Università degli Studi di Trieste
Dipartimento di Fisica

Corso di Laurea Triennale in Fisica

A Comparative Analysis of Neural Network Methods for Source Detection and Classification in Astrophysics

Laureando:
Omar Cusma Fait

Relatore:
prof. Elena Orlando

ANNO ACCADEMICO 2022-2023

Abstract

Machine learning enables researchers to make discoveries and better understand the universe. However, as the amount of data increases, the tools they use to analyze, manage, and store data must adapt accordingly. The importance of developing new tools is therefore paramount. Machine learning has gained significant importance due to its potential to enhance the efficiency and accuracy of source detection and classification tasks.

This thesis compares different methods for source detection, classification, localization, and deblending in astrophysics. It also investigates various state-of-the-art machine learning models, including Region-based Convolutional Neural Networks (R-CNN), Fast R-CNN, Faster R-CNN, and Mask R-CNN, and the datasets they have been trained on. The study of these models reveals insights into their strengths, limitations, and suitability for different applications.

Contents

1 Introduction

Modern astrophysics is a data-rich field that involves analyzing massive datasets to explore and understand the universe at large scales. Due to advancements in technology, the amount of data generated by astrophysical observations has increased in recent years. This surge in data poses challenges for traditional source detection and classification methods in astrophysics. Neural networks have emerged as a powerful tool to handle the complexity and volume of astrophysical data. Machine learning-based methods can potentially improve the efficiency and accuracy of automated image analysis, enabling astrophysicists to extract valuable insights from large datasets.

Before the advent of computers, astronomers had to resort to manual observational methods to detect and classify astronomical objects. They visually inspected photographic plates or glass slides containing recorded night sky images. They meticulously compared these images with reference catalogs or star atlases to identify and classify stars based on their brightness, color, spectral characteristics, and other observable features. Manual analysis rapidly became time-consuming, impractical, and subjective and has fallen in favor of computer-aided techniques.

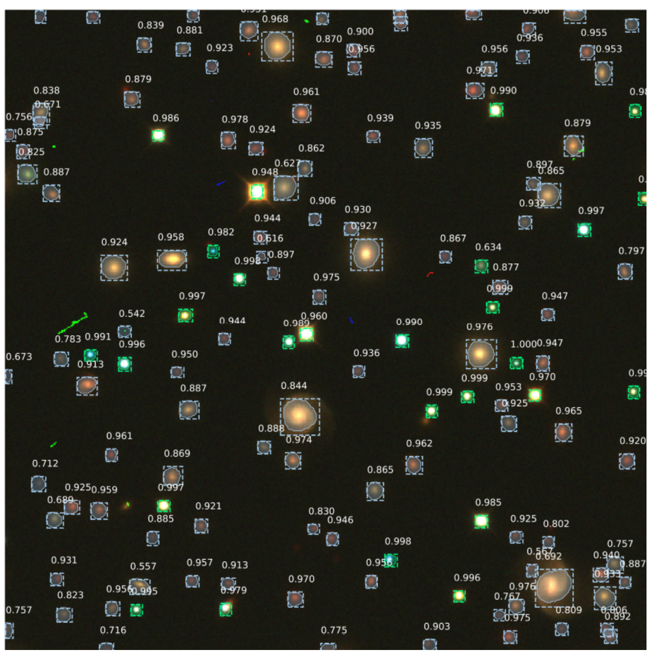

Figure 1: Example of detection inference in a simulated image (image taken from Burke et al. (2019)) Galaxies are shown in light blue regions, and stars are shown in green. The detection confidence that the object belongs to that particular class is shown above each mask.

New data analysis methods have greatly enhanced researchers' ability to extract useful information from massive amounts of data. Since machine learning and deep neural network models appeared in this field, the paradigm has shifted towards automated and data-driven approaches. These techniques demonstrate remarkable capabilities in handling complex and diverse astrophysical datasets, enabling researchers to extract valuable information and uncover new insights. The fig. 1 is an example of the input and output of such a model. It shows a simulated astronomical image with different sources in the optical band. Dashed squares are bounding boxes indicating that the model could detect the sources. The blue curve that outlines the silhouette of the sources delimitates the regions where the model predicts the presence of a source. Their color indicates that the model could classify the sources as galaxies (light blue) and stars (green). The model successfully identifies the sources as separate objects, even in the crowded lower right region.

This thesis aims to review a non-exhaustive sample of scientific publications that explore various neural network architectures used to perform source detection and classification on astrophysical images, exploring their capabilities and performance. The primary objective is to compare the U-Net, R-CNN, Faster RCNN, and Mask R-CNN architectures. Specifically, it describes the model that the publications in question are using, the task that the model tries to perform, the type of data that the model takes as input, and the training dataset. It also investigates the feasibility of applying the model proposed in one paper to the dataset of another.

This review briefly introduces the basics of astrophysics and machine learning to make it easier for readers to understand later sections. Section 2 will showcase some of the most commonly used datasets in astrophysics, which also feature in the scientific publications discussed in this thesis. Section 3 discusses the goals and applications of machine learning. Section 4 begins with simple examples of machine learning, followed by a list of techniques used in scientific papers and their limitations. Section 5 reviews the work done in each article, including a brief discussion about the model's limitations. At the end of this section there is a table summarizing all the relevant information about each article. Finally, Section 6 presents conclusions and remarks.

2 Description of the datasets

Astrophysics encompasses a wide selection of phenomena and requires various datasets to study the universe. Researchers in astrophysics utilize numerous types of datasets to unravel the mysteries of celestial objects and phenomena. Sky surveys provide comprehensive sky maps, facilitating the identification of new objects and the study of large-scale structures. Time-domain surveys capture the dynamic nature of astronomical phenomena, such as transient events and periodic variations. Additionally, simulated datasets play a vital role in testing theoretical models and numerical simulations, aiding in interpreting and validating observational data. The diverse range of datasets in astrophysics ensures a comprehensive approach to studying the universe and deepening our knowledge of its intricacies.

The articles in this study focus on various regions of the electromagnetic spectrum. The images comprising each dataset contain various sources, such as stars, galaxies, active galactic nuclei (AGN), and pulsars (PSR). An important variable to consider is the density of sources in each image. Source-dense images will result in more frequent overlapping, making classification very challenging.

The following chapters present some of the most important observational instruments that provided the data in the discussed publications, along with a short description.

2.1 Vera C. Rubin Observatory (LSST)

The Vera C. Rubin Observatory[1], previously called the Large Synoptic Survey Telescope (LSST), is an astronomical observatory in Chile. Its main task is carrying out a board-view astronomical survey in the optical range, the Legacy Survey of Space and Time.

"Synoptic" refers to observations that provide a broad view of a subject at a specific moment. The LSST Base Facility, located on a 2682-meter-tall mountain named Cerro Pachón in Chile, is named after Vera Cooper Rubin, a pioneering American astronomer renowned for her discoveries related to galaxy rotation rates.

The data produced by this observatory will include images of stars, galaxies, asteroids, and other celestial objects. Repeated observations of the same region in the sky enable the study of transient phenomena (events that exhibit a noticeable change in brightness or characteristics over a relatively short period), such as supernovae.

[1]Source: `https://en.wikipedia.org/wiki/Vera_C._Rubin_Observatory`

2.2 Fermi LAT

The Large Area Telescope[2] (LAT) is an instrument on board the Fermi Gamma-ray Space Telescope (FGST), a space observatory launched by NASA in 2008 and performing gamma-ray astronomy in low earth orbit. The data from the LAT consist of measurements of gamma-ray photons over a wide energy range, spanning from 20 MeV to 300 GeV. The primary scientific objective of the LAT is to explore the gamma-ray universe, studying a wide range of celestial sources and phenomena.

In more detail, the main goal of the FGST is to conduct a comprehensive all-sky survey to study various astrophysical and cosmological phenomena, including active galactic nuclei, pulsars, and dark matter. In addition to the LAT, the FGST is equipped with the Gamma-ray Burst Monitor (GBM) to investigate gamma-ray bursts at lower energies. FGST is a collaborative project involving NASA, the United States Department of Energy, and government agencies from France, Germany, Italy, Japan, and Sweden. It is named in honor of Enrico Fermi, a prominent figure in high-energy physics.

