A MobileNetV2 model of transfer learning is employed for remote sensing image classification

Leyu Cao

School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China.

21211234@bjtu.edu.cn

Abstract. Remote sensing image classification is an important and complicated problem in deep learning field. In order to achieve better classification effect, predecessors have proposed different kinds of models. A transfer learning-based MobileNetV2 model is suggested in this paper. First, we take UC-Merced dataset as input and introduce lightweight network MobileNetV2 for scene identification. Secondly, we combined the transfer learning training method and pre-trained the MobileNetV2 model to realize the high-performance classification of deeper remote sensing images. The UC- Merced dataset was classified with a model that achieved 91.43% accuracy,91.00% kappa index, and 91.50% F1-score. These results demonstrate the model's impressive performance in remote sensing image classification and its potential for practical scene identification applications.

Keywords: Scene classification; Convolutional neural network; MobileNetV2; Transfer learning.

1. Introduction

With the continuous development of remote sensing technology, high resolution remote sensing images can be obtained through advanced remote sensing technology and equipment [1], so that more information can be obtained for scene classification. However, the acquisition of remote sensing image cannot touch the object itself, which makes it difficult to obtain the same object information through remote sensing image [2]. The feature design of classification methods based on artificial feature description depends on relevant professional knowledge and experience, and the descriptive ability of these features is very limited in the face of complex images. The classification method of machine learning belongs to shallow learning network, it is difficult to establish complex function representation, and can not adapt to the classification of complex remote sensing images.

Deep learning is inspired by how neural networks work in the human brain. With large amounts of training data and multi-level neural network models, deep learning can learn more efficient feature representations, thereby improving the accuracy of classification tasks. In recent years, deep learning has made remarkable achievements in the field of image classification and is widely used in many fields. Convolutional neural Network (CNN) is an important branch of deep learning. At present, CNN has been widely used in agriculture [3, 4], Medical image [5, 6], Automatic driving [7, 8], fault diagnosis [9, 10] and other fields. At present, many researches have applied CNN to scene identification. Rakshitha et.al[11] use CNNS and artificial neural networks to classify recordings based on the scene or environment in which they are recorded. angyang Li et.al[12] use a deep feature fusion model for remote sensing scene classification. Xiaobin Li et.al[13] use the decision level fusion of features extracted by convolutional neural networks for remote sensing scene classification. However, the above research has such problems as: (1) the number of parameters is large; (2) When the amount of data is limited, it is often insufficient to adequately train CNN.

In order to solve the above problems, this study used MobileNetV2 neural network and introduced transfer learning training method to pre-train the model, and verified it on UC-Merced dataset. Finally, the experiment obtained satisfactory results. The continuous improvement and development of these research results and technical methods will help to further improve the accuracy and efficiency of remote sensing image classification.

The remaining structure of this paper is as follows: The second section is the theoretical part, which mainly introduces the algorithm flow of this paper, as well as the model and method used. The third

chapter deals with the datasets, hyperparameters and the obtained data results. The fourth chapter is the summary of this paper and the prospect of the future.

2. Methods

2.1 Framework

The dataset used in this paper is UC-Merced Land Use dataset. By using mobilenet network for image recognition and introducing transfer learning to extract network features, UC-Merced Land Use 21 labels are finally obtained.

