MMF-Net: Multi-model Fusion Network for Hyperspectral Image Classification Based on Transfer Learning

Yanfen Sun[a], Bian Bawangdui[a]

[a]Tibet University

Abstract

Deep learning has demonstrated outstanding performance in hyperspectral image (HSI) classification. However, its generalization ability is limited by the high cost of annotations and strong cross-scene heterogeneity. To address these challenges, this paper proposes a multi-model fusion network (MMF-Net) based on transfer learning. First, a linear mapping layer is used to convert HSI data into a three-channel representation to mitigate the modality gap between the source domain (ImageNet) and the target domain (HSI). Then, three heterogeneous pre-trained models—Inception, VGG16, and Xception—are fine-tuned, with a channel attention mechanism incorporated to enhance discriminative features. Finally, high-level semantic features from multiple models are fused to construct a joint spectral-spatial-semantic representation, and a lightweight logistic regression classifier is employed for efficient pixel-wise classification.

Experimental results show that the proposed method achieves good classification performance and generalization ability even with limited samples.

Introduction

- Compared to traditional shallow models, deep networks can more effectively utilize high-level nonlinear feature representations.
- However, these algorithms may encounter overfitting issues during training, require large amounts of data, and are time consuming. Data scarcity is a key factor contributing to these challenges.
- Contribution:

1. In the feature extraction stage, three existing CNN pre-trained models were selected and fine-tuned using transfer learning
2. An attention mechanism was introduced to adaptively reweight feature maps.

Methodology

To resolve the issues of limited labeled samples and cross-domain feature adaptation in HSI classification, a Multi-model Fusion Network (MMF-Net) based on transfer learning is proposed, as illustrated in Fig. 1. MMF-Net is structured around four key components: (1) channel adaptation mapping, (2) multi-model feature extraction, (3)attention-weighted fusion (Fig.2), and (4) classifier design.

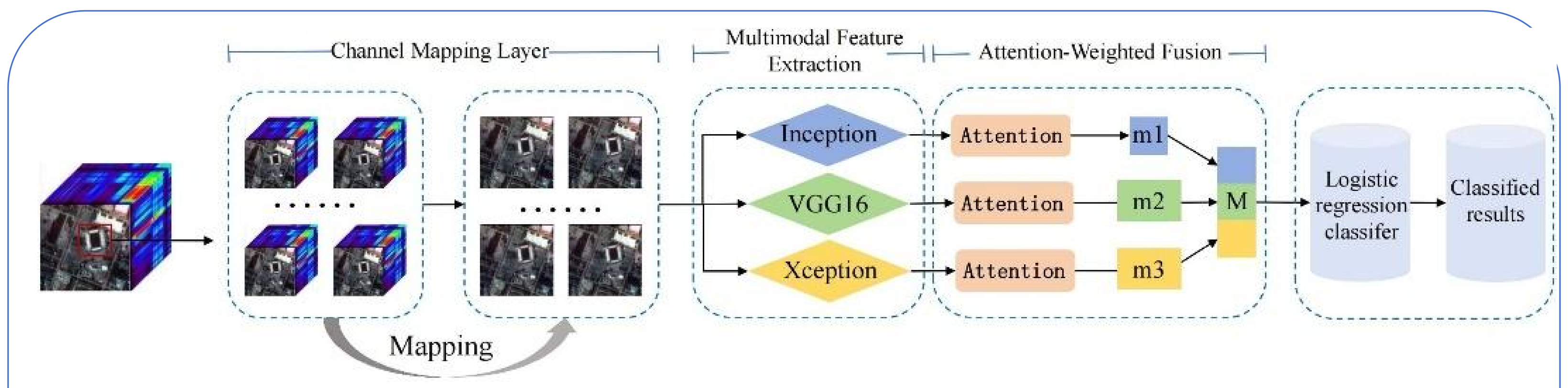

Fig. 1 The proposed MMF-Net model structure.

