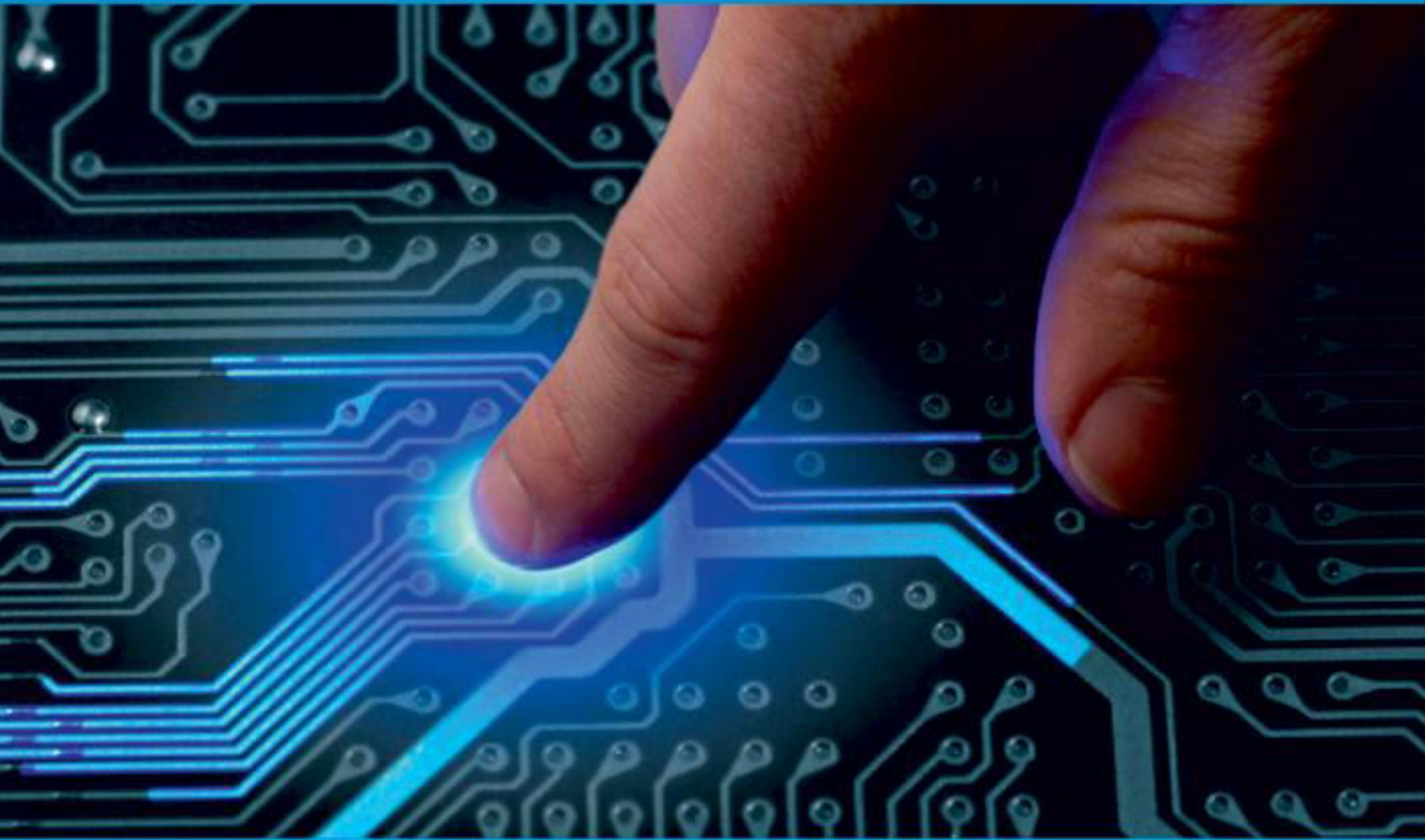
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907137 |

CASTOZERON: Galaxy Detection and Identification using YOLO

Koushik B, Vinay Kumar B L, Prof. Divyashree P

Department of Computer Science and Technology, Bangalore Institute of Technology, Bengaluru, Karnataka, India

ABSTRACT: The Inception of Everything starts at the single point called The Big Bang Theory. According to Georges Lemaitre in 1927, the expansion of the Universe is always increasing at the rate of cosmic acceleration. As a result, Astronomical objects took birth. One of the astronomical matters is The Galaxy.

