Galaxies classification using Deep learning Algorithm in Convolutional Neural Networks

P. Archana[1] , B. Saketh[2] B. Rishi kiran reddy[2] K. Uday kumar[2] P. Manesh[2]
[1]Assistant Professor, Department of CSE,[2]Final year B.Tech. Students, Department of CSE,
Sreyas Institute of Engineering and Technology, Hyderabad, Telangana, India

Abstract

We present a framework to compose artificial neural networks in cases where the data cannot be treated as independent events, our particular motivation is star galaxy classification for ground based optical surveys. Due to a turbulent atmosphere and imperfect instruments, a single image of an astronomical object is not enough to definitively classify it as a star or galaxy. Instead the context of the surrounding objects imaged at the same time need to be considered in order to make an optimal classification. The model we present is divided into three distinct ANNs: one designed to capture local features about each object, the second to compare these features across all objects in an image, and the third to make a final prediction for each object based on the local and compared features. By exploiting the ability to replicate the weights of an ANN(Artificial Neural Networks), the model can handle an arbitrary and variable number of individual objects embedded in a larger exposure. We train and test our model on simulations of a large up and coming ground based survey, the Large Synoptic Survey Telescope (LSST) and compare to the state of the art approach, showing improved overall performance as well as better performance for a specific class of objects that are important for the LSST.

Keywords: Deep learning Algorithm,Neural Networks, Galaxy Image Classification.

1.Introduction

Galaxy morphological classification is done on large databases of information to help astrophysicists in testing theories and finding new conclusions for explaining the physics of processes governing galaxies,star-formation, and the analysis of the universe. Historically, galaxies classification could be a matter of visually inspecting 2-dimensional pictures of galaxies and categorizing them as they seem. This classification was thought of a long-run goal for astrophysicists. However, the difficult nature of galaxies and quality of pictures have created the classification of galaxies difficult and not correct. Galaxy classification system helps astronomers within the method of grouping galaxies per their visual form. The foremost notable being the Hubble sequence is considered one among the foremost used schemes in galaxy morphological classification. The Edwin Powell Hubble sequence was created by Hubble in 1926. In the past few years, advancements in machine tools and algorithms have begun to enable automatic analysis of galaxy morphology. They were unable to scale to a lot of larger pictures. With the massive increase within the process power, memory size and also the convenience of powerful GPUs and huge datasets, it absolutely has potential to coach deeper, larger and a lot of advanced models.

2.Aim

Galaxy Image Classification using a Deep Convolutional Neural Network is presented. The galaxy can be classified based on its features into three main categories, namely: Elliptical, Spiral, and Irregular. The proposed deep galaxy architecture consists of one input convolutional layer having 16 filters, followed by 3 hidden layers, 1 penultimate dense layer and an Output Softmax layer. It trained over 3232 images for 200 epochs and achieved a testing accuracy 97.38% which outperformed conventional classifiers like Support Vector Machine and previous research contributions in the same domain of Galaxy Image Classification.

3.Motivation

In the past few decades, the desire of humans to know more about other galaxies has increased and so has their efforts. We want to know the most fundamental 3 questions about our existence, how and why. A part of the answer to our question lies in how the galaxies originated and evolved over time. Different galaxies have varying shapes, sizes, colors and features and to solve the puzzle of formation and evolution of galaxies, we need to understand how we can infer the distribution, location and type of galaxies on the basis of their shapes, size and color. This, in turn, requires us to classify the galaxy images based on their shapes, sizes and other features. In earlier successful projects, hundreds of thousands of volunteers helped classify shapes of some millions of these images by eye. But with growing data, it became difficult to do this manually any more. So, an initiative was launched to find good automated metrics that could potentially be used to analyze the images of the galaxies and answer these questions. The Research will be helpful for astronomical scientists and cosmologists. It will help to classify a huge collection of Galaxy images without manual effort of viewing each image individually.

ISSN: 2278-4632
Vol-10 Issue-5 No. 6 May 2020

4.Proposed System

In this project, Galaxy Image Classification using a Deep Convolutional Neural Network is presented. The galaxy can be classified based on its features into three main categories, namely: Elliptical, Spiral, and Irregular. The proposed deep galaxy architecture consists of one input convolutional layer having 16 filters, followed by 3 hidden layers, 1 penultimate dense layer and an Output Softmax layer. It trained over 3232 images for 200 epochs and achieved a testing accuracy 97.38% which outperformed conventional classifiers like Support Vector Machine and previous research contributions in the same domain of Galaxy Image Classification.

