frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 January 2025
DOI 10.3389/fpls.2024.1458978

OPEN ACCESS

EDITED BY
Mohsen Yoosefzadeh Najafabadi,
University of Guelph, Canada

REVIEWED BY
Xin Wu,
Beijing University of Posts and Telecommunications (BUPT), China
Peter Yuen,
Cranfield University, United Kingdom

*CORRESPONDENCE
Ling Zhou
zhouling401618@163.com

RECEIVED 03 July 2024
ACCEPTED 13 December 2024
PUBLISHED 16 January 2025

CITATION
Wang B, Chen G, Wen J, Li L, Jin S, Li Y, Zhou L and Zhang W (2025) SSATNet: Spectral-spatial attention transformer for hyperspectral corn image classification. *Front. Plant Sci.* 15:1458978. doi: 10.3389/fpls.2024.1458978

SSATNet: Spectral-spatial attention transformer for hyperspectral corn image classification

Bin Wang[1], Gongchao Chen[2], Juan Wen[3], Linfang Li[2], Songlin Jin[2], Yan Li[4], Ling Zhou[2*] and Weidong Zhang[2]

[1]School of Life Sciences, Henan Institute of Science and Technology, Xinxiang, China, [2]School of Information Engineering, Henan Institute of Science and Technology, Xinxiang, China, [3]School of Art, Henan University of Economics and Law, Zhengzhou, China, [4]School of Software, Henan Institute of Science and Technology, Xinxiang, China

Hyperspectral images are rich in spectral and spatial information, providing a detailed and comprehensive description of objects, which makes hyperspectral image analysis technology essential in intelligent agriculture. With various corn seed varieties exhibiting significant internal structural differences, accurate classification is crucial for planting, monitoring, and consumption. However, due to the large volume and complex features of hyperspectral corn image data, existing methods often fall short in feature extraction and utilization, leading to low classification accuracy. To address these issues, this paper proposes a spectral-spatial attention transformer network (SSATNet) for hyperspectral corn image classification. Specifically, SSATNet utilizes 3D and 2D convolutions to effectively extract local spatial, spectral, and textural features from the data while incorporating spectral and spatial morphological structures to understand the internal structure of the data better. Additionally, a transformer encoder with cross-attention extracts and refines feature information from a global perspective. Finally, a classifier generates the prediction results. Compared to existing state-of-the-art classification methods, our model performs better on the hyperspectral corn image dataset, demonstrating its effectiveness.

KEYWORDS

corn identification, hyperspectral image classification, deep learning, morphology, image classification

1 Introduction

Hyperspectral imaging technology comprehensively measures an object's spectral properties by recording its absorption and reflection across various spectral bands (Li et al., 2024c; Zhang et al., 2024b; Li et al., 2024a). The resulting hyperspectral images, composed of multiple consecutive bands, are rich in feature information and can

thoroughly reveal the nature of the object. This technology advances intelligent agriculture by utilizing the detailed feature information in hyperspectral images, thereby avoiding the destructive methods of traditional seed identification. Hyperspectral imaging has gradually been applied to intelligent agriculture, geological exploration, and medical treatment, offering new development opportunities and technical capabilities.

The increasing variety of corn seeds available in the market presents a significant challenge to the cereal farming industry, making the accurate identification of corn varieties especially crucial. Recently, researchers have been investigating hyperspectral image classification techniques using machine learning and deep learning approaches (Zhang et al., 2023c; Wu et al., 2022). Ahmad et al. (Ahmad et al., 2019). utilized a self-encoder paired with a multilayer extreme learning machine to mitigate high computational overhead and the Thuesian phenomenon in hyperspectral images, which improved the accuracy of hyperspectral image classification. Okwuashi et al. (Okwuashi and Ndehedehe, 2020) introduced a deep support vector machine algorithm incorporating four kernel functions and demonstrated its effectiveness in hyperspectral image classification using publicly available datasets. Zhang et al. (Zhang et al., 2020) employed a deep forest model with hyperspectral imaging to classify rice seeds with different levels of frost damage in small sample datasets. Su et al. (Su et al., 2022) introduced a new semi-supervised method for hyperspectral image classification that integrates normalized spectral clustering with kernel learning, effectively addressing the issues of relevant features appearing in non-adjacent regions and the lack of non-Euclidean spatial correlation. Jin et al. (Jin et al., 2023) developed a cost-sensitive K-neighborhood algorithm to reduce noise interference, enhance spatial information utilization, and achieve robust performance in hyperspectral wheat image classification. Farmonov et al. (Farmonov et al., 2023) employed wavelet transform for feature extraction, combined with random forests and support vector machine algorithms, to localize crops in farmland and classify crop hyperspectral images, playing a significant role in crop growth monitoring and harvest prediction. Sim et al. (Sim et al., 2024) combined machine learning algorithms with hyperspectral imaging for fast, non-destructive detection of coffee origin without sample processing. Wang et al. (Wang et al., 2024b) proposed a cross-domain few-shot learning strategy utilizing a two-branch domain adaptation technique to mitigate distortion caused by enforcing different domain alignments, achieving effective cross-domain transfer learning for low/high spatial resolution data. Although machine learning methods have demonstrated exemplary performance in hyperspectral image classification, their reliance on manual or semi-automatic feature extraction limits their potential. The emergence of deep learning methods has enabled the automatic extraction of spectral, spatial and spatial-spectral features from hyperspectral images, leading to significant advancements in this field.

Zhang et al. (Zhang et al., 2019) created a straightforward 1D convolutional capsule network to tackle the high dimensionality and limited labeled samples in hyperspectral images, achieving effective feature extraction and classification. Wang et al. (Wang et al., 2020) developed an end-to-end cubic convolutional neural network that integrates Principal Component Analysis with 1D convolution for efficient extraction of spatial and spectral features. Roy et al. (Roy et al., 2020) proposed an improved residual network using an adaptive spatial-spectral kernel with attention mechanisms, utilizing 3D convolutional kernels to simultaneously extract spatial and spectral features, achieving excellent classification results. Cui et al. (Cui et al., 2021) introduced a lightweight deep network using 3D deep convolution to classify hyperspectral images with fewer parameters and lower computational costs. Ortac et al. (Ortac and Ozcan, 2021) evaluated the performance of 1D, 2D, and 3D convolutions in hyperspectral image classification, demonstrating that 3D convolution offers superior feature extraction capabilities. Ghaderizadeh et al. (Ghaderizadeh et al., 2021) employed depth-separable and fast convolutional blocks in combination with 2D convolutional neural networks to effectively tackle data noise and insufficient training samples. Paoletti et al. (Paoletti et al., 2023a) proposed a channel attention mechanism to automatically design and optimize convolutional neural networks, reducing the computational burden in feature extraction while obtaining effective classification outcomes. Sun et al. (Sun et al., 2023) introduced an extensive kernel spatial-spectral attention network designed to efficiently leverage 3D spatial-spectral features, maintaining the 3D structure of hyperspectral images. Jia et al. (Jia et al., 2023) developed a structure-adaptive CNN for hyperspectral image classification, which employs structure-adaptive convolution and mean pooling to extract deep spectral, spatial, and geometric features from a uniform hyperpixel region. Gao et al. (Gao et al., 2023) designed a lightweight 3D-2D multigroup feature extraction module for hyperspectral image classification, which mitigates the loss of crucial details in single-scale feature extraction and the high computational expense of multiscale extraction. Zhang et al. (Zhang et al., 2023b) introduced a method combining 3D and 2D convolution to fully utilize the spatial, texture and spectral features of hyperspectral data for the task of identifying wheat varieties. In conclusion, while 2D and 3D convolutions effectively capture spectral and spatial features from hyperspectral data, traditional convolutional neural networks are limited by high computational complexity and insufficient feature utilization, impacting their classification performance.