Castozeron is the astronomical project used to identify galaxies and classify based on its type. This project is manually trained by using data annotation techniques and these data is enhanced in data augmentation step. This project uses YOLO algorithm and deep learning concepts to predict the results. It draws a bounding box with confidence value greater than threshold value and prints the class of the galaxy. The state-of-the- art feature of this project is the Star-ship Detection.

KEYWORDS: Galaxy, Spiral, Elliptical, Edge-on, starship,Galaxy detection, YOLO, classes, bounding boxes, annotations, confidence level, augmentation.

I. I. INTRODUCTION

The Inception of Everything starts at the single point called The Big Bang Theory. According to Georges Lemaitre in 1927, the expansion of the Universe is always increasing at the rate of cosmic acceleration. As a result, Astronomical objects took birth. One of the astronomical matter is The Galaxy.

The dawning point of all the questions about this Brobdingnagian Brunet is - "Look up the sky!". Those tiny effulgent dots fascinate you. What are those which looks small but maintains the laws of Nature? Yes! The answer is Stars and Galaxies. The Galaxy is a gravitational bound system of stars, stellar remnants, interstellar gas, dust and dark matter. The word Galaxy is derived from the Greek word 'Galaxias' literally 'Milky', a reference to the Milky Way. The dimension of galaxy is from dwarfs having 108 stars to giants having 1014 stars orbiting the galaxy's centre of mass. They are categorized according to their visual morphology as elliptical, spiral and irregular. Many galaxies are thought to have supermassive Black Holes at their centers. On September 25th , 2020, Hubble views a galaxy on the dark side ie., NGC 5585 which extends over 35,000 light-years across. When compared with galaxies of a similar shape and size, NGC 5585 stands out by having a notably different composition: it contains a far higher proportion of dark matter.

'Castozeron' provides an Astronomical Bonanza- The identification of galaxies which handcuffs the user to the application. Castozeron uses YOLO algorithm and Deep learning. The state-of-the-art feature of this project is the Star-ship Detection. The Castozeron is devoted completely to The Kingdom of Stars to twinkle in the tip of the eyes.

II. LITERATURE SURVEY

Jorge de la Calleja, Olac Fuentes.[1] The authors proposed a model uses machine learning and image analysis for performing automated morphological galaxy classification. It uses a neural network, and a locally weighted regression method, and implements homogeneous ensembles of classifiers.The performance of machine learning depends on the quality of the data. Difficulty in high precision analysis of galaxy morphology. ANN results will have less efficiency and accuracy.John Jenkinson, Artyom Grigoryan, Mehdi Hajinoroozi.[2] The authors proposed an idea to investigates the development of an automated classification system for galaxies in astronomical images based on the method of sparse representation.Main drawback of this method is pre-processing of unit galaxies.Evan Kuminski, Joe George, John Wallin and Lior Shamir.[3] This model uses citizen science data for training machine learning systems, and show experimental results demonstrating that machine learning systems can be trained with citizen science data.The performance of machine learning depends on the quality of the data. Difficulty in high precision analysis of galaxy morphology. I.M.Selim, Arabi E. Keshk, BassantM.ElShourbugy.[4] The authors proposed the model that uses Non-

