SSTNet: Spatial, Spectral, and Texture Aware Attention Network using Hyperspectral Image for Corn Variety Identification

Weidong Zhang, Zexu Li, Hai-Han Sun, Qiang Zhang, Peixian Zhuang, and Chongyi Li

***Abstract*—Currently, most existing methods using hyperspectral image to assist seed identification only consider the spectral information but ignore the spatial information resulting in unsatisfactory classification results. To cope with this issue, we propose a spatial, spectral, and texture-aware attention network to identify corn varieties, called SSTNet. Specifically, we first employ 3D convolution to extract the spatial and inter-spectral features. Subsequently, we utilize 2D convolution to extract the spatial and texture features. Meanwhile, we embed an attention mechanism into the 2D convolution module to further refine the spatial and texture features. The advantageous complementary properties of 3D and 2D convolutions allow the spatial and textural features of hyperspectral images to be fully exploited. Besides, we construct a hyperspectral image dataset including 1200 samples of 10 corn varieties. Experiments on our proposed dataset demonstrate that our SSTNet outperforms the state-of-the-art methods for identifying corn varieties.**

***Index Terms*—Hyperspectral image, corn variety, spatial features, spectral features, texture features.**

I. Introduction

DIFFERENT varieties of corn seeds adapt to the diverse needs of the market, but it also poses the risk of mixing corn varieties. Therefore, variety identification of seeds is of great value in improving the purity of corn varieties [1, 2]. Unfortunately, the traditional method of identifying the purity of corn varieties by physicochemical analysis takes a long time to identify and destroys the seeds, so it has been unable to meet the urgent needs of modern agriculture [3, 4]. Hyperspectral imaging as a fast, efficient and nondestructive identification technique has achieved remarkable results in seed variety identification [5, 6].

In the early stage, some researchers rely on the characteristics of the spectrum to manually or semi-automatically extract features and apply them to hyperspectral image classification [7–9]. For instance, collaborative sparse regression [10], sparse matrix decomposition [11], covariance matrix [12], and weighted sparse representation [13] were utilized for hand-crafted features. He *et al*. [14] proposed a manual feature extraction method based on multi-scale covariance maps to semi-automatically integrate spatial and spectral features. Zhang *et al*. [15] devised a superpixel-guided variable 3-D Gabor phase coding strategy to exploit the spatial information of the spectra. However, the richness of hand-crafted features is insufficient and depends on some priors.

Recently, deep learning has been gradually applied to image visualization [16], saliency detection [17, 18], image enhancement [19–22], and semantic segmentation [23, 24] thanks to the deep network structure and powerful feature extraction capabilities. For example, Paoletti *et al*. [25] proposed a residual-based depth pyramid network to increase the diversity of spatial properties of spectra. Zhang *et al*. [26] integrated automatic encoders and classifiers into a two-branch deep network for end-to-end collaborative learning. Hu *et al*. [27] presented a multiobject convolutional neural network to implement hyperspectral multi-task classification. Roy *et al*. [28] proposed a 3-D-2-D CNN model to extract spatial and spectral features. Yu *et al*. [29] designed a two-channel network to exploit global and multiscale features of the hyperspectral image, which fully uses feature information at different levels. In short, deep learning has better feature representation capability compared to traditional methods. Meanwhile, the effective utilization of spatial and spectral information of hyperspectral images is the key to the strong representation capability of the network model.

In this letter, we propose a **S**patial, **S**pectral and **T**exture-aware **A**ttention **Net**work using hyperspectral image for corn variety identification, called SSTNet. We highlight the contributions of this work as follows.

