Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/33/20230245

Accurate and efficient galaxy classification based on mobile vision transformer

Xinrui Tan

School of Electrical Engineering and Computer Science, The University of Queensland, Brisbane, QLD, 4072, Australia

xinrui.tan@uqconnect.edu.au

Abstract. Understanding the formation and evolution of galaxies in observational cosmology heavily relies on galaxy morphological classification. Nevertheless, the continuously growing volume of astronomical data has surpassed human capacity for manual classification. In this context, deep learning presents a promising approach to enhancing classifying galaxies. In this paper, the Mobile Vision Transformer (MobileViT) is introduced to construct an efficient and accurate galaxy classifier. Transfer learning is introduced to assist in model fine-tuning. MobileViT combines the features of MobileNet and Visual Transformer (ViT). A lightweight model is used to effectively analyse the relationships between sequences for efficient and accurate classification. Experiments are built on Galaxy10 DECals dataset. Excellent performance is achieved in identifying galaxy types compared to other lightweight models. The model achieves an accuracy of over 87% and maintains a high speed of inference of less than 50 milliseconds per step. Experimental results show that the introduction of MobileViT is the best solution for efficient galaxy classification. The model can be deployed on any portable device for instant observation and classification.

Keywords: morphological classification, mobile vision transformer, transfer learning, lightweight.

1. Introduction

The galaxy morphological classification plays a key role in the field of observational cosmology. It is the cornerstone for building a comprehensive catalogue of galaxies. By systematically examining the different shapes and structures exhibited by galaxies, astronomers can gain valuable insights into the underlying mechanisms of their formation and evolution. However, the vast amount of data available from modern astronomical observations has far exceeded the ability of humans to classify galaxies comprehensively and manually. In response, machine learning (ML) has been introduced for decades to enhance the classification of galaxy morphology. Nowadays, deep learning models have become prominent contributors. Notably, recent advances have produced powerful models leading to the development of multiple highly accurate classifiers. However, this has been accompanied by increasingly complex time requirements associated with training and inference procedures. Efficient models easily available on portable devices can substantially aid astronomy study and instruction, saving a great deal of time and energy.

DOI: 10.54254/2755-2721/33/20230245

The incorporation of ML into the field of galaxy classification has a significant and extended history. Artificial neural networks (ANNs), decision trees, and other basic machine learning techniques were used in the early research into this topic [1, 2]. Deep learning techniques, notably convolutional neural networks (CNN), were increasingly popular for galaxy classification as large-scale models became more prevalent in the field of image classification in the 2010s. Dieleman et al. first used a seven-layer CNN and exploited the translational and rotational invariance of galaxy images to classify galaxy morphology [3]. Building on this work, Kim et al. extended their investigation of galaxy classification by employing a more extensive model similar to VGG [4]. In addition, Zhu et al. modified the ResNet architecture to classify galaxies into five different classes, achieving an impressive accuracy of over 95% [5]. Lately, Lin et al. made a significant breakthrough by introducing Vision Transformers (ViT) to the field of galaxy classification, demonstrating the superior performance of Transformers in analyzing small-sized and faint galaxies [6].

This study identifies an efficient and accurate classifier for galaxy classification, enabling it to be implemented directly on the observatory's data reception equipment. The focus is on achieving high efficiency while maintaining high accuracy by using a lightweight model with fewer parameters. In this study, Mobile Vision Transformer (MobileViT) is fine-tuned [7]. This is a hybrid model combining CNN and transformer features. MobileViT is pre-trained on ImageNet [7]. It is subsequently fine-tuned for the specific task of dividing galaxies into ten discrete classes. During the training of this model, transfer learning is utilised to save time and effort. Additionally, to lessen the danger of overfitting, data augmentation techniques are used on the dataset. Meanwhile, a comprehensive analysis and comparison of the predictive performance and inference speed of models of similar size are performed. The experimental findings demonstrate that MobileViT performs better at accurately recognising galaxy types than other lightweight CNN or ViT models.

2. Methodology

2.1. Dataset description

This project is based on the Galaxy10 DECals dataset, derived from the Galaxy Zoo (GZ) data version 2. The dataset consists of approximately 270,000 SDSS galaxy images, which were meticulously classified by volunteers. Among these images, around 22,000 were selected based on the top 10 categories as determined by the votes of the volunteers [8, 9]. The Galaxy10 DECals dataset comprises a compilation of 17,736 color galaxy images, each with a pixel size of 256x256 and representing the g, r, and z-bands. These images are divided into 10 distinct classes, which are listed in Table 1 [9].

Table 1. Architecture of dataset.

