Galaxy Morphology Classification Based on DenseNet-SE4 Algorithm Postprint

Authors: Yu Mao, Liangping Tu, Zhenyang Xu, Yue Jiang and Mingyu Zheng

Date: 2025-09-28T12:32:54+00:00

Abstract

As massive amounts of image data are generated by large-scale sky survey projects, research on the morphological classification of galaxy images has become increasingly important. Deep learning, with the capability of automatic feature extraction, exhibits remarkable performance in image classification algorithms. In the past, most of the excellent algorithm models proposed by astronomers focused on the classification of major categories and often ignored the subtle differences between galaxy categories. For this purpose, based on the DenseNet-121 model, this paper attempts to introduce a variety of improvement strategies such as dynamic multi-scale convolution, learnable grouped convolution, and the squeeze-and-excitation module to optimize the performance of the model. After numerous exhaustive experimental comparisons, the DenseNet-SE4 network with excellent performance is proposed. Subsequently, we conduct comparative experiments between this network and multiple advanced convolutional models on the dataset consisting of Galaxy10 DECaLS and GZD-5. We select the data of eight galaxy categories with similar morphologies, such as round smooth galaxies and barred spiral galaxies, to comprehensively test the classification ability of the model. The experimental results illustrate that the DenseNet-SE4 network achieves an accuracy of 88.96%, a precision of 89.00%, a recall rate of 89.44%, and an F1-score of 89.21% on the test set, thus it has reached the highest level among the eight comparison algorithms. Moreover, the model was tested on data within different redshift intervals, demonstrating good robustness. Finally, the visualization method was employed to further validate the effectiveness and rationality of the DenseNet-SE4 network.

Full Text

Preamble

Research in Astronomy and Astrophysics, 25:085010 (19pp), 2025 August

https://doi.org/10.1088/1674-4527/ade22c
CSTR: 32081.14.RAA.ade22c

Galaxy Morphology Classification Based on DenseNet-SE4 Algorithm

Yu Mao[1], Liangping Tu[1],[2], Zhenyang Xu[3], Yue Jiang[3], and Mingyu Zheng[3]

[1] School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan 114051, China; tuliangping@ustl.edu.cn
[3] School of Science, University of Science and Technology Liaoning, Anshan 114051, China

Received 2025 May 13; revised 2025 May 28; accepted 2025 June 5; published 2025 July 1

Abstract

As massive amounts of image data are generated by large-scale sky survey projects, research on the morphological classification of galaxy images is growing increasingly important. Deep learning, with its capability for automatic feature extraction, exhibits remarkable performance in image classification algorithms. In the past, most excellent algorithm models proposed by astronomers focused on the classification of major categories and often ignored the subtle differences between galaxy categories. For this purpose, based on the DenseNet-121 model, this paper attempts to introduce a variety of improvement strategies such as dynamic multi-scale convolution, learnable grouped convolution, and the squeeze-and-excitation module to optimize model performance. After numerous exhaustive experimental comparisons, the DenseNet-SE4 network with excellent performance is proposed. Subsequently, we conduct comparative experiments between this network and multiple advanced convolutional models on a dataset consisting of Galaxy10 DECaLS and GZD-5. We select data from eight galaxy categories with similar morphologies, such as round smooth galaxies and barred spiral galaxies, to comprehensively test the classification ability of the model. The experimental results illustrate that the DenseNet-SE4 network achieves an accuracy of 88.96%, a precision of 89.00%, a recall rate of 89.44%, and an F1-score of 89.21% on the test set, thus reaching the highest level among the eight comparison algorithms. Moreover, the model was tested on data within different redshift intervals, demonstrating good robustness. Finally, a visualization method was employed to further validate the effectiveness and rationality of the DenseNet-SE4 network.

Key words: methods: data analysis –techniques: image processing –galaxies: general

1. Introduction

The morphological classification of galaxies (Masters 2025) is a crucial task in astrophysics research, facilitating a profound grasp of the formation and evolutionary patterns of galaxies. Its essence lies in revealing the dynamical history and physical mechanisms of galaxies through structural features.

The morphology of galaxies can be classified based on various criteria. Among numerous galaxy classification systems, the Hubble classification system is currently the most widely used. In the Hubble classification system (Hubble 1926), galaxies are mainly categorized into elliptical galaxies, spiral galaxies, barred spiral galaxies, lenticular galaxies, and irregular galaxies. Subsequently, with the continuous refinement of the classification system, galaxies under observation were further classified as merging galaxies and disturbed galaxies according to whether they exhibited signs of ongoing mergers or the presence of tidal debris. In the Hubble classification system, elliptical galaxies and spiral galaxies are two of the most classic categories. Elliptical galaxies are the simplest in morphology within the Hubble sequence. They lack prominent disk structures or spiral arms, exhibiting a smooth luminosity distribution and an approximately ellipsoidal geometric shape.

Spiral galaxies are distinctly marked by their prominent disk structures, winding spiral arms, and central bulges. Meanwhile, significant morphological disparities exist among the subclasses of both elliptical and spiral galaxies. For example, among elliptical galaxies, some assume a nearly spherical form, whereas others exhibit a highly flattened elliptical shape. Based on the degree of flattening, elliptical galaxies are meticulously categorized into round smooth galaxies, in-between round smooth galaxies, and cigar-shaped smooth galaxies. For spiral galaxies, some have a bulge that occupies a substantial proportion of their entire structure, with tightly coiled spiral arms exhibiting a compact and closely intertwined configuration. In contrast, some spiral galaxies possess an exceedingly tiny bulge, accompanied by loose spiral arms that are highly branched. Additionally, contingent upon their relative orientation to Earth, they are categorized as edge-on galaxies.

