UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

COS700 Research Report

Detecting odd radio circles with Siamese Convolutional Neural Networks

Student number: u17021627

Supervisor(s):

Dr. Anna Bosman
Dr. Kshitij Thorat

Jaime Matthew Tellie

Abstract

Radio astronomy is a field in astrophysics that involves the study of celestial objects that emit radio light rather than optical light. Odd radio circles (ORCs) are recent type of celestial radio sources discovered in 2019 which have a circular/spherical shape and emit electromagnetic synchrotron radiation at radio frequencies. Although they share similar properties to other radio sources, objects like galaxies, black holes, stars, or pulsars, they do not currently have any association to these objects that we know of. This study will apply modern computer vision models for the purpose of classifying ORCs in radio-astronomical data. In particular convolutional neural networks (CNNs) are the core model to computer vision and its associated tasks such as object identification and classification. Although CNNs excel at computer vision tasks, they require large amounts of data as input. Few-shot learning is a technique used to circumvent the lack of data in an attempt to perform the task required with a much smaller dataset. Our approach will attempt to make use of few-shot learning and in particular, a specific type of CNN called the siamese neural network (SNN), in order to identify ORCs.

Keywords:

Convolutional neural networks, odd radio circles, few-shot learning, siamese neural networks, radio galaxies.

I. Introduction

Odd radio circles (ORCs) are recently discovered radio sources that form part of the radio wave spectrum [1] [2]. The class of radio waves emitted from ORCs have not yet been classified [3] as there is very little information [4] regarding this radio source [5]. The name is derived from the fact that ORCs take on a circular form and shape [6]. ORCs were firstly discovered in 2019 by the Australian Square Kilometer Array Pathfinder (ASKAP) radio telescope [2] [5]. Due to its recent discovery, there are only a few known to existence as these objects are extremely rare and form the core concept behind this research project. The scarcity of ORCs means that there is very little that we know about them [7]. Recent studies are still able to derive certain features that ORCs posses. Some of the most recently discovered ORCs show double ring and ear-like features [8]. The current ORCs found in existence stand at 6. With one instance being an ORC pair and the other 5 being single ORCs [8]. In addition to the knowledge we have, the data is equally as limited [5].

Fig. 1. Radio image of an ORC, overlaid with optical background

Objects such as radio galaxies [6], supernovae remnants [7], planetary nebulae and circumstellar disks could aid in the understanding of ORCs [9]. Radio galaxies, as defined for this study, are active galaxies which show presence of large-scale (typically 100s of kiloparsecs) radio jets powered by the central AGN. Supernovae remnants on the other hand, are the remains of supernovae explosions [7] [10] and are vital in the study to better understand stellar life-cycles [9]. Although ORCs resemble supernova remnants, their location at high Galactic latitudes suggests such origin to be unlikely [8]. Planetary nebulae do not have any relation to planets, however, they are dust and gas remnants of a star and have their definition due to early research in astronomy misidentifying them with gas planets. This misidentification could possibly hinder this study of identifying ORCs. Finally, circumstellar disks, also known as protoplanetary disks, are gas, dust and asteroids around a planet and are essential in the formation evolution of planets. These objects constitute a small subset of radio sources that could be misinterpreted with ORCs [7].

Machine learning is a discipline in computer science that attempts to aid computers to make decisions, perform tasks and move towards a state of "self-thinking". This is done by mimicking the biological neuron with an artificial neuron. Neural networks (NN) mimic the brain of biological organisms. NNs consists of artificial neurons and synapses which are designed to process information in a non-linear manner [11]. Computer vision is a domain in machine learning that focuses on image and video processing for the purpose of making decisions [1]. Computer vision models use image data to be trained in order to learn and become familiar with identifying and classifying objects in its world [12]. The main machine learning technique this report will look to investigate, is few-shot learning. The few-shot learning technique was designed to train NNs with limited data to identify the similarity between images. Thus, instead of being able to determine what object is portrayed in an image, the NN will simply be able to distinguish whether or not two or more images are closely related or not.

Convolutional neural networks (CNNs) are heavily intensive NN models that typically operate on very large datasets [13]. Traditional NNs, in particular CNNs, function on datasets with sample size of approximately more than 100 000 instances [13]. Few-shot learning is used in deep learning models with data sets of less than 100 instances. In addition to few-shot learning, data augmentation can be used to enhance a scarce dataset. Data augmentation techniques manipulate data instances in small ways to create new instances that can be added to the dataset. Due to the limited data surrounding ORCs and its unknown properties, these are the main reasons why the above mentioned models [14], few-shot learning and data augmentation will be employed in the research study to be conducted.

The remainder of this report will be layed out in the following structure. Section II deals with the problem statement. This gives a basic outline of the core issue this research will attempt to investigate. Section III provides a literature study with background information to the radio sources to be investigated and the models to be used, whereas Section IV explains the partnership of deep learning in radio astronomy, along with specific references to work done before and knowledge that can be added to this report. Section V, methodology, outlines the manner in which the research will be conducted. It gives more details, regarding models, data and the implementation approach. Finally, Section VI will provide the results and an in-depth analysis of the investigation conducted.

II. Problem Statement

ORCs are radio sources with limited to no information regarding their nature [3], origin [10] and properties [5] [9]. The problem at hand is to help the radio-astronomy community further discover, identify and possibly look at classifying ORCs [6]. With identification comes a subset of problems.

The hardest part of the identification concept is in fact that there are too few odd radio circles found in existence [5] [9]. In order to further study these objects and gain more insightful information, we would need more instances. As ORC identification is the core study to be conducted, the main feature to be considered would be the radio images of these objects at a single frequency [5] [14].

The challenge we may encounter in such an identification problem is the similarity to other radio sources [9]. As stated before, ORCs currently form a circular shape and emit electromagnetic synchrotron radiation at radio frequencies [5]. These properties exist across a range of other radio sources in the universe [9]. A possible solution to this issue could be to consider the size of these objects [6] [7]. The current odd radio circles known in existence can take on a very apparent large size [9], some spanning across multiple galaxies [3] [7]. This could aid in the identification problem. And finally, should the identification problem be a success, we could look towards investigating the nature of these sources even further [3] [5]. One of the next steps could even lead to classification of different types of odd radio circles [6]. As such, this report will aim to investigate the effectiveness of the use of siamese neural networks as a few-shot learning method for classifying radio stellar objects. That is, can it identify specific radio galaxies from each other, can it perform ORC classification amongst a set of radio galaxies, and whether the methods to be set out can be shared through transfer learning to other radio astronomy classification tasks.

III. Background

The following section will share background information regarding various different radio astronomy topics and objects of interest. In addition to radio astronomy, it will also outline some fundamental concepts regarding machine learning, its associated models and their applications.

A. Radio Astronomy

The radio astronomy section will provide more information regarding the most important of our radio stellar objects, radio galaxies, and the core object being ORCs. It will further explain the difficulty in identifying ORCs from other radio objects, along with their basic properties.

1) Radio Objects: As scientists are still studying ORCs and trying to get a better understanding of their nature [5], the most logical conclusion would be to try and identify its overlapping properties [14] with similar objects known to existence [13]. To date, the most common objects that are closely used in the investigative studies of ORCs are supernovae remnants [7] [10], planetary nebulae, circumstellar shells, protoplanetary discs and lastly radio galaxies [2] [9]. Figure 2, shows a optical illustration of Cassiopeia A, a very well-known supernova remnant. From the figure, it is clear how to

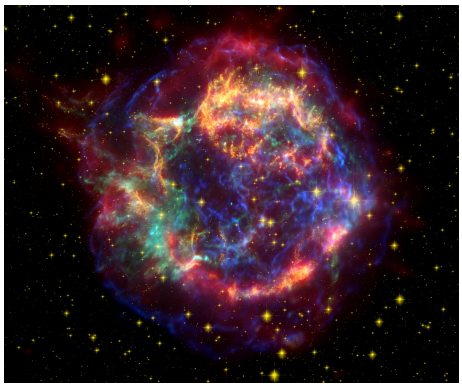

Fig. 2. "Cassiopeia A, a supernova remnant. Note the circular morphology very similar to ORCs."

see how it possesses a circular shape mimicking that of an ORC [9]. Radio galaxies in particular might prove to be the objects of most interest, as shown in other studies [5] [7]. Radio galaxies are active galaxies that emit in the radio part of the electromagnetic spectrum [7], typically via the synchrotron emission [14] produced from electrons that are accelerated by magnetic fields [15]. Figure 3, shows Cygnus A, the most well-studied radio galaxy discovered. Other influential factors to be considered in the study of ORCs are their surrounding environment [5] [7]. From studies conducted we have noticed that ORCs span across hundreds of light years [3] [7]. Factors like this help in the study of these objects as it can potentially answers [3] the questions regarding their nature and origin [1] [5]. As mentioned before, ORCs are only visible in radio wavelengths [7]. There are however, some surveys that serve as the driving point for future studies. Surveys done through the ASKAP [5] and EMU-PS have shown great insight into the initial investigative phases [6] and surveys like the MeerKAT Galaxy Cluster Legacy Survey (MGCLS) have been able to image ORCs.