2.3 Planck

Planck[3] was a space observatory operated by the European Space Agency (ESA). It mapped the anisotropies of the cosmic microwave background (CMB) at a higher resolution than its predecessor, the WMAP probe. The CMB is the residual radiation from the early universe and consists predominantly of microwave radiation. The Planck dataset has provided unprecedented insights into the universe's composition, evolution, and large-scale structure. Planck has also detected many galactic and extragalactic astronomical objects, such as different types of galaxies, stars surrounded by dust, galaxy clusters, and unidentified objects.

2.4 ASKAP

The Australian Square Kilometre Array Pathfinder[4] (ASKAP) is a radio telescope array in Western Australia. It is a technology demonstrator for the international Square Kilometre Array (SKA) project. It is operated by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and consists of 36 parabolic antennas, each 12 meters in diameter, working together as a single interferometer. It has a large collecting area and uses phased array feeds to enhance survey speed and sensitivity. ASKAP focuses on radio emissions from galaxies, pulsars, and radio transients.

[2] Source: `https://en.wikipedia.org/wiki/Vera_C._Rubin_Observatory`

[3] Source: `https://en.wikipedia.org/wiki/Planck_(spacecraft)`

[4] `https://en.wikipedia.org/wiki/Australian_Square_Kilometre_Array_Pathfinder`

2.5 ATCA

The ATCA[5] (Australian Telescope Compact Array) is a radio interferometer located at the Paul Wild Observatory near Narrabri, New South Wales, Australia. It is operated by the CSIRO. The ATCA consists of six 22-meter diameter dish antennas, which can be arranged in different configurations to provide different levels of angular resolution. The antennas are movable along a set of tracks, allowing for various configurations ranging from a compact array with the dishes close together to an extended array with the dishes spread over several kilometers.

This observatory can capture a variety of radio sources, including radio galaxies, radio supernova remnants, galactic radio jets, and more. Among other achievements enabled by this observatory, in 1996, scientists could use the data captured by this array to create the most detailed maps of hydrogen in the Magellanic Clouds. In 2001 they created the first three-dimensional structure of the Large Magellanic Cloud.

2.6 Radio galaxy zoo dataset

Radio Galaxy Zoo[6] (RGZ) is a citizen science project that aims to locate supermassive black holes in distant galaxies by engaging the public in classifying radio sources. The project helps identify black hole/jet pairs and their host galaxies, providing insights into black hole evolution. RGZ was initiated to cross-identify millions of extragalactic radio sources expected to be discovered by future surveys. Led by scientists Julie Banfield and Ivy Wong, RGZ collects valuable contributions from citizen scientists to enhance our understanding of black holes and their connections to galaxies. Data were taken from the Faint Images of the Radio Sky at Twenty-Centimeters (FIRST) survey and from the Australia Telescope Large Area Survey (ATLAS), taken with the ATCA Array.

2.7 Gaia

The Gaia[7] mission, conducted by the European Space Agency (ESA), aims to create a highly precise three-dimensional map of the stars in our neighborhood in the Milky Way galaxy and a less precise map of the entire galaxy. It involves the Gaia spacecraft, equipped with astrometric, photometric, and spectroscopic instruments to observe and measure billions of stars' positions, distances, motions, and properties in our galaxy.

The Gaia dataset is an extensive collection of astrometric and photometric data gathered by the Gaia spacecraft. It includes measurements of positions, parallaxes, proper motions, and photometric information such as brightness

[5]Source: `https://en.wikipedia.org/wiki/Australia_Telescope_Compact_Array`

[6]Source: `https://en.wikipedia.org/wiki/Radio_Galaxy_Zoo`

[7]Source: `https://en.wikipedia.org/wiki/Gaia_(spacecraft)`

and colors for a vast number of stars. This comprehensive dataset allows astronomers to study the Milky Way's structure, composition, and dynamics in unprecedented detail.

2.8 DECam

The Dark Energy Camera[8] (DECam) is a cutting-edge instrument at Chile's Cerro Tololo Inter-American Observatory. Its capabilities have been harnessed in numerous ambitious survey projects. Notably, it played a pivotal role in the Dark Energy Survey (DES), a groundbreaking initiative investigating the nature of dark energy and the expansive structure of the universe. This instrument allows astronomers to control and monitor observations from distant locations, streamlining data acquisition and maximizing efficiency. Through its contributions to studies on galaxy evolution, cosmology, supernovae, and gravitational lensing, DECam has become an important tool for researchers worldwide.

2.9 MeerLICHT

The South African Astronomical Observatory[9] (SAAO) houses numerous instruments, including MeerLICHT, an optical wide-field telescope for optical transient surveys. With a 60cm aperture, its main objective is to instantly identify optical counterparts of radio transients, such as fast radio bursts and gamma-ray bursts, for the MeerKAT radio telescope.

[8]Source: `https://en.wikipedia.org/wiki/Dark_Energy_Survey`
[9]Source: `https://en.wikipedia.org/wiki/South_African_Astronomical_Observatory`

3 Types of data analysis tasks

3.1 Image classification

Image classification is the process of assigning one or more labels to an image based on its visual content. The goal is to train a model to recognize and classify the object represented in an image accurately. It is a fundamental task in the field of artificial intelligence.

Various evaluation metrics are used to assess an image classification model's performance. One commonly employed metric is the confusion matrix, which provides a detailed breakdown of the model's predictions and the actual labels. It allows us to easily visualize the model's strengths and weaknesses by showing the categories in which the model is most confident and what classes the model mixes up. For example, in fig. 2, the values represent the frequency of how many times the model guessed the label in the row when faced with the label in the column. The colors are darker for low-frequency guesses and lighter for high-frequency guesses. When the model scores perfectly, only the confusion matrix is an identity matrix (1 on the diagonal, 0 elsewhere).

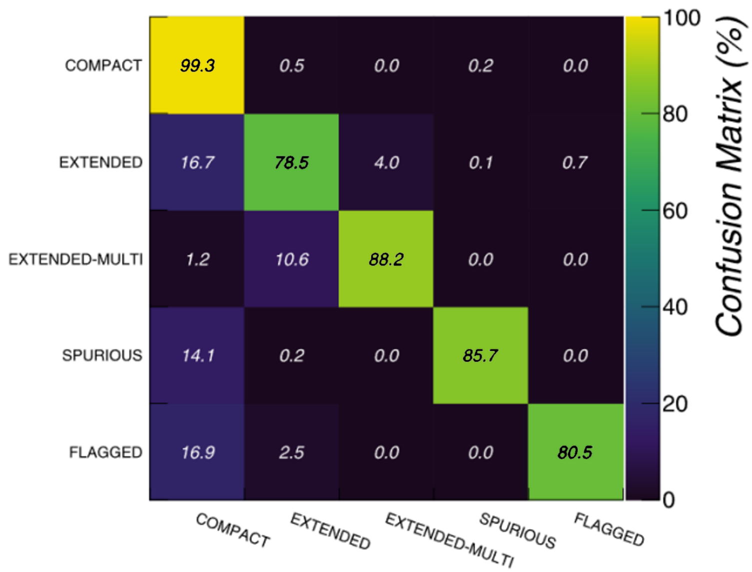

Figure 2: Confusion matrix. Correct classifications are counted on the diagonal (most frequent, yellow/green), misclassifications appear everywhere else (least frequent, dark blue) (Riggi et al., 2023)

3.2 Source detection

Source detection in astrophysics refers to the process of identifying and extracting astronomical sources, such as stars, galaxies, or other celestial objects in an image. It involves identifying regions or pixels in an image or dataset that contain significant signals or emissions from these sources, distinguishing them from the background noise or unrelated features. In the articles reviewd in this thesis, source detection is sometimes a preliminary step before classifying the detected sources. It is commonly done using some variation of a Convolutional Neural Network (CNN), as show in the following sections.