With the continuous emergence of various sky survey projects, the field of astronomy has transitioned from relying on relatively small datasets to having access to vast and abundant databases. For instance, there are the Sloan Digital Sky Survey (York et al. 2000) on a large scale, the Dark Energy Spectroscopic Instrument (DESI; Cheng et al. 2021) sky survey project, and the Vera C. Rubin Observatory sky survey project (LSST; Ivezić et al. 2019) that is about to be fully operational. While the sharply increasing galaxy image data provides astronomers with a great deal of astronomical data of significant scientific research value, it has also rendered the previous method of manually assigning morphological types no longer feasible. Against the backdrop of massive amounts of data, it is crucial to develop novel, more intricate, and highly productive tools to handle and analyze such an enormous amount of information.

In recent years, as machine learning and convolutional neural networks advance rapidly and vigorously, approaches to the autonomous classification of galaxy morphologies have been greatly promoted. In the field of machine learning, Mukundan et al. (2024) proposed a new method for galaxy morphological classification that combines non-parametric statistics from galaxy images with the K-Nearest Neighbors algorithm. This approach uses statistical metrics computed from galaxy images via STATMORPH (Rodriguez-Gomez et al. 2018), including concentration (C), asymmetry (A), smoothness (S), and Gini coefficient (G), to classify galaxies into early-type and late-type categories or multiple classes such as elliptical galaxies, spiral galaxies, and irregular galaxies. The results show that the method is effective and has the potential to be extended to more in-depth future surveys. Aguilar-Argüello et al. (2025) adopted the XGBoost machine learning method to classify galaxy morphologies into early galaxies, late galaxies, and five other classes, namely E, S0-S0a, Sa-Sb, Sbc-Scd, and Sd-Irr, based on color gradients, asymmetry gradients, global colors, geometries, and parameters. The results demonstrate that this method is consistent with existing research, but issues such as class imbalance and the feasibility of introducing additional parameters still persist. Furthermore, numerous outstanding research methods and techniques have contributed to the field of galaxy morphology classification (Barchi et al. 2017, Selim & Abd El Aziz 2017, Vavilova et al. 2022, Jiang et al. 2023 and Zhou et al. 2022).

Current research indicates that studies based on traditional machine learning methods exhibit higher accuracy and interpretability in small-sample datasets. However, these approaches heavily rely on manual sample selection, feature design, and feature selection, making them less effective when applied to larger datasets and difficult to guarantee the optimality of learned features.

As a branch of machine learning, deep learning has remained a focal point in model research and applications in recent years. Deep neural networks, with their large number of parameters, exhibit stronger fitting capabilities. Therefore, the larger the data scale, the more pronounced the improvement in model performance. For instance, Pan et al. (2024) introduced a Wide-field Infrared Survey Explorer-based galaxy classification network (WGC) to classify spiral and elliptical galaxies in wide-field infrared survey images. The WGC achieved an accuracy of 89.03%, outperforming the combination of K-means or SVM with color-color methods in more accurately identifying galaxy morphologies. The enhanced variant, WGC-mag, integrated amplitude parameters with image features, further boosting the accuracy to 89.89%. However, this study exclusively employed three bands for composite classification and did not further investigate the information or utilization of other bands. Wu et al. (2022) designed a lightweight EfficientNet-G3 framework and selected the Galaxy Zoo (GZ) dataset for comparative evaluation of the model. Experimental results showed that, with only one-fifth of the parameters of ResNet-26, the network increased the accuracy by 1%. However, lightweight network designs may exhibit weaker capabilities in capturing deep-level features and lower model robustness. Li et al. (2022) introduced the RegNet algorithm into the research of galaxy morphological classification. They conducted experiments with a new dataset composed of seven-category samples selected from the Extraction de Formes Idéalisées de Galaxies en Imagerie (EFIGI) and Galaxy Zoo 2 (GZ2) databases, and proved that the classification efficiency of the RegNet model outperforms other neural networks. Meanwhile, based on this model, they studied improvement strategies from the perspectives of attention mechanism and regularization, and determined the optimal classification model, the RegNetX-CBAM4 network. They also considered the influence of classification bias on galaxy morphological classification and the impact of redshift variation on the algorithm' s performance. Li et al. (2023) proposed a Multi-Scale Convolutional Capsule Network (MSCCN) model for galaxy morphological classification. This model adopts a multi-branch structure to improve the convolutional layer for extracting multi-scale hidden features. Subsequently, the extracted features are encapsulated and delivered to the capsule layer. The sigmoid function is employed to substitute the softmax function in the dynamic routing process. This substitution significantly enhances the robustness of the MSCCN. The results prove that it has good generalization ability and robustness. Aniyan & Thorat (2017), Ai et al. (2022), Zhang et al. (2022), Zhang et al. (2024), and Deng et al. (2024) have also conducted research on the classification of galaxy morphology using neural network methods.

Current studies on galaxy morphology classification, whether based on machine learning or deep learning, still face several unresolved challenges, though some novel solutions are being explored. For example, Ghaderi et al. (2025) proposed a technique based on computing Zernike moments and demonstrated its effectiveness in improving model performance, though it may be limited to enhancing certain models. To address challenges such as category similarity and inter-class data imbalance in existing galaxy image recognition studies, Ma et al. (2022) proposed a novel learning method named "Hierarchical Imbalanced Data Learning with Weighted Sampling and Label Smoothing" (HTWL). This approach employs weighted sampling and label smoothing techniques to tackle these issues and explores the recognition of galaxy morphologies ranging from round smooth, in-between round smooth, cigar-shaped, edge-on, to spiral. HTWL provides an effective solution to the data imbalance and category similarity problems commonly encountered in most galaxy morphology classification research. Medina-Rosales et al. (2024) demonstrated that data may exhibit biases in observable parameters, and models trained on such biased data can transfer these biases to predictions of new models. To address this, they introduced a novel method for modeling biased data, which was shown to yield bias-free models on biased datasets through comparisons with other debiasing models. Additionally, Ye et al. (2025) explored an unsupervised domain adaptation method, fine-tuning models trained on DECaLS images with GZD-5 labels for Beijing–Arizona Sky Survey (BASS) and the Mayall z-band Legacy Survey (MzLS) images. This approach reduced bias in galaxy morphology classification within the BASS survey and MzLS survey. Fang et al. (2025) proposed a USmorph framework for automatic galaxy morphology classification by integrating unsupervised and supervised machine learning methods.