2) Odd Radio Circles: ORCs can be viewed through infrared imagery as it possesses radio properties [14], but also through radio images. Recently, a polarization study of one ORC using the MeerKAT radio telescope [5] revealed tangential magnetic field along the ring periphery, suggesting the ring as an expanding spherical

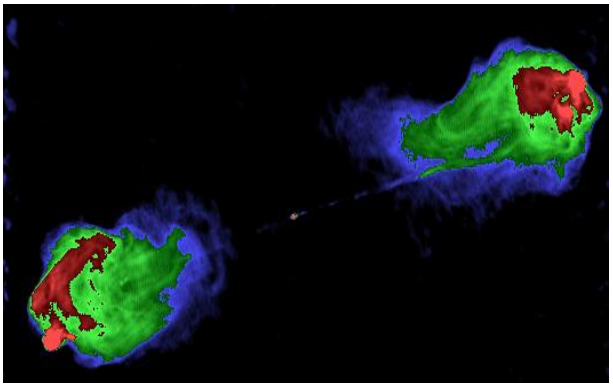

Fig. 3. Cygnus A

shock wave similar to that in the expanding supernova remnants [2]. Presently five sources have been identified with morphologies corresponding to those of ORCs. The geometrical centers of three ORCs seem to have a galaxy as the optical counterpart at photometric redshifts between 0.27 and 0.55. The optical colors of these galaxies suggest that the galaxies are quiescent and presently not forming stars. If ORCs are actually associated with these galaxies, the physical sizes of ORCs will be several hundred kiloparsecs. Two ORCs are without optical counterparts and separated by 1.5 arcmin from each other. A potential sixth ORC was recently identified as a supernova remnant in the outskirts of the Large Magellanic Cloud [2] [9]. Due to

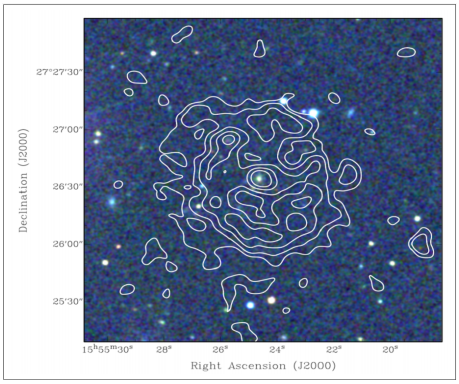

Fig. 4. Image of an ORC from ASKAP survey

their detections at high galactic latitudes (above or near +/-40 degree) with an estimated density of 1.6 sources per 100 square degrees [7], ORCs cannot belong to the classical population of galactic supernova remnants, which is known to be strongly concentrated near the galactic plane [9]. Various studies conducted attempt to better understand the origins of ORCs [3]. Some considered various possibilities for their origins such as mergers of black holes [5] [14], winds from starburst galaxies and end-on lobes of radio galaxies [2] [14].

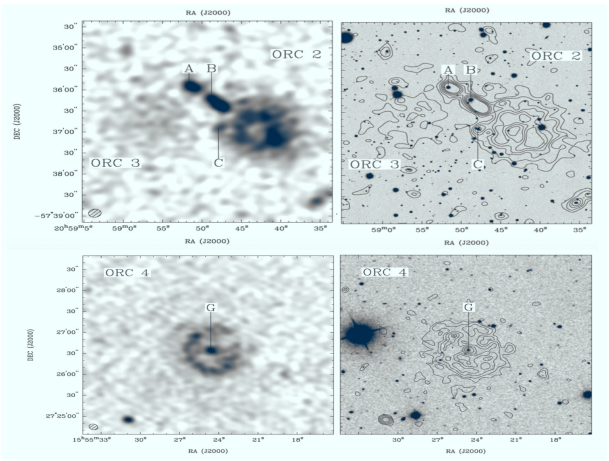

Fig. 5. Multiple ORCs and their coordinates

B. Machine Learning

The following section of the report will provide a basic understanding of machine learning and its associated models. It will discuss the inner workings of machine learning models, and the logic behind their success in multidimensional optimization problems.

1) Artificial Neural Network: With the enhancement of data and the increased complexity of problems mathematics and computer science have defined new mathematical models to circumvent such complexity. Currently, one of the most successful models are called neural networks (NN). NNs are designed to mimic the biological brain simulating neurons and their processing of input and output [16]. The simplest form of a NN is the perceptron. Figure 6 is a graphical illustration of the biological neuron and its working components. Dendrites serve as input terminals that are able to receive information from various sources. Various sources could provide different information to the dendrites. Our ears provide audio information whereas our eyes provide visual information. The cell body is where the core logic of a neuron gets processed. This is the part of the neuron that extracts the most useful information and proceeds to pass it on using an Axon.

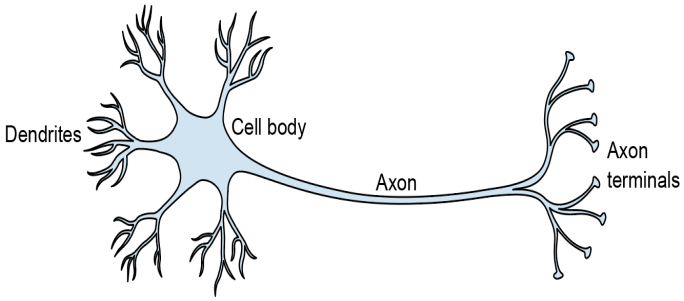

Fig. 6. The Biological Neuron

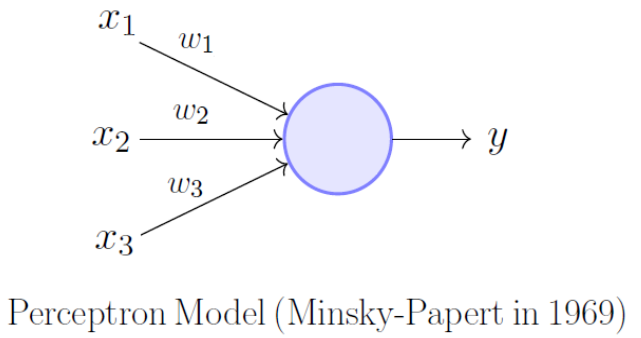

Fig. 7. Artificial Neuron: The Perceptron

Figure 7 gives us an illustration of how the artificial neuron was designed to mimic the biological neuron. A single perceptron is a function that can decide whether or not an input, represented by a vector of numbers, belongs to some specific class. From Figures 6 and 7, we can see the mapping from a biological neuron to an artificial neuron. Each X value along with the corresponding W mimics the dendrites whilst the circular shape represents the core cell body being the neuron (node). Once finished processing, the output value Y represents the axon passing information from one neuron to the next.

An MLP is a type of feed-forward NN, and is composed of multiple layers of perceptrons as shown in figure 8.