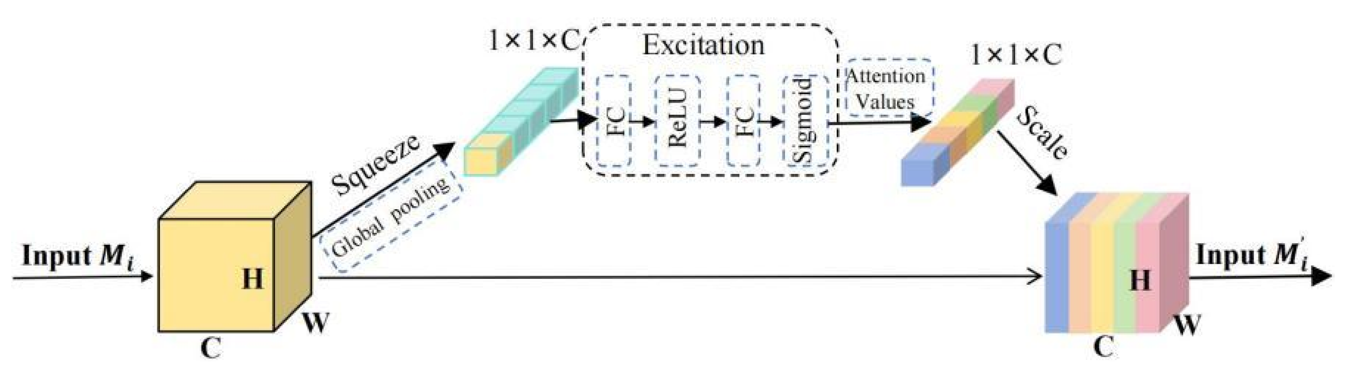

Fig. 2 Channel Attention Block.

Results

Table 1. Classification results on the SA dataset using 10% of the training samples.

Method	EMP-SVM	VGG16	Inception	Xception	T-CNN	MMF-Net
OA(%)	86.95	87.37	78.96	82.75	92.71	95.77
AA(%)	87.56	91.36	79.23	83.54	89.43	93.73
k ×100	85.51	86.15	80.33	84.16	91.61	94.62
Training Time	3.72	29.99	63.49	141.75	8.23	20.44
Test Time	2.30	83.25	88.42	66.96	6.49	85.84

Table 2. Classification results on the KSC dataset using 10% of the training samples.

Method	EMP-SVM	VGG16	Inception	Xception	T-CNN	MMF-Net
OA(%)	92.62	95.91	90.12	93.26	95.41	99.13
AA(%)	88.26	93.74	89.61	90.28	93.30	98.74
k ×100	91.94	94.57	90.02	93.21	94.96	99.21
Training Time	5.57	31.17	115.34	19.91	10.67	25.48
Test Time	1.70	8.09	6.12	8.83	1.94	4.80

Conclusion

To address the challenges of limited labeled samples and cross-domain feature adaptation in HSI classification, this paper proposes a Multi-Model Fusion Network (MMF-Net) based on transfer learning. The network efficiently transfers the ImageNet pre-trained models to the HSI domain through learnable mapping layers, combining Inception, VGG16, and Xception multi-model parallel fine-tuning strategies to extract complementary features. A channel attention mechanism is used to dynamically calibrate cross-domain feature responses, and finally, a cascade fusion is applied to construct a discriminative joint representation.

Experimental results show that MMF-Net achieves significant performance improvements on the Salinas, Pavia, and KSC datasets. Compared to existing methods, MMF-Net demonstrates superior performance, validating its effectiveness and advantages in transfer learning tasks.

References

[1]Zhou Y, Wang M. Remote sensing image classification based on AlexNet network model[C]//Frontier Computing: Theory, Technologies and Applications (FC 2019) 8. Springer Singapore, 2020: 913-918.
[2]Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
[3]Ballester P, Araujo R. On the performance of GoogLeNet and AlexNet applied to sketches[C]//Proceedings of the AAAI conference on artificial intelligence. 2016, 30(1).
[4]He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
[5]Wang W, Li H, Zhao C, et al. Interval estimation of motion intensity variation using the improved inception-V3 model[J]. Ieee Access, 2021, 9: 66017-66031.
[6]Zhou S, Wu H, Xue Z. Grouped subspace linear semantic alignment for hyperspectral image transfer learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-16.
[7]Xia J, Yokoya N, Iwasaki A. Ensemble of transfer component analysis for domain adaptation in hyperspectral remote sensing image classification[C]//2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2017: 4762-4765.
[8]Ye M, Qian Y, Zhou J, et al. Dictionary learning-based feature-level domain adaptation for cross-scene hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(3): 1544-1562.
[9]Shen J, Cao X, Li Y, et al. Feature adaptation and augmentation for cross-scene hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(4): 622-626.
[10]Yang J, Zhao Y Q, Chan J C W. Learning and transferring deep joint spectral–spatial features for hyperspectral classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(8): 4729-4742
[11]Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.
[12]He X, Chen Y, Ghamisi P. Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 58(5): 3246-3263.
[13]Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.
[14]Keras. Cholletf. (2015-03-28) [2021-10-27]. https://github.com/fchollet/keras.
[15]Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
[16]Benediktsson J A, Palmason J A, Sveinsson J R. Classification of hyperspectral data from urban areas based on extended morphological profiles[J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(3): 480-491.
[17]Kang K, Li H, Yan J, et al. T-cnn: Tubelets with convolutional neural networks for object detection from videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 28(10): 2896-2907.