were needed: one with RFI and one without RFI. To accomplish this, multiple observation days within the I-LOFAR archive were reviewed, and images containing embedded RFI were selected. In total, the RFI dataset consisted of 3,760 images. The band of RFI typically seen in the frequency range between 0-20 MHz was also included in the dataset, see Figure 7.2. The reasoning for including the 0-20MHz frequency band was to see if the RFIRGAN model could pick up and remove the most noticeable RFI within the image. To create the dataset without RFI, the same dataset that contained RFI was used. Data augmentation methods, such as cropping and blurring, were employed to remove both the embedded and frequency band RFI in the RFI dataset. Figure 7.3 shows this process for a particular image in the dataset. Once complete, two datasets were created that were used to train the RFIRGAN model.

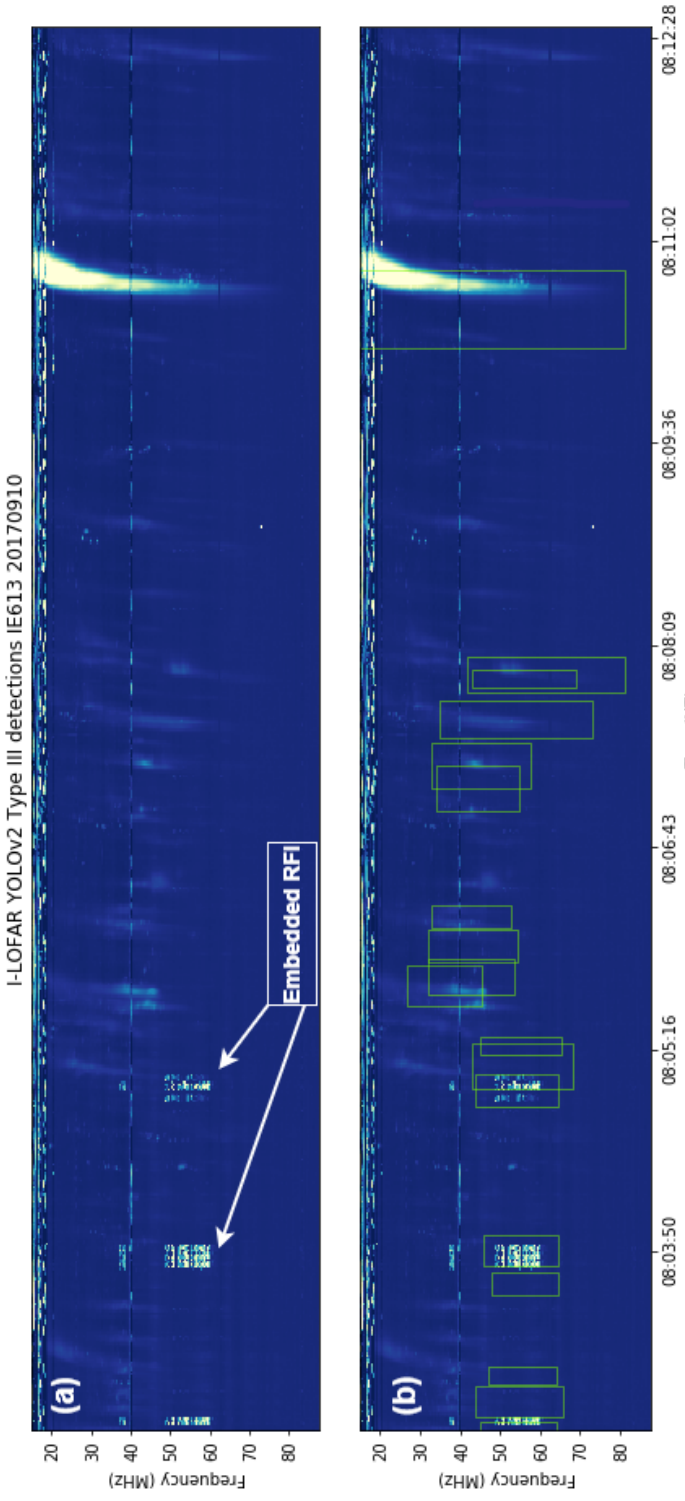


Fig. 7.1 An example of RFI hindering the detections of YOLO. Plot (a) illustrates some embedded RFI within a 10 minute segment during an observation on the 10th of september 2017. The YOLO model's detections have been compromised in plot (b) as the model assumes the embedded RFI is a Type III SRB. This was often the case when a presence of embedded RFI was in an observation.

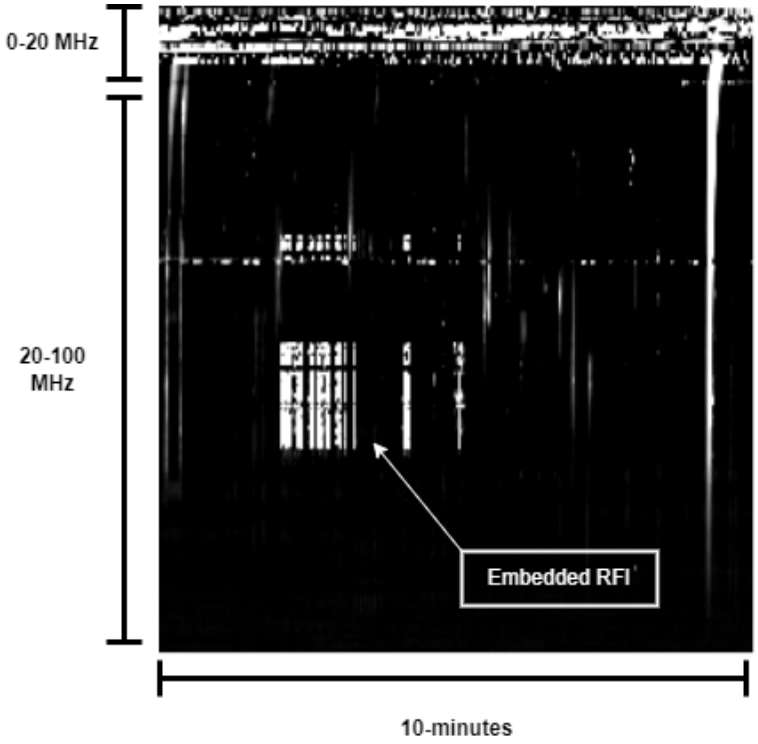


Fig. 7.2 The two types of RFI seen in an I-LOFAR observation. The RFI is located in the 0-20 MHz band and the embedded RFI is located in the 20-100 MHz band.

7.3 Radio Frequency Interference Removal Generative adversarial Network (RFIR-GAN)

Once the dataset was created, the model needed to be configured. The RFIRGAN model was constructed using the same architecture used to remove eyeglasses from people's faces, known as Eyeglasses Removal GAN (ERGAN) (Hu et al. 2021), see Figure 7.4. The ERGAN shares some similiar concepts to the cycle GAN (Zhu et al. 2017), coupled GAN (Liu et al. 2017) and MUNIT GAN (Huang et al. 2018) in terms of its architecture, however, its approach in terms of object masking has many differences. The proposed ERGAN focuses on manipulating the eye region by introducing an eye-area loss and an invertible eye generation mechanism. It adopts an instance generation mechanism that can swap the eye area of two facial images, generating more diverse images compared to Cycle GAN-based methods. The face

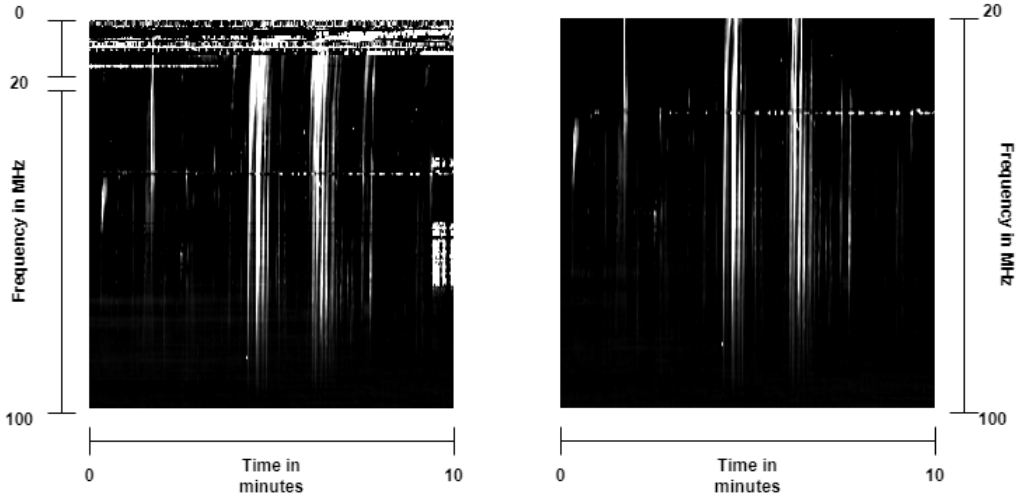


Fig. 7.3 The two datasets used to train RFIR-GAN before (a) and after (b) data augmentation was applied.

images are viewed as one domain with two codes: face appearance and eye-area attribute. The ERGAN network needed to be slightly altered to mitigate the effects of RFI.