4.1.Advantages of proposed system

- Higher Accuracy
- Least Error
- Faster Classification of Galaxies
- Time Saving
- Handle larger datasets

4.2.Architecture

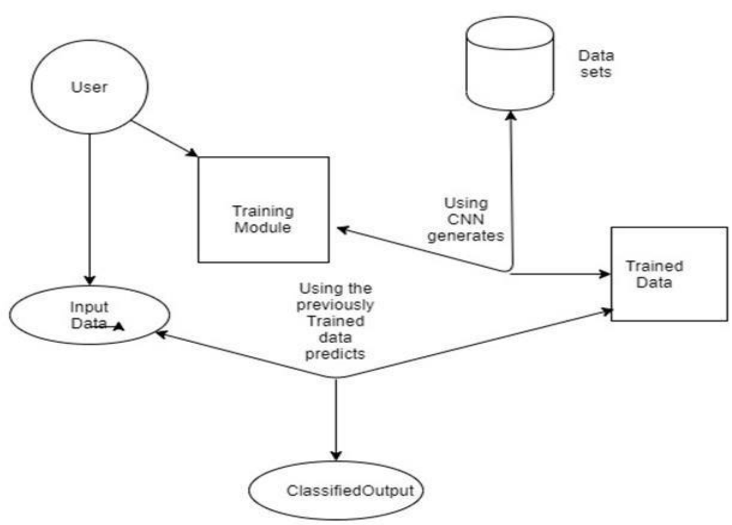


Figure 1: Architecture

5. Implementation

In this project, we classified Galaxy Image data into its 3 corresponding major classes - Elliptical type, Spiral type and Irregular type using a Deep Convolutional Neural Network (CNN) architecture.

5.1.Data Pre-Processing

The Dataset containing the Galaxy Images was obtained from Kaggle and NASA Hubble-Space Gallery Websites. The Dataset was categorized into 3 classes with images kept in two main folders: training and validation. The training folder is further subdivided into three subfolders for 3 classes: spiral, elliptical and irregular. Similarly, the validation folder is also having three subfolders with the same name. The number of images in these folders are listed in the table below:

Table 1: Different classes with number of images for training, validation and testing.

Classes	Total Images	Training Set	Validation Set	Testing Set
Spiral	1464	1000	400	64
Elliptical	1464	1000	400	64
Irregular	1686	1232	390	64

Initially, we had only 11 images for the Irregular class. We used the Augmenter Package of Python to perform Image Augmentation on those 11 images and generated 1615 augmented images.

5.2.Model Selection

We built our Convolutional Neural Network (CNN) model in Python using the Keras framework. The CNN architecture consisted of 1 Input Convolution 2D layer followed by 4 hidden layers, 1 penultimate dense layer and finally 1 output layer. We used a filter size of 3 x 3 in each layer. All the images were resized to 128 x 128 pixels. The batch size used was 64 and was trained for 40 epochs with 10 timestamps per epoch. We used Dropout

ISSN: 2278-4632
Vol-10 Issue-5 No. 6 May 2020

Regularization after the hidden layers and before the output layer. The detailed architecture diagram is shown in the figure below:

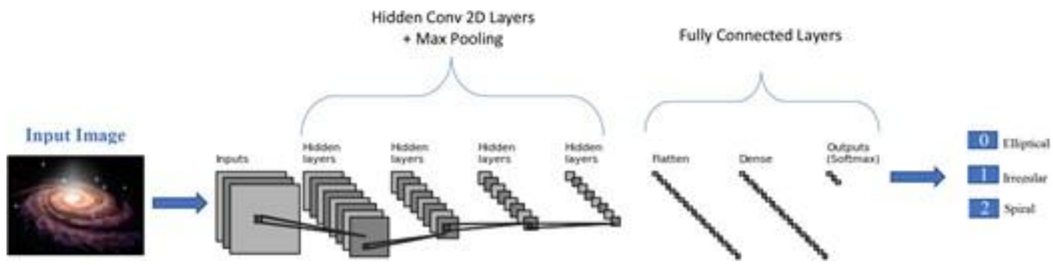


Figure 2: The architecture of Convolution Neural Network for galaxy classification.

5.3.Training Procedure

The model was trained on NVIDIA 960MX GPU followed by intensive training on the NVIDIA DGX 1 Octa Tesla V100 Supercomputer servers using technologies like Putty and WinSCP. On training for 40 epochs, it was observed the training accuracy was at 95.00% with training loss at 15.37% while Validation accuracy was at 94.75% and Validation loss at 15.31%. The Training set containing 3 classes were atotal of 3232 images while the Validation set containing the same number of classes contained 1190 images.After training the CNN model, we saved the weights of the model in weights.h5 file as we could now perform testing easily as many times we needed using the already trained model weights. This step was also performed as training took nearly 1.5 hours to complete and it was not feasible to train the model every-time for testing. Thus, this time was saved by making the weights.h5 file for testing.