Inspired by (Vaswani et al., 2017), researchers have suggested a Transformer-based network model for image classification (Zhang et al., 2024a). Hong et al. (Hong et al., 2021) effectively classified hyperspectral remote sensing images by leveraging spectral local sequence information from neighboring bands, considering the temporal properties, and designing cross-layer skipping connections combined with the Transformer structure. Roy et al. (Roy et al., 2021) introduced an innovative end-to-end deep learning framework, using spectral and spatial morphological blocks for nonlinear transformations in feature extraction. Yang et al. (Yang et al., 2022) integrated convolutional operations into the Transformer structure to capture local spatial context and subtle spectral differences, fully utilizing the sequence attributes of spectral features. Sun et al. (Sun et al., 2022b) developed a spatial-spectral feature tokenization converter to capture both spectral-spatial and high-level semantic features, achieving hyperspectral image classification through a feature transformation module, a feature extraction module, and a sample label learning module. Kumar et al. (Kumar et al., 2022) developed a novel morphology-expanding convolutional neural network that connects the

morphological feature domain with the original hyperspectral data, reducing computational complexity and achieving good classification results. Peng et al. (Peng et al., 2022) developed a two-branch spectral-spatial converter with cross-attention, using spatial sequences to extract spectral features and capture deep spatial information to establish interrelationships among spectral sequences. Tang et al. (Tang et al., 2023) introduced a dual-attention Transformer encoder based on the Transformer backbone network for hyperspectral image classification, effectively extracting global dependencies and local spatial information between spectral bands. Qi et al. (Qi et al., 2023a) embedded 3D convolution in a two-branch Transformer structure to capture globally and locally correlated spectral-spatial domain features, demonstrating good performance for hyperspectral image classification through validation. Qiu et al. (Qiu et al., 2023) proposed a cross-channel dynamic spectral-spatial fusion Transformer capable of extracting multi-channel and multi-scale features, using multi-head self-attention to extract cross-channel global features and enhancing spatial-spectral joint features for hyperspectral image classification. Sun et al. (Sun et al., 2024) converted the spatial-spectral features into a memory marker storing *a priori* knowledge into an in-memory tagger, using a memory-enhanced Transformer encoder for the hyperspectral image classification task. Ahmad et al. (Ahmad et al., 2024) designed a Transformer-based network for hyperspectral image classification by combining wavelet transform with downsampling. The wavelet transform performs reversible downsampling, enabling attentional learning while preserving data integrity. Based on these studies, we propose utilizing a combination of 2D-3D convolution and Transformer, leveraging spectral-spatial morphological features to identify hyperspectral corn seed varieties. The contributions of this paper can be summarized as follows:

- We developed a 3D-2D convolutional cascade structure that autonomously extracts contextual features, reduces data complexity and efficiently captures high-level abstract features for integration into the Transformer architecture.
- We introduced a spectral-spatial morphology structure that employs expansion and erosion operations for spectral-spatial morphology convolution, enhancing the understanding of the data's intrinsic properties.
- We employed a Transformer Encoder with CrossAttention to comprehensively extract and refine feature information from hyperspectral corn images on a global scale using the attention mechanism.

2 Related works

Currently, researchers have proposed a variety of methods for classifying hyperspectral remote sensing images and hyperspectral seed images. We classify these approaches into deep learning methods, machine learning methods and traditional methods. The deep learning methods are further divided into hybrid CNN-Transformer methods, Transformer-based methods, and CNN-based methods. Next, we overview and summarize these research outcomes.

Traditional methods for hyperspectral image classification primarily rely on analyzing physical and statistical features. These methods typically include spectral feature extraction, pixel-based classification, and target-based classification. For example, Cui et al. (Cui et al., 2020) introduced a super-pixel and multi-classifier fusion approach to tackle the challenges of limited labeled samples and substantial spectral variations. Similarly, Chen et al. (Chen et al., 2021a) introduced a feature extraction means that combines PCA and LBP, optimized using the Gray Wolf optimization algorithm for hyperspectral image classification. While these methods perform well for simpler classification tasks, their effectiveness diminishes when faced with complex backgrounds and highly mixed pixels.

Machine learning methods effectively classify hyperspectral images by learning the features of sample data. With the advancement of machine learning technology, researchers increasingly utilize machine learning algorithms for hyperspectral image classification. For example, Pham et al. (Pham and Liou, 2022) developed a push-sweep hyperspectral system using a support vector machine to date surface defects, addressing the problem of insufficient accuracy and speed in detecting date skin defects with traditional methods. Sun et al. (Sun et al., 2022a) constructed a network integrating multi-feature and multi-scale extraction with a swift and efficient kernel-extreme learning machine for rapid classification, significantly enhancing hyperspectral image classification accuracy. Wang et al. (Wang et al., 2023b) proposed a capsule vector neural network that combines capsule representation of vector neurons with an underlying fully convolutional network, achieving good classification performance with insufficient labeled samples. Compared to traditional methods, machine learning approaches handle high-dimensional data more effectively and achieve higher classification accuracy. However, these methods still rely on human-designed feature extraction and selection, preventing them from fully utilizing all the information in hyperspectral data.

Deep learning methods excel in hyperspectral image classification due to their automatic feature extraction and end-to-end learning capability (Zhang et al., 2024c; Hong et al., 2023). These methods can be categorized into hybrid CNN-Transformer methods, Transformer-based methods, and CNN-based methods.

CNN-based methods are designed to capture spectral and spatial features through convolutional layers specifically tailored for hyperspectral data, significantly improving classification performance (Wu et al., 2021). Yang et al. (Yang et al., 2021) introduced a spatial-spectral cross-attention network that suppresses redundant data bands and achieves robust, accurate classification. Yu et al. (Yu et al., 2021) developed a spectral-spatial dense convolutional neural network framework with a feedback attention mechanism to tackle issues of high complexity, information redundancy, and inefficient description, thereby improving classification efficiency and accuracy. Zheng et al. (Zheng et al., 2022) developed a rotationally invariant attention network for pixel feature class recognition, leveraging spectral features and spatial information. Paoletti et al. (Paoletti et al., 2023b) created a channel attention mechanism to automatically design and

optimize a CNN, integrating 1D and spectral-spatial (3D) classifiers to process data from various perspectives while reducing computational overhead. Guo et al. (Guo et al., 2023) introduced a dual-view global spatial and spectral feature fusion network that efficiently extracts spectral-spatial features from hyperspectral images, accounting for global and local information.

Transformer-based methods excel at capturing long-range dependencies and complex features in hyperspectral images through a self-attention mechanism. Huang et al. (Huang et al., 2022) introduced a 3D swin transformer that captures rich spatial-spectral information, learns semantic representations from unlabeled data, and overcomes traditional methods' limitations regarding receptive fields and labeling requirements. Yu et al. (Yu et al., 2022) proposed a multilevel spatial-spectral transformer network that processes hyperspectral images into sequences, addressing issues faced by CNN-based methods such as limited receptive fields, information loss in downsampling layers, and high computational resource consumption. Zhang et al. (Zhang et al., 2023d) developed a location-lightweight multi-head self-attention module and a channel-lightweight multi-head self-attention module, allowing each channel or pixel to associate with global information while reducing memory and computational burdens. Zhao et al. (Zhao et al., 2023) proposed an active learning hyperspectral image classification framework using an adaptive super-pixel segmentation and multi-attention transformer, achieving good classification performance with small sample sizes. Wang et al. (Wang et al., 2023a) introduced a trispectral image generation channel that converts hyperspectral images into high-quality trispectral images, mitigating the spatial variability problem caused by complex imaging conditions. Compared to CNNs, transformers have significant advantages in processing global and multi-scale features, allowing for better handling of global information in hyperspectral images.

