The Crop Journal 10 (2022) 1460–1469

Contents lists available at ScienceDirect

The Crop Journal

journal homepage: www.keaipublishing.com/en/journals/the-crop-journal/

Stacked spectral feature space patch: An advanced spectral representation for precise crop classification based on convolutional neural network

Hui Chen [a], Yue'an Qiu [a], Dameng Yin [b], Jin Chen [a,*], Xuehong Chen [a], Shuaijun Liu [a], Licong Liu [a]

[a] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[b] Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

ARTICLE INFO

Article history:
Received 30 August 2021
Revised 5 December 2021
Accepted 11 December 2021
Available online 3 February 2022

Keywords:
Crop classification
Convolutional neural network
Handcrafted feature
Stacked spectral feature space patch
Spectral information

ABSTRACT

Spectral and spatial features in remotely sensed data play an irreplaceable role in classifying crop types for precision agriculture. Despite the thriving establishment of the handcrafted features, designing or selecting such features valid for specific crop types requires prior knowledge and thus remains an open challenge. Convolutional neural networks (CNNs) can effectively overcome this issue with their advanced ability to generate high-level features automatically but are still inadequate in mining spectral features compared to mining spatial features. This study proposed an enhanced spectral feature called Stacked Spectral Feature Space Patch (SSFSP) for CNN-based crop classification. SSFSP is a stack of two-dimensional (2D) gridded spectral feature images that record various crop types' spatial and intensity distribution characteristics in a 2D feature space consisting of two spectral bands. SSFSP can be input into 2D-CNNs to support the simultaneous mining of spectral and spatial features, as the spectral features are successfully converted to 2D images that can be processed by CNN. We tested the performance of SSFSP by using it as the input to seven CNN models and one multilayer perceptron model for crop type classification compared to using conventional spectral features as input. Using high spatial resolution hyperspectral datasets at three sites, the comparative study demonstrated that SSFSP outperforms conventional spectral features regarding classification accuracy, robustness, and training efficiency. The theoretical analysis summarizes three reasons for its excellent performance. First, SSFSP mines the spectral interrelationship with feature generality, which reduces the required number of training samples. Second, the intra-class variance can be largely reduced by grid partitioning. Third, SSFSP is a highly sparse feature, which reduces the dependence on the CNN model structure and enables early and fast convergence in model training. In conclusion, SSFSP has great potential for practical crop classification in precision agriculture.

1. Introduction

Accurate and timely crop mapping is fundamental to precision agriculture. Compared with conventional time-consuming and labour-intensive *in situ* surveys [1], remotely sensed (RS) data has been widely adopted as an irreplaceable data source for mapping crop types and monitoring their dynamics [2–4]. This is because RS data contain rich spectral features that characterize crop leaf pigment, leaf water content and canopy structure [5], and spatial features that reflect crop planting morphology and texture, particularly when high spatial resolution images are considered [6,7].

To reflect the unique characteristics of particular crop types for mapping, efforts have been made to design features based on prior knowledge or statistical analysis, thus called handcrafted features. In terms of spectral features, spectral similarity [8] and spectral indices [9] have been designed to maximize the differences among different crop types as well as other land cover types, and have been successfully applied to classification. In terms of spatial features, a typical way is to generate grey-level co-occurrence matrix (GLCM) texture so that the neighbouring pixels are taken into account and contribute to distinguishing crops with similar spectra but different spatial properties [6,10,11]. Furthermore, spectral and

* Corresponding author.
E-mail address: chenjin@bnu.edu.cn (J. Chen).

https://doi.org/10.1016/j.cj.2021.12.011

spatial features can be integrated, e.g., by employing high spatial resolution hyperspectral data [12–15] or combining high spatial resolution unmanned aerial vehicle (UAV) images with high spectral resolution Sentinel-2 satellite images [16]. However, despite the well-established handcrafted features, designing or selecting such features valid for specific crop types requires prior knowledge and thus remains an open challenge.

In the last decade, convolutional neural networks (CNNs) have shown superiority over feature handcrafting [17,18] because of their advanced ability to automatically extract high-level domain-oriented features from input data that are discriminative enough for classification [19–21]. Nevertheless, different types of CNNs, using 1D, 2D or 3D convolution, have their trade-off in feature extraction. In general, 1D-CNNs can extract spectral features but do not utilize spatial information [22,23]. By contrast, 2D-CNNs can perform convolutional operations in spatial dimensions to fully mine spatial features but are insufficient in mining spectral features [24,25]. Although 3D-CNNs can integrate both spatial and spectral dimensions [26], they require a larger sample size and the training of a large number of parameters, which significantly increases the computational burden [27]. In recent years, new strategies have been introduced, such as multi-scale input [28], multi-temporal input with attention mechanism [29] and modeling the relationship between low-level and high-level crop types [30]. However, the same training challenge lies in these CNNs with sophisticated structures besides 3D convolution, owing to the general trade-off between model complexity and the difficulty of training. Moreover, selecting an appropriate CNN poses challenges to users with less deep-learning experience, such as farmers and crop scientists who expect to apply the techniques to crop mapping. It would be more beneficial to them if simple CNNs can achieve similar results as complex CNNs in terms of feature extraction and final classification.

Recently, several studies have achieved satisfactory classification accuracy using handcrafted features, instead of raw data, as input to CNNs, even for the ones with simple structures [31–33]. For example, in the land cover classification task [34], a handcrafted spatial feature called Multiscale Covariance Maps (MCM) was used as the input of a 2D-CNN model, resulting in better classification results than using raw data. Two other handcrafted spatial features, the extended morphological profiles (EMP) and the Gabor features, were also successfully combined with CNNs [35–37]. These successful attempts suggest that combining handcrafted spatial features with the 2D-CNNs that automatically extract high-level spatial features is a practical classification strategy. Unfortunately, combining handcrafted spectral features with 2D-CNNs to exploit spatial and spectral features simultaneously has not been widely explored yet.

In this study, we developed an enhanced spectral feature called Stacked Spectral Feature Space Patch (SSFSP) as the input for 2D-CNN-based crop classifiers, by which the shortcomings of 2D-CNN in the mining of spectral features can be significantly addressed. Our aim is to better utilize spatial and spectral features simultaneously, reduce the dependence of feature extraction on CNNs, accelerate the training process, and eventually improve crop classification.

2. Methodology

2.1. Stacked spectral feature space patch (SSFSP)

2.1.1. Main idea

In remote sensing, features can be defined in terms of the measurable properties of a phenomenon observed in RS data [38], which mainly include spectral and spatial properties. These features can be incorporated into a classification system, and each class has its unique feature pattern. The feature vector V_{FS}, usually called the feature space, can be expressed as a multi-dimensional coordinate system:

$$V_{FS} = \{f_1, f_2, \cdots, f_n\} \quad (1)$$

Take the 2D spectral feature space as an example. Two spectral bands (Fig. 1a) are transformed collectively into a spectral space $\{(f_1, f_2)\}$, where the two axes mark the original spectral values. The new feature space is composed of two spectral bands, and thus the space mainly shows the spectral distribution characteristics of various land cover types in this two-band feature space. Accordingly, the pixels of different crop types in RS data can be projected to the new spectral feature space to show the spectral distribution characteristics of crop types (Fig. 1b): (1) different crop types are distributed in relatively fixed areas with their clusters and (2) within these clusters, the density of points also varies spatially. These two characteristics can be regarded as describable features in the new spectral feature space, the former is referred to as the spatial-distribution feature and the latter as the intensity-distribution feature. The new spectral feature space can be combined with a 2D-CNN to support simultaneous mining of these two describable features if that spectral feature space can be converted to a 2D image that the CNN can process.

