RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2022, **12**, 8750

Effective band selection of hyperspectral image by an attention mechanism-based convolutional network

Zengwei Zheng,[ac] Yi Liu,[d] Mengzhu He,[ac] Dan Chen,[ac] Lin Sun[ac] and Fengle Zhu *[b]

The selection of effective and representative spectral bands is extremely important in eliminating redundant information and reducing the computational burden for the potential real-time applications of hyperspectral imaging. However, current band selection methods act as a separate procedure before model training and are implemented merely based on extracted average spectra without incorporating spatial information. In this paper, an end-to-end trainable network framework that combines band selection, feature extraction, and model training was proposed based on a 3D CNN (convolutional neural network, CNN) with the attention mechanism embedded in its first layer. The learned band attention vector was adopted as the basis of a band importance indicator to select effective bands. The proposed network was evaluated by two datasets, a regression dataset for predicting the relative chlorophyll content (soil and plant analyzer development, SPAD) of basil leaves and a classification dataset for detecting the drought stress of pepper leaves. A number of calibration models, including SVM, 1D-CNN, 2B-CNN (two-branch CNN), 3D ResNet and the developed network were established for performance comparison. Results showed that the effective bands selected by the proposed attention-based model achieved higher regression R^2 values and classification accuracies not only than the full-spectrum data, but also than the comparative band selection methods, including traditional SPA (successive projections algorithm) and GA (genetic algorithm) methods and the latest 2B-CNN algorithm. In addition, different from the traditional methods, the proposed band selection algorithm can effectively select bands while carrying out model training and can simultaneously take advantage of the original spectral–spatial information. The results confirmed the usefulness of the proposed attention mechanism-based convolutional network for selecting the most effective band combination of hyperspectral images.

Received 16th October 2021
Accepted 11th March 2022

DOI: 10.1039/d1ra07662k

rsc.li/rsc-advances

1. Introduction

A hyperspectral image consists of a series of narrow spectral bands usually within the visible and near-infrared spectral region.[1,2] Compared with RGB images, contiguous spectral images provide high-resolution spatial information as well as rich descriptions about biochemical components of the studied objects.[3] As of today, hyperspectral imaging has been widely used in various fields, such as soil environmental monitoring,[4] plant elements analysis,[5,6] plant species classification,[7,8] medical disease recognition,[9,10] *etc.* However, a typical hyperspectral image usually contains hundreds or even thousands of spectral bands, and the excessive number of bands may give rise to the Hughes phenomenon.[11] In addition, the high dimensional hyperspectral data introduce higher complexity to data processing and pose greater challenges to data storage and analysis.[12] To reduce redundancy or noise of the hyperspectral images, it is of vital significance to select representative spectral bands[13] for retaining the most informative spectral bands with minimum collinearity. The selected representative spectral bands are expected to perform comparatively or superiorly to the full-spectrum data in the subsequent modelling process.

In order to effectively select bands, a great number of researchers have devoted to making their efforts. At present, the most popular and widely applied methods are still the traditional spectral feature selection algorithms, such as genetic algorithm (GA),[14] successive projections algorithm (SPA)[15] and uninformative variable elimination (UVE).[16]

In recent years, with the rapid development of deep learning technology and its wide application in various fields, many researchers also start to select effective spectral bands based on CNN (convolutional neural network, CNN). In addition to the average spectrum, CNN can also directly take the original

[a]*School of Computer & Computing Science, Zhejiang University City College, Hangzhou 310015, China*

[b]*College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou 310023, China. E-mail: zhufl@zjut.edu.cn*

[c]*Intelligent Plant Factory of Zhejiang Province Engineering Lab, Hangzhou 310015, China*

[d]*College of Computer Science & Technology, Zhejiang University, Hangzhou 310027, China*

hyperspectral images as input without flattening the sample in advance as the traditional method, and it is effective in both classification and regression tasks.[17] Acquarelli *et al.*[18] stated that CNN could effectively identify important spectral regions and the iterative training based on deep learning could successfully realize the selection of effective bands. Although both the traditional methods and the approach proposed by Acquarelli *et al.*[18] contributed greatly to the effective band selection, they were all based on the average spectrum manually extracted by ROI (region of interest) from each sample and did not fully take advantage of the rich spatial information of hyperspectral images.

Recently, a few studies proposed the effective band selection based on spectral–spatial integrated CNN model. Feng *et al.*[19] combined mathematical method with CNN and pointed out that the importance of band could be differentiated by the use of the hard thresholding function. Liu *et al.*[1] proposed the 2B-CNN (two-branch CNN) model, consisting of 1D CNN and 2D CNN for extracting spectral and spatial features respectively. The weights of the first convolution layer in the 2D spatial branch were used as indicator of the effective bands, and was successfully applied to three classification datasets, achieving better results than the traditional selection methods based on extracted mean spectra. Similar to 2B-CNN, Torres-Tello *et al.*[20] introduced SHAP (SHapley Additive exPlanations) to identify the effective spectral bands based on the spectral–spatial fusion neural network for predicting moisture content in canola and wheat. For plant disease identification based on hyperspectral images, Nagasubramanian *et al.*[21] used the 3D CNN model to extract rich and consecutive information from spectral–spatial domains simultaneously. At the same time, they implemented band selection based on the calculated gradient magnitude of each band which reflected the band importance. In addition, Ortiz *et al.*[22] proposed the ILFS (Integrated Learning and Feature Selection) framework based on FCN (Fully Convolutional Networks) to automatically screen important input features while training the model. In this process, they used the chain rule to calculate the differential value of the loss function relative to the selected effective bands as the basis for determining the importance of bands. The presentation of the above five models demonstrated that CNN enables us to perform band selection and end-to-end modelling tasks simultaneously, and our method also makes full use of this advantage.