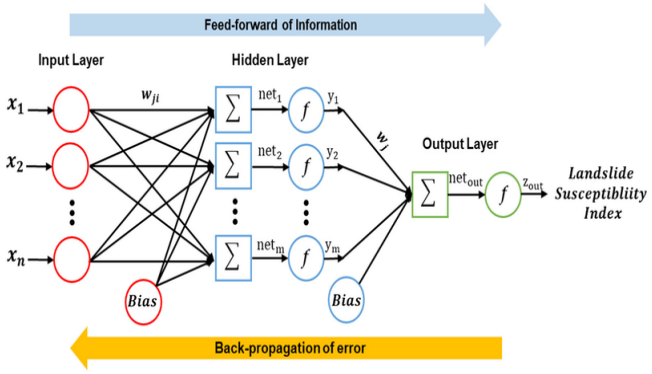

Fig. 8. Multilayer Perceptron

A feed-forward NN is a network in which nodes do not form a cycle of connections, but only propagate forward. There are at least three levels of nodes in an MLP: an input layer, a hidden layer, and an output layer [17].

2) Activation Functions: NNs are networks of neurons, where each neuron receives weighted inputs. The sum of the weighted inputs is passed through a non-linear activation function to determine whether the neuron should transmit the signal further through the network or not. This functionality is designed to emulate the functioning of a biological neuron. Common activation functions include the Rectified Linear Unit (ReLu) function. It takes the form of

$$f(x) = \max(0, x)$$

With this formula, the input is calculated and provides output for values between [0,x]. ReLu is known for aiding the efficient training of deep networks without the need for pre-training [12]. One strength of ReLu is that it will naturally promote sparsity as all negative vales are set to 0, as can be seen in Figure 9.

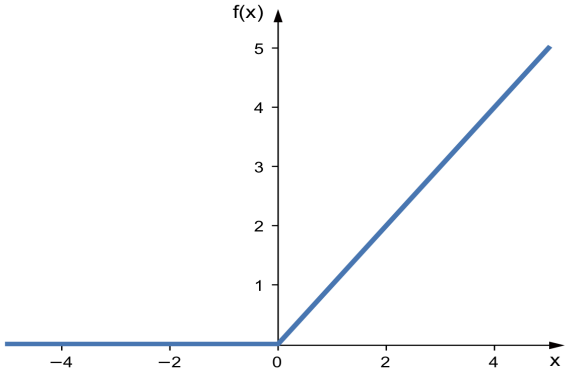

Fig. 9. ReLu Graph

3) Optimizers: NNs learn from the data via the process of training the NN weights. Weights can be adjusted iteratively by computing the error yielded by the current set of weights, and optimizing (minimizing) the error function with respect to the weights [17] [11]. The most prominent of these optimization techniques used in NNs of today include stochastic gradient descent (SGD) and the Adam optimizer. Gradient descent is a numeric optimization technique where the gradient of the error function is used to determine the direction of the search at each iterative step. Adam optimizer is a variation of gradient descent where the step along each gradient dimension is dynamically adjusted per each individual weight, which results in faster convergence [18].

4) Convolutional Neural Network: A CNN is a NN model that extends the perceptron and MLP architecture. CNNs have shown most success in computer vision applications and have shown great reconstruction performances on natural images [17] [12] [19]. This is due to the fact that the CNN model can have multiple inputs fed into the network through the input layer neurons. Basic CNN architectures make use of what is known as convolutional layers, pooling layers and fully connected layers. Convolutional layers comprise sets of feature maps, where each feature map is generated by applying a kernel to the previous layer's outputs. A kernel is defined as a square matrix designed to serve as a mathematical operator applied to an image in order to extract a specific features. A pooling layer is a layer in a CNN that performs

what is known as down sampling, reducing the spatial dimensions (being width and height) of the input data while retaining important information. Max pooling is a specific type of pooling operation where the maximum value in each local region of the input is retained and the other values are discarded. Finally, fully connected layers are the endpoint of a CNN where the output of the previous layers serves as input to each neuron of the last layer. The purpose of the fully connected layer is to perform classification or regression on the features extracted by the convolutional and pooling layers [12]. Each layer in a CNN performs feature extraction, further refining and fine tuning the internal representation of the image to maximise information retrieval [15]. Figure 10 illustrates the architecture of a CNN, its layers and their functionality [11] [20]. Various models have been created

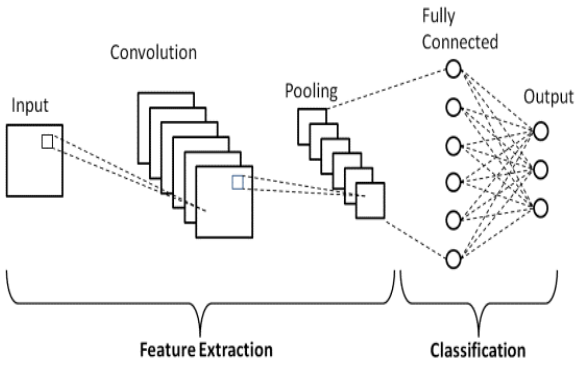

Fig. 10. Convolutional Neural Network

to perform computer vision tasks. All models starting from AlexNet, VGG, and Inception to deep Residual networks and more recently Wide Residual Networks have shown to be some of the most computationally efficient models [20].

Previous architectures that have been deployed for astronomical image classification includes residual networks (ResNets), wide residual networks (WRN), densely connected residual network (DenseNet) and dual path network (DPN) [20]. ResNet was designed to create a residual connection between convolutional layers in an attempt to combat loss of signal back-propagation [20]. The creation of WRN was due to the fact that there was a performance decline in deeper ResNets. The authors of the WRN model therefore changed the order of convolution, batch normalization, and activation, and add dropout to tune their architecture to train faster and perform competitively with deeper ResNet models.

IV. Related Work

The following section will outline all the most recently published and relevant studies and work conducted to aid in the investigation of few-shot learning methods in ORC detection. In particular Section IV-A will discuss recent publications surrounding the support of deep learning in radio astronomy. It will further expand on the role of computer vision and its related work as well as the Inception model. Section IV-B will provide more information of recently published studies conducted on radio galaxies and odd radio circles.

A. Deep Learning

The following section will provide more understanding regarding the role of deep learning. In particular section IV-A1 will expand on some recent studies conducted providing a greater support for computer vision models and their diverse abilities for applications in different domains. Sections IV-A2 and IV-A3 will provide some insights relating to the Inception and VGG models in general.

1) Computer Vision: The radio astronomy community has made use of machine learning techniques for a very long time. We have seen an increase in this overlap between the respective fields in an attempt to aid one another. In fact, using deep learning techniques for object identification and classification is not a new problem for the radio astronomy community. Some studies conducted have attempted to locate astronomical objects [21]. Various methods such as weighted linear regression, random forests and neural networks were used [21]. Although the research produced methods with accuracy scores of 93% and above, future works on unbalanced datasets still need be conducted.

Even though we know papers exist that showcase the overlap between deep learning and radio astronomy, the specific techniques in handling the data need also be explored and the variations of these techniques. Information sources like "Object Classification in Astronomical Images" provide a benchmark for processing astronomical image data [22].

Likewise, there exists new studies conducted which have made use of The Square Kilometre Array (SKA) to automate the object detection and classification problem [23]. The SKA, the largest telescope in the world, provides extreme amounts of image data that takes several hours to interpret and analyse. Semantic segmentation is another example of a deep learning methods used in radio astronomy. By making use of models like R-CNN, Mask R-CNN and U-Net, we are able to get an overview of existing techniques with regards to their performance and success relating to the specific task [23].

Whilst being able to perform object detection and classification, deep learning techniques can also be used to remove noise from image data. Convolutional denoising auto-encoders have been used to denoise synthetic images

from radio telescopes. This is done to help detect the small, faint radio sources in the universe [24]. Although the study conducted in this paper, was performed on synthetic data, further extensions to this study could be done on real telescope data [24]. Applying these techniques to different wavelengths and frequencies could provide insights in the generalizability of the models. The study could also be extended to apply these techniques on optical and infrared images. Through this, various different domains in astronomy could make more use of these models.