Class Label	Name	# Images
0	Barred Spiral Galaxies	2043
1	Cigar Shaped Smooth Galaxies	334
2	Disturbed Galaxies	1081
3	Edge-on Galaxies with Bulge	1873
4	Edge-on Galaxies without Bulge	1423
5	In-between Round Smooth Galaxies	2027
6	Merging Galaxies	1853
7	Round Smooth Galaxies	2645
8	Unbarred Loose Spiral Galaxies	2628
9	Unbarred Tight Spiral Galaxies	1829

2.2. Proposed approach

This study introduces MobileViT as a backbone network. The network is further fine-tuned by weight adjustment to facilitate galaxy classification. In addition, preprocessing layers are added for data augmentation. Figure 1 visualizes the overall experimental process. The overall process is divided into

DOI: 10.54254/2755-2721/33/20230245

five steps. First, in the preprocessing stage, images are rescaled and randomly cropped, flipped, and rotated to accomplish data augmentation. Second, each batch of images after preprocessing is fed into the core network, i.e., the MobileViT architecture. Third, MobileViT performs feature extraction and feature fusion on the input data. Additionally, the output is discarded during the training process according to a predetermined discard rate, and fourth, the output features are forwarded to the classifier. The classifier consists of a fully connected layer using Softmax activation. The resultant output is the probability associated with each individual category.

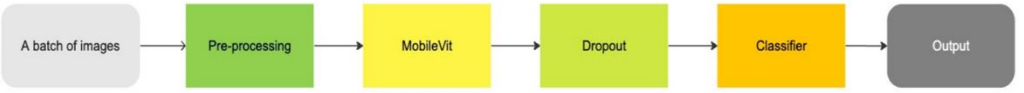


Figure 1. Process of network.

2.2.1. Pre-processing. The model's pre-processing layers serve the purpose of data augmentation and encompass several operations, namely random crop, rescaling, random flip, random rotation, and random zoom (see Figure 2). The picture pixel values are normalised using x=x/127.5-1 (where x indicates the pixel values), translating all pixel values into the range of [-1, 1], to guarantee prevention of exploding gradients, convergence speed, and model accuracy enhancement. These pre-processing layers are integrated at the bottom of the model and are executed on the GPU to avoid potential CPU bottlenecks. Additionally, the initial 3-dimensional matrix representation of the images is compressed and transposed into tensors, which represent the pixel values.

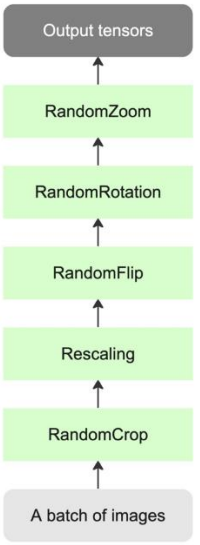


Figure 2. Architecture of pre-processing module.

2.2.2. MobileViT. In this study, MobileViT lightweight network is used as the basis for galaxy classification. MobileViT combines the advantages of MobileNet and ViT. MobileNet is a CNN with a lightweight backbone structure. ViT, based on the self-attention mechanism, is good at capturing global feature information. Figure 3 illustrates the network architecture of MobileViT. In the MobileViT block, the kernel size "n" is typically set to three. Downward arrow-labelled structures indicate down-sampling operations. MobileViT offers three distinct configurations, namely MobileViT-S, MobileViT-XS, MobileViT-XXS in descending order of the number of parameters. In this study, MobileViT-S is

DOI: 10.54254/2755-2721/33/20230245

employed. During this procedure, the pre-processed batch of images undergoes a down-sampling 3-by-3 convolution before being fed into several MobileNetV2 blocks for feature extraction and down-sampling. Subsequently, the batch passes through a sequence of combinations of MobileNetV2 blocks and MobileViT blocks. The MobileNetV2 blocks facilitate down-sampling, while the MobileViT blocks are responsible for capturing the fusion of local and global representations. Finally, the processed batch undergoes a 1×1 convolution along with a global average pooling, generating the logits for output.

