Morphology based Galaxy Classification

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

MASTER OF TECHNOLOGY IN COMPUTER SCIENCE

BY

Ayan Mukherjee
Roll. No. CS2305

UNDER THE SUPERVISION OF

Dr. Sarbani Palit
Computer Vision and Pattern Recognition Unit (CVPRU)
Indian Statistical Institute, Kolkata

Indian Statistical Institute, Kolkata

Certificate

This is to certify that the dissertation entitled "**Morphology based Galaxy Classification**" submitted by **Ayan Mukherjee** to the Indian Statistical Institute, Kolkata, in partial fulfillment of the requirements for the degree of Master Of Technology in Computer Science, is an authentic and genuine record of the research work carried out by the candidate under my supervision and guidance. I affirm that the dissertation has met all the necessary requirements in accordance with the regulations of this institute.

Sarbani Palit 11-6-2025

Dr. Sarbani Palit
CVPR Unit
Indian Statistical Institute
Kolkata- 700108
India

Declaration

I, **Ayan Mukherjee**, holding Roll. No. **CS2305**, hereby declare that the content presented in the dissertation **Morphology Based Galaxy Classification** represents the original work carried out by me for the degree of Master Of Technology in Computer Science at Indian Statistical Institute, Kolkata. I affirm that no sections of this report have been sourced or copied from external references without proper attribution. I am aware that any instance of plagiarism or the use of unacknowledged materials from third parties will be treated with the utmost seriousness and consequences.

Ayan Mukherjee
11/06/2025

Ayan Mukherjee
M.Tech. CS
Roll. No. CS2305
Indian Statistical Institute, Kolkata

Acknowledgments

I express my sincere appreciation to Dr. Sarbani Palit, my supervisor in the Computer Vision and Pattern Recognition Unit of Indian Statistical Institute, Kolkata, for her guidance, continuous support and inspiration. Her profound knowledge and creative suggestions have taught me a lot about every subject and have shown me how to conduct solid research.

I sincerely thank Ankita Sarkar, Senior Research Fellow at Indian Statistical Institute, for her invaluable assistance in gathering the data sets and landmark research articles essential for this research. Her consistent provision of ideas and unwavering support have been instrumental in the success of this project.

I am deeply grateful to all teachers in Indian Statistical Institute for their invaluable advice, insight and instruction, which provided a crucial perspective for my research.

Finally, I want to express my gratitude to my family for their unwavering support. I also extend my sincere appreciation to all my friends for their continuous assistance and encouragement. I am thankful to everyone who has contributed to my growth and success, even if I have inadvertently missed mentioning them on the above list.

Abstract

Galaxy evolution is an area of vital importance in current research as it is believed to hold vital clues of the past as well as the future of the universe. The structure or morphology of a galaxy acts as an indicator of its stage of evolution and may also shed light on the course of its future evolution. The deployment of high-resolution telescopes like James Webb Telescope has made available large amount of high-resolution images, thereby, facilitating the deployment of high-performance Machine Learning and Deep Learning techniques. In the proposed work, two different methods have been assessed to achieve better classification accuracy. The first work explores Zoom-Dilate Convolution and the second work explores equivariant convolution. In the first work, we have used cascaded simple Zoom-Dilate Convolution blocks to achieve most of the classification accuracy while transformer blocks are used to raise it further but not too much in order to keep the model reasonably lightweight. In the second work, we have explored equivariant convolution blocks to achieve most of our accuracy followed by a transformer block. The models have been trained on Galaxy10 SDSS datatset.

Keywords: Zoomed Convolution, Dilated Convolution, Transformer, Galaxy Morphology, Equivariant Convolution

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Dataset

1.1 Introduction

Galaxies are vital indicators for deciphering the extensive history of the universe, with their shapes, colors, and distinctive features providing critical clues to their evolutionary paths. Galaxy morphology is the study of categorizing galaxies based on their visual characteristics. This endeavor was introduced by Edwin Hubble, who developed the "tuning fork" diagram to classify the galaxies. In his scheme, galaxies were organized into twelve categories within four main types: elliptical, spiral, lenticular, and irregular.