In 2014, the Google Mind team proposed the attention mechanism to classify images and achieved good performance.[23] Attention mechanism is a problem-solving method proposed by imitating human attention. It is a technology that enables the model to focus on important information then fully learn and absorb it, realizing the rapid screening of high-value information from a large amount of data. In recent years, the attention mechanism has been increasingly applied to various scenes in the field of deep learning, such as image recognition, semantic segmentation, machine translation.[24] Nowadays, some researchers have pointed out that the effective band selection of hyperspectral images can be realized through the attention mechanism. For example, Lorenzo and Tulczyjew *et al.*[12] proposed an attention-based CNN architecture, in which the attention module was embedded in each convolutional layer to extract attention heat map, which reflected the importance of each band in the training process and thus served as the basis for the selection of effective bands. Also, Cai and Liu *et al.*[13] proposed a BS-Net architecture consisting of attention module and reconstruction module, through which they explicitly simulated the nonlinear interdependence between spectral bands. Experimental results showed that this method was superior to the most popular band selection methods. The above mentioned two studies focused on analysing remote sensing images at pixel level, whereas the attention-based band selection method for object-scale hyperspectral images analysis was not reported.

In this paper, we explored a method of attention mechanism-based 3D CNN to implement the selection of effective bands along with model training while taking advantage the spatial–spectral continuity of hyperspectral images in object-scale analysis. To evaluate the effectiveness of the proposed band selection approach, both regression and classification tasks were carried out on two hyperspectral image datasets respectively. The former was on the basil leaves to predict the leaf SPAD (soil and plant analyzer development) value, which is highly correlated with the chlorophyll content of plants.[25,26] Chlorophyll plays a central role in the light absorption of plant photosynthesis. The latter was on the pepper leaves to detect the leaf drought stress, which greatly affects photosynthesis and plant growth. A number of calibration models, including SVM, 1D-CNN, 2B-CNN, 3D ResNet and the developed network were established for comparing the performance of selected effective bands with the full-spectrum data. The effectiveness of the proposed band attention method was also assessed by comparing with the traditional selection algorithms of SPA and GA as well as the latest algorithm of 2B-CNN based on the performance of calibration models.

2. Material and methods

2.1 Data acquisition

The image acquisition of datasets was carried out using the hyperspectral imaging system (Fig. 1). The whole equipment consisted of a SNAPSCAN VNIR (visible and near-infrared) hyperspectral imaging camera (IMEC, Leuven, Belgium), a lighting module containing one 150 W studio halogen lights and an image acquisition platform. During the measurement, the blade sample was placed on the platform and kept a vertical distance of 34.5 cm from the camera, ensuring that the blade sample was always in the camera's field of view and both the camera and the sample remained stationary throughout the acquisition process. During the collection, SNAPSCAN sensor translated inside the camera to obtain a 3D data cube. The spectral dimension of each acquired image covered 140 wavelengths in the 470–900 nm range. In order to reduce the influence of varying intensity and distribution of light during the acquisition, all the collected images were calibrated. The reflectance value was calibrated by the following equation:

$$R = \frac{R_0 - D}{W - D} \times 100 \quad (1)$$

where R is the calibrated image, R_0 is the raw hyperspectral image, D is the dark reference image and W is the white reference image.

2.2 Dataset

Basil leaf. The dataset of basil leaf was measured to predict the SPAD value reflecting the relative content of chlorophyll. Since lighting intensity greatly affects the chlorophyll content in plants, sweet basil was cultivated under artificial LED light with three different lighting intensity of 200 ± 5, 135 ± 4, 70 ± 5 μmol m^{-2} s^{-1} respectively. A total of 120 basil crops were planted with 40 ones for each lighting intensity treatment. The photoperiod of 16/8 h (light/dark), the red : blue light ratio of 3, the temperature of 25/20 °C (day/night) were all kept the same for all treatments. For acquiring leaf samples with diverse SPAD values, one or two leaves in the upper, middle, and lower positions of plant canopy were randomly sampled from each basil crop, resulting in 540 samples altogether. The SPAD value of each sample was measured using a SPAD-502 meter (Minolta Camera Co., Osaka, Japan) on six sites on leaf surface to be averaged for the final value. The size of hyperspectral image of each leaf sample was 600 × 800 × 140.

Pepper leaf. The dataset of pepper leaf was targeted at detecting the drought stress of leaves. A total of 20 pepper plants were cultivated for 50 days, then split into two treatments with 10 plants for each. For the drought stressed group, progressive drought with a very low amount of water was applied to maintain the soil moisture content at a highly deficient level for 7 days. For the well-watered group, plants were irrigated with sufficient water of about 200 ml each day. For each treatment with 10 plants, 300 leaf samples were collected respectively, resulting in 600 samples altogether. The size of hyperspectral image of each leaf sample was 120 × 200 × 140.

Dataset partition. In this experiment, for basil leaf dataset, all samples were divided into two different sets, the training set and the test set, which included 420 and 120 samples respectively. The test set was used to test the predictive ability of the trained model. The same division method was used for the pepper leaf dataset, and the division ratio was 3 : 1.

2.3 Image preprocessing

Image preprocessing can eliminate the irrelevant information in the image, enhance the detectability of related information and simplify the data to the maximum extent, so as to improve the reliability of feature extraction. The preprocessing steps mainly included removing the background interference and resizing image, so as to transform the raw hyperspectral image data into the small dataset that could be input into CNN. Fig. 2 is a representative image of basil leaf sample before and after preprocessing.

Background interference removal. As images collection were affected by various factors such as imaging angle and light intensity, there was a lot of noise interference in the background of collected hyperspectral images. In this experiment, as there was a large difference between blade and background at 800 nm, this band was taken as the threshold band with threshold value of 0.15 to retain complete blade information to the maximum. Therefore, pixels with reflectance less than 0.15 at 800 nm were considered as the background and set as 0.

Image resize. The basil leaf hyperspectral images were resized into smaller images with spatial dimension reduced from 600 × 800 to 160 × 160. This operation was not performed on the pepper leaf dataset because the original spatial dimension of each pepper leaf was 120 × 200, which was already small enough to be input into CNN model.

2.4 Overall architecture of the proposed model

As shown in Fig. 3, the proposed attention mechanism-based predictive model was constructed as follows: (1) fed the preprocessed object-scaled hyperspectral image cube into the band attention module for spectral feature extraction; (2) used element-wise multiplication between the original data cube and

Fig. 2 A representative image of basil leaf before (a) and after (b) image preprocessing.

