International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842
Issue 08, Volume 09 (August 2022) https://www.irjcs.com/archives

MEGACOSM – DETECTION AND CLASSIFICATION OF ASTRONOMICAL OBJECTS

B V Vinay
Assistant Professor, Dept. of ISE
Vemana Institute of Technology, Bangalore
Poonam Bhaskar
Dept. of ISE
Vemana Institute of Technology, Bangalore
Rohan Keshav H
Dept. of ISE
Vemana Institute of Technology, Bangalore
Sirvi Priyanka Hiralal
Dept. of ISE
Vemana Institute of Technology, Bangalore
Vidya P
Dept. of ISE
Vemana Institute of Technology, Bangalore

Publication History
Manuscript Reference No: IRJCS/RS/Vol.09/Issue08/SPAUCS10093
Research Article | Open Access
Peer-review: Double-blind Peer-reviewed
Article ID: IRJCS/RS/Vol.09/Issue08/SPAUCS10093
Volume 2022 | Article ID SPAUCS10093 http://www.irjcs.com/volumes/Vol09/iss-08/13.SPAUCS10093.pdf
Received: 25, July 2022 | Accepted: 04, August 2022 | Published: 12, August 2022
doi: https://doi.org/10.26562/irjcs.2022.v0908.13
Citation: Vinay,Poonam,Rohan,Sirvi,Vidya (2022).MEGACOSM – Detection and Classification of Astronomical objects. International Research Journal of Computer Science (IRJCS), Volume IX, Issue VIII, 2022, Pgs217-223
BibTeX key:
Academic Editor-Chief: Dr.A.Arul Lawrence Selvakumar

Abstract: The concept of existence started with the bigbang theory, which was a phenomenon where multiple objects collided to create multiple other things. This creation is said to be increasing at the rate of the cosmic acceleration due to which multiple new objects came into existence called as the astronomical objects. Space object (SO) detection, classification, and characterization are significant challenges in many research fields. In recent years, deep learning and other forms of artificial intelligence (AI) have drawn the attention of many astronomers and academics. Megacosm is project that is used for the classification and the identification of those newly created celestial objects. It works on the process of manually training the model with the data annotation techniques and the dataset is enhanced using the data augmentation technique. It uses YOLO as the core algorithm and also deep learning concepts like CNN (Convolution Neural Network) to predict results. It gives the output as a bounding box around the detected object along with the accuracy of that prediction.
Keywords: Astronomical Objects, Megacosm, SOs, AI, YOLO, Confidence value

I. INTRODUCTION

A naturally available solid entity, combination, or construction that exists in the known universe is considered to as an extra solar planet or celestial object. The galactic is the basic building block of synthesis at the ultimate sizes. A web that spans the observable universe is created by the organisation of galaxies into groups and clusters, frequently with in bigger super clusters, which are strung along vast filaments between virtually empty voids.

A. Deep Learning

The ability of deep neural networks to interpret visual data has made them famous. They have developed into a vital part of numerous computer vision applications during the last few years. In actuality, object detection has evolved dramatically as a result of deep learning. The performance in object detection significantly improved with the advent of the YOLO (You Only Look Once) and R-CNN (Region Based Convolutional Neural Network) families. Convolutional layers are the most common type of neural network utilized for image-related tasks. CNNs are the name for these neural networks (Convolutional Neural Networks). These CNN struly execute as liding window method in a natural and effective manner. It is a step in the neural network's process of learning representations of your photographs.

Researchers in AI at the University of California, Berkeley in 2014, proposed the Region-based Convolutional Neural Network (R-CNN).Three essential parts make up the R-CNN. An algorithm called "selective search" is used by a region selector to first find "regions of interest"—areas of pixels in a picture that might be representing items (RoI). The region selector generates around 2,000 regions of interest for each image. After being stretched to a predetermined size, the RoIs are input into a convolutional neural network. The CNN analyses each region individually to extract the features using a series of convolution operations. By just using interconnected layers, the CNN encodes the feature maps it into a mono vector of data variables. The encrypted values extracted from the CNN are therefore linked with the output classes to use a classifier machine learning model. Everything that is not an object is labeled as "background" in a different output.