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907137 |

Negative matrix factorization for morphological classification of galaxies.Main drawback of this method is pre-processing of unit galaxies. With the increase in classes, there is a decrement in the accuracy.Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. [5]YOLO provides an object detection system as a regression problem to detect objects in a image in a spatially separated bounding boxes and associated class probabilities. A single convolutionalneural network simultaneously predicts multiple bounding boxes and class probabilities for those boxes. YOLO trains on full images and directly optimizes detection performance. This spatial constraint limits the number of nearby objects that the model can predict. The model struggles with small objects that appear in groups. The model has a low recall and more localization error. Struggles to detect small objects.K. Shubhankar Reddy, K. Sreedhar.[6] This paper provides a review of the present technique that is relevance feedback schemes (RF) and analysis of Retrieval Systems based on Relevance feedback Techniques in content-based image retrieval (CBIR).The basic idea behind relevance feedback is to integrate human perception subjectivity into the query and involve user to evaluate the retrieval results.Nour Elden M. Khalifa, Mohamed Hamed N. Taha, Aboul Ella Hassanien, I. M. Selim. [7] The authors proposes a model using a Deep Convolutional Neural Network architecture for galaxy classification. The galaxy can be classified based on its features into main three categories Elliptical, Spiral, and Irregular.Main drawback of this method takes more time to make an inference.Mohamed Abd El Aziz, I. M. Selim, ShengwuXiong. [8] The author proposes an image-retrieval method to detect the type of galaxies within an image and return with the most similar image. The proposed methods take long inference time. Roberto E. Gonzalez, Roberto P. Munoz, Cristian A. Hernandez. [9] The author presents a method for automatic detection and classification of galaxies which includes a data augmentation procedure to make trained models more robust against the data taken from different instruments. This model trained using CNN and deep learning techniques.There is no morphological analysis and the model is less accurate in detection.Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, Xindong Wu. [10]Provides a introduction to deep learning and focuses on various object detection models to improve detection. It provides a review on deep learning based object detection frameworks.Manuel Jiménez, Robert John, Isaac Triguero. [11]This model performs automated classification of galaxy images exploring two distinct machine learning strategies, convolutional neural networks or feature extraction. This method pre-processes of unit galaxies.Sazida B. Islam, Damian Valles. [12] Theauthor proposes an automated wildlife monitoring system by image classification using computer vision algorithms and machine learning techniques.The model would perform lower for collected camera trap images in the site that has not encountered during model training.Hong Liu, Pinhao Song, Runwei Ding. [13] The author proposes to build a General Underwater Object Detector with small underwater dataset with limited types of water quality. It is highly difficult to do domain shifts.Mohana, HV Ravish Aradhya. [14]The author proposesa convolution Neural Network (CNN) model that is designed for urban vehicle dataset for single object detection and YOLOv3 for multiple object detection on KITTI and COCO dataset. This model requries large amount of training data.Akihiro Demachi, Shin Matsushima and Kenji Yamanishi. [15]The authors develop a novel algorithm for sparse non-negative matrix factorization (SNMF), which can discover patterns of web behaviours.

III. EXISTING SYSTEM

There exists no official system for this purpose, there are machine built for the purpose of exploration of the capabilities of machine learning in the field of astronomy and the use of astronomical dataset. The classification of galaxies helps in mapping the space and explores new frontiers. The main drawback is the data collection which will be solved through further development of technology within the next decade.

IV. PROBLEM STATEMENT

To develop a model that will detect and identify the Galaxies from the astronomical data using YOLO, an object-detection model.

V. PROPOSED SYSTEM

Castozeron proposes an Astronomical Bonanza for detection and identification of galaxies that takes astronomical images as input and performs the detection and identification on the input data which handcuffs the user to the application.

The data for Castozeron assemblages from Hubble Deep Fields through manual download. The collected data were images of galaxies mainly spiral, edge-on and elliptical. Along with galaxy starship images collected from various sources on the internet, the star ships that were collected were Starship USS Enterprise and Millenium Falcon. The collected images were subjected to annotation using a LableImg tool. LableImg tool is used to annotate images with multiple objects in them. Using this tool the collected data was annotated with following classes: spiral, elliptical, edge-

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907137 |

on, starship. The annotation of image produces a text file that is named after the image, and it contains the coordinates of the bounding box and the associated class with the index of the class elements.

The annotated image is subjected to data preprocessing using Roboflow. Roboflow is an online tool that provides functionality to perform data preparation to train the model. The images are resized to 416×416. The resized images are split into 3 sets: train, test and validation. In Castozeron 79%, 14%, 7% of the data are split among the train, validation and test respectively. The test data is then subjected to data augmentation procedure to increase the dataset size and its horazine. The images are rotated between -15° and +15°. Some images were rotated 90° clockwise, counterclockwise and upside down. The images were flipped horizontally and vertically to cover all visual appearance of the galaxy. Some of these images were applied shear effect to distort their appearance to increase the model's robustness. Some images were brightened and darkened, cropped, and saturated. Through this data augmentation procedure, the train data was scaled by thrice.