The objective is to identify the presence of sources and estimate their initial properties, such as position and flux. Source detection methods typically utilize statistical techniques and thresholding algorithms to detect significant deviations from the background noise level.

3.3 Source localization

Source localization (sometimes called “source finding”) is the subsequent step to source detection. It focuses on precisely determining the positions or coordinates of the detected sources. It involves refining the initial estimates obtained from source detection and providing accurate positional information. Source localization methods can include centroiding, fitting point spread functions, or employing advanced image processing algorithms to estimate the exact location of the sources within the data. An example of such a task performed on a crowded field is shown in fig. 3.

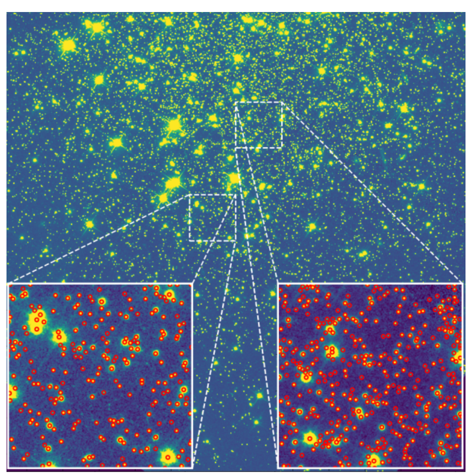

Figure 3: Star cluster image retrieved from the Hubble Space Telescope archive. In red are the localized sources (Stoppa et al., 2022)

3.4 Image segmentation

Image segmentation is the process of partitioning an image into multiple segments or regions, each corresponding to a particular object or region of interest. The goal is to simplify the representation of an image by producing a pixel-wise segmentation map of the image, where each pixel is assigned to a specific object instance, making it easier to read the image. This task is not exclusive to astronomy, as it sees applications in medicine, video surveillance, autonomous driving, and for object detection in general. It can be used to outline people in an image, as can be seen in 4. This task is challenging when dealing with occlusion (when two objects partially overlap), as shown in the section about deblending.

Figure 4: Example of instance segmentation. In all of these images, each region corresponding to a different person has been separated from the others (Burke et al., 2019)

Importantly, while image segmentation is a broad term that refers to a technique that partitions an image, there are a couple of more specific terms that is worth mentioning. Semantic segmentation assigns a label to each pixel in an image, where all pixels belonging to the same class are labeled with the same category. An example of semantic segmentation is an algorithm that outlines all people in the image. Instance segmentation takes this a step further, by not only assigning class labels to pixels, but also by distinguishing between individual instances of objects within the same class. As an example, finding each individual person in the image.

3.5 Deblending

Deblending refers to the process of separating the contributions of multiple sources when they overlap in an image. This can be particularly important in astrophysics, where galaxies or other objects can be difficult to distinguish when they are close together or overlapping. Crowded images result in a high number of blended sources which, if not considered, may end up being undetected or misclassified. An example of deblending can be seen in fig. 5.

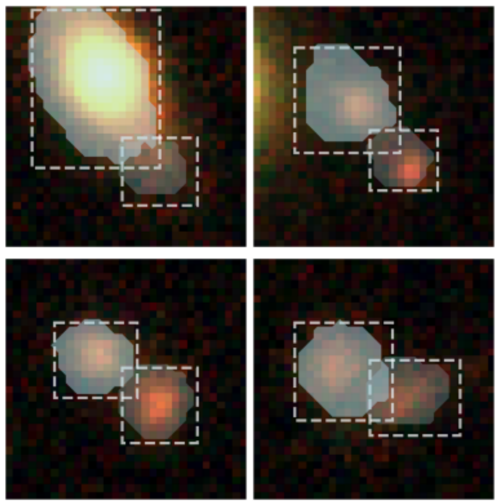

Figure 5: Examples of close blends identified as separate galaxies in a real astronomical image using a deep learning technique. Bounding box edges are shown as white dashed lines, and masks are shown as transparent white regions (Burke et al., 2019)

3.6 Denoising

Although not a very common choice in astrophysics, noise can be removed from an image with the help of neural networks. One downside is that a model trained to perform such a task can be subject to overfitting and generalize poorly. Another problem is that the training data can be tricky to obtain. A preferred approach is the simpler convolution with a Gaussian kernel, as shown in the next chapter.

A denoising model can be trained on pairs of noisy and denoised images. Such a dataset can be easily obtained with a simulation. Determining the noise characteristics in an image obtained from real-world observations, on the other hand, can be challenging, as it is not directly known or controlled like in simulated data. It can be done by comparing an image with a shorter exposure time against the same image with a longer exposure time. This type of data, of course, is only sometimes available.

4 Methods and limitations

This section provides the basic concepts of Machine Learning, focusing on the neural network architectures mentioned in the scientific articles reviewed in Section 5.

4.1 Kernels and convolutions

A kernel is a small matrix applied to an image for various operations, such as filtering and smoothing. The convolution is computed using the kernel as a sliding window (fig. 6) that scans the whole image and performs a weighted average of the pixels within that region.

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 3 & 3 & 0 \\ 0 & 3 & \boxed{3} & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 3 & 3 & 0 \end{bmatrix}$$

Figure 6: Example of edge detection with convolutions. Each combination of a 3x3 matrix in the matrix on the left gets multiplied elementwise with the kernel (in the middle), and the sum of those values is assigned to the appropriate element of the output matrix (on the right)

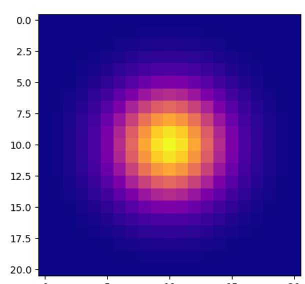

Figure 7: Gaussian kernel, plotted with the matplotlib[10]Python library

A Gaussian kernel (fig. 7), specifically, is a type of kernel that follows a Gaussian distribution. It is commonly used for image smoothing or denoising tasks. This choice of the kernel has been proven to be well-suited for denoising images of stars.

[10]Source: `https://matplotlib.org/3.1.1/gallery/index.html`

A kernel can also be trained using reinforcement learning. It can be asked to predict the original version of images given some noisy ones. A simpler approach can be to optimize the radius of the Gaussian distribution of the kernel to suit the specific dataset better. Training, though, requires having some data with known noise, which can be obtained through simulations.

4.2 Neural networks

A neural network (NN) is a complex mathematical function based on a collection of units called neurons (or nodes), connected to each other through edges (or synapses). Each neuron performs a simple computation based on the output of the connected neurons, and produces a result that gets fed to other neurons. This process is inspired by biological neural networks.

The operation that each neuron performs is the weighted sum of its inputs. The weights are the parameters of the neural network, one for each node. A bias is also applied, which roughly corresponds to a threshold that the inputs need to overcome for the neurons to fire. An activation function is then applied to the sum. This operation determines the output of the neuron.

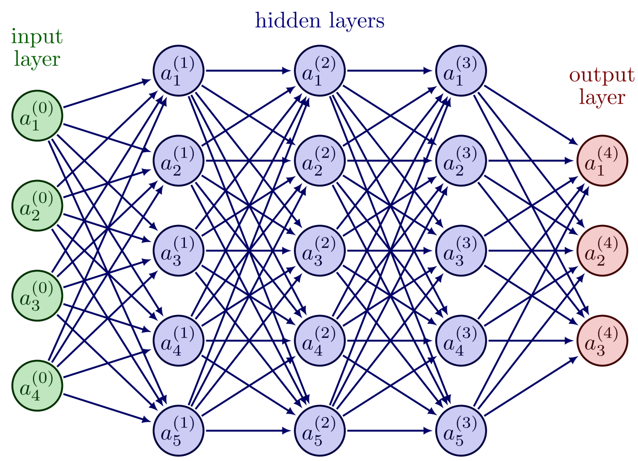

Figure 8: A simple feed-forward neural network with fully connected layers. The input nodes (green) represent the input values. Each subsequent node (blue or red circles) takes as input (arrows) the values from the previous nodes, and produces an output that is then propagated until the final layer (output, in red). (image taken from tikz.net[11])

Most of the machine learning models are variations of the Feed Forward Neural Network (FFNN, fig. 8), which is a type of neural network where the

[11]Source: `https://tikz.net/neural_networks/`

neurons are arranged into layers, and the neurons of all the neurons of two adjacent layers are connected. The input is propagated from the input layer, throughout the hidden layers, to the output layer.

