iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** V **Month of publication:** May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5215

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

Classification of Galaxies using Convolutional Neural Networks (CNN)

Teena Varma[1], Darryl Fernandes[2], Vishal Dube[3]
[1, 2, 3]*Xavier Institute of Engineering, Mumbai University*

Abstract: ***The universe is a gigantic ever-expanding mess. To classify it, Is a cosmologist's nightmare. There are numerous classes and subclasses of galaxies. Previously hundreds of thousands of volunteers helped classify millions of these images by eye. But with growing data, it becomes unfeasible to this manually. By using machine learning, we are automating the process of classifying galaxies using deep convolution neural network. The galaxies can be classified based on its features into 3 main categories: elliptical, Spiral and Irregular. By using machine learning we can reduce the human labor and the time required to complete such a herculean task.***

Currently this project only classifies for 3 broad classes and not for all the subclasses, so we are planning to expand this project to categorizing these classes of galaxies into their respective subclasses. We also hope to extend this project to include the classification of various other entities that make up our universe such as supernovae, pulsar stars, etc.

Keywords: Galaxy Classification, Hubble Classification, Convolutional Neural Networks (CNN), Image Pre-processing, Machine Learning (ML), Deep Learning.

I. INTRODUCTION

A galaxy is a set of stars , dust and interstellar gases whose cohesion is ensured by gravitation . The galaxies have a great diversity in size (between 2,000 and 500,000 light years in diameter) and in shape. The radiation from the galaxies makes it possible to distribute the latter into normal and active galaxies, among which are the quasars . The groupings of galaxies that we observe in the universe are called clusters and super clusters .

There exist hundreds of billions of galaxies scattered throughout the cosmos which basically define the structure of the universe on the macroscopic scale. Astronomers and cosmologists are dependent on acute observational studies and classification of these properties to inform their theoretical models of the Universe and propel the field of Astronomy further into previously uncharted territories.

Galaxy classification has gone beyond the realm of a few thousand galaxies to that of a million galaxies through the Galaxy Zoo project. Galaxy Zoo has taken morphology from the exclusive practice of a few experts to the public at large, thus facilitating citizen science at its best.

The advent of new technology such as powerful new telescopes and advanced camera modules has made large scale survey of the universe a straightforward task and has catalogued extra-terrestrial objects at an unparalleled rate. The mind bogglingly variety of galaxies and objects ought to be classified for it to be helpful in any context. The methods for classification need to be enhanced rapidly as the next generation of surveys such as Large Synoptic Survey Telescope (LSST) will accumulate more data exponentially at a neck breaking pace.

One of the chief way's cosmologist extract information from the luminous intensities of images of galaxies is by looking at its morphology. Cosmologist bifurcate galaxies into morphological categories; the distribution of morphologies across space-time tells us about the evolution of the cosmos. Out of the many classification methods, the most famous method is the "Hubble Tuning-Fork" model. The Hubble classification is used to categorize galaxies into elliptical, spiral, or irregular morphologies.

In this project we are using the same algorithms used by cosmologist to classify galaxies. In our project we will be using image preprocessing to reduce the background noise of the image and then feeding it to a convolutional neural network which will appropriately classify the image into its type.

Convolutional Neural Network are a deep learning concept that are designed to process data through multiple layers of arrays. The main difference between CNN and other neural networks is that CNN works on 2D arrays and operates directly on images as opposed to the emphasis on feature extraction by other neural networks while working on images.

CNN uses spatial correlation. Each concurrent layer of the neural network connects some input neurons. The hidden neurons process the input, not realizing te changes outside the specific boundary

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

II. PROPOSED WORK

A. Overview

The main aim of this project is to automate the classification of galaxies. We have implemented it by gathering our dataset from the galaxy zoo challenge and the NASA Hubble site. We then clean the dataset and manually classify it into the 3 categories. Then we run a simulation we run on a convolutional neural network can be multi-layer with 3-4 hidden layers and 3 classes or categories with RELU (Rectified Linear Unit) activation function. The loss function used will be Adams optimizer and categorical cross entropy.

