A Scale Sequence Object-based Convolutional Neural Network (SS-OCNN) for crop classification from fine spatial resolution remotely sensed imagery

Huapeng Li [a] [*], Ce Zhang [b], Yong Zhang [c], Shuqing Zhang [a], Xiaohui Ding [d], Peter M. Atkinson [b]

[a] *Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China*

[b] *Lancaster Environment Centre, Lancaster University, Lancaster, UK*

[c] *School of Electrical and Information Engineering, Changchun Guanghua University, Changchun, China*

[d] *Guangzhou Institute of Geography, Guangzhou, China*

Abstract:

Accurate crop distribution mapping lays an important scientific basis for crop yield prediction, field management, and balancing ecosystem services. Benefiting from the rapid progress of remote sensing technology, fine spatial resolution (FSR) remotely sensed imagery now offers great opportunities to map crop classes in detail. However, the highly dynamic nature of agro-ecosystems in space and time makes crop classification using FSR imagery an extremely difficult task. In this research, a novel

[*] Corresponding author.
E-mail addresses: lihuapeng@neigae.ac.cn (H. Li).

Scale Sequence Object-based Convolutional Neural Network (SS-OCNN) was developed for crop classification from FSR remotely sensed imagery. Different from the standard pixel-wise CNN, the SS-OCNN classifies images at the object level by taking segmented objects (crop parcels) as basic units of analysis, thus, ensuring that the boundaries between crop parcels are delineated precisely. These segmented objects were subsequently classified using a CNN model integrated with an automatically generated scale sequence of input patch sizes (i.e. a range of input windows for the CNN). This scale sequence can fuse effectively the features learned at different scales by transforming progressively the information extracted at small scales to larger scales, thus, increasing classification accuracy. The effectiveness of the SS-OCNN was investigated using two heterogeneous agricultural areas with FSR SAR and optical imagery, respectively. Experimental results revealed that the SS-OCNN consistently achieved the most accurate classification results over the two sites, increasing the overall accuracy by around 9% and 3% in comparison with the pixel-wise CNN and single-scale OCNN, respectively. By examining the class-wise accuracies, we found that the increase in overall accuracy was contributed mainly by small biomass crop classes (such as Hay and Winter wheat) which have a weak remote-sensing signature (Li et al., 2019a), relatively high heterogeneity and low signal-noise ratio. The SS-OCNN, thus, provides a new paradigm for crop classification over heterogeneous areas using FSR imagery, and has great potential and a wide application prospect.

Key words: Convolutional neural network, multi-scale deep learning, OBIA, crop classification, fine spatial resolution imagery;

1. Introduction

The world's population is predicted to increase to almost 10 billion by 2050 (FAO, 2018) which, together with economic development, is greatly increasing the demand for food. Detailed crop distribution data are of great importance to assess and forecast agricultural yield and, thus, further ensure food security both at local and global scales (Bastiaanssen et al., 2003; Debats et al., 2016). Moreover, such information is very useful for precision agriculture and crop management. For example, accurate identification of crop types is crucial for estimation of agricultural water use, and to avoid excessive use of water resources (Zheng et al., 2015). In addition, detailed information on crop distribution can be valuable for balancing global ecosystem services, since over-extensification of agriculture can threaten or damage the functions of other landscapes across the world, such as forests, wetlands and river environments (Nobre et al., 2016; Yan and Zhang, 2019).

Remote sensing has become a leading tool for crop monitoring and classification as it provides timely and repeated observations over relatively large areas (Wardlow et al., 2008; Biradar and Xiao, 2011; Dong et al., 2015; Li et al., 2019a; Liu et al., 2020). With the rapid progress of remote sensing technology, a very large number of fine spatial resolution (FSR) images are now commercially available, offering great opportunities for detailed crop mapping and classification (Sidikea et al., 2019; Li et al., 2019a).

Small farmland parcels mixed in other landscapes can be identified using FSR images with rich spatial information, such as RapidEye, Quickbird, Gaofen (GF), and Worldview (Persello et al., 2019). However, the highly dynamic nature of agro-ecosystems in space and time usually leads to high intra-class variance and low inter-class separability in the FSR imagery (Belgiua and Csillik, 2018; Li et al., 2019b; Hu et al., 2019; Dey et al., 2020). This is further complicated by diversified farming practices (Azar et al., 2016; Li et al., 2019a). Such complexity of spatial and temporal patterns over agricultural fields makes crop mapping from FSR imagery an extremely challenging task (Salehi et al., 2017; Li et al., 2020; Zhao et al., 2020). Advanced and robust methods are, therefore, needed to mine effectively and extensively the rich information hidden in FSR imagery to achieve highly accurate crop mapping and classification.

Over the past four decades, a great number of classification methods have been proposed for agricultural land cover classification. These methods can be broadly categorised into parametric (e.g. maximum likelihood classifier) and non-parametric approaches (e.g. decision trees). Non-parametric methods are generally superior to parametric ones due to their greater dependence on the data (Lu and Weng, 2007). Amongst non-parametric methods, machine learning (ML) methods (e.g. support vector machines, SVM) have become increasingly popular owing to their capacity to solve complex, non-linear problems (Duro et al., 2012; Cai et al., 2018). In these ML methods, a range of hand-coded features need to be generated and employed as prediction variables, such as texture, spectral features and vegetation indices (Wardlow and Egbert, 2008; Wang et al., 2015; Essa et al., 2017). However, these features are low and/or mid-level descriptors, which are insufficient to represent the rich and multi-level information in the FSR imagery. Besides, the extracted features are essentially hand-crafted, and they rely heavily on user experience and expertise.

Recently, deep learning (DL) has received enormous interest, whereby representative features can be learnt automatically in an end-to-end manner (LeCun et al., 2015). Amongst the DL architectures, deep convolutional neural networks (CNNs) have attracted great attention in both the academic and industrial communities in view of its great capacity to solve a variety of computer vision tasks, such as image processing (Krizhevsky et al., 2012), pattern recognition (He et al., 2016) and object detection (Girshick et al., 2016). Thanks to its superiority in high-level feature representation, CNNs have achieved impressive and promising results in the remote sensing field (Sylvain et al., 2019; Jiang et al., 2020), such as for change detection (Wang et al., 2019), scene classification (Zheng et al., 2019) and image segmentation (Chai et al., 2019). The CNN has also seen application in a wide range of remotely sensed image classification tasks, in which all pixels within a scene of image are classified into several categories. For example, Chen et al. (2016) presented a regularized deep feature extraction approach that uses both spectral and spatial features in hyperspectral images in classification. Zhang et al. (2018a) proposed a hybrid land cover classification method, whereby the patch-based CNN and the pixel-wise MLP (Multilayer Perceptron) were combined using a decision fusion strategy. Sidike et al. (2019) provided an expanded CNN for land cover classification over heterogeneous areas. These efforts demonstrated that CNNs consistently outperformed other benchmark ML algorithms (e.g. SVM), thanks to its hierarchical feature extraction capability. Different from traditional ML algorithms which can only extract spectral features, a standard CNN network learns features via a fix-sized patch using several filters. As such, both spectral and spatial features of the raw imagery can be learned at multiple levels, which is beneficial for image classification. The patch size of the CNN has a considerable impact on the scale of representations, and further on the accuracy of image classification (Chen et al., 2019). Previous studies have demonstrated that patch size (scale) is one of the most important parameters of a CNN classification model (Zhang et al., 2018b; Lv et al., 2018; Chen et al., 2019). Generally, a particular sized ground object needs to be characterised by an appropriate observational scale. However, the sizes of ground objects in a landscape often vary greatly due to the high complexity and diversity of natural and anthropogenic influences on the Earth's surface, especially over cropland areas. As a result, it is very difficult to confirm a certain scale that is suitable for all ground objects. Some studies adopted multiple scales in a CNN model to increase the accuracy of land cover classifications (Zhang et al., 2018b; Lv et al., 2018). A major drawback of such multi-scale CNN approaches is that the combination of multiple scales is very difficult to determine, and these scale values may vary greatly across regions. Consequently, multi-scale methods can lack generalizability, and hard to extend to other study areas.

To solve this scale issue, by mimicking the hierarchical processing mode of human cognition, we very recently proposed a new scale sequence joint deep learning (SS-JDL) method, in which the value of each scale as well as the number of total scales can be determined automatically (Zhang et al., 2020). Classification results confirmed the superiority of the SS-JDL over single- or multiple-scale CNN approaches for land cover/use classification. However, the SS-JDL was developed for land cover/use classification, where a scale sequence is incorporated into a pixel-wise CNN network. Such per-pixel CNN classification often blurs the boundaries between adjacent ground objects because patches overlap (Pan et al., 2019), such that some objects are over-expanded or shrunk. The SS-JDL is, thus, not suitable for image classification over agricultural fields, where the boundaries between field parcels need to be identified accurately (Li et al., 2019b). Furthermore, the SS-JDL was applied for hierarchical classification at different levels of semantic meaning (e.g., land cover and land use), which is different to a specific mapping task (e.g. crop classification).