This paper presents an improved network for galaxy morphological classification based on the DenseNet algorithm. An attention module is introduced subsequent to the feature concatenation operation of the denseblock of this network, enabling the selection of the combined features. The experimental results have proven that this model shows outstanding classification performance within the dataset of this paper. Moreover, this network has been compared with other design strategies and several Convolutional Neural Networks (CNNs), including AlexNet (Krizhevsky et al. 2017), VGG (Simonyan & Zisserman 2015), ResNet (He et al. 2016), MobileNet (Sandler et al. 2018), EfficientNetV2 (Tan & Le 2021), RegNet (Radosavovic et al. 2020), EfficientNetB0 (Tan & Le 2020) (Liu et al. 2022). The experimental outcomes indicate that this network demonstrates the best performance on the test set, with the accuracy, precision, recall, and F1-score reaching 88.96%, 89.00%, 89.44%, and 89.21%, respectively.

The structure of this paper is as follows. In Section 2, we introduce the source of the dataset used in this paper, as well as the process of dataset division and preprocessing. Section 3 elaborates in detail on the DenseNet network model selected in this paper, along with the improvement strategies for various models. Section 4 presents the implementation details of the training, and simultaneously shows the experimental comparison results of different models, and expounds on the experimental results of the optimal model. Section 5 employs visualization methods to analyze the experimental results in order to enhance the interpretability of the model. Section 6 summarizes the achievements, draws conclusions, and lists some useful references.

2. Data

The imaging datasets employed in this paper mainly come from the Galaxy10 DECaLS dataset released on the astroNN (Leung & Bovy 2018) website. Second, a part of the data is selected from the GZ DECaLS project for supplementation. In this section, we describe the dataset, mainly including data selection, dataset division, and data preprocessing.

2.1. Data Selection

The Galaxy10 DECaLS dataset is a screened standard dataset. Its image data is sourced from the DESI Legacy Imaging Surveys, and the labels are from Galaxy Zoo. The images in the dataset have been standardized in terms of size, resolution, and spectral bands. This data has a high level of annotation consistency. The ten galaxy morphological categories it contains are respectively round smooth galaxies, in-between round smooth galaxies, cigar-shaped smooth galaxies, barred spiral galaxies, unbarred tight spiral galaxies, unbarred loose spiral galaxies, edge-on galaxies without bulge, edge-on galaxies with a bulge, disturbed galaxies, and merging galaxies.

The improved model in this paper is expected to show a greater focus on the subtle differences in information and structural features between different categories of galaxy images. For example, in the dataset, both the round smooth galaxies and the in-between round smooth galaxies exhibit approximately circular symmetrical contours. However, the luminosity of the former decreases from the center outward, while the latter has an increase in brightness at a certain distance from the center. This enables us to concentrate more on the change in the brightness curve rather than the obvious change in the overall shape, thus strengthening the model' s aptitude to identify the early stages of galaxy evolution. Both the unbarred tight spiral galaxies and the unbarred loose spiral galaxies have no central bar structure, and their spiral arms extend directly outward from the bulge. This allows the model to distinguish the evolutionary stages based on the tightness of the spiral arms, instead of relying on the number or direction of the spiral arms for learning. Both the edge-on galaxies without bulge and the edge-on galaxies with bulge present an extremely edge-on and elongated morphology. This empowers the model to grasp the features of central symmetry and the bulge contour to identify the presence of a bulge. Therefore, we have selected the data of eight categories out of the ten categories in this dataset as the main part of our dataset, so as to train and test the model's capability of concentrating its attention on the subtle differences in information and structural features between different categories of galaxy images.

Some categories in the Galaxy10 DECaLS dataset have relatively small amounts of data, resulting in an unbalanced dataset. Therefore, we also used the classification criteria and data from the GZ DECaLS project to supplement the data of four categories: cigar-shaped galaxies, unbarred tight spiral galaxies, edge-on galaxies without a bulge, and edge-on galaxies with a bulge. Compared with the GZ2 images, the images used in DECaLS are deeper and have a higher resolution, which can reveal morphologies that were previously invisible and improve the morphological clarity of galaxy images. For faint features, such as tidal features, low surface brightness spiral arms, or more complex features, such as flocculent spiral arms, etc., as shown in Figure 1. All these can greatly improve the model' s competence in learning and extracting features. In the GZD-5 volunteer classification activity of the GZ DECaLS (Walmsley et al. 2021) project, an improved decision tree was used to better identify merging galaxies and dim bar structures in barred spiral galaxies. This study selected 253,286 galaxy samples from the GZ DECaLS.

However, the number of classification votes for most of the data in GZD-5 is only about 5. Therefore, we hoped to exclude the data with low persuasiveness. At the same time, we referred to the cutting method given by (Walmsley et al. 2021) to obtain some of the desired data. We selected a part of the screened data through visual inspection to supplement and balance the dataset. The specific screening criteria and the number of selected samples are shown in Table 1.

After integrating the above two groups of datasets, we obtained a total of 2645 round smooth galaxies, 2027 in-between round smooth galaxies, 2374 cigar-shaped smooth galaxies, 2043 barred spiral galaxies, 2185 unbarred tight spiral galaxies, 2628 unbarred loose spiral galaxies, 2213 edge-on galaxies without bulge, and 2409 edge-on galaxies with a bulge, amounting to a total of 18,524 pieces of image data. Examples of images of the eight categories are shown in Figure 2.