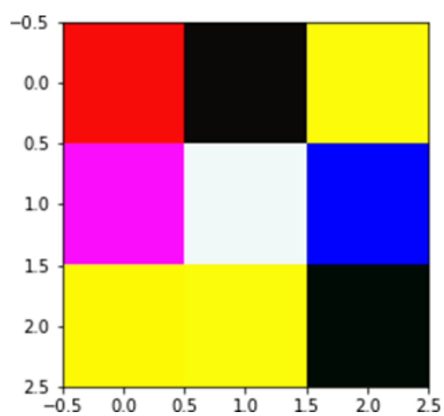


Figure 3: First Layer First Filter Visualization.

6. Implementation

We tested our trained Galaxy Classifier CNN architecture model on some new images of the 3 corresponding classes - Irregular, Elliptical and Spiral. After 200 epochs, running on the NVIDIA DGX 1 Octa Tesla V100 Supercomputer servers, we obtained:

Training Accuracy = 99.530%, Validation Accuracy = 98.480%, After training and saving the model weights, we tested the model and obtained: Testing accuracy = 97.398%. The model accuracy curve is shown below:

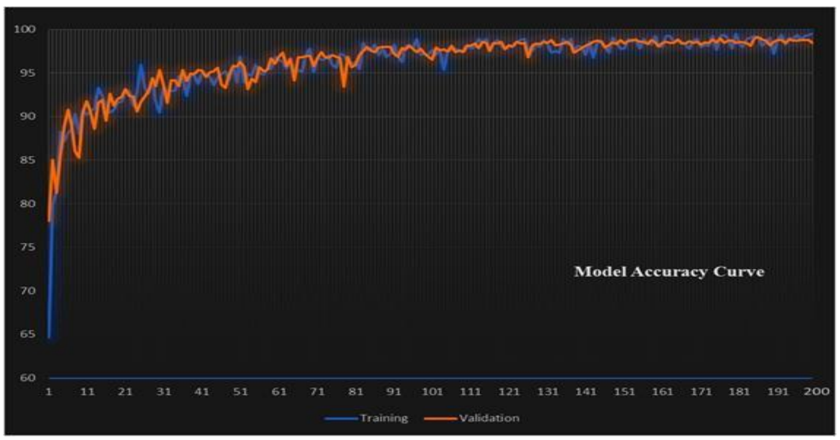
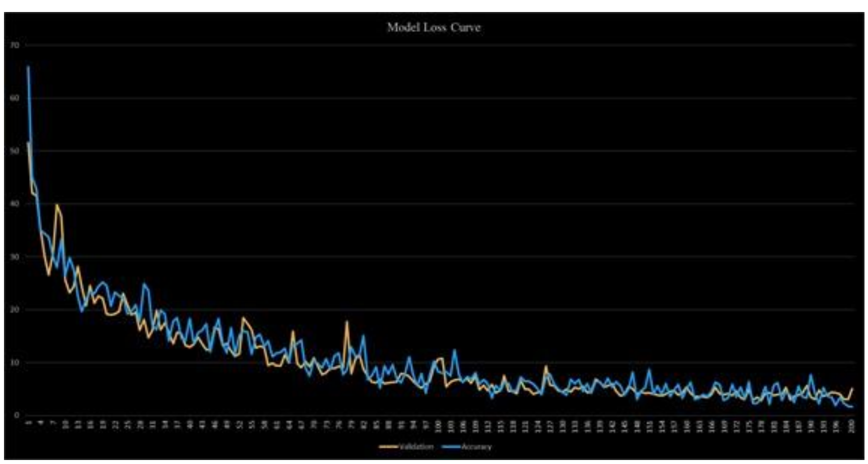


Figure 4: Model Accuracy Curve.



ISSN: 2278-4632
Vol-10 Issue-5 No. 6 May 2020

Figure 5: Model Loss Curve.

We used a Windows 10 machine having the following specifications: Intel dual- core i5 processor clocked at 2.40 GHz with 8 GB RAM and a dedicated Nvidia GeForce 940M GPU. We also trained the model on the Nvidia DGX 1 (8X Tesla V100) Supercomputer servers having 5120 Nvidia Tensor Cores and a computing power of 960 TeraFlops. The python code was written on Spyder 3.0.0, Jupyter 1.0.0 and Python 3.5.2. using Keras 2.1.3 framework. Many more python packages were needed and used time to time.

Upon testing, the results were obtained in the form of a list with each image having labels as 0, 1 and 2, where 0 stands for Elliptical type, 1 stands for Irregular type and 2 stands for Spiral type.

The sample output for 64 Elliptical images is shown below:

array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
dtype=int64)

Similarly, the outputs for all classes are found and accuracy is calculated.