- We first construct a corn seed hyperspectral image dataset (*i.e*., ***CSHID***), which includes 1200 samples of 10 varieties. Specifically, each variety consists of 120 samples with 128 bands. Meanwhile, the raw resolution of each band is 696×520 and the resolution after cropping the region of interest is 210×200. CSHID also promotes the application of deep learning methods for identifying hyperspectral images of agricultural seeds to a certain extent. It also has significant applications and economic value for breeding and screening different corn grades.
- We propose a 3D spatial and spectral-aware convolution module, which utilizes 3D convolution to extract the spatial and spectral features of the target spectrum. Besides,

This work was supported in part by the National Natural Science Foundation of China under Grants 62171252, the Postdoctoral Science Foundation of China under Grant 2021M701903. (*Corresponding author: Hai-Han Sun.*)

Weidong Zhang and Zexu Li are with the School of Information Engineering, Henan Institute of Science and Technology, Xinxiang, China (e-mail: zwd_wd@163.com, lizexu_19@163.com).

Hai-Han Sun is with the School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798 (e-mail: hannah.h.sun@outlook.com).

Qiang Zhang is with the School of Information Science and Technology, Dalian Maritime University, Dalian, China (e-mail: qzhang95@dlmu.edu.cn).

Peixian Zhuang is with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China (e-mail: zhuangpeixian0624@163.com).

Chongyi Li is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore (e-mail: lichongyi25@gmail.com).

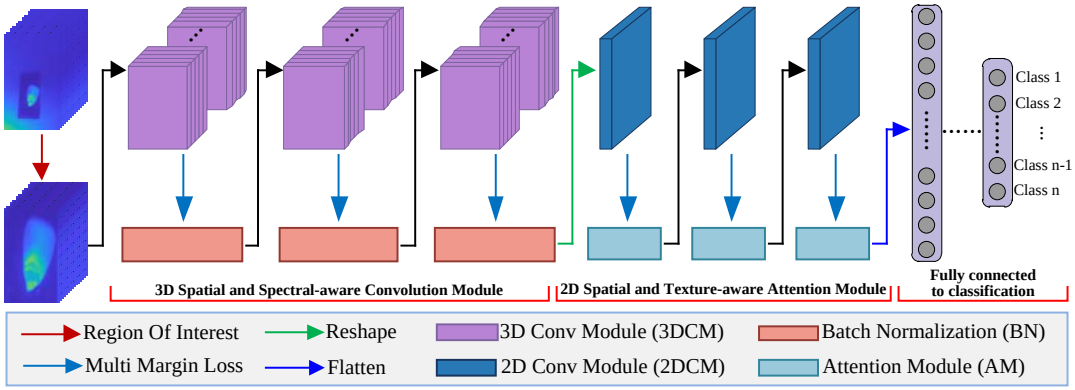

Fig. 1. **Flowchart of our SSTNet method.** In our SSTNet, the region of interest is first derived from the input hyperspectral image of corn seed. Subsequently, the regions of interest of each band of these corn seeds were passed to the 3D spatial and spectral-aware convolution module to capture each band's spatial and spectral features. Whereafter, we utilize the 2D spatial and texture-aware attention module to further refine the spatial and texture features to serve the identification task of corn seeds.

we also propose a 2D spatial and texture-aware attention module, which extracts the spatial and textural features of the target spectrum using 2D convolution module with channel and spatial attention.

This letter introduces the proposed spatial and texture-aware attention network in Section II. Section III presents the experimental results and analysis. Section IV summarizes the conclusions.

II. Methodology

Fig. 1 presents the flowchart of SSTNet. Our SSTNet mainly consists of three 3D convolution modules and three 2D convolution modules with attention mechanism. The resolution of the hyperspectral image determines the input node of SSATNet, and the discriminated class determines its output node. Our SSTNet includes three cascaded 3D convolutional modules with batch normalization, three cascaded 2D convolutional modules, channel and spatial attention modules, and three fully connected layers. SSTNet has a total of 163,896 training parameters and ten outputs.