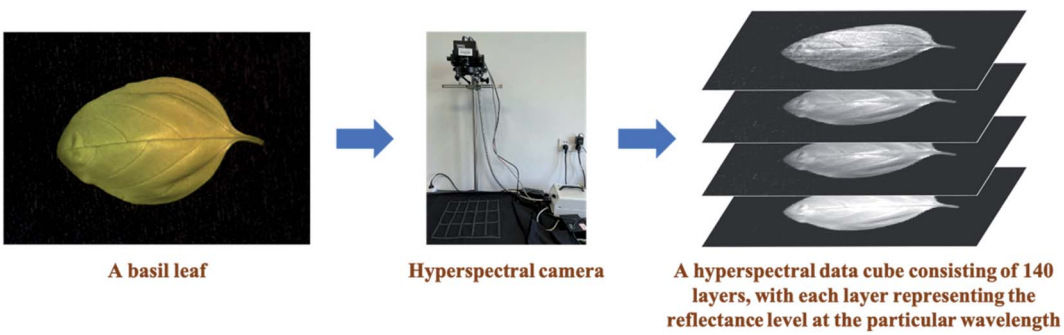

Fig. 1 Illustration of the hyperspectral image acquisition procedure in this study.

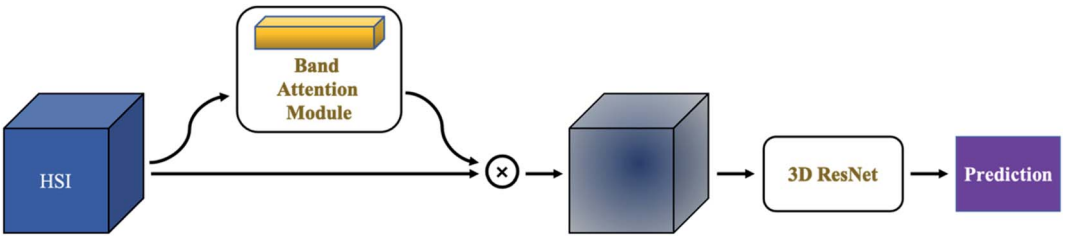

Fig. 3 Overall architecture of the proposed model with attention mechanism in its first layer.

the extracted one-dimensional spectral feature (namely, band attention map) to obtain a new data cube that containing the band weights information; (3) fed the new cube into the 3D ResNet for prediction.

Note that the regression and classification analysis had exactly the same network structure except for the output size of the last linear layer, more specifically, the output size of regression was 1, and for classification it was 2.

2.5 Implementation details of 3D ResNet

3D CNN is developed from 2D CNN. In recent years, an emergence of many studies on related models have been witnessed in the field of image recognition. Given the consecutiveness of spectral images in band dimension, a 3D CNN structure was employed in the prediction task as shown in Fig. 4, with the aim of making full use of the information from both spatial and band dimensions. Since the purpose of this paper was to prove the usefulness of the attention mechanism for effective band selection, the architectural selection of CNN was not the focus. ResNet is one of the most popular network architectures of CNN as it can increase accuracy by adding considerable depth[27] while alleviating the training degradation problem of deep neural network to a large extent. In this experiment, we adopted a 18-layer 3D ResNet architecture as the representative 3D CNN framework.

2.6 Implementation details of band attention module

The attention mechanism in computer vision, from the perspective of implementation methods, mainly includes spatial attention, channel attention and time attention.[28] In this study, in order to fully distinguish the importance of bands, the attention mechanism was applied to the spectral dimension of hyperspectral images. It is worth mentioning that during the process of convolution, the information of different bands were fused together, interfering with the selection of effective spectral bands. Therefore, as shown in Fig. 3, the attention module was embedded in the first layer of the overall prediction model framework.

The implementation details of band attention module are shown in Fig. 5, the procedure worked as follows: (1) computed two different spectral context descriptors H^{b}_{avg} and H^{b}_{max} through average-pooling and max-pooling based on the original input hyperspectral image cube respectively; (2) fed the obtained descriptors into the shared network and generated the corresponding one-dimensional spectral features in turn. The shared network was consisted of MLP (Multilayer Perceptron) with a hidden layer (note that the output dimension of the shared network was consistent with the dimension of the input descriptor); (3) added up the output vectors of the shared MLP for band attention map generation; (4) used the obtained attention map to generate a band weights adjusted hyperspectral image cube by element-wise multiplication. The generation process of band attention map could be described as:

$$M_b(H) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(H)) + \mathrm{MLP}(\mathrm{MaxPool}(H))) = \sigma(W_1(W_0(H^{b}_{avg})) + W_1(W_0(H^{b}_{max}))) \quad (2)$$

where H was the original hyperspectral image data, σ represented the sigmoid function. W_0 and W_1 were the weights in the shared MLP network. For two inputs of H^{b}_{avg} and H^{b}_{max}, the weights were shared. Each layer in the MLP network used the ReLU activation function.

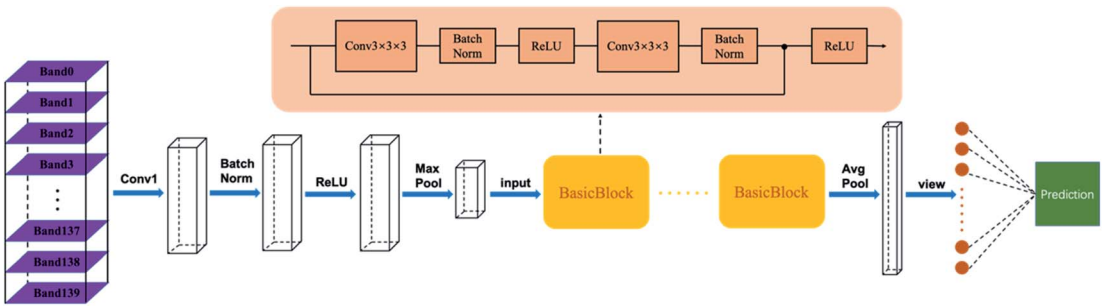

Fig. 4 Implementation details of the 3D ResNet section in the proposed model.