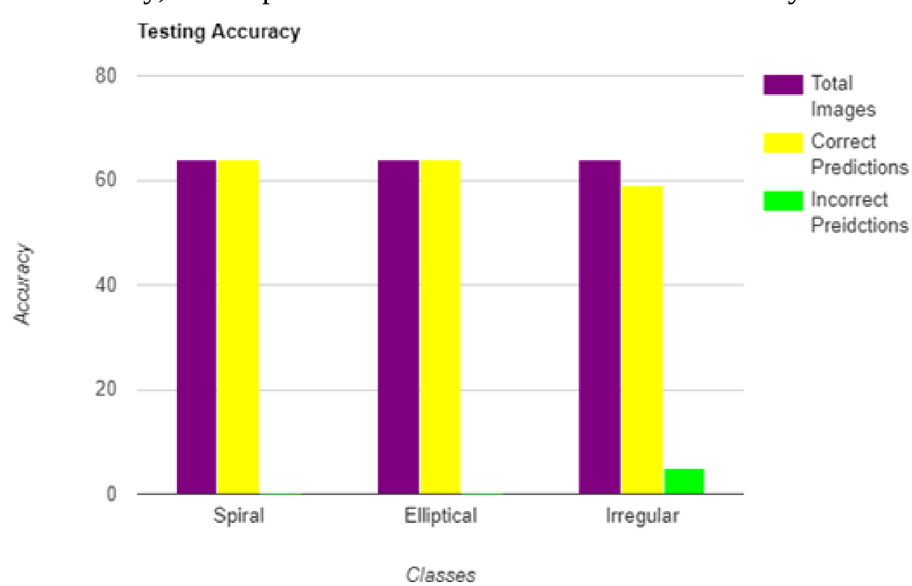


Figure 6: Testing Accuracy Graph for Different Classes.

7.Conclusion

The Research will be helpful for astronomical scientists and cosmologists. It will help to classify a huge collection of Galaxy images without manual effort of viewing each image individually. The Research can be fine-tuned for further classification of galaxies into their subclasses as explained in the Background section. The testing time was reduced to a few seconds by saving the CNN weights file and thus it will be working on real time scenarios also.

7.1.Limitations

- The project is classified only for three broad classes, not for all the subclasses.
- It has to be run on high speed GPUs in order to get maximum accuracy. Thus, the cost of computation is high and a good computation server is needed.
- Data is imbalanced.
- The model takes good time to train.

7.2.Future Enhancements

- The model can be fine-tuned more to make it more efficient.
- Data can be made balanced to avoid misclassification.
- Model could be made such that it can be run on normal CPUs, rather than running on GPUs in supercomputers.
- More accurate models can be prepared in the future using even fine parameters or new techniques.

8.References

[1] M. Abd Elfattah, N. El-Bendary, M. A. Abu Elsoud, A. E. Hassanien, and M. F. Tolba, An intelligent approach for galaxies images classification, in 13th International Conference on Hybrid Intelligent Systems (HIS 2013), 2013, pp. 167172.

[2] M. Abd Elfattah, N. Elbendary, H. K. Elminir, M. A. Abu El-Soud, and A. E. Hassanien,

Galaxies image classification using empirical mode decomposition and machine learning techniques, in 2014 International Conference on Engineering and Technology (ICET), 2014, pp. 15.

[3] Adams and A. Woolley, Hubble classification of galaxies using neural networks, Vistas Astron., vol. 38, pp. 273280, Jan. 1994.

[4] Dominguez, A History of the Convolution Operation [Retrospectroscope], IEEE Pulse, vol. 6, no. 1, pp. 3849, Jan. 2015.

[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE, vol. 86, no. 11, pp. 22782324, 1998.

[6] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. S. Kruthiventi, and

[7] R. V. Babu, A Taxonomy of Deep Convolutional Neural Nets for Computer Vision, Front. Robot. AI, vol. 2, p. 36, 2016.

[8] H. Jiang and E. Learned-Miller, Face Detection with the Faster R-CNN, in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017, pp. 650657.

[9] J. De La Calleja and O. Fuentes, Machine learning and image analysis for morphological galaxy classification, Mon. Not. R. Astron. Soc., vol. 349, no. 1, pp. 8793, 2004.

[10] 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.

[11] 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.

[12] 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.

[13] Baillard, A., Bertin, E., de Lapparent, et. al.: Astron Astrophys 532 (2011)

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

[15] 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,

[16] The Astrophysical Journal Supplementary Series., 217(2):32, 2015

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

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

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

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

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

[22] I.M.Selim., Mohamed Abd El Aziz. Automated morphological classification of galaxies using projection gradient non negative matrix factorisation algorithm, Experimental Astronomy, 43(2):131-144, 2017