Methods that hybrid CNN and Transformer aim to utilize the strengths of both to enhance hyperspectral image classification performance. These hybrid methods typically employ Transformers to capture global dependencies and CNNs to extract local spatial features. Zhang et al. (Zhang et al., 2022a) designed a dual-branch structure combining Transformer and CNN branches, effectively extracting both global hyperspectral features and local spectral-spatial features, resulting in high classification accuracy. Zhang et al. (Zhang et al., 2023a) proposed a network that integrates Transformer and multiple attention mechanisms, utilizing spatial and channel attention to focus on salient information, thereby enhancing spatial-spectral feature extraction and semantic understanding. Qi et al. (Qi et al., 2023b) introduced a global-local 3D convolutional Transformer network, embedding a dual-branch Transformer in 3D convolution to simultaneously capture global-local correlations across spatial and spectral domains, addressing the restricted receptive field issue of traditional CNNs. Xu et al. (Xu et al., 2024) proposed a two-branch convolutional Transformer network based on 3D CNN and an improved Transformer encoder, integrating spatial and local-global spectral features with lower computational complexity. Chen et al. (Chen et al., 2024) developed the TCCU-Net, a two-stream collaborative network that learns spatial, spectral, local and global information end-to-end for effective hyperspectral unmixing. This integration enables the model to leverage both spectral and spatial information from hyperspectral images more comprehensively, enhancing classification robustness and accuracy.

3 Methodology

The network flowchart of our proposed Spectral-Spatial Attention Transformer for hyperspectral corn image classification is shown in Figure 1. It contains 3D-2D Convolutional Module, Spectral-Spatial Morphology, Transformer Encoder with CrossAttention, and Classifier.

3.1 Motivation

With the development of intelligent agriculture, the integration of hyperspectral imaging technology and deep learning has gained widespread application in crop research, particularly in seed classification and identification. As a globally important food crop, the classification of corn seeds is significant for improving agricultural productivity and preserving crop genetic resources. Hyperspectral images can capture reflectance features at different wavelengths, providing researchers with rich spectral information for more precise seed classification and quality assessment (Chang et al., 2024).

In recent years, transformer models have emerged as popular in computer vision due to their powerful feature extraction and representation capabilities (Han et al., 2023; Li et al., 2024b). Compared to traditional convolutional neural networks, transformers are better at handling high-dimensional data and capturing long-range dependencies, which are crucial for extracting complex features from hyperspectral images. Additionally, the self-attention mechanism of Transformers enables the model to flexibly focus on important areas within the image, thereby enhancing classification accuracy. Consequently, choosing Transformer-based methods allows for more effective utilization of hyperspectral data, providing more reliable support for corn seed classification.

3.2 3D-2D convolution module

In hyperspectral image classification, effective feature extraction is vital for improving accuracy. Both 3D and 2D convolutions are widely used in this domain due to their unique advantages. 3D convolution simultaneously operates in spectral and spatial dimensions, capturing their correlation. Unlike traditional 2D or 1D convolutions, 3D convolution provides richer feature descriptions and retains more original spectral and spatial information, thus enhancing classification accuracy. It fully leverages the three-dimensional data structure of hyperspectral images, avoiding information loss or oversimplification. However, as network depth and input data size increase, the computational complexity and memory requirements of 3D convolution rise significantly, demanding higher hardware resources and more

10.3389/fpls.2024.1458978

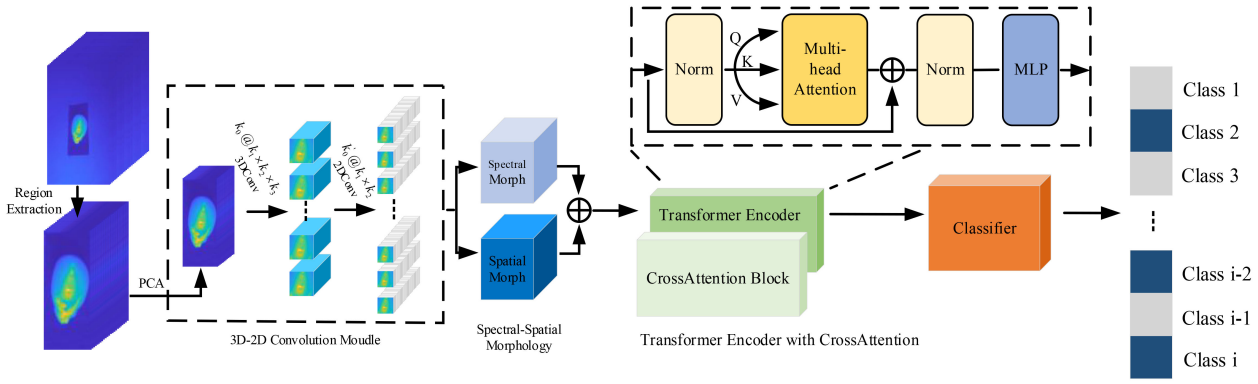

FIGURE 1
Flowchart of the spectral-spatial attention Transformer for hyperspectral corn image classification. Initially, the data are preprocessed with region of interest extraction and PCA dimensionality reduction. Subsequently, local spatial, spectral, and texture features are extracted using 2D and 3D convolutions. The spectral and spatial morphology modules further analyze the internal structure of the data. The Transformer encoder with cross-attention then extracts and refines the feature information from a global perspective. Finally, the classifier provides the prediction results.

training time. 2D convolution, on the other hand, has lower computational complexity and high efficiency, as it operates on two-dimensional space (width and height). It effectively utilizes spatial and texture information, making it suitable for handling local features and texture details in hyperspectral images. Combining 3D and 2D convolutions can efficiently leverage the strengths to extract features from hyperspectral corn images. 3D convolution captures complex spectral-spatial relationships, while 2D convolution extracts local spatial features and texture information, maintaining computational efficiency. This combination optimizes feature extraction, leading to improved classification performance.

3D convolution is mainly used for three-dimensional data processing, extracting features by sliding a convolution kernel across the three dimensions of the input data. Suppose the input data is $I^{D\times H\times W\times C}$, where C is the number of channels, W is the width, H is the height, and D is the depth (spectral dimension).The dimensions of the 3D convolution kernel are $K_d \times K_h \times K_w \times C \times N$, where N is the number of output channels (i.e., the number of convolution kernels), C is the number of input channels, K_w is the size in the width direction, K_h is the size in the height direction, and K_d is the size of the convolution kernel in the depth direction. For an input tensor I and a convolution kernel W, the output tensor Y of the 3D convolution can be expressed as

$$Y(n,d,h,w) = \sum_{c=0}^{C-1}\sum_{k_d=0}^{K_d-1}\sum_{k_h=0}^{K_h-1}\sum_{k_w=0}^{K_w-1} I(c,d+k_d,h+k_h,w+k_w) \times W(n,c,k_d,k_h,k_w) + b(n) \quad (1)$$

where $I(c,d+k_d,h+k_h,w+k_w)$ is the value of the input tensor I at channel c and position $(d+k_d,h+k_h,w+k_w)$. $W(n,c,k_d,k_h,k_w)$ represents the weight of the convolution kernel W at output channel n and input channel c, positioned at (k_d,k_h,k_w). $b(n)$ is the bias term for each output channel n in the convolutional layers. It is initialized with random values (typically small values close to zero) and then adjusted during training via backpropagation. The gradient of the loss with respect to the bias is computed and used to update $b(n)$, just like the weights of the convolutional filters. This adjustment allows the model to shift the activations of each channel, enabling the network to adapt to various patterns in the data and improve its representation of features.

2D convolution is applied to 2D data processing, extracting features by sliding a convolution kernel (filter) across the two dimensions of the input data. Assuming the input data is $I^{H\times W\times C}$, the 2D convolution kernel has dimensions $K_h \times K_w \times C \times N$, with the parameter presentation consistent with that of 3D convolution. For an input tensor I and a convolution kernel W, the output tensor Y of the 2D convolution can be expressed as

$$Y(n,i,j) = \sum_{c=0}^{C-1}\sum_{k_h=0}^{K_h-1}\sum_{k_w=0}^{K_w-1} I(c,i+k_h,j+k_w) \times W(n,c,k_h,k_w) + b(n) \quad (2)$$

where $I(c,i+k_h,j+k_w)$ is the value at position $(i+k_h,j+k_w)$ in the input tensor I at channel c. $W(n,c,k_h,k_w)$ represents the weight of the convolutional kernel W at position (k_h,k_w) for output channel n and input channel c.