However, with progression of modern research, it has become clear that Hubble's simple classification fails to encompass the full spread of galactic forms. Contemporary approaches to classification take into account a broader array of attributes, including a galaxy's roundness or elongation, surface smoothness, the presence of a central bar, and the configuration of its spiral arms (whether they are tightly wound or more dispersed). Moreover, factors such as the existence of a central bulge, evidence of interactions or mergers with neighboring galaxies, and the locations of star-forming regions contribute additional layers of complexity to galaxy classification.

In our work, we aim to develop a comprehensive approach to galaxy classification using RGB imaging data and neural networks to achieve a classification system. This work contributes not only to a deeper understanding of galaxy diversity but also offers a scalable solution for classifying large volumes of observational data from modern astronomical surveys using smaller image sizes hence achieving good accuracy with a relatively lightweight model and low computational resources. The most significant highlight is the development of a Zoomed and Dilated Convolutional Block which is used multiple times in a cascaded manner leading to an appreciable increase in performance. Furthermore, the effect of equivariant convolutions on our model performance has also been studied in this work. Cyclic and Dihedral equivariant convolution have been explored in-place of the standard convolution in the cascaded multi block setup. The incorporation of transformer blocks further enhances the performance.

1.2 Dataset

Galaxy Zoo [2] is a citizen science initiative that enlists volunteers to classify galaxies by morphology using images from astronomical surveys like the Sloan Digital Sky Survey (SDSS). Since its launch in 2007, the project has relied on public participation to identify galaxy shapes—such as spirals, ellipticals, and mergers—contributing valuable data to the field of astronomy. One of the datasets derived from this effort is the Galaxy10 SDSS dataset [3], released around 2019. It contains approximately 21,785 galaxy images categorized into ten broad morphological classes. Each image measures 69×69 pixels and is labeled based on the consensus of Galaxy Zoo volunteers, with a minimum agreement threshold of 55% to ensure label reliability.

Originally, the images were 424×424 pixels in size. They were centrally cropped to

207×207 and subsequently downscaled three times using bilinear interpolation to produce 69×69 pixel images, making them suitable for use with standard computer and GPU memory constraints. Although the dataset nominally includes 10 classes, one class contains only 17 samples and has therefore been excluded from this study. As a result, our analysis focuses on the remaining 9 morphological classes, defined as follows:

- Class 0: Disk, Face-on, No Spiral
- Class 1: Smooth, Completely round
- Class 2: Smooth, in-between round
- Class 3: Smooth, Cigar shaped
- Class 4: Disk, Edge-on, Rounded Bulge
- Class 5: Disk, Edge-on, No Bulge
- Class 6: Disk, Face-on, Tight Spiral
- Class 7: Disk, Face-on, Medium Spiral
- Class 8: Disk, Face-on, Loose Spiral

A sample from each class is shown in Fig: 1.2. As the data is already processed no further pre-processing has been applied on the images. All the models were trained on the original images.

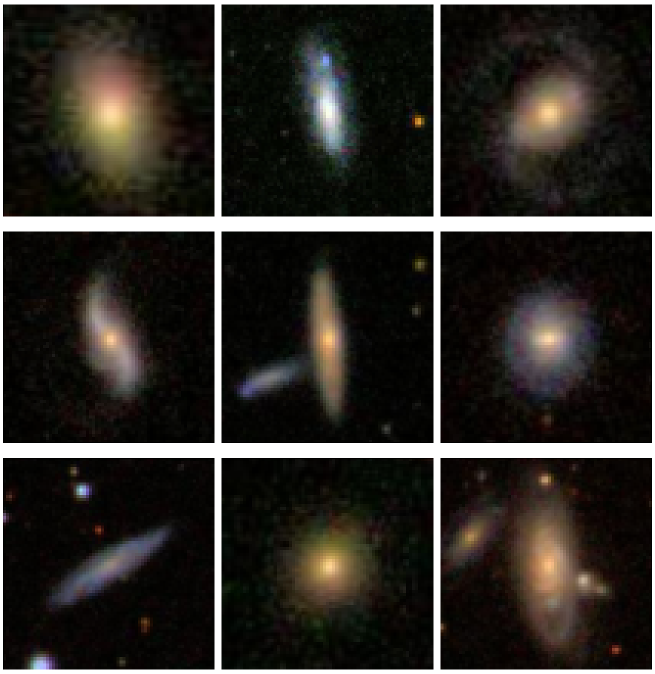

Figure 1.1 – Sample Images

Chapter 2

Related Works

Astroformer [4] by Dagli R combines convolutional and transformer layers to leverage the strengths of both, enhancing generalization and performance even with limited data. The authors draw inspiration from the CoAtNet and MaxViT architectures and propose a new stack design along with a novel relative self-attention layer. The model achieves stateof-the-art results on the Galaxy10 DECals (image size 256×256) dataset, witha top-1 accuracy of 94.86%. But all this at the cost of very heavy CoAtNet-4 [1] and very high computation resources four NVIDIA Tesla V100 GPUs or a TPUv3-8 cluster with a batch size of 256.

