Journal of Physics:
Conference Series

PAPER • **OPEN ACCESS**

Galaxy Morphology Classification with DenseNet

To cite this article: Wuyu Hui *et al* 2022 *J. Phys.: Conf. Ser.* **2402** 012009

Galaxy Morphology Classification with DenseNet

Wuyu Hui[1, a†], Zheng Robert Jia[2, b†], Hansheng Li[3, c*, †], Zijian Wang[4, d†]

[1]School of Electronic and Automation, Anhui University, Hefei, China

[2]College of Liberal Arts, Computer Science, University of Minnesota, Minnesota, U.S.A

[3]Division of Science and Technology, Data Science, BNU-HKBU United International College, Zhuhai, China

[4]School of Mathematical Sciences, Capital Normal University, Beijing, China

[a]Z01914072@stu.ahu.edu.cn

[b]jia00129@umn.edu

[c*]Corresponding author: p930026065@mail.uic.edu.cn

[d]1190504018@cnu.edu.cn

[†]These authors contributed equally.

Abstract – Galaxy classification is crucial in astronomy, as galaxy types reveal information on how the galaxy was formed and evolved. While manually conducting the classification task requires extensive background knowledge and is time-consuming, deep learning algorithms provide a time-efficient and expedient way of accomplishing this task. Hence, this paper utilizes transfer learning from pre-trained CNN models and compares their performances on the Galaxy10 DECals Dataset. This paper applies opening operation, data augmentation, class weights, and learning rate decay to further improve the models' performance. In our experiments, DenseNet121 outperforms the other models and achieved approximately 89% test-set accuracy within 30 minutes. The second best-performing model, EfficientNetV2S, takes double the time achieving 2.43% lower test set accuracy.

1. INTRODUCTION

Attempts to classify galaxies have been mainly based on three different approaches. The earliest method is to rely on manual classification, which results in inconsistent standards and inefficiency. In order to classify a large number of images in a short period of time, machine learning algorithms based on the analysis of physical parameters of the galaxies were introduced [1]. Nevertheless, this technique requires complicated feature engineering. Hence, our work addresses these problems by utilizing neural networks, which classify the galaxies efficiently and do not require extensive background knowledge.

This paper makes predictions based on the Galaxy10 DECals Dataset, separating galaxies into ten categories based on their morphology. This paper approaches the problem by applying different CNN models and comparing their performances. Many related works directly regard validation accuracy as model accuracy. Nevertheless, this approach is biased since it is adjusting its hyperparameters based on

the validation dataset. To resolve this problem, our experiments compare model performances based on their test accuracy. In this experiment, our work investigates 11 light-weighted models: the ResNet50, the MobileNetV2, the MobileNetV3L, the EfficientNetV2S, the EfficientNetB0, the EfficientNetB1, the EfficientNetB2, the EfficientNetB3, the EfficientNetB4, the EfficientNetB5, the DenseNet121 and show that the DenseNet121 outperforms other models. Transfer learning proves effective in feature extraction and significantly reduces training time. This paper utilizes pertained models with ImageNet weights and compares their performances.

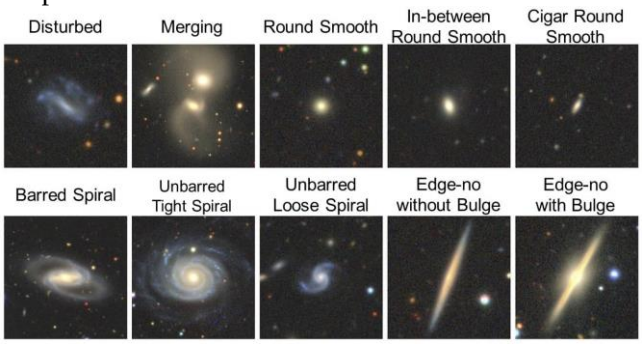

Figure 1. Example images from each of the 10 classes in Galaxy10 DECals dataset.

2. RELATED WORK

2.1 Galaxy Classification

There are three commonly taken approaches for the galaxy classification task.

The first is to use visual classifications. Previous classification tasks were conducted by a small group of individual astronomy experts. Since manually classifying galaxies is time-consuming, their attempts have been limited to a small subset of the entire SDSS dataset.