3.3 Spectral-spatial morphology module

Hyperspectral images contain abundant textural, spatial, and spectral information. Morphology, a nonlinear image processing technique, is mainly used to analyze and manipulate the shape and structure of images. In hyperspectral image processing, morphological methods can effectively extract spatial and spectral features, enhancing the robustness and accuracy of image classification. Building on this, we integrate morphology with 2D convolution to locally manipulate images using structural elements, which can highlight or suppress specific shape features.

Spatial features can be extracted from each spectral band of a hyperspectral corn image through morphological operations like dilation and erosion. The dilation operation can emphasize the bright areas in the image and expand the edges of the target object,

10.3389/fpls.2024.1458978

making the morphological features of the corn seed more pronounced. The computational expression for dilation is as

$$D(I) = I \oplus B = \bigcup_{b \in B}(I + b) \tag{3}$$

where I denotes the input image, B is the structural element (a small template used to detect the morphological features of the image), $\oplus$ stands for the dilation operation, $\bigcup_{b \in B}()$ represents the union of all structural element positions to take the maximum value, and + denotes the pixel displacement operation. b influences the dilation and erosion operations. These operations involve shifting and adjusting the shape of features within the image, where b helps control the degree of expansion (dilation) or contraction (erosion). Like the convolutional biases, the values of b in these operations are also learned during training, refining the model's ability to capture spatial relationships and remove irrelevant details in the data. Conversely, the erosion operation removes noise and small bright spots, resulting in a smoother and more uniform target area. The computational expression for erosion is as

$$E(I) = I \ominus B = \bigcap_{b \in B}(I - b) \tag{4}$$

where $\ominus$ denotes the erosion operation, $\bigcap_{b \in B}()$ represents the intersection operation to take the minimum value for all structural element positions, and − indicates the negative displacement operation of pixels. Performing these operations on each spectral band extracts subtle spatial variations and enhances the representation of spatial features. Subsequently, these spatial features are combined with spectral features to fully utilize the spectral and spatial information in hyperspectral images. Specifically, we apply morphological operations to each spectral band to extract spatial features. These spatial features are merged with the original spectral information to construct high-dimensional feature vectors. This method preserves the spectral information of the hyperspectral image while enhancing the representation of spatial structure information. The feature extraction and classification effectiveness is further improved by integrating these morphological operations with 2D convolution. 2D convolution extracts local spatial features within each spectral band and enhances the representation of spatial information. These two convolutional operations complement each other, allowing the features, preprocessed through morphological operations, to be input into the convolutional neural network for more accurate classification.

The bias b in these equations plays a crucial role in adjusting the output activations, improving the feature extraction process. In the convolutional operations (Equations 1, 2), it allows the network to adapt to various activation patterns, enhancing the model's ability to learn more complex relationships in the data. In the morphological operations (Equations 3, 4), it enhances spatial feature representation by refining the shapes and structures in the image. This combination of accurate feature extraction and refinement leads to better corn seeds classification performance.

By integrating morphological and convolutional techniques, we substantially enhance hyperspectral corn image classification accuracy and robustness. This combined approach boosts classification performance and improves resilience against complex backgrounds and noise.

3.4 Transformer encoder with CrossAttention module

The Transformer encoder enhances input data representation through a sophisticated attention module that captures dependencies among different parts of the input sequence. Figure 2 depicts the detailed structure of this attention module, consisting of two primary components: multi-head self-attention and scaled dot-product attention.

Originally, the Transformer architecture was designed for natural language processing, particularly for handling sequence data, and it excels in this domain due to its multiple self-attention core blocks. Unlike conventional Convolutional Neural Networks and Recurrent Neural Networks, the Transformer exclusively utilizes the attention mechanism, enabling efficient capture of global dependencies in sequential data. The input sequence is initially converted into a fixed-dimensional vector representation via an embedding layer, with positional information preserved through positional encoding, which is generated by sine and cosine functions.

Each encoder layer includes multiple self-attention heads, each independently processing the input sequence to generate an attention representation, which is then concatenated and integrated through a linear transformation. The multi-head self-attention mechanism enables the model to attend to multiple parts of the input sequence simultaneously. Specifically, the input sequence is represented as a key (K), query (Q), and value (V). Multiple sets of Q, K, and V are created through the linear projection of a learned weight matrix. Each set of Q, K, and V is passed to the scaled dot-product attention mechanism, where attention scores are calculated and applied to the values. The Q is multiplied by the transposed key K^T to obtain the raw attention score, which is then divided by the square root of the key's dimension, $\sqrt{d_k}$, to maintain gradient stability. The computational process of self-attention can be summarized as

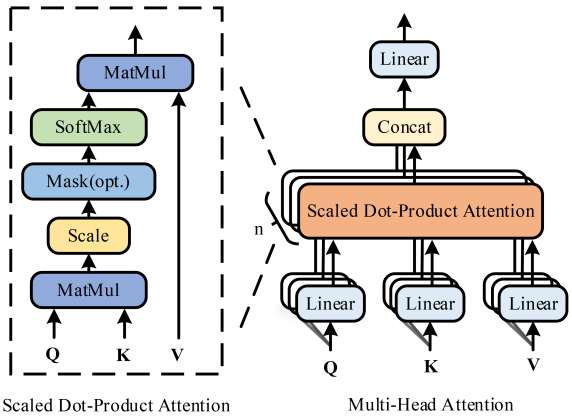

FIGURE 2
Diagram illustrating the structure of the multi-head attention mechanism and scaled dot-product attention.

$$\mathrm{SA} = \mathrm{Attention}\,(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V \tag{5}$$

Through its unique multi-head self-attention mechanism and feed-forward neural network, the Transformer structure efficiently captures global dependencies and improves the classification accuracy of hyperspectral corn images.

3.5 Loss function

In this paper, we propose a method that combines spectral-spatial morphology with a 3D-2D convolutional Transformer network to classify hyperspectral corn images. This approach fully utilizes the spatial and spectral features of hyperspectral images. To optimize model performance, we employ the CrossEntropyLoss function.

The CrossEntropyLoss function is commonly used in classification tasks, especially for multi-class classification problems. It measures the discrepancy between the true category distribution and the predicted probability distribution by computing the negative log-likelihood between the actual labels and the predicted probabilities. This function ensures numerical stability by converting the output into a probability distribution using the Softmax function. Additionally, the gradient of the CrossEntropyLoss function is relatively easy to compute, facilitating the implementation of the back-propagation algorithm and model optimization. By directly quantifying the alignment between predicted probabilities and actual labels, it accurately reflects the performance of the classification model. Consequently, we apply the CrossEntropyLoss function to the hyperspectral corn image classification task. Its computational expression is as

$$\mathrm{CrossEntropyLoss} = -\sum_{i=0}^{N} y_i \log\left(\hat{y}_i\right) \tag{6}$$

where y_i represents the true label of the sample, N is the total number of samples, and $\hat{y}_i$ is the predicted probability from the model. The network model converts the output to a probability distribution using the Softmax function

$$\hat{y}_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{7}$$

where z_i represents the linear output of the model. For a given category c, the true label $y_c = 1$ while the labels for all other categories are 0. The predicted probability $\hat{y}_i$ corresponding to the true label y_i is substituted into Equation 6, and the loss value for each sample is

$$\mathrm{Loss} = -\sum_i y_i \log\left(\hat{y}_i\right) \tag{8}$$

By measuring the difference between actual and predicted labels and updating the model parameters through the backpropagation algorithm to minimize the loss, this approach effectively guides the model in learning to handle complex hyperspectral corn image features. Consequently, it improves both the classification accuracy and robustness.

4 Experiment and analysis

In this section, we will first discuss the dataset used, detail the specific implementation of SSATNet, and then present the evaluation metrics, multi-classification results, and ablation study.