The aim of this paper was threefold: (1) to develop a novel Scale Sequence Object-based CNN (SS-OCNN) for land cover classification from FSR imagery, (2) to validate the effectiveness of the proposed SS-OCNN in mapping crop categories across heterogeneous agricultural landscapes, and (3) to investigate the generalizability of the SS-OCNN using both optical and radar FSR remotely sensed imagery. In the SS-OCNN, an object-based CNN model with an automatically-generated scale sequence was designed to classify ground objects at the object level. By doing so, different-sized objects can be identified accurately, with the boundary information delineated precisely. To the best of our knowledge, this is the first effort to classify crop categories at the object level using an autonomous multi-scale CNN model. The proposed method was tested on two heterogeneous agricultural fields with different crop compositions using FSR Synthetic Aperture Radar (SAR) and optical imagery, respectively.

2. Method

2.1 Convolutional Neural Network (CNN)

The CNN, a biologically inspired multi-layer model, was developed originally to process data with multiple arrays. It is, therefore, arguably suitable for handling remotely sensed imagery in which pixels are arranged regularly. The CNN is essentially a forward neural network involving a cascade of multiple convolutional, pooling and fully connected layers (Zhang et al., 2018a). Specifically, a convolutional layer convolves across the entire image to capture multi-level feature representations using convolutional filters. An activation function, for example, the Rectified Linear Unit (ReLU), is employed outside the convolutional layer to gain nonlinear representations of the input data (Hinton et al., 2012). A max-pooling layer follows to strengthen the generalisation capacity of the CNN by reducing the resolution of the input data. A fully connected layer is subsequently added on top of the last max-pooling layer. Finally, the weights of the CNN are optimised using a stochastic gradient descent algorithm.

2.2 Object-based Convolutional Neural Networks (OCNN)

The object-based CNN is trained with labelled patches like the standard pixel-wise CNN. However, unlike the pixel-wise CNN which labels image patches which are densely overlapping at the pixel level, the OCNN classifies the image by predicting the class of each object obtained from image segmentation. The convolutional window of the OCNN is located at the centroid of each segmented object during the process of model inference (Zhang et al., 2018b). This strategy can not only maintain the boundary of each object, but also significantly increase computational efficiency during the model inference process. Once the class of each segmented object is labelled by the OCNN, the final thematic classified map is produced. An optimal input window size (scale) needs to be determined for the OCNN through trial and error.

2.3 Scale Sequence Object-based Convolutional Neural Networks (SS-OCNN)

The proposed Scale Sequence Object-based Convolutional Neural Network (SS-OCNN) method classifies remotely sensed imagery at the object level using a CNN combined with a scale sequence process as follows. The scale sequence consists of a set of scales (i.e. convolutional window or patch sizes) which transforms information learned from smaller scales to larger scales, so that detailed information about each object is convoluted into and represented across a broader context sequentially by adopting increasingly larger convolutional windows. The general procedure of the presented SS-OCNN methodology is shown in Fig. 1, in which the classification results are improved gradually over the scale sequence.

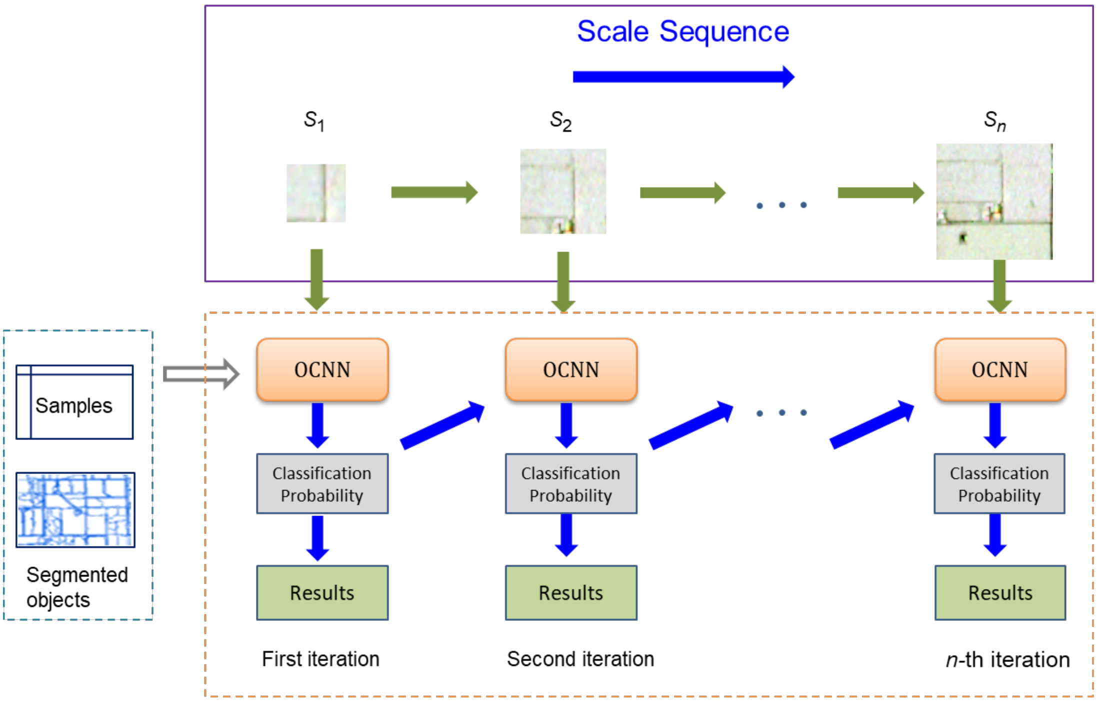

Fig. 1. Workflow of the presented SS-OCNN method

In the SS-OCNN, a scale sequence $\mathbf{S}$ consisting of a set of scales is first formulated to characterise segmented objects at different scales. In total, three parameters, including the total number of scales (n), the minimum scale (s_1), and the maximum scale (s_n), are needed to create a scale sequence as follows:

$$\mathbf{S} = \text{interpolate}(s_1, s_n, n) \tag{1}$$

where interpolate denotes the function of interpolation (a linear interpolation approach is adopted here). A scale sequence $\mathbf{S} = (s_1, s_2, \dots, s_i, \dots, s_n)$ where $i \in (1,2, \dots, n)$ is, thus, achieved using Eq. (1). The values of s_1 and s_n can be determined according to the geometric sizes of segmented objects (Zhang et al., 2020).

The segmented objects are classified at each scale. The classification results ($\mathbf{X}$) of the current (i-th) iteration are conditional on the outputs of the previous iteration, which formulates a Markov process as follows:

$$P\big(\mathbf{X}(s_i)^i\big) = P\big(\mathbf{X}(s_i)^i \big| \mathbf{X}(s_{i-1})^{i-1}\big) \tag{2}$$

where i represents the number of iterations within the Markov process, $P\big(\mathbf{X}(s_i)^i\big)$ are the classification probabilities at the i-th iteration, and s_i and s_{i-1} refers to the scale values adopted by the OCNN classifier at i-th and i-1-th iterations, respectively.

Suppose $\mathbf{M}$ is a scene of remotely sensed imagery with a total number of m classes to be classified. Let $\mathbf{O}$=$(o_1, o_2, \dots, o_j, \dots, o_v)$ represent the set of segmented objects from $\mathbf{M}$, where o_j and v are the j-th object and the total number of objects, respectively. Let $\mathbf{T}$= $(t_1, t_2, \dots, t_k, \dots, t_u)$ denote the set of training samples, where t_k and u are the k-th sample and the total number of samples, respectively. Herein, $\mathbf{T}$ is employed to train the OCNN model and, thus, estimate class probabilities per object within each scale through the iterative process. Fig. 1 gives a general flowchart of the developed SS-OCNN and the main steps are described in the following text.

