GACNet: Generate Adversarial-Driven Cross-Aware Network for Hyperspectral Wheat Variety Identification

Weidong Zhang, *Member, IEEE*, Zexu Li, Guohou Li, Peixian Zhuang, *Member, IEEE*, Guojia Hou, Qiang Zhang, *Member, IEEE*, and Chongyi Li, *Senior Member, IEEE*

Abstract—Wheat variety identification from hyperspectral images holds significant importance in both fine breeding and intelligent agriculture. However, the discriminatory accuracy of some techniques is limited due to insufficient datasets, data redundancy, and noise interference. To address these issues, we propose a wheat variety identification framework called generate adversarial-driven cross-aware network (GACNet), comprising a semi-supervised generative adversarial network (GAN) for data augmentation and a cross-aware attention network (CAANet) for variety identification. First, the semi-supervised GAN (SSGAN) alleviates data scarcity by generating fake hyperspectral images as realistically as possible through learning the distribution hypothesis of real hyperspectral images, while the discriminator distinguishes between real and fake hyperspectral images. Subsequently, the CAANet is employed for wheat variety identification, which leverages a cascading cross-learning of 3-D and 2-D convolutions to fully exploit spectral, spatial, and texture features and refines the features through an embedded attention mechanism in the cross-convolutional module. Additionally, we constructed a hyperspectral wheat variety dataset (HWVD) comprising 4560 samples of 19 categories. Extensive experiments on our dataset demonstrate that our GACNet outperforms state-of-the-art methods for wheat variety identification. The HWVD will be made available.

Index Terms—Cross-aware attention, generative adversarial network (GAN), hyperspectral image, wheat variety identification.

Weidong Zhang, Zexu Li, and Guohou Li are with the School of Information Engineering and the Institute of Computer Applications, Henan Institute of Science and Technology, Xinxiang 453003, China (e-mail: zwd_wd@163.com; lizexu_19@163.com; liguohou6@163.com).

Qiang Zhang is with the School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China (e-mail: qzhang95@dlmu.edu.cn).

Guojia Hou is with the College of Computer Science and Technology, Qingdao University, Qingdao 266071, China (e-mail: guojiahou@qdu.edu.cn).

Peixian Zhuang is with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100084, China (e-mail: zhuangpeixian0624@163.com).

Chongyi Li is with the School of Computer Science, Nankai University, Tianjin 300071, China (e-mail: lichongyi@nankai.edu.cn).

I. INTRODUCTION

WITH a wide variety of wheat varieties entering the market and meeting the needs of different populations, the risk of wheat variety mixing has increased. However, mixing seeds presents many challenges for fine breeding in intelligent agriculture [1]. Traditional chemometric methods for seed identification face limitations regarding seed destruction and time-consuming identification processes, failing to meet the demand for efficient and rapid identification of seed varieties in modern agriculture. Hyperspectral imaging, on the other hand, offers non-destructive, fast, and efficient identification capabilities by capturing and analyzing spectral information point by point in a spatial region, making it increasingly attractive for agricultural seed identification [2], [3].

Currently, the main focus to improve the accuracy of hyperspectral wheat variety identification is through data augmentation and the development of effective identification methods [4]. Data augmentation benefits from the generative adversarial network (GAN), which excels in data generation thanks to the constant game characteristics between the generator and discriminator. It is worth noting that rich and diverse data significantly improve the robustness and effectiveness of deep learning methods [5]. In terms of identification methods, traditional, machine learning, and deep learning methods have been widely used in hyperspectral image classification. Traditional discrimination methods in the early years relied on hand-crafted features for hyperspectral image classification [6], [7], while machine learning methods employ hand-crafted or semi-automatically extracted features for the hyperspectral image classification [8], [9], [10]. In contrast, deep learning methods can automatically extract valuable features, leading to improved accuracy in hyperspectral image classification [11], [12].

The primary challenges in hyperspectral wheat seed variety identification are the scarcity of hyperspectral seed image datasets and the effective extraction and utilization of hyperspectral image features for seed variety identification. Consequently, the current focus on improving the accuracy of hyperspectral seed variety identification revolves around data augmentation and the development of robust identification methods.

Regarding data augmentation, many strategies have been employed to derive more representations from the raw data, thereby enhancing both the quantity and quality of raw
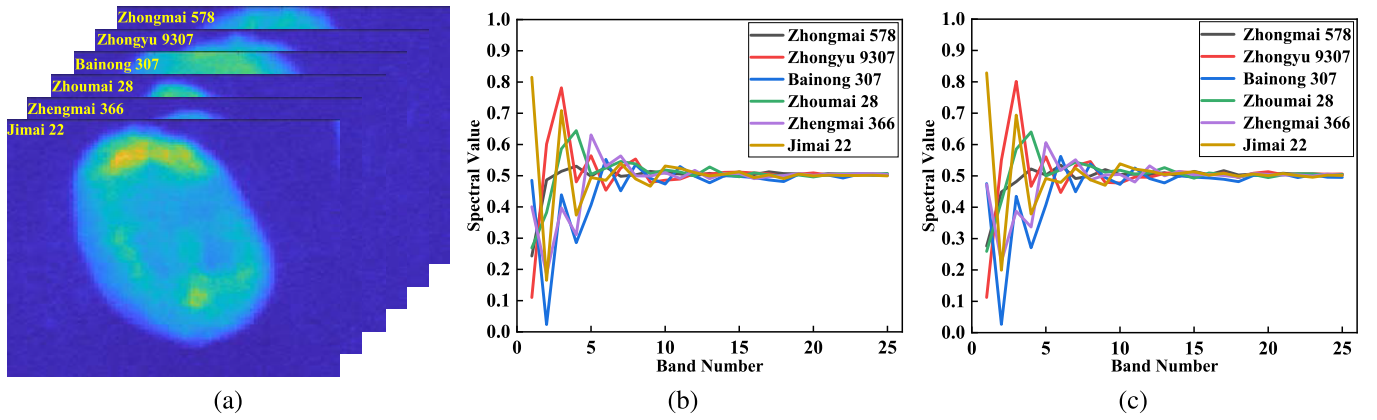
Fig. 1.   Spectral band of six wheat varieties, real, and fake spectra generated by the SSGAN. It is worth noting that we only exhibit representative bands of six wheat varieties. (a) 60th band in the hyperspectral image of wheat varieties with a resolution of 80×100 and 128 bands. (b) Spectral values correspond to each band of the six wheat varieties after dimensionality reduction by the principal component analysis (PCA). (c) Real spectra of six wheat varieties correspond to the fake spectra generated by the SSGAN.

data to harness the benefits yielded by increased data volumes. Traditionally, researchers primarily employed geometric transformations, color adjustments, and noise addition for data augmentation [13], [14]. Additionally, some researchers performed data augmentation using priori knowledge or transforming multiple samples [15], [16]. However, traditional data augmentation methods only amplify the data based on external sample attributes, without fully considering the intrinsic data characteristics. Recently, the prominence of conditional GAN [17], CycleGAN [18], StyleGAN [19], diffusion model [20], and Wasserstein GAN (WGAN) [21] in data augmentation has risen due to their capacity to capture inherent distributional properties during learning and inference. Hyperspectral images have higher dimensions, and high-dimensional data leads to more complex data distributions, increasing the risk of pattern collapse. Notably, WGAN uses Wasserstein distance as the loss function can better solve the pattern collapse issue. Hence, this study aims to investigate the potential of WGAN in generating hyperspectral wheat seed images.

Regarding hyperspectral seed image classification, various methods including traditional, machine learning, and deep learning approaches have found application. Traditional discrimination methods relied on manually crafted features for hyperspectral image classification [22], [23], whereas machine learning techniques utilized hand-crafted or semi-automatically extracted features for classification. In contrast, deep learning methods automatically extract valuable features, resulting in enhanced accuracy for hyperspectral image classification [24]. Traditional convolutional neural networks predominantly concentrate on image texture features, often overlooking the spatial and spectral features of hyperspectral images. With the evolution of CNNs, researchers [25], [26] utilized the deep network structure of CNNs to extract hierarchical deep spatial features from hyperspectral images. Recent deep learning methods [27], [28] leverage both spectral and spatial features to boost classification accuracy. For instance, Roy et al. [27] employed a 3D-CNN to extract spatial and spectral features, along with a 2D-CNN for spatial features. In our recent work [28], we employed 3-D convolution mod-

ules to extract spatial and spectral features of hyperspectral corn seed images, along with 2-D convolution modules for spatial and textural features. Thus, the effective utilization of spatial, spectral, and textural features within hyperspectral images remains pivotal in enhancing the performance of hyperspectral seed variety identification.

In this work, we propose a generate adversity-driven cross-aware network, called GACNet, for identify wheat varieties. The GACNet employs a GAN to generate realistic data and an effective discriminatory network model for hyperspectral wheat variety identification. In the data augmentation stage, our focus lies in the collaborative work between the generator with multilayer perceptrons, the discriminator with multilayer perceptrons, and the classifier to produce more realistic hyperspectral wheat images. In the wheat variety identification stage, we explore how to fully utilize 3-D convolution, 2-D convolution, and attention modules to construct an efficient identification model. In Fig. 1, we showcase the real spectra of the different wheat varieties after dimensionality reduction, along with their corresponding fake spectra generated by the semi-supervised GAN. Due to space constraints, only six varieties are shown in Fig. 1. The fake spectra generated by our proposed method closely resemble the real spectra, indicating the effectiveness of our data generation approach. These generated samples significantly contribute to the performance of our proposed cross-aware attention network (CAANet), leading to higher classification accuracy and improved robustness in identifying wheat varieties.

The main contributions of this work are summarized as follows.

1) We construct a comprehensive HWVD, comprising 4560 samples of 19 wheat varieties. Among these, 2280 samples are real hyperspectral wheat samples, and the remaining 2280 samples are generated hyperspectral wheat samples. Each wheat variety includes 120 samples with 128 bands, and the spatial resolution of each sample is 80×110. This dataset not only facilitates the application of deep learning methods in hyperspectral agricultural seed identification but also holds significant

value for screening and breeding different wheat varieties.

2) We introduce a semi-supervised GAN (i.e., SSGAN) that utilizes generator and discriminator networks with multilayer perceptrons to generate hyperspectral wheat images as realistically as possible. This is achieved by constantly gaming the spatial and spectral distribution characteristics between the input noises and the real samples.

3) We propose a CAANet that employs cross-cascaded three 3-D convolutional modules, three 2-D convolutional modules, and three attention modules. The network is specially designed to fully exploit the spatial, spectral, and textural features embedded in the hyperspectral wheat variety images to enhance the model's effectiveness and robustness in wheat variety identification.