It must also be mentioned that even though few-shot learning can provide success in certain problem domains, we have noted that data was and still is the most important factor when dealing with image tasks. Work with ECA Relation Network, which is simply an improved version of the Relation Network, showed accuracy results of 52% and 67% when performing few-shot image classification [25]. In particular these results were reported for 5-way 1-shot and 5-way 5-shot learning methods, respectively. The limitations behind this paper however, seem to stem from the fact that the experiments conducted were performed on the mini-ImageNet dataset [25]. As such these accuracy could fall short of being considered and effective solution for real-world data and applications.

More work on CNNs applied to the radio signal domain include the work like Convolutional Radio Modulation Recognition Networks [26]. This papers combines naively learned features with expert-based methods for radio modulation classification to showcase the strength of CNNs applied to the domain. The results show the success of deep CNNs as radio modulation classifiers [26]. Although the data used in this study is different to a computer vision task, as the data used was encoded time series data, future works are suggested to apply the results obtained to different types of datasets and more modern architectures. One of these exploratory options could look at investigating the effectiveness of transfer learning could be used to explore these methods even further [26].

2) Inception: Although deep learning is a broad subfield in machine learning, it is designed for specific tasks. Some common examples of use of siamese neural networks include person identification [27]. These networks are specifically useful in various application areas but are most frequently seen in intelligent video surveillance systems. The siamese inception architecture has also before been designed to learn effective semantic features for person re-identification [27].

Facial recognition methods have seen an increase in the past 20 years. This is due to the importance of it as a biometric identification technology. This is done to ensure public and personal safety of individuals, organizations and governments. A major benefit of facial recognition technology is the insurance of public and personal safety. These methods are employed by simple business, entities, public organizations and governments. A large part of its success is due to classical facial recognition algorithms such as Principal Component Analysis [28]. This algorithm is used to reduce the dimensions of features effectively, in particular with image data. Although this research paper will not make use of the entire Inception architecture proposed by Google, studies have shown that simple inclusions of an inception model combined with the siamese network have also proven to show great results of success [28]. With this inclusion of inception modules, the network itself is able to increase its width and achieve cross-channel connection of information. Studies conducted by Xian-Feng on CASIA-webface dataset has produced accuracy scores of 99%+ with the use of such methods. Although it must be noted that this study conducted made use of a dataset of approximately 400 000 and more images and performed experiments by decreasing sample sizes [28].

3) VGG-16: The VGG-16 cnn also has made great impact in solving radio image classification tasks as outlined from the study done by [29]. This study shows methods for classifying solar radio spectrum images using the VGG-16 model with the intent to examine its effectiveness for transfer learning. The methods used in [29] shows support for the use of VGG-16 and its abilities to solve a task for classifying radio images. This comes through its approximate 12% increase in true positive rate compared to traditional deep learning models.

B. Radio Astronomy

Section IV-B1 will discuss 2 recent studies conducted using the VGG-16 neural network model. Both studies provide some of the most recent results related to radio galaxy classification using deep learning models. Both studies investigate FRI and FRII radio galaxies, with the second study adding a third FR0 radio galaxy and the YOLOv5 model.

1) Radio Galaxies: The use of the VGG-16 net model has also been used in radio galaxy classification before. Through semi-supervised learning, autencoders based on VGG-16 were designed and pre-trained to extract general features of galaxies. After fine-tuning and providing the model with labeled data, it was able to successfully classify different morphologies [30]. This study also makes use of FRI, FRII and bent tailed radio galaxies. In this study, the semi-supervised model is compared to transfer learning. It is able to perform and analysis on the limitations of transfer learning due to the lack of training with astronomical image data [30].

Even though a part of this paper focuses on few-shot learning in classifying FRI and FRII radio galaxies, other studies have made use of different models in an attempt to perform the same task. Studies conducted have made use of the YOLOv5 object detection model for detecting and classifying FR0, FRI and FRII radio galaxies [31]. With precision scores of 89.4% on generalization set, we can see the effectiveness of deep learning models in classifying radio galaxies. Not only has this paper shown success in the before mentioned 3 radio galaxies, but can also be applied to detection of other galaxy types. The strength in this research, lies in the combination of different data types. The combination of radio and optical image data, allowed the researchers a more comprehensive analysis on the classification task [31]. The variety of data also allows for more data to be fed into the models used.

V. Methodology

This research report takes on an experimental approach whereby it runs multiple experiments with various datasets in order to determine the potential for few-shot learning methods, in particular the use of SNNs, in classification tasks with respect to radio-astronomy images. The dataset to be used in this study contains a subset of radio galaxy images used from the study conducted by Aniyan and Thorat [16]. The study conducted by Aniyan and Thorat [16] performed classification of 3 types of radio galaxies, namely Bent tail, FRI and FRII radio galaxies. For our study, we shall only make use of Bent tail and FRII galaxies.

A. Image conversion

- Sigma Clipping: First, the sigma-clipped statistics of each image are estimated in order to calculate the background noise and flux levels. With sigma-clipped statistics, pixels above a certain sigma level from the median are discarded or nulled. In this study, all values below a 3σ level of the background were cut - off by suppressing those values to zero so as to highlight the contribution of the source and remove any unwanted back- ground noise [16]. For all images we selected a sigma clipped value of 3σ
- FITS to JPG: After performing sigma-clipping, it then performs normalization to normalize the data to the range [0,255]. After normalization, we convert each image to an unsigned 8-bit integer. This is a common data type for representing pixel values between the range [0,255]. The final step in converting the FITS image to JPG, is the expansion of the last dimension. Thereafter, the 2D array returned by the sigma clipped FITS image adds a final third dimension to it. This is done to ensure our images are converted to colour RGB images to be fed into our SNN.

The report conducts 4 experiments. Each experiment will look to test the limits of SNNs based on specific scenarios. These experiments will respectively be labelled, **RG_NA** (radio galaxy non-augmented), **RG_AG** (radio galaxy augmented), **ORC_NA** and (ORC non-augmented) **ORC_AG** (ORC augmented). The experiments were labelled in such a way that the description preceding the underscore, describes the specific radio source to be investigated, whereas the description following the underscore describes the dataset to be used. Experiments RG_NA and RG_AG will in particular investigate the abilities of SNNs to classify different radio galaxies from one another. As such, experiments RG_NA and RG_AG will measure the difference in classifying Bent tail vs FRII radio galaxies. Experiments ORC_NA and ORC_AG will look at an SNNs ability to identify and distinguish ORCs from radio galaxies.

Experiment RG_NA will attempt to perform a similar study to Aniyan and Thorat [16]. Specifically, the report shall look at classifying radio galaxies. Instead of classifying each different variation and type of radio galaxy, we shall only look at differentiating bent tail galaxies from FRII galaxies. Experiment RG_AG will share the same study, but will include data augmentation techniques. The core investigation of this research paper is to determine the strengths and limits of few-shot learning methods. As such data augmentation will allow us to investigate those limits.

B. Data Augmentation

As machine learning models require data to be able to learn and perform certain tasks, various different data formats can be considered. The model generally consumes hundreds of thousands of instances of images to analyze and learn from. Various data augmentation techniques exist in order to manipulate the input data in such a way that it generates a new instance to be used in the data set. The following data augmentation techniques will be used in the experiments:

- Flipping: Flipping is a technique which provides a reflected mirror version of the image.
- Rotating: Rotating is a simple technique which involves rotating an image a certain amount of degrees [12]. This "newly" generated image will thus serve as an added instance in our data set.
- Cropping: Cropping, is the data augmentation technique that allows us to redefine the dimensions of an image to that of a smaller image. This is done to gain a better view and insight of the center of the image.

Each of these techniques generates a new image to be added to the data set. Although each image might not differ that much from the original image(s), they still serve a valuable purpose in the training of the machine

learning model in the sense that the data set has increased and the NN has more data to learn. These techniques are employed to avoid overfitting. Overfitting occurs in ML when models are able to accurately predict the output on training data but not on new data. This can easily occur with a small dataset. ORC images are very few, hence why the fore mentioned techniques are desirable to be used. To increase the dataset as well as to avoid overfitting.