B. YOLO

In 2016, Facebook AI research and the Institute for AI proposed the algorithm.2016 saw the introduction of the "You Only Look Once" (YOLO) family of neural networks, which improved and extended the accuracy of deep learning-based object recognition. The major development in yolo is the collaboration of entire classification and detection process into a sole network. You Only Look Once (YOLO) accomplishes everything in a single run through a single network as opposed to extracting features and regions separately. YOLO is appropriate for applications that need real-time inference and can conduct object identification at frame rates for video streaming. Deep learning object detection has advanced significantly over the past few years, going from a patchwork of various parts to a single, effective neural network. Object-detection networks are One of the key components in many applications today. It's in your computer, phone, automobile, camera, and other devices.

C. Convolution Neural Network

In a regular Neural Network there are three types of layers:
Input Layer: The input layer transmits data to the hidden layer.
Hidden Layer: There might be a huge amount of hidden levels, depending on our model and the volume of data. Each hidden layer might just have a distinct number of neurons, but they are frequently greater than the number of parameters. The network is nonlinear because the output of each layer is determined by multiplying the output of the layer underlying it by its learnable weights, adding learnable biases, and then computing the activation function.
Output Layer: The output from the hidden layer is translated into the probability score for each class after being passed into a logistic function likes igmoidors of tmax. Convolutional layer, pooling layer, fully connected layer, dropout and activation functions are the typical 5 layers of CNN, which is comparable to neural networks but has a few extra layers.
Convolutional layer, this layer, a convolutional layer, is utilised to extract different information from the input images. Between the input image and a filter with a specific sized matrix, the mathematical operation of convolution is carried out.
Pooling layer, The major goal of the pooling layer is to cut the size of the binarized feature map in decreasing the computational requirements. This is conducted separately on each previous layer and by diminishing the relationships between layers. Fully Connected layer, consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. These layers are usually placed before the output layer and form the last few layers of a CNN Architecture.
Dropout, Dropout can contribute to clustering in the training examples when all the attributes are coupled to the Fully connected layers. When a given model performs so well on training data that it has a negative effect on the model's performance when applied to new data, this is known as over fitting. A dropout layer, which reduces the size of the model by removing a few neurons from the neural network during training is used to solve this issue..
Activation function, The activation function is among the most key parameters in the CNN model. They are employed to discover and approximation any type of continuous and complex link between network variables. In essence, it determines which model data should be transmitted ahead to the output layer and which data must be sent back to the network for additional processing.

II. LITERATURESURVEY

Shivani Dere et al. [1] Using the deep learning approach, the objective of galaxy classification and detection has been Accomplished. The galaxy can be classified as having a wholly spherical shape, acigar-like appearance, or a hybrid of the two. It has spiral, spiral barred, spiral, elliptical, and irregular features on the edge. A powerful transfer learning model will be developed and used to accomplish this goal. The objective is to investigate the immensity of the cosmos by comprehending the enormous amount of astronomical data and utilising a variety of machine learning approaches.
Yonghui Lu et al.[2] In order to convert a large object detection network into a single category object detection network, Yonghui Lu looks into a number of different techniques. The down sampling layer should be kept separate from the other network modules using a method for constructing a network topology. A strong YOLO-compact network is created for real-time item recognition in one category. The experimental findings point to the YOLO-compact network's excellent performance and comparatively tiny size. Jimenez, Manuel, et al. [3] The results allow us to conclude that convolutional neural networks offer the best trade-off between runtime and accuracy, even

though, depending on the classification problem math and, adding a significant amount of depth and complexity to the network does not always provide a significant improvement in its ability to predict.

Auto encoders also provide a workable alternative to feature extraction-based classification of these images. This is because they can separate the feature extraction and learning processes, which could be advantageous as the amount of data that needs to be sorted grows in the future. Last but not least, it has been shown that using amateur and professional label sets provided by citizen science projects to learn from them might ultimately produce extremely promising outcomes. In light of the work shown here, we hope to enhance the learning phase by in corporating both unlabeled data and varying levels of confidence in the labeling of the images.

M. Bahar Ulla and others [4] CPU Based YOLO is superior for real-time object recognition on non-Computers. Bahar Ulla comes to the conclusion that the right framework must be initially selected in order to successfully accomplish object identification job sona CPU. The YOLO approach must then be used correctly. The model can then be applied with a typical desk topor laptop to execute YOLO. As a result, real-time object detection can be applied to a number of activities, including facial recognition, traffic monitoring, and surveillance. The machine will complete the tasks by providing it with a special dataset in the future.