B. Proposed Steps of the Project

Our project consists of 5 Steps namely:

1) *Data Pre-Processing*: Cleaning the images, reducing background noise and choosing suitable images.
2) *Model Selection*: Manually classify images into the 3 categories and selecting the suitable parameters for the CNN.
3) *Training the dataset*: Feeding the classified images to the CNN and running the simulation.
4) *Testing the dataset for appropriate accuracy*: Using the remainder of the dataset to predict the class of the galaxy and check for accuracy.
5) *Implementation:* Using the model to predict the class of a galaxy of an image previously unseen by the model.

a) *Data Pre-Processing:* The dataset containing the images of the galaxies was obtained from the Galaxy Zoo – The Galaxy Zoo Challenge from Kaggle and the NASA Hubble-Space Galaxy Website. The dataset was then manually classified into 3 sets (Irregular, Spiral, Elliptical) and into 2 Types: Training set and Test Set. These 2 folders are also divided into 3 categorized folders.

TABLE I
SUMMARY OF DATASET

Classes	Total Images	Training Set	Test Set
Spiral	1464	1000	464
Elliptical	1464	1000	464
Irregular	1686	1232	454

b) *Model Selection*: Our Convoluted Neural Network (CNN) model was implemented using python as a programming language and the Keras framework. The architecture of our model is as follows:

The model consists of 1 input Convolution 2D layer. 4 hidden layers and finally 1 penultimate dense layer before the output layer followed by the output layer. Each image was altered and rescaled to a resolution of 150 x 150 pixels. We used a batch size of 128 and trained it for 40 epochs with 15 timestamps per epoch. Dropout regularization was used after the hidden layers and before the output layer to increase the accuracy.

c) *Training the dataset:* For training our models we used an AMD Radeon R5 M330 2GB GPU followed by rigorous and in-depth training on 2 x Zotac GTX Titan X AMP edition running in SLI using an SSH shell to transfer data and communicate between the 2 machines. While training for 40 epochs, it was observed that the training accuracy is 97.43% with a loss of 2.57%. The training set contains a total of 3232 images of the 3 classes. Due to large dataset which is being used, it is not practical to train the model every time we use the models. Therefore, we trained the models once and saved the results in a weight.h5 file which will be used during the testing phase of our project.

d) *Testing the dataset for appropriate accuracy:* For the testing out our trained Convoluted Neural Network model we used a new unknown set of images of the 3 corresponding classes - Irregular, Elliptical and Spiral. After running our model on the Titan GPU for about 40 epochs we attained the following results:

©IJRASET: All Rights are Reserved

Training accuracy = 97.43%
After training the models and saving the model weights, we tested the model and attained the following results:
Testing accuracy = 94.89%

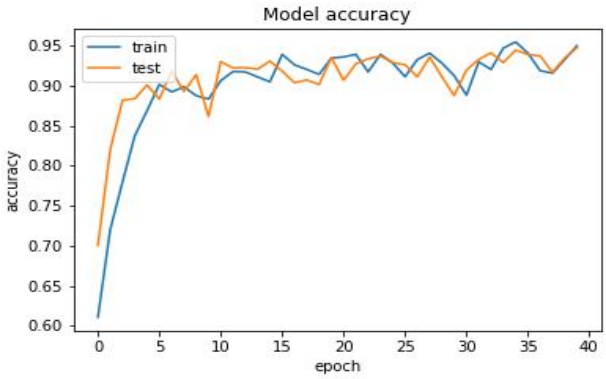


Fig. 2.4.1 Graph illustrating the Model accuracy

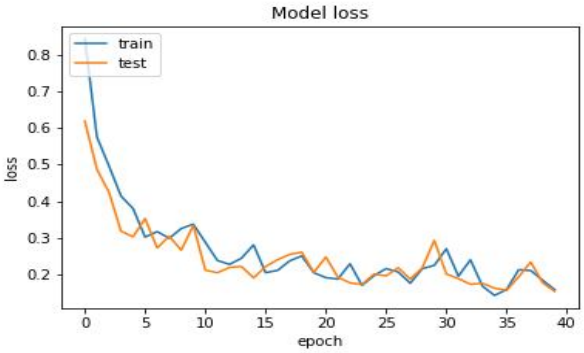


Fig. 2.4.2 Graph illustrating the Model loss

e) Implementation: The whole implementation of the cycle can be summarized in this flowchart.