Thresholding is a technique that is frequently used to extract important pixels from an image by removing noise and background pixels [15]. In previous studies conducted, rotational data augmentation was used to evaluate the amount of time that it takes to train a CNN whilst using this approach. It was determined that with small rotational intervals, data augmentation leads to near-perfect accuracy on the samples in the testing data, which means a large portion of the training time might have been wasted [15].

With experiments RG_NA and RG_AG investigating the abilities of SNNs to classify the two classes of radio galaxies, data augmentation will play a crucial role in this study. The initial dataset size for experiments RG_NA and RG_AG will be the same. However, after performing the before-mentioned data augmentation techniques, we hope that the increase in size of dataset for experiment RG_AG, will provide more insight into a SNNs abilities to perform the classification task based on the amount of samples it receives. Experiments ORC_NA and ORC_AG will look to perform the same tasks, but with ORC data. That is, the aim is to distinguish ORCs from radio galaxies. In particular, we will make use of FRII radio galaxies, as they could serve as a good base for training a NN to identify radio galaxy type features to that compared of an ORC. Each experiment will follow a 80/20 rule whereby 80% of the dataset will be used for the neural network training and validation sets, and the remaining 20% will be used for the generalisation set to measure the true success of the proposed SNN. Furthermore, when the SNN is trained, the training and validation datasets are split using the 70/30 rule, whereby 70% of the data is used for training the neural network and 30% is used for the validation set. As an example, experiment RG_NA will initially contain 150 images. Of these 150 images, 120 will be used for training and validation and 30 will be used for generalisation, when using the 80/20 rule. As SNNs work with different classes when dealing with images, we would also need to split these classes equally. This comes out as 120 split between our 3 classes comes out to produce 40 images per class. That is 40 images for the anchor, positive and negative class. As for the generalisation set, we would have the remaining 30 images also even split across the 3 classes. We then have 10 images per class, anchor, positive and negative. This logic will also be applied to every other experiment including the experiments investigating the classification with ORCs. The remaining experiments, along with their dataset sizes, are outlined in Table I. The above mentioned logic was replicated on each dataset for each experiment. The initial ORC datasets will be much smaller due to the limited amount of image data surrounding ORCs. Both sets of experiments, experiments designed for radio galaxy classification and experiments designed for ORC classification, will initially share the same size dataset as its counterpart experiment. This is done to ensure consistency and to avoid any bias from the neural network. What this means is that the dataset used in experiment RG_NA will be used in RG_AG, with the difference being that augmentation was applied to the images in RG_AG.

Experiment	*Init size*	*Train size*	*Val size*	*# per class*
RG_NA	150	120	30	(40, 10)
RG_AG	3150	2520	630	(840, 210)
ORC_NA	12	9	3	(3, 1)
ORC_AG	252	189	63	(63, 21)

TABLE I
SAMPLE SIZES OF EACH EXPERIMENT.

In order to gain more insight into our image data, we made use of principal component analysis (PCA). PCA is an algorithmic technique used to which specializes in dimensional reduction. This allows us to map our image data to a 2-dimensional plane and gain insights based on their features. By applying PCA to experiments RG_NA through to ORC_NA, we were able to plot the anchor, negative and positive datasets against each other on a scatter plot. From this scatter plot, and through PCA we were able to visualize how the datasets correlate to one another. PCA offers a linear projection where the individual components (=axis) are ranked in terms of variance explained, and can help us identify the relation between datasets. That is, should data points be very tightly and closely plotted, they share very similar features, indicating that the images are fairly similar in nature. Should the points be more sparsely scattered, this would indicate that the images are distinctly different from each other.

As can be seen from Figure 11, the information obtained by performing PCA on experiment RG_AG's training dataset, shows very tightly plotted data points. The anchor and positive classes overlap a lot as this is expected. The both share the same class of radio galaxy being Bent tail radio galaxies. From the chart of the negative data points, we can see that data points are more sparse. This is good as it indicates that the algorithm does show that the dataset shows characteristics of differences. This still presents a problem though as many of the data points overlap. We have chosen to keep this dataset

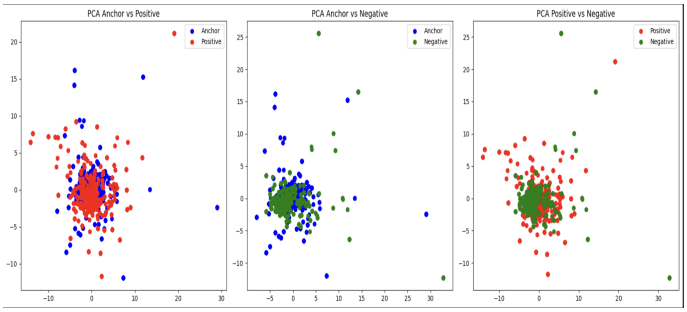

Fig. 11. PCA applied to experiment RG_AG dataset

as it can give insight into a SNN's ability to classify radio galaxies from one another, even with datasets with samples that are very similar to each other.

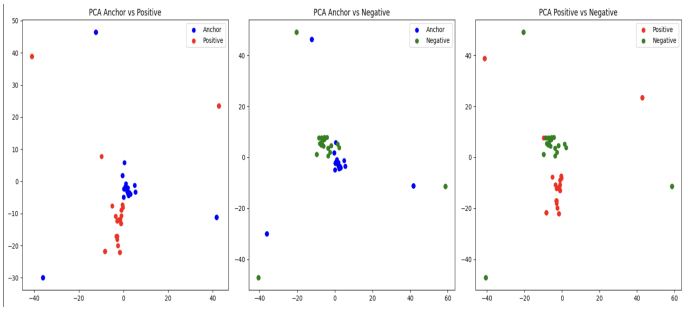

Fig. 12. PCA applied to experiment ORC_AG dataset

From Figure 12, we can see a more sparse separated scatter plot showcasing a differentiated dataset, for the PCA applied to experiment ORC_AG's training dataset. This is good as we have a healthy dataset which displays a distinction from ORCs and FRII radio galaxies. Although many data points can still be seen to overlap, we are aware of this as the classification task of distinguishing ORCs from radio galaxies can still be challenging. The interesting thing to note is that the anchor and positive datasets also show some diversity after applying PCA. This is due to the specific images selected. The study inserted differentiating FRII images in order to allow the NN to learn general features of FRII radio galaxies compared to ORCs.

C. Siamese Neural Network Architecture

The architecture of a siamese neural network was designed to include a twin-pair convolutional neural network in the underlying architecture. This is done as each CNN will learn the specific features of one class of image and thereafter produce the output to the resulting distance layer. The underlying CNN architecture in a SNN is the exact same. The weights between both CNNs are shared in order to ensure that the overall neural network performs the same task based on the same learned features. The output produced by each CNN is consumed by the final distance layer in the SNN. This distance layer then computes the Euclidean distance between the features learned by each CNN. The smaller the distance, the more similar the images are to each other. The greater the output produced by the distance layer, the more different each image is from one another. As such we hope to construct a SNN that produced maximal distances between Bent Tail and FRII radio galaxies, as well as maximal distances between any radio galaxy and our target radio source, ORCs.

Fig. 13. SNN Architecture Flow

Figure 13 shows the SNN architecture. The inner blocks portray the architectural design of the internal CNN VGG-16. The outer layers are the added fully connected layer which flows towards the single output given by the SNN.

Table II is a technical outline of the SNN architecture and the features it learns. It shows the output of each layer and the number of parameters each layer learns.

D. Metrics

With the study of machine learning, come various mathematical performance metrics. These metrics are designed to measure the success of the models. The Receiver Operating Characteristic (ROC) curve is a plot that is commonly used to evaluate the performance of binary classification models in machine learning [13]. The Area Under Curve (AUC) measures the overall performance of a model across all possible classification thresholds, by calculating the area under the ROC curve. Other performance metrics that are often considered are precision, recall and the F1-score [15] [20].

