Trip-GhostNet for Hyperspectral Image Classification

Zitong Zhang1a, Xiaoqing Hu2*, Siqing Wang2, Lei Kang2, Qiaoyu Ma2
1School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing 100083, China
2School of Science, China University of Geosciences (Beijing), Beijing 100083, China
a3001200116@cugb.edu.cn *Corresponding author’s e-mail: xiaoqinghu_321@163.com

Abstract. The classification of hyperspectral remote sensing images (HSI) is an important task in the processing and application of hyperspectral images. Convolutional neural network (CNN) has significant advantages in the extraction and fusion of hyperspectral image spectrum and spatial features, so it has become a common method in the field of HSI classification. However, the spectral domain feature redundancy of hyperspectral images is very high, and the spatial domain feature structure is complex, which makes CNN more time-consuming and higher memory requirements in the feature extraction process. Based on this problem, lightweight networks with fewer parameters have gradually become the main application method in the field of HSI classification. In order to reduce the computational complexity of deep feature extraction and take into account the shallow features, a high-order nonlinear Ghost module is proposed on the basis of the original Ghost linear transformation module. Furthermore, in view of the independent characteristics of each dimension of HSI, a Trip-GhostNet is proposed, which simultaneously extracts and fuses features from three dimensions in a lightweight manner. According to the distribution characteristics of HSI, the influence of attention embedding methods in the high-order Ghost module of each branch on the feature extraction is compared and analyzed. The results show that the proposed model can reduce model calculations and improve classification accuracy, and is suitable for HSI classification problems.

1. Introduction
HSI have the characteristics of multiple bands, wide spectral band coverage, and rich information. Hyperspectral imaging is a technology that uses hyperspectral remote sensing sensors to obtain hyperspectral images by collecting reflected or radiated electromagnetic wave signals from different ground objects. The wavelength range and spatial resolution of different sensors are different. A wider wavelength range and higher spatial resolution can often reflect the object information in more detail. However, as the spectral bands and the spatial resolution increases, the data volume of HSI also increases exponentially. Although high-resolution remote sensing images measure the earth's surface more accurately, a large number of richly structured images also bring many new problems to the intelligent processing of HSI. In the context of the ‘big data’ era, the development of a fast or even real-time remote sensing image processing system that can greatly improve work efficiency is now attracting a considerable number of people's attention. Among them, the high complexity of spatial structure patterns in large-scale HSI makes the understanding and classification of intelligent features more challenging[1].

In recent years, in order to make the model more efficient, innovative lightweight network models have emerged one after another. On the one hand, for the purpose of feature extraction, how to meet the
needs of lightweight. On the other hand, in terms of important feature extraction, the attention mechanism has also been introduced into a variety of feature extraction modules[2]. The goal of the attention mechanism is to increase the focus on important features and weaken the focus on unnecessary features.

In the field of HSI classification, data cubes with image spatial and channel information is often used as the input for the classification of ground objects. CNN has the characteristics of image feature extraction in the spatial domain, feature shuffling and fusion in the channel domain, so that it can perform both spatial and spectral domain feature extraction in the HSI classification[3]. In 2020, GhostNet[4] proposed by Han et al. innovatively replaced part of the convolution operation with a less computationally intensive affine transformation, greatly reducing the amount of model calculations, and at the same time, the embedding of attention can make the model converge faster. On the basis of GhostNet, we proposed a high-order Ghost block. And combined with the spatial-spectrum features of HSI, we proposed a Trip-GhostNet, which pays attention to the important features of the spectral and spatial domains at the same time. This method performs better on the classification of three HSI datasets, with an accuracy increase of up to 1.83%, which provides a more suitable feature extraction idea for HSI.