2.2 MobileNetV2

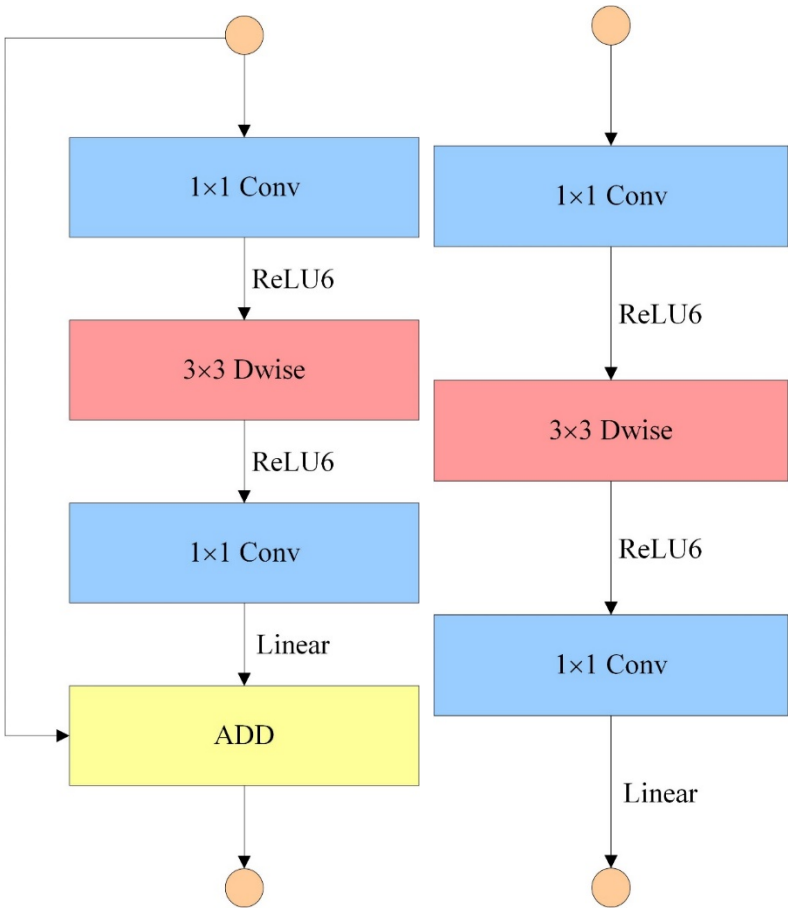

Figure 1. MobileNetV2 architecture block diagram.(a) stride=1 block.(b) stride=2 block.

To meet the needs of mobile AI applications, Google has introduced MobileNetV1, a lightweight convolutional neural network (CNN) for mobile devices and embedded systems.In spite of this, MobileNetV1's performance and computational efficiency could be enhanced. So, Google introduced MobileNetV2, which improves on MobileNetV1.

Compared with MobileNetV1, MobileNetV2 introduces Pointwise Convolution before Depthwise Separable Convolution to dynamically adjust feature channels.The second point-by-point convolution's activation function was eliminated, while MobileNetV2, taking inspiration from ResNet mode, implemented a short circuit connection to decrease the parameters and calculations, thus enhancing the model's performance.

As shown in Figure 1, MobileNetV2 adopts the inverse residual structure, first raising dimension, convolution, and then reducing dimension. Use Shortcut connection when the input and output shapes are the same and the step size is 1. Its deep separable convolutional network structure mainly includes

Pointwise Convolution (PW) and Depthwise Convolution (DW) to reduce the computation and parameter number.

Overall, MobileNetV2 uses a lightweight neural network architecture with lower model complexity and computational effort, uses residual connections to make the network more stable, converges more easily during training, has higher accuracy and faster running speed, while maintaining a small model size and memory footprint, and is suitable for a variety of hardware platforms. Both mobile and embedded devices.

2.3 Transfer Learning

Transfer learning is a machine learning method that aims to use knowledge learned on one task to improve learning on another related task [14]. Transfer learning involves the utilization of knowledge or characteristics acquired in a single domain (the source domain) to tackle a problem in a related yet distinct domain (the target domain).

Transfer learning's fundamental concept is that there may be a likeness or association between the origin domain and the goal domain, and by exploiting the source domain's knowledge to assist the learning mission of the target domain.When there is scant data in the target domain or it is hard to acquire a large quantity of labeled data, this technique is usually employed. By transferring the knowledge from the source domain, the model's generalization capacity and performance in the target domain can be augmented.

We will create the base model based on the MobileNet V2 model developed by Google. This model has been pre-trained on the ImageNet dataset, a large dataset of 1.4 million images and 1,000 classes. ImageNet is a research training data set. With transfer learning, the original features learned by the pre-trained network can be utilized. Despite some differences between the pre-trained images and the target classification images, the convolutional neural network comprehensively trained on the large-scale high-quality ImageNet dataset can still migrate successfully, thereby improving the recognition of the UC Merced Land Use dataset.

Since the pre-trained network is used to classify 1,000 kinds of images, its neurons in the fully connected layer are 1,000. In order to make the network meet the requirements of UC Merced Land Use feature image recognition, the output neurons in the fully connected layer are changed to 21.

3. Experiment

3.1 Dataset

In 2010, the University of California, Merced, released the UC-Merced dataset, which is a classic remote sensing image scene classification dataset[15]. This RGB image collection encompasses 21 distinct feature types, such as agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis court. A total of 2100 images are encompassed in the dataset, with 100 per class. The images have a spatial resolution of 0.3m, and each image has a resolution of 256x256 pixels. The inclusion of various spatial patterns, as well as the presence of highly overlapping scenarios, such as densely distributed residential areas, medium residential areas, and sparsely distributed residential areas, makes the dataset richer and more challenging. As a benchmark for image classification algorithms, this dataset is regularly employed to evaluate and contrast the efficacy of various algorithms. Figure 2 shows a partial sample of the UC Merced dataset.