2.2. Data Set Division

For the purpose of ensuring that the sample sizes of the validation set and the test set are sufficient, and simultaneously ensuring that the data in the training set does not affect the validation and testing of the model, we first selected 8000 galaxy images out of the 18,524 galaxy images using the classification sampling method. We divided 8000 images equally between the validation set and the test set. After that, the remaining 10,524 images were assigned to the training set. Meanwhile, data augmentation was carried out on the images in the training set. The data augmentation methods included random small-angle rotation, as well as horizontal and vertical flips of the images. The final total number of samples was 39,496, and the ratio of the training set, the validation set, and the test set was 8:1:1. The basic information of each category is shown in Table 2.

2.3. Pre-processing of Data

When a convolutional model utilizes image data for model training, the quality of the images directly affects the experimental results of the algorithm. Consequently, we implement a preprocessing process for the image data to enhance the model' s feature extraction capability. The initial pictures in the Galaxy10decals dataset have a resolution of 256 × 256. Nevertheless, there is a substantial amount of sky background within these images, as depicted in Figure 3. Utilizing these images for model training will enable the network to learn a significant amount of unnecessary information. As a result, it will be rather difficult for the model' s performance to reach its optimal state. Therefore, prior to the data division, we first conducted central cropping on the original images. This operation removed some of the background information and simultaneously placed the target celestial object of interest in the central region of the image. In the second step, the images were reduced to a size of 128 × 128 through bilinear interpolation, aiming to decrease the computational workload and boost the training speed. It can be seen that the processed images have retained the contour and feature information of the galaxies and highlighted the classification targets. Finally, data augmentation was performed on the divided training set. We employed methods such as image rotation, vertical flipping, and horizontal flipping to augment the dataset, which is conducive to improving the model' s ability to generalize and its robustness.

Finally, we carried out normalization and standardization on the data. This can accelerate the convergence speed during model training and also increase the accuracy of the model. The images extracted from GZD-5 have a pixel size of 424 × 424. We applied the same processing method. First, we carried out central cropping to make the image size 224 × 224, which can retain the effective information in the original images to the greatest extent. Then, through interpolation processing, the images were reduced to the size of 128 × 128. Finally, standardization processing was performed. The data processing flow is shown in Figure 4.

3. Deep CNNs

CNN is a deep learning network that is currently widely applied. Its core concept is to efficiently extract spatial feature information via local perception and weight sharing. A standard CNN primarily consists of six components: the input layer, the convolutional layer, the pooling layer, the activation function layer, the fully connected layer, and the output layer. The network is capable of forward propagation, during which it calculates the output results by utilizing the training data and weight parameters. In addition, it conducts backpropagation to update the weights based on the gradients. Currently, CNN has achieved a series of breakthrough research results and has become a popular subject of extensive research. In this paper, we mainly introduce the DenseNet network (Huang et al. 2017) and other classic neural networks, such as AlexNet, VGG, ResNet, MobileNet, EfficientNetV2, RegNet, EfficientNetB0 and ConvNeXt-t.

3.1. DenseNet Network

DenseNet (Huang et al. 2017) is a deep learning network proposed in 2017. Its design inspiration comes from the concepts of ResNet (Residual Network) and its predecessor, Highway Networks. ResNet enables the training of deeper CNN models, thereby achieving higher accuracy. The essence lies in establishing a "shortcut connection" between the previous layer and the subsequent layer. This is helpful for the backpropagation of gradients during the training process, thus making it able to train deeper CNN models. The output of a traditional network at layer i is:

For ResNet, an identity function from the input of the previous layer has been added:

DenseNet utilizes a densely-connected architecture, where each layer has connections to all the subsequent layers. This approach ensures that feature maps from all the previous layers are received as input by each layer, promoting feature reuse and enhancing gradient flow throughout the network. Mathematically, the output of the layer incorporates the concatenated feature maps of all preceding layers, enabling richer feature representation and improved model performance.

Another significant characteristic of DenseNet is achieving feature reuse by the association of features in the channel dimension. These characteristics enable DenseNet to outperform ResNet in terms of performance, and it accomplishes this by using fewer parameters and reducing computational costs. There are four design structures for the DenseNet model, namely DenseNet-121, DenseNet-161, DenseNet-169, and DenseNet-201. The differences among these structures are reflected in the changes of the hyperparameter growth rate K value, the configuration of the hierarchical depth, and the differences in the initial feature dimensions.

The network architecture of DenseNet predominantly comprises denseblocks and transitions. Within a denseblock, the feature maps of every layer share identical spatial dimensions, allowing them to be concatenated along the channel dimension. Following the convolution operation of each layer within all DenseBlocks, K feature maps are output. In DenseNet, K is defined as the growth rate, and it represents a hyperparameter. Suppose the number of channels of the feature map of the input layer is K0, then the number of channels of the input of the layer is:

Therefore, with the growth in the number of layers, the input of the DenseBlock will reach an extremely large scale. Due to the very large input of the subsequent layers, a bottleneck layer is adopted inside the denseblock to reduce the computational amount. The primary functions of the Transition layer are to link two consecutive DenseBlocks and downsample the feature map. Additionally, the Transition layer is able to compress the model. The network structure of the DenseNet-121 backbone network used in this paper is shown in Figure 5.

4. Methods

In Section 3.1, we presented the design concept of the DenseNet model and elaborated on its model structure in great detail. In this section, our emphasis will be on depicting the strategies for network enhancement and the specific experimental details.