The second is to use machine learning with physical properties such as color, brightness, and concentration index. In 2003, Abraham et al. came up with models that predict galaxies based on those physical properties [2]. However, since each galaxy is distinct from others, it is difficult to quantify these measurements. Introducing these physical features might lead to potential biases or issues. Due to its limitation, this method has been deemed outdated and is less commonly used nowadays.

Lastly, neural networks provide an automatic process to classify these galaxies and are able to learn complex features from the images. Compared to the aforementioned methods, neural networks could classify an enormous number of images within a short period of time and do not require complex feature engineering.

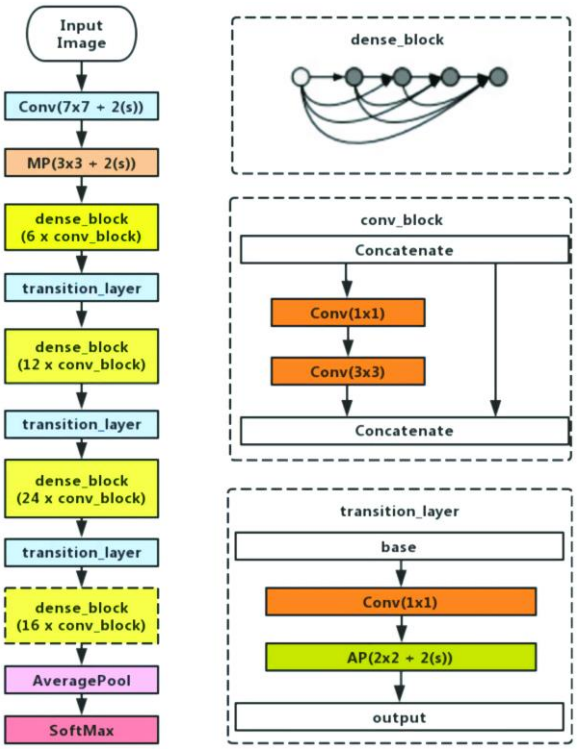

Figure 2. The Architecture of DenseNet121.

2.2 Galaxy10 Dataset

In 1926, Edwin Hubble proposed the first galaxy classification diagram, "The Hubble Sequence," which classifies the galaxies into three broad classes based on the shape observed from the earth. Later, Gerard' de Vaucouleurs and William Wilson Morgan proposed new systems: the De Vaucouleurs system and the MK system, respectively [3][4].The Galaxy10 DECals Dataset this paper uses is release d by Galaxy Zoo, which refers to the Hubble Sequence for galaxy classification.

The Galaxy Zoo project was launched on 11th July 2007 to visually classify nearly one million galaxies from the SDSS [5]. Galaxy10 Dataset labels come from Galaxy Zoo, following the "Hubble Sequence" classification method. The goal is to predict which galaxy class the input galaxy belongs to. Both datasets are labeled by "volunteer citizens," who label the tasks based on given instructions. Galaxy Zoo then records the percent agreement amongst users for each image. Images with more than 95% agreement are marked "superclean," and those with 80% are marked as "clean" [6].

The original Galaxy10 SDSS (Sloan Digital Sky Survey) dataset, with labels form Galaxy Zoo and images from SDSS, contains 21785 69x69 pixels colored galaxy images separated into ten classes. The resolution of the original dataset is low, with a significant imbalance between classes. For example, class 5 (Disk, Edge-on, Boxy Bulge) includes merely 17 images.

The new dataset comes from DECals (Dark Energy Camera Legacy Survey) and consists of 17736 256x256 pixels colored galaxy images, which are selected in 10 broad classes using volunteer votes with more rigorous filtering, for example, weighting users according to how often they agreed with the majority and adding bias corrections [6]. The new version introduced new classes and adjusted all classes to be more balanced, abandoning the previous class that consisted of only 17 images.

In comparison with the previous dataset, the new version provided fewer images yet with higher resolution. Low-resolution images limit the amount of useful information the model could extract and prevent the model from achieving the best performance. Research has shown that upscaling from lower resolution leads to higher loss and higher resolution images have predominant advantages over lower resolution ones [7]. Especially for our task, in which the pivotal information concentrated in the center of

the image, higher resolution provides more useful information during the classification task. Since our work uses pertained architectures, we resize/crop images into the shape (224,224,3). In this case, cropping the outer part of the new dataset image from (256,256,3) to (224,224,3) does not lose much useful information, since the targeted galaxy is located in the center of the image, while upscaling the previous dataset from (69,69,3) to (224,224,3) can harm model performance.

