Remote Sensing Scene Classification Based on Improved GhostNet

Biyun Wei¹²*, Xiaole Shen¹³ and Yule Yuan¹²

¹ College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China
² College of Applied Technology, Shenzhen University, Shenzhen, China
³ College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
Email: weibiyun2018@email.szu.edu.cn

Abstract. Nowadays, the design of convolutional neural network (CNN) models is getting deeper and wider. When traditional CNN is used to process limited data of remote sensing images, it will lead to overfitting. We will use lightweight and efficient models to classify remote sensing images. In order to improve the classification accuracy and reduce the intermediate parameters, we improved GhostNet and proposed a smaller CNN named Improved GhostNet. Meanwhile, we use image enhancement methods to enlarge the datasets and dropout, it will reduce the amount of parameters. We experimented on three datasets, such as AID, UC Merced, NWPU-RESISC45. Then, we used MobileNetV3-Small and GhostNet to compare with our CNN model. The classification accuracy of improved GhostNet achieves more than 91%, and the accuracy on the AID is improved by 2.05% compared to the original GhostNet. These results demonstrate the effectiveness and efficiency of improved GhostNet.

1. Introduction
With the rapid development of earth observation technology, the amount of remote sensing images has increased significantly. The valuable information contained in a large amount of accumulated remote sensing images urgently needs to be fully excavated and utilized. Recognition and classification of complex scenes of remote sensing images is one of the important contents for extracting and analyzing this information. It can be widely used in land use, global environmental pollution detection and military target detection, etc. It has important theoretical significance and practical value.

At present, convolutional neural network (CNN) has become a key research direction in image recognition, object detection, and semantic segmentation. Traditional image classification CNNs include AlexNet [1], VGGNet [2], ResNet [3], DenseNet [4], etc. Those networks all perform very well in image classification, however, they produce a lot of parameters and floating point operations (FLOPs) during the training process. Thus, lightweight and more efficient network has become a research focus. Typical lightweight networks, such as SqueezeNet, ShuffleNet, MobileNetV1, MobileNetV2, MobileNetV3, GhostNet [5-10], can not only obtain better accuracy, but also significantly reduce the forward inference time of the network.

In this paper, we proposed a lightweight network named improved GhostNet, which is an improvement based on GhostNet. Improved GhostNet is lighter than GhostNet and more suitable for training small sample datasets. Experimental results show that improved GhostNet has greatly improved the accuracy compared to the original GhostNet.
2. Datasets

2.1. Datasets Collection
The datasets used in this paper include three datasets, UC Merced, NWPU-RESISC45 [11], AID. We use three datasets to verify the generalization ability of Improved GhostNet on small remote sensing datasets.

UC Merced was released by the UC Merced Computer Vision Lab in 2010. The dataset is divided into 21 categories. Most categories in UC Merced overlap with categories in NWPU-RESISC45. They have the same size, both are 256*256.

NWPU-RESISC45 is a large benchmark dataset provided by Northwestern Polytechnical University. It is the most frequently used remote sensing image dataset for verifying the quality of the network. The dataset contains 45 categories of remote sensing image data, each with 700 images. These 45 categories are shown in figure 1.

AID is a remote sensing image dataset released by Wuhan University and Huazhong University of Science and Technology in 2017. The dataset contains 30 categories of data, and the size of each image is 600*600. The number of samples of each category of data in the dataset is inconsistent, which will cause the network to have uneven weight distribution in training. However, this part is not the focus of this paper. For this dataset, we only focus on the average classification accuracy using different network models.

2.2. Datasets Preprocessing
Because of the relatively small number of samples in each dataset, the network model will overfit probably. We use the image enhancement methods provided by the Keras to expand the image data, including image flip, image rotation, image contrast adjustment, image zoom, etc. The image enhancement methods can not only increase the amount of images, but also improve the generalization ability of the network. As we know, the classified scene information will be affected by weather, environment and other factors. The scene information can be enriched through image enhancement to adapt to the scene in different environments.

We take the intersection category in NWPU-RESISC45 as an example, the images after image enhancement are shown in figure 2.

3. CNN for Remote Sensing Image Classification
In order to deploy CNN on embedded platforms, we use a lightweight and efficient network to train and predict remote sensing images. Compared with the traditional deep neural network, the lightweight network model has fewer parameters and lower FLOPs. It can satisfy the requirements of
low storage and low energy consumption of embedded devices.