Cao J. et al. [5] enhanced the Vision Transformer (ViT) [6] architecture by integrating convolution into two key components. First, they introduced a hierarchical transformer structure with convolutional token embeddings, mimicking the down-sampling approach of CNNs. This design increases token width while reducing the total number of tokens. Second, they replaced the standard position-wise linear projections used for queries, keys, and values in ViT with a convolutional transformer block that employs depth-wise separable convolutions, referred to as convolutional projections. Their proposed Convolutional

Vision Transformer (CvT) model is applied to classify galaxies into five distinct categories, using data from the Galaxy Zoo 2 (GZ2) project for low-redshift galaxies and the Galaxy Zoo CANDELS dataset for high-redshift galaxies.

Barchi P.H. et al. [7] conducted a study comparing machine learning(ML) and deep learning approaches. For ML, they applied decision trees, support vector machines, and a multilayer perceptron with three hidden layers using ReLU activation. In the deep learning category, they utilized the GoogleNet Inception model. Their study examined how various datasets influenced training outcomes by altering the number and size of objects in sample images and introducing a parameter K, defined as the ratio of the area of galaxy's Petrosian ellipse to the area of the full width at half maximum. In an 11-class classification task, with labels sourced from Galaxy Zoo 2 and data from SDSS DR7, they achieved a maximum overall accuracy of 65.2% using a CNN when K>20, utilizing approximately 30% of the dataset.

Zoomed Convolution [8] was introduced by Jeong et al. in 2023. It enhances traditional convolutional methods to capture multi-scale features in the images in a better manner. The receptive field changes based on whether the image is zoomed in (upsampled) or zoomed out (down-sampled) before applying convolution. Up sampling increases the input size, allowing the convolution to focus on finer details within a smaller region. The receptive field covers a smaller portion of the original image's overall structure, emphasizing high-resolution features. On the contrary, down sampling reduces the input size, effectively increasing the receptive field to cover a greater area in the original image. This helps the model capture more global context with each convolution.

Dilated convolution [9] was introduced by Yu F. et al. in 2016. Its strength lies in the fact that it captures large contextual information without losing spatial resolution. In this type of convolution filter, gaps (or dilation) are introduced in between its weights. This enables the kernel to capture a wider region of the input without adding more parameters or increasing computational cost.

Equivariant CNNs [10] are a class of neural network architecture deviced to preserve symmetries in data. These symmetries can include translations, rotations, reflections, and more. In simpler terms, an equivariant CNN ensures that if the input is transformed (e.g., rotated), the output will transform in a predictable and consistent way. In our case, we want it to predict the same label for the original and, rotated and/or translated image.

Weiler M. et al. [11] introduced generalized convolutional networks that are equivariant under broad symmetry groups. These networks exhibit translation equivariance, meaning their operations commute with translations of the input, ensuring independence from specific spatial locations. Beyond translations, these models also generalize to other geometric transformations such as rotations and reflections. Compared to standard convolutional networks, equivariant convolutional networks have shown notable improvements in data efficiency, faster convergence, and overall performance.

Sneh Pandya et al. [12] proposed using group CNNs (GCNNs) that are equivariant to the 2D Euclidean group, *E(2)*, for morphology based galaxy classification. The core idea is to incorporate the inherent symmetries found in galaxy images to our advantage by using it as an inductive bias in the model architecture.

A. Sarkar et al.[13] showed that Tidal Features targeted the tightness and prominence of the arms of the galaxy and gave better performance on different state-of-the-art models. This feature was obtained by first converting the image to grayscale followed by a Gaussian blur with a kernel of size 23. The kernel size was determined experimentally. This blurred image when subtracted from original image gives us the required tidal features.