II. BACKGROUND

This section provides an overview of the related work in data augmentation and hyperspectral seed identification. The details of each related work are as follows.

A. Data Augmentation

GANs have gained widespread adoption in diverse image-related tasks such as orientation estimation [29], image restoration [30], and data generation [31], primarily due to their exceptional data generation capabilities. In the domain of hyperspectral image augmentation and classification, GANs have emerged as a valuable tool. Zhu et al. [32] made a pioneering contribution by successfully applying GANs to hyperspectral image classification, utilizing a 1-D GAN as a spectral classifier and a 3-D GAN as a spatial–spectral classifier, thereby introducing a novel approach to hyperspectral image classification. Wang et al. [33] designed a GAN-assisted CapsNet for hyperspectral image classification, exploring the capabilities of a 1-D structure triple GAN in generating hyperspectral images and the classification capabilities of CapsNet. Zhang et al. [21] investigated the positive effects of GANs on smooth pre-training and classification performance in hyperspectral classification models. Xie et al. [34] proposed a self-spectral learning GAN for target detection of hyperspectral images, fully utilizing spatial and spectral features. Li et al. [35] introduced a sparse coding-inspired GAN for anomaly detection in hyperspectral images, effectively combining sparse coding and GANs to enhance detection performance. Meanwhile, Wang et al. [36] extended the application of GANs in hyperspectral images by proposing a frequency-to-spectral mapping GAN for anomaly detection. Inspired by the diffusion model, Chen et al. [20] propose a generative framework for hyperspectral image classification, which efficiently mines the distribution of highly and highly redundant data by iterating, denoising, and explicitly building the data generation process. Additionally, Yuan et al. [37] proposed an efficient and controllable framework for generating remote sensing pseudo-samples based on diffusion models. However, the diffusion model is dedicated to high-dimensional

images with a small amount of data, and its performance in generating diverse samples is unsatisfactory. In summary, GANs have demonstrated wide utility in various hyperspectral image applications and have played a prominent and positive role in data augmentation and improving the performance of deep learning classification models. Their ability to provide more diverse and richer data has proven instrumental in enhancing the effectiveness and robustness of such models.

B. Identification Methods

Currently, hyperspectral image identification methods can be categorized into traditional techniques, machine learning techniques, and deep learning techniques.

Traditional methods depend on the spectral characteristic to employ manually extracted features for hyperspectral image classification. For instance, He et al. [6] proposed a manual feature extraction method based on multiscale covariance maps that absorb and integrate spatial and spectral feature information. Peng et al. [38] introduced a self-paced joint sparse representation model using a self-paced learning strategy to learn the weights of neighboring pixels, outperforming the joint sparse representation model in terms of accuracy and robustness. Additionally, Gan et al. [39] developed a multifeature kernel sparse representation method to address the limitations of linear sparse models in handling hyperspectral images with highly nonlinear distributions. Fang et al. [40] proposed a local covariance matrix method for characterizing the correlation between different spectral bands and spatial information in the scene. Gradually, methods such as covariance pooling [41], low-rank regularization [42], and covariance metric [43] have also been applied to hyperspectral image classification. However, such methods are constrained by their inadequate feature extraction performance and feature redundancy.

Machine learning methods rely on manually or automatically extracted features for hyperspectral image classification. Gao et al. [44] proposed a method that integrates support vector machines and a probabilistic joint sparse model to compute the posterior probabilities of test samples, outperforming most classification methods on three datasets. Zabihzadeh et al. [45] introduced a sparse Bayesian approach for metric learning in the potential space, which is widely used in hyperspectral image classification, handwritten digits, and face recognition. Okwuashi et al. [8] designed a deep support vector machine for hyperspectral image classification, verifying various kernel functions and achieving higher classification accuracy. Su et al. [46] proposed a KNN method based on random subspace for hyperspectral image classification, utilizing a shape adaptive neighborhood constraint in the framework of random subspace integration to improve the traditional KNN's classification ability. Almoujahed et al. [47] classified hyperspectral images of health and scab infection in the canopy of eight wheat varieties using SVM, obtaining good classification accuracy. However, this method struggled to achieve higher accuracy with increased samples. In summary, machine learning methods face challenges in obtaining good classification results for large-scale spectral data due to their insufficient feature extraction capability.
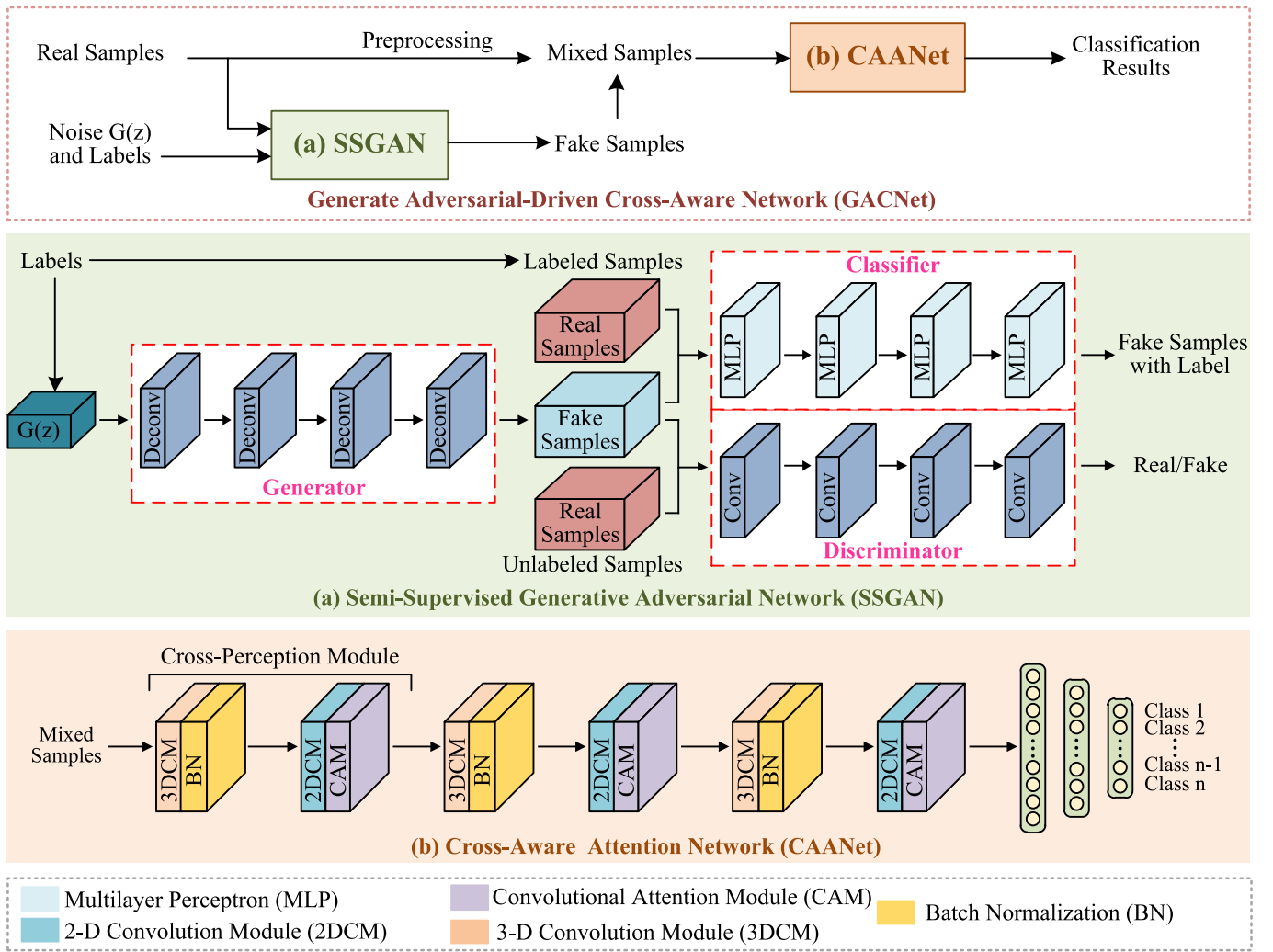
Fig. 2. Flowchart of the proposed GACNet for hyperspectral wheat variety identification. In our identification method, it includes an SSGAN and a CAANet. In our SSGAN, it consists of a generator, a discriminator, and a classifier. Our SSGAN is mainly used to generate hyperspectral wheat images as actual as possible by learning the distribution properties between noises and real samples. In our CAANet, it includes the 3DCM and the 2DCM with attention. Our CAANet utilizes cross-cascade learning of 3-D and 2-D convolution modules to fully exploit spectral, spatial, and textural features of hyperspectral wheat images. (a) SSGAN. (b) CAANet.

Deep learning methods typically employ unsupervised or semi-supervised strategies to automatically extract valuable features to enhance the effectiveness and robustness of discriminative models. These methods have found extensive applications in various image-related tasks, including underwater image enhancement [48], shadow removal [49], low-light image enhancement [50], super-resolution [51], [52], [53], [54], and remote sensing application [55], [56]. In hyperspectral image classification [57], [58], deep learning can fully exploit and utilize pixel dependence for seed identification. For instance, Sellami et al. [59] designed a semi-supervised 3-D convolutional neural network (CNN) for hyperspectral image classification, effectively reducing data dimensions and achieving good results in small sample classification. Haut et al. [60] introduced a deep pyramidal residual network that improved the utilization of spectral and spatial features in hyperspectral images. Zhang et al. [61] developed a low-weight 3-D CNN with transfer learning for hyperspectral image clas-

sification, effectively utilizing spatial, spectral, and textural features to improve classification performance. Cao et al. [62] leveraged the advantages of activation learning and deep learning to improve the classification ability of traditional CNN for hyperspectral images. Zhang et al. [63] designed a spectral–spatial fractal residual to further explore the effects of spectral and spatial features on classification performance. Furthermore, Zhang et al. [28] explored the effectiveness of spatial, spectral, and textural features for hyperspectral maize seed identification. Recently, Huang et al. [64] have fully integrated transformers with spectral and spatial features for hyperspectral image classification. Additionally, Zhang et al. [65] presented a cross-domain self-taught network that utilizes an attention module and four residual modules to extract deep spatial and spectral features. Overall, efficiently utilizing spatial, spectral, and textural features in hyperspectral images holds great significance for improving the performance of deep learning methods.

III. METHODOLOGY

In our work, we propose a hyperspectral wheat variety identification method called GACNet in Fig. 2, which comprises an SSGAN for data augmentation and a CAANet for seed variety identification. Notably, the data generated by the SSGAN is used to improve the accuracy of identifying hyperspectral wheat varieties.