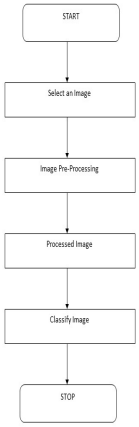


Fig. 2.5.1 Implementation flowchart

The Convolutional Neural Network constructed by us for this consists of 1 input Convolution 2D layer. 4 hidden layers and finally 1 penultimate dense layer before the output layer followed by the output layer.
While training, the classifier performs forward passes through our dataset (one complete pass through all data objects is referred to as epoch) multiple times over. Input images are being operated on by the layers of the network which, at this point, all have random weights; the CNN produces its first attempts at predicting class labels and scores.

©IJRASET: All Rights are Reserved

Then, based on the previous values received, the model computes the error using the loss function. It then compares its first outputs, the results of random initialization, to the actual labels to see how close (or far) it is from the desired values.

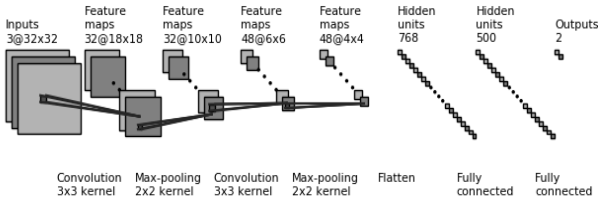


Fig. 2.5.2

Once the values are compared the error is being back propagated from the end to the start of the network so that the CNN can update its weights accordingly. This is typically done through one of Gradient Descent optimization (Adamax Algorithm was using by us in this project) algorithms.

The weights change only once per epoch and that's why training a CNN model might turn out computationally very expensive. Given that datasets for modern CNNs tend to be large, it takes lots of time and resources to get through each sample. The images from the dataset are passed thought the following filter in the first layer of the Convolutional Neural Network. This layer acts as a filter and helps to reduce noise and help in enhancing the desirable features of image which will be useful for classification.

The following image is the filter:

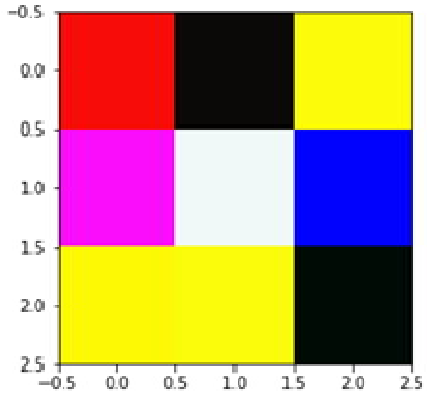


Fig. 2.5.3

The images from the training set will be passed through this filter and the resultant image will then we given to the CNN which will classify the images.

The following image is a representation of each type of galaxy after passing through the above filter.

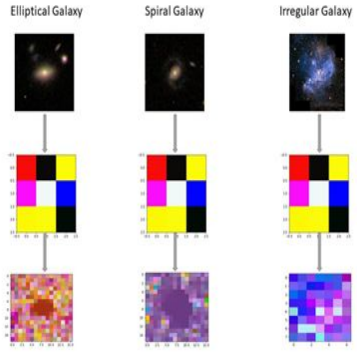


Fig. 2.5.4

©IJRASET: All Rights are Reserved

The images after passing through the filter will then be analyzed in the following manner by the CNN.