The augmented data was then transferred to the development environment. The data was stored in 3 different files: test, train and valid. The YOLO model was used to train Castozeron. The YOLO architecture is a very efficient and fast network that produces highly accurate object detection. The model was trained with the test data of images of size 416×416, and the four classes spiral, elliptical, edge-on and starship, along with the configuration of YOLO for over 500 epochs. The training took over an hour to complete. The trained model was then saved in a weights file which was further used to perform validation and use it to perform galaxy detection and identification. Castozeron uses this model to provide an efficient model to perform detection and identification of galaxies and starships.

There are **four modules** in Castozeronare:

- **I. Astronomical data assemblage**
- **II. Preprocessing of data**
- **III. Training the model**
- **IV. Galaxy Detection and Identification**

I. Astronomical data assemblage

This module requires the manual collections of Galaxy images in the form of .jpg and .jpeg format.The galaxy images were collected from Hubble Deep Field images.Hubble Deep Field provides images of astronomical bodies such as galaxies, comets, asteroids, planets and nebulas.These astronomical images are captured through the Hubble Telescope.Through the data assemblage component, the astronomical images gathered were 435 which includes spiral, elliptical and edge-on galaxies with star-ships.

II. Preprocessing of data

The aim of pre-processing is an improvement of the image data that suppresses unwilling distortions or enhances some image features important for further processing.LabelImg, a graphical image annotation tool is used to make bounding boxes on the collected dataset.LabelImg annotates images produces a text file that contains the coordinates for particular object of the class. Roboflow, is a platform which provides tools to perform data augmentation which scaled the dataset to 889 images.The data has been split into train, valid and test with 79%, 17% and 4% of the augmented data respectively.

III. Training the model

Once we have training data with annotated objects, it is loaded from Roboflow to the collab notebook.The model is trained by inputting the image size, batch size, the epochs, data, configuration.The YOLO algorithm is used to train the model.The number of training epochs is 500.The configuration file contains the anchors and the backbone of the neural network.

IV. Galaxy Detection and Identification

The trained model is then used to detect galaxies present in the image and will identify the class it belongs to.The test data is passed with the trained model, size, source of the test data and the confidence level of 0.4.The galaxies are detected and identified if the confidence level is above 0.4.These detected galaxies are annotated with bounding boxes with its class label and confidence level.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907137 |

. VI. SYSTEM FRAMEWORK

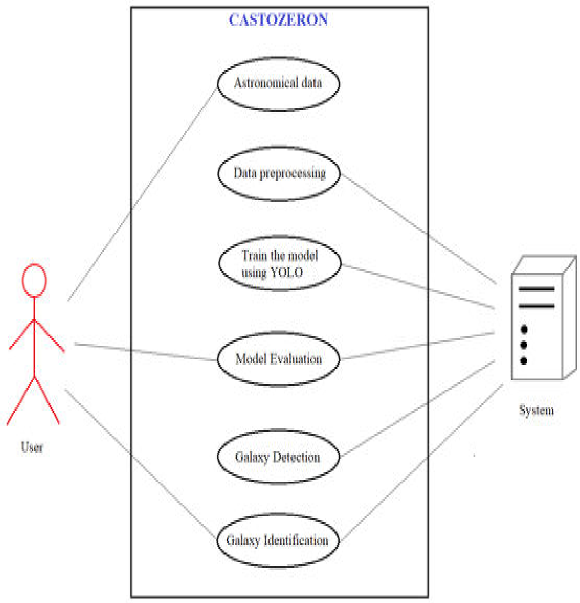


Figure 1. Use Case Diagram.

The use case diagram where the system and user are the two actors. The user should be able to upload the astronomical image on the user interface. The model will perform object detection on the input and user will be able to view the detected and identified galaxy.