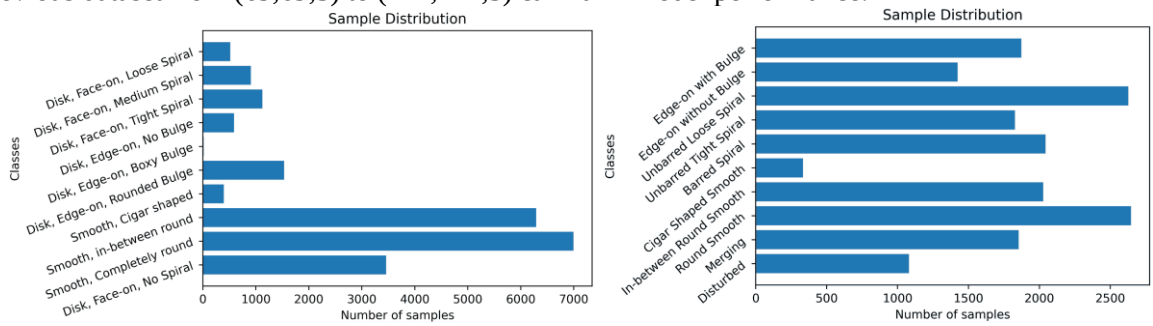

Figure 3. Comparison between the two sample distributions. The sample distribution for the old dataset, Galaxy10 SDSS, is shown on the left. The sample distribution for the new dataset, Galaxy10 DECals, is shown on the right.

2.3 DenseNet

A vanishing gradient is a common problem for large models. When more layers are added to the model, it is able to learn more complex features, and yet the gradients were difficult to be propagated back to the initial few layers. The DenseNet resolves this issue using similar theories as the ResNet.

There are small differences among these two similar structures, the ResNet adding previous layers up while the DenseNet concatenates them. According to Figure 2, the DenseNet proposes two changes to the standard CNN model: a Dense block that allows gradients to flow through and a transition layer that improves model efficiency by reducing parameters. Within each Dense Block, layers and layers have dense connectivity and concatenate the information. The concatenation step could be understood as the model reusing learned features. After each dense block, the concatenation process makes the model architecture too wide. Consequently, transition layers are introduced to reduce dimensions and extract important features. In the experiment applying the DenseNets to CIFAR, SVHN, and ImageNet conducted by Huang et al., the DenseNets resulted in consistent improvement in accuracy with the growing number of parameters, without any signs of performance degradation or overfitting [8].

2.4 Transfer Learning

Transfer learning utilizes knowledge gained from a previous model and applies it to a new problem [9]. The initial few CNN layers from the previous model are capable of extracting low-level features. Thus, the target model copies these initial layers and then randomly initializes the rest of the layers. The remaining layers of the target network are then randomly initialized and trained toward the target task. In this way which takes advantage of the knowledge learned from the pre-trained data, the early and intermediate layers are frozen, and we only retrain the latter layers. This technique is effective when the features extracted are general. In other words, the features learned from the previous model are applicable to the target model.

One can choose to either back-propagate the weights to the previous layers, which is called fine-tuning, or freeze the previous layers. Through the two options, transfer learning can train on a dataset much smaller than the previous dataset without overfitting the small dataset. When training a complicated model on a comparatively small dataset, transfer learning saves much training time and prevents overfitting. Also, we can take advantage of model architecture developed by the deep learning research community, such as GoogLeNet and ResNet. Despite improvements in the model during the training phase, transferring layers from a pre-trained model enhances the generalization ability of the target model.

3. METHODOLOGY

3.1 Morphological Opening

Previous research has shown that morphological openings are increasing, anti-extensive, translation invariant, and idempotent [10]. In this task, the morphological opening operation helps remove unrelated small objects that are not at the center but might distract the model's focus and can emphasize the targeted galaxy. Figure 4 clearly shows how the opening operation helps eliminate unrelated noises in the input data.