For making this type of spectral feature space directly usable by CNNs, the 2D spectral feature space is required to be converted into a grid image (called gridded spectral feature image hereafter) so that the spatial-distribution feature and intensity-distribution feature can be represented explicitly. The spatial-distribution feature can be described by the locations of pixels at the grids of the new gridded spectral feature space (Fig. 1c). Note that in the gridded spectral feature space, there are multiple points in some grids (Fig. 1c) because many pixels hold similar values in both spectral bands. This phenomenon corresponds to the frequency of the point distribution, namely the intensity distribution that we wish to exploit. Accordingly, we calculate the frequency of the points within each grid as the grid values, forming the final gridded spectral feature image (Fig. 1d).

Such a transformation can be performed for any two spectral bands to generate a new set of 2D gridded spectral feature images and form a Stacked Spectral Feature Space (SSFS) V_{SSFS} as:

$$V_{SSFS} = \{(f_1, f_2), (f_1, f_3), \cdots, (f_{n-1}, f_n)\} \quad (2)$$

SSFS forms a patch at the local scale. It can be used in the patch-based classification process and the pixel-based classification process if focusing on the central pixel of the patch (one patch corresponding to a moving window). In this way, the spectral information is not only further exploited in CNN with its powerful spatial feature mining capability, but the neighbourhood information within a patch is also taken into consideration.

2.1.2. Procedure for generating SSFSP

In a standard CNN-based pixel-wise classification framework (Fig. 2), the patch input is obtained by cropping the neighbouring pixels with a regular geometric shape from the image (referred to as the traditional feature patch (TFP)). In our approach, TFP is further transformed into a stack of gridded spectral feature images, which is referred to as the SSFS patch (SSFSP). The detailed procedures to generate SSFSP are given below.

Band reduction: Since hyperspectral imagery (HSI) may suffer from high collinearity and redundant information [39], band reduction is usually needed to obtain the optimal set of input bands. There are many methods to achieve this goal, e.g., principal component analysis (PCA). Accordingly, the complete satellite image $Z \in R^{i \times j \times c}$ is reduced to its subset image $X \in R^{i \times j \times c'}$, where

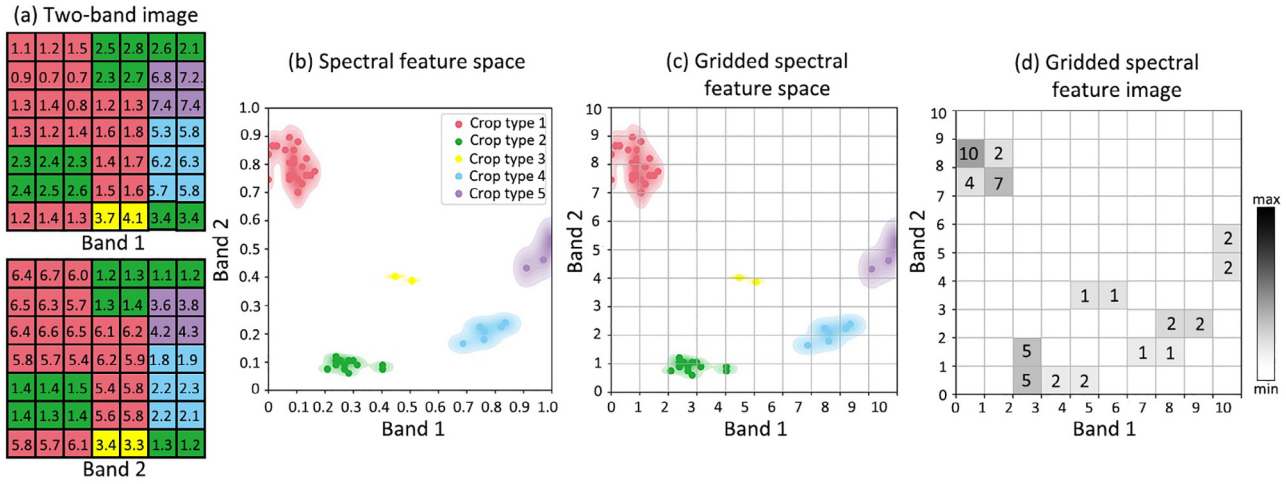

Fig. 1. Schematic diagram of converting a two-band image into a gridded spectral feature image.

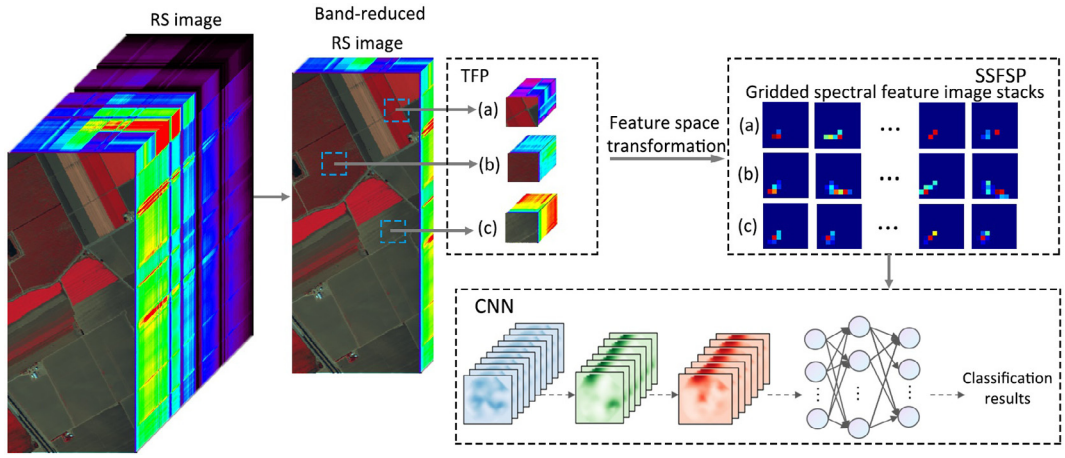

Fig. 2. Flowchart of generating Stacked Spectral Feature Space patch (SSFSP) from traditional feature patch (TFP) cropped from remotely sensed (RS) image for convolutional neural network (CNN)-based classification.

i and j represent the size of an image (height and width), c and c' are the band numbers of Z and X with c' smaller than c.

Generating TFP: TFP $\in X^{W\times W\times c'}$ is generated by cropping X with a user-defined parameter window size W. The class of its central pixel labels the class of a TFP, and the training and testing datasets are generated with TFPs with the labelled central pixels. To make spectral values comparable between bands, X is standardized by the global maximum max and the minimum min in this step.

$$X_{\text{stand}} = S(X) = \frac{X - \min}{\max - \min} \tag{3}$$

Generating SSFSP: SSFSP is generated from all the combinations of any two bands of TFP. The coordinate of the mth pixel of TFP in one gridded spectral feature image (one band of SSFSP) is:

$$\text{coord}^{\text{SSFSP}}(m) = \text{Round}\left(R\left(I_a^{\text{TF}}(m), I_b^{\text{TF}}(m)\right)\right), \tag{4}$$

where $I_a^{\text{TF}}(m)$ and $I_b^{\text{TF}}(m)$ are the standardized values of the mth pixel of TFP at bands a and b. R is a scaling factor that decides the coordinate ranges in each gridded spectral feature image of SSFSP, such as 10 in Fig. 1c. *Round ()* is the rounding function to convert decimal numbers into their nearest integers. Here, coupled with the R and the rounding function, SSFSP can significantly reduce intra-class variability. After that, the number of points that fall in the same grid is summarized and assigned to the value of that grid. All the gridded spectral feature images constitute SSFSP $\in X^{R\times R\times c''}$ with the band number of $c'' = \frac{c'\times(c'-1)}{2}$.