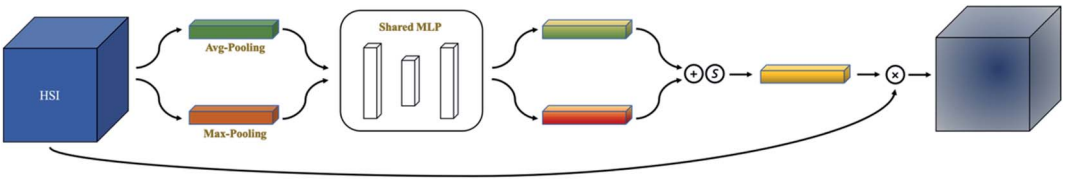

Fig. 5 Implementation details of the attention-based band selection module.

After the generation of band attention map, the final multiplication operation could be summarized as:

$$H' = M_b(H) \otimes H \quad (3)$$

where $\otimes$ denoted element-wise multiplication and H' was the final refined output. During multiplication, the band attention values propagated along the spatial dimension.

2.7 Effective bands selection with the proposed attention-based model

The trained band attention module was used for effective band selection. Here we focus on the band attention maps with the dimension of $m \times 1 \times 1$ (m is equal to the number of bands) obtained by feeding the original hyperspectral image cubes into the band attention-based model. Each band of original hyperspectral image had its corresponding value in these maps, specifically, the values of these maps at index b corresponded to the b-th band of input hyperspectral images. In a nutshell, the importance of bands was evaluated by the indicator I_b defined as:

$$I_b = \mathrm{Softmax}\left(\sum_{i=1}^{n} |w_i^b|\right), \quad b = 1, 2, 3, \ldots\ m. \quad (4)$$

where w_i^b represented the value of the i-th obtained band attention map at index b, n was the number of samples in dataset, m was the total number of bands, which was 140 in this study. I_b was calculated in two steps: firstly, summing up the absolute values at index b of these generated attention maps; secondly, mapping the obtained sum to value between 0 and 1 by Softmax. It's naturally that in the training process of the proposed attention-based 3D CNN model, the embedded band attention mechanism enabled the learning of the weight of each band to adjust the importance among different bands, so as to achieve better modeling performance. Therefore, the value of I_b indicated how much the corresponding band was highlighted by the band attention module and how many contributions it would make to the later prediction task. In this paper, the bands with higher I_b value were selected as effective bands.

In order to prove the usefulness of the proposed attention-based method for selecting effective bands, we compared this method with the common traditional algorithms of SPA, GA and the novel algorithm of 2B-CNN based on the training of SVM, 1D-CNN, 3D ResNet and the proposed network. For the sake of fairness, the number of effective bands selected by the comparison methods were kept consistent with the attention-based model.

2.8 Methods for comparison

SVM for modeling. Support vector machine is a popular chemometric method usually trained on the average spectrum. In this study, the average spectrum of each sample in the dataset was extracted and used to establish SVM models. The radial basis function was employed as the kernel function, and the regularization parameter and the width parameter of the kernel function were determined by a grid search and 10-fold cross validation process.

1D-CNN for modeling. The 1D-CNN proposed by Liu *et al.*[1] is a typical CNN architecture composed of one-dimensional convolution layers, pooling layers and a fully-connected layer, which considering only spectral information. The detailed structural parameters of the 1D-CNN model were listed in the paper of Liu *et al.*[1] and the same structure was used in this study.

SPA for effective band selection. Successive projections algorithm is a classical variable selection algorithm to select the combination of effective bands with the least collinearity. It is now widely used in the field of agriculture and food to select the effective bands. The R^2 was taken as the evaluation metric for band selection of basil leaf dataset and for pepper leaf dataset it was accuracy.

GA for effective band selection. The selection of effective bands always involves combinatorial optimization, in which the genetic algorithm can usually get better results. For better comparison, an improved genetic algorithm with modified mutation and crossover algorithm functions was adopted in this study. On the premise of ensuring fairness, the probability of 0 and 1 appearing randomly was changed, so that the expected number of band combinations was selected in each generation. The R^2 and accuracy were used as the evaluation metrics of basil leaf and pepper leaf datasets respectively, the number of epochs of the algorithm was set as 300, and LightGBM (light gradient boosting machine) framework was used to quickly obtain the results.

2B-CNN for modeling and effective band selection. 2B-CNN is a novel method developed by Liu *et al.*[1] This network was constructed by two branch, a 1D convolutional branch for spectral feature extraction and a 2D convolutional branch for spatial feature extraction. The better classification performance was obtained by the fusion of spectral and spatial features. Based on this model, the weights learned by the 2D convolutional branch of 2B-CNN were used as the indicator of effective bands. The same network structure and experimental settings were used in this study.

2.9 Experimental setup

During the training of the proposed attention-based 3D CNN model, all the hyper-parameters were set in the same way. The batch size was 8 and the number of training epochs was 70, the value of momentum of batch gradient descent optimizer and weight decay were set to 0.9 and 0.001 respectively, the learning rate automatically decayed according to the number of iterations. The models were implemented in Python 3.7 with Pytorch 1.0.0. In terms of evaluation strategy, standard metrics of regression and classification tasks were used. Specifically, for regression analysis, the metrics were RMSE and R^2, while for classification analysis, accuracy, precision and sensitivity were used. These metrics were calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} w_i(y_i - \hat{y}_i)^2} \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} w_i(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} w_i(y_i - \bar{y}_i)^2} \quad (6)$$

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \quad (7)$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (8)$$

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (9)$$

in which TP, TN, FP, FN represented the true positive, true negative, false positive and false negative of the confusion matrix, respectively.