- Precision

  Precision is a metric used to measure the performance of classifiers. It is defined as the percentage of number of instances correctly classified in relation to the total number of correctly classified instances. The TP variable represents the number of instances the model was able to predict the correct class in a positive manner. FP is the number of instances the model incorrectly predict as the correct class. It is given by the following formula:

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

Layer	*Output Shape*	*# Params*
InputLayer	[(None, None, None, 3)]	0
	Block 1	
Conv2D	(None, None, None, 64)	1792
Conv2D	(None, None, None, 64)	36928
MaxPooling2D	(None, None, None, 64)	0
	Block 2	
Conv2D	(None, None, None, 128)	73856
Conv2D	(None, None, None, 128)	147584
MaxPooling2D	(None, None, None, 128)	0
	Block 3	
Conv2D	(None, None, None, 256)	295168
Conv2D	(None, None, None, 256)	590080
Conv2D	(None, None, None, 256)	590080
MaxPooling2D	(None, None, None, 256)	0
	Block 4	
Conv2D	(None, None, None, 512)	1180160
Conv2D	(None, None, None, 512)	2359808
Conv2D	(None, None, None, 512)	2359808
MaxPooling2D	(None, None, None, 512)	0
	Block 5	
Conv2D	(None, None, None, 512)	2359808
Conv2D	(None, None, None, 512)	2359808
Conv2D	(None, None, None, 512)	2359808
MaxPooling2D	(None, None, None, 512)	0
	Output to L1 Distance Layer	
Dense	(None, None, None, 4096)	0
	L1 Distance Layer	
Dense	(None, None, None, 1)	0

TABLE II
SNN Architecture

- Recall

  Recall is defined as the ratio of positive instances classified correctly in comparison to the total number of positive instances. This measurement gives insight into our models' ability to detect positive instances. A higher value is more desirable as it means our model finds more positive samples. FN is a variable where the model incorrectly predicts the incorrect class. It is given by the following formula:

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

- F1-Score

  F1-score is the combination of precision and recall. We attempt to compute the harmonic mean. This in turn gives a very broad overview of the models performance. Both recall and precision measure the models' ability to identify positivity. This means that a higher recall and precision lead to a higher F1-score, whereas a lower recall does the opposite. The F1-score is given by the following formula:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

- Accuracy:

  Accuracy is defined as the percentage of correctly classified instances to the entire set of classified instances. It is given by the following formula:

$$Accuracy = \frac{TP + FP}{TP + TN + FP + FN}$$

- Confusion Matrix

  A confusion matrix is a 2x2 grid that outputs the performance of a mathematical algorithm in its ability to correct classify true positives, true negatives, false positives and false negatives. This is included in the analysis of our SNN specifically to measure its performance on the generalization set.

- ROC-AUC

  The ROC-AUC is a graph showing the performance of a classification model at all classification thresholds. It can simply be understood as the visual representation of the confusion matrix. It produces an output score in the form of a percentage, indicating how well a model has performed. With the objective to get an ROC-AUC curve score closest to 100%, it combines the true positive rate with the false positive rate to provide insight. The true positive rate is a synonym for recall, where as the false positive rate can be defined as the following:

$$FPR = \frac{FP}{FP + TN}$$

VI. Experimentation

The final section of this report outlines each of the experiments and their individual results and evaluation of their performance compared to their ability to perform the core task, i.e. classifying ORCs and FRII radio galaxies. Section VI-A reports on the results obtained for experiments RG_NA and RG_AG as outlined by the Section V. Specifically the success of SNNs to classify FRII and bent tail radio galaxies from each other is investigated. Section VI-B expands on this investigation by conducting the classification of ORCs from FRII radio galaxies and performing experiments ORC_NA and ORC_AG. To provide an overview analysis of the internal architecture for our model, we will produce the results for 3 different internal models in the SNN in section VI-C. This will show the journey towards the VGG-16 model. From starting out with old reliable models such as AlexNet, with shifts to modern architectures like VGG-16 and Inception, it will

show the average performances of each model based on specific hyperparameters.

A. Few-Shot Learning for Radio Galaxy Classification

Experiment RG_NA was designed to further add upon previous work done with radio galaxy classification such as that from [16]. In particular we wanted to measure the strength and performance of SNNs in their few-shot learning abilities to a dataset containing very similarly featured data. On average we tested a set of hyperparameters for 5 runs in order to gain insight into the models performance. Tables III to V, show specific metrics that is taken into account as defined by the methodology in Section V. The internal CNNs that was used, in general tend to perform well on both training data and the validation sets. But in order to find an optimal model, we analyzed its performance on a generalization set. That being, a dataset which it has never seen before. As the classification task for experiment RG_NA is one to try and differentiate similar celestial objects we need to ensure that the model correctly identifies each instance of a FRII and bent tail radio galaxy. For this we mainly focused on precision as our core metric.

Experiment RG_AG, is the extension of RG_NA. RG_NA makes use of a raw datasets of 150 images. We know the CNNs in general are data driven models that require large samples in order to achieve successes. As such experiments RG_NA and RG_AG measure the SNN's extent to perform few shot learning. With the addition of data augmentation to experiments RG_AG, we increased the dataset to sample sizes of approximately 3000. Although, we could have added more data augmentation techniques and applied these techniques to the dataset for more rounds, which would allow sample sizes of tens of thousands of images, we kept the number of augmentation rounds low as the core study involves an investigation regarding SNNs with few-shot learning methods. Image 14 is a radio image to show one of the FRII radio galaxies that will be used in the dataset for experiments RG_NA and RG_AG, as well as ORC_NA and ORC_AG.

B. Few-Shot Learning ORC Classification

Experiment ORC_NA displays the first set of results with regards to ORC classification using few shot learning methods. The nature of these experiments are the same as that of RG_NA. With the core idea being, to investigate the SNN model on very few images. Through the PCA analysis conducted in section V-B we notice that the dataset displays more diverse properties compared to the investigation done by experiments RG_NA and RG_AG.

Experiment ORC_AG, serves as the final experiments. With the addition of data augmentation we are able to increase our sample size. It must be noted, that even with the use of data augmentation, experiments ORC_NA and ORC_AG serve as our smallest datasets. With sample sizes containing images as little as a single image in a dataset, we can truly test the extent to which a SNN can perform classification using few-shot learning. Image 15 is a radio image to show one of the ORCs that will be used in the dataset for experiments ORC_NA and ORC_AG.

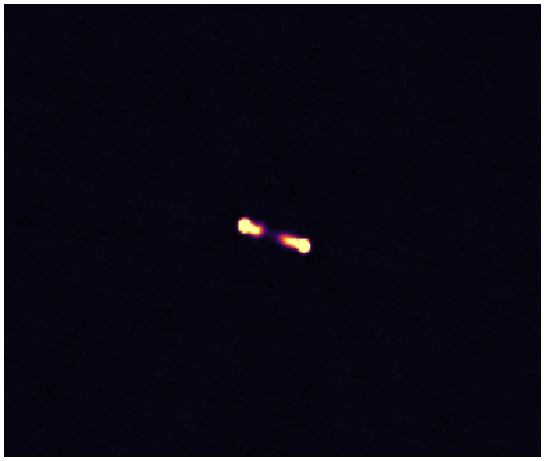

Fig. 14. Radio image of FRII radio galaxy

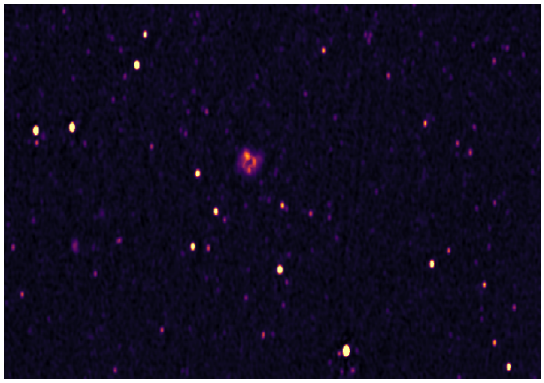

Fig. 15. Radio image of ORC

C. Model Performances

The following section will tabulate the performance of each investigation showcasing the average results obtained by each internal CNN, based on the specific experiments. The hyperparameters were always shared with slight deviations. All experiments were run with both the Adam and RMSProp optimizer. The report also made use of a batch size of 32, since the amount of images we had for training was 120. It kept oscillating between a 1e-4 and 1e-5 learning rate, but found the 1e-5 to be the best as the 1e-4 converged too quickly. The following was observed when running experiments RG_NA through to ORC_AG.