Chapter 3

Methodology

3.1 Zoomed-Dilated Convolution and Galaformer

3.1.1 Zoomed-Dilated Convolution

By combining the advantages of both techniques, we developed Zoomed-Dilated Convolution (ZDconv). Given the small size of the images, zoomed convolution offers a zoom-in view to capture fine local details and a zoom-out perspective for broader global patterns. Meanwhile, dilated convolution further enlarges the receptive field without compromising resolution. In this layer, we perform three main operations:

1. perform Bilinear interpolation of the image by a factor of "Zoom Factor".

2. perform 3×3 dilation convolution with the chosen "Dilation Factor".

3. perform Bilinear interpolation to resize the image back to its original size.

3.1.2 Galaformer

The model Galafomer draws inspiration from CoAtNet [1] which combines convolutional (C) and transformer (T) blocks. The convolution block of CoAtNet is integrated using MBConv (Mobile Inverted Bottleneck Convolution) layer. MBConv block [14] typically has three convolutional layers per block when expansion is greater than one: the first point-wise (1x1), the depthwise (3x3), and the final point-wise (1x1) linear projection. When expansion=1, it skips the initial pointwise convolution, resulting in two convolutions (depth wise and final point-wise). The transformer used is similar to the one in Vit [6]. This network was originally designed for the ImageNet dataset where the image size was 224×224. Several versions of CoAtnet were proposed. Among them, we have chosen C-C-T-T as this worked best for our model. Again in terms of size, there are 5 variations S0 to S4. S0 is the lightest and S4 is the heaviest. We have choosen S0 to keep our model lightweight.

In the Zoomed convolution block, the input is billinear interpolated to a size of original image size * zoom factor. Then, Conv2D is applied on this interpolated image. The output of Conv2D is billinear interpolated to original size if downsample is False and to half of original size (downsampled) if downsample is True.

In the Dilated convolution block, we have a Conv2D block with the choosen dilation factor. The output of Conv2D is billinear interpolated to half of original size (downsampled) if downsample is True.

Galaformer differs from CoAtNet in the convolutional block. We have used ZDconv instead of MBConv. ZDconv only has one convolution operation per layer as opposed to three convolutions in MBConv. When downsample is false, the output of the block is returned. When downsample is true, the input is passed throught a pooling layer followed by a projection (Conv2D with kernel_size = 1) and the summation of output (downsampled) and pooling-projection operation is returned. The Model has 4 main blocks ZDconv Block 1 which uses 3 ZDconv convolution with max-pooling, ZDconv Block 2 with 2 ZDconv convolution without max-pooling, ZDconv Block 3 comprising 1 ZDconv convolution with max pooling and finally a transformer block. The proposed architecture of the model is shown in 3.1

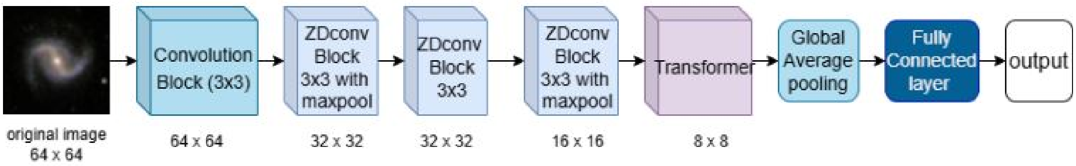

Figure 3.1 – Block Diagram of Galaformer

3.2 Equivariance in CoAtNet [1] and Galaformer

A model that is not equivariant needs to be trained on images that are transformed in different ways in order to make it equivariant. This leads to longer training time and the performance is also not satisfactory. So, a better approach can be incorporating equivariance in the model itself by construction. In this way, the model's complexity will be decreased, thereby, reducing the training time and improving performance.

In this work, an attempt has been made to study the effects of Group Convolutional Neural Networks, specifically, cyclic and dihedral groups of different orders on model performance. We have tested both CoAtNet [1] and Galaformer by converting the convolution present in these models with Equivariant convolution. We have also tested its effects on Zoomed-Dilated Convolution (ref. Sec 3.1).

The types of equivariances tested were cyclic (Ceq) and dihedral (Deq). The number of axis tested were 4, 8, 16 and 32. We represent each block as Ceq_n or Deq_n, where n is the number of axis.

A new block was prepared by adding some of the features of Ceq_32 into ZDConv (from Galaformer). This block is similar to ZDConv. The difference is that the Convolution used in ZDConv has been replaced with the one used in Ceq_32. The projection and pooling operation used when downsample is true is same as the one in ZDConv block. This block can be called ZDCEQ block. Continuing in a similar vein, ZDCEQ2 was also prepared by combining Ceq_32 and ZDConv2 in a similar fashion as in ZDCEQ.