2. Research area and datasets
Three HSI datasets, Indian Pines[5], Tea Farm[6] and Salinas[7] are used in the experiment to study the application effect of the models on these three datasets with different resolutions. The related pixel, band, category and spatial resolution information of the datasets are shown in Table 1. During model learning, 3% sampling of Indian Pines and 1% sampling of Tea Farm and Salinas are used as the training sets, and Principal Component Analysis (PCA) is used to reduce dimensionality on the channel, and the first 8 principal components are taken as input.

| Dataset       | Year | Pixel number | Bands number | Spatial resolution (m) | Category number |
|---------------|------|--------------|--------------|------------------------|-----------------|
| Indian Pines  | 1992 | 145×145      | 220          | 22.19                  | 16              |
| Tea Farm      | 2017 | 512×348      | 80           | 2.25                   | 10              |
| Salinas       | 1998 | 512×217      | 224          | 3.7                    | 16              |

Table 1. The information of different HSI datasets.

![Hyperspectral remote sensing images and real land cover map.](image)

Figure 1. Hyperspectral remote sensing images and real land cover map.

3. Methods

3.1. High-order Ghost block
With the increase of convolutional layers, the features extracted by CNN gradually become more abstract [8]. CNN's abstract expression of features is an effective way to extract deep nonlinear features, but it will also cause the loss of shallow linear and nonlinear features. Furthermore, internal detailed features are more important for HSI, but layer-by-layer convolution can easily cause the loss of this part of the shallow features. Therefore, in order to obtain both deep and shallow features, a high-order Ghost block as shown in Figure 2 is designed for feature extraction of HSI.
3.1.1. **High-order linearity and nonlinearity.** After the input image is obtained, feature extraction and linear combination transformation are performed on the input image through the layer-by-layer Ghost module. The cascade of the three-layer Ghost modules can bring a feature extraction method with richer features. Each layer of Ghost module includes linear affine transformation and nonlinear deep-wise convolution. In this way, the input information is retained in a linear transformation, and abstract features are further extracted in a non-linear way. The cascade of the three-layer Ghost modules will obtain 4 different feature extraction results. They are deep convolutional abstraction (Equation 1), deep linear cascading transformation (Equation 2), shallow non-linear abstraction (Equations 3, 5 and 7), and shallow non-linear transformation (Equations 4, 6 and 8).

\[
Y_0 = f_3^{3x3}(f_2^{3x3}(f_1^{3x3}(X_0) + b_{10}) + b_{20}) + b_{30}
\]

(1)

\[
Y_1 = w_3(w_2(w_1X_0 + b_{11}) + b_{21}) + b_{31}
\]

(2)

\[
Y_2 = f_3^{3x3}(f_2^{3x3}(w_1X_0 + b_{11}) + b_{20}) + b_{30}
\]

(3)

\[
Y_3 = w_3(w_2(f_1^{3x3}(X_0) + b_{10}) + b_{21}) + b_{31}
\]

(4)

\[
Y_4 = f_3^{3x3}(w_2(w_1X_0 + b_{11}) + b_{21}) + b_{30}
\]

(5)

\[
Y_5 = w_3(f_2^{3x3}(f_1^{3x3}(X_0) + b_{10}) + b_{20}) + b_{31}
\]

(6)

\[
Y_6 = f_3^{3x3}(w_2(f_1^{3x3}(X_0) + b_{10}) + b_{21}) + b_{30}
\]

(7)

\[
Y_7 = w_3(f_2^{3x3}(w_1X_0 + b_{10}) + b_{20}) + b_{31}
\]

(8)

3.1.2. **Feature reusing.** The high-order linear cascade keeps the input features always in the feature extraction process, so that the model can eventually have both shallow and deep features at the same time. Equation 1 corresponds to the part of the original input features retained in the model. In the classification of HSI, for the ground objects with relatively homogeneous spatial structure, the distinguishing features of the ground objects are often well extracted in the shallow layer. Therefore, by appropriately retaining the features from shallow layers, the loss of the shallow features in the abstraction process can be avoided. This layer-by-layer feature reusing method increases the diversity of feature extraction with the linear computation of the input image. At the same time, an attention mechanism with linear computation can be embedded between layers to guide the learning direction of the model.
3.2. High-order Ghost block

