Hyperspectral Image Classification Method Based on Multi-scale DenseNet and Bi-RNN Joint Network

Lianhui Liang ¹, Jun Li ² and Shaoquan Zhang ¹*

¹College of Electrical and Information Engineering, Hunan University, Changsha, Hunan, 410082, China
²School of Geography and Planning, Sun Yat-sen University, Guangzhou, Guangdong, 510275, China
*Corresponding author’s e-mail: zhangshaoquan1@163.com

Abstract. With the depth of the Convolutional neural network (CNN) increases, CNN may lead to the problem of gradient disappearance. Simultaneously, single scale convolutional kernel may not reflect the complex spatial structural information in hyperspectral image (HSI). In addition, the CNN based approach regards the spectral band data on a single pixel of the HSI as a disordered high dimensional vector for processing, which does not meet the characteristics of the spectral band data. To tackle these aforementioned issues, a novel classification approach based on multi-scale densely connected convolutional network (DenseNet) and bi-direction recurrent neural network (Bi-RNN) with attention framework is introduced in this study. Specifically, multi-scale DenseNet is exploited to fully extract the multiple scales complex spatial structural information and utilize the strong complementary yet correlated spatial feature information between convolution layers, and Bi-RNN with attention is designed to obtain inner spectral correlations within a continuous spectrum. For comparison and verifying the effectiveness of our proposed method, we test the proposed method with nine other recently proposed methods on Salinas dataset, and the experimental results demonstrate that the proposed method can sufficiently exploit spectral and spatial information and outperforms other competitive methods.

1. Introduction

Hyperspectral images, which consist of hundreds of contiguous spectral bands, contain plentiful information. With the abundant information both in spectral and spatial dimension [1,2], HSI has been applied in many practical applications, such as Anomaly detection [3], water monitoring [4] and among others. HSI classification is to assign each pixel vector to a specific class, is one of the main tasks in the processing of HSI.

Deep learning methods have been introduced in hyperspectral image classification due to their strong predictive power, they can extract more discriminative features from the plentiful spectral features and spatial backgrounds of HSI, and achieve better performance than traditional shallow classifiers. Using deep learning methods to obtain the discriminated high-level feature for HSI classification has become a hot topic in remote sensing community [5]. In generally, deeper networks can capture finer features, but it will be hard to train deeper networks, which can easily lead to the phenomenon of gradient disappearance and explosion. The emergence of the DenseNet alleviates those problems [6].

DenseNet uses concatenation for feature aggregation and densely connections to ensure the maximum flow of HSI information between layers in the network by directly connecting all layers to
each other. Furthermore, the effect of gradient disappearance in DenseNet is reduced while maintaining the expressive power of the network, thereby enabling deeper networks to produce better HSI classification results\cite{6}. 3D DenseNet was proposed to HSI classification to learn more spectral-spatial information\cite{7}. \cite{8}proposed a double-branch multi-attention mechanism network(DBMA) motivated by a DenseNet to extract spectral and spatial features separately. Based on DenseNet, \cite{9}introduced a fast densely connected spectral-spatial convolution network (FDSSC), which greatly reduced the training time in HSI classification. \cite{10}utilized a double-branch dual-attention mechanism(DBDA) for HSI classification and obtained better performance. However, the abovementioned DenseNet methods only adopt the single scale convolutional kernels and thus are not conducive to reflect the complex spatial structural information in HSI. Thus, a multi-scale DenseNet framework was proposed to sufficiently exploit multiple scales information for the HSI classification\cite{11}.

Furthermore, the spectral data of HSI intrinsically is a sequence-based data structure. However, the CNN based approach regards the spectral band data on a single pixel of the HSI as a disordered high dimensional vector for processing, which does not meet the characteristics of the spectral band data and may led to information loss. \cite{12}adopted an RNN for HIS classification and achieved outstanding performance at first. \cite{13}proposed a spectral-spatial attention network (SSAN) for HSI by means of using Bi-RNN attention network to learn inner spectral correlations within a continuous spectrum and using CNN attention network to focus on spatial relevance between neighboring pixels in the spatial dimension, which results demonstrate that the full use of spectral and spatial information can considerably enhance performance.

To address the abovementioned problems, a novel framework based multi-scale DenseNet and Bi-RNN attention network(MD-RNN) is proposed. Multi-scale DenseNet is utilized to extract complex spatial structural information and utilize the correlated spatial feature information between convolution layers, and Bi-RNN with attention is exploited to obtain inner spectral correlations within a continuous spectrum. In this paper, the major contributions of this study are summarized three steps. Firstly, to sufficiently utilize complex spatial structure information, through principal component analysis(PCA), the dimensions of HSI are reduced to four low-dimensional subspace. Then, multi-scale convolution kernels DenseNet is exploited to extract the complex spatial feature information of HSI. Secondly, a Bi-RNN attention network is used for obtaining spectral information, and additional spectral attention parameters assign a greater weights to the key spectral data through the attention network to enhance the spectral correlation between adjacent spectrum bands. Finally, in order to extract the integrated spectral-spatial features, we concatenate the output of the two branches to form a new fully connected layer, and feed it into a mish activation function to predict the probability distribution of each class.