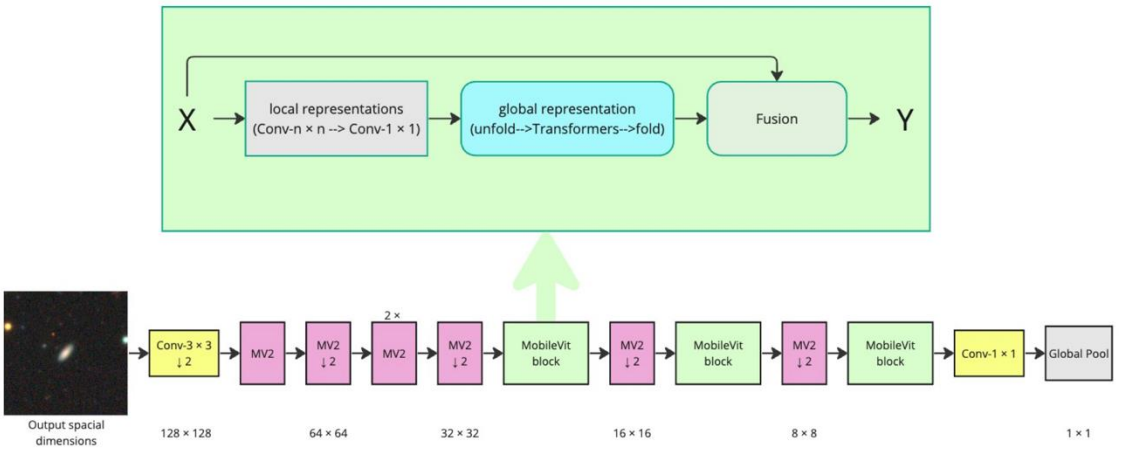


Figure 3. Architecture of MobileViT, in which Conv-n × n denotes a normal n × n convolution and MV2 indicates the MobileNetV2 block. Blocks responsible for down-sampling are designated with ↓ 2 [10].

2.2.3. Loss function. In this work, sparse categorical cross-entropy is employed. It is frequently used in classification applications. With only one accurate class considered for each input sample, it calculates the difference between the anticipated probability distribution and the actual class label. When the classes are mutually exclusive and the target class is represented as an integer index rather than a one-hot encoded vector, as it is in this study, this loss function is very suitable. In order to help the model make accurate class predictions during training, the sparse categorical cross-entropy seeks to reduce the discrepancy between projected probability and the actual class label. The mathematical expression is as follows,

$$Loss_{SCCE} = -\sum_{i}^{C} t_i \times log(p_i), \tag{1}$$

where C is the collection of class labels, t_i is the truth label for the i-th class, and p_i is the class's Softmax probability.

2.3. Implementation details

The model training process utilizes Nvidia's Tesla A100 GPU, along with 80 GB of system memory and 12 vCPUs. An exponentially declining learning rate scheduler is used to optimise the training process. The initial learning rate, decay rate, and number of decay steps were all set to 0.002, 0.01, and 10,000 respectively. The learning rate is displayed beneath:

$$lr = init\_lr \times decayed\_rate^{(steps/decayed\_steps)}, \tag{2}$$

DOI: 10.54254/2755-2721/33/20230245

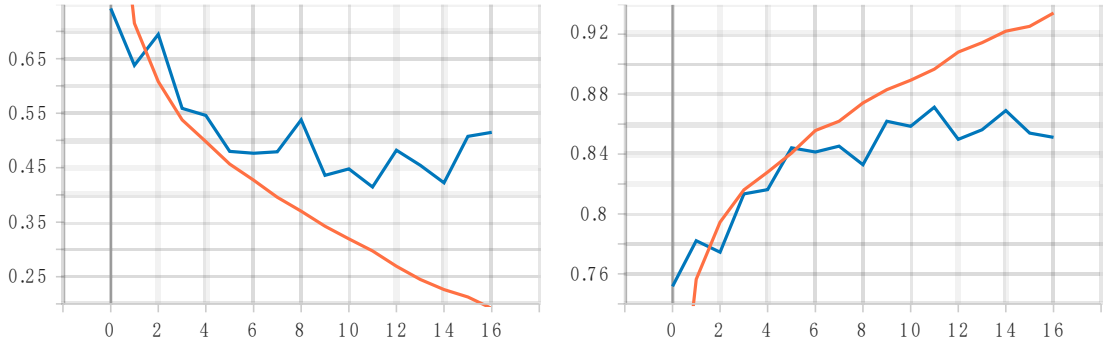
where $init\_lr$ is the initial learning rate, the $decayed\_rate$, in this case, is 0.01, $steps$ is the number of steps been taken, and the value for $decayed\_steps$, in this case, is 10,000. During the training process, the Adam optimizer is employed, incorporating a weight decay rate of 0.01 [10]. In addition, an early stop callback mechanism is implemented to monitor the accuracy of the validation. The early stop mechanism consists of a 5-calendar-time patience. Starting from the third calendar time, the optimal model weights will be automatically restored.