The rest of the proposed classification framework is the same as a typical CNN, namely that a proper CNN model is selected. SSFSP is used as the model input to extract high-level features for classification.

2.2. Adopted CNN models

Seven typical CNN models and a simple multilayer perceptron model (MLP) were used in our study to test the effectiveness of SSFSP, including Multi-OCNN [40] (Fig. 3), DeepNet [24], 3DCNN [27], and four pre-trained versions of the Resnet family (Resnet-18, -34, -50 and -101) [41]. The detailed topologies of the first three and the last models are listed in Table S1. Multi-OCNN is proposed for object-based classification [40], and we adopted its main architecture except that we used a patch input strategy to classify at the pixel level instead of the prior image segmentation. The present study was based on this benchmark model unless otherwise stated. The logarithmic form of Softmax (Log-Softmax) was used on all models as it showed a better performance than Softmax [42]. The same training settings were adopted for all models. The

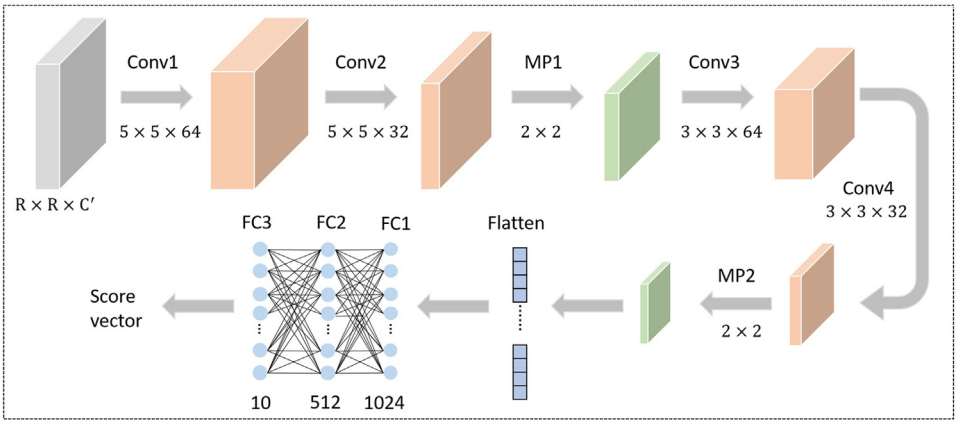

Fig. 3. The architecture of the benchmark model (Multi-OCNN). Conv, MP and FC mean the convolutional layer, max-pooling layer, and fully-connection layer, respectively, and the numbers indicate the size of kernels of Convs and MPs or the nodes number of FCs.

cross-entropy loss function and Adam optimizer [43] with a batch size of 64 and weight decay of 0.0008 (other parameters of Adam referred to [43]) were used to update the parameters of models. The training epoch for all the eight models was set to 100 epochs, and the learning rate was changed to half every 40 epochs.

2.3. Evaluation metrics

The classification accuracies in the following experiments were evaluated by three commonly used metrics, namely the overall accuracy (OA), the average accuracy (AA), and the Cohen's Kappa coefficient (Kappa) [44–46]. OA represents the proportion of the correctly classified samples to the total samples. It measures the model's ability to classify overall samples. However, it does not reflect the classification accuracy of each class. AA is the average of the classification accuracies (proportion of samples in a class that are correctly classified) of all classes and measures how good the model can distinguish between classes. Kappa measures the classification agreement beyond the chance agreement, which is not considered by OA [47], and is another commonly used metric in a multi-class classification task.

3. Data and experiment

3.1. Experiment data

We used two publicly available HSI datasets (Salinas (SA) and WHU-Hi dataset) to test the accuracy of crop classification of SSFSP as crops predominate in both datasets (Fig. 4). The SA [48] with 224 contiguous spectral bands across wavelengths from 0.4 to 2.5 μm was collected at a plot of farmland with 16 crop types in the Salinas Valley in California, USA. The image size of SA in pixels is 512 × 217 and has a high spatial resolution of 3.7 m. The WHU-Hi dataset was captured by a UAV with a Headwall Nano-Hyperspec imaging sensor with 270 bands ranging from 0.4 to 1.0 μm [24,49]. The Honghu and Longkou sub-datasets of WHU-Hi dataset were chosen. The Honghu dataset was collected in Honghu of Hubei province, China, consisting of 940 × 475 pixels and 22 land cover types, while the Longkou dataset was collected in Longkou Town of Hubei province, China, consisting of 550 × 400 pixels and 9 land cover types. The spatial resolution of the Honghu and Longkou datasets is 0.043 m and 0.463 m, respectively. In this study, the Longkou dataset was used as supplementary experimental data because of the length limit, and its experimental results are shown in Table S2 and Fig. S1. Only 5% of the labelled samples were randomly selected as the training set due to the strong prediction capacity of the CNNs.

3.2. Experimental design

The proposed SSFSP-based classification framework was compared to the TFP-based framework, and they are denoted as SSFSP-CNN and TFP-CNN, respectively. To make operation easier and ensure generality in the step of band reduction, unless otherwise stated, we selected only five bands with the same central wavelength as in the Landsat 8 OLI sensor (Table S3) as input, considering that these bands are available for most sensors. This simple band reduction strategy is referred to as ORI. Three experiments were carried out for comprehensive evaluation as described below.

3.3. Experiment I: The benchmark in various scenes

The first experiment aims to compare the two inputs (SSFSP and TFP) for the benchmark CNN classifier in different scenes with the evaluation metrics. To compare their computational efficiencies, the dynamics of the accuracy and loss in their CNN training processes were also investigated.

3.4. Experiment II: Robustness against model selection

The second experiment aims to evaluate the robustness of the SSFSP against the selection of CNN models. As numerous CNN models with various architectures have been proposed, it is challenging for users to choose the most suitable model for their missions. The SSFSP is thus expected to mitigate the effect of model selection on classification accuracy. Therefore, this experiment adopted all the aforementioned CNN models and MLP and compared the accuracy variations of the two inputs (SSFSP and TFP) given different classifiers.

3.5. Experiment III: Comparison to other handcrafted features with refinement

Since both pre-processing and post-processing are expected to improve classification accuracy, it is interesting to compare SSFSP to the combinations of other handcrafted features and pre/post-processing techniques to see whether the proposed SSFSP input still yields the best results. In this experiment, SSFSP was compared with the two typical handcrafted features mentioned in the introduction, MCM and Gabor [34,50]. The window sizes *W*