3. Results and discussion

In this section, we objectively evaluated the performance of the attention-based band selection model for hyperspectral images according to the datasets of basil leaf and pepper leaf. Firstly, the performance of the proposed network and comparison models established with full-spectrum data was analysed, along with discussing the impact of the band attention module on the overall model effectiveness. Next, the selected effective bands by the proposed method were analysed both in chemical significance and modelling performance, comparing with full bands. Finally, the effectiveness of the proposed attention-based selection method was evaluated by comparing with other band selection methods.

3.1 Comparison of prediction performance on full-spectrum

The regression and classification performance of the proposed attention-based 3D ResNet, SVM, 1D-CNN, 2B-CNN and 3D ResNet models established with full-spectrum data are presented in Table 1. For both datasets, the proposed attention-based 3D ResNet models achieved the best results. In terms of regression analysis on basil leaf dataset, compared with the SVM, 1D-CNN, 2B-CNN and 3D ResNet models, the R^2 improvement obtained by the proposed network was 0.065, 0.527, 0.298 and 0.002, respectively. The gain was quite significant when compared with 1D-CNN and 2B-CNN, indicating that these two kinds of convolutional network architectures were not suitable for regression analysis in this study. In terms of classification analysis on pepper leaf dataset, the accuracy improvement based on the attention-based 3D ResNet was 17.10%, 16.51%, 13.89% and 2.78% compared with SVM, 1D-CNN, 2B-CNN and 3D ResNet models respectively. During the progression of drought stress on pepper leaves, the biochemical and structural changes occurred in stressed leaves altered their spectral responses, the morphological (*e.g.*, shrivel, curl) and textural variations of leaves induced by drought stress altered their spatial information. Hence the mere learning of spectral information by SVM and 1D-CNN performed the worst, the learning of both spectral and spatial information by 2B-CNN manifested better results. And the joint extraction of spectral–spatial information by 3D ResNet and attention-based 3D ResNet performed the best.

In addition, the effect of the added band attention module in 3D ResNet on the consistency and convergence of model training needed in-depth discussion. For basil leaf and pepper leaf datasets, the original hyperspectral images were fed into the 3D ResNet with and without the band attention module for comparison, in which other hyper-parameter configurations were kept the same. Results showed that the addition of attention module would not increase the number of epochs or training time, instead, the performance of regression and classification tasks could even be improved to some extent. In addition, the embedding of this band attention module allowed us to select the most effective bands while training. Therefore, it was demonstrated that the quality of the overall modeling performance was improved through this embedding.

3.2 Effective bands selection by the proposed attention-based network

We extensively analysed the selected effective bands in this section. As described in the Section 2.6, the importance of spectral bands was distinguished by the band attention module,

Table 1 Prediction results of the attention-based 3D ResNet and the comparison models on full-spectrum data

Dataset	Metric	SVM	1D-CNN	2B-CNN	3D ResNet	Proposed model
Basil	RMSE	2.890	5.552	4.458	2.420	2.379
	R^2	0.825	0.354	0.583	0.878	0.881
Pepper	Accuracy (%)	56.79	57.38	60.00	71.11	73.89
	Precision (%)	55.65	56.19	59.44	63.91	77.63
	Sensitivity (%)	62.31	68.14	66.06	95.51	66.29

and the value of I_b was used for quantitative assessment. Fig. 6 shows the relative I_b value of each band obtained by feeding the original hyperspectral images into the band attention module. In these two heat maps, the horizontal axis represents the band; the vertical axis represents the randomly sampled samples; while the colour bar represents the value of I_b. As the band attention method shows considerable stability that the locations of important bands were basically unchanged for all samples in the dataset, the overall distribution can be represented by a small number (10 in Fig. 6) of samples. The top 13 bands with higher I_b values were selected, and their distribution on the average spectrum is displayed in Fig. 7.

It can be clearly seen from Fig. 7 that the effective band subset obtained by the band attention module had few continuous bands and the distribution of these bands was relatively uniform. Based on the fact that adjacent spectral bands are usually highly correlated, the selected bands should contain less redundancy, which might be beneficial for the prediction performance of the proposed model. In addition, these bands were distributed in positions with large fluctuations in the average spectrum curve, indicating that the most important information of hyperspectral images was captured by the proposed band selection method.

To analyse the selected effective bands in more depth, wavelengths assignment was carried out. The total bands were labelled from 0 to 139. For basil leaf dataset, the selected band subset was [5, 6, 8, 25, 30, 45, 58, 73, 78, 82, 89, 111, 116], corresponding to the wavelengths of 486 nm, 489 nm, 496 nm, 553 nm, 569 nm, 625 nm, 665 nm, 713 nm, 728 nm, 741 nm, 764 nm, 828 nm and 841 nm respectively. These wavelengths were consistent with the strong reflection of green light (553 nm, 569 nm), absorption of red light (625 nm, 665 nm), abrupt reflection increment at the red edge region (728 nm, 741 nm, 764 nm) of a typical leaf's spectral profile, due to the porphyrin ring in chlorophyll molecules in basil leaves (Walsh 2020). Besides, the effective wavelengths of 728 nm, 741 nm, 764 nm could be attributed to the fourth overtone of the methyl ($-CH_3$), methylene ($-CH_2$) and methine (–CH) groups stretching vibration of the chlorophyll molecules.[29] For pepper leaf dataset, the selected subset was [0, 2, 6, 10, 19, 31, 58, 59, 65, 70, 79, 94, 114], corresponding to wavelengths of 468 nm, 476 nm, 489 nm, 501 nm, 534 nm, 573 nm, 665 nm, 667 nm, 688 nm, 703 nm, 732 nm, 780 nm and 836 nm. These wavelengths were correlated with the biochemical (moisture, pigments, *etc.*) and cell structural changes in pepper leaves undergoing drought stress. The 468 nm and 476 nm in blue light region were attributed to the absorption of chlorophylls, carotenes and xanthophylls. The 534 nm and 573 nm in green light region were due to the combined effects of chlorophylls refection and anthocyanins absorption. The 665 nm, 667 nm and 688 nm in red light region were ascribed to the absorption of chlorophylls.[30] The 732 nm, 780 nm and 836 nm were related to the third overtone of O–H group, forth overtone of C–H group and third overtone of N–H group respectively in various biochemical components (moisture, amino acids, carbohydrates, *etc.*) in pepper leaves.[29] The correlation analysis of the effective wavelengths with the absorption of chemical functional groups in leaf biochemical components demonstrated that our proposed attention-based convolution network could select the key spectral wavelengths which were not only representative and interpretable, but also oriented at solving specific analytical problems.