For the first iteration, the training process of the OCNN classifier with a scale value of s_1 can be represented as:

$$\mathrm{OCNN}^1 = \mathrm{OCNN.Train}(\mathbf{M}, \mathbf{T}, s_1) \tag{3}$$

The trained OCNN model is employed to calculate the classification probabilities $P(\mathbf{X})^1$ as follows:

$$P(\mathbf{X})^1 = \mathrm{OCNN}^1.\mathrm{Predict}(\mathbf{M}, \mathbf{O}, s_1) \tag{4}$$

From the i-th iteration where $i \geq 2$, the original image ($\mathbf{M}$) and the classification probabilities at the previous iteration (i-1) are combined as conditional probabilities for the current classification as:

$$\mathbf{M}_{\mathrm{con}}{}^{i} = \mathrm{Concate}(\mathbf{M}, P(\mathbf{X})^{i-1}) \tag{5}$$

where $\mathrm{Concate}$ is a function used to concatenate the original image $\mathbf{M}$ with the probabilities outputted at the previous iteration. The training process at the i-th iteration can, thus, be represented as:

$$\mathrm{OCNN}^{i} = \mathrm{OCNN.Train}({\mathbf{M}_{\mathrm{con}}}^{i}, \mathbf{T}, s_{i}) \tag{6}$$

Note that the OCNN model is rebuilt at each iteration and trained from scratch. The classification probabilities at the current iteration can be predicted subsequently using the trained OCNN model as:

$$P(\mathbf{X})^{i} = \mathrm{OCNN}^{i}.\mathrm{Predict}({\mathbf{M}_{\mathrm{con}}}^{i}, \mathbf{O}, s_{i}) \tag{7}$$

Using Eq. (7), the probability of being assigned to each class for each segmented object is predicted at each iteration. Note that the space size of $P(\mathbf{X})^{i}$ is the same as the image $\mathbf{M}$, and the number of bands contained in $P(\mathbf{X})^{i}$ is equal to the number of classes m to be differentiated.

The final classification result can be achieved according the probabilistic output of the last iteration ($P(\mathbf{X})^{n}$) as:

$$\mathbf{M}_{\mathrm{result}} = \arg\ \max(P(\mathbf{X})^{n}) \tag{8}$$

where $\arg\ \max$ is a function assigning the class label with the maximum membership to each object of the imagery, and the object-based final classification map can, thus, be generated.

By combining the scale sequence with the OCNN classifier, the proposed SS-OCNN methodology for crop classification essentially has three major advantages:

1. The SS-OCNN can maintain the boundary of crop parcels precisely in the classification map using the object-based CNN classifier, whereas the standard pixel-wise CNN blurs crop boundaries.

2. The SS-OCNN increases classification accuracy gradually using the scale sequence, through which critical information can be transferred from local scales to larger scales.

3. The SS-OCNN is implemented automatically without conducting laborious trial and error experiments to choose the optimal scale, which significantly increases the computational efficiency.

3. Experimental results and analysis

3.1 Study area and materials

Two agricultural areas, lying in the central region of the Sacramento Valley, California, were selected for this research (Fig. 2). California possesses about 15% of the national receipts for crops and has long been considered the most varied and productive agricultural region across the USA (California Agricultural Statistics, 2011). Both study sites are strongly heterogeneous and highly distinctive in crop compositions and are, therefore, suitable to test the effectiveness of the developed approach. To further investigate the generalisation of the SS-OCNN, fine spatial resolution SAR and optical remotely sensed imagery were adopted in S1 and S2, respectively.

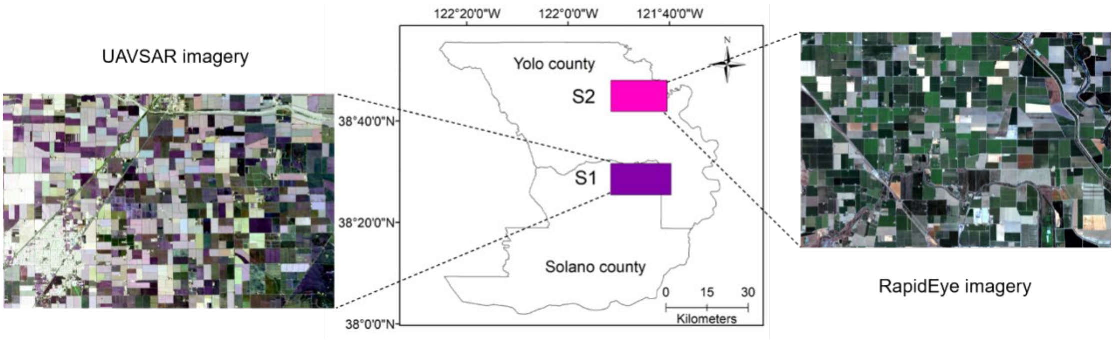

Fig. 2. The two study sites located in the agricultural district of California.

For S1, an L-band Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR) image composed of three linear polarizations (HH, HV, and VV) was acquired on 29 Aug, 2011. The UAVSAR data (3474×2250 pixels) were provided in the GRD (georeferenced) format with a fine spatial resolution of 5 m. In addition to the linear polarizations, three polarimetric parameters (i.e. entropy, anisotropy, and alpha angle from the Cloude-Pottier decomposition) which are sensitive to biophysical parameters of crops (Li et al., 2019), were also derived from the original dataset and used for crop classification. In total 10 dominant crop types were identified in S1 (Table 1).

For S2, an optical RapidEye image consisting of five bands (Blue, Green, Red, Red-edge, and Near-infrared) was captured on 10 July 2016. The image (Level 3A Ortho product) was delivered with sensor, radiometric and geometric corrections already applied. The spatial extent of the image is 3222×2230 pixels, with a fine spatial resolution of 5 m. To acquire surface reflectance, the image was further atmospherically corrected using atmospheric and topographic correction (ATCOR) method. A total of nine crop types were found within S2 (Table 1).

Cropland Data Layer (CDL) datasets are generated annually by the United States Department of Agriculture (USDA, Boryan et al., 2011), and have been adopted widely as a reference for crop monitoring and classification owing to its very high accuracy (Zheng et al., 2015; Cai et al., 2018; Li et al., 2019a). For example, the overall classification accuracy of the CDL is reported to be as high as 88.3% for all crops across California in 2016. To collect typical samples, crop parcels in the remotely sensed scene with an area larger than 5 ha were targeted and delineated manually according to the

CDL layer (Li et al., 2019a). These digital polygons were divided subsequently randomly into three parts: 40% for model training, 10% for model validation, and the remaining 50% for model testing. Training and validation sample points were collected within the training and validation polygons, respectively, using a stratified random sampling scheme to ensure they come from different crop polygons. The number of samples for each crop type was, thus, generally proportional to the total area of each crop. For crop classes with relatively small area (e.g. Pepper in S1), the sample sizes were increased to make them more comparable with the sample sets of other classes. In total, 1415 and 1262 sample points (training and validation) were acquired within S1 and S2, respectively, as shown in Table 1. To test comprehensively the classification accuracy, a wall-to-wall assessment was adopted for each study site (Zhong et al., 2019). That is, all pixels falling in the testing polygons were employed for accuracy assessment.

Table 1 Crop categories over both study areas with sample size for each crop.

Study area	Crop category	Training sample	Validation sample	Total sample
S1	Walnut	110	30	140
	Almond	110	30	140
	Alfalfa	125	32	157
	Hay	101	25	126
	Clover	110	28	138
	Winter wheat	120	30	150
	Corn	108	27	135
	Sunflower	122	32	154
	Tomato	120	30	150

	Pepper	106	28	134
S2	Walnut	108	27	135
	Almond	115	30	145
	Fallow	90	22	112
	Alfalfa	124	31	155
	Winter wheat	116	30	146
	Corn	93	24	117
	Sunflower	130	32	162
	Tomato	141	36	177
	Cucumber	93	24	117

3.2 Model architecture and parameter settings

The SS-OCNN needs to segment the input image into homogeneous objects (i.e. crop parcels). In this research, a multi-resolution segmentation (MRS) algorithm was performed using the eCognition Developer software to achieve segmented objects. The control parameters of the MRS algorithms were optimised using a systematic trial-and-error procedure (Duro et al., 2012) until the segmented objects matched well with crop boundaries according to visual inspection. The adopted parameter combinations for both study areas are listed in Table 2. Note that the value of scale is tuned to be relatively small to generate a small amount of over-segmentation and ensure that the segmented objects are homogeneous.

Table 2 Parameters adopted by the multi-resolution segmentation algorithm for the UAVSAR and RapidEye imagery.

Study site	Imagery	Scale	Shape	Compactness	Number of objects	Mean area of objects (ha)
S1	UAVSAR	30	0.2	0.7	3040	6.43
S2	RapidEye	180	0.3	0.6	3867	4.65

To classify each segmented object into specific crop types, a standard CNN was applied by taking the centre point of each object as the convolution location to extract both within-object and contextual information (Zhang et al., 2018b; Li et al., 2019b). The segmented objects were subsequently labelled using the trained CNN, such that the entire image was classified at the object level. The structure of the CNN employed in the SS-OCNN method was similar to the AlexNet with six hidden layers (i.e. three pairs of convolution and max-pooling layers (Fig. 3). Small filters were designed in the convolutional layers (5×5 for the first layer and 3×3 for the others), and the number of filters for each layer was set to 64 to learn within-object and contextual information for each segmented object. A rectified linear unit (ReLU) was adopted as the activation function for all layers. To prevent the CNN from overfitting, a regularization technique called "dropout" which randomly drops a number of neurons contained in a layer was applied before the fully connected layer (Srivastava et al., 2014). As suggested by Zhong et al. (2019), the dropout value was selected from {0.1, 0.2, 0.3, 0.4, 0.5} and 0.3 was found to be optimal at both sites. The number of epochs was set to 600 to allow the deep network to fully converge through backpropagation.