Chapter 4

Experiments and Results

4.1 Galaformer

The dataset was divided into 80% for training and 20% for testing. The Adam optimizer was employed with a learning rate of 0.001.

4.1.1 Ablation study for Zoomed, Dilated and Zoomed-Dilated Convolution

To evaluate all three techniques and determine the corresponding hyperparameters, a simple model was used comprising a 2D convolution layer (implementing the technique under test) with ReLU activation, followed by a fully connected layer. For the Zoomed convolution, the best accuracy with the simple model was 73.34% for a zoom factor of 0.5. For the Dilated convolution, the best accuracy obtained was 71.15% for a dilation factor of 2.

Next, the 2D covolution block is updated with Zoomed-Dilated convolution and both the Zoom factor and Dilation factor are tuned. We get the best accuracy of 73% for Zoom factor of 0.5 and Dilation factor of 2.

4.1.2 Experiments on CoAtNet[1] and Galaformer

The dataset split, optimizer, and learning rate remained consistent with the previous test. Additionally, 10-fold cross-validation was performed on the 80% training portion. For each fold, the model was trained for 200 epochs using a batch size of 1000.

The coatnet_0 model was selected out of the 4 variations mentioned in [1] because it was the lightest of all.

Model	Dataset	Test Accuracy
CoAtNet_0	Galaxy10 SDSS	79.97%
CoAtNet_0	Galaxy10 SDSS Zoomed-In	76.23%

Table 4.1 – CoAtNet_0 Results

The MBConv blocks of coatnet_0 were changed to Zoomed convolution block. Then, the MBConv block was replaced by Dilated convolution block.

Model	Dataset	Test Accuracy
CoAtNet_0 with Zoomed convolution	Galaxy10 SDSS	80.04%
CoAtNet_0 with Dilated convolution	Galaxy10 SDSS	80.22%

Table 4.2 – CoAtNet_0 with Zoomed and Dilated convolution blocks Results

Next, the MBConv block was replaced by Zoomed-Dilated Convolution (ZDConv) block. We also modified the transformer (T) by removing max pooling from it. We also tested the model after removing max pooling from the ZDConv blocks (referred to as ZDConv2 hereafter). Next, we tested the proposed model, Galaformer with ZDConv-ZDConv2-ZDConv-T as the blocks.

Model	Test Accuracy
CoAtNet_0 with ZDConv	79.01%
CoAtNet_0 with ZDConv after removing max-pooling from T	79.72%
CoAtNet_0 with ZDConv after removing max-pooling	81.3%
Galaformer	**82.27%**

Table 4.3 – Performance comparison of different models

4.2 Equivariant CNN

The train-validation-test split, optimizer, learning rate, no. of cross-validation folds, no. of epochs were same as in the previous test. The batch size was set to 500 because of memory constraints.

4.2.1 Incorporating equivariance in CoAtNet[1] and Galaformer

The Galaformer model was choosen as the starting point. Different types of equivariances and number of axis (order) were tested by changing the third block implementing Equivariant Convolution with different combinations.

The type of equivariance blocks tested were Ceq_4, Ceq_8, Ceq_16, Ceq_32, Deq_4, Deq_8, Deq_16 and Deq_32. The best performing equivariant convolutions was selected and further tested, this time, with convolution only models. We selected VGG16 [15] for this purpose. The equivariant convolution Deq_32 was also tested as it had strong theoretical basis of putting up a good performance. It covered flipping as well as rotation along the axis.

4.2.2 Experiments on CoAtNet [1] and Galaformer with equivariant convolution

After these tests, Ceq_32 was finalized for further modifications and incorporations.