A. 3D Spatial and Spectral Aware Convolution Module

Hyperspectral images need to capture spectral and spatial information encoded in multiple bands. 3D convolution can extract both spatial and spectral hierarchical features from hyperspectral images, but at the cost of increased computational complexity. In our work, we use 3D convolution kernels to generate feature maps on multiple successive bands in the input layer to obtain features at both spatial and spectral levels. In the 3D convolution, the activation value of the j^{th} feature map at the spatial coordinates (x, y, z) of the i^{th} layer is defined as:

$$f_{i,j}^{x,y,z} = \mathrm{LR}\left(\sum_{j=1}^{J_{(i-1)}}\sum_{l=0}^{L_i}\sum_{w=0}^{W_i}\sum_{h=0}^{H_i} k_{i,j}^{l,w,h} \times f_{(i-1)j}^{(x+l)(y+w)(z+h)} + \theta_{i,j}\right), \quad (1)$$

where L, W, and H represent the layer, width and height, respectively. l, w, and h represent the indices of the three dimensions in the convolution process, respectively. LR represents the $LeakyRelu$ activation function. $\theta_{i,j}$ denotes the bias value of the j^{th} feature map of the i^{th} layer. $J_{(i-1)}$ represents the number of feature maps at layer $L-1$. $k_{i,j}^{l,w,h}$ represents the weight of the convolution kernel connecting the j^{th} feature image at the spatial coordinates (l, w, h). Subsequently, we perform batch normalization for each pixel value in the 3D convolution module to prevent gradient explosion or disappearance. Meanwhile, we designed three cascaded 3D convolution modules to utilize spatial and spectral features fully.

B. 2D Spatial and Texture-aware Attention Module

2D convolution can be used to capture the texture features of an image, which estimates valuable feature maps by covering all the feature information of the previous layer. In 2D convolution, we complete the 2D convolution operation by calculating the sum of dot products between data and kernel. In our work, the kernel spans the input data to cover the texture features of the entire image. We pass the feature maps through the $\mathrm{Leaky}Relu$ activation function, where the activation function of the j^{th} feature map at spatial location (x, y) in the i^{th} layer is mathematically expressed as:

$$f_{i,j}^{x,y} = \mathrm{LR}\left(\sum_{j=1}^{J_{(i-1)}}\sum_{l=0}^{L_i}\sum_{w=0}^{W_i} k_{i,j}^{l,w} \times f_{(i-1)j}^{(x+l)(y+w)} + \theta_{i,j}\right), \quad (2)$$

where the details of LR, J, L, and W refer to Eq. (1), and l and w are indices in two dimensions during the convolution process. $k_{i,j}^{l,w}$ represents the weight of the convolution kernel connecting the j^{th} feature image at the spatial coordinates (l, w). Subsequently, we embedded a spatial and channel attention mechanism [30] in the 2D convolution module to further refine the texture and spatial features. As shown in Fig 2, it is mainly composed of spatial attention and channel attention.

Channel attention first compresses the feature map in the spatial dimension to obtain a one-dimensional vector, which considers both mean pooling and max pooling. Average pooling and max pooling are used to aggregate the spatial information of feature maps and feed it to a shared network. It produces a channel attention map by compressing the spatial

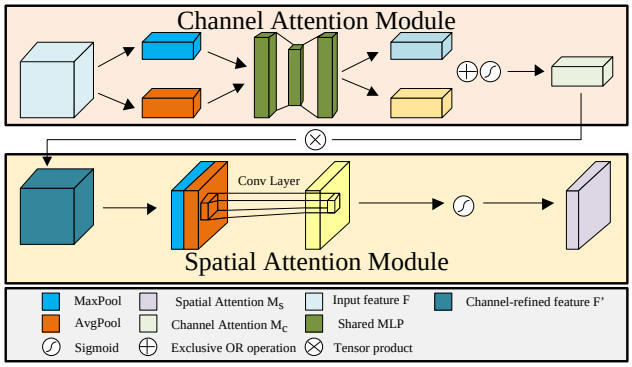

Fig. 2. The feature refinement process of the spatial and channel attention modules.

dimension of the input feature map and summing it element-wise. Average pooling has feedback for each pixel on the feature map, while max pooling only has gradient feedback at the most responsive part of the feature map when performing gradient backpropagation calculations. In summary, the channel attention mechanism is mathematically expressed as

$$F_{cam}(\mathbf{F}) = \mathrm{Sg}(\mathrm{MLP}(\mathrm{AP}(\mathbf{F})) + \mathrm{MLP}(\mathrm{MP}(\mathbf{F}))), \quad (3)$$

where MLP denotes the multi-layer perceptron, AP denotes the average pooling, MP denotes the max polling, $\mathbf{F}$ denotes the feature matrix output by the 2D convolution module, and Sg denotes the $Sigmoid$ activation function.