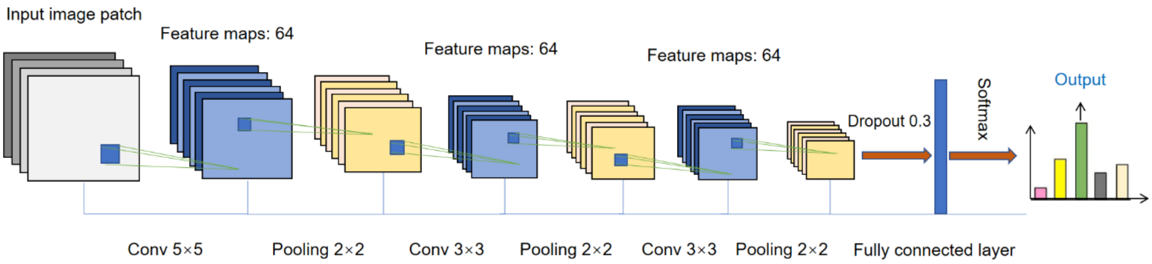

Fig. 3. Model architecture and parameter settings of the deep network employed in the SS-OCNN method.

3.3 Benchmark methods and parameter settings

To further assess the effectiveness of the developed SS-OCNN, four typical methods, including the standard pixel-wise CNN (PCNN), object-based image analysis (OBIA), the object-based CNN (OCNN), and the multiscale object-based CNN (MOCNN) were used as benchmarks. To make a fair comparison, the number of hidden layers for the three CNN-based benchmarks (i.e. PCNN, OCNN, and MOCNN) were set as six (the same as the network structure of the proposed SS-OCNN) through cross-validation. Parameters including dropout and number of iterations of the three benchmarks were also set in line with those of the SS-OCNN. Descriptions and parameter settings of these benchmarks are detailed as follows:

PCNN: the PCNN predicts the labels of all pixels within the entire image using densely overlapping patches. The input window size was set as 24×24 for both study sites according to previous experience (Li et al., 2019b). The number of filters for each layer was optimised as 32. Settings of the other parameters were the same as for the OCNN.

OBIA: The OBIA was implemented based on the segmented objects generated by the MRS algorithm. Several hand-crafted features were acquired from each object, including spectral features (i.e. mean and standard deviation) and texture variables. These hand-crafted features were subsequently fed into a parameterised SVM classifier for object-based classification.

OCNN: In contrast to the pixel-wise CNN, the OCNN was also applied based on segmented objects. Unlike the SS-OCNN which uses a range of scales, the OCNN adopts only one input window (i.e. scale) selected from a range of sizes through cross-validation, including {16×16, 24×24, 32×32, 40×40, 48×48}. The optimal scale was found to be 40×40 for both sites.

MOCNN: The multiscale OCNN (MOCNN) is a variant of the OCNN in which three input scales (window sizes) were adopted to enhance the generalisation of the network (Lv et al., 2018). As suggested by Lv et al. (2018), three CNN window sizes at 30 × 30, 40 × 40, and 50 × 50 were employed as the input windows to achieve predictions for each segmented object. Those predictions were subsequently fused with a majority voting strategy to achieve the final classification results.

3.4 Classification results and analysis

3.4.1 The SS-OCNN results

As described above, the SS-OCNN requires predefined minimum and maximum scales. In this research, the two values were respectively set as 8×8 and 48×48 for both study sites, which approach the major axis lengths of the minimum and maximum crop parcels, respectively. A range of scales were interpolated between the two end-points, thus, achieving a sequence of scales (i.e. input window sizes of the CNN). To demonstrate how the number of scales influences the classification results, the SS-OCNN was implemented with different numbers of scales (equivalent to iterations) over both sites (Fig. 4). Note that the smallest and second smallest numbers of iterations were one and two, indicating the minimum scale only (8×8) and the minimum and maximum scales (i.e. 8×8 and 48×48), respectively. Obviously, the accuracies of the SS-OCNN increased rapidly and significantly over both sites (from 78.69% to 85.50 for S1 and 82.91% to 88.43% for S2) as the number of iterations increased from one to three. The accuracies then increased gradually with the number of scales and reached a maximum at 6 scales for both sites, with up to 87.79% for S1 and 89.46% for S2, respectively. The accuracies tended to be stable (or slightly decreased) as the number of scales exceeded six (the optimal number of scales).

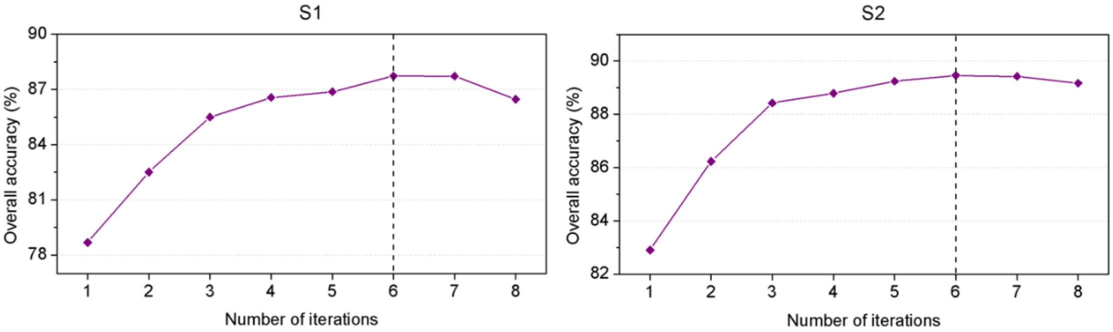

Fig. 4. Variations in overall accuracy of crop classifications achieved by the SS-OCNN with iteration over both sites. The black dash line highlights the highest accuracy acquired at iteration 6.

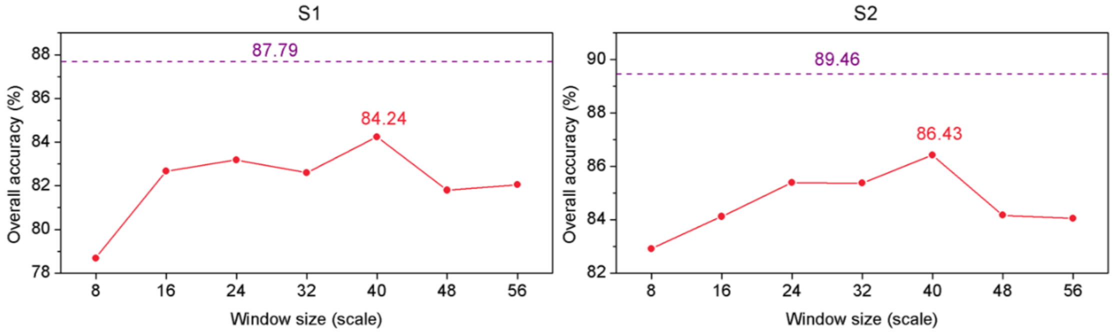

Fig. 5. Influence of window size on crop classification accuracies using the OCNN (red solid lines) and the proposed SS-OCNN (violet dashed lines) for S1 and S2.

The SS-OCNN method classified the remotely sensed imagery using the scale sequence and, thus, did not require specification of an optimal scale like the OCNN. Fig. 5 demonstrates the influence of these scales (window sizes) on the overall accuracy of OCNN, with the scales varying from 8×8 to 56×56 with a step size of 8. As shown in the figure, the SS-OCNN (violet dashed line) consistently outperformed the OCNN (red solid line) for crop classification (87.79% and 89.46%) for both S1 and S2. For the OCNN, the optimal scale was found to be 40×40 for both sites. The greatest accuracies of the OCNN using this optimal scale were only 84.24% and 86.43% for S1 and S2, respectively; over 3% smaller than those of the SS-OCNN.

To demonstrate visually how the SS-OCNN increased classification accuracy with iteration, the crop classification results are shown against iteration for three subsets of S1 (Fig. 6) and S2 (Fig. 7), respectively. The yellow and red circles in the figures indicate correct and incorrect classifications, respectively.