Galaformer block configuration	Test Accuracy
ZDConv-ZDConv2-Ceq_4-T	80.71%
ZDConv-ZDConv2-Ceq_8-T	80.52%
ZDConv-ZDConv2-Ceq_16-T	80.48%
ZDConv-ZDConv2-Ceq_32-T	81.63%

Table 4.4 – Performance comparison of Cyclic Equivariant Convolution Blocks

Galaformer block configuration	Test Accuracy
ZDConv-ZDConv2-Deq_4-T	81.24%
ZDConv-ZDConv2-Deq_8-T	80.85%
ZDConv-ZDConv2-Deq_16-T	80.43%
ZDConv-ZDConv2-Deq_32-T	80.66%

Table 4.5 – Performance comparison of Dihedral Equivariant Convolution Blocks

Model	Test Accuracy
VGG16	81.49%
VGG16 with convolutions replaced with Ceq_32	83.26%
VGG16 with convolutions replaced with Deq_32	82.66%

Table 4.6 – Performance comparison of VGG16 with Ceq_32 and Deq_32 Equivariant Convolution Blocks

CoAtNet[1] was modified by replacing MBConv with Ceq_32 blocks and tested. Also, Ceq_32 was changed to Dilated Ceq_32 in the model and tested.

Model	Dataset	Test Accuracy
CoAtNet_0 with Ceq_32	Galaxy10 SDSS	81.56%
CoAtNet_0 with Dilated Ceq_32	Galaxy10 SDSS	81.99%

Table 4.7 – CoAtNet_0 with Ceq_32 and Dilated Ceq_32 blocks Results

A common trend noted in the confusion matrix was that the model's failure to significantly differentiate between the spiral classes (class 6, 7 and 8). The tidal feature [13] was used along with the RGB channels for the further tests that were performed.

We tested ZDCEQ and ZDCEQ2 blocks by replacing the ZDConv and ZDConv2 block

Model	Dataset	Test Accuracy
Galaformer	Galaxy10 SDSS classes 6, 7 and 8	66.47%
Galaformer	Galaxy10 SDSS classes 6, 7 and 8 with tidal features	73.92%

Table 4.8 – Spiral Classes Results

of Galaformer with them. These models were trained and tested on Galaxy10 SDSS with tidal features.

Galaformer block Configuration	Test Accuracy
Galaformer	82.18%
ZDCEQ-ZDCEQ-ZDCEQ-T	82.77%
ZDCEQ-ZDCEQ-ZDCEQ-T with ReLU in ZDCEQ	82.87%
ZDCEQ-ZDCEQ2-ZDCEQ-T	82.93%

Table 4.9 – Galformer with ZDCEQ and ZDCEQ2 Results

Chapter 5

Conclusion and Future Work

The proposed GalaFormer in Section 3.1.2: is compact and lightweight, yet observed to perform robustly on the dataset Galaxy10 SDSS comprising of small image sizes. The most significant contribution of this work is the design of the Zoom-Dilation Convolution block, which is at the heart of the model. Along with a transformer block, it can successfully achieve good accuracy with low computational cost. The design of the proposed model leverages the availability of a relatively small dataset with images of a size smaller than usual. Further, tailored to work with a batch size of 250, it consumes only 5.5 GB of memory. While the model's maximum peak accuracy (across the folds) is only marginally lower, yet it demonstrates a promising efficiency-to-performance ratio. Future improvements will aim to enhance accuracy through structural modifications but without compromising model size.

Different equivariant models were tested, the results of which are shown in Section 4.2. The use of Equivariant CNNs has improved the model accuracy compared to the one without Equivariant CNNs. A common observation was that the models were not able to distinguish between the different classes of spiral galaxies. (classification based on the tightness of its arms). A sample Confusion matrix has been shown Fig. 5.1. The use of Zoomed-Dilation along with Equivariance improves the model performance further.