A. Semi-Supervised GAN

Insufficient hyperspectral data poses serious challenges to the identification of wheat varieties, leading to limitations in the robustness and classification performance of the identification models. To address this issue, we explore data augmentation techniques for hyperspectral wheat images to enhance the effectiveness and robustness of the identification model. Inspired by WGAN [21], we propose an SSGAN for hyperspectral data augmentation by leveraging the correlation between individual bands in hyperspectral images.

The detailed workflow of the SSGAN is in Fig. 2(a). It includes a generator with four deconvolution layers, a discriminator with four convolution layers, and a classifier with four multilayer perceptrons. Initially, the generator generates fake samples by mapping random noise from a potential spatial mapping distribution. Subsequently, the discriminator is employed to distinguish between real and fake samples generated by the generator. If the generated sample meets the convergence criteria of a real sample in the discriminator, it is accepted as output. Otherwise, the network continues training the generator and discriminator until convergence is achieved. In other words, the generator inputs noise and class labels and innovatively introduces conditional information to guide the generator, so that the generator can generate images that meet certain conditions or attributes and forcibly connects the class labels input by the generator with the classifier to label the generated images. Finally, the classifier assigns a specific category label to the samples generated by the generator. The loss functions employed to train the discriminator, generator, and classifier are described as follows.

1) Discriminator Loss: The discriminator computes the distance between the real distribution and the generated distribution. To avoid gradient explosion or vanishing, we confine the discriminator's gradient variance between real and generated instances using a gradient penalty, which imposes the Lipschitz condition [66] on the discriminator. The gradient penalty interpolates between the real and generated samples. The interpolated sample is defined as

$$\hat{x} = wx + (1 - w)\tilde{x} \tag{1}$$

where x represents the real sample, $\tilde{x}$ represents the sample produced by the generator G processing the noise z, and $\hat{x}$ denotes the interpolated sample. Furthermore, the weighting factor w, which lies in the range [0,1], determines the extent of interpolation between the real and generated samples. Subsequently, the interpolation points are utilized to compute the gradient norm, which is defined as

$$\|G_N\|_2 = \sqrt{\sum_{i=1}^{B} (\nabla D(\hat{x}))^2 + \delta} \tag{2}$$

where B is the batch size, $\nabla D(\hat{x})$ is the interpolation point between the real and generated samples, and D is the discriminator. Besides, δ is a constant that approximates 0 to avoid the error problem of the gradient norm. Whereafter, we take the interpolation of the comparison between the gradient norm and the target norm of the preset term as the gradient penalty term, which is expressed as

$$G_{\text{NP}} = w_{\text{GP}} \times (\|G_N\|_2 - 1)^2 \tag{3}$$

where w_{GP} is the weight of the gradient penalty. The loss function of the discriminator combines the error and gradient interpolation between the real sample distribution and the generated sample distribution, which is expressed as

$$\text{Loss}_D = - \mathop{\mathbb{E}}_{x \sim P_r} [D(x)] + \mathop{\mathbb{E}}_{\tilde{x} \sim P_g} [D(\tilde{x})] + \mathop{\mathbb{E}}_{\hat{x} \sim P_{\hat{x}}} [G_{\text{NP}}] \tag{4}$$

where P_r and P_g are the distributions of real samples and generated samples, respectively, and $\mathbb{E}$ denotes the expectation operator.

2) Generator Loss: The generator's objective is to minimize the discrepancy between the distribution of real and generated samples. During the initial training stage, the discriminator can easily distinguish real samples from the generated ones, causing $1 - D(G(Z))$ to with $D(G(Z)))$ approaches 0. Hence, we maximize $D(G(z))$ during the training of the generator training, aiming to maximize the distance between the noise distribution and the generated samples' distribution. In this work, the generator's loss function is defined as

$$\text{Loss}_G = - \mathop{\mathbb{E}}_{z \sim P_z} [D(G(z))] \tag{5}$$

where z is random noise, and P_z is the distribution of random noise.

3) Classifier Loss: The classifier's task is to label the samples generated by the generator with labels corresponding to real wheat varieties. As the generator's samples are unlabeled, we introduce an additional classifier that receives the generated samples as input and assigns specific labels to them. In the training phase of the classifier, we utilize the Mixup data augmentation technique [67], which randomly applies linear weighted fusion to the features and labels of two samples to create a new sample with labels. This linear combination requires the model to consider both original samples' labels during classification. Implementing Mixup augments the diversity of training samples, enhancing the model's generalization across different samples. Therefore, the classifier's loss function is defined as

$$\text{Loss}_C = -\frac{1}{N}\frac{1}{C}\sum_{m=1}^{N}\sum_{n=1}^{C}(\lambda y_i + (1 - \lambda)y_j)\log(q_{\text{mn}}) \tag{6}$$

where λ is a randomly sampled weight from the beta distribution, controlling the degree of linear weighting between two samples and taking values in the range [0, 1]. N and C denote

the number and class of wheat samples, respectively. y_1 and y_2 represent the real labels of the first and second samples, respectively, and q_{mn} denotes the predicted probability of the nth category of the mth sample.

Finally, **the overall objective loss function** of our proposed SSGAN is defined as

$$\text{Loss}_T = \text{Loss}_D + \text{Loss}_G + \alpha \times \text{Loss}_C \tag{7}$$

where α is the weight used to control the importance of the classifier's loss in the overall optimization process.

B. Cross-Aware Attention Network

The 3-D convolution module [27], [28] primarily focuses on extracting spectral and spatial features from hyperspectral images, while the 2-D convolution module [27], [28] focuses on spatial and textural features. To fully leverage the benefits of both 3-D and 2-D convolution modules for hyperspectral feature extraction, we propose a CAANet for hyperspectral wheat variety identification. The architecture of the CAANet is shown in Fig. 2(b), which consists of three cascaded cross-perception modules and three fully connected layers. Each cross-perception module is composed of a 3-D convolution module, a 2-D convolution module, and an attention module. First, the 3-D convolutional module extracts spatial and spectral features from hyperspectral wheat images. Subsequently, the 2-D convolution module further captures spatial and textural features. Lastly, the attention module refines and optimizes the extracted features. The sequential arrangement of these three modules enables a more comprehensive exploitation of deeper features. Afterward, the fully connected layers maps the features extracted by the cross-perception modules to probability distributions for different categories. The first two fully connected layers employ the LeakyReLU activation function to introduce nonlinearity, effectively accommodating complex relationships between features and seed varieties. In the output layer, the softmax function is utilized to transform category scores into probability values for the final classification decision. The details of the 3-D convolution module, the 2-D convolution module, and the attention mechanism are described as follows.

1) Three-Dimensional Convolution Module: Hyperspectral images require capturing both spectral and spatial information encoded in multiple frequency bands. Therefore, the 3-D kernel is employed to generate a convolved feature map across multiple contiguous bands in the input layer, enabling the extraction of spatial dimensions and interspectral features. In the 3-D convolution operation, the activation function of obtaining the jth feature map at the spatial location (x, y, z) of the ith layer is formulated as

$$F_{i,j}^{x,y,z}$$
$$= f_a\left(\sum_{j=1}^{J_{i-1}} \sum_{l=0}^{L_i} \sum_{w=0}^{W_i} \sum_{h=0}^{H_i} w_{i,j}^{l,w,h} \times F_{i-1,j}^{x+l,y+w,z+h} + \delta_{i,j} \right) \tag{8}$$

where $f_a()$ is the LeakyRelu activation function, $w_{i,j}^{l,w,h}$ is the weight of the convolution kernel connected to the jth feature map at the position (l, w, h), J_{i-1} is the number of feature

maps of the $(i-1)$th layer, $\delta_{i,j}$ is the bias value of the jth feature map of the ith layer, L, W, and H are the dimensions, width, and height of the 3-D convolution kernel, respectively, and l, w, and h are the indexes in the three dimensions during the convolution process.

2) Two-Dimensional Convolution Module: For hyperspectral images, it is essential to extract both spatial and textural features from each band. To achieve this, we first reshape the 3-D hyperspectral data into a 2-D format by flattening it. The 2-D convolution operation obtains a feature map by calculating the dot product between the data and the convolution kernel to compute spatial and textural features of hyperspectral images of wheat. LeakyRelu is used as the activation function. Mathematically, the process is defined as

$$F_{i,j}^{x \times y, z}$$
$$= f_a\left(\sum_{j=1}^{J_{i-1}} \sum_{l \times w=0}^{(L \times W)_i} \sum_{h=0}^{H_i} w_{i,j}^{l \times w, h} \times F_{i-1,j}^{(x+l) \times (y+w), z+h} + \delta_{i,j} \right) \tag{9}$$

where $L \times W$ is the width of the reshaped hyperspectral data, H is the height of the reshaped hyperspectral data, $f_a()$ is the LeakyRelu activation function, J_{i-1} is the number of feature maps of the $(i-1)$th layer, $\delta_{i,j}$ is the bias value of the jth feature map of the ith layer, $w_{i,j}^{l \times w, h}$ is the weight of the convolution kernel connected to the jth feature map at the position $(l \times w, h)$, and J_{i-1} is the number of feature maps of the $(i-1)$th layer. Meanwhile, $l \times w$, and h are the indexes in the two dimensions during the convolution process.