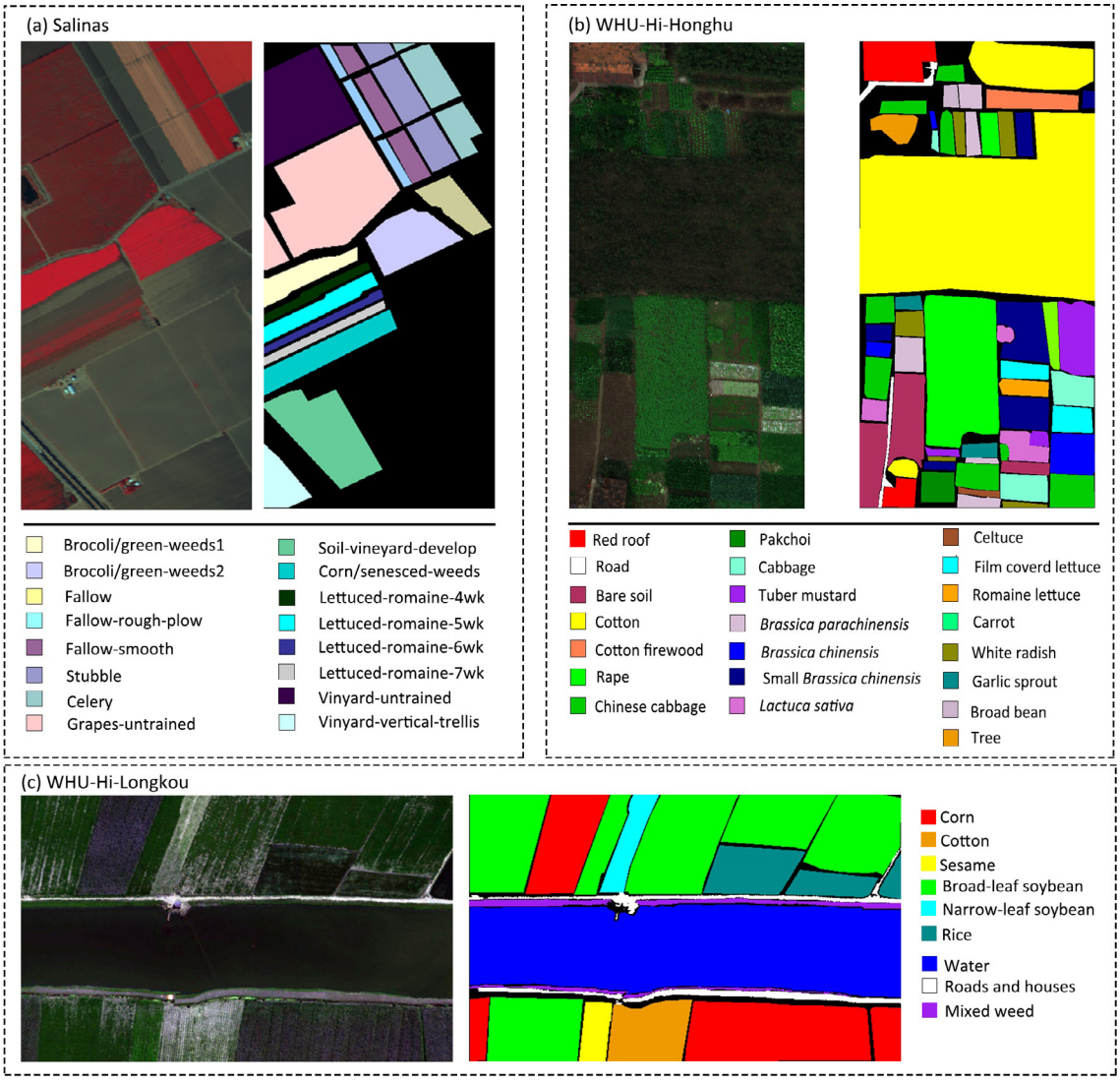

Fig. 4. Hyperspectral datasets. (a) Salinas. (b) WHU-Hi-Honghu. (c) WHU-Hi-Longkou. Each dataset is demonstrated with a false-color composite image and ground-truth map.

of all the features were set to 15. Gabor filtering was performed in four directions on each of the five selected bands (see Table S3), resulting in 20 input bands in total. For MCM, the input bands were also set to 20 as in Gabor, namely 20 scales for obtaining the neighbouring vectors. As for pre-processing, data augmentation was carried out using horizontal flip, vertical flip, and random angle rotation. As for post-processing, the typical conditional random field (CRF) was adopted, which refined object boundaries in classification maps. Their different combinations were compared using the Honghu dataset and the benchmark CNN model. To avoid the effect of sample imbalance and highlight the effects of pre-processing and post-processing, small and balanced samples (100 samples per crop type) were used as training samples, different from Experiments I and II.

3.6. Parameter settings

As described in Section 2, the SSFSP is transformed from the TFP, so the three parameters, the window size W, the scaling factor R, and the reduced band number c', need to be specified in advance. W should be determined based on the landscape complexity. Although a larger W incorporates more neighbour information, it can also introduce undesired noise that is more likely to happen in scenes with heterogeneous landscapes. Therefore, it was empirically set to 15 for both the SA and the Honghu datasets. As for R, it is less sensitive to the landscape and was empirically set to 25 for the two datasets, indicating that the size of the gridded feature image is 25×25. As for the reduced band number c', it was set to 5, meaning only five bands were used in experiments. The sensitivity for these three parameters was analyzed in Section 5.2.

4. Results

4.1. Results for experiment I

4.1.1. Results for the Honghu dataset

In the Honghu dataset, all crop types showed promising and better classification results in SSFSP-CNN than TFP-CNN (Fig. 5). For crop types with much fewer labelled samples (pakchoi, celtuce, carrot, and broad bean), SSFSP-CNN was more efficient to distinguish the land cover, and its accuracy was much higher than that of TFP-CNN (Table S4). In addition, the salt-and-pepper appeared

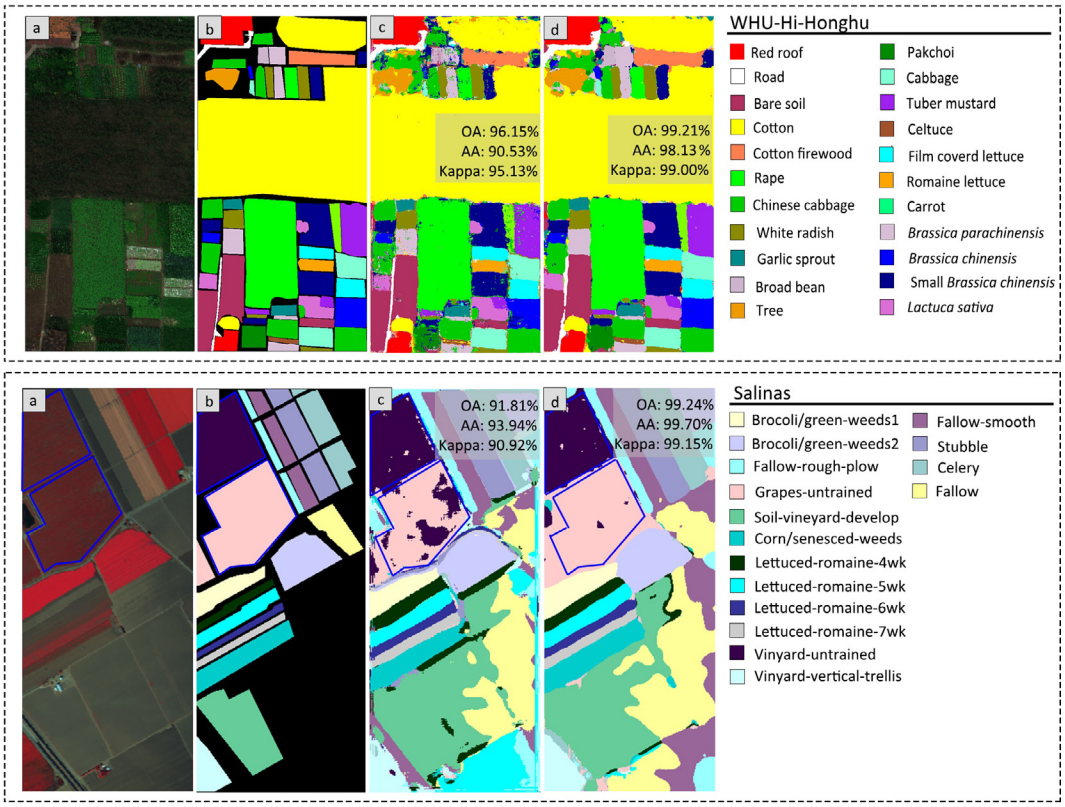

Fig. 5. Classification maps and accuracies for the SA and Honghu datasets. (a) False-color composite image. (b) Ground truth. (c) TFP-CNN. (d) SSFSP-CNN.

in the TFP-CNN result but was significantly reduced in the SSFSP-CNN result.