4.1. Optimized DenseNet Structure

Derived from the DenseNet network, our model incorporates enhancements to achieve superior performance. As described in Section 5.2, the DenseNet model has a total of four different model structures. The different parameters among each structure determine the complexity of the model and its applicable scenarios. To select the most suitable model for our galaxy image dataset, we carried out experiments on the four models with the dataset in this paper and compared the four groups of models from various perspectives, as stated in Section 5.2. After comprehensive consideration, we chose the DenseNet-121 model, which is the most appropriate one.

In the DenseNet-121 model, a dense connection method is utilized to reuse the extracted features. This design compels feature reuse, boosts the flow of gradients during the training process, and can mitigate the issue of vanishing gradients. However, not all early-stage features are beneficial to the subsequent layers. Under such circumstances, there exists a substantial amount of redundant information within the internal connections of the model. Moreover, DenseNet-121 expands the receptive field step by step by stacking fixed 3×3

convolutions. This may lead to the model failing to correctly capture the effective feature information of the input data in the shallow network. Additionally, the receptive field of the fixed convolutions grows slowly. These issues cause the possibility that the model is unable to effectively capture the global context information. Therefore, we have considered the following solutions to improve the model' s capabilities.

This paper incorporates the attention mechanism to enhance the model's performance without altering the model structure. As a module widely incorporated into networks to boost the performance of networks, the attention mechanism enables the network to learn global information, selectively emphasize significant features, and suppress useless features. It can decrease the weights of the ineffective feature channels thereby generated by the dense connections in DenseNet, improving the feature representation. We attempted to introduce the Squeeze-and-Excitation (SE; Hu et al. 2011) module, the Efficient Channel Attention (ECA) module (Wang et al. 2020), and the Convolutional Block Attention Module (CBAM; Woo et al. 2018). The SE and ECA modules are one-dimensional channel attention mechanisms, which are capable of focusing on the channel information of the model. The structure of the SE module is shown in Figure 6. On the other hand, the CBAM module constructs a dual-channel mechanism covering both the channel and spatial dimensions.

Although the attention mechanism has demonstrated significant performance improvements in various visual tasks, its effectiveness highly depends on the structural characteristics of the target network and the requirements of the tasks. On the other hand, the fixed 3×3 convolutions in the shallow network of the DenseNet model have a small receptive field, and these fixed convolutions have poor robustness to variations in the input scale. We made an attempt to substitute the 3×3 convolutions in the dense layers of the model with dynamic multi-scale convolutions as illustrated in Figure 7. The sizes of the three convolutions are 1×1, 3×3, and 5×5 respectively. Adjustable weight values are allocated to the three branches of the convolutions. Moreover, by referring to the learnable grouped convolutions in the CoDenseNet (Huang et al. 2018) shown in Figure 8, we replaced the 1×1 convolutions in the dense layers so as to combine features more efficiently.

Based on the improvement concepts mentioned above, we have designed the following seven optimization strategies. We respectively attempted to incorporate attention modules at five positions of the DenseNet model, and also introduced dynamic multi-scale convolutions and learnable grouped convolutions into the dense layers:

1. Introducing an attention module after the feature module at the end of the model and before the global average pooling is also an option worth trying. Here, it enables the final calibration of the deeply abstracted features, highlighting the channel responses that are strongly related to the categories.

2. Add the module between the 1×1 convolution and the 3×3 convolution in the DenseLayer block. As the core unit for feature reuse and growth, the 1×1 convolution in the dense layer itself serves the function of reducing the dimensionality of channel features. Adding an attention module between the convolutions can directly act on the features after dimensionality reduction, screening the features. As a result, the subsequent 3×3 convolution operation is performed on the features weighted by the attention mechanism. This ensures that high-value information is fused during the subsequent dense connections.

3. The Transition layer compresses the number of channels via a 1×1 convolution and subsequently downsamples the spatial dimensions of the feature map. Hence, embedding an attention module at this location allows for the recalibration of channel weights at this stage. Moreover, it enables the selection of crucial spatial information prior to downsampling, preserves key features, and enhances the input quality of the subsequent dense block.

4. Add an attention module at the terminal part of denseblock, that is, after the feature concatenation. The traditional concatenation method is to concatenate the outputs of all the previous layers. Although processed by dropout, there is still a large amount of useless information. Adding the module at this position can optimize the sensitivity of useful features from a large number of features and improve the overall performance of the model. After comparing the experiments of the above five schemes, we select the optimal model for further improvement.

5. In the ResNet model, the SE module is introduced at the end of the residual structure, and in RegNetY and EfficientNetV2, the attention module is also added after the 3×3 convolution. Because the features after this convolution contain more spatial context information, the features can be adjusted more comprehensively at the channel or spatial level. Following the same design concept, we also attempt to introduce the module after the 3×3 convolution in this paper.

6. We try to add the dynamic multi-scale convolution to the model. Our purpose is to allow the model to enhance its receptive capacity for input features at the feature scale level, thus boosting the model' s performance.

7. The learnable group convolution is an improved traditional group convolution structure, which can dynamically learn the channel grouping strategy to enhance the feature representation capability and computational efficiency of the model. According to the CoDenseNet model proposed by Huang et al. (2018), this structure is used to optimize the cross-layer feature fusion. Therefore, we also attempt to apply this structure to the model.

The above are all our improvement strategies based on the DenseNet model, and all the experimental results can be found in Section 5.2. Through comparing the experimental results, we finally determined that the optimal designed model among the improvement schemes is DenseNet-SE4, which means that the SE module is introduced into the model structure in Scheme 4. From the results of all our optimization strategies, we have summarized some experiences: The location of the attention mechanism within the network framework has a marked effect on the model' s operational efficiency and output quality. Adding it in an inappropriate position has the potential to lead to a deterioration of the model' s performance. Moreover, for the data samples in this paper, it is inappropriate to introduce models with complex structures and a large number of parameters, as this is likely to cause overfitting, resulting in poor generalization performance of the model and relatively poor performance in the test set. Therefore, we did not consider adding multiple attention mechanisms. Meanwhile, the introduction of multiple structures will cause these structures to interact with each other, which will make the adjustment of the model' s parameters difficult and complicate the process of the model achieving convergence during training.