3) Convolutional Attention Module: This module is designed to refine the spatial, spectral, and textural features extracted from the hyperspectral images. The convolutional attention module (CATM) includes the channel and spatial attention modules, as shown in Fig. 3. The channel attention module compresses the feature map in the spatial dimension to obtain a 1-D vector. It employs average pooling and max pooling to aggregate the spatial information of the feature map and forward it to a shared network. The spatial dimension of the input feature map is compressed and the elements are summed up to produce the channel attention feature map. The average pooling operation and max polling operation of the input matrix $\mathbf{X}_{i,j}$ with a dimension of $H \times W$ are expressed as

$$F_{avg}^c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbf{X}_{i,j} \tag{10}$$

and

$$F_{max}^c = \max(\mathbf{X}_{i,j}), i \in \{1, \ldots, H\}, \quad j \in \{1, \ldots, W\} \tag{11}$$

where max() is the max pooling function. Therefore, we can redefine the expression of channel attention as

$$F_{CAM} = \text{Sigmoid}\left(\text{MLP}\left(F_{avg}^c \right) + \text{MLP}\left(F_{max}^c \right) \right) \odot F \tag{12}$$

where F is the input feature, $\odot$ is an element-by-element multiplication operation, MLP() is a multilayer perceptron. The spatial attention module compresses in the channel domain and thus applies average pooling and max pooling in the channel
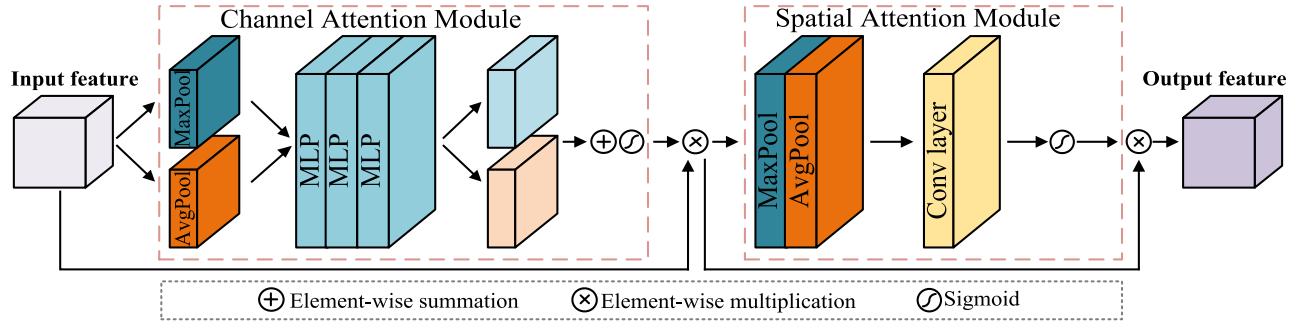
Fig. 3. Flowchart of the convolutional attention module. It consists of a channel attention module and a spatial attention module. The former module is mainly used to extract texture features of wheat hyperspectral images, and the latter is mainly used to extract spatial features of wheat hyperspectral images.

TABLE I
DETAILS OF THE RAW AND GENERATED WHEAT HYPERSPECTRAL SAMPLES

| Wheat lines	Zhengmai lines			Bainong lines				Zhoumai lines		Other lines									
Wheat variety	Zhengmai 101	Zhengmai 366	Zhengmai 7698	Bainong 207	Bainong AK58	Bainong 307	Bainong 4199	Zhoumai 28	Zhoumai 36	Jimai 22	Xinmai 32	Zhongyu 9307	Zhongmai 578	Zhongmai 18	Shangmai 167	Xinong 511	Yuyuan 916	Luomai 28	Fengde 21
Origin	Henan	Henan	Henan	Henan	Henan	Henan	Henan	Henan	Henan	Shandong	Henan	Henan	Henan	Henan	Henan	Shanxi	Henan	Henan	Henan
Real number	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120
Generated number	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120
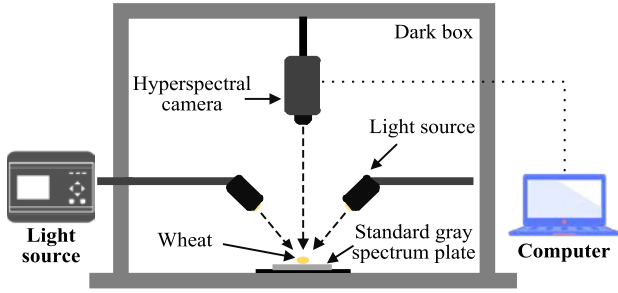
Fig. 4. Acquisition device of hyperspectral imaging system.

dimension. Max pooling extracts the maximum value in the channel, while average pooling computes the average value in the channel. Afterward, we concatenate the compressed feature maps and pass them to another convolution layer. The mathematical expression of the spatial attention module is

$$F_{\text{SAM}} = \text{Sigmoid}\Big(\text{Conv}\Big(\text{Concat}\Big(F_{\text{avg}}^s, F_{\text{max}}^s\Big)\Big)\Big) \odot F. \quad (13)$$

Finally, CATM performs an elementwise multiplication of the channel attention feature map and the spatial attention feature map to obtain the final attention feature map, which is expressed as

$$F_{\text{CATM}} = F_{\text{SAM}}(F_{\text{CAM}}(F)). \quad (14)$$

4) Objective Loss: As a multiple-class wheat variety identification problem, the cross-entropy is adopted as the loss function [68]. The cross-entropy calculates the distance between two probability distributions and determines the degree of similarity between two probability distributions by the magnitude of the distance. The formula for cross-entropy loss is expressed as

$$\text{Loss}_{\text{CE}} = H(p) + D_{\text{KL}}(p||q). \quad (15)$$

In (15), $H(p)$ is the entropy of the true distribution calculated as $H(p) = -\sum_{i=1}^{N} p(x_i) \log p(x_i)$, where $p(x_i)$ is the distribution of real labels. $D_{\text{KL}}(p||q)$ is the KL scatter of the true label distribution and the predicted probability distribution, which is calculated as

$$D_{\text{KL}}(p||q) = \sum_{i=1}^{N} p(x_i)\big[\log p(x_i) - \log q(x_i)\big] \quad (16)$$

where $q(x_i)$ is the predicted probability distribution, and N is the total class of wheat varieties. The Loss_{CE} of the proposed discrimination model CAANet is redefined as

$$
\begin{aligned}
\text{Loss}_{\text{CE}} &= -\sum_{i=1}^{N} p(x_i) \log p(x_i) + \sum_{i=1}^{N} p(x_i)\big[\log p(x_i) - \log q(x_i)\big] \\
&= -\sum_{i=1}^{N} p(x_i) \log q(x_i). \quad (17)
\end{aligned}
$$

IV. EXPERIMENT AND ANALYSIS

In this section, we first introduce the constructed hyperspectral wheat variety dataset (HWVD), followed by the implementation details of the proposed CAANet, the evaluation metrics, the identification results, and the ablation analysis.

A. Hyperspectral Wheat Variety Dataset

We created an HWVD consisting of 4560 samples from 19 wheat varieties, including 2280 real hyperspectral wheat samples and 2280 synthetic hyperspectral wheat samples generated using our proposed SSGAN. Table I provides details about the hyperspectral images of the 19 wheat varieties in the dataset.
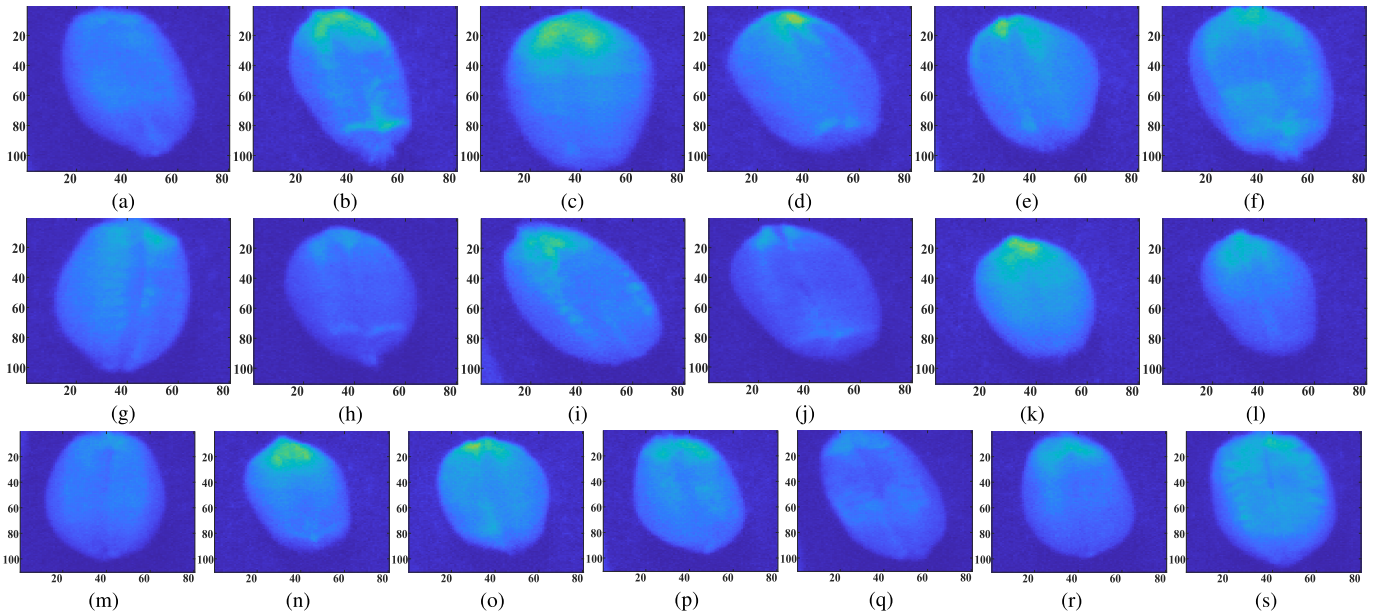
Fig. 5. 75th spectral band of 19 wheat seeds. (a)75th spectral band of Zhengmai 101. (b) 75th spectral band of Zhengmai 366. (c) 75th spectral band of Zhengmai 7698. (d) 75th spectral band of Bainong 207. (e) 75th spectral band of Bainong AK58. (f) 75th spectral band of Bainong 307. (g) 75th spectral band of Bainong 4199. (h) 75th spectral band of Zhoumai 28. (i) 75th spectral band of Zhoumai 36. (j) 75th spectral band of Jimai 22. (k) 75th spectral band of Xinmai 32. (l) 75th spectral band of Zhongyu 9307. (m) 75th spectral band of Zhongmai 578. (n) 75th spectral band of Zhongmai 18. (o) 75th spectral band of Shangmai 167. (p) 75th spectral band of Xinong 511. (q) 75th spectral band of Yuyuan 916. (r) 75th spectral band of Luomai 28. (s) 75th spectral band of Fengde 21.
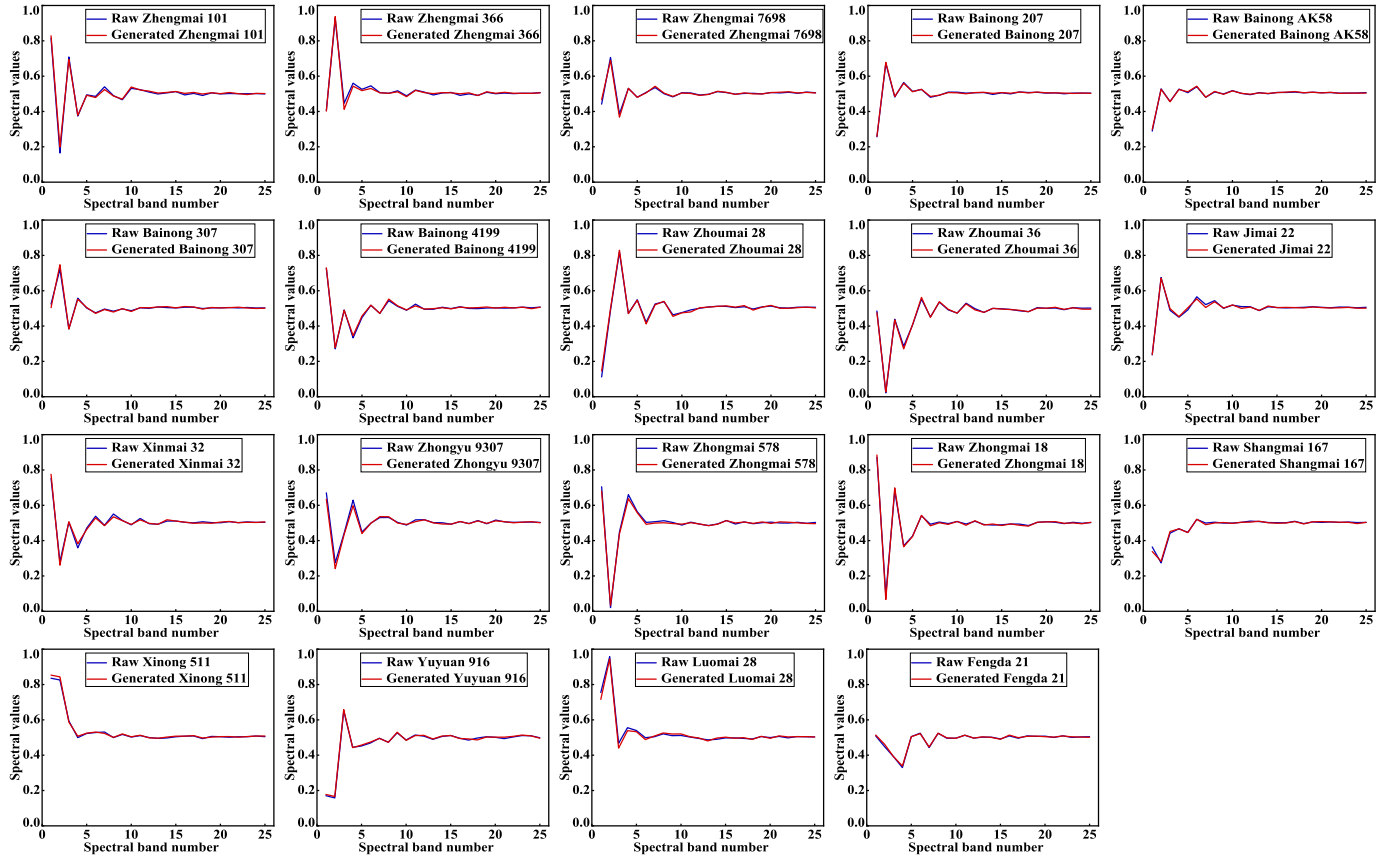