A thorough analys is of deep learning-based object recognition frameworks that address various sub-problems, such as occlusion, clutter, and low resolution, with varying degrees of R-CNN modifications is provided by Zhong-quin Zhao et al. [5] in their publication advancements in neural networks and related learning systems, which offer in sight ful information and direction for future development.

Noor Eldeen and others [6] the classification of galaxies is one of the strange issues that has drawn their of astronomers recently. Deep convolutional neural network science's break through has drawn more researchers into this field. Deep convolutional neural network architecture for classifying galaxies was presented in this paper. After training on 1356 images, it received a score of 97.272 Robert E.and others. [7] As demonstrated in this research, increasing the amount of the training data set is not always the most important objective to create a strong galaxy detector/ classificator; rather, a well-based data augmentation may provide more observable gains. The training set is more reliable and may be used with any telescope and band to apply to photos.

Muhammad Abd Elaziz and others [8] this work present a new method for the automatic recognition of galaxy form datasets leveraging a picture method. The proposed scheme is split into two stages: the first stage entails the recovery of a set of characteristics based on current form, hue, and values of parameters, and the second stage entails the identification of the most important characteristics using a binary sine-cosine algorithm. The correlations seen between qualities of the request galactic photograph and the properties of many other galactic imagery are determined in the second stage.

Including Joseph Redmon [9] A model for object detection already exists: YOLO. It is easy to build and can be quickly trained on uncompressed pictures. The complete model is simultaneously trained, and YOLO is trained using a loss function that is mostly related to detection performance. I.M.Selim and others [10] Using a supervised machine learning system based on the Non-Negative Matrix Factorization algorithm, the research proposes a computer-based method for classifying the morphology of galaxies. This system can automatically recognize images of spiral, elliptical, and lenticular galaxies based on similarities. This model's accuracy in classifying spirals and ellipses, as well as the remaining objects, is97percent.

John Jenkinson et al. [11], Image processing is automated using the proposed paradigm. Prior to classification, the images underwent background removal using threshold in gandstarandarte fact removal using the morphological opening technique, both of which provide sparse at a sets.

III. PROPOSED SYSTEM

Megacosm uses YOLO algorithm and deep learning to detect and identify celestial objects using the data sets provided. A real-time object detection technique called YOLO makes use of neural networks. The prominence of this algorithm is due to its accuracy and rapidity. Deep learning is a sub hierarchy under Artificial Intelligence and an important telement of data science which deals with automating predictive analytics. Software architecture consists of 4 stages: The first stage is the astronomical data collection, where relevant images of cosmic or celestial objects are taken from the Hubble deep field in the JPG format as it provides highly compressible images. The second stage is Pre-processing of Data, having two sub stages namely Image Annotation and Data Augmentation.

In image annotation we use the Label Img image annotation tool to annotate the obtained dataset and in Data Augmentation we can resize the image as per desired manner and also add image properties such as enhancement and saturation. The third stage is Training the model, here we will be using Yolo algorithm to train the model and it considers the properties like image-size, batch-size, configuration which is to make a file that support the yolo to train, also take the consideration of data and epochs (show many times the machine learning algorithm has gone over the full training dataset.) The last stage is to detect and identify the object, the trained model from the third stage is now put up against the test data including the criteria of image-size and confidence level that is intersection over union between the predicted box and the ground truth its value is zero if no matched parameter is found. Once these processes are done the model is ready to detect and identify the object.