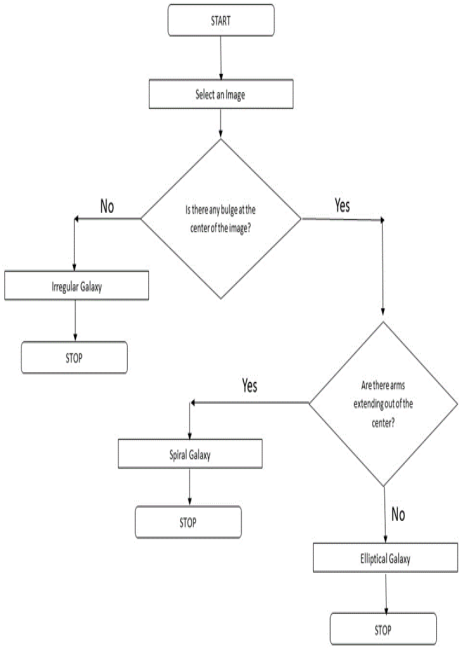


Fig. 2.5.5

Once the background is removed and the image is pre-processed, the CNN will scan the image for a central bulge.
If there no central bulge (bright central spot) observed in the image, then it will be classified as an Irregular galaxy as shown in the figure.
If there is a central bulge observed (bright central spot) observed then the algorithm check if there are arms extending from the central bulge outwards (spirals of the spiral galaxy), the image will be classified as a spiral galaxy.
If there are no arms and there is central bulge, then the image will be classified as an elliptical galaxy as evident in the figure.
After running the CNN over the complete Training and Test set, the following graph was obtained comparing the accuracy of the training and test set.

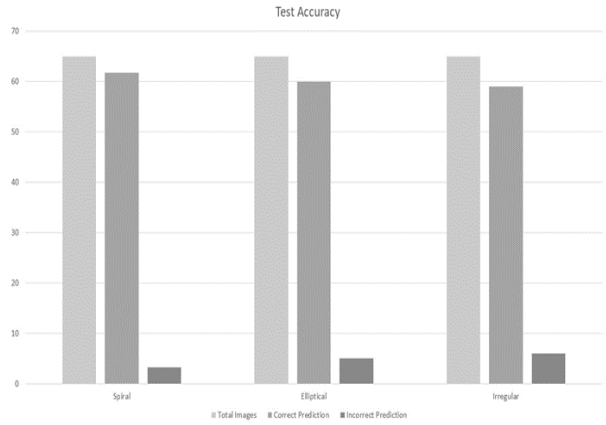


Fig. 2.5.6

The model took a total of 2.5 hours to run for 40 epochs on the Training set. The total validation time for the model to run over the complete test set was about 30 minutes. For prediction however, it takes about 5 minutes to classify 50 images as we have already stored the weights in a h5 file and use to for predictions.

©IJRASET: All Rights are Reserved

III. CONCLUSIONS

Due to the ever-increasing amount of data that will be generated in the future due to advancements in technology and other means this software will be a saving grace for the astrologist and cosmologist as it will reduce human labor and automate the process of classifying galaxies to a great extent.

Due to its high accuracy and superior speed as compared to other models available, the project can be used in scientific work as it reliable and fast.

The project can further be fine-tuned to classify galaxies based on the subclasses as well. Since we saved the training results on a separate file, this application can we used in real time scenarios as well where speed and accuracy are critically important.

IV. ACKNOWLEDGMENT

We would like to express our sincere and heartfelt gratitude to our teacher Prof. Teena Varma who gave us the golden opportunity to do this wonderful project on the topic Classification of Galaxies using Convolutional Neural Networks, which also helped us in doing a lot of Research and we learned about so many new things. We are thankful to her for sharing with us her knowledge and assisting us throughout this project.

REFERENCES

[1] John Kormendy and Ralf Bender., A revised parallel-sequence morphological classi_cation of galaxies: structure and formation of s0 and spheroidal gakaxies, The Astrophysical Journal Supplementary Series.,198(1):2,2011

[2] Ronald J. buta., Kartik Sheth., E. Athanassoula., A. Bosma., Johan H. Knapen., Eija Laurikainen., Heikki Salo., Debra Elmegreen., Luis C. Ho., Dennis Zaritsky, A Classical Morphological Analysis of Galaxies in the Spitzer Survey of Stellar Structure in Galaxies, The Astrophysical Journal Supplementary Series., 217(2):32, 201

[3] Lior Shamir., Automatic morphological classi_cation of galaxy images, Monthly Notices of the Royal Astronomical Society, 399(3):13671372,2009

[4] Edward J. Kim and Robert J. Brunner., Stargalaxy classi_cation using deep convolutional neural networks, Monthly Notices of the Royal Astronomical Society, 464(4), 1: 44634475,2017