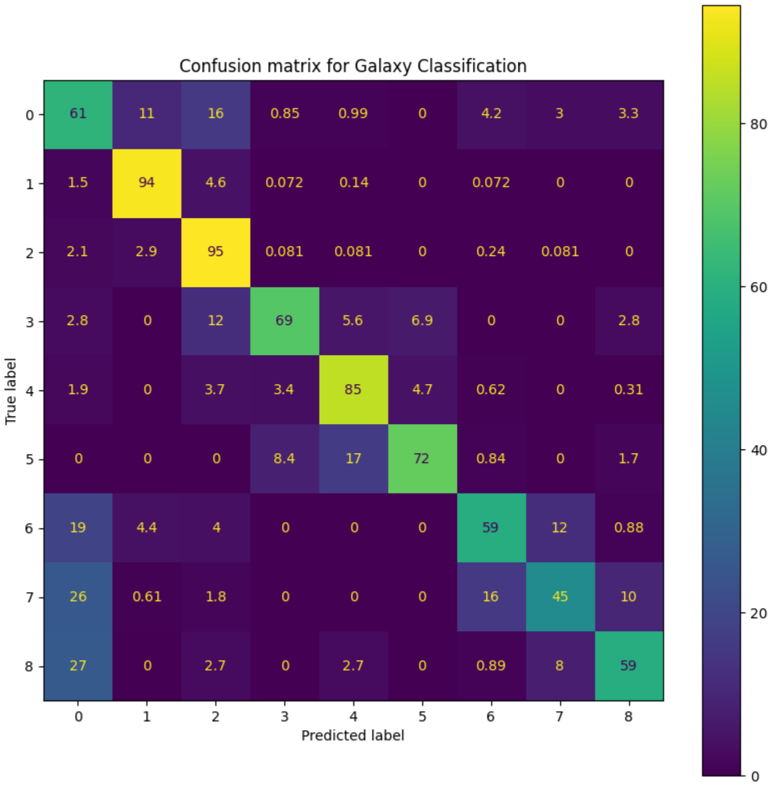

Figure 5.1 – Confusion Matrix of ZDCEQ-ZDCEQ2-ZDCEQ-T

Future work could explore additional methods to enhance dataset quality and investigate the potential of using other astronomical datasets. Such efforts could further improve the classification accuracy and generalizability of the proposed approach, broadening its applicability in the analysis of galaxy morphology.

List of References

[1] Z. Dai, H. Liu, Q. V. Le, and M. Tan, "Coatnet: Marrying convolution and attention for all data sizes," 2021. [Online]. Available: https://arxiv.org/abs/2106.04803

[2] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu, P. Murray, and J. Vandenberg, "Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey," *Monthly Notices of the Royal Astronomical Society*, vol. 389, no. 3, p. 1179–1189, Sep. 2008. [Online]. Available: http://dx.doi.org/10.1111/j.1365-2966.2008.13689.x

[3] H. Leung and J. Bovy, "Galaxy10 sdss dataset," https://astronn.readthedocs.io/en/latest/galaxy10sdss.html, 2017.

[4] R. Dagli, "Astroformer: More data might not be all you need for classification," 2023. [Online]. Available: https://arxiv.org/abs/2304.05350

[5] J. Cao, T. Xu, Y. Deng, L. Deng, M. Yang, Z. Liu, and W. Zhou, "Galaxy morphology classification based on convolutional vision transformer (cvt)," *Astronomy amp; Astrophysics*, vol. 683, p. A42, Mar. 2024. [Online]. Available: http://dx.doi.org/10.1051/0004-6361/202348544

[6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021. [Online]. Available: https://arxiv.org/abs/2010.11929

[7] P. Barchi, R. de Carvalho, R. Rosa, R. Sautter, M. Soares-Santos, B. Marques, E. Clua, T. Gonçalves, C. de Sá-Freitas, and T. Moura, "Machine and deep learning applied to galaxy morphology - a comparative study," *Astronomy and Computing*, vol. 30, p. 100334, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213133719300757

[8] C. Y. Jeong, K. Moon, and M. Kim, "An end-to-end deep learning approach for real-time single image dehazing," *Journal of Real-Time Image Processing*, vol. 20, no. 1, Feb. 2023. [Online]. Available: http://dx.doi.org/10.1007/s11554-023-01270-2

[9] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015. [Online]. Available: https://arxiv.org/abs/1511.07122

[10] T. S. Cohen and M. Welling, "Group equivariant convolutional networks," 2016. [Online]. Available: https://arxiv.org/abs/1602.07576

[11] M. Weiler, P. Forré, E. Verlinde, and M. Welling, *Equivariant and Coordinate Independent Convolutional Networks*, 2023. [Online]. Available: https://maurice-weiler.gitlab.io/cnn_book/EquivariantAndCoordinateIndependentCNNs.pdf

[12] S. Pandya, P. Patel, F. O, and J. Blazek, "E(2) equivariant neural networks for robust galaxy morphology classification," 2023. [Online]. Available: https://arxiv.org/abs/2311.01500

[13] A. Sarkar, S. Palit, and U. Bhattacharya, "Demystifying galaxy classification: An elegant and powerful hybrid approach," in *2024 39th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2024, pp. 1–6.

[14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," 2018. [Online]. Available: https://arxiv.org/abs/1801.04381

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1409.1556