Being a GAN, the ERGAN framework shares some concepts described in Chapter 4, the obvious ones being the generator and discriminator networks. The generator of the ERGAN, however, is based on an autoencoder framework and consists of face appearance encoders (E_X^f, E_Y^f), eye-area attribute encoders (E_X^e, E_Y^e) and decoders (G_X, G_Y). The encoders encode a given face image into the face appearance code ($f.$) and eye-area attribute code ($e.$), respectively. The codes are combined and used to generate a new image with the decoder G. The decoder G is a deterministic function and has inverse encoders. G_X generates face images without glasses and G_Y generates face images with glasses. For RFIRGAN, one alteration is made to the generator network. The mask size area of the eye attribute encoders is increased to the size of the entire image to try and incorporate both types of RFI , see Figure 7.5.

RFIRGAN has two discriminators (D_X, D_Y) for images X (without RFI) and Y (with RFI), respectively. The discriminators aim to distinguish between generated

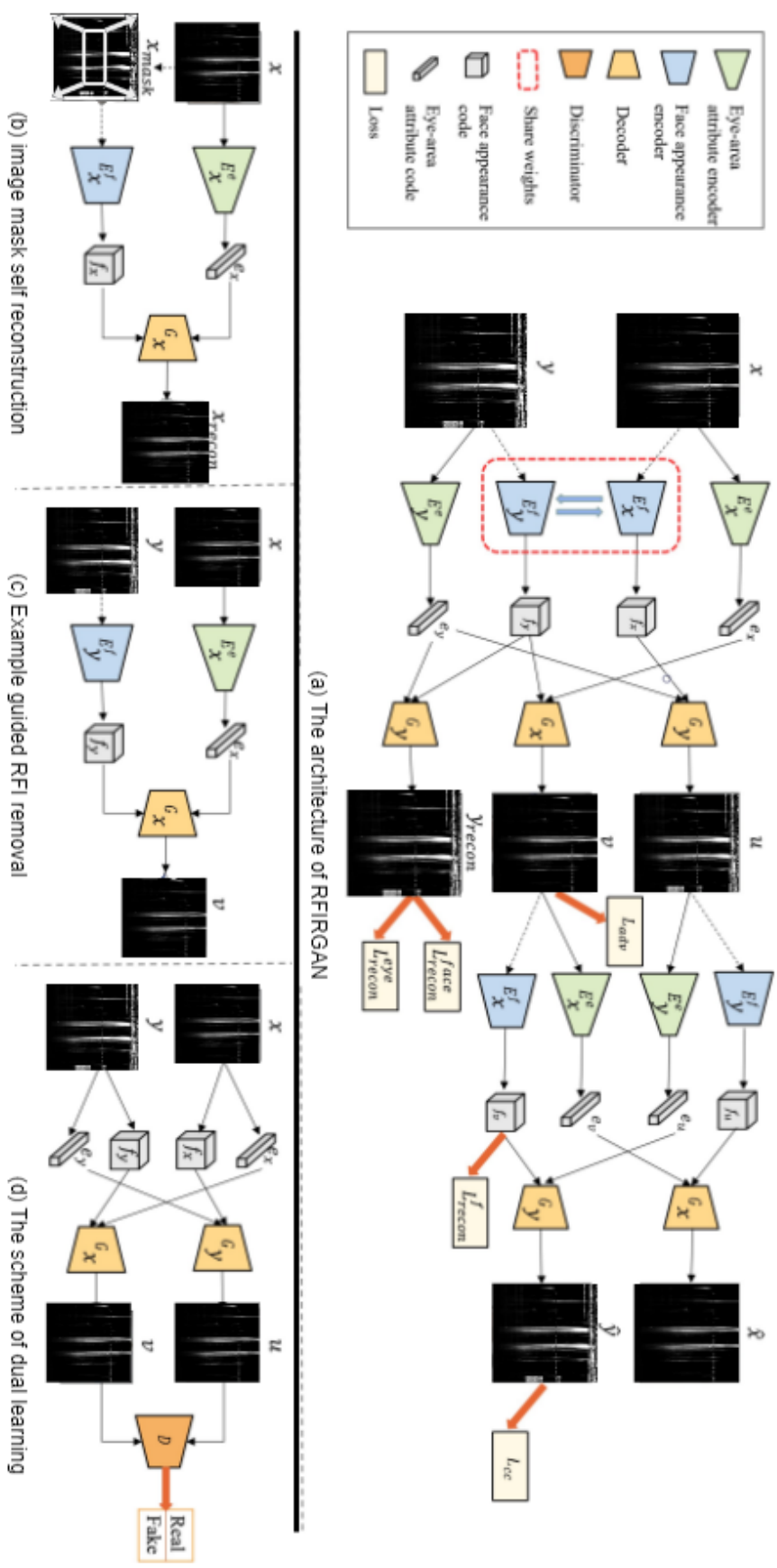


Fig. 7.4 An overview of Eyeglasses Removal Generative Adversarial Network (ERGAN) applied to SRBs. The model is altered by increasing the mask size to the full scale of the image input (modified from ERGAN image in Hu et al. 2021).

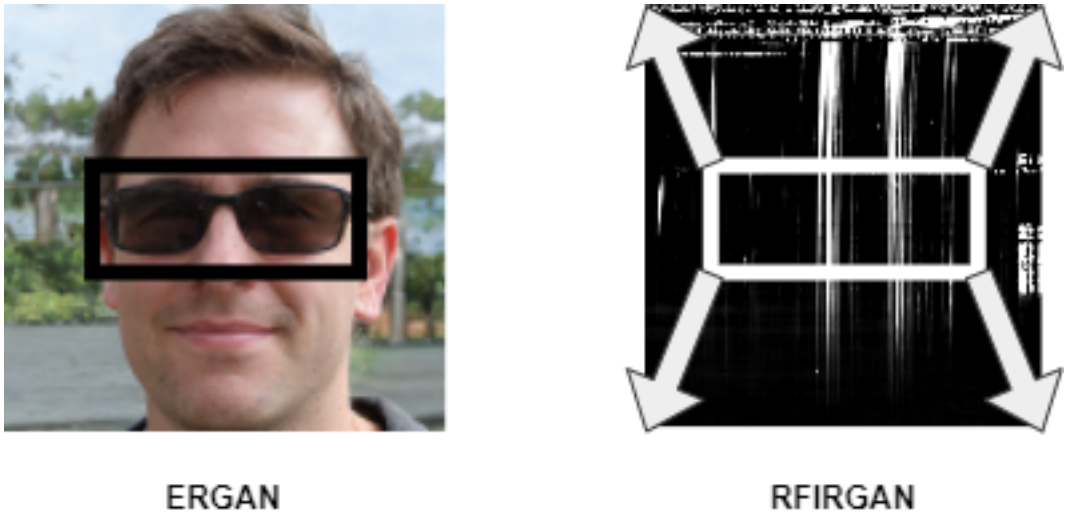


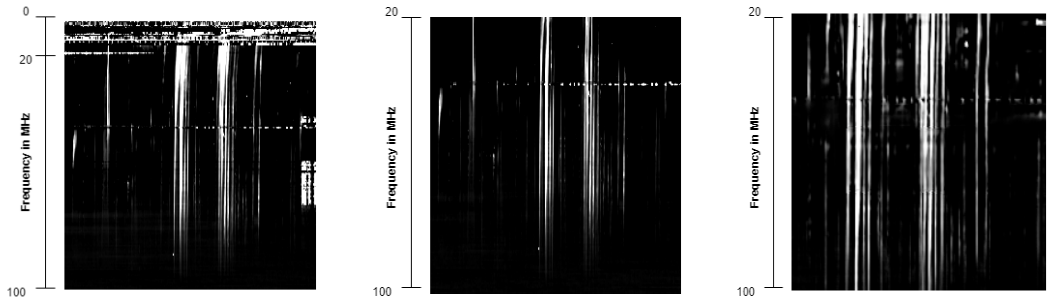
Fig. 7.5 The size of the masking area is increased to match the image dimensions to try to mitigate the effects of RFI. ERGAN image taken from MeGlass (Guo et al. 2018)

and real images. For example, D_X distinguishes between images generated by decoder G_X and real images in X, while D_Y distinguishes between images generated by decoder G_Y and real images in Y. The images generated by decoder G_X have the same without RFI appearance code as image x and the same with RFI attribute code as image y. On the other hand, the images generated by decoder G_Y have the same without RFI appearance code as image Y and the same with RFI attribute code as image X.

Once the model was trained using this configuration and dataset, the model was evaluated using the same methods described in Chapter 4.

7.4 Model evaluation using human perception

As mentioned in Section 4.6.1, manual assessment or human interpretation of images produced by a generator model is the common way to evaluate a GAN (Borji 2019). The Generator was used to create a batch of fake images, then an evaluation of the quality and diversity of the images in relation to the target domain, in this case, the removal of RFI, was performed. To accomplish this, RFIRGAN was trained over a



(a) with RFI **(b)** without RFI **(c)** RFIRGAN

Fig. 7.6 The comparison of the training sets (a) with RFI in the range of 0-100 MHz, (b) without RFI and (c) mitigated RFI from RFIRGAN plotted in the frequency range of 20 and 100 MHz.

number of epochs. Once a training instance was complete, the generated images with and without RFI were compared to a number of images in both training sets. When visually comparing the training sets' images, with and without RFI, to the images produced by RFIRGAN, see Figure 7.6, it is noteworthy that the RFIR-generated images almost completely removes the RFI seen in the 0-20MHz frequency band leaving only some remnants of this type of RFI just above this band in the 20-25MHz range where some Type III SRBs start in observed data (possible distortion by the Type III). The RFIRGAN removes most of the large examples of embedded RFI seen in Figure 7.6. However, it fails in some cases to remove small elements of embedded RFI most notably around SRBs which could illustrate a possible distortion in RFI removal when the embedded RFI is located near an SRB.