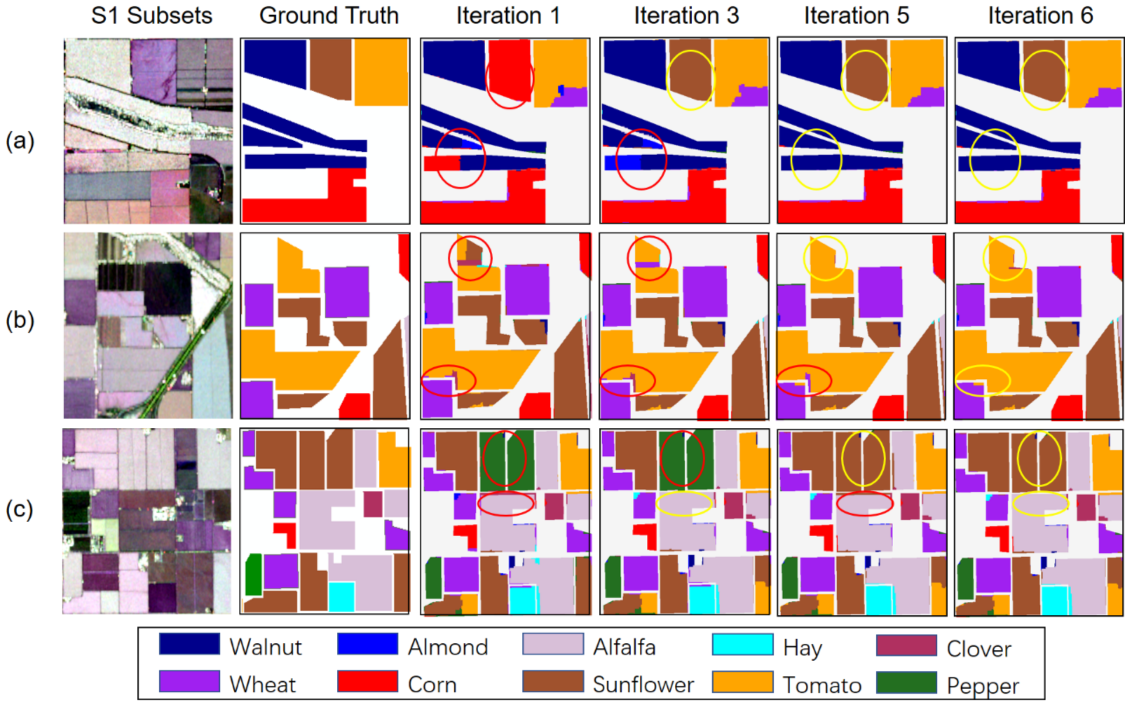

Fig. 6. Subset crop classification results in S1 achieved using the SS-OCNN along with iteration, respectively. Correct and incorrect classifications are marked using yellow and red circles, respectively.

In S1, both iterations 1 and 3 failed to identify Sunflower from Pepper, as shown by the red circles in Fig. 6 (c). At the same time, parts of Walnut and Tomato were, respectively, misclassified as Almond and Sunflower (Fig. 6 (a) and (b)), because of the similarity of spectral reflectance between them. These problems were solved through iteration by including more scales in the SS-OCNN model. As shown in Fig. 6, those misclassifications were rectified in the classification map of iteration 5. For example, the two Sunflower parcels in the upper part of Fig. 6 (c) were correctly identified. However, iteration 5 falsely classified Alfalfa as Sunflower, as shown in Fig. 6 (c) (red circles). Further, parts of Tomato were misidentified at both iterations 3 and

5 (Fig. 6 (b)). Fortunately, all of the aforementioned misclassifications were finally revised in the classification map of iteration 6. For example, parts of Walnut that were misclassified throughout iterations 1 to 3 were distinguished precisely at iteration 6, as shown by the yellow circles of Fig. 6 (a). Besides, the classification errors for Tomato and Alfalfa at iteration 3 were also resolved, as illustrated in Fig. 6 (b and c). In short, the proposed SS-OCNN method achieved a desirable result at iteration 6, where crop parcels were classified accurately.

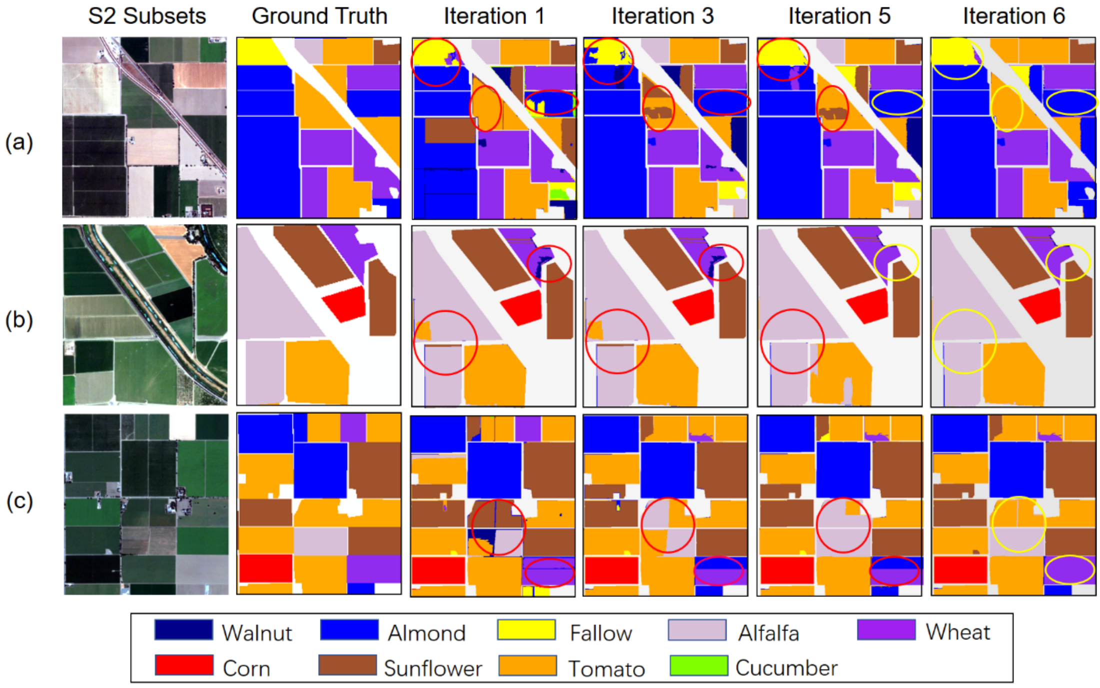

Fig. 7. Subset crop classification results in S2 achieved using the SS-OCNN along with iteration. Correct and incorrect classifications are marked using yellow and red circles, respectively.

Similar to S1, the classification results of S2 also presented obvious increases in accuracy with iteration. For example, the linear noise within and between crop parcels at iteration 1 (Fig. 7 (a and c)) were eliminated after iteration 1. Another significant increase in accuracy gained through iteration was the differentiation between crops with similar spectral reflectance (e.g. Sunflower and Tomato). For example, a falsely classified Tomato parcel at iterations 1 to 5 was correctly identified at iteration 6, as illustrated in Fig. 7 (c). Besides, some misclassified crop parcels were gradually classified accurately with iteration. For example, crop parcels misclassified as Almond in the upper left corner of Fig. 7 (a) and lower right corner of Fig. 7 (c) were gradually rectified to Fallow and Winter wheat, respectively, with iterations increasing from 1 to 6.

3.4.2 Benchmark comparison for crop classification

The proposed SS-OCNN was assessed against a range of comparators, including the PCNN, OBIA, OCNN and MOCNN. Tables 3 and 4 illustrate the classification accuracy assessment for S1 and S2, respectively, using the overall accuracy (OA), Kappa coefficient (κ), and per-class mapping accuracy. As can be seen from the tables, the SS-OCNN consistently produced the largest classification accuracies, with the OA up to 87.79% for S1 and 89.46% for S2, greater than for the MOCNN (85.85% and 87.27%), OCNN (84.24% and 86.43%), OBIA (83.72% and 82.81%) and PCNN (79.21% and 80.72%). The Kappa coefficient results are consistent with the OA, with the κ of SS-OCNN reaching 0.86 for S1 and 0.87 for S2, more accurate than those of the MOCNN (0.83 and 0.85), OCNN (0.82 and 0.84), OBIA (0.81 and 0.79) and PCNN (0.76 and 0.77), respectively.

The superiority of the proposed method was also demonstrated by per-class classification accuracy. As illustrated by Tables 3 and 4, the SS-OCNN achieved the most accurate mapping accuracy for half of the crop categories in S1 and most categories in S2 (highlighted in bold face). For S1, the most notable accuracy increase was seen for Hay with an accuracy of 75.87%, dramatically greater by 17.18%, 20.95%, 37.76% and 36.31% in comparison with the MOCNN, OCNN, OBIA and PCNN, respectively (Table 3). Likewise, an impressive increase was obtained for Winter wheat, with an increase of around 10% in accuracy. In addition, a marked accuracy increase can be seen in Pepper in comparison to the MOCNN and OCNN classifications, with about a 19% increase in accuracy. Besides, Almond, Tomato, and Corn, presented only slight increases in accuracy in comparison to the benchmarks. Other crop classes, such as Walnut and Sunflower, did not show obvious increases in average accuracy.