4.1.2. Results for the SA dataset

In the SA dataset, the OA, AA and Kappa of TFP-CNN are significantly lower than those of SSFSP-CNN (Fig. 5). The accuracies of nearly all crop types are above 99% by SSFSP-CNN, whereas several crop types have poor classification accuracy by TFP-CNN, such as Fallow (68.57%) and Grapes-untrained (76.23%) (Table S5). Noticeably, Grapes-untrained and Vinyard-untrained (highlighted by the blue polygons in Fig. 5) are largely misclassified by TFP-CNN. In contrast, despite their similar spectra, SSFSP-CNN still classifies them excellently (Fig. 5).

4.1.3. Dynamics of accuracy and loss

The dynamics of the classification accuracies and losses of SSFSP-CNN and TFP-CNN during model training is shown in Fig. 6. SSFSP-CNN reaches the highest accuracy after a few iterations in both datasets, and the curves are smooth in the subsequent iterations without a large magnitude of change. In contrast, the accuracy curves of TFP-CNN have more significant variation and reach the peaks only in the later stage of training. Similarly, the loss values of SSFSP-CNN decrease rapidly and become stable at the early stage, while TFP-CNN does not converge until the later stage. The results highlight the robustness of SSFSP and its benefit to the acceleration in model training.

4.1.4. Results for experiment II

The accuracy variations using SSFSP and TFP as input for various CNN models and MLP are demonstrated in Fig. 7 (see Table S6 for detailed statistics and Figs. S2 and S3 for mapping results). Under the seven CNN models, the OA, AA, and Kappa values for SSFSP do not significantly differ from those under the benchmark model, and satisfactory accuracy is achieved even with the MLP model (corresponding to the outliers in the box plots). In contrast, TFP shows significant variations in OA, AA and Kappa values under the seven CNN models. Such variations indicate a strong dependence of TFP on the choice of CNN models, which usually requires extensive experience in practical applications. In short, SSFSP relies less on the selection of CNN models, which significantly saves the time required in selecting and tuning the models.

4.1.5. Results for experiment III

The results are shown in Table 1. According to original works [34,35], which firstly used MCM and Gabor as input of CNN, PCA was used to generate MCM, and ORI was used as band reduction method for SSFSP, TFP and Gabor in the study. In addition, 100 samples per crop type were used as training samples to avoid the effect of sample imbalance, which is different from Experiments I and II. All the three features provide further accuracy improvements compared to TFP, with SSFSP the most pronounced improvements and Gabor slightly better than MCM. When using the data augmentation strategy, the accuracies for the Honghu dataset with Gabor, MCM and TFP all improved, while there is no significant effect (or even a slight decrease) for SSFSP. This is because these augmented and original samples obtain completely similar SSFSP, resulting in minor accuracy improvement before and after data augmentation. However, it can still be observed that SSFSP yields the highest accuracy. After being post-processed by CRF, they all further improved, with SSFSP reaching an impressive 99.51% OA and ranking first. Therefore, there is no need to perform data augmentation on SSFSP, and we recommend post-processing for SSFSP by using CRF if possible.

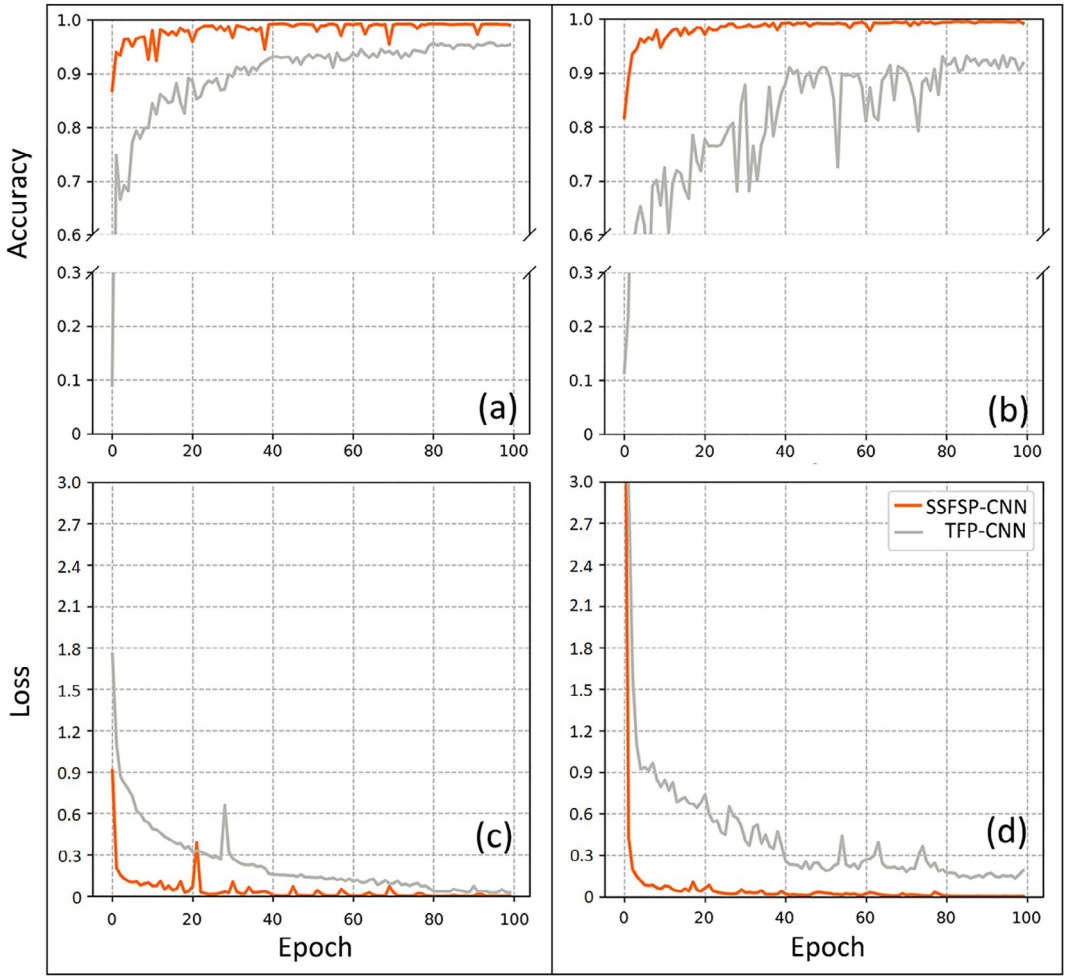

Fig. 6. Accuracy and loss curves of SSFSP-CNN and TFP-CNN during training. (a) Accuracy curves of SSFSP-CNN and TFP-CNN for Honghu. (b) Accuracy curves of SSFSP-CNN and TFP-CNN for SA. (c) Loss curves of SSFSP-CNN and TFP-CNN for Honghu. (d) Loss curves of SSFSP-CNN and TFP-CNN for SA.

5. Discussion

5.1. The superiority of SSFSP

The experiments demonstrated that the proposed feature SSFSP yielded better results than TFP in accuracy, robustness and training efficiency. Its outperformance can be attributed to two aspects.