Fig. 6. Average spectra of the raw and generated wheat hyperspectral samples by SSGAN. Notably, the average spectra of samples generated by our proposed SSGAN fit the average spectra of the raw samples better.

For the 2280 real samples of wheat varieties, we used the SOC 710 portable visible/near-infrared imaging spectrometer to capture spectra from the dorsal and ventral surfaces of wheat seeds. The setup of the acquisition device of the imaging system is shown in Fig. 4. The spectrometer's spectral range is 400–1000 nm with a spectral resolution of 4.6875 nm, providing 128 bands. The image resolution is 696 × 520 pixels. We utilized the ENVI software to convert

TABLE II
EVALUATION MATRIX OF IDENTIFICATION RESULTS OF WHEAT HYPERSPECTRAL IMAGES

Item		Reference		Total	Evaluation metrics
		True	False		
Identification result	True	TP	FN	TP + FN	$A_R = \text{TP}/(\text{TP} + \text{FN})$
	False	FP	TN	FP + TN	-
Total		TP + FP	FN + TN	N	-
Evaluation metrics		$A_P = \text{TP}/(\text{TP} + \text{FP})$	-	-	$\text{F1} = (2 \times A_R \times A_P)/(A_R + A_P)$ $A_T = (\text{TP} + \text{TN})/N$

the raw hyperspectral images into a standard data format and then extracted regions of interest with a resolution of 80×110 pixels. Specifically, each real wheat variety comprises 120 samples with 128 bands, and the spatial resolution of each sample is 80×110 pixels. As a demonstration, Fig. 5 shows the 75th band of the 19 wheat varieties in the dataset, namely Zhengmai 101, Zhengmai 366, Zhengmai 7698, Bainong 207, Bainong AK58, Bainong 307, Bainong 4199, Zhoumai 28, Zhoumai 36, Jimai 22, Xinmai 32, Zhongyu 9307, Zhongmai 578, Zhongmai 18, Shangmai 167, Xinong 511, Yuyuan 916, Luomai 28, and Fengde 21.

For the generated 2280 samples of wheat varieties, we first employed the PCA to preprocess the raw wheat hyperspectral image by compressing it from 128 bands to 25 bands. Subsequently, SSGAN inputs noise and labels into the generator to generate fake samples, while it inputs real samples into the discriminator to determine whether the generated fake sample conforms to the data distribution characteristics of the real sample. Finally, the classifier labels the fake samples to their corresponding wheat variety categories. Fig. 6 illustrates the average spectrum of the raw and generated wheat hyperspectral samples by the SSGAN. It is worth noting that the samples generated by our proposed SSGAN show exceptional agreement with the hyperspectral images of 19 wheat varieties.

B. Implementation Details

Our GACNet is trained and tested on the constructed HWVD presented in Section IV-A. Specifically, it is tested on 2280 samples of real wheat varieties, 2280 samples of generated wheat varieties, and 4560 samples of a mixture of real and generated wheat varieties, respectively. The training and testing data are split by a 4:1 ratio. The GACNet is implemented on a Windows 11 PC with an AMD Ryzen 7 3700X 8-Core CPU running at 3.6 GHz, 48 GB memory, an NVIDIA GeForce GTX 3060 GPU, and the PyTorch Deep Learning Toolbox. The GACNet consists of two main subnetworks: a data augmentation subnetwork (SSGAN) and a wheat variety identification subnetwork (CAANet). For the training of the SSGAN, the batch size is set to 8, and the optimizer uses RMSprop with a smoothing parameter of 0.99 and a constant parameter of $\varepsilon = 10^{-8}$ to prevent the denominator from being zero. The learning rate is set to 5×10^{-5}, and the network is trained for 1000 iterations. For the training of the CAANet, the batch size is set to 16, the optimizer uses Adam with attenuation rate parameters of 0.9 and 0.999, the learning rate is set to 5×10^{-3}, and the network is trained for 200 iterations. We also introduced the Dropout mechanism in the fully connected layer of CAANet to prevent overfitting by setting 20% of the nodes to 0.

C. Evaluation Metrics

To comprehensively evaluate the performance of our GACNet for identifying hyperspectral images of wheat varieties, we used four commonly- metrics: precision (A_P), recall (A_R), F-score (F_1), accuracy (A_T), and kappa coefficient (K_A). For the identification results, it includes four cases: true positive (TP), false positive (FP), false negative FN, and true negative (TN). A_P represents the proportion of samples with correct prediction results in the samples that have been predicted to be positive. A_R denotes the proportion of correctly predicted positive samples out of all positive samples. F_1 is the reconciled value of the precision A_P and recall A_R, which can better compensate for the one-sidedness of evaluating the prediction results of the model only by the precision A_P or recall A_R. A_T represents the proportion of the number of correctly predicted samples in the total number of test samples. K_A is a metric to measure classification accuracy, and the higher its value, the better the classification result. Table II provides the evaluation matrix of identification results of wheat hyperspectral images.

D. Identification Results

We conducted extensive experiments and analysis on the constructed dataset to evaluate the identification performance of our GACNet for wheat varieties. The evaluation was performed on the only real dataset, only generated dataset, and mixed dataset of real and generated samples of wheat varieties. Table III presents the discriminative results of our GACNet and nine comparison methods on the three experimental datasets. The comparison methods include **machine learning methods**: KNN [10], DT [70], RFA [69], and FSVM [2]; and **deep learning methods**: MSDNet [71], 3DCNN [72], HybridSN [27], A^2S^2K-ResNet [73], SSTNet [28], and MorphFormer [74]. For the comparison methods, we implemented the source code and parameters provided by the authors.

1) Quantitative Evaluation on the Real Dataset: As shown in Table III, traditional machine learning methods such as KNN [10], RFA [69], and DT [70] achieve relatively poor identification performance. FSVM [2], a machine learning method that utilizes spectral and spatial features, outperforms KNN [10], RFA [69], and DT [70]. Among the deep learning methods, MSDNet [71], 3DCNN [72], HybridSN [27], A^2S^2K-ResNet [73], and MorphFormer [74] demonstrate better identification ability by fully exploiting spectral and spatial features. SSTNet [28] and our GACNet, which consider texture features in addition to spectral and spatial features, exhibit even better identification performance. In comparison, our proposed GACNet outperforms all compared methods in terms of A_P, A_R, F1, A_T, and K_A scores, achieving the highest or

TABLE III

QUANTITATIVE SCORES OF THE IDENTIFICATION RESULTS OF OUR GACNET WITH THE COMPARED METHODS TESTED ON THE REAL, GENERATED, AND MIXED DATASETS. THE BEST RESULTS IN THE TABLE WE MARK IN RED, WHILE THE SECOND BEST RESULTS WE MARK IN BLUE