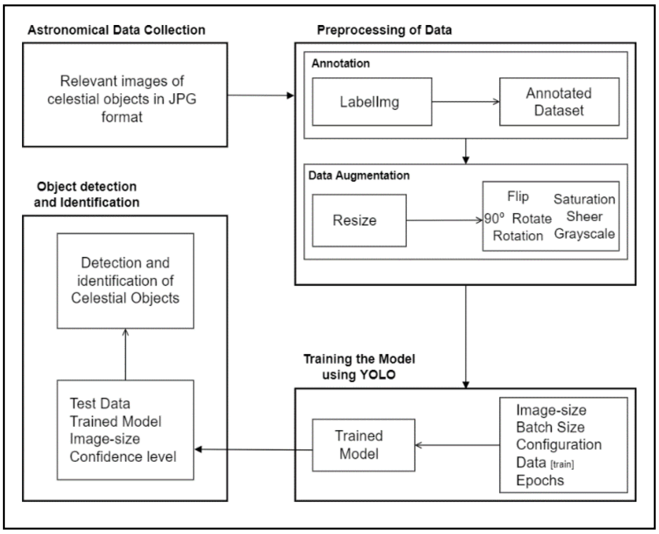


Fig1: System Architecture

A. Implementation

Astronomical Data Collection Megacosm assembly data manually downloaded from Hubble Deep Fields. Photos of stars, star clusters, planets, galaxies, satellites, comets, fireballs, asteroids, blackholes, and nebulae were among the images of collected data that were gathered from various online sources.

Data Preprocessing Building a machine learning model requires the preprocessing of data, and the quality of the preprocessing determines the model's performance. A Lable Img tool was used to annotate the collected photos. Multiple item photos can be annotated using the Lable Img tool. The data was labelled with the following classes using this tool: stars, star clusters, planets, asteroid, etc. An image is annotated by creating a text file with the same name as the image that contains the bounding box coordinates, the class that the bounding box belongs to,

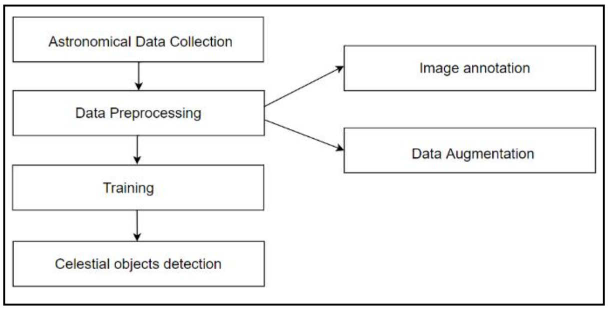


Fig2: Methodology

B. Dataset

Data preprocessing using Robo flow is applied to the annotated image. A functioning on line application called Roboflow offers the ability to prepare data in order to train a model.416416 is the new picture size. Three sets of the scaled photos are created: train, test, and validation. 87, 8, and 4 percent, respectively, of the data in Megacosm are divided across the train, validation, and test. To expand the dataset size and its horazine, the test data is then put through a data augmentation technique. The rotation ranges from -15° to +15° for the photos. A few photographs were upside down, 90 degrees counter clock wise, and rotated. To capture the entire visual appearance of the galaxy, the photos were flipped both horizontally and vertically. As hear effect was used on several of these photographs to alter their look. Some of these photographs had the shear effect applied to them to distort their appearance and make the model more resilient. Some pictures were cropped, saturated, and made brighter and darker. The train data was scaled by three using this data augmentation process

Training The development environment was then given the enhanced data. Three distinct files—test, train, and valid—were used to hold the data. Megacosm received instruction using the YOLO model. Highly precise object identification is produced using the YOLO design, which is an extremely effective and quick network. The model was trained using test images with a size of 416416 and the classes, as well as the YOLO configuration for more than 500 epochs. The entire training process took more than an hour. The weights file containing the trained model was then used to validate it and perform astronomical object recognition and identification. This model is used by Megacosm to give an effective model for astronomical object detection and identification.

Detection and Classification of Astronomical Objects: The model is now ready to detect the celestial object present in the image and classify the same for the class it belongs to, along with the confident score.

Since object detection falls within the category of supervised machine learning, you must train our models on samples that have labels. The boundaries and classes of the objects in each image in the training dataset must be included in a separate file. There are numerous open-source programmes available for making annotations for object detection. The annotated data is used to train the object detection network to identify regions in images that correspond to different types of objects. The model makes use of the unprocessed picture data from "The Hubble Deep field," where photos were taken with the Hubble telescope and annotated with Label Img.