7.5 Model evaluation using FID

The task became somewhat more complex when assigning an appropriate metric value to RFIRGAN's performance. Usually, when evaluating a GAN, the generator network's newly produced images are compared to the discriminator's training set using the FID score described in Section 4.6.2. In the case of RFIRGAN, the generator isn't producing new similar images to the training set but images exactly

the same as the discriminator's training set without any identifiable RFI feature. The approach seen in (Hu et al. 2021) was used, where FID calculates the percentage increase in image quality compared to the discriminator's training set. To do this, the FID score of the RFI and no-RFI training set is compared with the generated RFIRGAN images. The FID scores of both RFI and no-RFI datasets were compared to removed RFI-generated images by RFIRGAN, as seen in Table 7.1. Using this method, a 31.83% increase in image quality or FID relative to the dataset with RFI was achieved using RFIRGAN.

Table 7.1 A comparison of FID scores between the RFIR method compared to the dataset with RFI and without RFI. Again, the objective is to get a FID score as close to 0 as possible indicating a good similarity between simulated and real data.

Dataset	Comparison	FID Score
No RFI dataset	RFIR	16.00
No RFI dataset	With RFI dataset	23.47
With RFI dataset	RFIR	41.92

7.6 Summary

The GAN described in Chapter 5 was modified to of mitigating the effects of RFI in solar observations by I-LOFAR. By using real observed data, 3,760 images were collected from the I-LOFAR archive and augmented to isolate two types of RFI; the RFI seen at the 0-20 MHz range and the embedded RFI seen in the 25-100 MHz range. The two resulting training sets consisted of one containing RFI and one without RFI. These two training sets were then fed into RFIRGAN for training.

The ERGAN's framework was altered by increasing the mask area of the glasses to the full scale of the image to capture the two types of RFI. By doing so, the new RFIRGAN learns to identify and mitigate the effects of RFI presented in solar observations. The resulting FID score illustrated a 31.82% increase in image quality compared to images containing RFI.

This RFIRGAN-based approach for RFI mitigation has laid a solid foundation for future endeavors in this field. By harnessing the power of Generative Adversarial Networks, this method has showcased its potential in effectively suppressing RFI artifacts and restoring signal integrity. The success of this approach opens the door to exciting possibilities for further refinement and innovation. Future work could explore fine-tuning the GAN architecture to enhance its performance in complex RFI scenarios, optimising training strategies to achieve even better noise reduction, and investigating the adaptability of the method to various radio communication environments.

Chapter 8

Conclusions

8.1 Synopsis

The automatic detection and classification of SRBs is a major challenge in solar physics due to their heterogeneous form. Near real-time detection and classification of SRBs has become a necessity in recent years due to large data rates generated by advanced radio telescopes such as the Low-Frequency Array (LOFAR). The research described in (Carley et al. 2020c) illustrated the potential of AI classification techniques for SRB characterisation. But could this research be taken a step further and not only answer the what but the where in observations in terms of frequency and time? Addressing this question was the main objective of this research.

8.2 Conclusions

It is now possible to answer all of the research questions raised in Chapter 1.

Sub Research Question 1: How effectively can deep learning methods be employed to classify and locally detect Type III solar radio bursts?

In the experiments described in Chapters 3 and 5, the YOLOv2 results showed that the model could not only accurately detect and classify Type III SRBs but also

do so in real time. In Chapter 3, the YOLOv2 model was trained using a simulated dataset of Type III SRBs embedded in RFI channels, created with parametric modeling techniques. The model efficiently detected and accurately classified Type III SRBs in real-time and achieved an F1 accuracy of 82.63%. Then, in Chapter 5, a combination of congruent deep learning models was used to generate, detect, and classify Type III bursts in real-time. By combining GAN-generated and real observed data, a training set was created to evaluate the YOLOv2 model. The model achieved a 77.71% mAP accuracy in detecting Type III bursts.

Sub Research Question 2: Is it possible for deep learning methods to generate simulations of Type II and Type III solar radio bursts that exhibit comparability with actual observed data?

Chapter 4 investigated the possibility of GAN as a means of SRB simulation. To create realistic models for classification and detection algorithms, a GAN was employed using real observations from the LOFAR archive. The GAN generated simulated Type III and Type II SRBs, which were then evaluated using human perception and the FID score. Comparing GAN-simulated Type IIs and Type IIIs to real observed Type IIs and Type IIIs, the achieved FID scores are 34.01 and 9.23, respectively. The results demonstrated the effectiveness of GANs in producing realistic images for accurate classification and detection algorithms.

Sub Research Question 3: How effective are deep learning methods in detecting and segmenting Type II solar radio bursts?

Chapter 6 showcased the effectiveness of combining deep learning models for the detection, classification, and segmentation of Type II SRBs. Specifically, the Mask R-CNN configuration achieved a mAP accuracy of 78.90% when applied to a dataset of 500 Type II examples. By extending the analysis beyond detection to include object segmentation, the models successfully captured the shape and intensity of the Type II bursts.

Sub Research Question 4: How effectively can deep learning techniques be applied to mitigate the effects of radio frequency interference in solar observations?

When working with both detection and simulation methods there was always one phenomena ever present, RFI. In Chapter 7, a GAN architecture (ERGAN) was modified for the purpose of RFI mitigation. By increasing the size of the mask area of the ERGAN and giving the model SRB data as input, a new model was created known as RFIRGAN. Using this model, the image quality was increased by 31.83% when compared to augmented data without RFI and in turn, opened a potential avenue for future developments.

Based on the literature review undertaken, this is the first time both object detection and generative models have been applied to SRB data. The automatic detection and classification of SRBs is a challenge in solar radio physics due to their heterogeneous form. For this reason, AI and particularly DL based systems are and will be a crucial processing pipeline for current and future radio telescopes.

8.3 Future development

This section proposes future avenues of research inspired by the work presented in this thesis.

8.3.1 Evaluation metrics

In this research, we introduced the computer vision metric mAP, which serves as a valuable tool for quantifying the model's ability to accurately detect local instances of Type III SRBs. The mAP metric, while widely accepted and utilised in various computer vision domains, may encounter limitations when applied to the evaluation of abstract and complex phenomena such as SRBs. The methodology

we adopted for calculating mAP, rooted in the principles of computer vision, aligns well with the prevailing standards observed in COCO dataset competitions. These competitions typically employ an IoU threshold of 0.5 to determine the adequacy of object detection.

However, it is important to acknowledge that the intricate nature of Type III SRBs and their visual manifestations could lead to instances where a more nuanced assessment is required. The IoU threshold of 0.5, although providing a robust benchmark for accurate model detection, might inadvertently lead to the exclusion of pertinent detection data. Consider the scenario where a Type III SRB detection is visually evident, but the bounding box slightly overlaps the actual occurrence, resulting in an IoU value just below the threshold. Regrettably, such instances could be erroneously classified as false detections, potentially undermining the comprehensive evaluation of the model's performance.

In light of these considerations, a compelling argument arises for the need to develop a novel metric system specifically tailored to the unique characteristics of Type III SRBs. This new metric should encapsulate the inherent complexities of abstract phenomena, enabling a more refined assessment of detection accuracy. Striking the right balance between the existing rigor of computer vision standards and the intricacies of SRB detection will be crucial in fostering a comprehensive and holistic evaluation framework. By embracing this nuanced approach, we aspire to achieve a model that not only maintains robustness and accuracy in detection but also attains a heightened sensitivity to the subtle variations and overlaps that characterise Type III SRBs. In doing so, we can lay the foundation for a metric system that is better aligned with the intricacies of these abstract phenomena and thus contribute to advancing our understanding of solar activity observations.

8.3.2 YOLO

In Chapters 3, 5, and 6, object detection was applied to the task of automatic SRB detection and classification. Using training sets and simulation methods, an accuracy of between 77-83% at detecting and classifying Type II and Type III SRBs was obtained. However, although each object detection method was successful to a point, each method had its limitations. The YOLOv2 model, proved to be very accurate and provided real-time capabilities, yet it was still limited by the shape of its bounding box predictions, as the true shape of the Type III bursts couldn't be extracted. The Mask R-CNN solved this problem via segmentation on Type IIs but the model itself lacked the capability of real-time detection.

Since the development of these models during this research, YOLO has iterated through a number of releases to the point where the capabilities of YOLOv2 and Mask R-CNN have been combined. The development of YOLOv7 (Wang et al. 2022) and YOLOv8 (Ultralytics 2022) have introduced new detection methods, including semantic and instance segmentation from polygon labelled inputs. By using these new YOLO methods, the shape from all SRB classes, not only Type II and Type IIIs, could be extracted and done so in near real-time with greater accuracy.