SSFSP is a more advanced representation of spectral information of crop types in comparison to TFP. To generate SSFSP, the spectral information is transformed from the 1D spectral vector into the 2D gridded spectral feature image. During the transformation process, not only the spectral values of a pixel in any two bands are retained within the coordinates of the gridded feature image, but also the interrelationship between any two bands is expressed through the 2D gridded spectral feature image. Nevertheless, such information is not available in common CNN inputs. Accordingly, SSFSP makes full use of the spectral information by generating it and using the powerful spatial feature mining capability of CNN. Furthermore, SSFSP includes spatial neighbour information uniquely. Given a set of pixel observations, the image patches with many possible spatial arrangements are transformed into an identical SSFSP regardless of their spatial distribution and the geometric transformations (e.g., rotation and flip) (Fig. 8a). Thus, one SSFSP naturally can correspond to several TFPs with different possible spatial arrangements if they share similar spectral characteristics. This advantage allows a classifier to focus on spectral differences, thus reducing the need for training samples with different spatial arrangements. In contrast, traditional CNN inputs require rotation and flipping operations to add samples of different spatial arrangements. As a result, for the crop types with limited samples (e.g., Pakchoi, Celtuce, Carrot, and Broad bean in the Honghu dataset), their TFP-based classification accuracies are largely lower than the others with enough samples, which is commonly known as the issue of sample imbalance. By contrast, SSFSP-based classification accuracies are significantly better because the problem is alleviated by SSFSP (Tables S4, S5). Although data augmentation should help TFP against sample imbalance, the training cost also increases and the improvement can be limited (Table 1).

Second, the reduction of intra-class variance by SSFSP also benefits the classification performance. A preferred classification feature should generally increase inter-class variance and reduce the intra-class variance because larger intra-class variance can bring uncertainty to the classification. In SSFSP, the intra-class variance is effectively reduced by adaptively smoothing the spectral details with the scaling factor *R*. As demonstrated in Fig. 8b, with a moderate *R*, the intra-class variance is smoothed out, which contributes to the final classification accuracy. Thus, compared with TFP, the intra-class variance is retained in a finite number of points on the class spectral curve resulting from flattening the SSFSP to a one-dimensional vector, and other points on the class spectral curve are shown to have zero values, which can be regarded as feature sparsity (Fig. 8c). The spectral sparsity of SSFSP

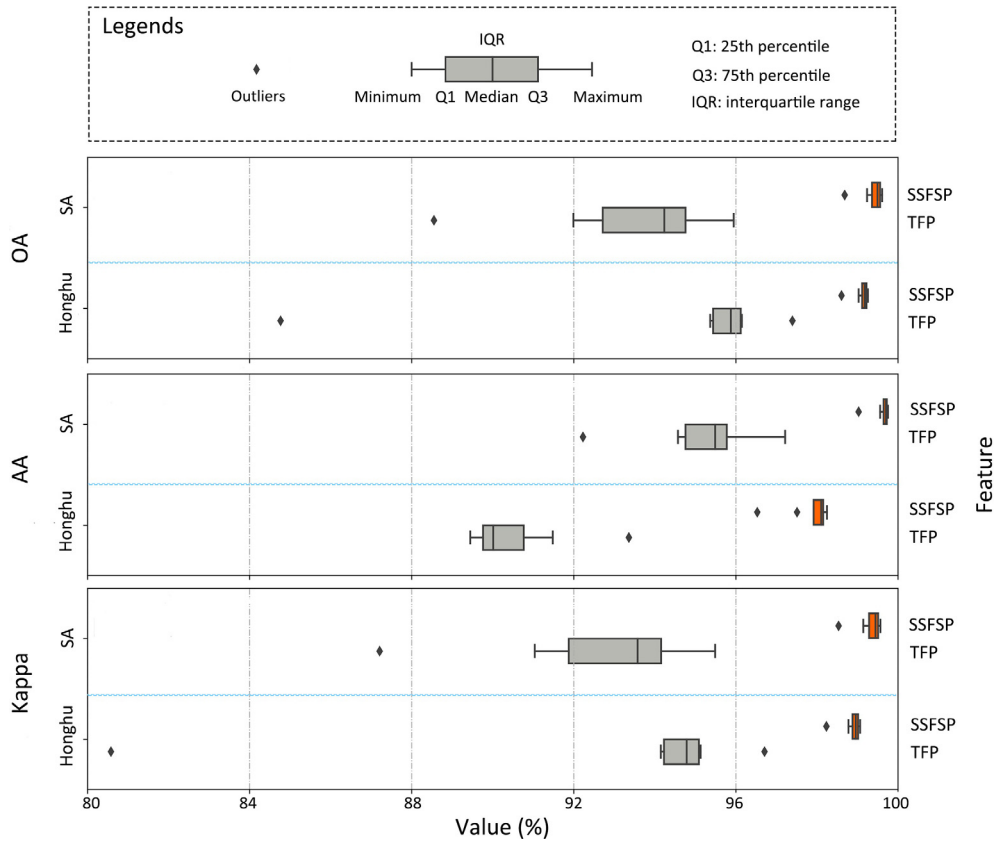

Fig. 7. Accuracy variations using SSFSP and TFP with different CNN models and MLP. AA, average accuracy; OA, overall accuracy; Kappa, Cohen's Kappa coefficient.

Table 1
Comparison of classification accuracies with data pre- and post-processing.

Processing	Accuracy metrics (%)	Feature			
		TFP	MCM	Gabor	SSFSP
Non Aug, non CRF	AA	88.91	87.73	88.07	96.01
	OA	87.14	87.74	88.18	95.38
	Kappa	84.09	84.72	85.30	94.19
Aug only	AA	90.22	88.15	88.19	96.15
	OA	89.39	88.51	90.55	95.19
	Kappa	86.76	85.45	88.19	93.96
Both Aug and CRF	AA	95.56	93.64	93.63	98.86
	OA	93.94	92.91	94.04	99.51
	Kappa	92.41	91.13	92.01	99.38

'Aug' and 'CRF' represent 'data augmentation' and 'conditional random field' respectively.
AA, average accuracy; OA, overall accuracy; Kappa, Cohen's Kappa coefficient.

benefits the CNN model in terms of noise exclusion and finding key features [51–53]. Hence, the need for complex networks is eliminated (Fig. 7), and higher training accuracy can be obtained in the early training stage (Fig. 6).

5.2. Suggestion for SSFSP users

Three parameters are available for users to determine when using SSFSP, namely, c', R, and W. We conducted the sensitivity analysis for the three parameters, of which results can be found in Fig. S4, Fig. S5 and Table S7. Accordingly, we recommend using ORI as the band reduction method for SSFS and setting c' to 4, considering the trade-off between accuracy and computational efficiency. When the spectral differences of crop types are prominent, R can be set to a small value, such as 15. When the spectra of crop types are very similar, a larger R (e.g., 20 or 25) is needed to increase the spectral differences for classification.

5.3. Limitation and future work

The major limitation of SSFSP is currently low computational efficiency when using a larger number of spectral bands, especially for pixel-wise classification. We consider developing the current SSFSP into a version suitable for fully end-to-end crop classification to improve computational efficiency [54]. Recently, the Spectral AttentionModule has received much attention in the deep learning community [54]. The proposed SSFSP mines the interrelationship between any two original spectral bands, while SAM mines the interdependencies between feature maps obtained from former convolutional layers. Therefore, SSFSP and SAM are in different