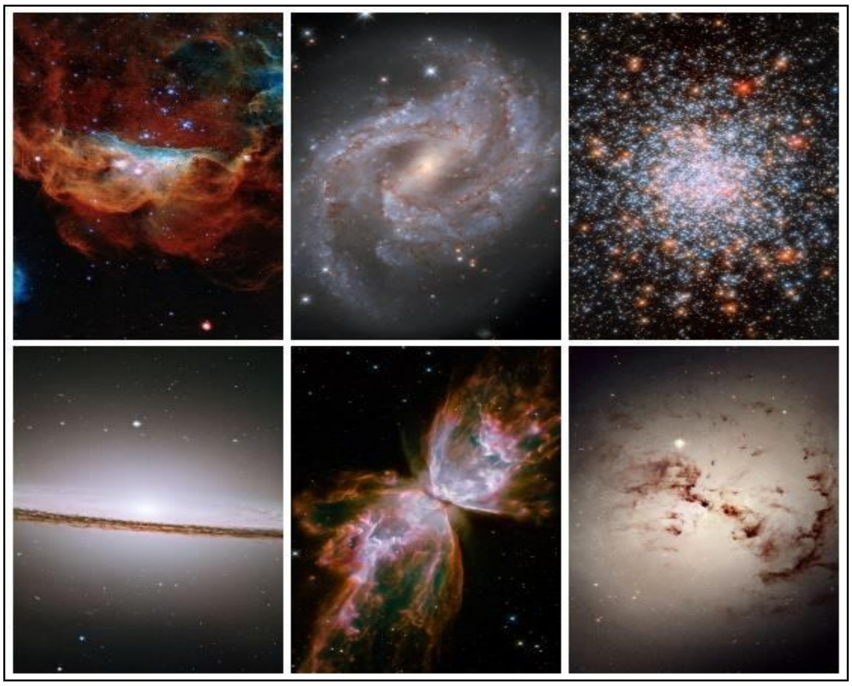


Fig3: Raw Images from Hubble Deep Field

IV. TESTING

Table1: Annotation

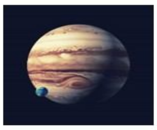
Test	Test Data	Expected Result	Actual Result	Pass/Fail
1.	StarCluster 	Coordinates of the particular class of star cluster	Coordinates of the particular class of star cluster	Pass
2.	Planet 	Coordinates of the particular class of planet	Coordinates of the particular class of planet	Pass

Table2: Augmentation

Test	Test Data	Expected Result	Actual Result	Pass/Fail
1.	Dataset of Images	Resizing of images to 416 X 416	Resizing of images to 416 X 416	Pass
2.	Dataset of Images	90° Rotate: Clockwise, Counterclockwise, Upside Down, Shear, Rotation, Grayscale,	90° Rotate: Clockwise, Counterclockwise, Upside Down, Shear, Rotation, Grayscale,	Pass
		Saturation, Hue, Flip, Brightness, Cutout, Bounding Box	Saturation, Hue, Flip, Brightness, Cutout, Bounding Box	

Table3: Detection and Classification

Test	Test Data	Expected Result	Actual Result	Pass/Fail
1.	StarCluster 	Accurate detection and identification of star cluster and its class	Approximate detection and identification of star cluster and its class	Intermediate
2.	Planet 	Accurate detection and identification of planet and its class	Approximate detection and identification of planet and its class	Pass

Annotation is a machine level way to explain the model through diagrams or text about the image. In our case, we usually draw boxes around the observed objects called the bounding boxes to mark the territory of the object. With this the model learns more about the things which it will be dealing with rather than the whole image which is usually useless at times when compared to the actual working space.

Augmentation is to deal with accuracy of the predicted object. One way to handle it is by increasing the size of the dataset is by augmentation. This involves orienting a particular image in multiple ways like increasing the saturation of the original image, gray-scaling and rotating the image in different quadrants so that the model can detect the same object in different orientation. To fully capture the visual aspect of the astronomical objects, the photos were rotated both horizontally and vertically. In order to strengthen the robustness of the model, several of these photographs had the shear effect applied to distort their look. When an image is annotated, a text file with the same name as the image is created that includes the bounding box coordinates, the class that the image is connected with, and the index of the class items. Data preprocessing using Roboflow is applied to the annotated image. A functioning online application called Roboflow offers the ability to prepare data in order to train a model. 416416 is the new picture size. Three sets of the scaled photos are created: train,test, and validation.

V. EXPERIMENTAL RESULTS

The model detects the given input data ad classify them according to its classes with confidence score which tells us the accuracy of the detected objects.