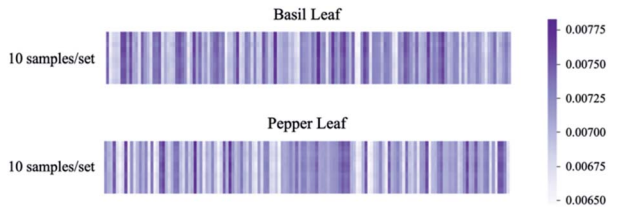

Fig. 6 The obtained band importance indicators using the attention mechanism-based network.

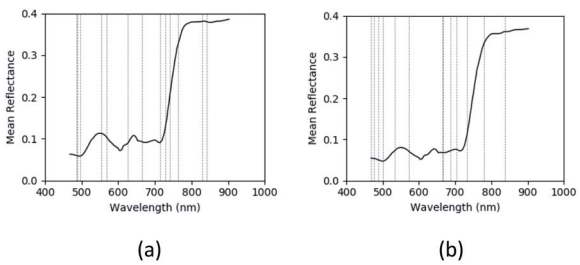

Fig. 7 Effective bands of basil leaf dataset (a) and pepper leaf dataset (b) identified by attention-based band selection model.

In order to better evaluate the efficacy and importance of extracted effective bands by the developed attention-based band selection model, hyperspectral image data before and after band extraction were put into the same models for performance comparison, including the spectral–spatial models of 2B-CNN, 3D ResNet and attention-based 3D ResNet and the spectral models of SVM and 1D-CNN. The results are shown in Table 2. It can be seen that, for all modelling methods in both datasets, the band subset selected by the proposed method could give better results than the original hyperspectral image data (without band selection). This phenomenon could be interpreted as that for the original hyperspectral image data, a lot of redundancy and noise were contained in the consecutive spectral bands, which would cause interference to the prediction analysis. The proposed band selection removed this interference information while retaining the most important and representative bands, which not only compressed data and improved training efficiency, but also improved the performance of prediction analysis.

3.3 Comparison with other band selection methods

As one of the main methods for dimensionality reduction of hyperspectral images, band selection has been actively researched in the field of hyperspectral image analysis. Thus, we compared our attention-based band selection algorithm with two traditional methods of SPA and GA as well as a newly proposed method of 2B-CNN. For fair comparison, the number

Table 2 Comparison on the performance of the original full spectral bands and the corresponding effective band subset obtained by the band attention module

		SVM		1D-CNN		2B-CNN		3D ResNet		Proposed model	
Dataset	Metric	Full	Subset	Full	Subset	Full	Subset	Full	Subset	Full	Subset
Basil	RMSE	2.890	2.600	5.552	5.206	4.458	3.499	2.420	2.133	2.379	2.046
	R^2	0.825	**0.858**	0.354	**0.432**	0.583	**0.743**	0.878	**0.904**	0.881	**0.912**
Pepper	Accuracy (%)	56.79	**60.95**	57.38	**61.14**	60.00	**65.56**	71.11	**76.67**	73.89	**76.90**
	Precision (%)	55.65	60.56	56.19	59.76	59.44	61.82	63.91	71.17	77.63	73.17
	Sensitivity (%)	62.31	61.22	68.14	68.26	66.06	70.23	95.51	88.76	66.29	85.31

of effective bands selected by all algorithms mentioned above would remain the same, which was 13 for both datasets in this study.

For the two datasets, the selected band subset (total bands labelled from 0 to 139) and their corresponding wavelengths by the three comparative methods and the attention-based method are shown in Table 3. It can be observed that the effective bands selected by above algorithms did not show great similarity. This was due to the difference in their selection principle. The SPA aimed to calculate the bands with maximal discriminative power in spectral domain, while the GA aimed to select the combination of bands with the best performance, in which the selection was random. Both of SPA and GA relied on the average spectrum and only the information of spectral domain was considered in band selection. Meanwhile, the 2B-CNN was constructed based on the spatial–spectral fusion of hyperspectral images when performing modeling task, but the effective bands were selected only based on the spatial branch of 2D CNN. In contrast, the method proposed in this study took both spatial and spectral information into account for effective band selection while carrying out prediction analysis.

To assess the effectiveness of these band subsets selected by different methods, calibration models were established, the results are shown in Table 4 and visualized in Fig. 8. For comparison among different calibration models, similar to the full-spectrum results in Table 1, for both datasets the proposed attention-based 3D ResNet model performed the best, and its performance difference with the 3D ResNet was insignificant. Overall, the 3D CNN models performed obviously better than others, demonstrating the merit of joint feature extraction from spectral–spatial dimensions of hyperspectral images.

For comparison among different band selection methods, obviously the proposed band attention method showed superiority to SPA, GA and 2B-CNN for all models on both regression and classification tasks. In specific, for basil leaf dataset, the R^2 obtained in SVM, 1D-CNN, 2B-CNN, 3D ResNet and the proposed model based on the spectral subset selected by the band attention module were 0.858, 0.432, 0.743, 0.904 and 0.912 respectively. And the averaged relative improvement of the attention-based method compared with SPA, GA and 2B-CNN calculated across all models was 10.39%, 7.91% and 2.48% respectively. For pepper leaf dataset, the accuracy achieved by SVM, 1D-CNN, 2B-CNN, 3D ResNet and the proposed model established with the spectral subset selected by the attention module were 60.95%, 61.14%, 65.56%, 76.67% and 76.90% respectively. And the averaged relative improvement of the attention-based method compared with SPA, GA and 2B-CNN calculated across all models was 5.98%, 4.19% and 2.12% respectively. Noticeably, although the performance of 2B-CNN was not good in regression analysis on the basil leaf dataset, the band subset selected by 2B-CNN still performed well on both datasets, superior to the traditional methods of GA and SPA. This was probably due to that the 2B-CNN was originally designed for classification and band selection.[1] Thus, the advantage of the proposed model over 2B-CNN could be seen not only on effective band selection, but also on both classification and regression tasks. Moreover, for both GA and SPA, the effective bands were selected based on average spectra, then the prediction analysis was carried out in a separated procedure, which required a lot of manual interventions. In comparison, our proposed band attention module was directly embedded in the modeling analysis of convolutional network, which could conduct model training and band selection simultaneously while achieving good results. Therefore, our proposed band selection model is of practical significance both in performance and convenience, and may be applicable to other hyperspectral