4.2. Details of Implementation

We have attempted different improvement strategies for various models, and finally selected the appropriate optimization strategies for each model based on multiple aspects such as the convergence speed and convergence accuracy. In this paper, the DenseNet model and other selected classification models employ the Stochastic Gradient Descent with Momentum (SGDM; Liu et al. 2020) optimization method. The batch size of all models is configured as 128, the number of epochs is set to 60, and the loss function is defined as the cross-entropy loss function. For the AlexNet and VGGNet, the Adaptive Moment Estimation (Adam; Kingma & Ba 2017) optimization approach is utilized, and appropriate learning rates are given. All our experiments are implemented based on PyTorch, and all the networks are trained from scratch without using pre-trained weights.

5. Results and Discussions

In this section, to measure the classification ability and generalization ability of the models more accurately, we selected some performance metrics for analysis. First, we introduced these metrics. Subsequently, we compared the classification effects of four DenseNet model architectures and multiple optimization schemes on the dataset of this paper. After that, we tested the performance of the models in different redshift intervals. Finally, we conducted a visual analysis of the models.

5.1. Classification Metrics

For the purpose of efficiently evaluating the performance of classification models, we have selected the following common performance classification indicators. We utilized a confusion matrix to assess the experimental results. Accuracy represents the ratio of accurately categorized instances to the cumulative number of observations. Precision represents the proportion of true positives among all positive predictions. Recall measures the proportion of correctly identified positive cases relative to all actual positive cases. Additionally, we also employed the F1 score and the Receiver Operating Characteristic (ROC) curve. The F1 value is a comprehensive performance metric, where a higher value indicates better model quality. For the ROC curve, the closer its value is to the upper left corner formed by the x-axis and the y-axis, the better the performance of the model.

5.2. Comparison and Discussion

All the training of our models was carried out on the dataset presented in this paper. We monitored the performance of the models in the validation set, and ultimately utilized the test set to assess the classification performance of the models. In the first experiment of this paper, we initially tested the four designed DenseNet models that had been trained on the test dataset, and the outcomes are presented in Table 3. We compared the four models from various aspects, encompassing the quantity of parameters, the count of floating-point operations (FLOPS), accuracy, precision, recall, and the F1 score. Among them, FLOPS represents the number of operations required for the model to complete one forward propagation, which serves to evaluate the computational complexity of the model. The parameters refer to the total number of all trainable weights in the model. Experimental outcomes indicate that the performance of the model has no proportional connection with the FLOPS and the quantity of parameters. The accuracy of DenseNet-121 is only 0.02% lower than that of DenseNet-169. Moreover, DenseNet-161 and DenseNet-201, which have a larger number of model parameters, exhibit lower accuracy. This demonstrates that, for the dataset in this paper, the increase in depth and width of the DenseNet architecture has a saturation effect on the performance improvement. In other words, once a certain scale is exceeded, the marginal gains in performance brought about by the increase in computational resources gradually decline. Taking various factors into comprehensive consideration, we finally selected DenseNet-121 as the basic model, which not only meets the demand for high precision but also features low resource consumption.

Based on the DenseNet-121 model, we carried out experiments in accordance with the improvement schemes described in Section 4.1. Initially, we made attempts to substitute two convolution structures on the foundation of the basic model. Subsequently, we endeavored to add the attention mechanism to the model and designed it as per the provided schemes. The ultimate experimental results are presented in Table 4. By comparing the experimental results, we discovered that incorporating dynamic multi-scale convolution has declined, whereas the performance with learnable grouped convolution has enhanced. Meanwhile, among the five schemes for introducing the attention mechanism, the four models, namely DenseNet121-ECA1, DenseNet121-SE4, DenseNet121-ECA5, and DenseNet121-SE5, have demonstrated remarkable improvements.

DenseNet121-SE4 boasts the highest accuracy, DenseNet121-ECA5 has the highest precision and F1 score, and DenseNet121-SE5 has the highest recall rate, all of which outperform the basic model. The performance of some other models showed no obvious changes, and some even experienced a significant decline.

Finally, based on the four models with improved performance, we attempted to add learnable grouped convolution for comparison. The final results are shown in Table 5. As can be observed from the experimental results, the performance of the four models after the introduction of convolution has all decreased. Our analysis indicates that this could be attributed to the heightened complexity of the models, resulting in overfitting on the dataset of this paper. This overfitting situation hinders the models from fully grasping the synergistic effect between the relevant components. Moreover, after the grouped convolution filters the features, the attention mechanism might make further adjustments. The mutual interference between the grouped convolution and the attention mechanism leads to the features being impacted, making it more challenging for the models to converge. Through a comprehensive comparison of all aspects, the DenseNet121-SE4 model exhibits the best performance and is ultimately identified as our optimal model.

In order to more prominently showcase the performance of the optimized model, we analyze the model results in the test by using Confusion Matrix in Figure 9 and the ROC Curve in Figure 10. The column axis of the confusion matrix represents the true labels of the data samples, and the horizontal axis represents the predicted labels. It can be discerned from the figure that the model is liable to make classification errors among barred spiral galaxies, non-barred compact spiral galaxies, and non-barred loose spiral galaxies. Specifically, a total of 35 barred spiral samples were misclassified as the two types of spiral galaxies, and 46 spiral galaxy samples were wrongly classified as barred spiral galaxies. Meanwhile, the total number of samples that were misclassified between loose spiral galaxies and compact spiral galaxies is 122. Figure 10 employs the ROC curve and calculates the Area Under the curve (AUC). The closer the ROC curve is to the upper left corner, the better the prediction performance of the model. The results indicate that the model has good classification performance for each category of galaxies. Among them, the prediction effects for cigar-shaped galaxies, edge-on galaxies, and in-between round galaxies are the best. The average AUC value of the model also reaches 0.987, which shows that our model exhibits superior overall forecasting performance.