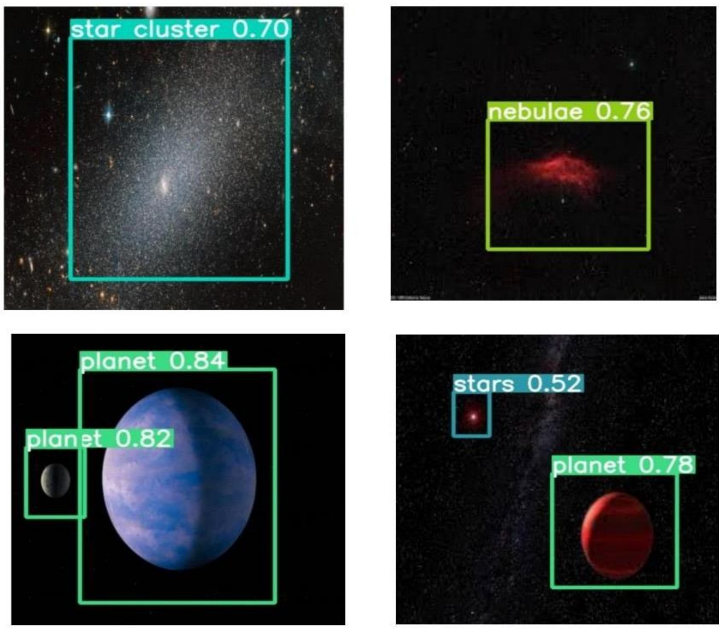


Fig4: Detection of Objects along with their confident score

VI. CONCLUSION

Megacosm analyses the provide draw astronomical photos to find the celestial objects. The use of this tool will aid integral arctic travelers in mapping their course and coordinates. It might beam an oeuvre for distant space travel. The discovery of celestial objects enables us to boldly travel to uncharted new worlds, search for new life, and discover new civilizations..These methods make a significant Contribution to several fields of research where it is necessary to remove enormous amounts of unnecessary data. With the help of this, we may take a closer look at pertinent observations and draw the appropriate conclusions. As a result, leveraging computer powerlessness the need for researchers to carefully examine each piece of acquired data. The models were only created to combine the capabilities of machine learning and astronomy, which may be their biggest short coming. There isn't a formal frame work in place for this, however there are machine screated to explore machine learning's potential in the study of astronomy and to utilize astronomical datasets.

REFERENCES

1. Shivani Dere, Maziya Fatima, Rutuja Jagtap, Unzela Inamdar,“Anomaly Detection in Astronomical Objects of Galaxies using Deep Learning”, 2021.
2. Yonghui Lu1, Langwen Zhang2, Wei Xie3, “YOLO-compact: An Efficient YOLO Network for Single Category Real-time Object Detection.”,2020
3. Manuel Jimenez, Mercedes Torres Torres, Robert John, Isaac Triguero, “Galaxy Image Classification based on Citizen Science Data: A Comparative Study”,2020
4. Md.Bahar Ulla,“CPU Based YOLO: A Real-Time Object Detection Algorithm”.2020
5. Zhong-QiuZhao, PengZheng, Shou-taoXu, and Xindong WuPeng Zheng, Shou-tao Xu, and Xindong Wu, Fellow, IEEE, “Object Detection with Deep Learning: A Review”,2019.
6. Nour Eldeen M. Khalifa, Mohamed Hamed N. Taha, Aboul Ella Hassanien “Deep Galaxy: Classification of Galaxies based on Deep Convolutional Neural Networks”
7. Rober to E.Gonz’alez, Roberto P.Mu˜noz, Cristian A.Hern´andez, “Galaxy detection and identification using Deep Learning and Data Augmentation”,2018.
8. Mohamed Abd Elaziz, I. M. Selim, Shengwu Xiong ,“Automatic Detection of Galaxy type from Galaxy datasets of Galaxies image based on Image Retrieval Approach”,2017
9. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You only look once: Unified, Real-Time Object Detection”,2016
10. I.M.Selim, Arabi E Keshk, Bassant M.El Shourbugy ,“Galaxy Image Classification using Non-Negative Matrix Factorization”,2016
11. John Jenkinson, Artyom Grigoryan, Mehdi Hajinoroozi, RaquelD́ıazHerná ndez, Hayde Peregrina Barreto, ArielOrtiz Esquivel, Leopoldo Altamirano and VahramChavushyan, “MachineLearning and Image Processing in Astronomy with Sparse Dataset”,2014