Hyperspectral image classification based on convolutional neural network

Xiwen Zhang

College of Engineering and Applied Science, Stony Brook University, Stony Brook, New York, 11794, United States

xiwen.zhang@stonybrook.edu

Abstract. Hyperspectral remote sensing is the acquisition of spatial and spectral data of objects by detecting the electromagnetic waves reflected from them using a hyperspectral sensor. It enables the evaluation and identification of material composition, morphology, structure, and other aspects. Currently, hyperspectral imaging is widely applied in fields such as agriculture, environmental monitoring, forestry, medicine, and geology. However, traditional hyperspectral image classification encounters the problem of high complexity and overfitting when dealing with small sample sizes, requiring the use of advanced convolutional neural network models to address this issue. This paper primarily employs literature analysis, review, and comparative analysis methods to summarize the main challenges encountered when using convolutional neural networks for hyperspectral image classification. It also selects popular models in recent years to introduce, namely 3D-CNN model, hyperspectral pyramidal ResNet model and HybridSN model. This article focuses on whether these three models can solve problems that traditional classification models cannot solve, and what room for improvement there is, which will provide a reference for future research.

Keywords: Hyperspectral Image Classification, Convolutional Neural Network, Feature Extract.

1. Introduction

Hyperspectral imaging (HSI) is an important remote sensing data source that plays a crucial role in earth observation and environmental research. Compared to traditional color images, hyperspectral images provide richer spectral information by capturing the reflectance variations of objects in the visible and near-infrared bands, enabling precise identification and quantitative analysis of surface materials. Composed hundreds of contiguous narrow bands, hyperspectral images offer detailed spectral signatures, allowing differentiation of different materials with similar spectral characteristics. Furthermore, hyperspectral images find wide applications in environmental monitoring, geological exploration, urban planning, and disaster management.

Since the classification of each image in HSI plays an important role in these applications, a large number of HSI classification methods have been explored over the decades. Traditional HSI classification is usually based on spectral information only, so their classification accuracy is often unsatisfactory. With the development of deep learning, Convolutional Neural Networks (CNN) and other algorithms have made great progress in extracting spectral and spatial features, and have great room for development. According to the data from Web of Science, the number of papers with keywords

"CNN" and "HSI" has increased year by year. In 2017, there were only 25 papers in this field, but in 2018 there were 58, in the following years 109, 117, 212, and in 2022, there were 283 papers a year[1].

CNNs are widely used in computer vision and image recognition areas, and they aim to simulate the processing of visual information to the human visual system. CNNs have demonstrated excellent performance in image processing tasks and are widely applied in image classification, object detection, facial recognition, and so on. Their strength lies in their ability to automatically learn feature representations from images and exhibit certain translational, rotational, and scale invariance. Additionally, CNNs can benefit from transfer learning and pre-trained models to accelerate the training process, making them essential tools in the field of computer vision today, as well as in HSI classification area. However, the increasingly mature hyperspectral imaging technology also brings new challenges to the task of hyperspectral image classification.

2. The difficulty of hyperspectral image classification

2.1. High complexity

The information reflected in the HSI spectrum dimension represents the reflectance of objects at different wavelengths, and this dimension often includes continuous spectral bands ranging from visible light to thermal infrared, totaling hundreds of bands. This results in high-dimensional spectral dimensions in current hyperspectral images, often reaching hundreds of bands, which leads to a large computational burden for hyperspectral image classification tasks [1]. It not only takes a long time but is also difficult to deploy on more devices due to the high redundancy. Therefore, before inputting hyperspectral image data into the classification network, some special processing is usually required on the original hyperspectral image, such as dimensionality reduction and band selection, to reduce useless information and improve classification efficiency.

2.2. Overfitting with small samples

Due to the increased feature dimension extracted by the model, the required number of samples for parameter training in hyperspectral images also increases dramatically. If the number of samples is too small, the precision of the estimated parameters cannot be guaranteed, and the parameters are not optimal. However, hyperspectral image classification usually requires pixel-level annotations, which are time-consuming and labor-intensive tasks that require professional annotation. The increasing complexity of hyperspectral images further increases the difficulty of annotation. Therefore, there is often a scarcity of annotated data in the field of hyperspectral images. Although the increase in the number of spectral bands implies more classification information, the overfitting problem arises because the estimated parameter values are not ideal, resulting in a significant deviation between the classification results and the ideal state. This phenomenon is known as the Hughes phenomenon, where the classification accuracy in hyperspectral image classification tasks decreases as the feature dimension increases.

3. Method

The traditional hyperspectral image classification method based on machine learning (ML) mainly consists of two steps: feature engineering and classifier training [2]. First, a specific manual design method is used to select the most representative features from raw HSI data with hundreds of spectral bands. Then, deeper features are extracted through nonlinear transformation, hoping to learn information unique to each category and then send it to subsequent classifiers for classification. Finally, the classifier will classify the categories based on this distinctive information.

However, traditional ML-based methods are usually unable to fully utilize the spectral spatial characteristics of ground objects due to the difficulty of manually designing feature extraction methods. In addition, traditional ML methods are difficult to effectively train on large data training samples, so dimensionality reduction operations need to be performed first, and inappropriate dimensionality reduction in the spectral domain may lead to the loss of much useful spectral information. Nowadays, with the continuous improvement of computing equipment and the rapid development of computing

resources, deep learning has shown great potential in extracting hierarchical and nonlinear features [3]. Therefore, next the paper will introduce the application of CNN, a deep learning method, in image classification tasks.