8.3.3 Super resolution GANs

In Chapter 4, the GAN was introduced for SRB simulation with a view to improving the detection algorithms accuracy and robustness. Using the GAN, a FID score of 9.23 for generating Type IIIs and 34.01 for generating Type IIs were obtained when compared to real observed data. This increase in image quality provided the detection algorithms with the accuracy and robustness they needed to account for the various forms each SRB may take. Although GAN proved to be important for SRB simulation and detection techniques, the images were still very small at 256 x 256.

(Ledig et al. 2016) has shown significant progress at taking small low quality images and upscaling them by a factor of four size while maintaining the image quality. By using the original GAN architecture as an input to the Super Resolution GAN (SRGAN) model, the SRB simulated images could be upscaled with more features for the detection models to train on. This could potentially add more accuracy and robustness to the detection models when tested on real observed data.

8.3.4 PointRend

In Chapter 6, Mask R-CNN was used for Type II detection, classification and segmentation. The GAN described in Chapter 4 was used to simulate and produce a large collection of Type II examples that was employed as a training set for Mask R-CNN. As a result, Type II SRBs could be detected with a mAP of 78.90%. However, as seen in Figure 8.1, the segmentation of the Type II suddenly stops when it's close to the boundary of the image. Furthermore, the segmentation itself appears very smooth over sharp edges of the Type II. One way in which the segmentation

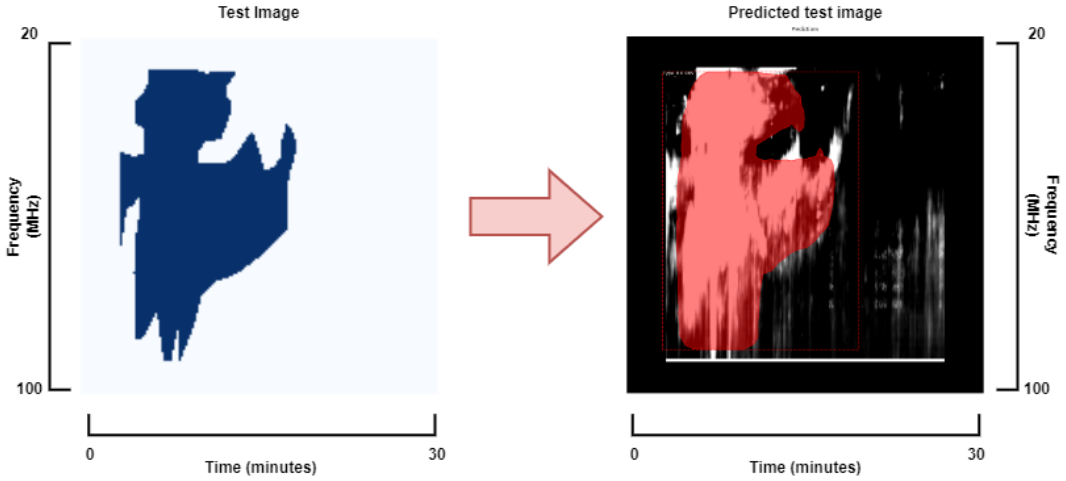


Fig. 8.1 An example case of the smooth segmentation being applied to a real Type II SRB. Notice how the segmentation doesn't fully encompass the Type II's sharp edges and borders. The application PointRend could be a solution to this.

could be improved is by applying PointRend to the Mask R-CNN model (Kirillov

et al. 2020). PointRend is a point-based rendering neural network module that can be added as an enhancement to the existing model. PointRend uses an iterative subdivision algorithm to upscale the predictions by predicting labels of points at selected locations by employing a small trained neural network. This method enables high-resolution output in an efficient way. With the addition of PointRend it could be possible to get an even more accurate segmentation of a Type II SRB.

8.3.5 RFIRGAN

Mitigating the effects of RFI in radio astronomy is a major challenge. In Chapter 7, the ERGAN framework was modified by expanding the size of the region of interest to encompass the entire image. In this way, the new RFIRGAN is trained to recognise the impacts of RFI that are present in solar observations and to find ways to lessen those effects. When compared to images that contain RFI, the resulting FID score demonstrates an improvement in image quality of 31.83%. Key to attaining significant improvement with the RFIRGAN model is the quality of the training dataset, which is made up entirely of real observed SRB data. However, although there was a significant amount of data within the training set, there was one aspect of the dataset that hindered the performance of the RFIRGAN. Within the training set, each image contained RFI in the 0-20 MHz range, so, there was an abundance of this RFI for training. This was not the case when it came to examples of embedded RFI at higher frequencies; Only a couple of hundred examples of this form of RFI could be sourced, well below the optimum number to train the RFIRGAN network.

For future work with this model, two options are presented to try and improve the model's training set. The first option, is to gather more observational data and search for more embedded RFI. Although quite tedious, data could be gathered within the I-LOFAR archive that contains embedded RFI to train the model again and further mitigate the effects of RFI within a given observation. The second option, is to

simulate embedded RFI. By employing the GAN described in Chapter 4, the ability to simulate and generate a collection of simulated images containing embedded RFI was achieved. These images seamlessly integrated into the existing training set for RFIRGAN, all accomplished within a short period of time. This method could train RFIRGAN with large quantities of both RFI examples to recognise and remove all forms of RFI within observational data.

Glossary

Artificial Intelligence The use by machines of techniques that mimic intelligent behaviour in humans.. 2, 3, 5, 7–10, 30, 33, 90, 99, 101

Artificial neural network is a type of machine learning algorithm that is designed to simulate the structure and function of biological neurons in the human brain.. 9

Computer Vision is an AI that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs.. 5, 10, 13, 15, 37, 144

Convolutional Neural Network is a type of DL architecture commonly used for image classification and recognition tasks.. xv, xviii, 2, 4–6, 10–15, 25–29, 31, 37, 39, 40, 42, 53, 62, 81, 82, 140, 141, 144–146

Coronal Mass Ejection is a massive burst of solar wind and magnetic fields that are released from the Sun's corona.. xiv, xviii, 19, 27, 79, 80

Deep Learning a type of machine learning based on artificial neural networks in which multiple layers of processing are used to extract progressively higher level features from data.. xiv, 1–3, 5–7, 9, 10, 13, 14, 26, 31, 33, 39, 90, 101, 139, 141, 145

Fast Radio Burst is a brief, bright, and highly energetic radio pulse that lasts only a few milliseconds. FRBs are detected as radio emissions from distant galaxies, and they are one of the most mysterious phenomena in astrophysics.. 5, 28, 29

Frechet Inception Distance is a metric used to evaluate the quality of generated images compared to a set of real images.. xvi, xxi, 58, 61–63, 96, 97, 103, 105

Generative Adversarial Network are a type of generative deep learning model that consists of two neural networks: a generator and a discriminator.. xv, xvi, 2–6, 15, 16, 18, 26, 50–52, 54–61, 63, 64, 66–69, 79, 80, 83, 92, 93, 95–97, 100, 101, 103, 104

InceptionV3 is a convolutional neural network (CNN) architecture that was developed by Google for image recognition and classification tasks.. 25, 62, 141, 144, 145

Intersection over Union is a metric used to measure how well the predicted bounding box of an object aligns with the ground truth bounding box.. xvii, xxi, 45, 72–75, 81, 85

LOFAR for Space Weather is software upgrade for LOFAR. It aims to improve our understanding of the Sun-Earth connection and the impact of space weather on our planet.. 5, 32, 33

Machine Learning the use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data.. 1, 5, 7–10, 23, 25, 26, 28, 33, 139–142

Mask R-CNN is a deep learning model for object detection and instance segmentation, which extends the popular Faster R-CNN object detection model by adding a mask prediction branch to also identify the exact shape of objects within an image.. xviii, xix, 13, 18, 28, 38, 79, 81–85, 87, 88, 103, 104

mean Average Precision is a performance metric used to evaluate the performance of object detection systems.. xxi, 6, 65, 72, 73, 75, 85, 104

object detection is a CV technique for locating instances of objects in images or videos.. 2, 5, 10, 13, 14, 28, 37–39, 56, 101, 103

Radio Frequency Interference is an unwanted electromagnetic energy that interferes with the proper functioning of a radio receiver or other electronic devices that rely on wireless communication.. xvii–xix, xxi, 3, 5, 6, 20, 21, 23, 29–32, 38, 55, 60, 66, 76, 77, 89, 90, 92, 93, 95–97, 101, 105, 106, 140, 141, 145

Radio Frequency Interference Removal Generative Adversarial Network is modified eyeglasses removal GANs network for the purpose of RFI mitigation.. xix, 6, 90, 92, 93, 95–97, 101, 105, 106

Random Forest is a supervised machine learning algorithm that is used for classification, regression, and other tasks. It is an ensemble method that combines multiple decision trees to create a more robust and accurate model.. xxii, 25, 26, 140, 142, 143

REALTA is a computational backend located at the Irish LOFAR station. xiii, xiv, 5, 17, 33, 34

Rectified Linear Unit is an activation function commonly used in artificial neural networks. It is a simple and efficient activation function that is popular because of its ability to speed up training of deep neural networks.. 42, 70

Region of Interest is a rectangular region of an input image that is likely to contain an object of interest. RoIs are identified by a region proposal network (RPN), which proposes a set of candidate RoIs based on the features extracted from the input image.. 81