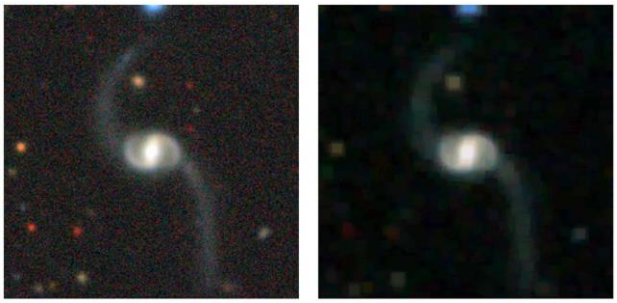

Figure 4. Comparing the image before opening on the left and image after opening on the right, the opening operation has removed unrelated noise and emphasized the target galaxy.

3.2 Data Augmentation

Data augmentation has been proven to be effective in preventing models from overfitting, thus increasing models' generalization ability, since it adds variability and flexibility to the input data. We conduct this data augmentation step on GPU to avoid CPU bottlenecks and GPU data starvation. Figure 5 shows the data augmentation procedure we go through during the training stage. Figure 6 shows an example of how images flow through the data augmentation process.

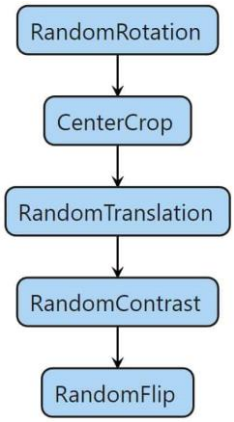

Figure 5. The Procedure for Data Augmentation.

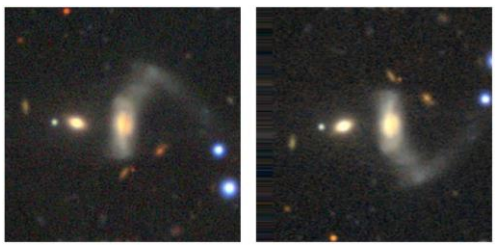

Figure 6. The figure on the left shows the input image before augmentation process. The figure on the right shows the input image after flowing through the augmentation process.

3.3 Model

1) Preprocessing: We normalized our data using $x = x/127.5 - 1$ (where x is the value for each pixel) to convert the pixel values into the range [-1,1]. This technique proves to increase convergence speed, improve model accuracy, and prevent exploding gradients.

2) Class Weights: Class weights can be used in the following two cases. The first one is a highly unbalanced sample, when a certain category is a small probability event, such as identifying whether the user violates the rules. The second case is when there is high importance of specific categories (usually accompanied by the first one), that is, the case requiring classification at a high significant level, such as cancer, credit card fraud transactions, and large online consumption in which the cost of misclassification is high.

As displayed in Figure 3, the dataset has a high imbalance between classes. Since the classes with fewer images are trained less, we added higher class weights to the classes with fewer samples to prevent the problem. This can force the loss function to pay more attention to the data with insufficient samples when dealing with unbalanced training data [11], which can improve the robustness of classification and accelerate convergence speed.

3) Backbone Model: We use DenseNet121 as our backbone model and compare it with the models mentioned in the abstract. Since transfer learning has proven to yield good results in feature extraction, our model then used the pre-trained weights on the ImageNet with a GAP(Global Average Pooling) layer and a final Dense layer. Because our input images are quite different from the ImageNet images, we also allow the weights in the backbone layers to be trainable.

4. EXPERIMENTS

4.1 Dataset

Figure 1 displays images from each class in the new Galaxy10 Dataset. The example images presented appear distinct, but throughout our experiment, we noticed that a lot of the images are confusing and even difficult for humans to identify. Manually classifying the galaxies is a complicated process that involves extensive consideration of different aspects, including but not limited to the roundness, number, and tightness of spiral arms, prominence of the central bulge, and bar feature of the galaxy. This reflects the advantage of using Neural Networks for the classification task, enabling us to construct the model with mere knowledge of astronomy and achieve decent model performance. The prediction process is also efficient and requires no manual involvement.

It is worth noting that some related work compares model performances based on validation accuracy, but this renders the problem that the model might overfit the validation set. We divided the data into train, validation, and test sets in the ratio of 70:15:15. The separate test set allows us to better compare the models.

4.2 Model Performance

Learning rate has always been one of the most important hyper-parameters and is crucial to model convergence and performance. We chose to manually decrease the learning rate after every ten epochs to accelerate convergence, as shown in Figure 7. The learning rate schedule follows (1), where the decay rate is 0.1, and the decay steps are 10.