In HSI, features from the same dimension are considered to be highly correlated, and features from different dimensions are considered to be weakly correlated or independent of each other. For the data cube of HSI, it has correlation in spatial neighborhood and spectral correlation between channels. Aiming at the feature independence between the spatial domain and the spectral domain, the three-branch feature extraction is used to extract the features from the spatial and the spectral domain respectively. Specifically, for a data cube with a size of \((H, W, C)\), feature extraction is performed on the \((H, C)\), \((W, C)\) and \((H, W)\) respectively. In the spatial domain of HSI, the spatial features of ground objects have different scales. For example, in Tea Farm datasets, Masson pine and Bamboo forests have larger spatial features than rice. And in the spectral domain, there are also differences in the changing trend of the HSI spectral signal corresponding to different ground objects with the increase of the wavelength. Therefore, according to the distribution characteristics of HSI, attention is paid to the different dimensional branches of the features.

As shown in Figure 3, the third-order Ghost feature extraction is performed independently in the \((H, W)\), \((H, C)\) and \((W, C)\), and spatial attentions are embedded between the latter two Ghost modules. The dimensions of the high-order Ghost feature extraction of each branch are different. When final fusing the features from each branch, the ResNeSt module can be used to add attention to global information fusion for each branch feature, so that the model can pay attention to the difference.

3.3. Spatial Attention Module

Spatial Attention Module (SAM) [9] aims to extract the significant features in the feature maps, focusing more on finding the most attractive target in an image. The global information evaluation of the channels often uses the global maximum or average value of the spatial domain. The maximum value represents the overall significant information of the spatial feature on the current channel, and the average value represents the overall performance. The corresponding implementation process of SAM is shown in Figure 2b, which is similar to the way the channel attention module (CAM) in SENet [10] compresses the spatial information of the feature maps (Figure 2a). The difference is that spatial attention is to perform mean pooling and maximum pooling on the channel information of the feature maps \(F\), to obtain the spatially significant features \(F_{\text{max}}\) and comprehensive soft features \(F_{\text{avg}}\) after channel compression. Then, merge the distribution information of these two spatial feature maps into a two-channel feature map \(F_{\text{total}}\), and use a 3×3 convolution and a sigmoid function \(f^{3×3}\) for nonlinear activation to obtain the SAM matrix after spatial excitation. Therefore, the calculation formula for extracting SAM can be expressed as:

\[
M^s = f_{\text{sigmoid}}(f^{3×3}(F^s_{\text{total}}))
\] (9)
4. Results

4.1. Model design and results
According to the size of the neighborhood window when HSI datasets sampling, a three-branch high-order GhostNet model as shown in Figure 3 is established, and SAM is embedded between the latter two Ghost Modules. For further comparing and studying the effects of different attention positions and methods on the classification of HSI, seven comparison models shown in Table 2 are designed. Among them, the Trip-Ghost (no attention) model refers to a Trip-GhostNet without any attention modules, and Trip-CNN is the model that replaces all ghost modules with convolution. In other attention comparison models, SAM1 refers to embedding SAM before the first Ghost module, SAM2 refers to embedding SAM between the first and second ghost modules, and SAM3 refers to embedding SAM between the second and third ghost modules. The model design and related representations of embedded CAM are similar to SAM.

Table 2. The results of different models

| Model                     | Accuracy | Flops (GB) |
|---------------------------|----------|------------|
|                           | Indian Pines | Tea Farm | Salinas |         |
| Proposed Trip-Ghost       | 84.73    | 97.81     | 98.89    | 30.70   |
| Trip-CNN                  | 82.45    | 97.68     | 98.70    | 32.12   |
| Trip-Ghost (no attention) | 82.16    | 97.04     | 98.83    | 30.70   |
| Trip-Ghost-CAM1           | 77.51    | 96.31     | 98.68    | 30.70   |
| Trip-Ghost-CAM2           | 81.22    | 97.15     | 98.60    | 30.71   |
| Trip-Ghost-CAM3           | 73.87    | 97.01     | 98.27    | 30.72   |
| Trip-Ghost-SAM1           | 79.47    | 96.46     | 97.85    | 30.70   |
| Trip-Ghost-SAM2           | 82.90    | 97.40     | 98.58    | 30.70   |