The connections' weights between the network nodes are initially set to random values. During training, the network adjusts these weights to minimize the difference between its predictions and the true values. Once the neural network is fully trained, it can make predictions on new data by feeding the input just like during training, as it is shown more in depth in the Training section. The performance of the model is evaluated using a loss function on a training dataset. The deterministic and non-recursive nature of the model allows the efficient computation of the gradient of the loss, making it possible for the model to learn from the data by updating the connections of the neurons in order to minimize the loss.

4.3 Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a type of NN that uses convolutions and pooling (reducing the size of the image by performing a local average or computing the local maximum) to improve the efficiency of the training and its ability to generalize. These operations exploit the symmetry of the object's position in the image to reduce the problem's complexity, making it a natural choice for computer vision tasks. It can learn low-level features and combine them into higher-level features, which enables the network to automatically learn how to perform otherwise difficult tasks such as classification.

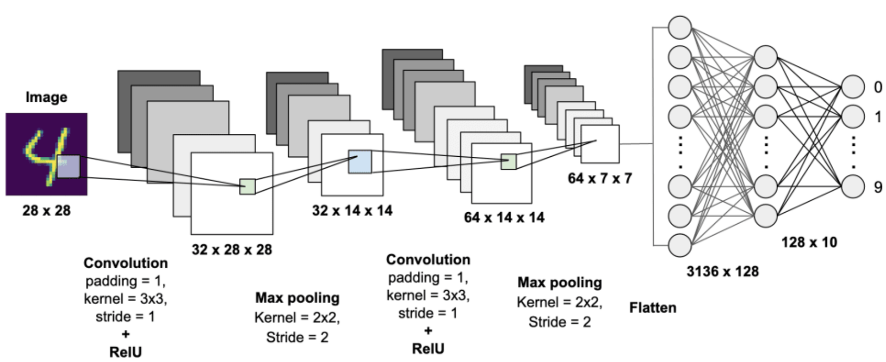

Figure 9: In a standard CNN, an input image undergoes several convolutions until it reaches an FFNN that performs classification (image taken from medium.com[12])

The simpler FFNN often struggles to capture the symmetries of the classification problem unless provided with an enormous amount of data. For instance,

[12] `https://medium.com/analytics-vidhya/introduction-to-convolutional-neural-networks-c50f41e3bc66`

an object's position, orientation, and scale should not, in principle, affect the model's output. By employing kernels, the CNN requires significantly fewer parameters and data and is easier to train in this type of problem.

Fig. 9 shows the inner structure of a CNN. A series of kernels is applied to the image to reduce its size and extract meaningful patterns. The output of each convolution gets passed down to the following convolutional layer.

Alternating convolutional and pooling layers helps reduce the input size during propagation. The final part of the CNN is usually a fully connected NN, as the relative position of the pixels gets lost in the abstraction process of the convolutional part of the network.

4.4 Transformer

The Transformer[13] is a powerful neural network architecture introduced in the "Attention Is All You Need" paper by Vaswani et al. (2017). It has gained significant popularity and has become a standard model for Natural Language Processing (NLP) tasks.

This model takes as input a tokenized version of some text, where each token might represent a character or a word. The goal of the model is to guess the following token of the text. The output is a distribution of probability over all the possible tokens. One of the most likely tokens gets added at the end of the input text. This process is repeated iteratively until the model decides to terminate the generation. Such a model tends to forget the context at the beginning of the text, so an attention mechanism ensures the relevant information is not lost.

The flexibility and adaptability of transformers make them a promising tool in astronomy, enabling more efficient and accurate analysis of astronomical data, automated processing of large-scale surveys, and discovery of new phenomena or patterns.

4.5 Residual Neural Network

A ResNet[14] is a type of neural network that uses residual connections (also known as shortcut, identity, or skip-connections, shown in fig. 10) that allow the network to propagate gradients more effectively and helps address the problem of vanishing gradients (the tendency of the gradient to suffer underflow and therefore be approximated to zero, resulting in no update in some parts of the network). These connections allow the network to learn residual mappings, which are the differences between the desired output and the current output of a layer. By learning these residuals, the network can focus on learning the fine-grained details rather than trying to learn the complete mapping directly. ResNet architectures typically comprise a series of residual blocks consisting of

[13] Source: `https://en.wikipedia.org/wiki/Transformer_(machine_learning_model)`

[14] Source: `https://en.wikipedia.org/wiki/Residual_neural_network`

convolutional layers and skip connections. The fig. 10 shows an example of an input x being propagated past some layers *F(x)*.

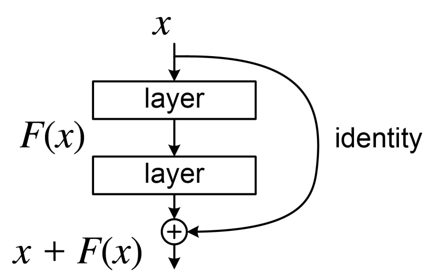

Figure 10: Skip-connection in a Residual Neural Network, the input x bypasses the two layers $F(x)$ and gets passed forward (image taken from Wikipedia[15])

ResNet has been widely used and achieved state-of-the-art performance in various computer vision tasks, including image classification (with the ResNet-101), object detection (Faster R-CNN and Mask R-CNN are two ResNet-based architectures that serve this purpose, as shown later), and image segmentation. The residual connections have proven to be effective in enabling the training of very deep networks and improving their overall accuracy and convergence speed.

4.6 Support Vector Machine

The Support Vector Machine[16] (SVM) is a supervised learning model for classification and regression tasks. It aims to find an optimal hyperplane that splits the input space into different classes based on training examples, and it maximizes the width of the gap between the classes in the feature space. SVMs are non-probabilistic binary linear classifiers and are considered robust prediction methods. On top of that, the kernel trick[17] allows an SVM to efficiently perform non-linear classification by implicitly mapping inputs into high-dimensional feature spaces. In addition to classification, SVMs are often applied to unsupervised learning tasks using support vector clustering. This algorithm utilizes the statistics of support vectors to categorize unlabeled data by finding natural clustering patterns and mapping new data according to these clusters.

A simple example where SVMs have been used to great effect is the problem of spam filtering. This task requires that the model decides whether an email is legit or spam. The trained model takes as input the words in the email and should, in principle, be able to correctly classify instances it has never seen before.

[15] Source: `https://en.wikipedia.org/wiki/Residual_neural_network`
[16] Source: `https://en.wikipedia.org/wiki/Support_vector_machine`
[17] Source: `https://en.wikipedia.org/wiki/Kernel_method`

4.7 R-CNN

R-CNN (Region-based Convolutional Neural Network) is an object detection algorithm. Unlike earlier similar algorithms that rely on the sliding window technique (see CNN) to find the objects, R-CNN uses selective search to propose regions that might contain an object. These regions are then passed through a CNN to compute a feature vector individually. This vector is then fed into a set of support vector machines (SVMs) to classify the object and predict its bounding box. R-CNN was one of the early successful attempts at using deep learning for object detection.

R-CNN can be slower in inference compared to other object detection models due to its multi-stage approach of generating region proposals and performing classification. The training process of R-CNN can be time-consuming and computationally intensive, and the algorithm has since been surpassed by faster and more accurate ones, like the next ones.

4.8 Fast R-CNN

The Fast R-CNN is also a deep-learning model for object detection. It is an improvement over the R-CNN model that has just been discussed. It improves upon the R-CNN by performing the feature extraction step more efficiently. Instead of extracting features from each region proposal individually, the entire input image is passed through the CNN to generate a feature map (the resulting output after the convolutional layers have performed feature extraction on the input image), and the region proposal step is integrated into the network itself. Then, each region proposal is associated with a fixed-size window, and features are extracted from this window (or bounding box) using the RoI pooling technique. This allows the features to be shared across all region proposals, making the process faster and more memory efficient. Finally, a set of SVMs are trained to classify each region proposal as containing or not an object and to predict the object's bounding box coordinates.