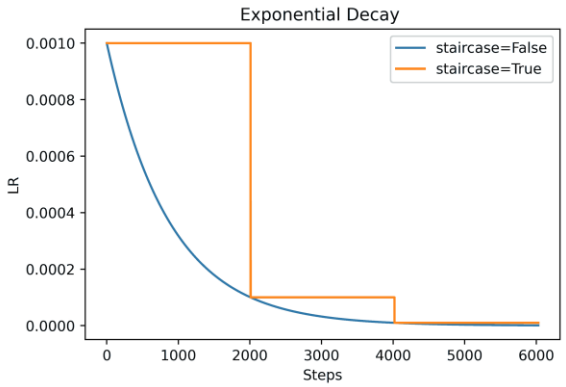

Figure 7. Learning rate schedule with exponential decay.

$$lr = initial\_lr * (decay\_rate^{(step/decay\_steps)}) \quad (1)$$

The technique turns out to be very effective, as presented in Figure 12. For our experiments, we trained the model for 30 epochs with a batch size of 64 using Adam optimizer and standard categorical cross-entropy as loss.

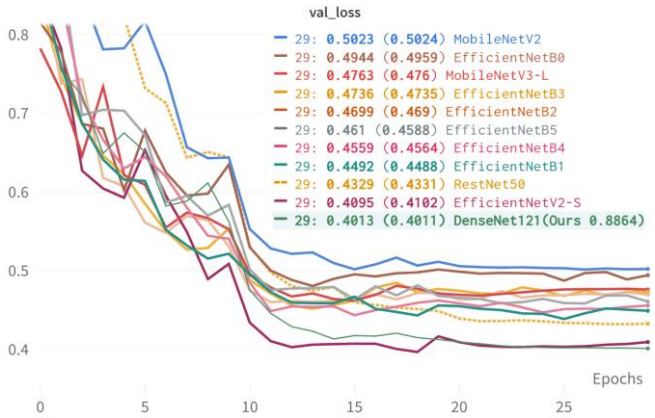

Figure 8. Validation loss comparison for different models.

Eventually, all models converge after 30 epochs of training. In Figure 8, there is a significant drop in validation loss. This is due to the learning rate schedule, which leads the model to converge to a better position. The validation losses for the EfficientNetV2S and the DenseNet121 are significantly lower than the other models.

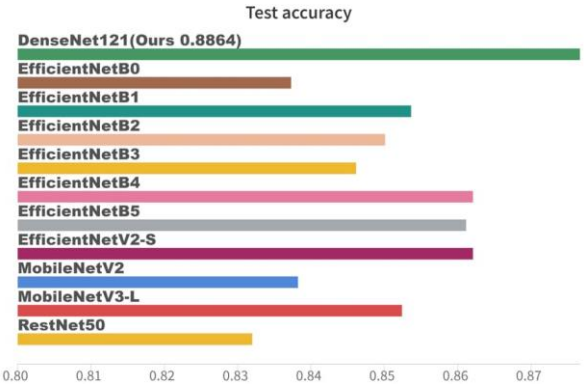

Figure 9. Test accuracy comparison for different models.

Figure 9 presents the test accuracy for each model. We evaluate the models' performance based on their test accuracy. We repeat the experiments several times and compare the average test accuracy. Though the validation loss for EfficientNetV2S is lower than DenseNet121, the latter achieved an accuracy of 88.64%, 2.43 percent higher than the second-best performing the EfficientNetV2S model, indicating that the DenseNet121 is more robust than other models.

4.3 Case Study

Presented in Figure 10, it is obvious that the model tends to confuse classes 0, 6, and 7. Furthermore, in Figure 11, we displayed two error cases where the model mislabels between class 6 and class 7.

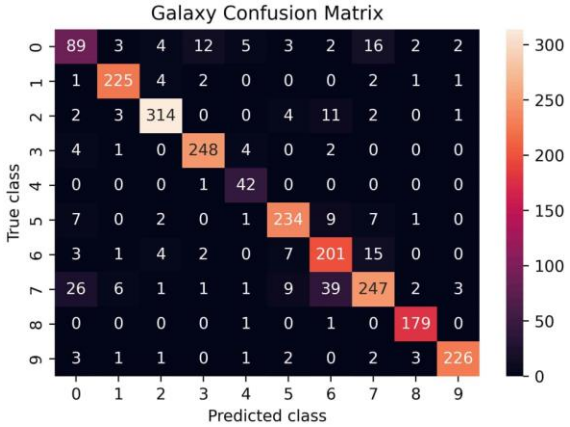

Figure 10. Confusion Matrix for DenseNet121.