Table 3 Band subsets of basil leaf and pepper leaf datasets obtained by different methods

Dataset	Method	Selected subset (labelled bands from 0 to 139)	Corresponding wavelengths (nm)
Basil	SPA	1, 4, 10, 11, 15, 36, 46, 47, 52, 59, 67, 115, 129	471, 483, 501, 505, 519, 591, 629, 632, 647, 667, 694, 839, 876
	GA	5, 25, 28, 38, 45, 55, 59, 68, 85, 96, 121, 129, 137	486, 553, 563, 596, 625, 656, 667, 697, 750, 785, 854, 876, 896
	2B-CNN	1, 9, 10, 27, 59, 64, 69, 77, 85, 92, 106, 111, 139	471, 500, 501, 559, 667, 684, 700, 726, 750, 774, 814, 828, 902
	Proposed	5, 6, 8, 25, 30, 45, 58, 73, 78, 82, 89, 111, 116	486, 489, 496, 553, 569, 625, 665, 713, 728, 741, 764, 828, 841
Pepper	SPA	0, 2, 3, 4, 6, 9, 12, 26, 47, 58, 65, 74, 129	468, 476, 479, 483, 489, 500, 508, 556, 632, 665, 688, 716, 876
	GA	12, 33, 41, 50, 51, 56, 65, 75, 85, 91, 104, 110, 114	508, 580, 606, 641, 643, 659, 688, 719, 750, 771, 808, 825, 836
	2B-CNN	14, 18, 34, 47, 74, 75, 91, 95, 96, 104, 116, 123, 127	515, 531, 583, 632, 716, 719, 771, 783, 785, 808, 841, 859, 871
	Proposed	0, 2, 6, 10, 19, 31, 58, 59, 65, 70, 79, 94, 114	468, 476, 489, 501, 534, 573, 665, 667, 688, 703, 732, 780, 836

Table 4 Comparison on the performance of multiple models trained on the selected band subset obtained by band attention method and other selection methods

Models	Band selection methods	Basil leaf dataset		Pepper leaf dataset		
		RMSE	R^2	Accuracy (%)	Precision (%)	Sensitivity (%)
SVM	SPA	2.829	0.832	56.61	55.60	59.89
	GA	2.792	0.837	57.38	58.49	60.32
	2B-CNN	2.698	0.847	59.76	57.38	65.53
	Proposed	2.600	**0.858**	**60.95**	60.56	61.22
1D-CNN	SPA	5.470	0.373	58.57	57.38	64.13
	GA	5.423	0.383	59.17	58.57	60.49
	2B-CNN	5.319	0.407	60.36	57.98	77.16
	Proposed	5.206	**0.432**	**61.14**	59.76	68.26
2B-CNN	SPA	4.399	0.594	61.11	65.93	65.64
	GA	4.241	0.623	63.32	59.67	77.95
	2B-CNN	3.575	0.732	64.45	61.67	72.31
	Proposed	3.499	**0.743**	**65.56**	61.82	70.23
3D ResNet	SPA	2.582	0.860	71.67	69.39	76.40
	GA	2.315	0.887	72.78	67.24	87.64
	2B-CNN	2.308	0.888	73.33	65.17	77.33
	Proposed	2.133	**0.904**	**76.67**	71.17	88.76
Proposed model	SPA	2.314	0.887	74.22	73.21	76.91
	GA	2.343	0.885	75.00	68.97	89.89
	2B-CNN	2.217	0.897	76.11	77.38	73.03
	Proposed	2.046	**0.912**	**76.90**	73.17	85.31

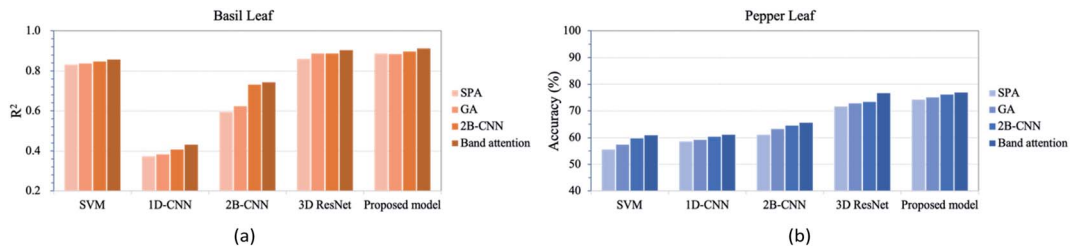

Fig. 8 Prediction results obtained by different models trained on the effective bands selected by different band selection methods. (a) Basil leaf dataset for regression analysis. (b) Pepper leaf dataset for classification analysis.

datasets in other plant science tasks, such as disease detection, heat stress identification and so on. In addition, with the largely reduced number of bands, more portable spectral imaging equipment could be developed to improve data collection efficiency in practical application. In future work, more samples need to be collected to further improve the performance of the proposed band selection model.

4. Conclusions

In this paper, we developed a network framework for band selection of hyperspectral images based on attention mechanism and 3D ResNet, and demonstrated how its trained band attention module can be used for effective band selection. Collected hyperspectral images of basil leaf and pepper leaf were used as the inputs of attention-based 3D ResNet for spectral–spatial information extraction and predictive task execution. The full-spectrum experimental results showed that the proposed model yielded superior classification and regression performance compared with SVM, 1D-CNN, 2B-CNN and 3D ResNet models.