Like R-CNN, Fast R-CNN can also be slow during inference due to the multi-stage process. The need for region proposal generation can limit its real-time performance. Also, the need for region proposal generation can limit its real-time performance.

4.9 RoI pooling

RoI pooling (short for Region of Interest pooling) is an algorithm used to perform the down-sampling of an image. The whole image is subdivided into a grid of a desired number of squares, and max pooling is performed on each sub-region. The pooled values are then combined to obtain a fixed-size feature map.

This technique is used on region proposals to obtain the fixed-size input to feed into a classifier. The max-pooling captures the most salient features within each sub-region and helps to preserve spatial information.

4.10 Feature Pyramid Network

A Feature Pyramid Network (FPN) is an architecture commonly used in computer vision tasks, particularly object detection and semantic segmentation. This architecture enables multi-scale feature extraction by combining feature maps from different levels in a CNN. It improves the performance of tasks like object detection and semantic segmentation, particularly when dealing with objects at different scales in an image. It addresses the challenge of effectively detecting objects at various scales in an image.

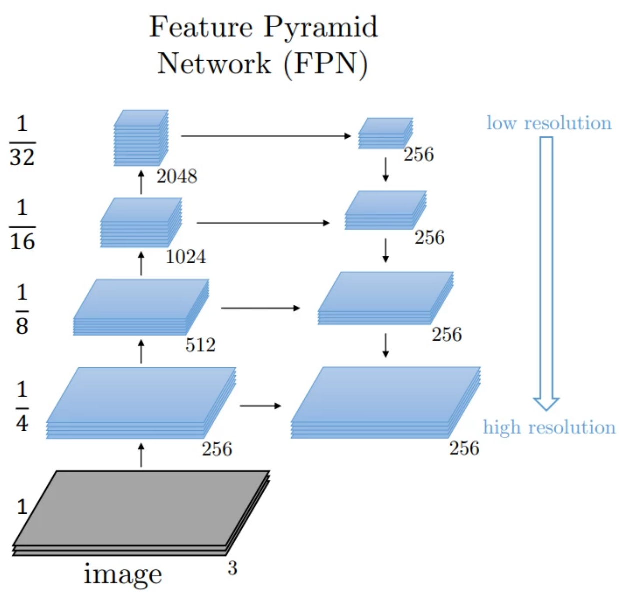

Figure 11: Feature Pyramid Network. The numbers on the left represent the fraction of the size of the layers compared to the input image. The numbers below each group of layers represent the number of layers in that group. (image taken from hasty.ai[18])

The FPN architecture (fig. 11) is designed to extract and leverage feature maps at different spatial resolutions or scales. It consists of a bottom-up pathway (on the left, going deeper into the network, the number of convolutional layers increases while their size decreases) and a top-down pathway (on the right, going deeper into the network, the number of convolutional layers decreases,

[18]Source: `https://hasty.ai/docs/mp-wiki/model-architectures/fpn`

while their size increases). The bottom-up pathway is typically a CNN that processes the input image to extract features at multiple levels, with each level representing a different scale or level of detail. The top-down pathway involves up-sampling the higher-resolution feature maps and merging them with the corresponding lower-resolution feature maps from the bottom-up pathway. This creates a feature pyramid that retains high-level semantic information and fine-grained spatial details. By combining feature maps at different scales through skip-connections (also called lateral connections, represented by the horizontal arrows in the figure), the FPN enables object detection and segmentation algorithms to detect objects of different sizes and effectively handle scale variations. The top-down connections help propagate contextual information from higher-resolution levels to aid in accurate localization and recognition of objects. The feature maps from a particular level (or a combination of those from different layers, depending on the complexity and scale of the objects you are interested in) are then used to extract the output from the model. Upscaling can be used to match the size of the input image. The next model makes specific use of this architecture.

4.11 Faster R-CNN

This is currently a state-of-the-art model used for object detection. The main component of the model is a Region Proposal Network (RPN) that proposes regions of interest (RoIs) for the Fast R-CNN to analyze later. Faster R-CNN uses a two-stage approach to object detection.

In the first stage, the RPN generates a set of bounding box proposals that are likely to contain an object. These proposals are generated using a sliding window approach on the feature maps generated by the convolutional layers of the network. The RPN predicts the probability of an object being present in each proposal, as well as the offsets that need to be applied to the proposal to obtain an accurate bounding box.

The second stage classifies the RoIs using a Fast R-CNN network. In this stage, the Fast R-CNN network takes the RoIs generated by the RPN and extracts features from each RoI using a Region of Interest pooling layer. These features are then fed into a fully connected network for classification and regression of the bounding box coordinates.

The Faster R-CNN model is computationally efficient because the RPN shares the convolutional layers with the Fast R-CNN network, reducing the computation needed to generate proposals. It also achieves state-of-the-art performance on several object detection benchmarks.

Faster R-CNN still relies on region proposal generation, which can be a bottleneck during inference. Faster R-CNN can also struggle with accurately detecting small objects due to the fixed anchor sizes used by the RPN.

4.12 Mask R-CNN

Mask R-CNN extends the Faster R-CNN model by adding a pixel-level segmentation capability. This means that it can classify each pixel that belongs to any given object in the image. This model combines the capability of object detection, classification, and instance segmentation on images. In other words, it can accurately detect objects and generate segmentation masks for each object instance.

Mask R-CNN can be slower than other object detection models due to its additional task of generating pixel-level segmentation masks. Training Mask R-CNN typically requires more computational resources and longer training time than other models.

4.13 U-Net

The U-Net (fig. 12) is a type of CNN architecture designed for biomedical image segmentation tasks, although it has also been applied in other domains. The U-Net follows an encoder-decoder structure where the image dimensions are successively reduced in the encoding path, detecting low-level features and progressively increasing the number of channels. These channels serve as feature detectors, capturing both low-level and high-level features. In the decoder path, the feature map is expanded back to its original size, allowing for precise image segmentation. The contracting path consists of repeated applications of convolutional and max pooling layers to reduce the spatial resolution of the input. In contrast, the expanding path uses transposed convolutional layers to increase the spatial resolution and generate the final segmentation mask.

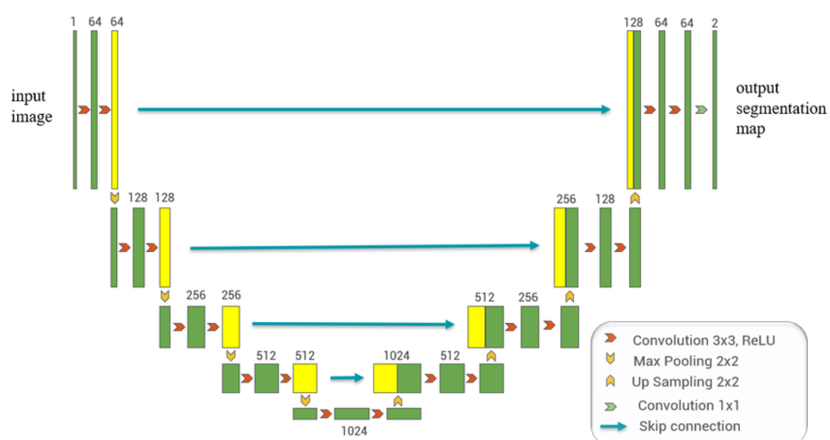

Figure 12: Visualization of the U-net architecture. The number above each rectangle represents the number of convolutional layers in that block (Guo et al., 2020)

Like with the FPN, the U-Net also includes skip connections between corresponding layers in the contracting and expanding paths, allowing the network to combine low-level and high-level features for accurate segmentation. On top of that, unlike the Mask-R-CNN, the output of the U-Net does not include bounding boxes. This means that the U-Net does not require bounding boxes as part of the dataset in the first place.