Since we are using a deep end-to-end model, it is difficult to interpret the model merely referring to its predicted labels. It is difficult for humans to explain which feature they rely on and how to make predictions. Grad-CAM addresses this problem by revealing which features positively influence the class of interest and is applicable to a wide range of CNN models [12]. The saliency map addresses the problem by calculating which pixel is the model most sensitive to while making predictions. However, the saliency map is often visually noisy and confusing. Smooth saliency map, which creates samples of the target image with noises and applies saliency maps on the average of these samples, is applied to increase the quality of the sensitivity map [13].

We then investigate the error cases to find out where our model performs wrong. In Figure 11, class 6 and class 7 represent Unbarred Tight Spiral Galaxies and Unbarred Loose Spiral Galaxies, respectively. The main difference between the two unbarred spiral galaxies is the tightness of their spiral arm. As shown in Figure 11, the GradCAM on the right shows that though the model mislabels the galaxy, it focuses on the correct part for classifying the image instead of the noises around. The smooth saliency map in the middle shows that the model directs attention to the spiral arms and attempts to distinguish their tightness. However, the different observation angles of the galaxies in the original images make it difficult for the models to visually recognize the tightness of spiral arms. The hazy band of light in the image on the second row hinders the model from extracting the spiral arms. Even humans suffer from classifying these images and require a significant amount of effort.

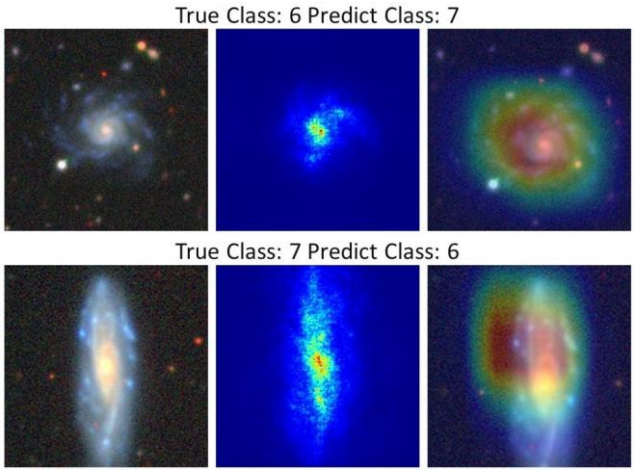

Figure 11. Examples of error cases: Class 6 corresponds to "Unbarred Tight Spiral Galaxies" and Class 7 corresponds to "Unbarred Loose Spiral Galaxies". The first column displays the original input image. The second and third column display smooth saliency map and the Grad-CAM, which explains the pixels that are most influential to the final decision.

4.4 Ablation Studies

Table 1. Results for our ablation studies

Ablation	Acc. %	Diff. %
Ours (DenseNet121)	88.64	-
-Class weight	87.22	-1.42
-Opening	86.30	-2.34
-Augmentation	83.24	-5.40
-LR Decay	82.45	-6.19
-Transfer Learning	78.78	-9.86

With ablation, we can easily isolate the effects each technique has on our model. After removing data augmentation, the validation loss for the model without augmentation gradually increases and does not recover after the 11th epoch. This indicates that the model is overfitting the training data and yet losing its generalization ability. Eliminating the data augmentation process leads to a 5.40% drop in accuracy. After removing the LR Decay, the model fails to converge smoothly and oscillates around the local minima. The oscillation harms model performances, decreasing the model accuracy by 6.19%. After removing transfer learning, the initial loss for the model becomes much bigger. This proves that transfer learning provides a better starting point for the weights compared to random initialization of these. Meanwhile, as presented in Figure 12, the model converges much slower and suffers from a 9.86% difference in accuracy without transfer learning.

The ablation studies demonstrate the techniques we employ positively influence our model performance.