All internal CNNs used the same set of hyperparameters to run each experiment. Experiment RG_NA, used a batch size of 8 for training data, with a batch size of 4 for validation data. Experiment RG_AG had a batch size of 64 and 4 respectively for training and validation. For experiment ORC_NA the report used the a batch size of 1 for both training and valdiation due to the small nature of both datasets. And finally, for experiment ORC_AG, the batch sizes to be used was 8 and 4 for training and validation.

Exp	***Gen Acc***	***Gen Loss***	***Prec***	***Recall***	**ROC-AUC**
RG_NA	51.25%	0,69	47.66%	0.74	51.2%
RG_AG	59.42%	1.12	62.16%	0.61	66.52%
ORC_NA	50%	0.69	50%	1	100%
ORC_AG	43.75%	0.69	45.67%	0.73	27.55%

TABLE III
METRICS FOR MODEL: ALEXNET

In general, the AlexNet was ran for a large amount of epochs. Experiment RG_NA ran for 50 epochs, experiment RG_AG for 30 and experiments ORC_NA and ORC_AG used 15 epochs. Each experiment was run for an average of 5 rounds whilst testing the Adam and RMSProp optimizer with learning rates of 1e-4 and 1e-5. This means that an average of 20 runs were observed for the AlexNet model based on the before mentioned hyperparameters. What was noted is that for experiment RG_NA the Adam optimizer learned faster but training results tended to oscillate. The RMSProp optimizer performs similarly yet with a more steady learning curve. With the 1e-4 learning rate, the network seemed to over fit, with the validation loss tending to increase and its accuracy decreasing.

Experiment RG_AG shows similar trends. Initially, the validation loss decreases, but the longer the model ran, we saw the loss increase, with the accuracies decreasing. Most runs saw an increase in accuracy initially, followed by a decrease. This could show that the model overlearns features and seems to stray away from the solution. A potential way to overcome this could be to run the model for less epochs, but that could lead it to high loss metrics with less features learned.

Experiments ORC_NA and ORC_AG overall seemed to produce weak results as the limited data seemed too hard of a task for the AlexNet model. The model trained very slowly, with some cases showing barely any improvement. The validation, training and generalization results produced similar metric scores.

The VGG-16 model was ran for less epochs. Experiment RG_NA ran for 30 epochs, experiment RG_AG for 10 and experiments ORC_NA and ORC_AG used 15 epochs. This model was evaluated on the same bases as the AlexNet. This means that an average of 20 runs were observed for the VGG-16 model based on the before mentioned hyperparameters. For the VGG-16 model it was noted that for experiment RG_NA the loss showed the same trends as that of AlexNet. The loss tends to decrease yet around epoch 15, the curve moves in the opposite direction and increases. In general the loss tended to be very high and kept increasing.

Exp	***Gen Acc***	***Gen Loss***	***Prec***	***Recall***	**ROC-AUC**
RG_NA	59.50%	1.18	54.20%	0.89	63.75%
RG_AG	56.53%	1.85	53.98%	0.98	72.35%
ORC_NA	100%	0.57	100%	1	100%
ORC_AG	91.63%	0.19	90.36%	1	97.55%

TABLE IV
METRICS FOR MODEL: VGG-16

Experiment RG_AG produced worse results surprisingly considering the datasets were much bigger. It could be noted that the the VGG-16 model simply could not distinguish much of a difference in the features of the two types of galaxies, as loss metric only increased, even with a much larger dataset. The most optimal performing set of hyperparameters were the RMSProp optimizer with a learning rate of 1e-4.

The most success of the report, seems to come from Experiments ORC_NA and ORC_AG which produce the strongest results. Although experiment ORC_NA contains the fewest samples, the NN VGG-16 seems to still learn the features of radio galaxies compared to ORCs quite well. Table IV produces the averages of the 20 runs for each experiment. As can be seen Experiment ORC_NA produces near perfect results, although there is still some caution as the loss is still somewhat high. Due to this, experiment ORC_AG produces the most desired model. With a minimal loss and good generalization metrics it shows that the VGG-16 model contends to be the best candidate for the ORC classification task amongst radio galaxies. The supporting evidence comes from experiment ORC_NA showing good results. Both these two experiments give evidence that a decently sized dataset could be used for the classification task.

Finally, the InceptionV3 model was used. Experiment RG_NA ran for 30 epochs, experiment RG_AG for 10 and experiments ORC_NA and ORC_AG used 15 epochs. Again, this model was evaluated on the same bases as the AlexNet and VGG-16. The InceptionV3 model it was noted that for experiment RG_NA the loss showed the same trends as that of AlexNet. For experiment RG_NA, the model InceptionV3 seems to

Exp	*Gen Acc*	*Gen Loss*	*Prec*	*Recall*	**ROC-AUC**
RG_NA	50.25%	0.70	49.82%	0.58	50.55%
RG_AG	60.19%	1.31	56.4%	0.94	69.58%
ORC_NA	50%	0.73	50%	1	0%
ORC_AG	54.13%	0.69	42.95%	0.42	54.4%

TABLE V
METRICS FOR MODEL: INCEPTIONV3

show the poorest results. During the runs the metrics tend to fluctuate the most. With training, we see results spiking and dipping drastically. There is no steady trend of learning.

Experiment RG_AG produced the only stable training results for the InceptionV3 model. The combination of learning rate 1e-5 and Adam gives the model its most steady training results. Unfortunately, the model tends to overfit by learning well on training data, but not being able to perform the same task on the generalization set.

In general the Adam optimizer seems to be the best performer with the InceptionV3 model. A steep learning rate such as that of 1e-5 works best for experiments ORC_NA and ORC_AG. Unfortunately, the model still seems to struggle in performing the classification of ORCs against radio galaxies. In most cases, the report shows that the AlexNet model tends to perform both the radio galaxy and ORC classification tasks better than the InceptionV3 model.

Figure 16 shows the learning curve for the VGG-16 model. As can be seen the VGG-16 model finds the solution early with steady decrease in loss. The RMSProp showed the best results leading into a steady learning with a smooth improvement in accuracy and a steady decrease in loss.

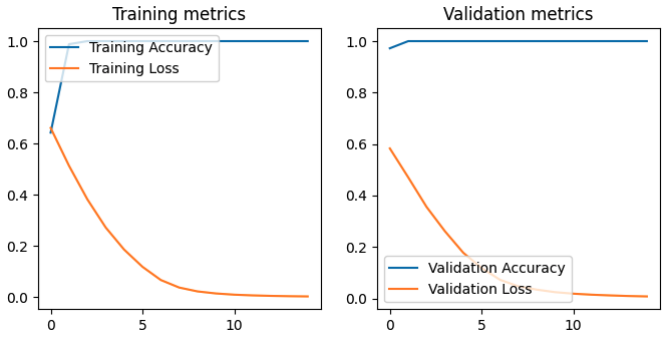

Fig. 16. VGG-16 metrics for experiment ORC_AG

VII. CONCLUSION

At the end of the conduction of experiments RG_NA through to ORC_AG, the results show that experiments ORC_AG produced promising results for ORC classification using few-shot learning methods. With training and validation accuracies of both 100% and minimal loss scores, the classification task seems to be a success when finding accuracy scores of around 95% and validation loss values of approximately the 0.1-0.15 range on the generalization set. In order to support the models success, the recall metric served as the core metric in experiments ORC_NA and ORC_AG. Not only were we able to achieve recall scores of 100, but also precision values of 90% and ROC-AUC scores of 100.

In conclusion, this report out to discover the abilities of a specific type of CNN, being the SNN and its strength in performing few-shot learning methods with regards to classification tasks. The main investigation involved finding an internal CNN to be used by our SNN for the classification. ORCs are and remain very hard to identify in the universe due to their extreme rarities and properties, of which we are still learning. However, recent studies published such as the findings presented in [6], help understand the nature of ORCs and their origins through various simulations.