U-Net is primarily designed for image segmentation tasks and may not be as effective for object detection as other models. It may need help detecting objects in complex or cluttered scenes where precise localization is required.

4.14 Training

Training a model involves iteratively updating its parameters to minimize a chosen loss function based on its performance in predicting the correct output from input data.

The network's output is compared to the ground truth labels using a chosen loss function. The loss function measures the discrepancy between the predicted output and the true labels and quantifies the network's performance. The next step is to compute the gradients of the loss function concerning the model's parameters. This is done using the backpropagation algorithm. Backpropagation calculates the gradient of the loss function by recursively applying the chain rule, starting from the final layer and moving backward through the network.

The model's parameters are updated using the gradients computed to minimize the loss function. This is typically done using an optimization algorithm such as Stochastic Gradient Descent (SGD). SGD adjusts the parameters in the direction opposite to the gradient, scaled by a learning rate. Other optimization algorithms, such as Adam, AdaGrad, or RMSprop, can also improve convergence speed and handling of different learning rates for different parameters. Each method has strengths and weaknesses, and deciding the most suited one is often a matter of trial and error.

4.15 Supervised and unsupervised learning

Supervised and unsupervised learning are two distinct approaches in the field of machine learning. These two approaches are not mutually exclusive and often complement each other, with unsupervised learning assisting in feature extraction or anomaly detection and supervised learning utilizing the insights gained from unsupervised learning for improved performance.

Supervised learning relies on labeled training data, where the algorithm learns to map input features to corresponding target labels. It excels in tasks requiring precise predictions or classifications but relies on a labeled dataset. The evaluation of supervised learning models can be based on well-defined metrics using the known target labels.

On the other hand, unsupervised learning operates on unlabeled data and aims to discover hidden patterns or structures. It is well suited for exploratory data analysis, clustering, and dimensionality reduction tasks. It can leverage readily available unlabeled data, making it more adaptable to scenarios with limited labeled samples or none whatsoever. Evaluation of unsupervised learning models can often be challenging due to the absence of ground truth labels.

4.16 Transfer Learning

Transfer learning is a powerful technique in machine learning that enables the transfer of knowledge from a pre-trained model to a related yet different task. This approach alleviates the challenges associated with training models from scratch by leveraging the learned representations and features captured by the pre-trained model. The main idea behind transfer learning is that the knowledge and representations learned by a model on a source task can be beneficial for solving a target task, even if the datasets or tasks are different. Instead of training a model from scratch on the target task, transfer learning allows us to start with a pre-trained model, which has already learned meaningful features from a large dataset, and fine-tune it on the target task with a smaller dataset.

4.17 Overfitting

Overfitting occurs when a machine learning model becomes excessively complex and starts to fit the training dataset too closely, resulting in poor generalization to unseen data. This happens when, instead of learning the patterns in the data, the model memorizes specifically the provided data points, resulting in high training accuracy but poor performance on validation or test data.

5 Literature review

This section will provide a comprehensive review of the chosen representative sample of works. The sample spans various datasets from radio to optical and gamma-rays, captured with the instruments presented in Section 3. It also includes the algorithms and techniques introduced in Section 4.

Burke et al. (2019) use Mask R-CNN to perform the deblending and classification of sources in infrared and optical bands. The authors mention several existing source detection and classification algorithms, emphasizing the need for efficient deblending algorithms in future surveys. The recent advancements in deep learning, multiband imaging, and spectral energy distribution information offer promising avenues for improving source classification, deblending, and photometric redshift estimation.

The dataset described in the article consists of simulated images of crowded extragalactic fields generated using PhoSim (the Photon Simulator software, a flexible open-source simulation software based on theoretical models of optics and on the knowledge of scientific instruments). The purpose of using simulated images is to overcome the limitations of real observational data, such as the lack of a sufficiently sized and unbiased training set. Simulated images provide a large training set with a known catalog of stars and galaxies, which serves as a ground truth for evaluating the network's performance. The simulations use the DECam (Dark Energy Camera) instrument options adapted from previous studies. The dataset has also been standardized and augmented.

In this work, the authors develop a new deep-learning method for classifying and deblending sources in astronomical images that can efficiently perform all source detection, classification, and deblending tasks. The network is robust to mask overlaps, resulting in clean deblends for significantly blended sources. The new deep learning method is based on the Mask R-CNN framework, which aims to perform source detection, classification, and deblending within a single machine learning framework. The Mask R-CNN framework is highly efficient and robust to occlusion (when two sources partially overlap).

The authors evaluate the network's performance using the average precision (AP) score metric and show the precision and recall curves for stars and galaxies. The network performs well at moderate IOU thresholds (i. e. the area of the intersection of the predicted and ground truth masks divided by the area of the union of the predicted and ground truth masks, see Appendix).

They measure a precision of 98% at 80% recall for galaxies in a typical field with about 30 galaxies/arcmin^2 with an IOU threshold of 0.5 and a minimum detection confidence threshold of 0.5. On the other hand, for stars in a typical field, they measure a precision of 92% at 80% recall. The authors plan to improve the framework by adding the detection of several features, including noise and contamination caused by cosmic rays and bleed trails. They also plan on exploring different architectures and using new data, such as DES or HSC deep coadd images.

An innovative framework, AutoSourceID-Light (ASID-L), is presented in Stoppa et al. (2022). It uses computer vision techniques that can naturally deal with enormous amounts of data and rapidly localize sources in optical images.

The dataset consists of wide-field optical images taken with the MeerLICHT telescope. The location of the sources in the images is provided by the Gaia Early Data Release 3 catalog.

The authors show that the ASID-L algorithm, based on U-Nets and enhanced with a Laplacian of Gaussian filter, performs localization of sources remarkably well. The U-Net is trained on pairs of optical images and their corresponding known segmentations (masks and images of continuous values in the range [0, 1]). The network learns to predict the segmentation mask based on the input image. The network also takes the Laplacian of the Gaussian-convolved image as input, making the sources' precise localization more accurate. Interestingly, they exploit the fixed size to find the standard deviation for the Gaussian kernel that best suits the problem.

The result of this work is a clear framework that can significantly increase the speed and accuracy of optical source localization, even in crowded fields. Another benefit of this pipeline is that it deals very well with cosmic rays, diffraction spikes, artificial trails, and especially faint sources. The authors suggest extending the model's capabilities to different resolutions and parts of the EM spectrum. The authors also suggest that the model can be trained from scratch on a different dataset of optical images and still perform well.

Panes et al. (2021) aim to identify point sources using the same base architecture in gamma rays. In this work, they developed an automatic deep-learning image segmentation pipeline that localizes and classifies gamma-ray point sources. Their work demonstrates that these two tasks can be accurately and robustly solved using neural networks.

The training data set is created by simulating the Fermi-LAT data, which consist of 9.5 years of exposure from Fermi-LAT, with active galactic nuclei (AGNs), pulsars (PSRs), and several models of background interstellar emission used as source properties. It includes hundreds of simulated sky instances based on the 4FGL Fermi-LAT catalog, considering the contributions of both point sources and diffuse emission. Each image consists of time-integrated photon number counts in pixels representing spatial and energy space.

This particular model consists of two primary components. The first is a U-shaped convolutional network (U-Net), while the second is a deep neural network trained to perform classification. The localization algorithm's output is fed into the classification neural network. In addition, k-means is utilized for source clustering and localization, along with the Centroid-Net algorithm for object counting.

The authors show that sources have "purity" and "completeness" higher than 90%. Regarding flux sensitivity, the model's performance is at least as good as that of the traditional algorithms while having the strong advantage of being robust concerning the diffuse emission across the range of models they tested. The authors also separately tested the classification neural network, which turned out to be capable of separating the sources.

The same authors continue improving this work, as recently presented in van den Oetelaar et al. (2021). In these proceedings, the authors discuss their progress in building a more robust data analysis pipeline using a more precise simulated dataset and including more source classes and time variability of blazar-type sources (BLL and FSRQ), again using the U-Net architecture.