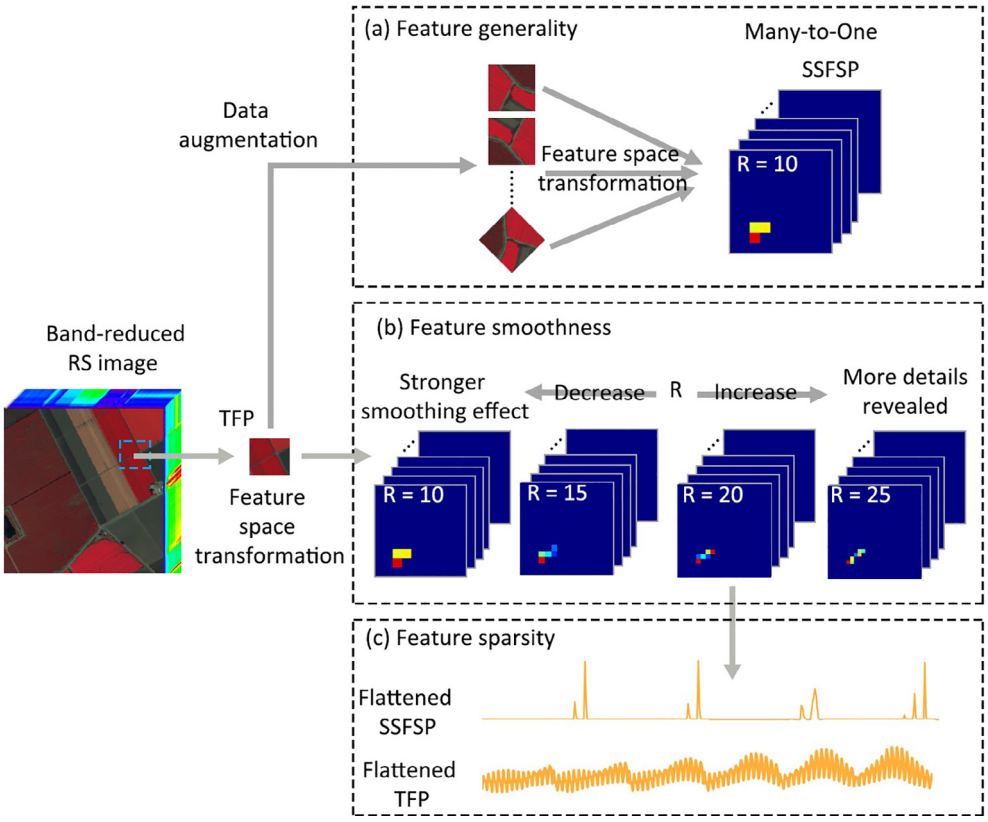

Fig. 8. Beneficial properties of SSFSP. (a) Advanced representation of spectral information. (b) Reduction of intra-class variance. The SSFSPs in (b) are generated from a TFP with *R* ranging from 10 to 25. The spectral-like curves in (c) are the average of flattened TFP and SSFSP for one class in the AS dataset.

stages in the classification framework, but they are not mutually exclusive. As they enhance the spectral features at different stages, we plan to combine SSFS and SAM in the future to obtain better classification results.

6. Conclusions

A promising spectral feature (SSFSP) for CNN-based crop classification is proposed in this paper. SSFSP is a set of two-dimensional (2D) gridded spectral feature images that record the spatial and intensity distribution characteristics of various crop types in a two-dimensional feature space consisting of any two spectral bands. SSFSP can be combined with 2D-CNN to support simultaneous mining of spectral and spatial features, as the spectral features are successfully converted to 2D images that can be processed by CNN. The comparative study using high spatial resolution hyperspectral datasets at three sites showed that SSFSP outperforms conventional spectral features in terms of accuracy, robustness, and training efficiency. Its excellent performance can be attributed to three aspects. First, SSFSP mines the spectral interrelationship with feature generality, which reduces the required number of training samples. Second, the intra-class variance can be largely reduced in the form of grid partitioning. Third, SSFSP is a highly sparse feature, which reduces the dependence on CNN model structure and enables early and fast convergence of model training. All three unique characteristics of SSFSP are connected by converting 1D spectral information into 2D information for convolution operation. As a corresponding guideline, we recommend users adopt SSFSP in precision agriculture applications, especially when spectral features are more important than spatial features in distinguishing crop types while paying attention to selecting several key parameters of SSFSP.

CRediT authorship contribution statement

Hui Chen: Data Curation, Formal analysis, Methodology, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing. **Yue'an Qiu:** Investigation, Writing – Review & Editing. Dameng Yin: Investigation, Writing – Review & Editing. **Jin Chen:** Conceptualization, Methodology, Funding acquisition, Supervision, Project administration, Writing – Review & Editing. **Xuehong Chen:** Methodology, Investigation. **Shuaijun Liu:** Investigation. **Licong Liu:** Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (67441830108 and 41871224). We thank Miss Yi Nam Xu for improving the quality of our manuscript.

Appendix A. Supplementary data

Supplementary data for this article can be found online at https://doi.org/10.1016/j.cj.2021.12.011.

References

[1] C. Zhang, H. Zhang, L. Zhang, Spatial domain bridge transfer: an automated paddy rice mapping method with no training data required and decreased