2. Related works

2.1. Multi-scale DenseNet for spatial feature

Generally, the dense block is the basic unit in DenseNet, which directly connects all layers to ensure maximum information flow between each layer. Each layer accepts information from the early layer as input and then conveys its own feature map to the later layer\cite{6}.

With the increase of DenseNet convolutional layers, fine (detailed) to coarse (more abstract) features can be extracted from different convolution layers, but a single-scale fixed convolution kernel of size cannot effectively obtain different scale spatial information of the HSI, the DenseNet method of multi-scale convolution kernels is proposed. Specifically, as shown in Fig.1, the spatial size of the input pixel of HSI is S×S, owns b channels, and each convolutional layer in a single dense block consists of k convolution kernels of size m×m(m=1,3,5,…), thus the number of feature maps generated by each convolutional layer is k. Meanwhile, as the number of DenseNet layers increase, the number of input feature maps can be formulated as:

\[
k_i = b + (l - 1) \times k
\]

(1)

where b is the number of the initial feature maps, that is, the number of input layer channels. Through the dense spatial block, the channel feature maps are merged to be k, and successfully obtain deeper
layers spatial feature information. Assuming the input of this block is $X_i$, and the output is $X_{i+1}$, then the extracted spatial feature information can be expressed as:

$$X_{i+1} = \text{Concat}[D_1(X_i), D_2(X_i), \ldots, D_j(X_i)]$$

where $D_1(\cdot), D_2(\cdot), D_j(\cdot)$ denote the output of DenseNet block under convolution kernels of different scale size, separately. Concat(\cdot) represents the concatenation operation, concatenating the feature maps which are outputted by all DenseNet blocks of different scale convolution kernels. In addition, to reduce the dimensionality of the channel and the network performance is improved while ensuring the accuracy, $1 \times 1$ convolution is exploited.

2.2. Bi-RNN attention network for spectral feature

Through regarding all spectra of a hyperspectral pixel as a sequence, we use a traditional Bi-RNN model, Bi-RNN is introduced to further utilize both latter and previous spectral information, which containing forward and backward hidden layer\cite{13}. The structure of Bi-RNN attention network is illustrated in Fig. 2. Its input is one spectral vector of hyperspectral vector $X$, $X=(X_1,X_2,\ldots,X_n)$, thus the bidirectional hidden vector is implemented as:

$$g = \text{Concat}(h_f, h_b)$$

where Concat(\cdot) represents the concatenation function between the forward and backward hidden layer. The update rule of forward and backward hidden layer can be respectively expressed as follows:

$$h_f = f(W_f x + U_f h_{f-1} + b_f)$$

$$h_b = f(W_b x + U_b h_{b+1} + b_b)$$

where $n$ represents the spectral band, $f(\cdot)$ represents the nonlinear activation function of the hidden layer, the coefficient matrices $W_f$ and $W_b$ are from the input at the previous step and the present step, $U_f$ and $U_b$ are from the hidden state $h_{f-1}$ at the previous step and $h_{b+1}$ at the succeeding step, respectively. Based on the Bi-RNN, an attention layer is added to obtain the weight of different spectral information to learn more features, the attention layer can be computed as:

$$\beta_i = \text{softmax} [W_i \tanh(W_i g_x + b_i) + b_i']$$

where $W_i$ and $W_i'$ represent coefficient matrices, $b_i$ and $b_i'$ are bias terms, $\tanh(\cdot)$ denotes the hyperbolic tangent function, and softmax(\cdot) is a activation function, whose output is attention weight and meets the probability distribution. Finally, after multiplying all the bidirectional hidden states $g_x$ and the corresponding attention weights $\beta$, we sum them to obtain a new spectral feature vector $y_n$. 

![Fig.1 Structure of multi-scale DenseNet network](image-url)
2.3. Spectral-Spatial feature fusion
The spatial feature information and spectral feature information are extracted by means of spatial branch network and spectral branch network, respectively. Subsequently, in order to fully utilize both spatial features and spectral correlation and obtain the fused spectral-spatial features, connecting the last fully connected(FC) layer in the multi-scale DenseNet and the one in the Bi-RNN to form a new FC layer, which is followed by another new FC layer to denote the fused spectral-spatial features. Finally, through the FC layer and mish activation function, the results of classification are achieved. The whole framework of this proposed method is presented in Fig 3.