The spatial attention mechanism compresses the channels and performs mean pooling and maximum pooling in the channel dimension. Subsequently, the module merges the previously extracted feature maps to obtain a two-channel feature map. The computational process of the spatial attention mechanism is expressed as:

$$F_{sam}(\mathbf{F}) = \mathrm{Sg}(\mathrm{k}^{7\times7}([\mathrm{MLP}(\mathrm{AP}(\mathbf{F})); \mathrm{MLP}(\mathrm{MP}(\mathbf{F})]))), \quad (4)$$

where $k^{7\times7}$ represents the operation of convolution using the convolution kernel of 7×7. For details of other parameters refer to Eq. (3). In our work, we designed three cascaded 2D spatial and texture-aware attention modules that can extract spatial and texture features better.

C. Loss function

In our SSTNet, we use the multi-margin loss (MML) [31] as the network's objective optimization loss function. MML can fully consider the error between the true class and other classes. For the corn seed data $Data(x, y)$ containing C samples, x is the neural network's output, and y is the true class label. We assume that the number of corn seed classes is n, then $0 \leq \mathrm{y}_n \leq C-1$. In short, the loss $Loss_n$ of the n^{th} sample in i^{th} class is defined as:

$$Loss_n = \frac{1}{C} \sum_{i=0, i\neq y_n}^{C-1} \max\left(0, \mathrm{Mar} - x_n[y_n] + x_n[i]\right)^{\eta}, \quad (5)$$

where Mar denotes a variable offset added to the loss, and [] denotes the maximum rounding operation not greater than y_n [32]. To deal with the issue of sample imbalance between multiple classes, a weight w controlling each class is introduced in Eq. (5). Thus, Eq. (5) can be reexpressed as:

$$Loss_n = \frac{1}{C} \sum_{i=0, i\neq y_n}^{C-1} \max\left(0, w[y_n]\mathrm{Mar} - x_n[y_n] + x_n[i]\right)^{\eta}. \quad (6)$$

Whereafter, the total loss for all classes is defined as:

$$Loss_{all} = \frac{Loss_1 + Loss_2 + ... + Loss_N}{C}, \quad (7)$$

where N represents the number of total samples, and $Loss_{all}$ represents the total loss value.

III. Experiments and Analysis

This section mainly introduces the proposed hyperspectral image dataset of corn seeds, classification results, and ablation study. Our SSTNet uses the Adam algorithm as the optimizer, the batch size of the train samples is set to 40, the batch size of the test samples is set to 16, the Dropout is set to 0.2, the learning rate is initialized to 10^{-2}, it decays linearly until it reaches 0 after 250 iterations, and the ratio of training samples to test samples is set to 4:1.

A. Proposed dataset

We collected a corn seed dataset of 10 varieties from the Henan region using the SOC 710 portable visible/NIR imaging spectrometer from Surface Optics. Its spectral range is 400-1000 nm, the spectral resolution is 4.6875 nm, the number of bands is 128, the raw image resolution is 696520 and the resolution after cropping the region of interest is 210×200. The corn varieties we collected were BaiYu607, BaiYu808, BaiYu818, BaiYu833, BaiYu879, BaiYu897, BaiYu918, BaiYu8317, BaiYu9284, and Feng-Da601, respectively. Meanwhile, each corn variety has 120 samples and each sample has 128 spectral bands. Besides, Fig. 3 shows some representative bands of BaiYu607 corn seeds.

B. Identification Results

In this letter, our SSTNet is compared with 8 classification methods, including **machine learning methods**: KNN [33], RFA [34], SGD [35], and FSVM [3]; and **deep learning methods**: DPRNet [25], MSDNet [31], HybridNet [28], and 3DCNN [36]. All methods are tested on our proposed dataset, and the ratio of training samples to test samples is 4:1. Additionally, we choose the F1-score, Recall, Precision and Accuracy as quantitative metrics to evaluate the classification performance of different methods. The higher the metric value, the better the classification performance of the method.