4.1 Experimental dataset

To verify the effectiveness of the SSATNet, we utilized the hyperspectral corn image dataset from SSTNet (Zhang et al., 2022b). This dataset contains 10 corn varieties, each with 120 samples. The collected images cover a spectral range from 400 to 1000 nm, encompassing 128 bands. To reduce computational overhead and focus on retaining only the core area of the corn seeds, the collected raw data resolution of 696 × 520 was reduced to 210 × 200 for feature extraction. The corn seed images were sourced from planting areas in Henan Province, including varieties such as FengDa601, BaiYu9284, BaiYu8317, BaiYu918, BaiYu897, BaiYu879, BaiYu833, BaiYu818, BaiYu808, and BaiYu607. Figure 3 shows different spectral band maps of a sample randomly selected from FengDa601, BaiYu818, and BaiYu833. This corn image dataset was obtained by contacting the authors.

4.2 Implementation details

The hyperspectral corn image dataset includes 10 varieties, totaling 1200 samples, divide into training and test sets in a 4:1 ratio. We conducted our experiments on a Windows 10 PC with an Intel® Xeon® Gold 5218 CPU @ 2.30GHz x64, an NVIDIA GeForce RTX 3090*2 graphics card, and 256 GB RAM. The Batch size is set to 16 for the training and 8 for the testing. We used Adamax as the optimizer with a learning rate of 0.01, an exponential decay rate of 0.9, a gradient squared moving average rate of 0.999, and 250 iterations. Additionally, we implemented a Dropout mechanism that randomly deactivates 10% of nodes, effectively preventing overfitting.

4.3 Evaluation metrics

To thoroughly assess the performance of our SSATNet in classifying hyperspectral corn images, we employ four standard evaluation metrics: F1-Score, Recall, Precision, and the Kappa coefficient(K_A). Precision assesses the accuracy of the classification model by evaluating the proportion of instances predicted to be positive that are actually positive. There exists a trade-off between Precision and Recall; increasing Precision may lead to a decrease in Recall and vice versa. Therefore, the F1-Score, derived as the harmonic mean of Precision and Recall, is often used for a more balanced evaluation of model performance, and its calculation expression is shown in Equation 9. The K_A is a consistency test metric that evaluates the agreement between the classified image and the reference image in hyperspectral remote

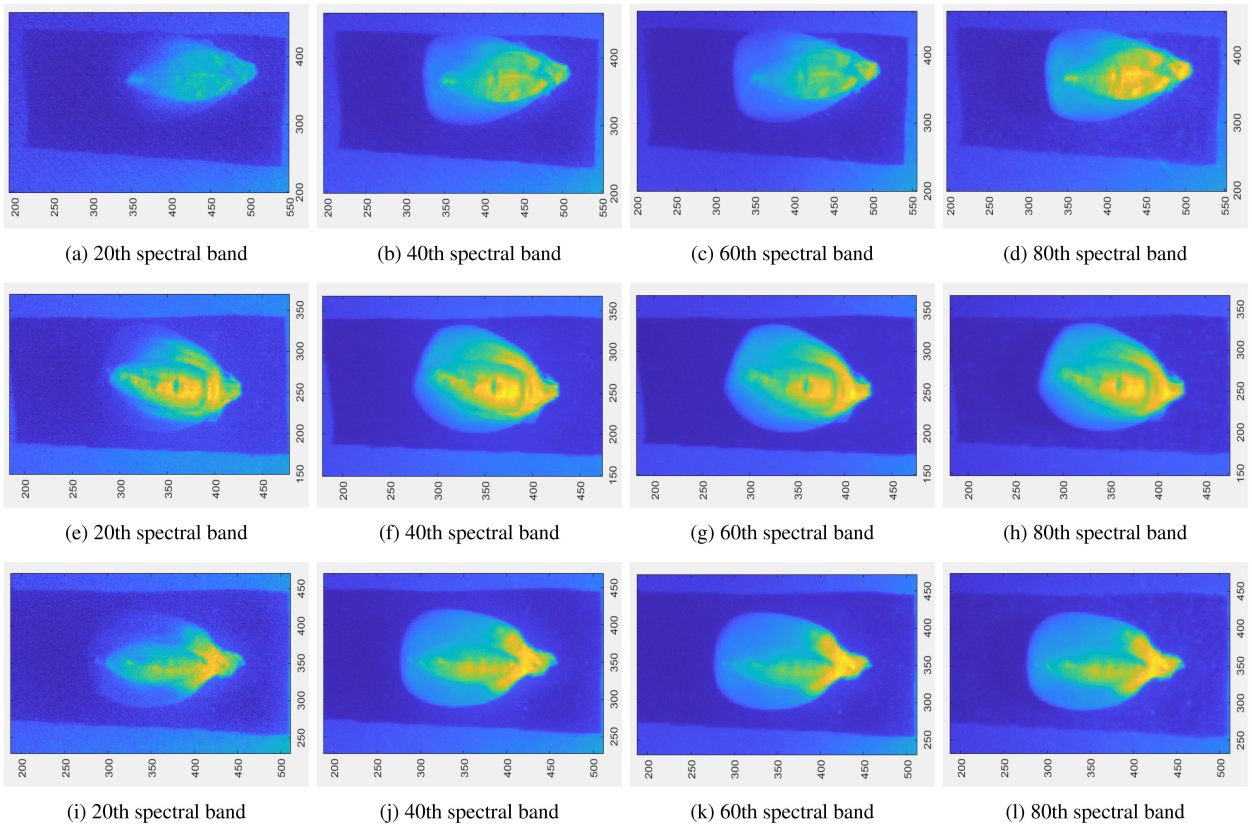

FIGURE 3
Randomly select a sample from three corn varieties, FengDa601 **(A–D)**, BaiYu818 **(E–H)**, and BaiYu833 **(I–L)**, and display their partial spectral bands.

sensing classification tasks, providing a more comprehensive reflection of the overall classification accuracy. Higher scores in these four evaluation metrics indicate better model performance. Figure 4 shows the confusion matrix of our model's classification results for hyperspectral corn images and the results of one of the training and testing sessions.

$$\text{F1} - \text{Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

4.4 Multi-classification results

Extensive experiments were performed to thoroughly test the generalization and effectiveness of our model for hyperspectral corn image classification. The comparison methods include KNN (Kumbure et al., 2020), SGD (Lei and Tang, 2021), RFA (Chen et al., 2021b), HybridNet (Roy et al., 2019), SSTNet (Zhang et al., 2022b), CTMixer (Zhang et al., 2022a), MSTNet (Yu et al., 2022), MATNet (Zhang et al., 2023a), and 3DCT (Wang et al., 2024a). The experimental results are presented in Table 1. The source code and parameters for the comparison methods were acquired from the original authors.

The results presented in Table 1 demonstrate the performance of various methods on the hyperspectral corn images dataset. Traditional machine learning models such as KNN (Kumbure et al., 2020), RFA (Chen et al., 2021b), and SGD (Lei and Tang, 2021) show subpar performance across all evaluation metrics, with RFA (Chen et al., 2021b) performing the worst across all metrics. These traditional models, lacking nonlinear activation mechanisms, struggle to extract deep spectral-spatial features effectively. In contrast, HybridNet (Roy et al., 2019), SSTNet (Zhang et al., 2022b), and 3DCT (Wang et al., 2024a), which integrate 3D convolution, demonstrate superior results due to their ability to capture spectral and spatial features simultaneously. Models like CTMixer (Zhang et al., 2022a), MSTNet (Yu et al., 2022), and MATNet (Zhang et al., 2023a) further leverage the Transformer architecture to address the complex relationships inherent in hyperspectral data. Our proposed model, which combines convolutional networks with Transformers and incorporates a novel spectral-spatial attention mechanism, achieves the best overall performance across all metrics. The integration of local and global feature extraction methods allows our model to substantially improve Precision, Recall, F1-Score, and K_A, surpassing existing state-of-the-art methods. These results validate the effectiveness of our design in capturing the complex spectral-spatial features of hyperspectral corn images and its superior ability to generalize to high-dimensional datasets.