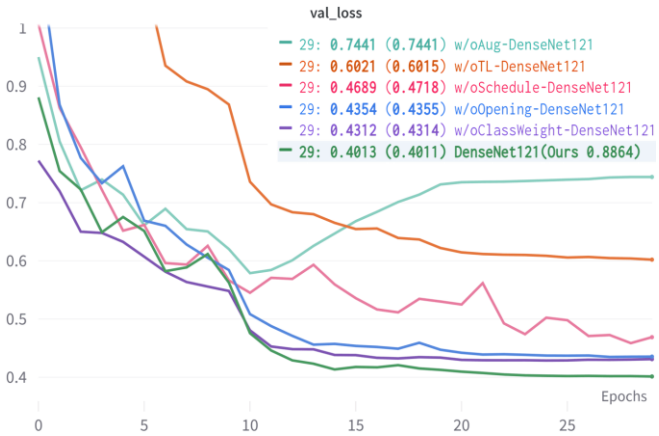

Figure 12. Illustrates the validation loss for our ablation experiments.

5. CONCLUSION AND FUTURE WORK

In our experiments, we compared the performances of different models and conducted ablation studies. Among the models we compared, the minor changes in model structure yielded similar test accuracy. Our work illustrates that the alleged advances proposed in their paper are not reflected in our experiments. Conversely, other techniques like using a finer dataset, learning rate schedule, morphological opening, and data augmentation greatly boosted our model performance. Hence, one should give more attention to preparing the data and resolving the error cases before applying specific tricks that optimize the model performance. For future works, we propose several aspects that could gather attention. Since biases within each batch might lead the model to learn in the wrong direction or information, our model implements class weights to make the entire training dataset balanced, and yet it does not guarantee the balance within each batch. Additionally, given that the images are very similar and obscure, our model tends to confuse those images. Thus, we could implement label smoothing to make the model less confident about its predictions and focal loss to force the model to learn harder samples.

REFERENCES

[1] Schawinski, Kevin, Daniel Thomas, Marc Sarzi, Claudia Maraston, Sugata Kaviraj, Seok-Joo Joo, Sukyoung K. Yi, and Joseph Silk. "Observational evidence for AGN feedback in early-type galaxies." Monthly Notices of the Royal Astronomical Society 382, no. 4 (2007): 1415-1431.

[2] Abraham, Roberto G., Sidney Van Den Bergh, and Preethi Nair. "A new approach to galaxy morphology. I. Analysis of the Sloan Digital Sky Survey early data release." The Astrophysical Journal 588, no. 1 (2003): 218.

[3] Vaucouleurs, G. de. "Classification and morphology of external galaxies." In Astrophysik iv: Sternsysteme/astrophysics iv: Stellar systems, pp. 275-310. Springer, Berlin, Heidelberg, 1959.

[4] Morgan, W. W., Philip C. Keenan, and Edith Kellman. An atlas of stellar spectra. Vol. 5. Chicago: University of Chicago Press, 1943.

[5] Lintott, Chris J., Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick et al. "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey." Monthly Notices of the Royal Astronomical Society 389, no. 3 (2008): 1179-1189.

[6] Lintott, Chris, Kevin Schawinski, Steven Bamford, Anže Slosar, Kate Land, Daniel Thomas, Edd Edmondson et al. "Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies." Monthly Notices of the Royal Astronomical Society 410, no. 1 (2011): 166-178.

[7] Thambawita, Vajira, Inga Strümke, Steven A. Hicks, Pål Halvorsen, Sravanthi Parasa, and Michael A. Riegler. "Impact of Image Resolution on Deep Learning Performance in Endoscopy Image Classification: An Experimental Study Using a Large Dataset of Endoscopic Images." Diagnostics 11, no. 12 (2021): 2183.
[8] Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. "Densely connected convolutional networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708. 2017.
[9] Bengio, Yoshua. "Deep learning of representations for unsupervised and transfer learning." In Proceedings of ICML workshop on unsupervised and transfer learning, pp. 17-36. JMLR Workshop and Conference Proceedings, 2012.
[10] Haralick, Robert M., Stanley R. Sternberg, and Xinhua Zhuang. "Image analysis using mathematical morphology." IEEE transactions on pattern analysis and machine intelligence 4 (1987): 532-550.
[11] Xu, Ziyu, Chen Dan, Justin Khim, and Pradeep Ravikumar. "Class-weighted classification: Trade-offs and robust approaches." In International Conference on Machine Learning, pp. 10544-10554. PMLR, 2020.
[12] Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-cam: Visual explanations from deep networks via gradient-based localization." In Proceedings of the IEEE international conference on computer vision, pp. 618-626. 2017.
[13] Smilkov, Daniel, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. "Smoothgrad: removing noise by adding noise." arXiv preprint arXiv:1706.03825 (2017).