For S2, the proposed SS-OCNN obtained satisfactory classification accuracy for most of the crop classes (Table 4), with accuracies of six crop classes (including Walnut, Winter wheat, Corn, Sunflower, Tomato, and Cucumber) being larger than 89%. The largest accuracy increases were obtained for Fallow, achieving an increase of 20.37%, 29.47%, 24.35% and 33.42% compared with the MOCNN, OCNN, OBIA and PCNN, respectively. The accuracy increases were also significant for Cucumber and Almond, with average increases of 11.82% and 9.73%, respectively. For the Alfalfa, Winter wheat, and Sunflower classes, a moderate accuracy increase (around 3%-6%) was achieved. Other crop classes, including Corn and Tomato, demonstrated a relatively small increase (<2%) in average accuracy in comparison to the benchmarks.

Table 3 Crop classification accuracy comparison amongst pixel-wise CNN, OBIA, object-based CNN, and the presented SS-OCNN on the first study area (S1). The largest accuracies are highlighted by bold font.

Crop Class (S1)	PCNN	OBIA	OCNN	MOCNN	SS-OCNN
Walnut	86.55	92.94	**95.32**	91.48	91.40
Almond	91.02	84.53	93.22	94.61	**94.69**
Alfalfa	83.75	88.27	**89.78**	89.42	88.58
Hay	39.56	38.11	54.92	58.69	**75.87**
Clover	62.57	76.33	72.41	**74.65**	73.28
Winter wheat	74.05	76.04	77.02	80.76	**87.76**
Corn	94.91	87.69	91.47	**95.97**	93.20
Sunflower	82.02	84.66	**89.88**	89.79	87.83
Tomato	85.72	90.91	88.22	89.21	**90.45**
Pepper	54.75	86.28	56.43	55.95	**75.20**
OA (%)	79.21	83.72	84.24	85.85	**87.79**
Kappa	0.76	0.81	0.82	0.83	**0.86**

Table 4 Crop classification accuracy comparison amongst pixel-wise CNN, OBIA, object-based CNN, and the presented SS-OCNN on the first study area (S2). The largest accuracies are highlighted by bold font.

Crop Class (S2)	PCNN	OBIA	OCNN	MOCNN	SS-OCNN
Walnut	75.73	83.10	**90.12**	89.34	89.37

Almond	70.65	74.76	78.76	80.55	**85.91**
Fallow	48.22	57.29	52.17	61.27	**81.64**
Alfalfa	76.46	81.11	84.64	85.50	**85.73**
Winter wheat	88.90	89.80	88.79	90.79	**94.77**
Corn	93.86	96.58	**99.17**	99.13	99.02
Sunflower	85.13	83.18	87.01	86.61	**89.71**
Tomato	84.89	85.76	91.60	**92.03**	89.80
Cucumber	78.85	75.27	78.86	77.27	**89.38**
OA (%)	80.72	82.81	86.43	87.27	**89.46**
Kappa	0.77	0.79	0.84	0.85	**0.87**

The classification maps of the proposed SS-OCNN and the benchmark methods were also compared for both S1 (Fig. 8) and S2 (Fig. 9) using three subsets, respectively. It is clear that the pixel-wise CNN (PCNN) achieved undesirable results caused by heavy salt-and-pepper noise. Besides, pixels near crop parcel boundaries were often misclassified, as shown in Fig. 8 (b and c) and Fig. 9 (a and c). In contrast, the OBIA method demonstrated more smooth classification results with precise boundary information. The classifications of Tomato as well as Alfalfa were improved to some extent, as illustrated by Fig. 8 (b) and Fig. 9 (c). However, the OBIA was inferior to the object-based CNN (OCNN) in eliminating some linear noise occurring between crop parcels, as demonstrated in Fig. 9 (a and b). Besides, the OCNN has certain advantages over OBIA in differentiating crop classes with similar spectra. For example, Winter wheat and Hay were more accurately distinguished from each other, as demonstrated in Fig. 8 (b) and (c). Nevertheless, severe confusion between Corn, Sunflower, and Pepper in S1, and between Sunflower and Tomato in S2, still exists in the OCNN classifications, as shown in Figs. 8 and 9, respectively. A slight increase in classification accuracy was achieved by the MOCNN in comparison with the OCNN, yet most of the misclassifications were not resolved. The SS-OCNN, surprisingly, revised all of the aforementioned misclassifications while achieving the smoothest results over both S1 and S2 by incorporating a scale sequence into the OCNN model.

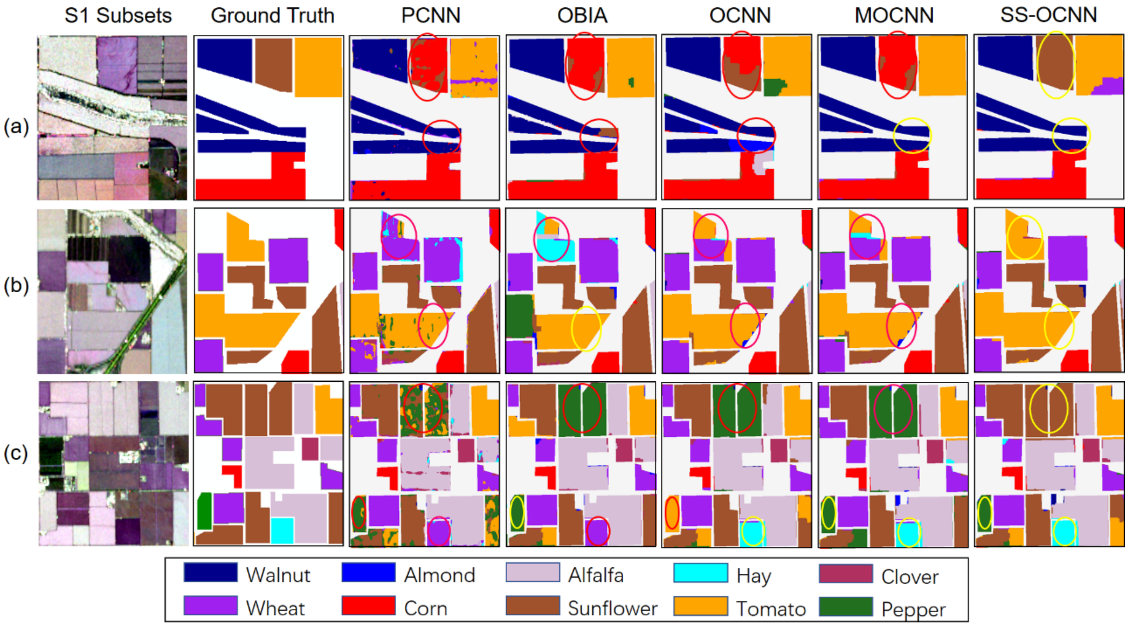

Fig. 8. Representative image subsets of S1 with the crop classification maps achieved by PCNN, OBIA, OCNN, and the proposed SS-OCNN, respectively. The incorrect and correct classifications are highlighted using red and yellow circles, respectively.

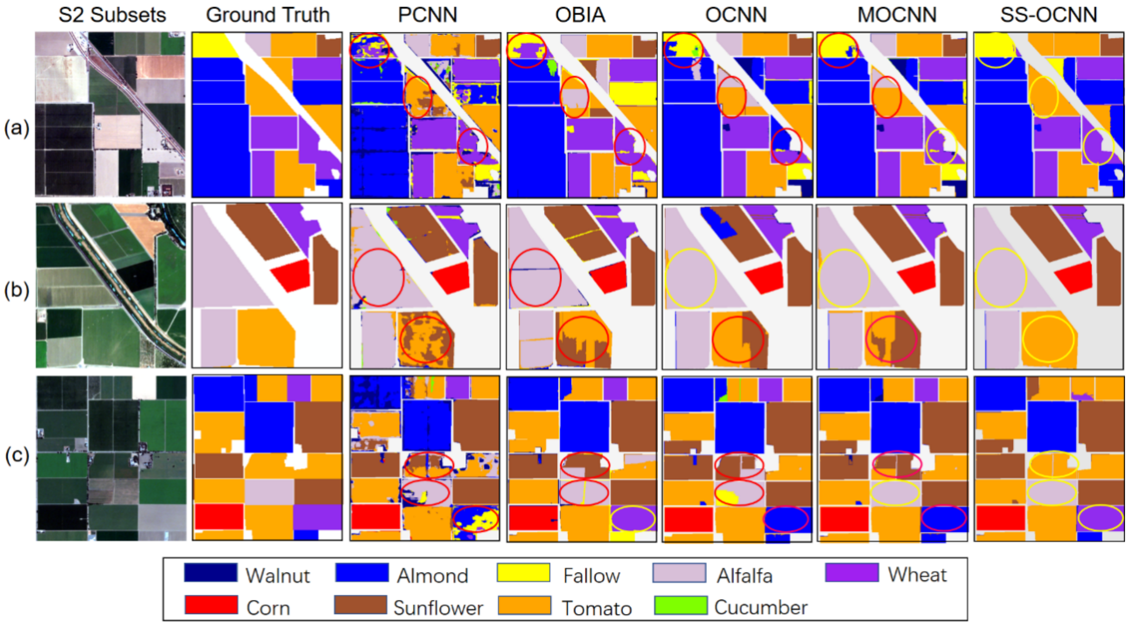

Fig. 9. Representative image subsets of S1 with the crop classification maps achieved by PCNN, OBIA, OCNN, and the proposed SS-OCNN, respectively. The incorrect and correct classifications are highlighted using red and yellow circles, respectively.