Scalable Link Interface is a technology developed by NVIDIA that allows multiple graphics processing units (GPUs) to be connected and used in parallel to improve graphics performance.. 42, 56

Search for Extraterrestrial Intelligence is a scientific effort to detect signs of intelligent life beyond Earth. The primary goal of SETI is to search for radio signals or other forms of electromagnetic radiation that may be transmitted by other civilizations in the universe.. 28, 29

Solar Radio burst is the solar radio emission associated with the sun's weather.. xv–xviii, xx, xxi, 1–3, 5, 6, 12, 15, 16, 18–21, 23, 25–28, 32, 33, 38, 42, 45, 46, 50, 55–69, 72, 74, 79, 80, 82, 96, 99–101, 103–105, 139, 142, 144–147

Support Vector Machine is a type of supervised machine learning algorithm that can be used for classification or regression analysis.. xxii, 13, 25, 26, 140–143

the LOw Frequency ARray the largest radio telescope operating at the lowest frequencies that can be observed from Earth.. xiii, xiv, xvi, xviii–xx, 1, 3–5, 16–18, 26, 32–34, 43, 55, 57, 60, 67, 68, 70, 71, 77, 80, 84, 85, 87, 89, 90, 92, 97, 99, 145, 146

Type II is a type of burst of electromagnetic radiation that originates from the Sun. They are characterized by a frequency drift from high to low frequencies, but at a slower rate than Type III bursts. The frequency drift is often accompanied by a "shock" signature in the radio emission, which suggests that the burst is associated with a shock wave propagating through the solar atmosphere.. xiv–xvi, xviii, xx, 2–6, 16, 19, 20, 22, 24–28, 50, 57–61, 63, 79, 80, 82–85, 87, 103–105, 140, 145, 147

Type III is a type of burst of electromagnetic radiation that originates from the Sun. They are characterised by a rapid frequency sweep from high to low frequencies, typically over a few seconds or less.. xiv–xvi, xx, 2, 4–6, 16, 19–21, 23–26, 28, 38, 41–46, 50, 55–61, 63, 65–73, 96, 99, 103, 140, 143, 145–147

You Only Look Once is a real-time object detection algorithm.. xiv, xv, xvii, xviii, xxi, 2, 4–6, 13, 18, 37–42, 44–46, 49, 56, 64–77, 99, 103

Nomenclature

Acronyms / Abbreviations

AI	Artificial Intelligence
ANN	Artficial Neural Network
CME	Coronal Mass Ejection
CNN	Convolutional Neural Network
CRAF	constant-false-alarm-rate
CV	Computer Vision
DCGAN	Deep Convolutional Generative Adversarial Network
DL	Deep Learning
FID	Fréchet Inception Distance

FN False Negative

FNN Feed-Forward neural network

FP False Positive

FRB Fast Radio Burst

GAN Generative Adversarial Network

HDCGAN High-resolution Deep Convolutional Generative Adversarial Networks

HFB High Frequency Band

HFR High Frequency Receiver

HOG Histogram of Oriented Gradients

IoU Intersection over Union

LFB Low Frequency Band

LOFAR4SW LOFAR for Space Weather

LOFAR LOw Frequency ARray

mAP mean Average Precision

MDI Michelson Doppler Imager

ML Machine Learning

MNIST Hand written digit set

RCNN Region Convolutional Neural Network

REALTA REAL-time Transient Acquisition backend

ReLU Rectified Linear Unit

RFI Radio Frequency Interference

RFIRGAN Radio Frequency Interference Removal Generative Adversarial Network

RF Random Forest

RoI Region of Interest

RPN Region proposal network

RSTN Radio Solar Telescope Network

SDO Solar Dynamics Observatory

SETI Search for Extraterrestrial Intelligence

SGD Stochastic Gradient Descent

SLI Scalable Link Interface

SOHO Solar and Heliospheric Observatory

SoI Signal Of Interest

SRB Solar Radio Burst

SSD Single Shot Detectors

SVM Support Vector Machine

TP True Positive

YOLO You Only Look Once

References

Abdulla, W. (2017). Mask r-cnn for object detection and instance segmentation on keras and tensorflow.

Aeronautics, N. and Administration, S. (2017). Sdo.

Afandi, N. Z. M., Abidin, Z. Z., Umar, R., Sabri, N. H., Ibrahim, Z. A., and Monstein, C. (2016). Implementation of frequency drift for identification of solar radio burst type ii. *International Journal on Advanced Science, Engineering and Information Technology*, 6(5):775–780.

Akeret, J., Chang, C., Lucchi, A., and Refregier, A. (2017a). Radio frequency interference mitigation using deep convolutional neural networks. *Astronomy and Computing*, 18:35–39.

Akeret, J., Seehars, S., Chang, C., Monstein, C., Amara, A., and Refregier, A. (2017b). Hide and seek: End-to-end packages to simulate and process radio

survey data. *Astronomy and Computing*, 18:8–17.

Aurass, H. (1997). Coronal mass ejections and type ii radio bursts. In *Coronal Physics from Radio and Space Observations*, pages 135–160, Berlin, Heidelberg. Springer Berlin Heidelberg.

Barnbaum, C. and Bradley, R. F. (1998). A new approach to interference excision in radio astronomy: Real-time adaptive cancellation. *The Astronomical Journal*, 116(5):2598–2614.

Bilik, S. and Horak, K. (2022). Sift and surf based feature extraction for the anomaly detection. (3).

Bonnin, X., Aboudarham, J., Fuller, N., Renie, C., Perez-Suarez, D., Gallagher, P., Higgins, P., Krista, L., Csillaghy, A., and Bentley, R. (2011). Automated detection and tracking of solar and heliospheric features in the frame of the European project HELIO. In Alecian, G., Belkacem, K., Samadi, R., and Valls-Gabaud, D., editors, *SF2A-2011: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, pages 373–377.

Borji, A. (2019). Pros and cons of gan evaluation measures. *Computer Vision and Image Understanding*, 179:41–65.

Brüggenwirth, S. and Wagner, S. (2021). Theoretical foundations of deep learning. *Deep Neural Network Design for Radar Applications*, pages 69–96.

Carley, E. P., Baldovin, C., Benthem, P., Bisi, M. M., Fallows, R. A., Gallagher, P. T., Olberg, M., Rothkaehl, H., Vermeulen, R., Vilmer, N., and et al. (2020a). Radio observatories and instrumentation used in space weather science and operations. *Journal of Space Weather and Space Climate*, 10.

Carley, E. P., Baldovin, C., Benthem, P., Bisi, M. M., Fallows, R. A., Gallagher, P. T., Olberg, M., Rothkaehl, H., Vermeulen, R., Vilmer, N., and et al. (2020b). Radio observatories and instrumentation used in space weather science and operations. *Journal of Space Weather and Space Climate*, 10.

Carley, E. P., Gallagher, P., Mccauley, J., and Murphy, P. (2020c). Using supervised machine learning to automatically detect type ii and iii solar radio bursts. *EGU General Assembly Conference Abstracts*, page 5109.

Carley, E. P., Reid, H., Vilmer, N., and Gallagher, P. T. (2015). Low frequency radio observations of bi-directional electron beams in the solar corona. *Astronomy and Astrophysics*, 581:1–10.

Cheng, J., Li, Y., Zhang, Y., Yan, Y., Tan, C., Chen, L., and Wang, W. (2022). Mitigation of radio frequency interference in the solar radio spectrum based on deep learning. *Solar Physics*, 297(4).

Connor, L. and van Leeuwen, J. (2018). Applying deep learning to fast radio burst classification. *The Astronomical Journal*, 156(6):256.

Curto, J. D., Zarza, I. C., de la Torre, F., King, I., and Lyu, M. R. (2017). High-resolution deep convolutional generative adversarial networks.

Deng, L. (2012). The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142.

Diaz-Ramirez, J. (2021). Machine learning and deep learning. *Ingeniare*, 29(2):182–183.

Du, Q.-f., Zhang, Q.-m., Li, X., and Gao, C.-l. (2021). Solar radio filtering algorithm based on improved long short-term memory. 21(4).

Dutta, A. and Zisserman, A. (2019). The VIA annotation software for images, audio and video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, New York, NY, USA. ACM.

Ekmekci, P. E. and Arda, B. (2020). History of artificial intelligence. *SpringerBriefs in Ethics*, (December):1–15.

ESA (2018). Space weather effects.

Evgeniou, T. and Pontil, M. (2001). Support vector machines: Theory and applications. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 2049 LNAI, pages 249–257.

Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, 7(1):1–9.

Garcia-Alonso, C. R., Perez-Naranjo, L. M., and Fernandez-Caballero, J. C. (2014). Multiobjective evolutionary algorithms to identify highly autocorrelated areas: The case of spatial distribution in financially compromised farms. *Annals of Operations Research*, 219(1):187–202.

Ghurah, M. A. (2018). Automated solar radio burst detection on radio spectrum: a review of techniques in image processing. 4:9–10.

Girshick, R. (2015). Fast r-cnn. *Proceedings of the IEEE International Conference on Computer Vision*, 2015 Inter:1440–1448.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2016a). Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):142–158.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2016b). Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):142–158.