Methods	Real dataset					Generated dataset					Mixed dataset				
	$A_P \uparrow$	$A_R \uparrow$	F1 $\uparrow$	$A_T \uparrow$	$K_A \uparrow$	$A_P \uparrow$	$A_R \uparrow$	F1 $\uparrow$	$A_T \uparrow$	$K_A \uparrow$	$A_P \uparrow$	$A_R \uparrow$	F1 $\uparrow$	$A_T \uparrow$	$K_A \uparrow$
KNN [10]	0.9530	0.9500	0.9515	0.9518	0.9491	0.9493	0.9417	0.9455	0.9452	0.9421	0.9649	0.9631	0.9640	0.9638	0.9618
RFA [69]	0.9576	0.9544	0.9560	0.9561	0.9537	0.9419	0.9351	0.9385	0.9386	0.9352	0.9786	0.9774	0.9780	0.9781	0.9769
DT [70]	0.9576	0.9470	0.9523	0.9518	0.9491	0.9426	0.9386	0.9406	0.9408	0.9375	0.9543	0.9517	0.9530	0.9529	0.9502
FSVM [2]	0.9671	0.9633	0.9652	0.9649	0.9630	0.9642	0.9614	0.9628	0.9627	0.9606	0.9803	0.9783	0.9793	0.9792	0.9780
MSDNet [71]	0.9745	0.9689	0.9717	0.9715	0.9699	0.9610	0.9558	0.9584	0.9583	0.9560	0.9750	0.9722	0.9736	0.9737	0.9722
3DCNN [72]	0.9801	0.9757	0.9779	0.9781	0.9769	0.9702	0.9569	0.9635	0.9627	0.9606	0.9811	0.9791	0.9801	0.9803	0.9792
HybridSN [27]	0.9618	0.9528	0.9573	0.9583	0.9560	0.9594	0.9489	0.9541	0.9539	0.9514	0.9818	0.9786	0.9802	0.9803	0.9792
A^2S^2K-ResNet [73]	0.9569	0.9365	0.9466	0.9474	0.9444	0.9569	0.9365	0.9466	0.9474	0.9699	0.9820	0.9784	0.9802	0.9803	0.9792
SSTNet [28]	0.9735	0.9681	0.9708	0.9715	0.9699	0.9689	0.9609	0.9649	0.9649	0.9630	0.9904	0.9898	0.9901	0.9901	0.9896
MorphFormer [74]	0.9860	0.9878	0.9866	0.9868	0.9861	0.9510	0.9611	0.9560	0.9561	0.9537	0.9934	0.9936	0.9935	0.9934	0.9934
GACNet	0.9878	0.9858	0.9868	0.9868	0.9861	0.9791	0.9767	0.9779	0.9781	0.9769	0.9946	0.9944	0.9945	0.9945	0.9942

approximate-highest scores of 0.9878, 0.9858, 0.9868, and 0.9861, respectively.

2) Quantitative Evaluation on the Generated Dataset: The identification results for the generated dataset are lower than those for the real dataset for all methods. This can be attributed to overfitting certain standard features, leading to less diverse feature representations in the generated samples compared to the real samples. Despite the reduction in discrimination results, our GACNet still exhibits the highest discrimination score among all methods.

3) Quantitative Evaluation on the Mixed Dataset: The identification scores for all methods on the mixed dataset are generally improved compared to the only real dataset and only generated dataset. RFA [69] and DT [70] consistently underperform other machine learning methods, likely due to the limited handling capability of highly correlated features of these tree-based machine learning algorithms. FSVM [2] achieves better performance by mapping nonlinear spectral data into high-dimensional space, and thus solving complex nonlinear problems with linear transformations. Among deep learning methods, A^2S^2K-ResNet [73] and MSDNet [71] achieve lower accuracy than other methods due to insufficient consideration of spatial information in hyperspectral images. 3DCNN [72] and HybridSN [27] fail to adequately incorporate attentional modules for dynamic feature selection, leading to unsatisfactory classification performance. In contrast, our proposed CAANet addresses these shortcomings by effectively integrating spatial and spectral information. This integration enables the model to learn more abstract and complex features, leading to superior identification performance compared to existing methods.

4) Quantitative Evaluation on the Houston Dataset [75]: To demonstrate the generalization performance of our method, we employed GACNet with the compared methods experimentally evaluated on the public Houston dataset [75]. The hyperspectral image has a resolution size of 340×1905 with 144 spectral bands, wavelengths of 0.38–1.05 *um*, and a spatial resolution of 2.5 meters per pixel. It contains 15 different classes and consists of disjoint training and test samples. From Table IV, our GACNet outperforms the compared methods in terms of classification performance for the Houston dataset. Due to limited space, we only present the test sample in Fig. 7(a) and the test results of our method in Fig. 7(b).
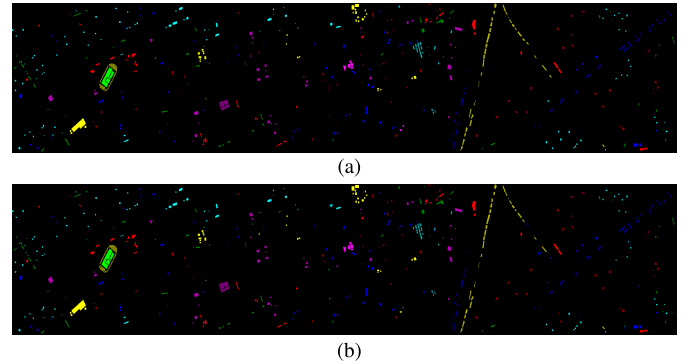

(a)

(b)

Fig. 7. Predicted results of our method experiments on the Houston dataset. (a) Test samples. (b) Predicted results of our method.

TABLE IV

QUANTITATIVE SCORES OF THE IDENTIFICATION RESULTS OF OUR GAC-NET WITH THE COMPARED METHODS TESTED ON THE HOUSTON DATASET [74]

Methods	Houston dataset				
	$A_P \uparrow$	$A_R \uparrow$	F1 $\uparrow$	$A_T \uparrow$	$K_A \uparrow$
KNN [10]	0.6422	0.7093	0.6563	0.7188	0.6957
RFA [70]	0.7789	0.9113	0.8399	0.7655	0.7462
DT [71]	0.4383	0.7072	0.5412	0.5954	0.5622
FSVM [2]	0.7222	0.6741	0.6973	0.6994	0.6748
MSDNet [72]	0.7322	0.7891	0.7596	0.7651	0.7470
3DCNN [73]	0.7469	0.8276	0.7852	0.7893	0.7732
HybridSN [27]	0.7919	0.8726	0.8303	0.8125	0.7983
A^2S^2K-ResNet [74]	0.7003	0.8107	0.7516	0.7606	0.7408
SSTNet [28]	0.7995	0.8430	0.8207	0.8225	0.8082
MorphFormer [75]	0.8606	0.9028	0.8812	0.8972	0.8700
GACNet	0.9404	0.9480	0.9442	0.9454	0.9409

Overall, our GACNet has good discrimination performance for both our constructed and publicly available datasets.

Fig. 8 illustrates the training loss and test accuracy of our GACNet on real, generated, and mixed datasets. Fig. 8(a) and (b) shows the results on real dataset. As shown in Fig. 8(a), our GACNet demonstrates faster convergence in the initial 50 epochs, followed by a stabilization of the loss from epochs 50–180. After 180 epochs, the model's loss continues to stabilize and eventually converges to 0. Fig. 8(b) shows that the test accuracy does not achieve satisfactory results due to the presence of more categorization categories and a limited number of training data, leading to difficulties for the model in learning more diverse features. Fig. 8(c) and (d) shows the results on generated dataset. As shown in Fig. 8(c), our GACNet converges rapidly before reaching 40 epochs, and it
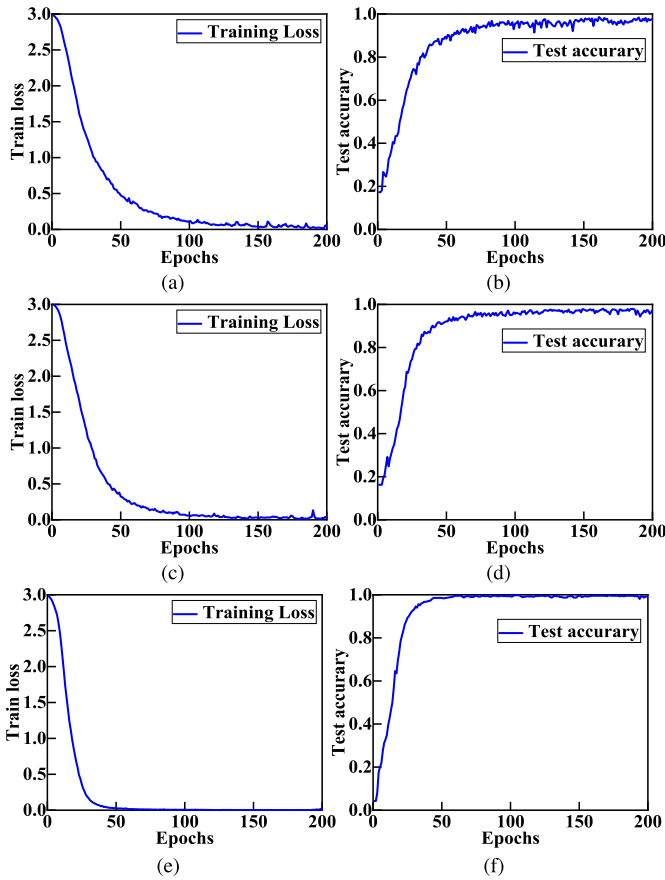
Fig. 8. Training loss and test accuracy of the GACNet experiments on the real, generated, and mixed dataset. (a)–(f) Loss and accuracy convergence versus the number of epochs of our GACNet training on the real, generated, and mixed datasets.

stabilizes and delivers better results after 150 epochs. However, the lower accuracy of the test set depicted in Fig. 8(d) suggests that the network may be overfitting to certain samples from the same class. Compared to Fig. 8(a)–(d), Fig. 8(e) and (f) demonstrates the faster convergence and higher test accuracy of our GACNet experiments on the mixed dataset. Overall, our analysis indicates that GACNet exhibits swift convergence performance and commendable discrimination capability.

Additionally, Fig. 9 shows the confusion matrices of the GACNet experiments conducted on real, generated, and mixed datasets. There are 24 testing data for each variety in both the real and generated datasets, and 48 testing data for each variety in the mixed dataset. In Fig. 9(a), the CACNet achieves an impressive 100% accuracy in identifying 15 wheat seeds and maintains an accuracy of over 90% for 4 other types of wheat seeds on the real dataset. Similarly, Fig. 9(b) illustrates that our CACNet achieves 100% accuracy in identifying 13 wheat seeds and maintains an accuracy of over 90% for 6 other types of wheat seeds on the generated dataset. Fig. 9(c) showcases our CACNet's exceptional performance, achieving 100% accuracy in identifying 14 wheat seeds and achieving over 95% accuracy for 5 other types of wheat seeds on the mixed dataset. Notably, the classification accuracies of all categories show significant improvements, providing compelling evidence that our CACNet effectively enhances the sample diversity and thereby improves the overall identification accuracy.