The core idea behind CNN is to extract features from images using convolutional and pooling operations. The convolutional operation involves sliding a small window (called a kernel) over the input image to extract local region features, capturing spatial structure information. The pooling operation is used to reduce the dimensionality of feature maps while preserving the most significant features. This hierarchical feature extraction process enables CNNs to automatically learn and represent abstract features in images, enabling understanding and classification of image content.

In addition to convolutional and pooling layers, CNNs also include components such as activation functions, fully connected layers, and output layers. Activation functions introduce non-linear transformations, increasing the expressive power of the network. Fully connected layers combine and integrate the features extracted from previous layers. The output layer uses appropriate activation and loss functions based on the specific task, such as classification, regression, and other objectives.

The layout of the convolutional neural network is the neural network closest to the physical biological brain, and has superiority in processing tasks. Compared with general neural networks, convolutional neural networks have outstanding performance in image processing: 1) Through the local connection and weight-sharing mode of neurons, the number of connections and training parameters of the network are reduced, and the operating efficiency is improved. At the same time, the simple network structure is more adaptable to a variety of classification tasks; 2) The weight sharing of the unified layer is conducive to the parallel operation of the network, and the parallel operation of multiple GPUs/CPUs has a very important impact on speeding up the training time of huge deep neural networks; 3) The network topology is suitable for the input of image data and can directly process the two-dimensional matrix of the image; 4) The feature extraction process and pattern classification process can be carried out simultaneously during the training process, avoiding complex, random, and inconsistent processes.

4. Popular models

Before 3D-CNN, the more commonly used is 2D-CNN, which performs well on traditional RGB image classification. Its main advantage is that features can be extracted directly from the original image. However, 2D-CNN can not cope with the problems of high complexity and small sample overfitting when dealing with HSI data set, so many new models have been proposed.

4.1. 3D-CNN model

3D-CNN uses a 3D kernel to carry out 3D convolution operation, and extracts the spatial and spectral features of the image at the same time, so as to retain the rich input spectral information of HSI. Unlike 2D-CNN, 3D-CNN contains no pooling layer, only two convolutional layers and one fully connected layer. 3D-CNN is based on pixel-level HSI classification, and the steps to extract pixel information are mainly divided into three steps. The first step is to preprocess the three-dimensional image, the second step is to extract the depth spectral space features, and the last step is to classify based on the depth spectral space features. Experimental results confirm that compared with traditional image classification methods, 3D-CNN can make full use of spectral information, and has fewer parameters to adjust, which makes it obtain better overall accuracy. In addition, because this model is light, it is not easy to overfit and easily train [4].

4.2. Hyperspectral pyramidal ResNet model

The hyperspectral pyramidal ResNet model consists of several stacked convolution layers where the output layer is larger than the input layer. In this way, the number of spectral channels is gradually increased on each block, creating the illusion of a pyramid. With the deepening of residual units, more feature maps can be extracted, so that the spatial and spectral information of HSI can be better utilized. Experimental results show that the accuracy of this model is better than other traditional models [5].

However, due to the construction of the pyramid, the computational cost of these HSI residual units is still high.

4.3. HybridSN model

Because the deep 3D-RNN model is still computationally complex, HybridSN mode is further improved. In order to simplify the model, the input spatial information and spectral information are combined by three-dimensional convolution and two-dimensional convolution. Specifically, three three-dimensional convolutions are first performed to save the spectral information of the input HSI data in the output. A 2D convolution is then performed, keeping in mind that it strongly differentiates spatial information within different spectral bands. This can reduce a bit of complexity while also making full use of the spatial and spectral characteristics of HSI data. The experimental results show that compared with simple 2D or 3D models, the mixed 3D and 3D convolution is not only more efficient, but also has excellent accuracy [6].

5. Conclusion

This paper mainly summarizes the difficulties of convolutional neural networks in hyperspectral image classification, which are the overfitting problems of large computational cost and small samples. In addition, this paper also mentions three popular models in recent years, which have made many improvements to solve the above difficulties, especially in the accuracy improvement and the optimization of the effect on small sample data sets. However, the high computational cost is still a problem, and it may be necessary to combine the dynamic neuron perception mechanism for further optimization. Although this paper focuses on the most popular models of CNN in HSI image classification, there are still some shortcomings. The next step will focus on the latest performance of graph neural networks and generative adversarial networks in HSI.

References

[1]   Y. Chen, H. Jiang, C. Li, X. Jia and P. Ghamisi, "Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 10, pp. 6232-6251, Oct. 2016, doi: 10.1109/TGRS.2016.2584107.

[2]   M. A. Hossain, Hasin-E-Jannat, B. Ahmed and M. A. Mamun, "Feature mining for effective subspace detection and classification of hyperspectral images," 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh, 2017, pp. 544-547, doi: 10.1109/ECACE.2017.7912965.

[3]   S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi and J. A. Benediktsson, "Deep Learning for Hyperspectral Image Classification: An Overview," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 9, pp. 6690-6709, Sept. 2019, doi: 10.1109/TGRS.2019.2907932.

[4]   Li, Ying, Haokui Zhang, and Qiang Shen. 2017. "Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network" Remote Sensing 9, no. 1: 67. https://doi.org/10.3390/rs9010067

[5]   M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza and F. Pla, "Deep Pyramidal Residual Networks for Spectral–Spatial Hyperspectral Image Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 2, pp. 740-754, Feb. 2019, doi: 10.1109/TGRS.2018.2860125.

[6]   S. K. Roy, G. Krishna, S. R. Dubey and B. B. Chaudhuri, "HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification," in IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 2, pp. 277-281, Feb. 2020, doi: 10.1109/LGRS.2019.2918719.