To further investigate the reasons for model misclassification, we randomly selected some samples with correct and incorrect classifications by the model in the three categories of barred spirals, unbarred loose spirals, and unbarred tight spirals. SHAP analysis was performed on these samples to compare and demonstrate the impact of sample features on model predictions, as shown in Figures 11–13. The SHAP analysis method calculates which pixel variations have the most significant impact on prediction results by comparing differences between input samples and background data, thereby assigning feature importance scores to pixels in the image. In SHapley Additive exPlanation (SHAP) plots, red regions represent areas identified by the model as important features, while blue regions indicate suppressed features. As evident from Figure 11, in correctly classified barred spiral galaxy samples, the bar structures are distinctly marked as key regions. In contrast, misclassified barred spirals show a focus on their spiral arm structures, and when the bar structures are indistinct, they are prone to being misclassified as spiral galaxies. Even when both bar and spiral arm structures are detected in a sample, the model tends to prioritize spiral arm features when determining the classification result. For spiral galaxy samples, Figures 12 and 13 show that the model primarily relies on the tightness of spiral arm winding to judge their categories. We argue that the primary reason for the frequent misclassification among these three categories is their morphological similarity, whereby indistinct bar features in barred galaxies can easily lead to misclassification as spiral galaxies. Additionally, misclassification may occur due to either the indistinguishability of barred spiral structures from tight spirals at certain viewing angles or the misjudgment of loose spiral arms due to their dimmer or more dispersed arm structures.

5.3. Comparative Analysis with Various Models

In this section, we compared the DenseNet121-SE4 model with other advanced convolutional models. The models for comparison include AlexNet, Convnext-t, RegNetX-4.0GF, Resnet-34, Vgg-16, Mobilenetv2-1, Efficientnet-B0, and Efficientnet-V2s. We conducted tests on all these models using the dataset in this paper, and made comparisons in terms of the models' accuracy, precision, recall, and F1 score. The final experimental results are presented in Table 6. It can be observed that these CNN models also exhibited good generalization ability on our dataset, with the accuracy rate being over 81%. Particularly, both Efficientnet-v2s and AlexNet had an accuracy rate of approximately 88%. Nevertheless, the Densenet121-se4 model topped the list in all the indicators.

We also argue that prediction errors of models can be attributed to the unclear quality of some data samples and the indistinctiveness of their key features. In the first three categories, the bar structures of some barred spiral galaxies are not obvious in the images, and the subtle morphological differences between loose and tight spirals make them difficult for models to distinguish. In the latter three categories, both cigar-shaped galaxies and edge-on galaxies exhibit elongated strip-like structures from edge-on perspectives, and sometimes the lack of additional characteristic information in the images makes it difficult to predict these samples correctly. Additionally, aside from the intrinsic limitations of the models and the quality of the data, classification prediction errors may also arise from subjective judgment biases in the true class labeling of some samples. During manual labeling, it is challenging to precisely demarcate boundaries, resulting in slight discrepancies between the labels and the actual morphology.

In addition, we extracted the redshift values of the test set and visualized the numerical distribution. Moreover, we investigated the result generated by the algorithm within the known different redshift intervals, as depicted in Figure 15. This figure reveals that the redshift values of the samples in the test set mainly range from 0.03 to 0.09, and the number of samples decreases as the redshift increases. Based on the analysis of the performance of the test set across the five redshift intervals, the model demonstrates high robustness overall, with the accuracy remaining steadily above 0.87. We opine that this might be attributed to the limitation of the sample size.

5.4. Feature Maps Visualization

Deep learning represents an end-to-end learning methodology. In the morphological classification task of galaxy images, we are only able to input image data into the network and subsequently obtain a single category label. For the purpose of exploring the intermediate processes of our network in greater depth, we have decided to use a visualization approach to explore the working mechanism of the model. In this section, we first selected a spiral galaxy sample as the input image for the model. We randomly extracted the variations in the activation intensity of 64 channels before and after incorporation of the attention mechanism, as illustrated in Figure 16. It can be observed that the majority of channels exhibit minimal or even no differences, whereas a small number of channels show substantial variations. We can reasonably infer that at this position, the activation intensity of the channels beneficial to the classification performance of the model increases. For example, in the spiral arm morphology channel, it focuses on the tightness of the spiral arm structure of the spiral galaxy, as well as some contour and texture information. The structure conforms to the structure of the spiral galaxy.

Subsequently, we presented the feature space distribution diagrams and Grad-CAM diagrams of all categories in the dataset before and after the addition of the attention mechanism, as shown in Figure 18. The color intensity in the diagrams intuitively shows the key regions of the images that the model focuses on during prediction. It can be seen that the heatmap of spiral galaxies has radial highlights along the distribution of the spiral arm structures. The heatmap of elliptical galaxies prominently displays the galaxy core and the external contour. The heatmap of barred spiral galaxies exhibits an obvious bar-shaped structure. There is also a clear difference in the presence or absence of a bulge in the two categories of edge-on galaxies. This shows that the model is already capable of quickly defining the core structure of the data for each category.

Additionally, to deepen the comprehension of galaxy morphological classification, we analyzed samples with discrepant prediction outcomes across multiple models and selected a representative subset of illustrative samples for display in Figure 14. Our statistics show that most models tend to confuse three categories: barred spirals, unbarred loose spirals, and unbarred tight spirals, with another three categories: cigar-shaped galaxies, edge-on galaxies, and in-between round galaxies. We argue that due to the unique algorithmic principles and feature learning capabilities of different models, these fundamental differences in principles will cause each model to focus on data features with different priorities. Some models may be better at capturing local features of samples, while others place greater emphasis on global features. Therefore, when faced with complex samples, such differences in feature learning capabilities can lead to divergent predictions for the same sample.