3. Result and discussion

This chapter aims to provide a comprehensive analysis of the results of the model, including visualization and discussion of various performance metrics. Specifically, loss and accuracy, classification reports, and confusion matrices for each cycle are examined in detail. In addition, a comparative analysis of the model's performance with other existing models deployed for the same galaxy classification task is presented. This comparative assessment helps to highlight the exceptional efficiency that the proposed model has shown in accomplishing the task at hand.

3.1. Performance analysis

For both the Training and Validation datasets, the Loss and Accuracy metrics are shown graphically in Figure 4. The Training loss initially exceeds 1 but gradually converges to a value below 0.20. On the other hand, the Validation loss exhibits a general convergence trend at approximately 0.49, albeit with minor oscillations. Notably, the Training Accuracy demonstrates a gradual and consistent increase, starting from below 74% and ultimately reaching a commendable 93%, whereas the Validation Accuracy oscillates at around 86% after convergence. Best weights are restored by early-stopping call-back from the 12th epoch, with a validation accuracy of 87.12%.



(a) Train-validation loss

(b) Train-validation accuracy

Figure 4. The curve of loss and accuracy, in which the orange and blue curve indicate train and validation metrices respectively.

Table 2 presents the classification report derived from evaluating the model on the test dataset. The overall performance is deemed satisfactory, except for class 2, where the model's predictive accuracy was notably lower. Notably, for classes 3 to 7, representing "Edge-on with Bulge galaxy," "Edge-on without Bulge galaxy," "In-between Round Smooth galaxy," "Merging galaxy," and "Round Smooth galaxy," the f1-scores exceeded 0.90. This indicates that the model excels at accurately identifying round bright spots and elongated structures characteristic of these classes. Conversely, for distributed galaxies that lack these distinctive attributes, the model's predictions were found to be less accurate.

DOI: 10.54254/2755-2721/33/20230245

Table 2. Classification report.

	Precision	Recall	F1-score	Support
Class 0	0.91	0.90	0.90	453
Class 1	0.75	0.76	0.76	68
Class 2	0.66	0.50	0.57	216
Class 3	0.92	0.95	0.93	343
Class 4	0.94	0.92	0.93	298
Class 5	0.91	0.94	0.93	407
Class 6	0.88	0.93	0.90	381
Class 7	0.92	0.95	0.93	530
Class 8	0.81	0.76	0.78	478
Class 9	0.80	0.85	0.83	374
Accuracy			0.87	3548
Macro avg	0.85	0.85	0.85	3548
Weighted avg	0.87	0.87	0.87	3548

The confusion matrix depicting the model's predictions on the test dataset is presented in Figure 5. Consistent with the observations made in the classification report, the model exhibited difficulties in distinguishing distributed galaxies from other galaxy types, particularly Unbarred Loss Spiral galaxies. This confusion can be attributed to the presence of similar blue and white spots in the centre of images for both Distributed galaxies and Unbarred Loss Spiral galaxies. However, the model encountered challenges in identifying the spiral arms characteristic of Unbarred Loss Spiral galaxies, which can be considered a comparatively subtle feature. Additionally, the model displayed some confusion among the three types of Spiral galaxies, thereby resulting in relatively lower f1-scores for classes 0, 8, and 9, as outlined in Table 2.

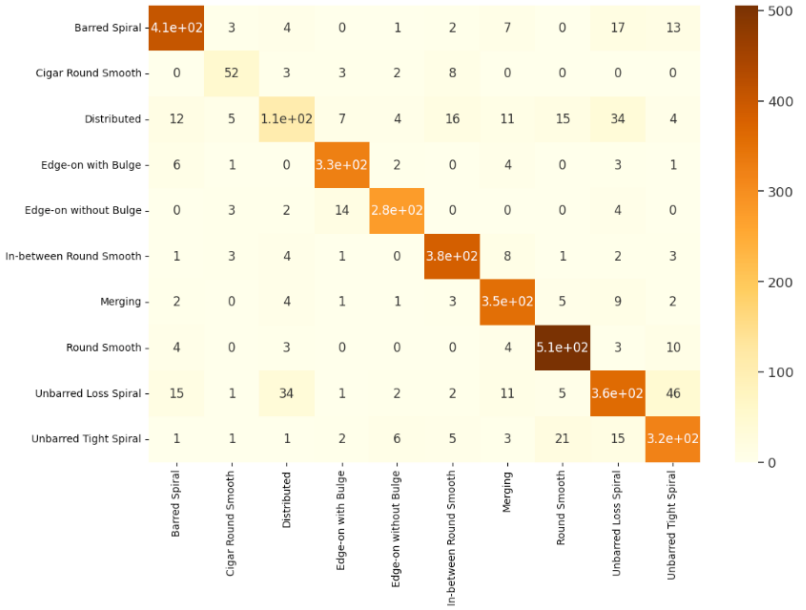


Figure 5. Confusion matrix.