4.5 Ablation study

To further evaluate the contribution of each module in SSATNet to the classification performance of hyperspectral corn seed images, we conducted ablation experiments on the dataset introduced by SSTNet

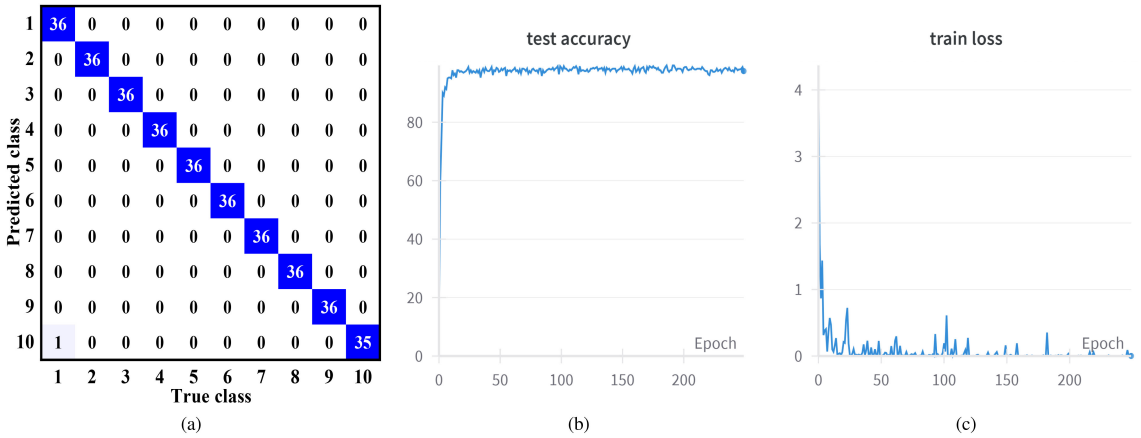

FIGURE 4
(A) The confusion matrix of our SSATNet classification results. **(B)** The results of one of the testing. **(C)** The results of one of the training.

TABLE 1 Test results of various methods on the hyperspectral corn images dataset.

Models	Hyperspectral Corn images			
	Precision	Recall	F1-Score	K_A
KNN (Kumbure et al., 2020)	96.12 ± 0.35	95.72 ± 0.32	95.90 ± 0.24	0.9675 ± 0.011
SGD (Lei and Tang, 2021)	96.98 ± 0.28	96.50 ± 0.18	96.70 ± 0.21	0.9721 ± 0.008
RFA (Chen et al., 2021b)	94.50 ± 0.40	94.10 ± 0.38	94.22 ± 0.39	0.9519 ± 0.009
HybridNet (Roy et al., 2019)	96.72 ± 0.30	96.44 ± 0.28	96.34 ± 0.21	0.9772 ± 0.007
SSTNet (Zhang et al., 2022b)	98.12 ± 0.18	97.78 ± 0.15	97.95 ± 0.17	0.9887 ± 0.005
CTMixer (Zhang et al., 2022a)	97.38 ± 0.33	97.75 ± 0.30	97.20 ± 0.32	0.9827 ± 0.008
MSTNet (Yu et al., 2022)	97.00 ± 0.38	96.95 ± 0.35	96.80 ± 0.36	0.9802 ± 0.009
MATNet (Zhang et al., 2023a)	98.27 ± 0.16	98.34 ± 0.14	98.25 ± 0.15	0.9930 ± 0.004
3DCT (Wang et al., 2024a)	98.30 ± 0.28	98.12 ± 0.25	98.19 ± 0.27	0.9928 ± 0.004
Our	**98.65 ± 0.18**	**98.57 ± 0.15**	**98.60 ± 0.17**	**0.9965 ± 0.003**

Optimal, bolded; Suboptimal, blue.

(Zhang et al., 2022b). In these experiments, we systematically removed individual components of the network while retaining the remaining modules unchanged. Specifically, we excluded the following components: 1) the 3D convolution module (-w/o 3DConv); 2) the 2D convolution module (-w/o 2DConv); 3) the spectral morphology structure (-w/o SpectralMorph); and 4) the spatial morphology structure (-w/o SpatialMorph). The Table 2 below illustrates the quantitative analysis metrics for each ablation experiment. The results demonstrate that the removal of the 3D convolution module leads to the most significant degradation in performance, underscoring its crucial role in capturing both spectral and spatial features in hyperspectral corn seed images. Without 3D convolution, the model's ability to integrate spatial-spectral correlations is substantially weakened. Similarly, the removal of the 2D

TABLE 2 Quantitative test results of ablation experiments.

Module	Precision	Recall	F1-Score	K_A
-w/o 3DConv	86.42 ± 0.31	87.33 ± 0.29	87.05 ± 0.36	0.8768 ± 0.006
-w/o 2DConv	89.51 ± 0.25	90.35 ± 0.25	90.52 ± 0.29	0.9117 ± 0.004
-w/o SpectralMorph	93.65 ± 0.22	93.27 ± 0.19	93.86 ± 0.25	0.9408 ± 0.004
-w/o SpatialMorph	92.59 ± 0.20	92.69 ± 0.21	92.31 ± 0.21	0.9332 ± 0.005
SSATNet (full model)	**98.65 ± 0.18**	**98.57 ± 0.15**	**98.60 ± 0.17**	**0.9965 ± 0.003**

Optimal, bolded.

convolution module also causes a noticeable decline in performance, although to a lesser extent compared to the absence of 3D convolution. This is because 2D convolution primarily focuses on extracting local spatial features and refining feature representations. The exclusion of the spectral morphology structure results in performance degradation, highlighting its importance in enhancing spectral feature representation and managing the complex spectral relationships inherent in hyperspectral data. Likewise, the spatial morphology structure significantly contributes to the model's performance by extracting and enhancing spatial features, enabling more accurate classification of corn seed images.

In summary, each module is crucial to the overall performance of SSATNet. The 3D convolution module provides the most significant enhancement to classification performance, followed by the spectral morphology structure and the spatial morphology structure. The 2D convolution module also provides substantial support in refining feature representation. Through the synergy of these modules, SSATNet excels in the hyperspectral corn seed classification task, demonstrating the effectiveness of its design.

5 Conclusion

In this paper, we propose the SSATNet method for non-destructive identification of hyperspectral corn varieties. First, we design a 3D-2D cascade structure to reduce image data complexity and effectively extract local feature information, facilitating the Transformer structure's processing. Additionally, we introduce a spectral-spatial morphology structure combined with 2D convolution to perform expansion and erosion operations on the data, providing a deeper understanding of the data's nature. Finally, we employ the Transformer structure to extract global feature information from hyperspectral corn images through the self-attention mechanism, achieving efficient capture of global dependencies between corn spectra. Ablation experiments highlight the effectiveness of each component of SSATNet in extracting features and classifying hyperspectral corn images. This method offers a new approach to non-destructive corn variety identification and significantly promotes the development of intelligent agriculture.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

BW: Investigation, Methodology, Resources, Writing – original draft. GC: Resources, Writing – original draft. JW: Validation, Visualization, Writing – review & editing. LL: Writing – review & editing, Formal analysis, Validation, Investigation. SJ: Supervision, Writing – review & editing. YL: Funding acquisition, Supervision, Writing – review & editing. LZ: Funding acquisition, Methodology, Supervision, Writing – original draft. WZ: Funding acquisition, Investigation, Supervision, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the China Postdoctoral Science Foundation project under Grant 2024M750747, in part by the Henan Provincial Science and Technology Research and Development Joint Foundation Project under Grant 235200810066, in part by the Teacher Education Curriculum Reform Research of Henan Province under Grant 2024-JSJYYB-099, and in part by the Key Specialized Research and Development Program of Science and Technology of Henan Province under Grants 242102210075, 232102210018, 242102211048, 242102211059, 242102211030, 242102210126.