Table I reports the results of the different methods. From Table I, it can be observed that FSVM [3] and RFA [34] exhibit poor classification performance for machine learning methods; thus their nonlinear modeling ability is weak. KNN [33] and SGD [35] show better classification performance than FSVM [3] and RFA [34]. For deep learning methods, DPRNet [25] and MSDNet [31] achieve unsatisfactory classification performance because they do not fully consider the spatial information of the spectrum. Although 3DCNN [36] and HybridNet [28] demonstrate better classification performance,

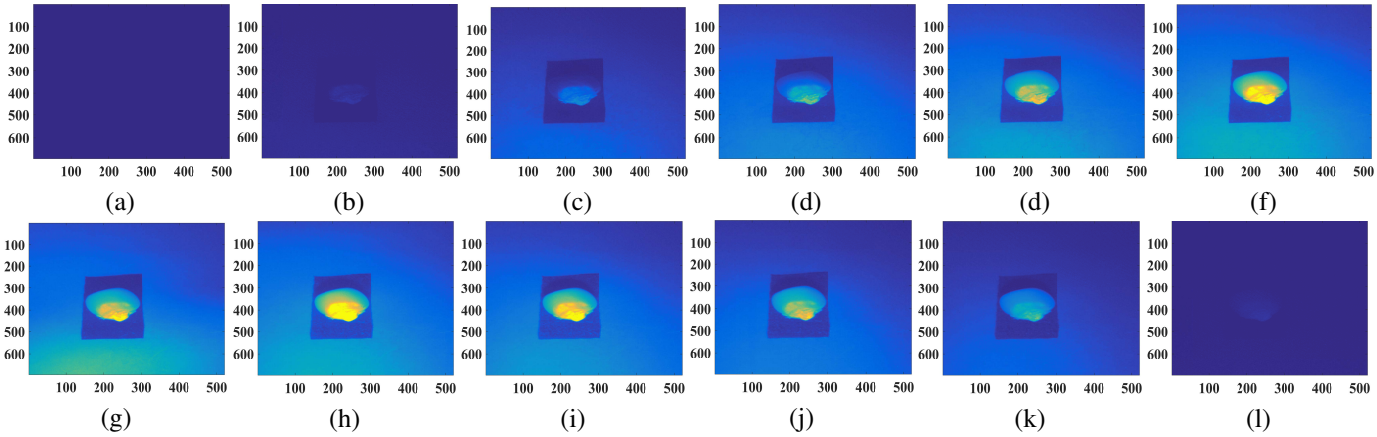

Fig. 3. Some representative spectral bands of BaiYu607 corn seeds. From left to right are (a) 1^{th} spectral band, (b) 10^{th} spectral band, (c) 20^{th} spectral band, (d) 30^{th} spectral band, (e) 40^{th} spectral band, (f) 50^{th} spectral band, (g) 60^{th} spectral band, (h) 70^{th} spectral band, (i) 80^{th} spectral band, (j) 90^{th} spectral band, (k) 100^{th} spectral band, and (l) 128^{th} spectral band, respectively.

they are still lower than our SSTNet in various metrics. In conclusion, our SSTNet fully considers the spatial, spectral, and texture features of hyperspectral images of corn seeds so that our model has a good discriminative capability.

TABLE I
IDENTIFICATION RESULTS OF 10 CLASSES OF CORN SEEDS TESTED BY DIFFERENT METHODS.(OPTIMAL: RED; SUBOPTIMAL: BLUE)

Method	F1-Score ↑	Recall ↑	Precision ↑	Accuracy ↑
FSVM [3]	0.9467	0.9417	0.9519	0.9458
KNN [33]	0.9610	0.9583	0.9637	0.9625
SGD [35]	0.9685	0.9663	0.9703	0.9705
RFA [34]	0.9426	0.9417	0.9435	0.9458
DPRNet [37]	0.9484	0.9583	0.9388	0.9583
MSDNet [31]	0.9398	0.9400	0.9396	0.9416
3DCNN [36]	0.9601	0.9576	0.9627	0.9583
HybridNet [28]	0.9621	0.9667	0.9673	0.9708
SSTNet	0.9795	0.9791	0.9800	0.9792