3. Experiment results and analysis
To verify the effectiveness of our proposed approach, several contrast experiments are conducted on the Salinas dataset. The image contains 16 land cover classes and is of size 512×217 pixels, and which consist of 204 spectral bands whose wavelength range from 0.4 μm to 2.5μm. In experiment, 20 water absorption bands are discarded and remaining 204 spectral bands are available to classification. The datasets are split into training, validation, and test sets. We randomly select 1% samples for training and validation from each class, respectively.

In the respects of parameter setting, the learning rate is set as 0.005, the size of patch is 29×29, and the dropout is 0.1. The batch size is 128, and the number of training epochs is 20000. Furthermore, in order to fully obtain different scales spatial information in multi-scale DenseNet, the size of convolutional kernels is set to 3×3 and 7×7. The number of convolution layers in two DenseNets is set to 3, and the number of convolution channels is separately set to 32, 32 and 32.

To demonstrate the superiority of MD_RNN approach, MD_RNN is compared with other widely used approaches and the state-of-the-art approaches, such as SVM, DBMA, FDSSC, DBDA,
ARNN[13], SSAN[13], SSRN[14], and 3DOC-SSAN[15]. In addition, the three metrics of overall accuracy (OA), average accuracy (AA), and Kappa coefficient are utilized to objectively evaluate the performance of measurements for each method[13]. For a fair comparison, all contrast experiments are conducted ten times, and then the average results are adopted to reduce the impacts of random selection. Meantime, all contrast methods are executed with the default parameters settings given in the papers. The categorized results of different methods on Salinas dataset are shown in table 1, where the bold indicates the best classification accuracy, and the classification maps of different methods are shown in Figs 4.

![Classification maps using different methods on the Salinas dataset](image)

Table 1 Categorized results of different methods on Salinas dataset

| Class | SVM       | SSRN      | DBDA     | FDSSC     | DBMA     | ARNN     | SSAN      | 3DOC-SSAN  | MD-RNN   |
|-------|-----------|-----------|----------|-----------|----------|----------|-----------|------------|----------|
| 1     | 98.27     | 98.97     | 100.00   | 100.00    | 100.00   | 93.09    | 95.73     | 100.00     | 99.54    |
| 2     | 98.88     | 99.73     | 99.67    | 99.94     | 100.00   | 94.60    | 98.99     | 99.97      | 100.00   |
| 3     | 98.81     | 95.86     | 97.37    | 98.80     | 97.85    | 99.64    | 98.61     | 99.03      | 98.86    |
| 4     | 99.19     | 95.64     | 94.31    | 98.16     | 94.20    | 98.39    | 96.12     | 94.57      | 99.27    |
| 5     | 98.13     | 95.71     | 99.76    | 99.37     | 98.60    | 98.82    | 99.89     | 99.70      | 99.81    |
| 6     | 99.46     | 99.97     | 99.99    | 99.98     | 99.19    | 99.97    | 99.72     | 100.00     | 99.97    |
| 7     | 99.43     | 99.93     | 99.80    | 100.00    | 99.70    | 93.47    | 94.72     | 99.69      | 99.97    |
| 8     | 99.33     | 98.88     | 99.88    | 99.69     | 99.85    | 99.09    | 99.51     | 99.72      | 99.52    |
| 9     | 98.93     | 98.88     | 98.88    | 99.10     | 98.79    | 99.66    | 100.00    | 96.95      | 100.00   |
| 10    | 98.93     | 98.05     | 98.05    | 99.10     | 98.79    | 99.66    | 100.00    | 96.95      | 100.00   |
| 11    | 90.44     | 98.08     | 96.42    | 97.55     | 97.78    | 90.54    | 94.84     | 100.00     | 99.71    |
| 12    | 99.21     | 99.35     | 99.70    | 99.78     | 99.81    | 98.04    | 99.95     | 98.16      | 99.84    |
| 13    | 98.44     | 98.54     | 99.58    | 99.67     | 99.33    | 89.06    | 88.62     | 96.58      | 99.78    |
14 88.55 99.10 98.71 98.93 98.52 89.03 92.27 96.41 99.90
15 54.72 75.08 90.09 87.25 95.89 91.27 85.55 95.15 99.06
16 92.37 99.90 99.93 99.97 99.02 94.52 97.17 100.00 98.42
OA /% 89.17 92.47 96.06 97.22 96.18 94.83 96.14 97.91 99.39
AA /% 87.90 96.81 97.74 98.57 97.96 95.02 96.47 98.25 99.52
Kappa /% 92.99 91.68 95.61 96.91 95.73 94.24 95.71 97.67 99.32