Acknowledgments

This brief text acknowledges the contributions of specific colleagues, institutions, or agencies that assisted the authors' efforts.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ahmad, M., Ghous, U., Usama, M., and Mazzara, M. (2024). Waveformer: Spectral–spatial wavelet transformer for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* 21, 1–5. doi: 10.1109/LGRS.2024.3353909

Ahmad, M., Khan, A. M., Mazzara, M., and Distefano, S. (2019). "Multi-layer extreme learning machine-based autoencoder for hyperspectral image classification." in *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)* - Volume 4: VISAPP. SciTePress, pp 75–82. doi: 10.5220/0007258000750082

Chang, H., Bi, H., Li, F., Xu, C., Chanussot, J., and Hong, D. (2024). Deep symmetric fusion transformer for multimodal remote sensing data classification. *IEEE Trans. Geosci. Remote Sens.* 62, 1–15. doi: 10.1109/TGRS.2024.3476975

Chen, H., Miao, F., Chen, Y., Xiong, Y., and Chen, T. (2021a). A hyperspectral image classification method using multifeature vectors and optimized kelm. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 2781–2795. doi: 10.1109/JSTARS.4609443

Chen, J., Yang, C., Zhang, L., Yang, L., Bian, L., Luo, Z., et al. (2024). Tccu-net: Transformer and cnn collaborative unmixing network for hyperspectral image. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 17, 8073–8089. doi: 10.1109/JSTARS.2024.3352073

Chen, Y., Zheng, W., Li, W., and Huang, Y. (2021b). Large group activity security risk assessment and risk early warning based on random forest algorithm. *Pattern Recognition Lett.* 144, 1–5. doi: 10.1016/j.patrec.2021.01.008

Cui, B., Cui, J., Hao, S., Guo, N., and Lu, Y. (2020). Spectral-spatial hyperspectral image classification based on superpixel and multi-classifier fusion. *Int. J. Remote Sens.* 41, 6157–6182. doi: 10.1080/01431161.2020.1736730

Cui, B., Dong, X.-M., Zhan, Q., Peng, J., and Sun, W. (2021). Litedepthwisenet: A lightweight network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2021.3062372

Farmonov, N., Amankulova, K., Szatmári, J., Sharifi, A., Abbasi-Moghadam, D., Nejad, S. M. M., et al. (2023). Crop type classification by desis hyperspectral imagery and machine learning algorithms. *IEEE J. selected topics Appl. Earth observations Remote Sens.* 16, 1576–1588. doi: 10.1109/JSTARS.2023.3239756

Gao, H., Zhu, M., Wang, X., Li, C., and Xu, S. (2023). Lightweight spatial-spectral network based on 3d-2d multi-group feature extraction module for hyperspectral image classification. *Int. J. Remote Sens.* 44, 3607–3634. doi: 10.1080/01431161.2023.2224099

Ghaderizadeh, S., Abbasi-Moghadam, D., Sharifi, A., Zhao, N., and Tariq, A. (2021). Hyperspectral image classification using a hybrid 3d-2d convolutional neural networks. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 7570–7588. doi: 10.1109/JSTARS.2021.3099118

Guo, T., Wang, R., Luo, F., Gong, X., Zhang, L., and Gao, X. (2023). Dual-view spectral and global spatial feature fusion network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 61, 1–13. doi: 10.1109/TGRS.2023.3277467

Han, D., Pan, X., Han, Y., Song, S., and Huang, G. (2023). Flatten transformer: Vision transformer using focused linear attention. In. *Proc. IEEE/CVF Int. Conf. Comput. vision.* 5961–5971. doi: 10.1109/ICCV51070.2023.00548

Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., et al. (2021). Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2021.3130716

Hong, D., Yao, J., Li, C., Meng, D., Yokoya, N., and Chanussot, J. (2023). Decoupled-and-coupled networks: Self-supervised hyperspectral image super-resolution with subpixel fusion. *IEEE Trans. Geosci. Remote Sens.* 61, 1–12. doi: 10.1109/TGRS.2023.3324497

Huang, X., Dong, M., Li, J., and Guo, X. (2022). A 3-d-swin transformer-based hierarchical contrastive learning method for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2022.3202036

Jia, S., Bi, D., Liao, J., Jiang, S., Xu, M., and Zhang, S. (2023). Structure-adaptive convolutional neural network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 61, 1–16. doi: 10.1109/TGRS.2023.3326231

Jin, S., Zhang, F., Zheng, Y., Zhou, L., Zuo, X., Zhang, Z., et al. (2023). Csknn: Cost-sensitive k-nearest neighbor using hyperspectral imaging for identification of wheat varieties. *Comput. Electrical Eng.* 111, 108896. doi: 10.1016/j.compeleceng.2023.108896

Kumar, V., Singh, R. S., and Dua, Y. (2022). Morphologically dilated convolutional neural network for hyperspectral image classification. *Signal Processing: Image Communication* 101, 116549. doi: 10.1016/j.image.2021.116549

Kumbure, M. M., Luukka, P., and Collan, M. (2020). A new fuzzy k-nearest neighbor classifier based on the bonferroni mean. *Pattern Recognition Lett.* 140, 172–178. doi: 10.1016/j.patrec.2020.10.005

Lei, Y., and Tang, K. (2021). Learning rates for stochastic gradient descent with nonconvex objectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 4505–4511. doi: 10.1109/TPAMI.2021.3068154

Li, Z., Chen, G., Li, G., Zhou, L., Pan, X., Zhao, W., et al. (2024c). Dbanet: Dual-branch attention network for hyperspectral remote sensing image classification. *Comput. Electrical Eng.* 118, 109269. doi: 10.1016/j.compeleceng.2024.109269

Li, X., Ding, H., Yuan, H., Zhang, W., Pang, J., Cheng, G., et al. (2024b). Transformer-based visual segmentation: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 46, 12 doi: 10.1109/TPAMI.2024.3434373

Li, C., Zhang, B., Hong, D., Jia, X., Plaza, A., and Chanussot, J. (2024a). Learning disentangled priors for hyperspectral anomaly detection: A coupling model-driven and data-driven paradigm. *IEEE Trans. Neural Networks Learn. Syst.* doi: 10.1109/TNNLS.2024.3401589

Okwuashi, O., and Ndehedehe, C. E. (2020). Deep support vector machine for hyperspectral image classification. *Pattern Recognition* 103, 107298. doi: 10.1016/j.patcog.2020.107298

Ortac, G., and Ozcan, G. (2021). Comparative study of hyperspectral image classification by multidimensional convolutional neural network approaches to improve accuracy. *Expert Syst. Appl.* 182, 115280. doi: 10.1016/j.eswa.2021.115280

Paoletti, M. E., Moreno-Álvarez, S., Xue, Y., Haut, J. M., and Plaza, A. (2023a). Aatt-cnn: Automatical attention-based convolutional neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 61, 1–18. doi: 10.1109/TGRS.2023.3272639

Paoletti, M. E., Moreno-Álvarez, S., Xue, Y., Haut, J. M., and Plaza, A. (2023b). Aatt-cnn: Automatical attention-based convolutional neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 61, 1–18. doi: 10.1109/TGRS.2023.3272639

Peng, Y., Zhang, Y., Tu, B., Li, Q., and Li, W. (2022). Spatial–spectral transformer with cross-attention for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2022.3203476

Pham, Q. T., and Liou, N.-S. (2022). The development of on-line surface defect detection system for jujubes based on hyperspectral images. *Comput. Electron. Agric.* 194, 106743. doi: 10.1016/j.compag.2022.106743

Qi, W., Huang, C., Wang, Y., Zhang, X., Sun, W., and Zhang, L. (2023a). Global-local three-dimensional convolutional transformer network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 61, 1–20. doi: 10.1109/TGRS.2023.3272885

Qi, W., Huang, C., Wang, Y., Zhang, X., Sun, W., and Zhang, L. (2023b). Global-local three-dimensional convolutional transformer network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 61, 1–20. doi: 10.1109/TGRS.2023.3272885

Qiu, Z., Xu, J., Peng, J., and Sun, W. (2023). Cross-channel dynamic spatial-spectral fusion transformer for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 61, 1–12. doi: 10.1109/TGRS.2023.3324730

Roy, S. K., Krishna, G., Dubey, S. R., and Chaudhuri, B. B. (2019). Hybridsn: Exploring 3-d–2-d cnn feature hierarchy for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* 17, 277–281. doi: 10.1109/LGRS.2019.2918719