Fig. 4 reports the confusion matrix of SSTNet for the test samples. From Fig. 4(a), it can be observed that the identification accuracy of SSTNet for 10 types of corn varieties is over 90%, and the identification accuracy of 6 types of corn seeds is 100%. Additionally, Figs. 4(a) and (b) show the accuracy and loss convergence curves during SSTNet training, it can be seen that our SSTNet converges quickly before 50 epochs. The SSTNet has better convergence and higher accuracy when the number of epochs is 150, SSTNet leveled off after 150 epochs.

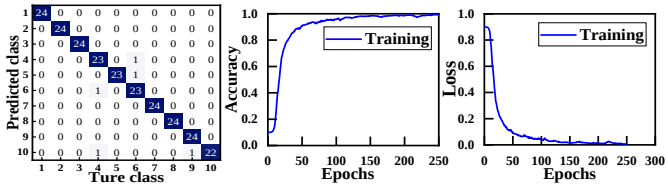

Fig. 4. (a) Confusion matrix of the SSTNet is tested on the proposed hyperspectral corn seed dataset. (b) Accuracy convergence versus number of epochs. (c) Loss convergence versus number of epochs.

C. Ablation Study

To prove the effectiveness of each module for our SSTNet for corn seed identification, we performed ablation studies on the proposed dataset. Concretely, (1) our SSTNet without 3D convolution module (-w/o 3DCM), (2) our SSTNet without batch normaliztion (-w/o BN), (3) our SSTNet without 2D convolution module (-w/o 2DCM), and (4) our SSTNet without attention module (-w/o AM).

TABLE II
RESULTS OF ABLATION STUDIES OF DIFFERENT MODULES.(OPTIMAL: RED; SUBOPTIMAL: BLUE)

Method	F1-Score ↑	Recall ↑	Precision ↑	Accuracy ↑
-w/o 3DCM	0.8741	0.8541	0.8952	0.8792
-w/o BN	0.9527	0.9458	0.9596	0.9583
-w/o 2DCM	0.9440	0.9417	0.9464	0.9458
-w/o AM	0.9541	0.9416	0.9670	0.9542
SSTNet (full model)	0.9795	0.9791	0.9800	0.9792

Table II exhibitions the F1-Score, Recall, Precision, and Accuracy scores corresponding to the ablated models, which can be shown that our SSTNet (Full Model) has the best score compared with other modules. Additionally, it also proves that each module has a positive effect on our SSTNet.

D. Running time

All methods are implemented on a Windows 11 PC with AMD Ryzen 7 3700X 8-Core CPU at 3.6 GHz, 16-GB Memory, NVIDIA GeForce GTX 1650 Super, Python 3.6, and Torch-GPU-1.10.1. Table III shows the total running time of 240 corn seeds tested by different methods. From Table III, Our SSTNet has the optimal and fast running time. In general, our method has good identification performance and high efficiency.

IV. CONCLUSION

The letter proposes an SSTNet for nondestructive identification of corn seeds using hyperspectral image. Our SSTNet fully accounts for the hyperspectral image spatial, spectral features, and texture features of corn seeds. Besides, the cooperation between the 3D module, 2D module, and attention module makes our SSTNet have better discriminative capability and robustness. In future work, we will expand the class and number of corn seeds to validate the performance of our SSTNet.

TABLE III
THE TOTAL RUNNING TIME OF DIFFERENT METHODS WERE TESTED ON 240 CORN SEEDS. (IN SECONDS; OPTIMAL: RED; SUBOPTIMAL: BLUE)

FSVM [3]	KNN [33]	SGD [35]	RFA [34]	DPRNet [37]	MSDNet [31]	3DCNN [36]	HybridNet [28]	SSTNet
0.032	0.035	0.031	0.032	0.029	0.048	0.197	0.025	0.023

REFERENCES

[1] M. Xie, Y. Li, S. Dong, B. Zhang, T. Gou, Fine-grained oil types identification based on reflectance spectrum: Implication for the requirements on the spectral resolution of hyperspectral remote sensors, IEEE Geoscience and Remote Sensing Letters 19 (2022) 1–5.