The overall improvement of the model performance of the DenseNet121-SE4 algorithm is limited. However, compared with the DenseNet-121 network, it has improvements in all indicators. The reason for the limited improvement may be due to the imbalance of the dataset and certain deviations existing in the original categories of the data, which are inevitable. Nevertheless, generally speaking, the model' s performance metrics are still quite good.

6. Conclusion

This paper utilizes the dataset composed of the Galaxy10 DECaLS dataset and the dataset from the GZ DECaLS project. The morphologies of galaxies are classified into eight categories: round smooth galaxies, in-between round smooth galaxies, cigar-shaped smooth galaxies, barred spiral galaxies, unbarred tight spiral galaxies, unbarred loose spiral galaxies, edge-on galaxies without bulge, and edge-on galaxies with a bulge. Before using the data, we carried out operations such as central cropping, bilinear interpolation, data augmentation, normalization, and standardization on both parts of the data. However, there are differences in the processing details of the central cropping operation and the linear interpolation operation. After that, we conducted a comparative analysis of the test results of four architectures of the DenseNet model, and selected the DenseNet-121 model as the basic model for the improved algorithm. We studied various improvement strategies for this network. Through comprehensive consideration, we determined that our best improvement plan is the DenseNet-SE4 model, that is, adding the SE module after the feature concatenation operation of the dense block. The experimental results show the DenseNet-SE4 network reaches an accuracy of 88.96%, a precision of 89.00%, a recall of 89.44%, and an F1-score of 89.21% on the test set, proving that this model has better generalization ability.

Acknowledgments

We thank the referees for greatly improving the manuscript. This work was supported by the National Natural Science Foundation of China (NSFC, grant No. U1731128).

References

Aguilar-Argüello, G., Fuentes-Pineda, G., Hernández-Toledo, H. M., et al. 2025, MNRAS, 537, 876

Ai, L. P., Xu, Q. F., Du, L. T., et al. 2022, AcASn, 63, 42

Aniyan, A. K., & Thorat, K. 2017, ApJS, 230, 20
Barchi, P. H., da Costa, F. G., Sautter, R., et al. 2017, arXiv:1705.06818
Cheng, T.-Y., Conselice, C. J., Aragón-Salamanca, A., et al. 2021, MNRAS, 507, 4425
Deng, Z.-J., Yu, S.-Y., Luo, A.-L., Kong, X., & Li, X.-R. 2024, RAA, 24,
Fang, G., Dai, Y., Lin, Z., et al. 2025, A&A, 693, A141
Ghaderi, H., Alipour, N., & Safari, H. 2025, ApJS, 277, 10
He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 770
Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. 2020, ITPAM, 42, 2011
Huang, G., Liu, S., van der Maaten, L., & Weinberger, K. Q. 2018, arXiv:1711.09224
Huang, G., Liu, S., van der Maaten, L., & Weinberger, K. Q. 2017, in 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 2261
Hubble, E. P. 1926, ApJ, 64, 321
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jiang, J., Zhang, J., Li, X., Li, H., & Du, P. 2023, RAA, 23, 115019
Kingma, D. P., & Ba, J. 2017, arXiv:1412.6980
Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2017, Commun. ACM, 60, 84
Leung, H. W., & Bovy, J. 2018, MNRAS, 483, 3255
Li, G., Xu, T., Li, L., et al. 2023, MNRAS, 523, 488
Li, J., Tu, L., Gao, X., et al. 2022, MNRAS, 517, 808
Liu, Y., Gao, Y., & Yin, W. 2020, arXiv:2007.07989
Liu, Z., Mao, H., Wu, C.-Y., et al. 2022, arXiv:2201.03545
Ma, X., Li, X., Luo, A., Zhang, J., & Li, H. 2022, MNRAS, 519, 4765
Masters, K. 2025, arXiv:2502.09610
Medina-Rosales, E., Cabrera-Vives, G., & Miller, C. J. 2024, MNRAS, 531, 52
Mukundan, K., Nair, P., Bailin, J., & Li, W. 2024, MNRAS, 533, 292
Pan, Z.-R., Qiu, B., Liu, C.-X., et al. 2024, RAA, 24, 045020
Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., & Dollár, P. 2020, in 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 10425
Rodriguez-Gomez, V., Snyder, G. F., Lotz, J. M., et al. 2018, MNRAS, 483, 4140
Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. 2018, in 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (Piscataway, NJ: IEEE), 4510
Selim, I. M., & Abd El Aziz, M. 2017, ExA, 43, 131
Simonyan, K., & Zisserman, A. 2015, arXiv:1409.1556
Tan, M., & Le, Q. V. 2020, arXiv:1905.11946
Tan, M., & Le, Q. V. 2021, arXiv:2104.00298
Vavilova, I. B., Khramtsov, V., Dobrycheva, D. V., et al. 2022, KosNT, 28, 03
Walmsley, M., Lintott, C., Géron, T., et al. 2021, MNRAS, 509, 3966
Wang, Q., Wu, B., Zhu, P., et al. 2020, in 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 11531

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. 2018, arXiv:1807.06521
Wu, D., Zhang, J., Li, X., & Li, H. 2022, RAA, 22, 115011
Ye, R., Shen, S., de Souza, R. S., et al. 2025, MNRAS, 537, 640
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579
Zhang, S., Fang, G., Song, J., et al. 2024, RAA, 24, 095012
Zhang, Z., Zou, Z., Li, N., & Chen, Y. 2022, RAA, 22, 055002
Zhou, C., Gu, Y., Fang, G., & Lin, Z. 2022, AJ, 163, 86

Note: Figure translations are in progress. See original paper for figures.