3.1. **Ghost Module**
GhostNet uses linear operation instead of partial convolution, which reduces the amount of inference calculation. Less computation is good for reducing the energy consumption and deploying the neural network on embedded devices. This module combining convolution operation and linear operation is called Ghost Module. Ghost Module is shown in figure 3. In the figure, \( \mathbf{Y} \) represents a non-redundant feature map generated by convolution, and \( \mathbf{Y}' \) represents a redundant feature map of \( \mathbf{Y} \) generated by linear operation.

![Figure 3. Ghost Module.](image)

3.2. **Bottleneck**
GhostNet contains two structures of bottleneck, with the stride=1 for feature extraction and the stride=2 for reducing the number of channels. Both types of bottleneck are formed by stacking two Ghost Modules. Figure 4 shows the structure of bottlenecks.

![Figure 4. Ghost Bottleneck.](image)

However, experiments show that using GhostNet to predict the three remote sensing image datasets is not ideal. The network structure is not suitable for direct training. We believe that the overfitting is because the dataset is too small.

3.3. **Improved GhostNet**
We adjust the network model according to the characteristics of the datasets, reduces the number of GhostNet layers, use dropout to reduce the amount of parameters, and adjusts the size of the filter in the network. It is helpful to reduce the network parameters and avoid overfitting in the network. We increase the size of the filter, so the reduction in the number of layers will have a certain effect on the extracted features, and more feature information can be extracted spatially.

The improved network structure is shown in table 1.
Table 1. Entire network structure of improved GhostNet.

| Input       | Operator | #exp | #out | SE  | Stride |
|-------------|----------|------|------|-----|--------|
| 224*224*3   | Conv2d 3*3 | -    | 16   | -   | 2      |
| 112*112*16  | GBnecka 3*3 | 16   | 16   | 1   | 2      |
| 56*56*16    | GBneck 3*3 | 72   | 24   | -   | 2      |
| 28*28*24    | GBneck 5*5 | 88   | 24   | -   | 1      |
| 28*28*24    | GBneck 5*5 | 96   | 40   | 1   | 2      |
| 14*14*41    | GBneck 5*5 | 240  | 40   | 1   | 1      |
| 14*14*40    | GBneck 5*5 | 120  | 48   | 1   | 1      |
| 14*14*48    | GBneck 5*5 | 144  | 48   | 1   | 1      |
| 14*14*48    | GBneck 5*5 | 288  | 96   | 1   | 2      |
| 7*7*96      | GBneck 5*5 | 576  | 96   | 1   | 1      |
| 7*7*96      | GBneck 5*5 | 576  | 96   | 1   | 1      |
| 7*7*96      | GBneck 5*5 | -    | 576  | 1   | 1      |
| 7*7*576     | GAP 7*7   | -    | -    | -   | -      |
| 1*576       | Conv2d 1*1 | -    | 1024 | -   | 1      |
| 1*1024      | Conv2d 1*1 | -    | Kd   | -   | -      |

a GBneck denotes Bottleneck. b #exp denotes expansion size. c SE indicates whether to use the Squeeze-And-Excite module. d K denotes the number of categories.

4. Experiments

4.1. Training Strategy

Due to the relatively small dataset, we train networks from scratch. Keras is used in the experiments as the deep learning framework, and the experimental system environment is ubuntu18.04. The experimental data are the three datasets mentioned in 2.1. We use three datasets to verify the generalization ability of improved GhostNet.

In this paper, the dataset is shuffled and divided into three parts, the training set, the verification set and the test set, and their ratio is 6:2:2. We expand the dataset by means of data enhancement. This is to prevent overfitting in network training, and at the same time improve the generalization ability of the network. The image enhancement methods used in this paper include image flip, image zoom, image rotation, etc. The experiments use Adam’s method to train the network, and the initial learning rate is 0.01. The training process is shown in figure 5.

![Figure 5. Training process from scratch.](image-url)
4.2. Comparison of the Experimental Results
In our experiments, we use MobileNetV3-Small and GhostNet for comparison, and use three datasets to verify the effectiveness and efficiency of improved GhostNet. The measurement standard of the experimental results includes three parts, which are the average accuracy of the test set prediction, the time required for the GPU to process an image, and the weight size. Considering that the relative inference time of each model is constant, we will test the processing time of each model on the GPU.