Similarly to the previous work, this scientific publication demonstrates the use of deep learning algorithms to detect and classify gamma-ray point sources, which can be applied to Fermi-LAT data and can potentially generalize to other parts of the electromagnetic spectrum. The pipeline uses the U-Net structure for precise source localization. It also introduces a robust training data generation method to exploit the full detector potential. The generated inputs consist of six images per sample, each representing the same portion of the sky but at different wavelengths and with different resolutions, complete with the position of the sources.

The U-Net aims to recreate the mask consisting of five-pixel radius disks that localize the sources in the image. As it has already been discussed, the earlier stages of the network take as input the higher resolution images, while the lower resolution images are fed later. A different network performs classification on the detected sources. It is a simple CNN with max-pooling layers.

The authors decided to round the output of the classification network to the closest integer (0 or 1) and compute the accuracy based on that value. This results in a binary accuracy of the network of 98.2%. However, it is worth noting that the images are mostly comprised of background pixels. A prediction of only background would score an accuracy of 96.5%, which is lower than the previous one. The authors also remind us that this is a work in progress. The authors' goal is to see if and how adding the time variability parameter to the classification algorithm through adding new data will affect classification capabilities. They also plan to apply the localization and classification method to the real Fermi data.

Both Panes et al. (2021) and van den Oetelaar et al. (2021) investigated using NNs for source detection by implementing the U-Net architecture on Fermi-LAT data. The U-Net architecture allows object detection and semantic segmentation only. Their method showed promising potential, though substantial limitations were noted, especially for faint sources and crowded regions.

A recent effort (Horangic et al., 2022a) has started investigating Fermi-LAT gamma-ray simulated data using the more promising and efficient Mask-R CNN, which can be used for instance segmentation. This is the first attempt to use the Mask R-CNN architecture on Fermi-LAT data. This method can potentially give more successful results than usual source detection and classification methods. As reported in their work, it would allow them to distinguish between sources and other emissions, even when sources are at the limit of the Fermi-LAT sensitivity or they are background dominated, and when sources are too close together to be detected as separate sources with the standard methods.

They have generated training and test datasets of simulated Fermi LAT images with different parameters, such as background noise and source photon count. Images are collected in five distinct energy ranges in gamma rays. The authors created a package that allows users to create and train an R-CNN. It can be trained on simulated data and can generalize on real data. They use a multi-task loss to simultaneously optimize the model for classification, localization, and segmentation, which are the task that the model aims to perform.

Horangic et al. (2022a) present their Python package, fermidetect, which is based on Detectron2 (Meta's open-source deep learning framework for object detection and instance segmentation) and uses neural networks to predict data simulated across five energy bins in the range of gamma rays. They plan to use this package to test the performance of different algorithms and hyperparameters on simulated LAT data and, ultimately, to detect sources in real LAT data.

The work is still in progress. Preliminary results have been presented in Horangic et al. (2022), where the authors have evaluated the performance of the model by using a small dataset. They obtain recall = 87% (see Appendix on statistics) and precision = 100%. This means that, even with their preliminary small dataset, not even one source was missed, and only a few sources were spurious.

In an earlier work with radio data, Riggi et al. (2023) proposed a new model, the CAESAR Mask R-CNN, capable of both detecting and classifying compact, extended, spurious, and poorly imaged sources in radio images. The model is implemented in Python using the TensorFlow and Keras libraries.

The datasets used to train the model are the ASKAP pilot surveys, the ATCA Scorpio survey, and the Radio Galaxy Zoo survey. In each image, the significant sources are located and labeled. Data contains, among other instances, multi-island sources and imaging artifacts, which make the problem even more challenging.

The complete algorithm is the most complex out of every paper in this study, as it is an ensemble of a Region Proposal Network, a Mask Branch that computes the segmentation mask, a feature extraction step, an ROI alignment step, a box regression step, and a network (ResNet) for classification. As discussed in the previous section, Mask R-CNN is simultaneously capable of performing object detection, classification, and segmentation in images. The team uses a ResNet-101 model for classification. They suggest training the model from scratch instead of using the default parameters.

The authors have extensively tested the pipeline. They found "completeness" of about 90% for compact sources, with a modest 60% reliability. Performance on extended sources is lower, at around 80%. The model especially needs improvement on multi-island sources.

The team plans to extend the dataset with new data and augment it with the help of generative adversarial networks. They also plan to address the poorer performance on extended objects by considering an alternative deep learning framework. In the longer term, they also plan to extend the number of types of sources that the model is capable of classifying.

Using the same radio telescopes, very recently, Sortino et al. (2023) have explored applications of deep learning in radio astronomy, specifically for automatic object detection and instance segmentation. The authors evaluate and compare the performance of different state-of-the-art deep learning approaches on radio interferometric images to provide insights to astrophysics researchers who want to use machine learning in their work.

The authors use radio astronomical images taken with the Australian Telescope Compact Array (ATCA), the Australian Square Kilometer Array Pathfinder (ASKAP), the Very Large Array (VLA), and the Radio Galaxy Zoo (RGZ) project. The data labels extended sources, compact sources, and imaging artifacts, which are the classes that the model in this paper aims to predict.

The authors explore the performance of many different architectures. For source detection tasks, Mask R-CNN, Detectron2, YOLO, EfficientDet, DETR, and YOLOS. The paper well documents the exact figures, but broadly speaking, the study found that DETR and Yolo achieved the highest F1 score (see Appendix) in compact and extended sources, respectively. Yolo achieved the highest reliability, while Detectron2 achieved the highest completeness and best overall score.

For image segmentation tasks instead, the authors consider a simple encoder-decoder architecture, a U-Net, Tiramisu, and PankNet. U-Net and Tiramisu achieved good performance with both compact and extended sources, with Tiramisu especially excelling with a high signal-to-noise ratio.

It is also worth mentioning the recent proposal of Diego-Palazuelos et al. (2023) about a machine learning approach to the detection of point sources in maps of the CMB temperature anisotropies. The authors train the neural network using simulations of the microwave sky and divide it into regions based on increasing levels of light from the Milky Way. They preliminarily found that their approach was more successful at detecting sources close to the plane of our galaxy than traditional detection methods.

Data is generated through simulations of the CMB at 143 GHz based on data provided by the Planck mission. They consider both instrumental noise and galactic and extragalactic foregrounds. The authors use the Planck Sky Model-1 to simulate the synchrotron radiation, thermal and spinning dust emissions, free-free diffuse Galactic emissions, and the extragalactic thermal and kinetic Sunyaev-Zeldovich effects.

The authors use supervised learning with binary labels to train independent simple CNNs that specialize in each region of the sky. They also point out that this approach might result in overfitting if the higher number of models is not met with an appropriate increase in the amount of training data.

The model's performance is estimated with various degrees of the light flux of the simulated images. As shown in the paper's results, the model requires much less flux to achieve high completeness. The authors plan to solve some issues with overfitting and extend the training to other frequencies.

Table 1 compares the scientific publications discussed. For each work it provides the name of the publication details, some tags that characterize the dataset used, the task that the authors aim to solve, and the model that they use.

Author and title	dataset	task	model
Burke et al. (2019) Deblending and classifying astronomical sources with Mask R-CNN deep learning	visible DECam images simulated with PhoSim, crowded extragalactic fields 3 energy bins	detection segmentation deblending	Mask R-CNN
Stoppa et al. (2022) AutoSourceID-Light - Fast optical source localization via U-Net and Laplacian of Gaussian	visible wide-field images Gaia, MeerLICHT telescope	detection localization segmentation	U-Net, AutoSourceID-Light ASID-L framework
Panes et al. (2021) Identification of point sources in gamma rays using U-shaped convolutional neural networks and a data challenge	gamma rays Fermi LAT 5 energy bins	detection classification localization clustering	U-Net + CNN, k-means for source clustering and localization, Centroid-Net algorithm for object counting
van den Oetelaar et al. (2019) Localisation and classification of gamma ray sources using neural networks	gamma rays, potentially to other wavelengths Fermi-LAT 6 energy bins	detection classification localization clustering	U-Net
Horangic et al. (2023) Investigating Detection of Sources in Fermi Gamma-ray Data with Neural Networks	gamma rays Fermi LAT satellite 5 energy bins, different resolution	detection classification	fermidetect, flexible neural architecture, based heavily on Meta AI's Detectron2 DL framework
Riggi et al. (2019) Astronomical source detection in radio continuum maps with deep neural networks	radio galaxy zoo ATCA Scorpio survey ASKAP Pilot survey	detection classification localization	CAESAR Mask R-CNN, main components are an RPN, CNNs, and ResNet
Sortino et al. (2023) Radio astronomical images object detection and segmentation: A benchmark on deep learning methods	radio ATCA, ASCAP, VLA, RGZ	detection segmentation	Mask R-CNN, Detectron2, YOLO, EfficientDet, DETR, YOLOS Encoder-decoder, U-Net, Tiramisu, PankNet
Diego-Palazuelos (2023) Machine learning approach to the detection of point sources in maps of the CMB temperature anisotropies	microwave, CMB at 143 GHz simulations of the microwave sky based on Planck	detection	CNN

Table 1: This table summarizes the details of the works used this review. The type of data, problem, and AI model are provided for each one.