4. Discussion

Crop classification using FSR remotely sensed image is a challenging task due to great spatial and temporal intra-class variation in remote sensing spectra or polarimetric signatures (Li et al., 2019a). This research illustrates that pixel-wise classifiers are inferior to object-based classifiers (i.e. OBIA and OCNN) in terms of crop classification from FSR images. This is because object-based methods classify the entire remotely sensed image at the object-level, thus, significantly reducing salt-and-noise effects while retaining precise boundary information on crop parcels (Figs. 8 and 9). Specifically, the OCNN was generally superior to the OBIA (Tables 3 and 4) since both within-object information (low-level features) and between-object information (high-level features) were employed for crop class discrimination (Zhang et al., 2019). Nevertheless, the OCNN did not achieve satisfactory accuracy for some crop parcels subjected to only one scale (i.e. input window of CNN) being adopted during the classification process.

Currently, an “optimal” scale (input window size) needs to be determined for CNN classifiers through a trial and error procedure. Such a scale selection process is rather tedious and labour-intensive, taking a lot of time. Further, even though the so-called “optimal” scale is acquired, it is just a compromise by which the deep learning models can achieve the relatively (not true) optimal classification accuracy (Zhang et al., 2020). In fact, one scale is far from sufficient to effectively and comprehensively capture the multi-level (i.e. multi-scale) features of crop parcels in consideration of the often great variation in spatial size. In principle, features learned by deep learning models at different scales provide unique information on the characteristics of ground objects. For example, local features of a certain crop parcel can be captured by deep learning models via a smaller scale, while global features can be acquired using a relatively larger scale. Such multi-scale information substantially enhances the observational dimension of ground objects, which is of great potential benefit for accurate classification amongst heterogeneous crop classes. In the SS-OCNN, features at smaller scales are conducted and integrated to features at larger scales gradually at the object level using the Markov model, thus, achieving multi-scale comprehensive observations on the ground objects. In other words, the SS-OCNN authentically and effectively fuses multi-scale features provided by the FSR remotely sensed imagery for increased classification accuracy.

Note that the SS-OCNN is fundamentally different from the MOCNN method, though they both adopt multiple scales in the classification process. In essence, the MOCNN is a variant of the single-window OCNN classifier as the final classification is based on the classification result at the optimal scale (i.e. 40 × 40). The remaining two scales (i.e. 30× 30 and 50× 50) generally achieved lower classification probabilities for most of crop parcels and they, thus, contribute little to the final classification accuracy of the MOCNN with a majority voting strategy. This explains why the MOCNN just slightly (0.5% to 1.5%) increased the overall classification accuracy in comparison to the OCNN for both sites. Besides, it is unreasonable to compare directly the probabilities of the OCNN achieved at different scales since they are not generated based on the same condition (scale). Some mislabeled crop parcels at a certain scale may have very high classification probability. By using the majority voting strategy, these misclassifications will transfer to the final classification map. For example, the MOCNN decreased the accuracy of Walnut in comparison to the OCNN.

Although the SS-OCNN significantly increased the overall classification accuracy for both study sites, it demonstrated clear differences in ability to increase the classification accuracy of different crop categories. In general, the accuracy increases of small biomass crop classes (e.g. forage and grain) were large, while those of the large biomass classes (e.g. summer field crops and fruit crops) were small. For example, the SS-OCNN surprisingly increased the accuracies of Hay in S1 and Fallow (with mixed pasture) in S2 by around 28% compared with those of OCNN, while the accuracies of Walnut and Corn were only increased slightly (0-2%) for both sites. A possible reason for this is that large biomass crops usually have strong signals in remotely sensed imagery and can, thus, be differentiated from each other relatively easily (Li et al., 2019a). In contrast, small biomass crops may only have relatively weak signals (Li et al., 2020) and they are, thus, difficult to classify accurately using a CNN at a specific scale. By using the SS-OCNN, multi-level features of small biomass crops were extensively learned and integrated for classification. In other words, the crops were observed at different scale dimensions which was extremely beneficial to capture their unique spectral or structural characteristics and, thereby, increase crop classification accuracy.

In this research, a forward scale sequence (i.e. start small, get larger) was adopted in the SS-OCNN for crop classification. Actually, there are a large varieties of sampling strategy with respect to orders of scale sequence, for example, the backward scale sequence (i.e. start large, get smaller) and random scale sequence (i.e. select a scale randomly at each iteration). However, we found that they were inferior to the forward scale sequence (not shown here), which is consistent with our previous work (Zhang et al., 2020). The reason for the superiority of the forward scale sequence might be that such a sampling strategy can allow local features (i.e. small scale features) to be effectively transferred and fused with global features (i.e. large scale features). Of course, the forward scale sequence can be further optimised, e.g. adopting a non-linear interpolation strategy in the process of scale sequence generation, which deserves further investigations.

The proposed SS-OCNN was employed to classify crop classes at the object level as an example, and achieved surprisingly accurate results with the help of the scale sequence method. In fact, real-world features are usually represented over a series of scales (e.g., small-scale house and large-scale park), and spatial scale is considered as a core issue in feature representation of remotely sensed imagery (Ming et al., 2015). In addition to agriculture, the proposed method should also be effective for image classification over a wide range of landscapes. For example, wetlands usually develop into parcels of different sizes affected by several factors, such as climate, soil and hydrogeological conditions. The parcel sizes of land use categories also vary greatly in consideration of functional requirements (Zhang et al., 2018b). As a result, the proposed method has great potential and a wide application prospect.

5. Conclusion

Crop classification using FSR remotely sensed imagery remains a great challenge due to the complex spatial and temporal patterns of croplands. Advanced deep convolutional neural network (CNN) has markedly increased the accuracy of crop classification by adopting an input patch (window) to extract multi-level features from raw imagery. Nevertheless, it is extremely difficult to determine an optimal window size for CNNs since areas of crop parcels often vary greatly. Besides, the standard pixel-wise CNN method tends to deform parcel geometry, which impairs the identification of crop parcels. In this paper, a novel Scale Sequence Object-based CNN (SS-OCNN) approach was proposed for crop classification using FSR remotely sensed imagery. In the SS-OCNN, segmented objects (crop parcels) were identified precisely using a CNN model combined with a range of automatically generated scales, thus, solving the problems associated with application of the standard CNN to FSR image-based crop classification. Experimental results on two heterogeneous agricultural fields with FSR SAR and optical imagery, respectively, demonstrated that the SS-OCNN achieved the most accurate crop classification results, consistently and significantly more accurate than the standard pixel-wise CNN, single-scale object-based CNN, and the multi-scale object-based CNN. Besides, the SS-OCNN also produced the smoothest results, with crop boundary information precisely delimited. Specifically, small biomass crop classes (e.g. forage and grain) that were extremely difficult to characterise using the benchmark methods were classified accurately. We, thus, conclude that the SS-OCNN is an effective method for crop classification from FSR remotely sensed image. Further, the SS-OCNN provides a novel and general solution for image classification of heterogeneous landscapes and it, therefore, has great potential and a wide application prospect.

Acknowledgements

This research was financially supported by the National Natural Science Foundation of China (41301465) and the National Key Research and Development Program of China (2017YFB0503602).

References

Azar, R., Villa, P., Stroppiana, D., Crema, A., Boschetti, M., Brivio, P.A., 2016. Assessing in-season crop classification performance using satellite data: a test case in Northern Italy. Eur J Remote Sens. 49, 361-380.

Bastiaanssen, W.G.M., Ali, S., 2003. A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. Agr Ecosyst Environ. 94 (3), 321-340.

Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens Environ. 204, 509-523.

Biradar, C.M., Xiao, X.M., 2011. Quantifying the area and spatial distribution of double- and triple-cropping croplands in India with multi-temporal MODIS imagery in 2005. Int J Remote Sens. 32 (2), 367-386.

Cai, Y.P., Guan, K.Y., Peng, J., Wang, S.W., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. Remote Sens Environ. 210, 35-47.

California Agricultural Statistic, 2011. USDA’s National Agricultural Statistics Service. Retrieved February 3, 2018, from California Field Office. www.nass.usda.gov/ca.

Chai, D., Newsam, S., Zhang, H.K.K., Qiu, Y., Huang, J.F., 2019. Cloud and cloud shadow detection in Landsat imagery based on deep convolutional neural networks. Remote Sens Environ. 225, 307-316.

Chen, Y.S., Jiang, H.L., Li, C.Y., Jia, X.P., Ghamisi, P., 2016. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. Ieee T Geosci Remote. 54 (10), 6232-6251.

Chen, Y.Y., Ming, D.P., Lv, X.W., 2019. Superpixel based land cover classification of VHR satellite image combining multi-scale CNN and scale parameter estimation. Earth Sci Inform. 12 (3), 341-363.

Debats, S.R., Luo, D., Estes, L.D., Fuchs, T.J., Caylor, K.K., 2016. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. Remote Sens Environ. 179, 210-221.

Dey, S., Mandal, D., Robertson, L.D., Banerjee, B., Kumar, V., McNairn, H., Bhattacharya, A., Rao, Y.S., 2020. In-season crop classification using elements of the Kennaugh matrix derived from polarimetric RADARSAT-2 SAR data. Int J Appl Earth Obs. 88, 102059.

Dong, J.W., Xiao, X.M., Kou, W.L., Qin, Y.W., Zhang, G.L., Li, L., Jin, C., Zhou, Y.T., Wang, J., Biradar, C., Liu, J.Y., Moore, B., 2015. Tracking the dynamics of paddy rice planting area in 1986-2010 through time series Landsat images and phenology-based algorithms. Remote Sens Environ. 160, 99-113.

Duro, D.C., Franklin, S.E., Dube, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sens. Environ. 118, 259-272.

Essa, A., Sidike, P., Asari, V., 2017. Volumetric Directional Pattern for Spatial Feature Extraction in Hyperspectral Imagery. Ieee Geosci Remote S. 14 (7), 1056-1060.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2016. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. Ieee T Pattern Anal. 38 (1), 142-158.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 770-778.

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., Kingsbury, B., 2012. Deep Neural Networks for Acoustic Modeling in Speech Recognition. Ieee Signal Proc Mag. 29 (6), 82-97.

Hu, Q., Sulla, D., Xu, B.D., Yin, H., Tang, H.J., Yang, P., Wu, W.B., 2019. A phenology-based spectral and temporal feature selection method for crop mapping from satellite time series. Int J Appl Earth Obs. 80, 218-229.

Jiang, M.H., Shen, H.F., Li, J., Yuan, Q.Q., Zhang, L.P., 2020. A differential information residual convolutional neural network for pansharpening. Isprs J Photogramm. 163, 257-271.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet Classification with Deep Convolutional Neural Networks. Commun Acm. 60 (6), 84-90.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature. 521 (7553), 436-444.

Li, H.P., Zhang, C., Zhang, S.Q., Atkinson, P.M., 2019a. Full year crop monitoring and separability assessment with fully-polarimetric L-band UAVSAR: a case study in the Sacramento Valley, California. Int. J. Appl. Earth Obs. Geoinf. 74 (02), 45-56.

Li, H.P., Zhang, C., Zhang, S.Q., Atkinson, P.M., 2019b. A hybrid OSVM-OCNN Method for Crop Classification from Fine Spatial Resolution Remotely Sensed Imagery. Remote Sens-Basel. 11 (20), 2370.

Li, H.P., Zhang, C., Zhang, S.Q., Atkinson, P.M., 2020. Crop classification from full-year fully-polarimetric L-band UAVSAR time-series using the Random Forest algorithm. Int J Appl Earth Obs. 87, 102032.

Liu, L., Xiao, X.M., Qin, Y.W., Wang, J., Xu, X.L., Hu, Y.M., Qiao, Z., 2020. Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google Earth Engine. Remote Sens Environ. 239, 111624.

Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. Int J Remote Sens. 28 (5), 823-870.

Lv, X.W., Ming, D.P., Lu, T.T., Zhou, K.Q., Wang, M., Bao, H.Q., 2018. A New Method for Region-Based Majority Voting CNNs for Very High Resolution Image Classification. Remote Sens-Basel. 10 (12), 1946.

Ming, D.P., Li, J., Wang, J.Y., Zhang, M., 2015. Scale parameter selection by spatial statistics for GeOBIA: Using mean-shift based multi-scale segmentation as an example. Isprs J Photogramm. 106, 28-41.

Nobre, C.A., Sampaio, G., Borma, L.S., Castilla-Rubio, J.C., Silva, J.S., Cardoso, M., 2016. Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. P Natl Acad Sci USA. 113 (39), 10759-10768.

Pan, X., Zhao, J., Xu, J., 2019. An object-based and heterogeneous segment filter convolutional neural network for high-resolution remote sensing image classification. Int J Remote Sens. 40 (15), 5892-5916.

Persello, C., Tolpekin, V.A., Bergado, J.R., de By, R.A., 2019. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. Remote Sens Environ. 231, 111253.

Salehi, B., Daneshfar, B., Davidson, A.M., 2017. Accurate crop-type classification using multi-temporal optical and multi-polarization SAR data in an object-based image analysis framework (vol 38, pg 4130, 2017). Int J Remote Sens. 38 (18), 5271-5271.

Sidike, P., Sagan, V., Maimaitijiang, M., Maimaitiyiming, M., Shakoor, N., Burken, J., Mockler, T., Fritschi, F.B., 2019. dPEN: deep Progressively Expanded Network for mapping heterogeneous agricultural landscape using WorldView-3 satellite imagery. Remote Sens Environ. 221, 756-772.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. J Mach Learn Res. 15, 1929-1958.

Sylvain, J.D., Drolet, G., Brown, N., 2019. Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. Isprs J Photogramm. 156, 14-26.

Wang, Q., Yuan, Z.H., Du, Q., Li, X.L., 2019. GETNET: A General End-to-End 2-D CNN Framework for Hyperspectral Image Change Detection. Ieee T Geosci Remote. 57 (1), 3-13.

Wang, T., Zhang, H.S., Lin, H., Fang, C.Y., 2016. Textural-Spectral Feature-Based Species Classification of Mangroves in Mai Po Nature Reserve from Worldview-3 Imagery. Remote Sens-Basel. 8 (1), 24.

Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the US Central Great Plains. Remote Sens. Environ. 112 (3), 1096-1116.

Yan, F.Q., Zhang, S.W., 2019. Ecosystem service decline in response to wetland loss in the Sanjiang Plain, Northeast China. Ecol Eng. 130, 117-121.

Zhang, C., Harrison, P.A., Pan, X., Li, H.P., Sargent, I., Atkinson, P.M., 2020. Scale Sequence Joint Deep Learning (SS-JDL) for land use and land cover classification. Remote Sens Environ. 237, 111593.

Zhang, C., Pan, X., Li, H.P., Gardiner, A., Sargent, I., Hare, J., Atkinson, P.M., 2018a. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. Isprs J Photogramm. 140, 133-144.

Zhang, C., Sargent, I., Pan, X., Li, H.P., Gardiner, A., Hare, J., Atitinson, P.M., 2018b. An object-based convolutional neural network (OCNN) for urban land use classification. Remote Sens Environ. 216, 57-70.

Zhang, C., Sargent, I., Pan, X., Li, H.P., Gardiner, A., Hare, J., Atkinson, P.M., 2019. Joint Deep Learning for land cover and land use classification. Remote Sens Environ. 221, 173-187.

Zhao, J., Zhong, Y.F., Hu, X., Wei, L.F., Zhang, L.P., 2020. A robust spectral-spatial approach to identifying heterogeneous crops using remote sensing imagery with high spectral and spatial resolutions. Remote Sens Environ. 239, 111605.

Zheng, B.J., Myint, S.W., Thenkabail, P.S., Aggarwal, R.M., 2015. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. Int. J. Appl. Earth Obs. Geoinf. 34, 103-112.

Zheng, X.T., Yuan, Y., Lu, X.Q., 2019. A Deep Scene Representation for Aerial Scene Classification. Ieee T Geosci Remote. 57 (7), 4799-4809.

Zhong, L.H., Hu, L.N., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens Environ. 221, 430-443.