Glover, A., Hilgers, A., Rosenqvist, L., and Bourdarie, S. (2008). Interplanetary proton cumulated fluence model update. *Advances in Space Research*, 42(9):1564–1568.

Gonzalez, T. F. (2007). Handbook of approximation algorithms and metaheuristics. *Handbook of Approximation Algorithms and Metaheuristics*, pages 1–1432.

Goodfellow, I. (2016). Nips 2016 tutorial: Generative adversarial networks.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2016). Generative adversarial networks. *Commun. ACM.*, 63(11):2234–2242.

Gordo, J. B. and Mateo, M. P. (2022). Automatic burst detection in solar radio spectrograms using deep learning: dearce method. pages 0–27.

Guo, J., Zhu, X., Lei, Z., and Li, S. Z. (2018). Face synthesis for eyeglass-robust face recognition. *arXiv preprint arXiv:1806.01196*.

Guo, J.-c., Yan, F.-b., Wan, G., and Hu, X.-j. (2022). A deep learning method for the recognition of solar radio burst spectrum. pages 1–14.

Gupta, S. (2015). A short introduction to heavy-ion physics. 14(5):771–780.

Haenlein, M. and Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61(4):5–14.

He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2020). Mask r-cnn. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):386–397.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem:770–778.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neur. In.*, 2017-Decem(Nips):6627–6638.

Hou, Y. C., Zhang, Q. M., Feng, S. W., Du, Q. F., Gao, C. L., Zhao, Y. L., and Miao, Q. (2020). Identification and extraction of solar radio spikes based on deep learning. *Solar Physics*, 295(10).

Hu, B., Zheng, Z., Liu, P., Yang, W., and Ren, M. (2021). Unsupervised eyeglasses removal in the wild. *IEEE Transactions on Cybernetics*, 51(9):4373–4385.

Huang, X., Liu, M. Y., Belongie, S., and Kautz, J. (2018). Multimodal unsupervised image-to-image translation. 11207 LNCS:179–196.

Jebaraj, I. C., Kouloumvakos, A., Magdalenic, J., Rouillard, A. P., Mann, G., Krupar, V., and Poedts, S. (2021). Generation of interplanetary type ii radio emission. *Astronomy and Astrophysics*, 654:1–15.

Jenkins, J., Paiement, A., Aboudarham, J., and Bonnin, X. (2020). Physics-informed detection and segmentation of type ii solar radio bursts. *BMVC (under review)*, 1.

Kalkan, E., Okur, F., and Altunisik, A. (2018). Applications and usability of parametric modeling. *Journal of Construction Engineering, Management and Innovation*, 1(3):139–146.

Khan, H. (2021). Types of ai | different types of artificial intelligence systems. *Ai*, 1(1):1–13.

Kirillov, A., Wu, Y., He, K., and Girshick, R. (2020). Pointrend: Image segmentation as rendering. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 9796–9805.

Lecacheux, A. (2000). The Nançay Decameter Array: A Useful Step Towards Giant, New Generation Radio Telescopes for Long Wavelength Radio Astronomy. *Geophysical Monograph Series*, 119:321.

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W. (2016). Photo-realistic single image super-resolution using a generative adversarial network.

Leshem, A., van der Veen, A., and Boonstra, A. (2000). Multichannel interference mitigation techniques in radio astronomy. *The Astrophysical Journal Supplement Series*, 131(1):355–373.

Li, S., Yuan, G., Chen, J., Tan, C., and Zhou, H. (2022). Self-supervised learning for solar radio spectrum classification. *Universe*, 8(12).

Lin, R. P. (2011). Energy release and particle acceleration in flares: Summary and future prospects. *Space Science Reviews*, 159(1):421.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2015). Microsoft coco: Common objects in context.

Linker, J. A., Mikić, Z., Riley, P., Lionello, R., and Odstrcil, D. (2003). Models of coronal mass ejections: A review with a look to the future.

Littman, M. L., Ajunwa, I., Berger, G., Boutilier, C., Currie, M., Doshi-Velez, F., Hadfield, G., Horowitz, M. C., Isbell, C., Kitano, H., and et al. (2022). Gathering strength, gathering storms: The one hundred year study on artificial intelligence (ai100) 2021 study panel report. (September).

Liu, M. Y., Breuel, T., and Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in Neural Information Processing Systems*, 2017-December(Nips):701–709.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9905 LNCS:21–37.

Lobzin, V. V., Cairns, I. H., Robinson, P. A., Steward, G., and Patterson, G. (2009). Automatic recognition of type iii solar radio bursts: Automated radio burst identification system method and first observations. *Space Weather*, 7.

Lobzin, V. V., Cairns, I. H., Robinson, P. A., Steward, G., and Patterson, G. (2010). Automatic recognition of coronal type ii radio bursts: The automated radio burst identification system method and first observations. *Astrophysical Journal Letters*, 710(1 PART 2):58–62.

Lobzin, V. V., Cairns, I. H., and Zaslavsky, A. (2014). Automatic recognition of type iii solar radio bursts in stereo/waves data for onboard real-time and archived data processing. *Journal of Geophysical Research A: Space Physics*, 119(2):742–750.

Lopez Pinaya, W. H., Vieira, S., Garcia-Dias, R., and Mechelli, A. (2019). Autoencoders. *Machine Learning: Methods and Applications to Brain Disorders*, pages 193–208.

Louppe, G. (2014). Understanding random forests: From theory to practice. *Machine Learning*, (July).

Lu, X., Wang, J., Dizaji, R., Ding, Z., and Ponsford, A. M. (2004). A switching constant false alarm rate technique for high frequency surface wave radar. *Canadian*

Conference on Electrical and Computer Engineering, 4(March 2016):2081–2084.

Luo, Y., Yang, F., and Lin, M. (2020). Detection method of solar radio bursts based on support vector machine model. pages 362–366.

Ma, L., Chen, Z., Xu, L., and Yan, Y. (2017). Multimodal deep learning for solar radio burst classification. *Pattern Recognition*, 61:573–582.

Ma, Q., Du, Q. F., Feng, S. W., Hou, Y. C., Ji, W. Z., and Han, C. S. (2022). Solar radio-burst forecast based on a convolutional neural network. *Solar Physics*, 297(10):130.

Meel, V. (2022). What is deep learning? an easy to understand guide.

Mijwel, M. M. (2015). History of artificial intelligence. *Computer Science*, (April 2015):3–4.

Murphy, P. C., Callanan, P., McCauley, J., McKenna, D. J., Fionnagain, D. O., Louis, C. K., Redman, M. P., Canizares, L. A., Carley, E. P., Maloney, S. A., and et al. (2021). First results from the real-time transient acquisition backend (realta) at the irish lofar station. *Astronomy and Astrophysics*.

Niamsuwan, N., Johnson, J. T., and Ellingson, S. W. (2005). Examination of a simple pulse-blanking technique for radio frequency interference mitigation. *Radio Science*, 40(5):1–11.

Offringa, A. R., de Bruyn, A. G., Biehl, M., Zaroubi, S., Bernardi, G., and Pandey, V. N. (2010). Post-correlation radio frequency interference classification methods. *Monthly Notices of the Royal Astronomical Society*, 405(1):155–167.

Offringa, A. R., Van De Gronde, J. J., and Roerdink, J. B. (2012). A morphological algorithm for improving radio-frequency interference detection. *Astronomy and Astrophysics*, 539.

Pankonin, V. and Price, R. M. (1981). Radio astronomy and spectrum management: The impact of warc-79. *IEEE Transactions on Electromagnetic Compatibility*, EMC-23(3):308–317.

Park, H. A. (2013). An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, 43(2):154–164.

Perez, L. and Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning.

Pick M., V. N. (2009). Sixty five years of solar radio astronomy: flares, coronal mass ejections and sun-earth connection. *The Astronomy and Astrophysics Review*.

Radford, A., Metz, L., and Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, pages 1–16.

Rafegas, I. and Vanrell, M. (2017). Color representation in cnns: Parallelisms with biological vision. *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, 2018-Janua:2697–2705.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December:779–788.

Redmon, J. and Farhadi, A. (2017). Yolo9000: Better, faster, stronger. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua:6517–6525.

Reid, H. A. S. and Ratcliffe, H. (2014). A review of solar type iii radio bursts. *Research in Astronomy and Astrophysics*, 14(7):773–804.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved techniques for training gans. *Adv. Neur. In.*, pages 2234–2242.

Salmane, H., Weber, R., Abed-Meraim, K., Klein, K. L., and Bonnin, X. (2018). A method for the automated detection of solar radio bursts in dynamic spectra. *Journal of Space Weather and Space Climate*, 8:0–18.

Scully, J. (2022a). Srb detection with yolov2. github.com/jeremiahscully/.

Scully, J. (2022b). Srb simulation with gans. github.com/jeremiahscully/GANs.

Scully, J., Flynn, R., Carley, E., Gallagher, P., and Daly, M. (2021). Type iii solar radio burst detection: A deep learning approach. In *Irish Signal and Systems Conference 2021*, page 1–6.

Scully, J., Flynn, R., Gallagher, P., Carley, E., and Daly, M. (2022). Simulating solar radio bursts using generative adversarial networks.

Shen, X. (2022). A survey of object classification and detection based on 2d/3d data.