6 Discussion and Conclusions

This thesis explores the state-of-the-art regarding the intersection of machine learning and astronomy, focusing on applying machine learning models to analyze data from current and future astronomical observatories enabling big data analyses. This thesis also reviews the basic machine learning concepts required to understand the current literature on this topic. Eventually, by analyzing the current state of research, this study identified some key insights and trends that highlight the impact and potential of machine learning in astronomy.

This work reviews a sample of recent studies on source detection and classification in astrophysics based on machine learning models, including R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN. Even though the sample is not exhaustive, it is representative of the state of the art in this field, covering data from the entire electromagnetic spectrum and t most updated models. Indeed, the models used in these studies are designed for various objectives and are trained on different datasets. For this reason, this thesis focuses more on comparing the approach of the authors rather than the models themselves. From this review, CNN methods appear very promising in source detection, localization, and classification in astronomy. However, this field is still young. The main limitation that all the models in this study have in common is that they have been trained on specific portions of the electromagnetic spectrum, which significantly restricts the selections of datasets that the models can be applied to. On top of that, the fully trained models might be accurate when dealing with data from the same observatory they have been trained on, but their performance is in question as soon as they are transferred to a different dataset. Of course, testing the model on a variety of datasets is essential to evaluate its robustness.

This chapter brings attention to the concept of transfer learning. This powerful machine learning technique enables knowledge transfer from a pre-trained model to a related yet different task. It alleviates the challenges associated with training models from scratch by leveraging the representations and features captured by the pre-trained model. The main idea behind transfer learning is that the knowledge acquired by a model on a source task can be beneficial for solving a target task, even if the datasets or tasks themselves are different. Instead of training a model from scratch on the target task, transfer learning allows us to start with a pre-trained model, which has already learned meaningful features from a potentially large dataset, and fine-tune it on the target task with a smaller one.

A technique that every paper in this study might benefit from is explainable AI. It aims to provide transparency and interpretability to the decision-making processes of deep learning models. For example, a common approach is to highlight the regions of the image that mainly contribute to the decision made by the model. This method might prove useful in astrophysics to understand how and why certain predictions or classifications are made, aiding in the validation and interpretation of results. Importantly, it can highlight whether classification has been made based on the actual source in the picture or whether the model has picked up some pattern from somewhere else in the image, for example, from the background. This difference is crucial when transferring the model (especially those trained on real data) to a different dataset, where such spurious patterns no longer hold, and the classification quality might suffer.

7 Appendix on statistics

In the context of classification, it is common to evaluate the performance of a model based on four factors (displayed in fig. 13). The true positive/negative rates (TP/TN) represent the portion of cases when the model correctly predicts a positive/negative outcome. The false positive/negative rates (FP/FN) are the fraction of cases the model erroneously identifies as positive/negative.

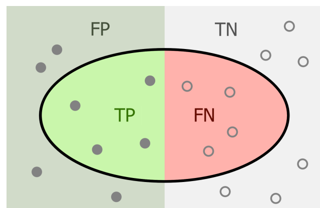

Figure 13: True positives, false positives, true negatives, false negatives

The table 2 contains these and some other important quantities.

$true\ positive\ ratio = TP$	$false\ positive\ ratio = FP$
$true\ negative\ ratio = TN$	$false\ negative\ ratio = FN$
$accuracy = T = TP + TN$	$inaccuracy = F = FP + FN$
$positives = P = TP + FP$	$negatives = N = TN + FN$
$real\ positives = P^* = TP + FN$	$real\ vegatives = N^* = TN + FP$
$precision = \frac{TP}{P}$ *	$(no\ name) = \frac{TN}{N}$
$recall = \frac{TP}{P^*}$ **	$efficacy = \frac{TN}{N^*}$
$F_1 = \frac{2\,TP}{2\,TP + F}$	$IOU = \frac{TP}{1 - TN}$

** Also called reliability, purity ** Also called completeness*

Table 2: Some important statistical quantities relevant for the study

References

Colin J Burke, Patrick D Aleo, Yu-Ching Chen, Xin Liu, John R Peterson, Glenn H Sembroski, and Joshua Yao-Yu Lin. Deblending and classifying astronomical sources with mask r-cnn deep learning. *Monthly Notices of the Royal Astronomical Society*, 490(3):3952–3965, 2019.

P Diego-Palazuelos, RB Barreiro, P Vielva, D Balbás, M López-Caniego, D Herranz, and B Casaponsa. Machine learning approach to the detection of point sources in maps of the cmb temperature anisotropies. *arXiv preprint arXiv:2302.14724*, 2023.

Yanan Guo, Xiaoqun Cao, Bainian Liu, and Mei Gao. Cloud detection for satellite imagery using attention-based u-net convolutional neural network. *Symmetry*, 12(6):1056, 2020.

Diana Horangic, Elena Orlando, and Andrew Strong. Investigating detection of sources in fermi gamma-ray data with neural networks. In *American Astronomical Society Meeting Abstracts*, volume 54, 2022.

Diana Horangic, Elena Orlando, Andrew Strong, and Saptashwa Bhattacharyya. Detecting fermi lat gamma-ray sources with neural networks. *Proceedings of the 27th European Cosmic Ray Symposium*, 2022a.

Boris Panes, Christopher Eckner, Luc Hendriks, Sacha Caron, Klaas Dijkstra, Guðlaugur Jóhannesson, Roberto Ruiz de Austri, and Gabrijela Zaharijas. Identification of point sources in gamma rays using u-shaped convolutional neural networks and a data challenge. *Astronomy & astrophysics*, 656:A62, 2021.

Simone Riggi, Daniel Magro, Renato Sortino, Andrea De Marco, Cristobal Bordiu, T Cecconello, AM Hopkins, J Marvil, G Umana, Eva Sciacca, et al. A comparative analysis of neural network methods for source detection and classification in astrophysics. *Astronomy and Computing*, 42:100682, 2023.

Renato Sortino, Daniel Magro, Giuseppe Fiameni, Eva Sciacca, Simone Riggi, Andrea DeMarco, Concetto Spampinato, Andrew M Hopkins, Filomena Bufano, Francesco Schillirò, et al. Radio astronomical images object detection and segmentation: a benchmark on deep learning methods. *Experimental Astronomy*, pages 1–39, 2023.

Fiorenzo Stoppa, Paul Vreeswijk, Steven Bloemen, Saptashwa Bhattacharyya, Sascha Caron, Guðlaugur Jóhannesson, Roberto Ruiz de Austri, Chris van den Oetelaar, Gabrijela Zaharijas, Paul Groot, et al. Autosourceid-light. fast optical source localization via u-net and laplacian of gaussian. *arXiv preprint arXiv:2202.00489*, 2022.

Chris van den Oetelaar, Saptashwa Bhattacharyya, Boris Panes, Sascha Caron, Gabrijela Zaharijas, Roberto Ruiz de Austri, and Guðlaugur Jóhannesson.

Localisation and classification of gamma ray sources using neural networks. *37th International Cosmic Ray Conference*, 2021.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.