Figure 2. Sample image of the UC-Merced dataset.

3.2 Hyperparameters

The operating system of the experiment was Windows 11, the memory was 16GB, Intel(R) Iris (R) Xe Graphics GPU was installed, PyCharm Community Edition 2022.3.2 programming platform was adopted. Programming uses Python3.10.10. The batchsize is 16, the epoch is 60, the Adam optimizer is used, the cross-entropy loss function is introduced, and the learning rate is 0.00001. Randomly selecting 80% of the UC-Merced dataset as the training set, the remaining 20% were then used as the test set.

3.3 Evaluation Metrics

In order to study the image recognition ability of neural network more directly, Accuracy, F1-score and Kappa are introduced to characterize the classification ability of the system. The formula is as follows:

$$P = \frac{T_p}{T_p + F_p} \tag{1}$$

$$R = \frac{T_p}{T_p + F_N} \tag{2}$$

$$F1 = \frac{2PR}{P + R} \tag{3}$$

$$ACC = \frac{T_p + T_N}{T_p + F_p + F_N + T_N} \tag{4}$$

$$kappa = \frac{p_0 - p_e}{1 - p_e} \tag{5}$$

where T_p indicates that the model determines the positive class as a positive class. F_p indicates that the model determines negative class as positive class. F_N indicates that the model determines the positive class as the negative class. T_N indicates that the model determines a negative class as a negative class. P denotes the accuracy rate. R denotes the recall rate. $F1$ represents the harmonic average of accuracy rate and recall rate. ACC refers to the proportion of correctly classified samples in the total samples, that is, model accuracy. kappa is used to test consistency. p_0 denotes the overall classification accuracy, and p_e represents the accidental consistency of classification tasks, and the formula is as follows:

$$p_0 = \frac{\sum_{k=1}^{K} A_k \cap B_k}{N} \tag{6}$$

$$p_e = \frac{\sum_{k=1}^{K} A_k * B_k}{N^2} \tag{7}$$

where A_k denotes the number of real labels of class k. B_k represents the number of predicted labels of class k. K denotes the number of total categories. N represents the number of total samples.

3.4 Effectiveness of Transfer Learning

Comparing the MobileNetV2 network with or without transfer learning, as shown in Table 1, the accuracy of the network without pre-training is only 25.71%, while the accuracy of the network after pre-training is 91.43%. The model's generalization capacity is significantly enhanced by transfer learning, and the precision of remote sensing recognition is greatly augmented.

Table 1. Comparison of the performance of the MobileNetV2 model using transfer learning.

Model name	Accuracy (%)	F1-score(%)	Kappa(%)	Params	FLOPs
MobileNetV2(Pretain = False)	25.71	23.24	22.00	2.2508M	426.0928M
MobileNetV2(Pretain = True)	91.43	91.50	91.00	2.2508M	426.0928M

3.5 Comparison with Classification CNNs

The performance of different models after using transfer learning is compared, as shown in Table 2. In terms of model accuracy, MobileNetV2 has the highest accuracy. It was 20.24% higher than ShuffleNetV2, the model with the lowest accuracy. Compared with the volume of the model, the convolutional parameter number Params and the complexity FLOPs of the model are smaller than ResNet18, VGG16 and AlexNet, which is a lightweight neural network.

Taking all the above parameters into account, MobileNetV2 has the highest accuracy, the number of convolutional parameters and the complexity of the model are only reduced by 0.9757M and 227.9524M respectively compared with ShuffleNetV2, but the accuracy is greatly improved compared with the latter. ShuffleNetV2 under transfer learning ensures network performance. The lightweight network is realized, the computation is reduced, and it is more suitable for satellite remote sensing image recognition.

Table 2. Performance comparison of different neural network models under transfer learning.

Model name	Accuracy(%)	F1-score(%)	Kappa (%)	Params	FLOPs
ResNet18[16]	89.52	89.71	89.00	11.1873M	2.3818G
ShuffleNetV2[17]	71.19	69.31	69.75	1.2751M	198.1404M
VGG16[18]	88.10	88.19	87.50	134.3466M	20.1643G
AlexNet[19]	81.19	81.21	80.25	57.0899M	923.1359M
MobileNetV2[20]	91.43	91.50	91.00	2.2508M	426.0928M

4. Conclusions

In order to realize high-resolution remote sensing image classification model, MobileNetV2 model based on transfer learning is used to classify UC-Merced dataset. After comparing various models, the model we used has achieved the best classification effect. The final result shows that the accuracy rate of the MobileNetV2 model using transfer learning reaches 91.43%. At the same time, lightweight classification is realized.