Singh, D., Sasikumar Raja, K., Subramanian, P., Ramesh, R., and Monstein, C. (2019). Automated detection of solar radio bursts using a statistical method. *Solar Physics*, 294(8):1–16.

Smith, F. (1986). Book review: Interferometry and synthesis in radio astronomy. *IEE Proceedings F Communications, Radar and Signal Processing*, 133(7).

Smolders, B. and Hampson, G. (2002). Deterministic rf nulling in phased arrays for the next generation of radio telescopes. pages 13–22.

Socarras Salas, Y., Vazquez Bermudez, D., Lopez Pena, A. M., Geronimo Gomez, D., and Gevers, T. (2012). Improving hog with image segmentation: Application to human detection. *Lecture Notes in Computer Science (including subseries*

Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7517 LNCS:178–189.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June:1–9.

Thompson, A. R., Moran, J. M., and Swenson, G. W. (2017). *Radio Frequency Interference*, pages 787–808. Springer International Publishing, Cham.

Trieu, T. H. (2018). Darkflow. *GitHub Repository. Available online: https://github.com/thtrieu/darkflow (accessed on 14 February 2019).*

Tzutalin (2015). Labelimg. Free Software: MIT License.

Ultralytics (2022). Ultralytics yolov8. https://github.com/ultralytics/ultralytics.

Van Haarlem, M. P., Wise, M. W., Gunst, A. W., Heald, G., McKean, J. P., Hessels, J. W., De Bruyn, A. G., Nijboer, R., Swinbank, J., Fallows, R., and et al. (2013). Lofar: The low-frequency array. *Astronomy and Astrophysics*, 556.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2022). Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. pages 1–15.

Warmuth, A. and Mann, G. (2004). The application of radio diagnostics to the study of the solar drivers of space weather. pages 49–68.

Weng, W. and Zhu, X. (2021). Inet: Convolutional networks for biomedical image segmentation. *IEEE Access*, 9:16591–16603.

Wild, J. P., Smerd, S. F., and Weiss, A. A. (1963). Solar bursts. *Annual Review of Astronomy and Astrophysics*, 1(1):291–366.

Wiley, V. and Lucas, T. (2018). Computer vision and image processing: A paper review. *International Journal of Artificial Intelligence Research*, 2(1):22.

Wu, J. (2017). Introduction to convolutional neural networks. *Introduction to Convolutional Neural Networks*, page 1–31.

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, 1168(2).

Zhan, X., Jiang, Z., Chen, Z., Zhang, M., and Song, C. (2014). Structures of gmcw37. *Research in Astronomy and Astrophysics*, 14(1):433–455.

Zhang, P. J., Wang, C. B., and Ye, L. (2018a). A type iii radio burst automatic analysis system and statistic results for a half solar cycle with nancay decameter array data. *Astronomy and Astrophysics*, 618:1–11.

Zhang, W., Yan, F., Han, F., He, R., Li, E., Wu, Z., and Chen, Y. (2021). Auto recognition of solar radio bursts using the c-dcgan method. *Frontiers in Physics*, 9(September):1–8.

Zhang, Y., Gao, J., and Zhou, H. (2020). Breeds classification with deep convolutional neural network. *ACM International Conference Proceeding Series*, pages 145–151.

Zhang, Y. G., Gajjar, V., Foster, G., Siemion, A., Cordes, J., Law, C., and Wang, Y. (2018b). Fast radio burst 121102 pulse detection and periodicity: A machine learning approach. *arXiv*.

Zhao, Z. Q., Zheng, P., Xu, S. T., and Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11):3212–3232.

Zhu, J. Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE*

International Conference on Computer Vision, 2017-October:2242–2251.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2021). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76.

Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T. Y., Shlens, J., and Le, Q. V. (2020). Learning data augmentation strategies for object detection. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12372 LNCS:566–583.

Appendix A

SRB classification methods

Carley et al. (Carley et al. 2020c) explores both machine and DL techniques for SRB classification.

A.1 Training and test data: Radio Solar Telescope Network and SRB Simulations

In order to construct the training data set that will be used as input into our ML algorithms, we first search the catalogues of the NOAA's Space Weather Prediction Centre. These catalogues provide solar radio burst event reports that include type, time, frequency, and a few other relevant metrics of the bursts. Our next step is to collect the raw data that will be used to construct the training data set. After that, we retrieve the pertinent information from the Radio Solar Telescope Network (RSTN) archive, limiting our search to occurrences that have occurred between 1996 and the present day because of the availability of data. In an ideal world, our library would have all five different types of radio bursts; however, types I, IV, and V do not occur very often, and in order to train ML algorithms effectively, there must be hundreds, if not thousands, of samples of a single classification feature. Only type IIs and

Type IIIs have a population that is measured in the thousands within the RSTN data archive.

We start by performing a background subtraction on each of the bursts, and then we scale the intensity of the bursts from their greatest to their minimum intensity. This ensures that the background "noise floor" of the spectrogram is not included in the scale. In order to decrease the dimensional input required by the ML algorithms, the dynamic spectra are downsampled to a 300 x 300 array (see below). We carefully review each spectrogram (image) that is included in the training data set, and we exclude any data that is especially tainted by RFI , has incorrect burst classification, is very weak, or is utterly ambiguous. An archive consisting of 900 images of Type II and 1800 images of Type III was produced as a final result. We added 1,500 more no detections or dynamic spectra that simply had RFI. These spectra did not contain any radio bursts. It is essential for the purposes of monitoring and space weather that the ML algorithms be trained to recognise an "all-clear" signal (also known as true negatives).

The total number of images is 3500, which is an amount that is suitable as input for traditional ML algorithms like SVM and RF but is insufficient for usage with deep neural networks like CNNs (see below). As a result, in addition to the data from the RSTN, we augment the training data that we use by simulating type II, Type III, and no-radio burst dynamic spectra. We generate Type III radio bursts by using parametric models, and these bursts have random characteristics in terms of their number, grouping, intensity, drift rate, inhomogeneity, start-end frequency, and start-end time. Figure X provides a good illustration of how we embed the bursts in a background of simulated and random RFI channels. We carry out the same procedure for the creation of a Type II radio burst, in which its drift rate, position in the spectrogram, inhomogeneity, and length in both time and frequency are each entirely randomised. We build 25,000 simulated Type III dynamic spectra, 25,000 simulated Type II dynamic spectra, and 25,000 dynamic spectra that are ostensibly

empty but actually include simulated RFI. For the purpose of training the RF CNN in particular, this simulated data set was utilised.

Using RSTN and simulation data sets, we are going to conduct an accuracy test on the machine and DL algorithms.

A.2 Support vector machines

SVMs are supervised ML algorithms that can be applied as discriminative classifiers (Evgeniou and Pontil 2001). An SVM finds a hyperplane with the dimension D-1 that separates the two (or more) groups of data points when it is given a collection of data points that have been grouped into various regions (various classes) of a vector space that has the dimension D. This is done when the SVM is presented with a vector space that has the dimension D. In this particular instance, one dimension of the space was represented by the pixel numbers of the radio burst image, which were 10,000 in total. Each pixel represented a value along one of the dimensions of the space. Because of this, a single point in this high-dimensional space can be represented by a single spectrogram. The SVM algorithm makes an effort to locate one or more hyper-planes in this space that are capable of efficiently dividing the three types of radio bursts into distinct groups. Dot products in this space are subjected to an application of a function known as a "kernel," and the typical method used to accomplish this goal is a transformation to a new space. The application of the kernel to the points in the initial space is a computationally inexpensive means of transforming to a higher dimensional space without the specific need to perform this transformation; as a result, this technique is referred to as the "kernel trick" in SVM. The SVM technique was implemented using the Scikit-learn package included in Python. All of the data were pre-processed to have a mean of zero and a standard deviation of one, as is customary for the vast majority of ML algorithms. When transforming to D+1 space, a polynomial kernel of degree 2 was utilised, which

resulted in a classification accuracy of 76% when applied to the test data. In most instances, hyperparameter tuning of the SVM algorithm resulted in overfitting of the data to the test set. In general, tweaking the hyperparameters of the SVM algorithm did not result in better classification. Table A.1 displays the precision, recall, and f1-scores for each of the burst classes.

SVM SRB Classification			
SRB Type	Precision	Recall	F1-Score
No burst	0.80	0.82	0.81
Type II	0.75	0.89	0.82
Type III	0.97	0.77	0.86

Table A.1 Classification metrics for the SVM model applied to the RSTN data set.

A.3 Random Forest

RF is a type of supervised ML method that may perform either regression or classification by employing an ensemble of decision trees and bagging the results of the program's computations (Louppe 2014). In this specific instance, we make use of RF as a classifier for the purpose of distinguishing radio bursts in the high-dimensional space that was previously presented. Once again, the Scikit-learn package for Python was used in its implementation. On the basis of the test data, an RF classifier consisting of 20 trees achieves a classification accuracy of 74%. Table A.2 displays the precision, recall, and f1-scores for each of the burst classes.

RF SRB Classification			
SRB Type	Precision	Recall	F1-Score
No burst	0.72	0.79	0.75
Type II	0.88	0.58	0.70
Type III	0.69	0.86	0.77

Table A.2 Classification metrics for RF on the RSTN data set.