[2] D. Li, F. Kong, J. Liu, Q. Wang, Superpixel-based multiple statistical feature extraction method for classification of hyperspectral images, IEEE Transactions on Geoscience and Remote Sensing 59 (10) (2021) 8738–8753.

[3] S. Jin, W. Zhang, P. Yang, Y. Zheng, J. An, Z. Zhang, P. Qu, X. Pan, Spatial-spectral feature extraction of hyperspectral images for wheat seed identification, Computers and Electrical Engineering 101 (2022) 108077.

[4] S. Weng, K. Han, Z. Chu, G. Zhu, C. Liu, Z. Zhu, Z. Zhang, L. Zheng, L. Huang, Reflectance images of effective wavelengths from hyperspectral imaging for identification of fusarium head blight-infected wheat kernels combined with a residual attention convolution neural network, Computers and Electronics in Agriculture 190 (2021) 106483.

[5] C. Liu, H. Li, A. Su, S. Chen, W. Li, Identification and grading of maize drought on rgb images of uav based on improved u-net, IEEE Geoscience and Remote Sensing Letters 18 (2) (2021) 198–202.

[6] P. Duan, P. Ghamisi, X. Kang, B. Rasti, S. Li, R. Gloaguen, Fusion of dual spatial information for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 59 (9) (2021) 7726–7738.

[7] P. Duan, X. Kang, S. Li, P. Ghamisi, J. A. Benediktsson, Fusion of multiple edge-preserving operations for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 57 (12) (2019) 10336–10349.

[8] R. Cai, C. Liu, J. Li, Phase-induced gabor-based multiview active learning for hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters 19 (2022) 1–5.

[9] X. Shang, S. Han, M. Song, Iterative spatial-spectral training sample augmentation for effective hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters 19 (2022) 1–5.

[10] T. Chen, Y. Liu, Y. Zhang, B. Du, A. Plaza, Superpixel-based collaborative and low-rank regularization for sparse hyperspectral unmixing, IEEE Transactions on Geoscience and Remote Sensing 60 (2022) 1–16.

[11] M. Song, T. Yang, H. Cao, F. Li, B. Xue, S. Li, C.-I. Chang, Bi-endmember semi-nmf based on low-rank and sparse matrix decomposition, IEEE Transactions on Geoscience and Remote Sensing 60 (2022) 1–16.

[12] L. Fang, N. He, S. Li, A. J. Plaza, J. Plaza, A new spatialspectral feature extraction method for hyperspectral images using local covariance matrix representation, IEEE Transactions on Geoscience and Remote Sensing 56 (6) (2018) 3534–3546.

[13] L. Gan, J. Xia, P. Du, Z. Xu, Dissimilarity-weighted sparse representation for hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters 14 (11) (2017) 1968–1972.

[14] N. He, M. E. Paoletti, J. M. Haut, L. Fang, S. Li, A. Plaza, J. Plaza, Feature extraction with multiscale covariance maps for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 57 (2) (2019) 755–769.

[15] S. Zhang, D. Tang, N. Li, X. Jia, S. Jia, Superpixel-guided variable gabor phase coding fusion for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 60 (2022) 1–16.

[16] P. Duan, X. Kang, S. Li, P. Ghamisi, Multichannel pulse-coupled neural network-based hyperspectral image visualization, IEEE Transactions on Geoscience and Remote Sensing 58 (4) (2020) 2444–2456.

[17] C. Li, R. Cong, S. Kwong, J. Hou, H. Fu, G. Zhu, D. Zhang, Q. Huang, Asif-net: Attention steered interweave fusion network for rgb-d salient object detection, IEEE Transactions on Cybernetics 51 (1) (2021) 88–100.

[18] C. Li, C. Guo, C. C. Loy, Learning to enhance low-light image via zero-reference deep curve estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (8) (2022) 4225–4238.