TABLE V
PERFORMANCE ANALYSIS OF EACH ABLATION MODEL EXPERIMENT IN OUR GACNET. THE BEST RESULTS IN THE TABLE WE MARK IN RED, WHILE THE SECOND BEST RESULTS WE MARK IN BLUE

Ablated models	$A_P \uparrow$	$A_R \uparrow$	F1 $\uparrow$	$A_T \uparrow$	$K_A \uparrow$
-w/o SSGAN	0.9791	0.9767	0.9779	0.9781	0.9769
-w/o Mixup	0.9810	0.9816	0.9712	0.9813	0.9803
-w/o 3DCM	0.9635	0.9619	0.9627	0.9627	0.9606
-w/o 2DCM	0.9758	0.9696	0.9727	0.9726	0.9711
-w/o CAM	0.9872	0.9864	0.9868	0.9868	0.9861
GACNet (Full model)	0.9946	0.9944	0.9945	0.9945	0.9942

E. Ablation Study

To assess the effectiveness of each module in our proposed GACNet for wheat variety identification, we conducted ablation studies. Specifically, we examined the performance of the following configurations: 1) the GACNet without the SSGAN (-w/o SSGAN); 2) the GACNet without the Mixup (-w/o Mixup); 3) the GACNet without the 3-D convolution module (-w/o 3DCM); 4) the GACNet without the 2-D convolution module (-w/o 2DCM); and 5) the GACNet without the convolutional attention module (-w/o CAM).

Table IV presents the quantitative classification results of the ablated models on the HWVD. Upon analyzing the results from Table V, the following observations can be made.

1) -w/o SSGAN: Removing the semi-supervised GAN from our GACNet leads to a reduction in classification performance. Notably, the ablation experiment was implemented without generated dataset. The main reason behind this is the limited generalization performance of the model due to insufficient samples. Notably, three subsequent ablation experiments were implemented on mixed dataset.

2) -w/o Mixup: The classification performance of our GACNet is weakened by removing the Mixup. The Mixup technology randomly applies linear weighted fusion to the features and labels of two samples to create a new sample with labels.

3) -w/o 3DCM: The classification performance of our GACNet is significantly affected by removing the 3-D convolution module. The 3DCM is crucial for extracting spatial and spectral features of wheat hyperspectral images.

4) -w/o 2DCM: The classification performance of our GACNet is diminished when excluding the 2-D convolution module. The 2DCM plays a vital role in capturing spatial and textural features of wheat hyperspectral images.

5) -w/o CAM: Removing the attention module from our GACNet results in decreased classification performance. The attention module is primarily used to re-refine features, but its impact on GACNet is relatively less pronounced.

In contrast, our full-model GACNet demonstrates the best results for each evaluation metric score, which further confirms that each module significantly contributes to the overall effectiveness of our GACNet.
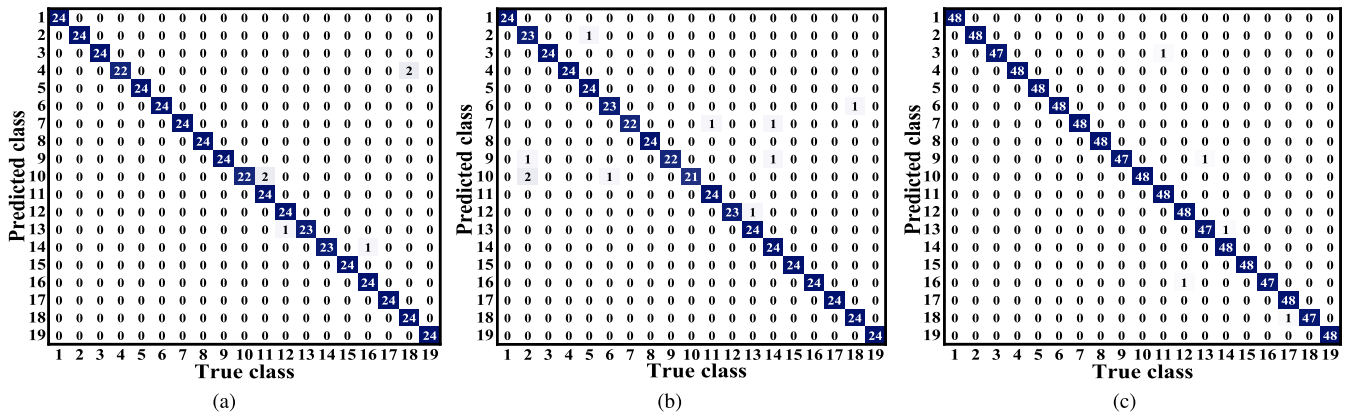
Fig. 9.   Confusion matrix of GACNet experiments on the real, generated, and mixed datasets. (a) Confusion matrix of GACNet experiments on the real dataset. (b) Confusion matrix of GACNet experiments on the generated dataset. (c) Confusion matrix of GACNet experiments on the mixed dataset.

V. CONCLUSION

In this article, we introduce GACNet, a generative adversarial-driven cross-aware network, for non-destructive hyperspectral wheat seed identification. The GACNet employs an unsupervised GAN (SSGAN) for data augmentation and a CAANet for seed variety identification. Thorough experimental validation demonstrates the superior performance of our GACNet in both data augmentation and seed variety identification tasks. Specifically, the SSGAN subnetwork effectively learns the distributional properties of real samples and generates realistic synthetic samples, thereby positively influencing GACNet's generalization performance. Meanwhile, the CAANet subnetwork demonstrates improved spatial, spectral, and texture extraction capabilities, further contributing to the enhanced classification performance of GACNet. Compared to nine state-of-the-art methods, our GACNet achieves superior quantitative evaluation scores in identifying wheat varieties, demonstrating the effectiveness and generalizability of our proposed network model.

Throughout the experimental design, we encountered several challenging issues. Hyperspectral data possesses high feature dimensions and multimodality, leading to potential inaccuracies in generated samples. Additionally, tuning the GAN becomes more complex due to the need for different hyperparameter settings for various hyperspectral image datasets and application scenarios. Moreover, factors like different years and origins may also impact the model's generalizability. To address these issues, we plan to expand the dataset to cover a more diverse range of seed varieties, origins, and ages, and explore the design of more efficient identification models in future research.

REFERENCES

[1] P. Duan, P. Ghamisi, X. Kang, B. Rasti, S. Li, and R. Gloaguen, "Fusion of dual spatial information for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7726–7738, Sep. 2021.

[2] S. Jin et al., "Spatial–spectral feature extraction of hyperspectral images for wheat seed identification," *Comput. Electr. Eng.*, vol. 101, Jul. 2022, Art. no. 108077.

[3] Y. Zhang, P. Duan, J. Mao, X. Kang, L. Fang, and P. Ghamisi, "Contour structural profiles: An edge-aware feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5545914.

[4] D. Zhu, B. Du, Y. Dong, and L. Zhang, "Spatial–spectral joint reconstruction with interband correlation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5529513.

[5] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3313–3332, Apr. 2023.

[6] N. He et al., "Feature extraction with multiscale covariance maps for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 755–769, Feb. 2019.

[7] G. Zhao, N. Li, B. Tu, G. Zhang, and W. He, "Density peak covariance matrix for feature extraction of hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 534–538, Mar. 2020.

[8] O. Okwuashi and C. E. Ndehedehe, "Deep support vector machine for hyperspectral image classification," *Pattern Recognit.*, vol. 103, Jul. 2020, Art. no. 107298.

[9] Q. Zhang, Q. Yuan, M. Song, H. Yu, and L. Zhang, "Cooperated spectral low-rankness prior and deep spatial prior for HSI unsupervised denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 6356–6368, 2022.

[10] M. Mailagaha Kumbure, P. Luukka, and M. Collan, "A new fuzzy k-nearest neighbor classifier based on the Bonferroni mean," *Pattern Recognit. Lett.*, vol. 140, pp. 172–178, Dec. 2020.

[11] X. Wang, K. Tan, P. Du, C. Pan, and J. Ding, "A unified multiscale learning framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4508319.

[12] Q. Liu et al., "Refined prototypical contrastive learning for few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5506214.

[13] L. Chen, G. Vivone, Z. Nie, J. Chanussot, and X. Yang, "Spatial data augmentation: Improving the generalization of neural networks for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5401711.

[14] W. Li et al., "EID-GAN: Generative adversarial nets for extremely imbalanced data augmentation," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 3208–3218, Mar. 2023.

[15] R. Rojas-Gómez, J. Yang, Y. Lin, J. Theiler, and B. Wohlberg, "Physics-consistent data-driven waveform inversion with adaptive data augmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[16] Q. Wang, J. P. Lohse, A. P. Doulgeris, and T. Eltoft, "Data augmentation for SAR sea ice and water classification based on per-class backscatter variation with incidence angle," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4205915.

[17] Y. Li et al., "StoryGAN: A sequential conditional GAN for story visualization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6322–6331.

[18] A. Almahairi, S. Rajeshwar, A. Sordoni, P. Bachman, and A. Courville, "Augmented CycleGAN: Learning many-to-many mappings from unpaired data," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 195–204.

[19] R. Abdal, Y. Qin, and P. Wonka, "Image2StyleGAN: How to embed images into the StyleGAN latent space?" in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4431–4440.

[20] N. Chen, J. Yue, L. Fang, and S. Xia, "SpectralDiff: A generative framework for hyperspectral image classification with diffusion models," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5522416.

[21] F. Zhang, J. Bai, J. Zhang, Z. Xiao, and C. Pei, "An optimized training method for GAN-based hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 10, pp. 1791–1795, Oct. 2021.

[22] H. Yu, H. Zhang, Y. Liu, K. Zheng, Z. Xu, and C. Xiao, "Dual-channel convolution network with image-based global learning framework for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[23] M. E. Paoletti, J. M. Haut, X. Tao, J. Plaza, and A. Plaza, "FLOP-reduction through memory allocations within CNN for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5938–5952, Jul. 2021.

[24] D. Ibañez, R. Fernandez-Beltran, F. Pla, and N. Yokoya, "Masked auto-encoding spectral–spatial transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5542614.

[25] S. K. Roy, J. M. Haut, M. E. Paoletti, S. R. Dubey, and A. Plaza, "Generative adversarial minority oversampling for spectral–spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5500615.

[26] S. L. Al-Khafaji, J. Zhou, X. Bai, Y. Qian, and A. W. Liew, "Spectral–spatial boundary detection in hyperspectral images," *IEEE Trans. Image Process.*, vol. 31, pp. 499–512, 2022.

[27] 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," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, Feb. 2020.

[28] W. Zhang, Z. Li, H.-H. Sun, Q. Zhang, P. Zhuang, and C. Li, "SSTNet: Spatial, spectral, and texture aware attention network using hyperspectral image for corn variety identification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[29] H.-H. Sun, Y. H. Lee, C. Li, L. F. Ow, M. L. M. Yusof, and A. C. Yucel, "The orientation estimation of elongated underground objects via multipolarization aggregation and selection neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[30] Q. Qi, K. Li, H. Zheng, X. Gao, G. Hou, and K. Sun, "SGUIE-Net: Semantic attention guided underwater image enhancement with multi-scale perception," *IEEE Trans. Image Process.*, vol. 31, pp. 6816–6830, 2022.