A.4 Dimensionality reduction and separability of SRBs in high-D space

SVM, RF, 1-Hidden Layer Neural Network (NN), and other methods such as k Nearest Neighbor (KNN) or Spanning Trees seek to build a function that distinguishes classes of data points in a high dimensional space. These algorithms are different from one another in the theoretical implementation, but ultimately accomplish the same goal when applied to the data. If the data are noisy or if they cannot be separated in the high-dimensional space, then there is no way to enhance the accuracy. We utilise a t-distributed stochastic neighbor embedding to project a 10,000-dimensional space onto a two-dimensional plane so that we can visualise the clustering and distribution of radio bursts in the high-dimensional space (Garcia-Alonso et al. 2014). This allows us to see how the radio bursts are distributed. tSNE is a method for maintaining distances in a space that has many fewer dimensions than the original. For instance, clusters of points that are geographically separated in the original space will continue to be geographically separated in the projection space. The results of such a projection are displayed in Figure A.1 for type II, Type III, and no radio burst. Even though these bursts can be separated into clusters at various points in the two-dimensional plane, they frequently segment and overlap one another. This is an indication that no simple function will be able to separate these three classes in their original space, as there will always be an overlap of points or a lack of simple clusters of each class that an ML algorithm can isolate. This is because there will never be a situation in which all of the points in each of the classes will be completely isolated from one another Because radio bursts can occur anywhere within the dynamic spectrum (image), there will never be a unique set of pixels that denote a particular radio burst class. This is due to the fact that radio bursts, while morphologically complex, can occur anywhere within the image. This results in the

radio blasts being dispersed all around high-D space and unable to be separated from one another.

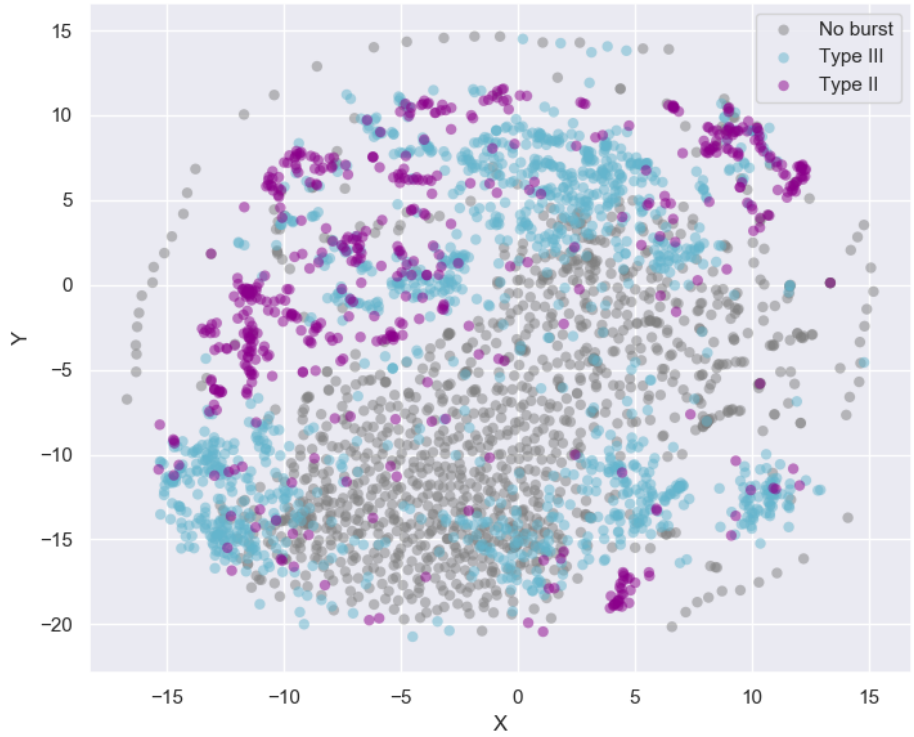


Fig. A.1 t-Distributed stochastic neighbour embedding of type II, III and no burst images onto a 2D space. While the different classes show some clustering, these clusters are not homologous and often overlap. The clusters may not be separable in the high-dimensional space of the original images.

In light of the results of the tSNE study, we have reached the conclusion that a more intricate representation of the images is required in order to correctly categorise the radio bursts. The use of CNN models is considered to be the cutting edge of CV and image classification. Next, we outline the training of the RF model for SRB classification.

A.5 Inception V3

Google's RF CNN is a DL model that was released in 2015. It is a subtype of the architecture known as Inception, which was developed to perform image classification tasks in a manner that is both more efficient and effective. The RF CNN architecture is a DL model that analyses and classifies images using a combination of convolutional and fully connected layers. It is made up of a number of inception modules, each of which has a number of convolutional and pooling layers that are used to extract features from the image that is being fed into the network. These extracted features are then processed through a succession of fully connected layers, which employ the extracted features to create image predictions.

The RF architecture also includes a number of methods for improving the performance of the model. These methods include the use of batch normalisation to stabilise the training process and the incorporation of auxiliary classifiers to improve the overall accuracy of the model. Both of these methods are included as part of the architecture. In addition, the RF architecture extracts features from the input image by using a combination of 1x1 and 3x3 convolutional filters. This enables it to collect both local and global features and, as a result, generate a classification score for a specific object in an image. In this case, we use the inception model to distinguish and classify SRBs within I-LOFAR observation images. The Inception model was implemented using the Tensorflow package included in Python. Using the concept of transfer learning, we take an advanced version of the Inception model that has already been trained on millions of images of generic scenes (such as cars and people) and retrain a smaller portion of the network on our own custom data SRBs. We extend and train the high-level portion of the inception model using the parametric modelling generated data. As stated in Section 3.1, we create three classes consisting of RSTN Type III dynamic spectra data, RSTN Type II dynamic spectra, and RSTN 'empty' dynamic spectra with RFI. These three classes are fed

directly into the Inception architecture for training. The Inception model was trained for 50 epochs and was evaluated using RSTN data as a test set but also real observed I-LOFAR data. The RSTN test set consisted of 300 images and the inception model achieved an accuracy score of 97%. However, this test set is quite similiar to the training set so we needed to test on some unseen data. So, we then evaluated the model's accuracy by applying it to an I-LOFAR observation day, 10/09/2017 to be exact, see Figure A.2. The model achieved an accuracy score of 90% when evaluated on LOFAR data, illustrating how robust the inception model is but more importantly how robust CNNs are at classifying SRBs.

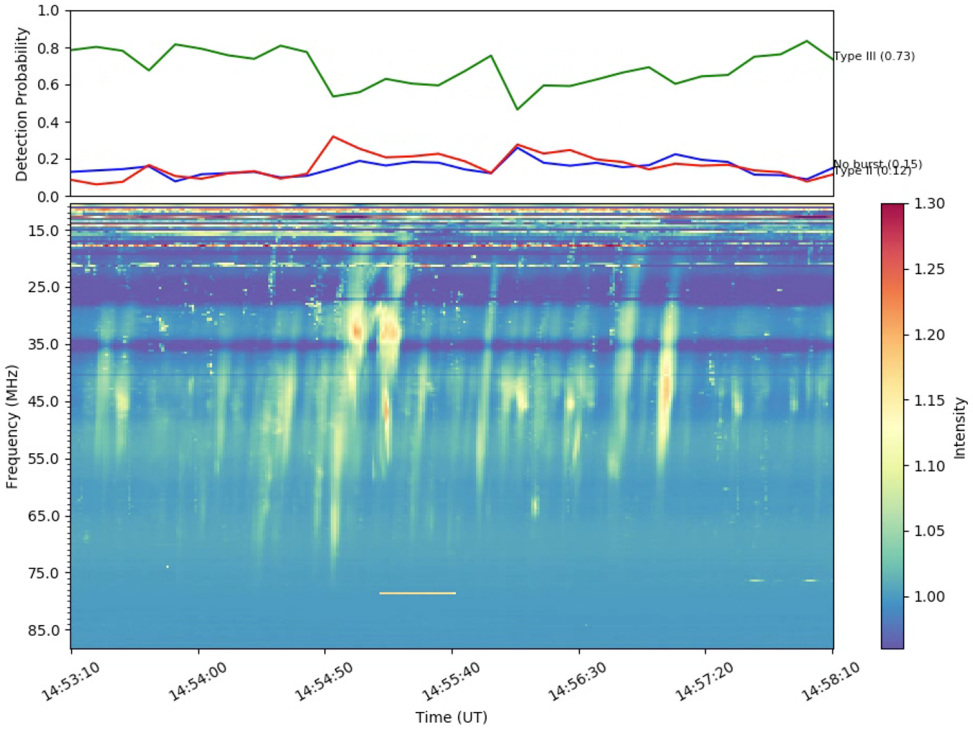


Fig. A.2 The Inception model tested on the 10/09/2017 observation made by I-LOFAR. Although the model proves successful at classifying Type IIIs within the image, it cannot locate exactly when and where in terms of time and frequency.

In spite of the fact that the Inception model is quite effective at classifying SRBs, it can only indicate the likelihood of the presence of an SRB within an image by means of the classification score, as seen in Figure X. The Inception model does not

take into account the precise location of a specific SRB, regardless of whether the SRB is of Type II or Type III, in terms of the frequency or the duration of time.