DOI: 10.54254/2755-2721/33/20230245

3.2. Comparison with other models

Several commonly employed models, namely ViT, the second version of 50-layer Deep Residual Networks (ResNet50V2), the lightest version of Dense Convolutional Network (DenseNet121), the second generation of the Small version of EfficientNets (EfficientNetV2S) with Self-Attention on top, and three generations of MobileNets (MobileNet, MobileNetV2, MobileNetV3-Large), are trained and evaluated for the identical task, utilizing similar architectural designs and implementation configurations as previously described. Table 3 provides a comprehensive overview of these models, highlighting their respective parameter counts, test accuracy achieved after 20 training epochs, and inference speed. Notably, MobileViT, the model employed in this study, exhibited significantly superior accuracy compared to other lightweight models, while maintaining a commendable inference speed. In contrast, when compared to heavyweight models, EfficientNet demonstrates the highest accuracy. However, the inference speed of EfficientNet is considerably slower than that of MobileViT. Consequently, MobileViT remains the most optimal solution for this particular task, striking a balance between extraordinary accuracy and efficient inference performance.

Table 3. Comparison of model performance and efficiency in galaxy classification.

Backbone Model	# Parameters	Test Accuracy (%)	Inference speed
ViT	85.8M	66.43	765ms/step
ResNet50V2	24.6M	62.26	59ms/step
DenseNet121	7.5M	64.49	63ms/step
EfficientNetV2-S +Self-Attention	21.0M	**88.16**	1s/step
MobileNet	3.7M	81.09	**29ms/step**
MobileNetV2	**2.9M**	79.74	35ms/step
MobileNetV3-Large	3.5M	14.18	34ms/step
MobileViT-S(Mine)	4.9M	87.12	45ms/step

4. Conclusion

This study introduces a transfer learning approach using MobileViT as a backbone network for the classification of galaxies based on morphology. To expedite the model's convergence speed, pre-trained weights from ImageNet are employed. Additionally, data augmentation techniques and various regularization methods are utilized to mitigate overfitting, while a decayed learning rate strategy is implemented to achieve optimal weight configurations. The proposed method is extensively evaluated through a series of experiments. The results demonstrate that, following fine-tuning on the galaxy dataset, the proposed model strikes a desirable balance between accuracy and inference speed, surpassing other efficient or lightweight models as well as some well-known approaches. In the future, further enhancements to the model architecture will be pursued as the primary research objective. Specifically, attention will be given to replacing and deploying specific network modules to optimize compatibility with mobile device hardware.

References

[1] Naim A Lahav O Sodre Jr L Storrie-Lombardi M C 1995 Automated morphological classification of APM galaxies by supervised artificial neural networks Monthly Notices of the Royal Astronomical Society 275(3): pp 567-590

DOI: 10.54254/2755-2721/33/20230245

[2] Owens E A Griffiths R E Ratnatunga K U 1996 Using oblique decision trees for the morphological classification of galaxies Monthly Notices of the Royal Astronomical Society 281(1): pp 153-157
[3] Dieleman S Willett K W Dambre J 2015 Rotation-invariant convolutional neural networks for galaxy morphology prediction Monthly notices of the royal astronomical society 450(2): pp 1441-1459
[4] Kim E J Brunner R J 2016 Star-galaxy classification using deep convolutional neural networks Monthly Notices of the Royal Astronomical Society 464(4): pp 4463–4475
[5] Zhu X P Dai J M Bian C J Chen Y Chen S Hu C 2019 Galaxy morphology classification with deep convolutional neural networks Astrophysics and Space Science 364: pp 1-15
[6] Lin J Y Y Liao S M Huang H J Kuo W T Ou O H M 2021 Galaxy Morphological Classification with Efficient Vision Transformer arXiv:2110.01024
[7] Mehta S Rastegari M 2021 Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer arXiv:2110.02178
[8] Leung H W Bovy J 2019 Deep learning of multi-element abundances from high-resolution spectroscopic data Monthly Notices of the Royal Astronomical Society 483(3): pp 3255-3277
[9] Willett K W Lintott C J Bamford S P et al 2013 Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey Monthly Notices of the Royal Astronomical Society 435(4): pp 2835-2860
[10] Loshchilov I Hutter F 2017 Decoupled weight decay regularization arXiv:1711.05101