In this paper, only UC-Merced dataset was tested. In the future, we will expand the scope of testing, further optimize and improve our model, and enhance the applicability of the model.

References

[1] M. Abolghasemi Najafabadi and I. Kazemi, "Systemic design of the very-high-resolution imaging payload of an optical remote sensing satellite for launch into the VLEO using an small launch vehicle," Heliyon, p. e27404, 2024/03/06/ 2024.

[2] S. Ai, A. S. Voundi Koe, and T. Huang, "Adversarial perturbation in remote sensing image recognition," Applied Soft Computing, vol. 105, p. 107252, 2021/07/01/ 2021.

[3] T. Akilan and K. M. Baalamurugan, "Automated weather forecasting and field monitoring using GRU-CNN model along with IoT to support precision agriculture," Expert Systems with Applications, vol. 249, p. 123468, 2024/09/01/ 2024.

[4] J. Sangeetha and P. Govindarajan, "Prediction of agricultural waste compost maturity using fast regions with convolutional neural network(R-CNN)," Materials Today: Proceedings, 2023/01/20/ 2023.

[5] S. Liu, L. Wang, and W. Yue, "An efficient medical image classification network based on multi-branch CNN, token grouping Transformer and mixer MLP," Applied Soft Computing, vol. 153, p. 111323, 2024/03/01/ 2024.

[6] Y. Ao, W. Shi, B. Ji, Y. Miao, W. He, and Z. Jiang, "MS-TCNet: An effective Transformer–CNN combined network using multi-scale feature learning for 3D medical image segmentation," Computers in Biology and Medicine, vol. 170, p. 108057, 2024/03/01/ 2024.

[7] M. Usman, M. Zaka-Ud-Din, and Q. Ling, "Enhanced encoder–decoder architecture for visual perception multitasking of autonomous driving," Expert Systems with Applications, vol. 246, p. 123249, 2024/07/15/ 2024.

[8] R. Shi et al., "CNN-Transformer for visual-tactile fusion applied in road recognition of autonomous vehicles," Pattern Recognition Letters, vol. 166, pp. 200-208, 2023/02/01/ 2023.

[9] F. Dao, Y. Zeng, and J. Qian, "Fault diagnosis of hydro-turbine via the incorporation of bayesian algorithm optimized CNN-LSTM neural network," Energy, vol. 290, p. 130326, 2024/03/01/ 2024.

[10] F. Yang, X. Tian, L. Ma, and X. Shi, "An optimized variational mode decomposition and symmetrized dot pattern image characteristic information fusion-Based enhanced CNN ball screw vibration intelligent fault diagnosis approach," Measurement, p. 114382, 2024/02/27/ 2024.

[11] Rakshitha, K. Karthik, A. D. Shetty, P. Sowmya, and R. Shettigar, "Performance Analysis of Acoustic Scene Classification Using ANN and CNN Techniques," in 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS), 2023, pp. 1-5.

[12] Y. Li, Q. Wang, X. Liang, and L. Jiao, "A Novel Deep Feature Fusion Network For Remote Sensing Scene Classification," in IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019, pp. 5484-5487.

[13] X. Li, B. Jiang, T. Sun, and S. Wang, "Remote Sensing Scene Classification Based on Decision-Level Fusion," in 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), 2018, pp. 393-397.

[14] L. P. Silvestrin, H. van Zanten, M. Hoogendoorn, and G. Koole, "Transfer learning across datasets with different input dimensions: An algorithm and analysis for the linear regression case," Journal of Computational Mathematics and Data Science, vol. 9, p. 100086, 2023/12/01/ 2023.

[15] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in ACM SIGSPATIAL International Workshop on Advances in Geographic Information Systems, 2010.

[16] K. He, X. Zhang, S. Ren, J. J. I. C. o. C. V. Sun, and P. Recognition, "Deep Residual Learning for Image Recognition," pp. 770-778, 2015.

[17] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 6848-6856.

[18] K. Simonyan and A. J. C. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," vol. abs/1409.1556, 2014.

[19] A. Krizhevsky, I. Sutskever, and G. E. J. C. o. t. A. Hinton, "ImageNet classification with deep convolutional neural networks," vol. 60, pp. 84 - 90, 2012.

[20] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," vol. abs/1704.04861, 2017.