Roy, S. K., Manna, S., Song, T., and Bruzzone, L. (2020). Attention-based adaptive spectral–spatial kernel resnet for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 59, 7831–7843. doi: 10.1109/TGRS.2020.3043267

Roy, S. K., Mondal, R., Paoletti, M. E., Haut, J. M., and Plaza, A. (2021). Morphological convolutional neural networks for hyperspectral image classification. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 8689–8702. doi: 10.1109/JSTARS.2021.3088228

Sim, J., Dixit, Y., Mcgoverin, C., Oey, I., Frew, R., Reis, M. M., et al. (2024). Machine learning-driven hyperspectral imaging for non-destructive origin verification of green coffee beans across continents, countries, and regions. *Food Control* 156, 110159. doi: 10.1016/j.foodcont.2023.110159

Su, Y., Gao, L., Jiang, M., Plaza, A., Sun, X., and Zhang, B. (2022). Nsckl: Normalized spectral clustering with kernel-based learning for semisupervised hyperspectral image classification. *IEEE Trans. Cybernetics*. 53 (10), 6649–6662. doi: 10.1109/TCYB.2022.3219855

Sun, L., Fang, Y., Chen, Y., Huang, W., Wu, Z., and Jeon, B. (2022a). Multi-structure kelm with attention fusion strategy for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–17. doi: 10.1109/TGRS.2022.3208165

Sun, G., Pan, Z., Zhang, A., Jia, X., Ren, J., Fu, H., et al. (2023). Large kernel spectral and spatial attention networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 61, 1–15. doi: 10.1109/TGRS.2023.3292065

Sun, L., Zhang, H., Zheng, Y., Wu, Z., Ye, Z., and Zhao, H. (2024). Massformer: Memory-augmented spectral-spatial transformer for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 62, 1–15. doi: 10.1109/TGRS.2024.3392264

Sun, L., Zhao, G., Zheng, Y., and Wu, Z. (2022b). Spectral–spatial feature tokenization transformer for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14. doi: 10.1109/TGRS.2022.3144158

Tang, P., Zhang, M., Liu, Z., and Song, R. (2023). Double attention transformer for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* 20, 1–5. doi: 10.1109/LGRS.2023.3248582

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30, 5998–6008. doi: 10.48550/ARXIV.1706.03762

Wang, J., Song, X., Sun, L., Huang, W., and Wang, J. (2020). A novel cubic convolutional neural network for hyperspectral image classification. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 13, 4133–4148. doi: 10.1109/JSTARS.4609443

Wang, X., Tan, K., Du, P., Han, B., and Ding, J. (2023b). A capsule-vectored neural network for hyperspectral image classification. *Knowledge-Based Syst.* 268, 110482. doi: 10.1016/j.knosys.2023.110482

Wang, Y., Yu, X., Wen, X., Li, X., Dong, H., and Zang, S. (2024a). Learning a 3d-cnn and convolution transformers for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* 21, 1–5. doi: 10.1109/LGRS.2024.3365615

Wang, D., Zhang, J., Du, B., Zhang, L., and Tao, D. (2023a). Dcn-t: Dual context network with transformer for hyperspectral image classification. *IEEE Trans. Image Process.* 32, 2536–2551. doi: 10.1109/TIP.2023.3270104

Wang, Z., Zhao, S., Zhao, G., and Song, X. (2024b). Dual-branch domain adaptation few-shot learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 62, 1–16. doi: 10.1109/TGRS.2024.3356199

Wu, X., Hong, D., and Chanussot, J. (2021). Convolutional neural networks for multimodal remote sensing data classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–10. doi: 10.1109/TGRS.2020.3040277

Wu, X., Hong, D., and Chanussot, J. (2022). Uiu-net: U-net in u-net for infrared small object detection. *IEEE Trans. Image Process.* 32, 364–376. doi: 10.1109/TIP.2022.3228497

Xu, R., Dong, X.-M., Li, W., Peng, J., Sun, W., and Xu, Y. (2024). Dbctnet: Double branch convolutiontransformer network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 62, 1–15. doi: 10.1109/TGRS.2024.3368141

Yang, X., Cao, W., Lu, Y., and Zhou, Y. (2022). Hyperspectral image transformer classification networks. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2022.3171551

Yang, K., Sun, H., Zou, C., and Lu, X. (2021). Cross-attention spectral–spatial network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14. doi: 10.1109/TGRS.2021.3133582

Yu, C., Han, R., Song, M., Liu, C., and Chang, C.-I. (2021). Feedback attention-based dense cnn for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–16. doi: 10.1109/TGRS.2020.3040273

Yu, H., Xu, Z., Zheng, K., Hong, D., Yang, H., and Song, M. (2022). Mstnet: A multilevel spectral–spatial transformer network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–13. doi: 10.1109/TGRS.2022.3186400

Zhang, B., Chen, Y., Rong, Y., Xiong, S., and Lu, X. (2023a). Matnet: A combining multi-attention and transformer network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 61, 1–15. doi: 10.1109/TGRS.2023.3254523

Zhang, W., Chen, G., Zhuang, P., Zhao, W., and Zhou, L. (2024a). Catnet: Cascaded attention transformer network for marine species image classification. *Expert Syst. Appl.* 256, 124932. doi: 10.1016/j.eswa.2024.124932

Zhang, W., Li, Z., Li, G., Zhou, L., Zhao, W., and Pan, X. (2024b). Aganet: Attention-guided generative adversarial network for corn hyperspectral images augmentation. *IEEE Trans. Consumer Electron.* doi: 10.1109/TCE.2024.3470846

Zhang, W., Li, Z., Li, G., Zhuang, P., Hou, G., Zhang, Q., et al. (2023b). Gacnet: Generate adversarialdriven cross-aware network for hyperspectral wheat variety identification. *IEEE Trans. Geosci. Remote Sens.* 62, 1–14. doi: 10.1109/TGRS.2023.3347745

Zhang, W., Li, Z., Sun, H.-H., Zhang, Q., Zhuang, P., and Li, C. (2022b). Sstnet: Spatial, spectral, and texture aware attention network using hyperspectral image for corn variety identification. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2022.3225215

Zhang, H., Meng, L., Wei, X., Tang, X., Tang, X., Wang, X., et al. (2019). 1d-convolutional capsule network for hyperspectral image classification. *arXiv preprint arXiv:1903.09834*. doi: 10.48550/arXiv.1903.09834

Zhang, J., Meng, Z., Zhao, F., Liu, H., and Chang, Z. (2022a). Convolution transformer mixer for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2022.3208935

Zhang, X., Su, Y., Gao, L., Bruzzone, L., Gu, X., and Tian, Q. (2023d). A lightweight transformer network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 61, 1–17. doi: 10.1109/TGRS.2023.3297858

Zhang, L., Sun, H., Rao, Z., and Ji, H. (2020). Hyperspectral imaging technology combined with deep forest model to identify frost-damaged rice seeds. *Spectrochimica Acta Part A: Mol. biomolecular Spectrosc.* 229, 117973. doi: 10.1016/j.saa.2019.117973

Zhang, W., Sun, X., Zhou, L., Xie, X., Zhao, W., Liang, Z., et al. (2023c). Dual-branch collaborative learning network for crop disease identification. *Front. Plant Sci.* 14, 1117478. doi: 10.3389/fpls.2023.1117478

Zhang, W., Zhao, W., Li, J., Zhuang, P., Sun, H., Xu, Y., et al. (2024c). Cvanet: Cascaded visual attention network for single image super-resolution. *Neural Networks* 170, 622–634. doi: 10.1016/j.neunet.2023.11.049

Zhao, C., Qin, B., Feng, S., Zhu, W., Sun, W., Li, W., et al. (2023). Hyperspectral image classification with multi-attention transformer and adaptive superpixel segmentation-based active learning. *IEEE Trans. Image Process.* 32, 3606–3621. doi: 10.1109/TIP.2023.3287738

Zheng, X., Sun, H., Lu, X., and Xie, W. (2022). Rotation-invariant attention network for hyperspectral image classification. *IEEE Trans. Image Process.* 31, 4251–4265. doi: 10.1109/TIP.2022.3177322