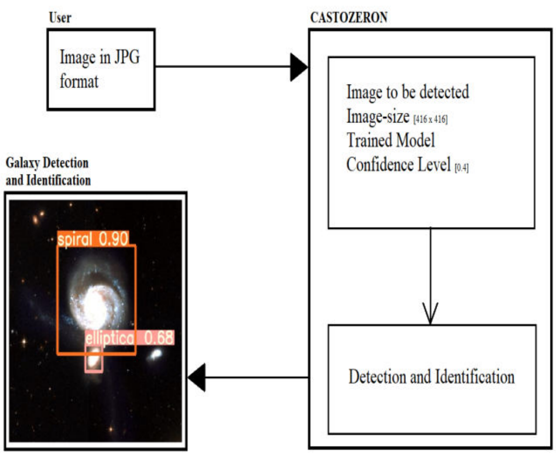


Figure 2. Interface Design.

Castozeron provides an interface through which the user can interact with the model. The user will be able to use the trained model to perform detection and identification. The user can input a image which will used to perform detection along with the trained model to detect and identify the galaxy or/and star ship with specified the confidence level or above.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907137 |

VII. SYSTEM ARCHITECTURE

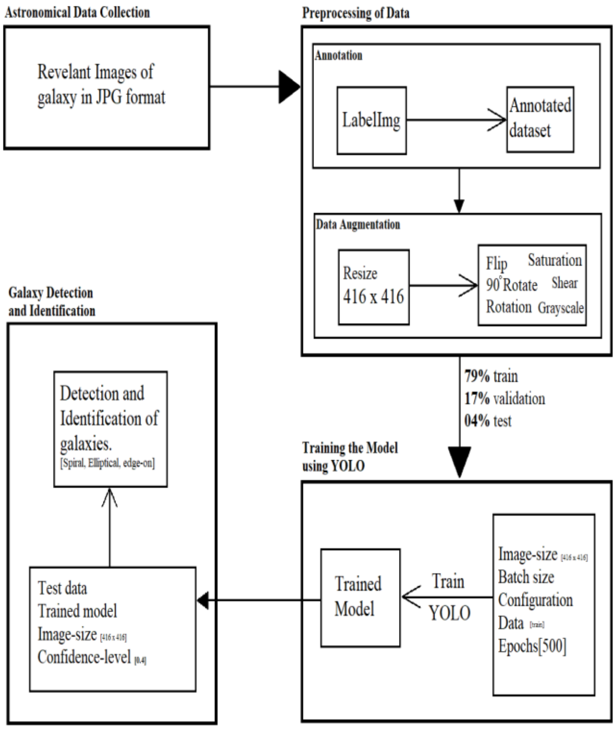


Figure 3. Architecture Diagram. The architecture diagram of Castozeron. The system takes astronomical images which is then annotated using LabelImg tool and then data augmentation is performed like resizing, rotating, shearing, etc., on training data using Roboflow to obtain data in 3 sets i.e., train, validation, and test. Later, the model is trained with image size 416 x 416, with the configuration of YOLO for over 500 epochs upon training data to obtain a trained model. The trained model is then used to perform detection and identification of test data with confidence level of 0.4.

VIII. RESULTS

The model was trained on 700 images over 500 epochs to get a precision of 0.543 and recall of 0.598 with mean average precision of 0.53. The model was able to perform multiple detections along with single detections of galaxies and starships.

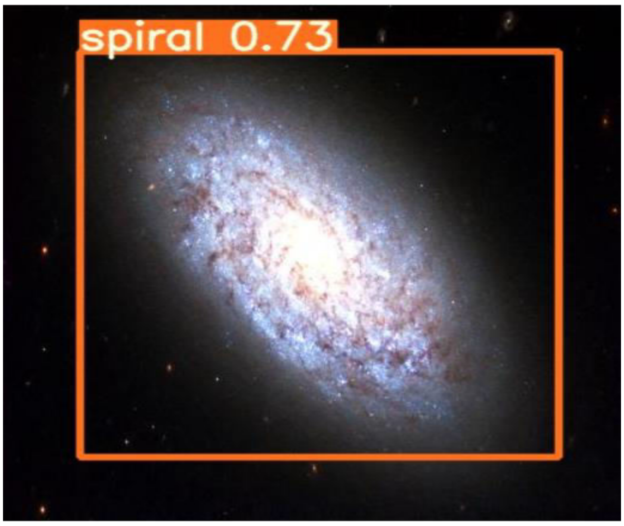


Figure 4. Detection of spiral galaxy. The detection of spiral galaxy with the confidence level of 0.73. The detected galaxy is identified with the bounding box around the detected galaxy along with the above confidence level which is