With the addition of deep learning models, we can see how machine learning has and will continue to aid the astronomy community and their respective spheres. Various models can be designed for various different tasks, but with this study conducted, we hope that it can aid researchers in the future when trying to answer any questions regarding ORCs.

As can be seen from the results obtained, SNNs show potential to perform classification with few-shot learning. The VGG-16 model showed that it was the best candidate for our internal CNN. In addition to the model, the RMSProp optimizer along with a steep learning rate of 1e-5 lead to the best performing model. With sample sizes of hundreds and thousands of images, to sample sizes containing a single image, future work could even look at extending this investigation to a one-shot learning methods, or alternatively even the use of transfer learning to other celestial objects.

REFERENCES

[1] N. Gupta, M. Huynh, R. P. Norris, X. R. Wang, A. M. Hopkins, H. Andernach, B. S. Koribalski, and T. J. Galvin, "Discovery of peculiar radio morphologies with askap using unsupervised machine learning," *Publications of the Astronomical Society of Australia*, vol. 39, p. e051, 2022.

[2] A. Omar, "Yet another odd radio circle?," *Research Notes of the AAS*, vol. 6, no. 5, p. 100, 2022.

[3] R. P. Norris, E. Crawford, and P. Macgregor, "Odd radio circles and their environment," *Galaxies*, vol. 9, no. 4, p. 83, 2021.

[4] M. Young, "New radio data reveal possible origins of odd radio circles," *Sky and Telescope*, vol. 144, no. 7, p. 8, 2022.

[5] R. P. Norris, J. D. Collier, R. M. Crocker, I. Heywood, P. Macgregor, L. Rudnick, S. Shabala, H. Andernach, E. Da Cunha, J. English, *et al.*, "Meerkat uncovers the physics of an odd radio circle," *Monthly Notices of the Royal Astronomical Society*, vol. 513, no. 1, pp. 1300–1316, 2022.

[6] K. Dolag, L. M. Böss, B. S. Koribalski, U. P. Steinwandel, and M. Valentini, "Insights on the origin of odd radio circles from cosmological simulations," *The Astrophysical Journal*, vol. 945, no. 1, p. 74, 2023.

[7] M. D. Filipović, J. L. Payne, R. Alsaberi, R. P. Norris, P. J. Macgregor, L. Rudnick, B. S. Koribalski, D. Leahy, L. Ducci, R. Kothes, *et al.*, "Mysterious odd radio circle near the large magellanic cloud–an intergalactic supernova remnant?," *Monthly Notices of the Royal Astronomical Society*, vol. 512, no. 1, pp. 265–284, 2022.

[8] K. Dolag, L. M. Böss, B. S. Koribalski, U. P. Steinwandel, and M. Valentini, "Insights on the origin of orcs from cosmological simulations," *arXiv preprint arXiv:2208.15003*, 2022.

[9] A. Omar, "Odd radio circles as supernovae remnants in the intragroup medium," *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 513, no. 1, pp. 101–105, 2022.

[10] S. K. Sarbadhicary, T. A. Thompson, L. A. Lopez, and S. Mathur, "On odd radio circles as supernova remnants," *arXiv preprint arXiv:2209.10554*, 2022.

[11] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.

[12] R. Flamary, "Astronomical image reconstruction with convolutional neural networks," in *2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 2468–2472, IEEE, 2017.

[13] A. Kimura, I. Takahashi, M. Tanaka, N. Yasuda, N. Ueda, and N. Yoshida, "Single-epoch supernova classification with deep convolutional neural networks," in *2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pp. 354–359, IEEE, 2017.

[14] M. Lochner, L. Rudnick, I. Heywood, K. Knowles, and S. S. Shabala, "A unique, ring-like radio source with quadrilateral structure detected with machine learning," *Monthly Notices of the Royal Astronomical Society*, vol. 520, no. 1, pp. 1439–1446, 2023.

[15] K. Brand, T. L. Grobler, W. Kleynhans, M. Vaccari, M. Prescott, and B. Becker, "Feature guided training and rotational standardization for the morphological classification of radio galaxies," *Monthly Notices of the Royal Astronomical Society*, vol. 522, no. 1, pp. 292–311, 2023.

[16] A. Aniyan and K. Thorat, "Classifying radio galaxies with the convolutional neural network," *The Astrophysical Journal Supplement Series*, vol. 230, no. 2, p. 20, 2017.

[17] V. Carruba, S. Aljbaae, R. Domingos, M. Huaman, and W. Barletta, "Machine learning applied to asteroid dynamics," *Celestial Mechanics and Dynamical Astronomy*, vol. 134, no. 4, p. 36, 2022.

[18] K. Schmidt, F. Geyer, S. Fröse, P.-S. Blomenkamp, M. Brüggen, F. de Gasperin, D. Elsässer, and W. Rhode, "Deep learning-based imaging in radio interferometry," *Astronomy & Astrophysics*, vol. 664, p. A134, 2022.

[19] O. Flasseur, T. Bodrito, J. Mairal, J. Ponce, M. Langlois, and A.-M. Lagrange, "deep paco: Combining statistical models with deep learning for exoplanet detection and characterization in direct imaging at high contrast," *arXiv preprint arXiv:2303.02461*, 2023.

[20] G. Harp *et al.*, "Machine vision and deep learning for classification of radio seti signals. arxiv e-prints," *arXiv preprint arXiv:1902.02426*, 1902.

[21] J. de la Calleja and O. Fuentes, "Automated classification of astronomical objects in multispectral wide-field images.," in *FLAIRS Conference*, pp. 610–612, Citeseer, 2006.

[22] R. L. White, "Object classification in astronomical images," in *Statistical Challenges in Modern Astronomy II*, pp. 135–151, Springer, 1997.

[23] R. Sortino, D. Magro, G. Fiameni, E. Sciacca, S. Riggi, A. DeMarco, C. Spampinato, A. M. Hopkins, F. Bufano, F. Schillirò, *et al.*, "Radio astronomical images object detection and segmentation: a benchmark on deep learning methods," *Experimental Astronomy*, pp. 1–39, 2023.

[24] C. Gheller and F. Vazza, "Convolutional deep denoising autoencoders for radio astronomical images," *Monthly Notices of the Royal Astronomical Society*, vol. 509, no. 1, pp. 990–1009, 2022.

[25] J. Yu, J. Liang, H. Mei, J. Fan, and S. Yu, "Improved few-shot learning for images classification," in *2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS)*, pp. 137–140, IEEE, 2022.

[26] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Engineering Applications of Neural Networks: 17th International Conference, EANN 2016, Aberdeen, UK, September 2-5, 2016, Proceedings 17*, pp. 213–226, Springer, 2016.

[27] S. Li and H. Ma, "A siamese inception architecture network for person re-identification," *Machine Vision and Applications*, vol. 28, no. 7, pp. 725–736, 2017.

[28] X.-F. Xu, L. Zhang, C.-D. Duan, and Y. Lu, "Research on inception module incorporated siamese convolutional neural networks to realize face recognition," *IEEE Access*, vol. 8, pp. 12168–12178, 2019.

[29] M. Chen, G. Yuan, H. Zhou, R. Cheng, L. Xu, and C. Tan, "Classification of solar radio spectrum based on vgg16 transfer learning," in *Image and Graphics Technologies and Applications: 16th Chinese Conference on Image and Graphics Technologies, IGTA 2021, Beijing, China, June 6–7, 2021, Revised Selected Papers 16*, pp. 35–48, Springer, 2021.

[30] Z. Ma, J. Zhu, Y. Zhu, and H. Xu, "Classification of radio galaxy images with semi-supervised learning," in *Data Mining and Big Data: 4th International Conference, DMBD 2019, Chiang Mai, Thailand, July 26–30, 2019, Proceedings 4*, pp. 191–200, Springer, 2019.

[31] Z. Zhang, B. Jiang, and Y. Zhang, "Automatic detection and classification of radio galaxy images by deep learning," *Publications of the Astronomical Society of the Pacific*, vol. 134, no. 1036, p. 064503, 2022.