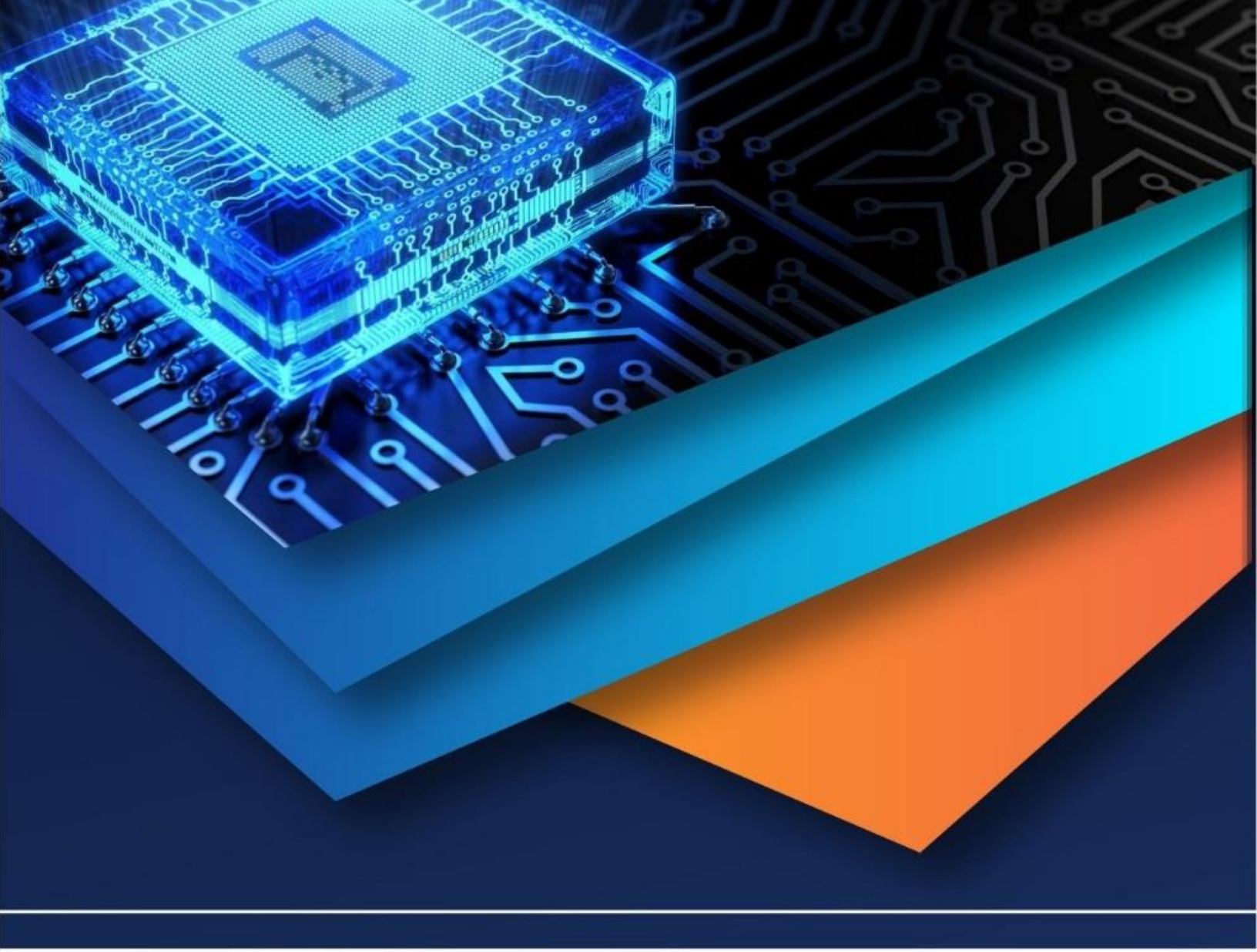
[5] Jorge De La Calleja and Olac Fuentes., Machine learning and image analysis for morphological galaxy classi_cationMonthly Notices of the Royal Astronomical Society, 349(1):8793, 2004

[6] Maribel Marin and L. Enrique Sucar and Jesus A. Gonzalez and Raquel Diaz., A Hierarchical Model for Morphological Galaxy Classi_cation, In FLAIRS conference, 2013

[7] I.M. Selim., Arabi E. Keshk., Bassant M. El Shourbugy., Galaxy Image Classi_cation using Non-Negative Matrix Factorization, International Journal of Computer Applications, 137(5), 2016

[8] I.M. Selim., Mohamed Abd El Aziz., automated morphological classi_cation of galaxies using projection gradient nonnegative matrix factorisation algorithm, Experimental Astronomy, 43(2):131-144, 2017

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)