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907137 |

above the threshold mentioned.

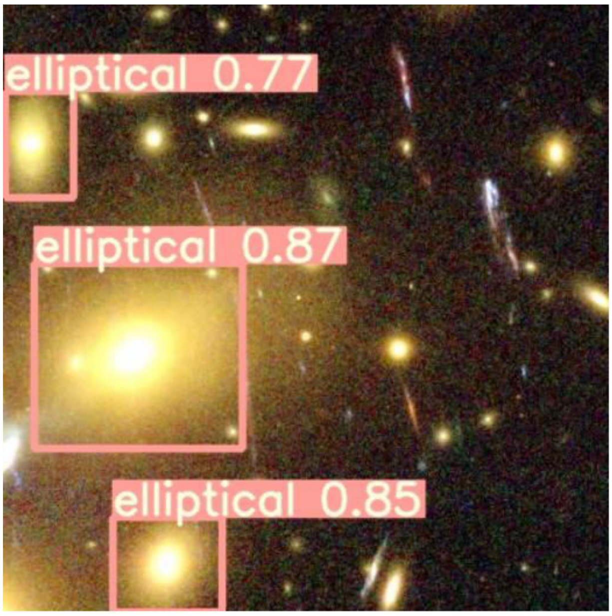


Figure 5. Detection of elliptical galaxies. Multiple detections of elliptical galaxies with the confidence levels of 0.77, 0.87 and 0.85.The detected galaxies are identified with the bounding box around each detected galaxy along with the above confidence level which is above the threshold mentioned.

Figure 6. Detection of Starship. The detection of starship with the confidence level of 0.50. The detected Starshipis identified with the bounding box around the detected starship along with the above confidence level which is above the threshold mentioned.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907137 |

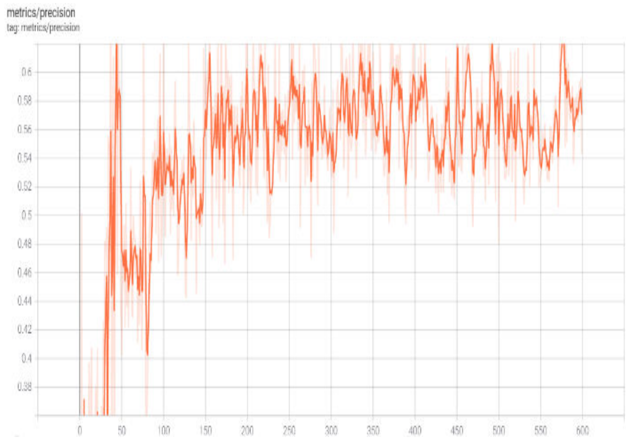


Figure 7. Precision graph.

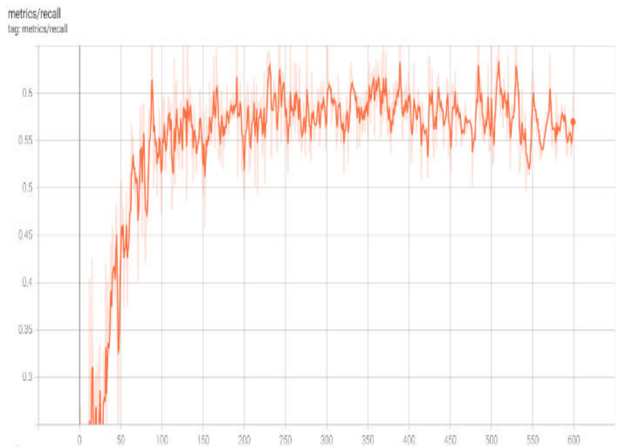


Figure 8. Recall graph

IX. APPLICATIONS

1. This tool could be the maneuver for deep spacefaring.

2. This tool will help space travelers to map course and coordinates during their intergalactic travel.

3.The detection of galaxies could help estimate the probability of finding alien lifeforms.