| CNN Model         | Dataset         | Average Accuracy (%) | GPU Processing Time (ms) | Weight (M) |
|-------------------|-----------------|----------------------|--------------------------|------------|
| MobileNetV3-Small | AID             | 92.45                | 21                       | 12.7       |
| GhostNet          | AID             | 92.75                | 25                       | 16.4       |
| Improved GhostNet | AID             | 94.8                 | 21                       | 5.6        |
| MobileNetV3-Small | UC Merced       | 92.38                | 22                       | 12.7       |
| GhostNet          | UC Merced       | 92.86                | 24                       | 16.4       |
| Improved GhostNet | UC Merced       | 93.1                 | 22                       | 5.6        |
| MobileNetV3-Small | NWPU-RESISC45   | 91.37                | 17                       | 12.7       |
| GhostNet          | NWPU-RESISC45   | 90.19                | 20                       | 16.4       |
| Improved GhostNet | NWPU-RESISC45   | 91.73                | 18                       | 5.6        |

4.2.1. Network Generalization Ability. In order to verify the generalization ability of Improved GhostNet on small sample remote sensing image datasets, we conduct experiments on the three general datasets. According to table 2, the accuracy of improved GhostNet on all three datasets achieves more than 91%. The results show that improved GhostNet is not only feasible on the classification of remote sensing images, but also can handle small sample data classification.

4.2.2. High Precision with Low Time Consumption. In our experiments, two classic lightweight convolutional neural networks, MobileNetV3-Small and GhostNet, are used to classify the three datasets. By comparing their results with the improved GhostNet classification results, it can be seen that improved GhostNet is more suitable for small sample remote sensing image dataset. On the AID dataset, the average classification accuracy of improved GhostNet achieves 94.8%. Compared with MobileNetV3-Small and GhostNet, the accuracy of improved GhostNet increases by 2.35% and 2.05%, respectively. As for UC Merced dataset, improved GhostNet raises accuracy by 0.72% and 0.24% compared with MobileNetV3-Small and GhostNet, respectively. The experimental results on NWPU-RESISC45 dataset show that Improved GhostNet raises accuracy by 0.36% and 1.54% compared with MobileNetV3-Small and GhostNet, respectively. In the aspect of GPU processing time, we can see that improved GhostNet’s inference speed is not reduced compared with the other two CNN models. In contrast, the inference speed is faster on NWPU-RESISC45 dataset.

4.2.3. Low Weight. In terms of weights, we can see the results shown in table 2 that Improved GhostNet has lower weight. The weight of Improved GhostNet is less than half that of the other two models. Since improved GhostNet is designed for embedded devices, lower weight is extremely necessary. Lower weight ensures that embedded devices have enough storage to build the CNN model.

5. Conclusion
To deploy CNN on embedded devices for remote sensing image classification, we proposed an improved GhostNet model. This paper verified the feasibility and generalization ability of improved GhostNet via experiments on different datasets. And by comparing with the classic lightweight and
efficient networks MobileNetV3-small and GhostNet, it is found that the accuracy of improved GhostNet is significantly improved while ensuring the same inference speed. This paper proved the effectiveness and efficiency of improved GhostNet on small sample remote image datasets.

References
[1] Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks Advances in Neural Information Processing Systems 1097-1105.
[2] Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition arXiv preprint arXiv:1409.1556.
[3] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp 770-778.
[4] Huang G, Liu Z, Van Der Maaten L and Weinberger K Q 2017 Densely connected convolutional networks Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp 4700-4708.
[5] Iandola F N, Han S, Moskewicz M W, Ashraf K, Dally W J and Keutzer K 2016 SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size arXiv preprint arXiv:1602.07360.
[6] Zhang X, Zhou X, Lin M and Sun J 2018 Shufflenet: An extremely efficient convolutional neural network for mobile devices Proceedings of the IEEE conference on computer vision and pattern recognition pp 6848-6856.
[7] Howard A G, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Adam H, et al. 2017 Mobilenets: Efficient convolutional neural networks for mobile vision applications arXiv preprint arXiv:1704.04861.
[8] Sandler M, Howard A, Zhu M, Zhmoginov A and Chen L C 2018 Mobilenetv2: Inverted residuals and linear bottlenecks Proceedings of the IEEE conference on computer vision and pattern recognition pp 4510-4520.
[9] Howard A, Sandler M, Chu G, Chen L C, Chen B, Tan M, Le Q V, et al. 2019 Searching for mobilenetv3 Proceedings of the IEEE International Conference on Computer Vision pp 1314-1324.
[10] Han K, Wang Y, Tian Q, Guo J, Xu C and Xu C 2020 GhostNet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp 1580-1589.
[11] Cheng G, Han J and Lu X 2017 Remote sensing image scene classification: Benchmark and state of the art Proceedings of the IEEE 105 (10) 1865-1883.