[19] P. Zhuang, J. Wu, F. Porikli, C. Li, Underwater image enhancement with hyper-laplacian reflectance priors, IEEE Transactions on Image Processing 31 (2022) 5442–5455.

[20] Q. Zhang, Q. Yuan, M. Song, H. Yu, L. Zhang, Cooperated spectral low-rankness prior and deep spatial prior for hsi unsupervised denoising, IEEE Transactions on Image Processing 31 (2022) 6356–6368.

[21] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, C. Li, Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement, IEEE Transactions on Image Processing 31 (2022) 3997–4010.

[22] X. Fu, W. Wang, Y. Huang, X. Ding, J. Paisley, Deep multiscale detail networks for multiband spectral image sharpening, IEEE Transactions on Neural Networks and Learning Systems 32 (5) (2021) 2090–2104.

[23] R. Zhang, J. Chen, L. Feng, S. Li, W. Yang, D. Guo, A refined pyramid scene parsing network for polarimetric sar image semantic segmentation in agricultural areas, IEEE Geoscience and Remote Sensing Letters 19 (2022) 1–5.

[24] H. Li, Y. Li, G. Zhang, R. Liu, H. Huang, Q. Zhu, C. Tao, Global and local contrastive self-supervised learning for semantic segmentation of hr remote sensing images, IEEE Transactions on Geoscience and Remote Sensing 60 (2022) 1–14.

[25] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, F. Pla, Deep pyramidal residual networks for spectralspatial hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 57 (2) (2019) 740–754.

[26] L. Zhang, J. Zhang, W. Wei, Y. Zhang, Learning discriminative compact representation for hyperspectral imagery classification, IEEE Transactions on Geoscience and Remote Sensing 57 (10) (2019) 8276–8289.

[27] Y. Hu, J. Zhang, Y. Ma, J. An, G. Ren, X. Li, Hyperspectral coastal wetland classification based on a multiobject convolutional neural network model and decision fusion, IEEE Geoscience and Remote Sensing Letters 16 (7) (2019) 1110–1114.

[28] S. K. Roy, G. Krishna, S. R. Dubey, B. B. Chaudhuri, Hybridsn: Exploring 3-d2-d cnn feature hierarchy for hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters 17 (2) (2020) 277–281.

[29] H. Yu, H. Zhang, Y. Liu, K. Zheng, Z. Xu, C. Xiao, Dual-channel convolution network with image-based global learning framework for hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters 19 (2022) 1–5.

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

[31] C. Zhang, G. Li, S. Du, Multi-margin based decorrelation learning for heterogeneous face recognition, The 28th International Joint Conference on Artificial Intelligence (2019) 680–686.

[32] Paisitkriangkrai, S., van, den, Hengel, A., Shen, C., A scalable stagewise approach to large-margin multiclass loss-based boosting, IEEE transactions on neural networks and learning systems 25 (5) (2014) 1002–1013.

[33] M. M. Kumbure, P. Luukka, M. Collan, A new fuzzy k-nearest neighbor classifier based on the bonferroni mean, Pattern Recognition Letters 140 (2020) 172–178.

[34] Y. Chen, W. Zheng, W. Li, Y. Huang, Large group activity security risk assessment and risk early warning based on random forest algorithm, Pattern Recognition Letters 144 (2021) 1–5.

[35] Y. Lei, K. Tang, Learning rates for stochastic gradient descent with nonconvex objectives, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (12) (2021) 4505–4511.

[36] Q. H. Lin, Y. W. Niu, J. Sui, W. D. Zhao, C. Zhuo, V. D. Calhoun, Sspnet: An interpretable 3d-cnn for classification of schizophrenia using phase maps of resting-state complex-valued fmri data, Medical Image Analysis 79 (2022) 102430.

[37] M. Haut, Juan, Fernandez-Beltran, Ruben, Plaza, Javier, J. Antonio, Paoletti, E. Mercedes, Pla, Deep pyramidal residual networks for spectralspatial hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 57 (2) (2019) 740–754.