4.The detection of galaxies allows us to explore strange new worlds, to seek outnew life and new civilizations, to boldly go where no man has gone before.

X. CONCLUSION

In this paper, we have provided theCastozeron which in turn provides the Astronomical Bonanza - The user's state will be catatonic to the application due to identification if galaxies.The Castozeron is devoted completely to The Kingdom of Stars to twinkle in the tip of the eyes.

XI. FUTURE ENHANCEMENT

Castozeron should be enhanced to detect and identify wide range of astronomical objects and celestial bodies.The performance aggrandizing of Castozeron with quality data.

REFERENCES

[1]. Lupton, R., Blanton, M.R., Fekete, G., Hogg, D.W., O'Mullane, W., Szalay, A., Wherry, N., 2004. Preparing Red-Green-Blue Images from CCD Data. PASP 116, 133–137. doi:10.1086/382245, arXiv:astro-ph/0312483.

[2]. K. Huang and S. Aviyente, "Sparse representation for signal classification," in In Adv. NIPS, 2006.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907137 |

[3]. J.-L. Starck and J. Bobin, "Astronomical data analysis and sparsity: from wavelets to compressed sensing," Proceedings of the IEEE, vol. 98, p. 6, June 2010.
[4]. M. Elad, M. A. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," Proceedings of the IEEE - Special Issue on Applications of Sparse Representation and Compressive Sensing, vol. 98, pp. 972–982, April 2010.
[5]. D. Cai, H. Bao, and X. He, "Sparse concept coding for visual analysis," Proc. IEEE Conf. Computer Vision & Pattern Recognition Machine Learning (CVPR'11), 2011
[6]. Machine Learning and Image Processing in Astronomy with Sparse Data Sets. Raquel Diaz Hernandez, Hayde Peregrina Barreto, Ariel Ortiz Esquivel, Leopoldo Altamirano and Vahram Chavushyan,2014.
[7]. Alam, S., Albareti, F.D., Allende Prieto, C., Anders, F., Anderson, S.F., Anderton, T., Andrews, B.H., Armengaud, E., Aubourg, E., Bailey, S., et al., 2015. The ´ Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III.
[8]. Dieleman, S., Willett, K.W., Dambre, J., 2015. Rotationinvariant convolutional neural networks for galaxy morphology prediction. MNRAS 450, 1441–1459.
[9]. Galaxy detection and identification using deep learning and data augmentation. Roberto E. Gonz´aleza,b, Roberto P. Mu˜noza , Cristian A. Hern´andezb.2018.
[10].Galaxy detection and identification using deep learning and dataaugmentation. Roberto E. Gonzalez, Roberto P. Munoz, Cristian A. Hernandez,2018.
[11].Detection with Deep Learning: A Review. Zhong-Qiu Zhao, Peng Zheng, Shou- tao Xu and Xindong Wu,2019.
[12] M. Marin, L. E. Sucar, J. A. Gonzalez, and R. Diaz, A Hierarchical Model for Morphological Galaxy Classification, in Proceedings of the TwentySixth International Florida Artificial Intelligence Research Society Conference, 2013, pp. 438443.
[13] I. M. Selim, A. E., and B. M.El, Galaxy Image Classification using Non-Negative Matrix Factorization, Int. J. Comput. Appl., vol. 137, no. 5, pp. 48, Mar. 2016.
[14] I. M. Selim and M. Abd El Aziz, Automated morphological classification of galaxies based on a projection gradient nonnegative matrix factorization algorithm, Exp. Astron., vol. 43, no. 2, pp. 131144, Apr. 2017.
[15] Baillard, A., Bertin, E., de Lapparent, et. al.: Astron Astrophys 532 (2011)

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi crossref

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details