For both datasets, the effective bands selected by the proposed method were representative and interpretable, achieving better prediction performance than the full-spectrum data. Also, the proposed attention-based selection method performed better not only than the traditional SPA and GA but also than the latest 2B-CNN algorithm, this proved the good band feature extraction ability of attention mechanism. The overall results indicated that the proposed framework not only performed well to effectively select the key and representative wavelengths, but also was convenient and flexible to be implemented while carrying on model training. To sum up, the proposed attention-based 3D ResNet is a promising band selection method of hyperspectral image and has great development potential.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive suggestions and criticisms. We would be grateful for the support by the Natural Science Foundation of Zhejiang Province (Grant No. LQ22C130004; LGN22F020002; LGN21F020002; LGN20F020003) and the Key Research and Development Program of Zhejiang Province (Grant No. 2022C03037).

References

1 Y. Liu, S. Zhou, W. Han, W. Liu, Z. Qiu and C. Li, *Anal. Chim. Acta*, 2019, **1086**, 46–54.
2 F. Taherkhani, J. Dawson and N. M. Nasrabadi, *International Conference on Biometrics*, 2019, 1–8.
3 Y. Zhan, D. Hu, H. Xing and X. Yu, *IEEE Geosci. Remote Sens. Lett.*, 2017, **14**, 2365–2369.
4 L. Wei, Y. Zhang, Z. Yuan, Z. Wang, F. Yin and L. Cao, *Appl. Sci.*, 2020, **10**, 2076–3417.
5 W. Liu, M. Li, M. Zhang, D. Wang, Z. Guo, S. Long, S. Yang, H. Wang, W. Li, Y. Hu, Y. Wei and H. Xiao, *Ecosys. Health Sustain.*, 2020, **6**, 1726211.
6 X. Zhou, J. Sun, Y. Tian, B. Lu, Y. Hang and Q. Chen, *Int. J. Rem. Sens.*, 2020, **41**, 2263–2276.
7 B. Zhang, L. Zhao and X. Zhang, *Remote Sens. Environ.*, 2020, **247**, 111938.
8 L. Bing, J. Sun, N. Yang, X. Wu and X. Zhou, *J. Food Process Eng.*, 2021, **44**, e13584.
9 J. Laimer, E. Bruckmoser, T. Helten, B. Kofler, B. Zelger, A. Brunner, B. Zelger, C. W. Huck, M. Tappert, D. Rogge, M. Schirmer and J. D. Pallua, *J. Biophotonics*, 2021, **14**, e202000424.
10 S. Lemmens, T. Van Craenendonck, J. Van Eijgen, L. De Groef, R. Bruffaerts, D. A. de Jesus, W. Charle, M. Jayapala, G. Sunaric-Megevand, A. Standaert, J. Theunis, K. Van Keer, M. Vandenbulcke, L. Moons, R. Vandenberghe, P. De Boever and I. Stalmans, *Alzheimer's Res. Ther.*, 2020, **12**(1), 144.
11 J. Wang, J. Zhou, W. Huang and J. F. Chen, *IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 3820–3823.
12 P. R. Lorenzo, L. Tulczyjew, M. Marcinkiewicz and J. Nalepa, *IEEE Access*, 2020, **8**, 42384–42403.
13 Y. Cai, X. Liu and Z. Cai, *IEEE Trans. Geosci. Rem. Sens.*, 2020, **58**, 1969–1984.
14 S. Li, H. Wu, D. Wan and J. Zhu, *Knowl. Base Syst.*, 2011, **24**, 40–48.
15 M. J. C. Pontes, R. K. H. Galvao, M. C. U. Araujo, P. Nogueira, T. Moreira, O. D. P. Neto, G. E. Jose and T. C. B. Saldanha, *Chemom. Intell. Lab. Syst.*, 2005, **78**, 11–18.
16 S. Wang, M. Huang and Q. Zhu, *Comput. Electron. Agric.*, 2012, **80**, 1–7.
17 W. Hu, Y. Huang, L. Wei, F. Zhang and H. Li, *J. Sens.*, 2015, **2015**, 258619.
18 J. Acquarelli, T. van Laarhoven, J. Gerretzen, T. N. Tran, L. M. C. Buydens and E. Marchiori, *Anal. Chim. Acta*, 2017, **954**, 22–31.
19 J. Feng, D. Li, J. Chen, X. Zhang, X. Tang and X. Wu, *IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 3804–3807.
20 J. W. Torres-Tello and S. Ko, *Biosyst. Eng.*, 2021, **210**, 91–103.
21 K. Nagasubramanian, S. Jones, A. K. Singh, S. Sarkar, A. Singh and B. Ganapathysubramanian, *Plant Methods*, 2019, **15**, 98.
22 A. Ortiz, A. Granados, O. Fuentes, C. Kiekintyeld, D. Rosario and Z. Bell, *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, 2018, pp. 1277–1286.
23 V. Mnih, N. Heess, A. Graves and K. Kavukcuoglu, *Advances in Neural Information Processing Systems*, 2014, vol. 27, pp. 2204–2212.
24 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, *Advances in Neural Information Processing Systems*, 2017, vol. 30, pp. 5998–6008.
25 F. H. Ruiz-Espinoza, B. Murillo-Amador, J. L. Garcia-Hernandez, L. Fenech-Larios, E. O. Rueda-Puente, E. Troyo-Dieguez, C. Kaya and A. Beltran-Morales, *J. Plant Nutr.*, 2010, **33**, 423–438.
26 J. Uddling, J. Gelang-Alfredsson, K. Piikki and H. Pleijel, *Photosynth. Res.*, 2007, **91**, 37–46.
27 K. He, X. Zhang, S. Ren and J. Sun, *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
28 S. Woo, J. Park, J.-Y. Lee and I. S. Kweon, *Proceedings of the European Conference on Computer Vision*, 2018, pp. 3–19.
29 H. Y. Cen and Y. He, *Trends Food Sci. Technol.*, 2007, **18**, 72–83.
30 K. B. Walsh, J. Blasco, M. Zude-Sasse and X. D. Sun, *Postharvest Biol. Technol.*, 2020, **168**, 111246.