image inputs for the large cloudy area, Comput. Electron. Agric. 181 (2021) 105978.
[2] G.M. Foody, Status of land cover classification accuracy assessment, Remote Sens. Environ. 80 (2002) 185–201.
[3] K.C. Seto, M. Fragkias, B. Güneralp, M.K. Reilly, J.A. Añel, A meta-analysis of global urban land expansion, PLoS ONE 6 (2011) e23777.
[4] J. Chen, J. Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, G. Han, S. Peng, M. Lu, W. Zhang, X. Tong, J. Mills, Global land cover mapping at 30 m resolution: a POK-based operational approach, ISPRS J. Photogramm. Remote Sens. 103 (2015) 7–27.
[5] X. Xiao, S. Boles, J. Liu, D. Zhuang, S. Frolking, C. Li, W. Salas, B. Moore, Mapping paddy rice agriculture in southern China using multi-temporal MODIS images, Remote Sens. Environ. 95 (2005) 480–492.
[6] G.H. Kwak, N.W. Park, Impact of texture information on crop classification with machine learning and UAV images, Appl. Sci. 9 (2019) 643.
[7] J. Bossu, C.H. Gée, G. Jones, F. Truchetet, Wavelet transform to discriminate between crop and weed in perspective agronomic images, Comput. Electron. Agric. 65 (2009) 133–143.
[8] J. Senthilnath, M. Kandukuri, A. Dokania, K.N. Ramesh, Application of UAV imaging platform for vegetation analysis based on spectral-spatial methods, Comput. Electron. Agric. 140 (2017) 8–24.
[9] J. Torres-Sánchez, J.M. Peña, A.I. de Castro, F. López-Granados, Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV, Comput. Electron. Agric. 103 (2014) 104–113.
[10] A. Pourreza, H. Pourreza, M.H. Abbaspour-Fard, H. Sadrnia, Identification of nine Iranian wheat seed varieties by textural analysis with image processing, Comput. Electron. Agric. 83 (2012) 102–108.
[11] Y. Zhang, L. Wu, Crop classification by forward neural network with adaptive chaotic particle swarm optimization, Sensors 11 (2011) 4721–4743.
[12] F. Garcia-Ruiz, S. Sankaran, J.M. Maja, W.S. Lee, J. Rasmussen, R. Ehsani, Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees, Comput. Electron. Agric. 91 (2013) 106–115.
[13] X. Jin, S. Liu, F. Baret, M. Hemerlé, A. Comar, Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery, Remote Sens. Environ. 198 (2017) 105–114.
[14] Ö. Akar, O. Güngör, Integrating multiple texture methods and NDVI to the Random Forest classification algorithm to detect tea and hazelnut plantation areas in northeast Turkey, Int. J. Remote Sens. 36 (2015) 442–464.
[15] L. Wei, K. Wang, Q. Lu, Y. Liang, H. Li, Z. Wang, R. Wang, L. Cao, Crops fine classification in airborne hyperspectral imagery based on multi-feature fusion and deep learning, Remote Sens. 13 (2021) 2917.
[16] L. Zhao, Y. Shi, B. Liu, C. Hovis, Y. Duan, Z. Shi, Finer classification of crops by fusing UAV Images and Sentinel-2A Data, Remote Sens. 11 (2019) 3012.
[17] N. Kussul, M. Lavreniuk, S. Skakun, A. Shelestov, Deep learning classification of land cover and crop types using remote sensing data, IEEE Geosci. Remote Sens. Lett. 14 (2017) 778–782.
[18] B. Xie, H.K. Zhang, J. Xue, Deep Convolutional Neural network for mapping smallholder agriculture using high spatial resolution satellite image, Sensors 19 (2019) 2398.
[19] Y. LeCun, K. Kavukcuoglu, C. Farabet, Convolutional networks and applications in vision, in: Proceedings of 2010 IEEE International Symposium on Circuits and Systems, IEEE, Paris, France, 2010, pp. 253–256.
[20] P. Sidike, V.K. Asari, V. Sagan, Progressively Expanded Neural Network (PEN Net) for hyperspectral image classification: a new neural network paradigm for remote sensing image analysis, ISPRS J. Photogramm. Remote Sens. 146 (2018) 161–181.
[21] Q. Yuan, H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, J. Wang, J. Gao, L. Zhang, Deep learning in environmental remote sensing: achievements and challenges, Remote Sens. Environ. 241 (2020) 111716.
[22] L. Zhong, L. Hu, H. Zhou, Deep learning based multi-temporal crop classification, Remote Sens. Environ. 221 (2019) 430–443.
[23] W. Li, G. Wu, F. Zhang, Q. Du, Hyperspectral image classification using deep pixel-pair features, IEEE Trans. Geosci. Remote Sens. 55 (2017) 844–853.
[24] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, L. Zhang, WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF, Remote Sens. Environ. 250 (2020) 112012.
[25] G.H. Kwak, C. Park, K. Lee, S. Na, H. Ahn, N.W. Park, Potential of hybrid CNN-RF model for early crop mapping with limited input data, Remote Sens. 13 (2021) 1629.
[26] S. Ji, C. Zhang, A. Xu, Y. Shi, Y. Duan, 3D convolutional neural networks for crop classification with multi-temporal remote sensing images, Remote Sens. 10 (2018) 75.
[27] B. Zhang, L. Zhao, X. Zhang, Remote sensing of environment three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images, Remote Sens. Environ. 247 (2020) 111938.
[28] H. Li, C.E. Zhang, Y. Zhang, S. Zhang, X. Ding, P.M. Atkinson, A Scale Sequence Object-based Convolutional Neural Network (SS-OCNN) for crop classification from fine spatial resolution remotely sensed imagery, Int. J. Digit. Earth 14 (2021) 1528–1546.
[29] J. Xu, Y. Zhu, R. Zhong, Z. Lin, J. Xu, H. Jiang, J. Huang, H. Li, T. Lin, DeepCropMapping: a multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping, Remote Sens. Environ. 247 (2020) 111946.
[30] H. Li, C. Zhang, S. Zhang, X. Ding, P.M. Atkinson, Iterative Deep Learning (IDL) for agricultural landscape classification using fine spatial resolution remotely sensed imagery, Int. J. Appl. Earth Obs. Geoinf. 102 (2021) 102437.
[31] W. Huang, Y. Huang, Z. Wu, J. Yin, Q. Chen, A multi-kernel mode using a local binary pattern and random patch convolution for hyperspectral image classification, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14 (2021) 4607–4620.
[32] T. Zhang, P. Zhang, W. Zhong, Z. Yang, F. Yang, JL-GFDN: a novel gabor filter-based deep network using joint spectral-spatial local binary pattern for hyperspectral image classification, Remote Sens. 12 (2020) 2016.
[33] M.M. Santoni, D.I. Sensuse, A.M. Arymurthy, M.I. Fanany, Cattle race classification using gray level co-occurrence matrix convolutional neural networks, Procedia Comput. Sci. 59 (2015) 493–502.
[34] N. He, M.E. Paoletti, J.M. Haut, L. Fang, S. Li, A. Plaza, J. Plaza, Feature extraction with multiscale covariance maps for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 57 (2019) 755–769.
[35] P. Ghamisi, E. Maggiori, S. Li, R. Souza, Y. Tarabalka, G. Moser, A. de Giorgi, L. Fang, Y. Chen, M. Chi, S.B. Serpico, J.A. Benediktsson, Y. Tarablaka, G. Moser, A. De Giorgi, L. Fang, Y. Chen, M. Chi, S.B. Serpico, J.A. Benediktsson, New frontiers in spectral-spatial hyperspectral image classification: the latest advances based on mathematical morphology, markov random fields, segmentation, sparse representation, and deep learning, IEEE Geosci. Remote Sens. Mag. 6 (2018) 10–43.
[36] Y. Chen, L. Zhu, P. Ghamisi, X. Jia, G. Li, L. Tang, Hyperspectral images classification with gabor filtering and convolutional neural network, IEEE Geosci. Remote Sens. Lett. 14 (2017) 2355–2359.
[37] M. Ghassemi, H. Ghassemian, M. Imani, Hyperspectral image classification by optimizing convolutional neural networks based on information theory and 3D-Gabor filters, Int. J. Remote Sens. 42 (2021) 4380–4410.
[38] C. Bishop, Pattern Recognition and Machine Learning, Springer, New York, NY, USA, 2007.
[39] W. Sun, Q. Du, Hyperspectral band selection: a review, IEEE Geosci. Remote Sens. Mag. 7 (2019) 118–139.
[40] V.S. Martins, A.L. Kaleita, B.K. Gelder, H.L.F. da Silveira, C.A. Abe, Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution, ISPRS J. Photogramm. Remote Sens. 168 (2020) 56–73.
[41] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: L. O'Conner (Ed.), 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, 2016, pp. 770–778.
[42] A. de Brébisson, P. Vincent, An exploration of softmax alternatives belonging to the spherical loss family, arXiv (2016) arXiv:1511.05042.
[43] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv (2014) arXiv:1412.6980.
[44] S.V. Stehman, Selecting and interpreting measures of thematic classification accuracy, Remote Sens. Environ. 62 (1997) 77–89.
[45] G.H. Rosenfield, K. Fitzpatrick-Lins, A coefficient of agreement as a measure of thematic classification accuracy, Photogramm. Eng. Remote Sens. 52 (1986) 223–227.
[46] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46.
[47] G. Foody, Assessing the accuracy of remotely sensed data: principles and practices, Photogramm. Rec. 25 (2010) 204–205.
[48] B.A. M Graña, MA Veganzons, Hyperspectral remote sensing scenes, Grupo de Inteligencia Computacional (GIC), http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes (Accessed on June 22, 2021).
[49] Y. Zhong, X. Wang, Y. Xu, S. Wang, T. Jia, X. Hu, J.i. Zhao, L. Wei, L. Zhang, Mini-UAV-Borne hyperspectral remote sensing: from observation and processing to applications, IEEE Geosci. Remote Sens. Mag. 6 (2018) 46–62.
[50] M.H. Shi, G. Healey, Hyperspectral texture recognition using a multiscale opponent representation, IEEE Trans. Geosci. Remote Sens. 41 (2003) 1090–1095.
[51] M.A. Ranzato, Y.L. Boureau, Y. Lecun, Sparse feature learning for deep belief networks, in: NIPS'07: Proceedings of the 20th International Conference on Neural Information Processing Systems, MIT Press, Cambridge, MA, USA, pp. 1185–1192.
[52] E. Doi, D.C. Balcan, M.S. Lewicki, Robust coding over noisy overcomplete channels, IEEE Trans. Image Process. 16 (2007) 442–452.
[53] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1?, Vision Res 37 (1997) 3311–3325.
[54] Z. Zheng, Y. Zhong, A. Ma, L. Zhang, FPGA: fast patch-free global learning framework for fully end-to-end hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 58 (2020) 5612–5626.