[31] J.-C. Burnel, K. Fatras, R. Flamary, and N. Courty, "Generating natural adversarial remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5513114.

[32] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5046–5063, Sep. 2018.

[33] X. Wang, K. Tan, Q. Du, Y. Chen, and P. Du, "Caps-TripleGAN: GAN-assisted CapsNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7232–7245, Sep. 2019.

[34] W. Xie, J. Zhang, J. Lei, Y. Li, and X. Jia, "Self-spectral learning with GAN based spectral–spatial target detection for hyperspectral image," *Neural Netw.*, vol. 142, pp. 375–387, Oct. 2021.

[35] Y. Li, T. Jiang, W. Xie, J. Lei, and Q. Du, "Sparse coding-inspired GAN for hyperspectral anomaly detection in weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5512811.

[36] D. Wang, L. Gao, Y. Qu, X. Sun, and W. Liao, "Frequency-to-spectrum mapping GAN for semisupervised hyperspectral anomaly detection," *CAAI Trans. Intell. Technol.*, vol. 10, pp. 1–16, Jan. 2023.

[37] Z. Yuan et al., "Efficient and controllable remote sensing fake sample generation based on diffusion model," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5615012.

[38] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1183–1194, Feb. 2019.

[39] L. Gan, J. Xia, P. Du, and J. Chanussot, "Multiple feature kernel sparse representation classifier for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5343–5356, Sep. 2018.

[40] L. Fang, N. He, S. Li, A. J. Plaza, and J. Plaza, "A new spatial–spectral feature extraction method for hyperspectral images using local covariance matrix representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3534–3546, Jun. 2018.

[41] J. Zheng, Y. Feng, C. Bai, and J. Zhang, "Hyperspectral image classification using mixed convolutions and covariance pooling," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 522–534, Jan. 2021.

[42] T. Chen, Y. Liu, Y. Zhang, B. Du, and A. Plaza, "Superpixel-based collaborative and low-rank regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5529216.

[43] B. Xi, J. Li, Y. Li, R. Song, D. Hong, and J. Chanussot, "Few-shot learning with class-covariance metric for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 31, pp. 5079–5092, 2022.

[44] Q. Gao and S. Lim, "A probabilistic fusion of a support vector machine and a joint sparsity model for hyperspectral imagery classification," *GISci. Remote Sens.*, vol. 56, no. 8, pp. 1129–1147, Nov. 2019.

[45] D. Zabihzadeh, R. Monsefi, and H. S. Yazdi, "Sparse Bayesian approach for metric learning in latent space," *Knowl.-Based Syst.*, vol. 178, pp. 11–24, Aug. 2019.

[46] H. Su, Y. Yu, Z. Wu, and Q. Du, "Random subspace-based k-nearest class collaborative representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6840–6853, Aug. 2021.

[47] M. B. Almoujahed et al., "Detection of fusarium head blight in wheat under field conditions using a hyperspectral camera and machine learning," *Comput. Electron. Agricult.*, vol. 203, Dec. 2022, Art. no. 107456.

[48] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Trans. Image Process.*, vol. 31, pp. 3997–4010, 2022.

[49] P. Duan, S. Hu, X. Kang, and S. Li, "Shadow removal of hyperspectral remote sensing images with multiexposure fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5537211.

[50] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4225–4238, Aug. 2022.

[51] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Aug. 2019.

[52] K. Jiang et al., "Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining," *IEEE Trans. Image Process.*, vol. 30, pp. 7404–7418, 2021.

[53] Y. Xiao, X. Su, Q. Yuan, D. Liu, H. Shen, and L. Zhang, "Satellite video super-resolution via multiscale deformable convolution alignment and temporal grouping projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5610819.

[54] Y. Xiao et al., "Space-time super-resolution for satellite video: A joint framework based on multi-scale spatial–temporal transformer," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 108, Apr. 2022, Art. no. 102731.

[55] H.-H. Sun, W. Cheng, and Z. Fan, "Learning to remove clutter in real-world GPR images using hybrid data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5113714.

[56] H.-H. Sun et al., "Estimating parameters of the tree root in heterogeneous soil environments via mask-guided multi-polarimetric integration neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5108716.

[57] J. Wang, W. Li, Y. Wang, R. Tao, and Q. Du, "Representation-enhanced status replay network for multisource remote-sensing image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, 2023.

[58] M. Zhang, W. Li, Y. Zhang, R. Tao, and Q. Du, "Hyperspectral and LiDAR data classification based on structural optimization transmission," *IEEE Trans. Cybern.*, vol. 53, no. 5, pp. 3153–3164, May 2023.

[59] A. Sellami, M. Farah, I. R. Farah, and B. Solaiman, "Hyperspectral imagery classification based on semi-supervised 3-D deep neural network and adaptive band selection," *Expert Syst. Appl.*, vol. 129, pp. 246–259, Sep. 2019.

[60] 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," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740–754, Feb. 2019.

[61] H. Zhang, Y. Li, Y. Jiang, P. Wang, Q. Shen, and C. Shen, "Hyperspectral classification based on lightweight 3-D-CNN with transfer learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5813–5828, Aug. 2019.

[62] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4604–4616, Jul. 2020.

[63] X. Zhang et al., "Spectral–spatial fractal residual convolutional neural network with data balance augmentation for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10473–10487, Dec. 2021.

[64] L. Huang, Y. Chen, and X. He, "Spectral–spatial masked transformer with supervised and contrastive learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5508718.

[65] M. Zhang, H. Liu, M. Gong, H. Li, Y. Wu, and X. Jiang, "Cross-domain self-taught network for few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4501719.

[66] T. Prashanth et al., "LipGene: Lipschitz continuity guided adaptive learning rates for fast convergence on microarray expression data sets," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 6, pp. 3553–3563, Nov. 2022.

[67] S. Pei, C. Wang, S. Cao, and Z. Lv, "Data augmentation for fMRI-based functional connectivity and its application to cross-site ADHD classification," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–15, 2023.

[68] X. Li, X. Zhang, W. Huang, and Q. Wang, "Truncation cross entropy loss for remote sensing image captioning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5246–5257, Jun. 2021.

[69] Y. Chen, W. Zheng, W. Li, and Y. Huang, "Large group activity security risk assessment and risk early warning based on random forest algorithm," *Pattern Recognit. Lett.*, vol. 144, pp. 1–5, Apr. 2021.

[70] Y. Lei and K. Tang, "Learning rates for stochastic gradient descent with nonconvex objectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4505–4511, Dec. 2021.

[71] B. Cao, N. Wang, X. Gao, J. Li, and Z. Li, "Multi-margin based decorrelation learning for heterogeneous face recognition," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 680–686.

[72] Q.-H. Lin, Y.-W. Niu, J. Sui, W.-D. Zhao, C. Zhuo, and V. D. Calhoun, "SSPNet: An interpretable 3D-CNN for classification of schizophrenia using phase maps of resting-state complex-valued fMRI data," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102430.

[73] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, "Attention-based adaptive spectral–spatial kernel ResNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7831–7843, Sep. 2021.

[74] S. K. Roy, A. Deria, C. Shah, J. M. Haut, Q. Du, and A. Plaza, "Spectral–spatial morphological attention transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 25503615.

[75] F. Pacifici, Q. Du, and S. Prasad, "Report on the 2013 IEEE GRSS data fusion contest: Fusion of hyperspectral and LiDAR data [technical committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 3, pp. 36–38, Sep. 2013.

Peixian Zhuang (Member, IEEE) received the Ph.D. degree from the School of Informatics, Xiamen University, Xiamen, China, in 2016.

He is currently an Associate Professor with the Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China. From 2017 to 2020, he was a Lecturer and a primary Supervisor with the School of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China. From 2020 to 2022, he was a Post-Doctoral Fellow and an Assistant Research Fellow with the Department of Automation, Tsinghua University, Beijing. His research interests include sparse representation, Bayesian hierarchical modeling, deep learning, and calcium signal processing.

Guojia Hou received the Ph.D. degree in computer application technology from the Ocean University of China, Qingdao, China, in 2015.

He is currently an Associate Professor with the College of Computer Science and Technology, Qingdao University, Qingdao, China. He has published more than 40 peer-reviewed articles in his research area. His current research interests include underwater vision, image/video processing, image quality assessment, and pattern recognition.

Weidong Zhang (Member, IEEE) received the Ph.D. degree in information and communication engineering from the School of Information Science and Technology, Dalian Maritime University, Dalian, China, in June 2022.

He is currently a Professor with the School of Information Engineering, Institute of Computer Applications, Henan Institute of Science and Technology, Xinxiang, China. He is also the Leader of the Institute of Computer Applications with the Henan Institute of Science and Technology. His current research interests include image restoration, intelligent agriculture, and deep learning.
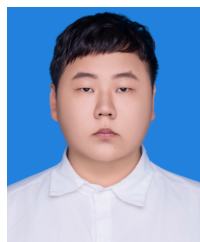
Zexu Li is currently pursuing the master's degree with the School of Information Engineering, Henan Institute of Science and Technology, Xinxiang, China.

His research interests include hyperspectral image processing and deep learning.

Guohou Li received the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in September 2011.

He is currently a Professor with the School of Information Engineering, Henan Institute of Science and Technology, Xinxiang, China. His research interests include intelligent control and signal and information processing.

Qiang Zhang (Member, IEEE) received the B.E. degree in surveying and mapping engineering and the M.E. and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2017, 2019, and 2022, respectively.

He is currently a Xinghai Associate Professor with the Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China. He has published more than ten journal articles on IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (TNNLS), IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), and *Journal of Photogrammetry and Remote Sensing* (ISPRS) P&RS. His research interests include remote sensing information processing, computer vision, and machine learning. More details could be found at https://qzhang95.github.io.

Chongyi Li (Senior Member, IEEE) received the Ph.D. degree from the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, in June 2018.

He is currently a Full Professor with the School of Computer Science, Nankai University, Tianjin. From 2016 to 2017, he was a joint training Ph.D. Student with Australian National University, Canberra, ACT, Australia. He was a Research Fellow with the City University of Hong Kong, Hong Kong, and Nanyang Technological University, Singapore, from 2018 to 2021. He was a Research Assistant Professor with the School of Computer Science and Engineering, Nanyang Technological University, from 2021 to 2023. His current research focuses on image processing, computer vision, and deep learning, particularly in the domains of image restoration and enhancement.