Table 1 and Fig 4 exhibit that the MD-RNN is effective in HSI classification. It can be seen that the MD-RNN gives the best performance in terms of OA, AA, and Kappa. The classification accuracy of SVM is 89.17, which is poorest performance compared to other methods based on deep learning. On the one hand, the classification performance of MD-RNN is better than SSRN, DBDA, FDSSC and DBMA, which indicates that the feature extraction network of Bi-RNN based spectral attention has more advantages in extracting spectral features. On the other hand, the MD-RNN has higher classification accuracy than ANN and SSAN, showing that multi-scale DenseNet can extract more spatial feature information than the single-scale CNN. Simultaneously, it can obviously see that SVM, SSRN, DBDA, DBMA, ARNN and SSAN indicate poor performance in terms of OA, which are all below 97%. The classification maps of them also present massively mislabelled areas. It also can be seen that FDSSC and 3DOC-SSAN show a smoother visual effect than other methods, but the MD-RNN has relatively less misclassification noise and can obtain a better balance in both homogeneous and structural regions.

4. Conclusion
In this study, a novel framework for HIS classification based on a multi-scale DenseNet and Bi-RNN joint network was proposed. It has two branch networks to extract spatial feature information and spectral feature information, separately adopting densely connected 2D convolution layer with kernels of different sizes, and Bi-RNN attention network. Then, spatial information and spectral information are integrated to better represent spectral-spatial features. Experiments conducted on Salinas dataset demonstrate that our proposed approach can obtain outstanding performance compared to other widely used classifiers.

The effectiveness of MD-RNN classification method has been proven on Salinas dataset. In the next work, we will extend this method to other more complex remote sensing scenes.

Acknowledgments
Authors would like to thank the editor and anonymous reviewers for their detail review. This work was supported in part by the Key R&D Program of Hunan Province, China(No.2019SK2102), and in part by General Program of China Postdoctoral Science Foundation(No. 2020M672483).

References
[1] F. Melgani, L. Bruzzone. Classification of hyperspectral remote sensing images with Support Vector Machines[J]. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42(8): 1778-1790
[2] Li Z, Huang L, He J. A Multiscale Deep Middle-level Feature Fusion Network for Hyperspectral Classification[J]. Remote Sensing, 2019, 11(6): 695.
[3] Zhao C, Wang Y, Qi B, et al. Global and Local Real-Time Anomaly Detectors for Hyperspectral Remote Sensing Imagery[J]. Remote Sensing, 2015, 7(4): 3966-3985.
[4] Pipitone, C.; Maltese, A.; Dardanelli, G., et al. Monitoring water surface and level of a reservoir using different remote sensing approaches and comparison with dam displacements evaluated via GNSS[J]. Remote Sensing, 2018, 10(1): 71.
[5] Li S, Song W, Fang L, et al, Deep Learning for Hyperspectral Image Classification: An Overview[J]. In IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(9): 6690-6709.
[6] Huang G, Liu Z, Maaten L., et al. K.Q. Densely connected convolutional networks[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017,
4700-4708.

[7] Zhang C, Li G, Du S, et al. 3D densely connected convolutional network for hyperspectral remote sensing image classification[J]. Remote Sensing, 2019, 13(1): 1-22.

[8] Ma W, Yang Q, Wu Y, et al. Double-Branch Multi-Attention Mechanism Network for Hyperspectral Image Classification[J]. Remote Sensing, 2019, 11(11): 1307.

[9] Wang W, Dou S, Jiang Z, et al. A Fast Dense Spectral–Spatial Convolution Network Framework for Hyperspectral Images Classification[J]. Remote Sensing, 2018, 10(7): 1068.

[10] Li R, Zheng S, Duan C, et al. Classification of Hyperspectral Image Based on Double-Branch Dual-Attention Mechanism Network[J]. Remote Sensing, 2020, 12(3): 582.

[11] Xie J, He N, Fang L, et al. Multiscale Densely-Connected Fusion Networks for Hyperspectral Images Classification[J]. In IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(1): 246-259.

[12] Mou L, Ghamisi P., Zhu X. Deep Recurrent Neural Networks for Hyperspectral Image Classification[J]. In IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7): 3639-3655.

[13] Mei X, Pan E, Ma Y, et al. Spectral-Spatial Attention Networks for Hyperspectral Image Classification[J]. Remote Sensing, 2019, 11(8): 963.

[14] Zhong Z, Jonathan L., Luo Z, et al. Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(2): 847–858.

[15] Tang X, Meng F, Zhang X, et al. Hyperspectral Image Classification Based on 3-D Octave Convolution with Spatial-Spectral Attention Network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(3): 2430-2447.