4.2. Lightweight performance
The amount of floating-point calculations and classification accuracy are used as the evaluation indicators of the model’s lightweight performance to compare the performance of these eight classification models. As shown in Figure 5, taking the performance of the CNN model as a comparison benchmark, the proposed new Trip-Ghost model can reduce the amount of calculation while improving the classification accuracy. First of all, according to Table 2 to compare the Trip-Ghost model with no attention and other models with attention, it can be seen that the embedding of attention does not bring more computation. When CAM is embedded in the shallow layer, the amount of calculations brought to the model is less, and it will increase with the deepening of the embedding position. Secondly, comparing the position of the attention embedding, it can be seen that the model performs best when the attention is embedded between the latter two ghost modules. Finally, compared to CAM, SAM
embedding brings less computation and does not increase as the layer goes deeper. Therefore, the proposed model is more suitable for the classification of HSI.

Figure 5. Lightweight performance of models

Based on the above comparative analysis, two reasons for the better performance of Trip-GhostNet are summarized. On the one hand, the Ghost module uses linear combination to replace the convolution to generate new feature maps, which reduces a part of the calculation from the feature map convolution. On the other hand, the Trip-Ghost model achieves high-order nonlinearity with multiple linear cascades, which is helpful for extracting the structural features from different dimensions of hyperspectral objects.

5. Conclusion
Taking the three-dimensional hyperspectral image data cube as the research target, using branch feature extraction, high-order Ghost lightweight extraction is performed on each dimension feature of the cube independently, and ResNeSt is used to fuse the branch features with different attention strengths.

(1) The Ghost block with high-order ghost has two advantages in feature extraction. On the one hand, it retains the compressed information of the original feature maps, and on the other hand, it realizes the extraction and fusion of multi-depth features of hyperspectral images with less computation;

(2) The embedding of the attention module can guide the feature extraction of the high-order ghost block, accelerate the model convergence, increase the focus on important information in the feature extraction process, and be more conducive to the accurate classification of HSI datasets;

(3) The multi-branch architecture can extract structural features from the three dimensions of the HSI data cubes, and the ResNeSt module can fuse the features after the multi-branch abstraction to strengthen the fusion of independent features in different dimension;

(4) Based on the deep feature extraction of the Trip-GhostNet model, the embedding of the attention mechanism should be related to the abstraction degree of the features. For HSI datasets with rich spatial structure features, embedding SAM to abstract features in various dimensions can effectively promote the model to capture important features.

In Trip-Ghost, the branch structure is used to extract features from various dimensions, which brings a more suitable idea to the image classification. In the next step, we will further discuss the application effects of the dimensional feature extraction architecture in other image classification models.

References
[1] Zhao, W., & Du, S. (2016). Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing, 54: 4544–4554.

[2] Mou, L. & Zhu, X. X. (2019). Learning to pay attention on spectral domain: A spectral attention module-based convolutional network for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 58: 110-122.
[3] Mou, L., Ghamisi, P., & Zhu, X. X. (2017). Unsupervised spectral–spatial feature learning via deep residual Conv–Deconv network for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 56: 391-406.

[4] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C. & Xu, C., (2020). Ghostnet: More features from cheap operations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle. pp. 1580-1589.

[5] Baumgardner, M. F., Biehl, L. L., & Landgrebe, D. A. (2015). 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3. Purdue University Research Repository. https://purr.purdue.edu/publications/1947/1.

[6] Zhang, X., Zhang, B., Zhang, L. F., & Sun, Y. L. (2017). Hyperspectral remote sensing dataset for tea farm. Global Change Data Repository. http://www.geodoi.ac.cn/doi.aspx?Id=720.

[7] San Wang. (2020). hyperspectral Dataset. IEEE Dataport. http://dx.doi.org/10.21227/eqk7-wa46.

[8] Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25: 1097-1105.

[9] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision. Munich. pp. 3-19.

[10] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake. pp. 7132-7141.