Remote Sensing of Environment 247 (2020) 111912

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution

Dujuan Zhang[a,d], Yaozhong Pan[a,b,*,1], Jinshui Zhang[c,d,**,1], Tangao Hu[e], Jianhua Zhao[f], Nan Li[a,d], Qiong Chen[b]

[a] State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing 100875, China
[b] Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining 810016, China
[c] State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China
[d] Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[e] Zhejiang Provincial Key Laboratory of Urban Wetlands and Regional Change, Hangzhou Normal University, Hangzhou 311121, China
[f] Data Management Center, National Bureau of Statistics of the People's Republic of China, Beijing 100826, China

ARTICLE INFO

Keywords:
Very high spatial resolution
Large-scale cropland mapping
Generalization
Deep convolutional neural networks
Fusion of low-level and high-level features

ABSTRACT

Timely and accurate delineation of the cropland extent over large area is crucial for operational agriculture monitoring and is also beneficial to address food security issues. Existing global datasets associated with cropland are limited by insufficient spatial resolution to properly represent areas with small parcel size distributions, and their less-than-ideal accuracies hamper application at regional and local scales. Diverse very high spatial resolution (VHSR) satellite systems are now available, offering sub-meter to five-meter resolution (e.g. Gaofen-1, Gaofen-2, and ZiYuan-3), and hence enabling explicit extraction of cropland areas from heterogeneous and fragmented landscapes. This study presented a generalized methodology for operational cropland mapping at very high resolution using a deep convolutional neural network to automatically learn the robust and discriminative features. Specifically, we slightly modified the pyramid scene parsing network (PSPNet) and combined deep long-range features with shadow local features to provide predictions with high level of detail. We demonstrated the modified PSPNet (MPSPNet) over four province-wide study areas (Heilongjiang, Hebei, Zhejiang and Guangdong) with diverse agrosystems across China from north to south using multi-source very high spatial resolution satellite images (mainly Gaofen-1 supplemented with Gaofen-2 and ZiYuan-3), with the overall accuracies ranging from 89.99% to 92.31%. Moreover, we compared MPSPNet with other CNN models and investigated the its behavior by visualizing the learned features on different layers, indicating that combining low and high level features for final classification was an efficient and accurate strategy for cropland mapping because the former capture edge information related to object boundaries and the latter could learn long-range spatial dependencies that helped recognize croplands. The temporal transfer and spatial transfer assessments from the respects of qualitative and quantitative corroborated the robust generalizability of the proposed method. And the contrast to the traditional object-based classification method also demonstrated the advantages and strong generalization capabilities of MPSPNet in extracting cropland using VHSR remote sensed images. We compared our results with the current cropland maps generated from FROM-GLC10, which further verified the effectiveness of the proposed approach for large-scale cropland mapping at very high resolution.

1. Introduction

Timely and accurate information regarding cropland extent and distribution is of world-wide interest for various applications, such as cropland dynamic monitoring, grain productivity investigation and evaluation, food security (Justice and Becker-Reshef, 2007; Thenkabail and Wu, 2012; See et al., 2015). Remotely sensed observations now provide extremely comprehensive sets of geographical images covering

* Correspondence to: Y. Pan, State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing 100875, China.
** Correspondence to: J. Zhang, State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China.
E-mail addresses: pyz@bnu.edu.cn (Y. Pan), zhangjs@bnu.edu.cn (J. Zhang).
[1] Present address: Beijing Normal University No. 19, XinJieKouWai St., HaiDian District, Beijing 100,875, P. R. China.

https://doi.org/10.1016/j.rse.2020.111912
Received 14 October 2019; Received in revised form 1 May 2020; Accepted 21 May 2020

almost every part of the earth's surface (Lefsky et al., 2002), and have long been important tools to identify and acquire cropland spatial distributions at local, regional and global scales (Waldner et al., 2015; Xiong et al., 2017; Phalke and Özdoğan, 2018; Massey et al., 2018; Yu et al., 2013; Fritz et al., 2015). Existing global cropland datasets have middle (10 m – 30 m) (e.g. Gong et al., 2013; Chen et al., 2014; Chen et al., 2015; Gong et al., 2019) or coarse resolution (250 m – 10 km) (e.g. Loveland et al., 2000; Friedl et al., 2002; Bartholomé and Belward, 2005; Ramankutty et al., 2008; Biradar et al., 2009; Thenkabail et al., 2009; Friedl et al., 2010; Pittman et al., 2010; Bontemps et al., 2011; Waldner et al., 2016), and commonly have the drawback that spatial resolution is not fine enough to accurately map areas with small parcel size distributions, such as commonly occur in agricultural landscapes around the world (Ozdogan and Woodcock, 2006; Griffiths et al., 2013). Besides, their comparatively poor overall accuracy (approximately 66–80%) limits their operational applications (Herold et al., 2008; Lu et al., 2016), particularly at regional and local scales.

Recent development of remotely sensing technologies providing very high spatial resolution (VHSR) images (e.g. Gaofen-1 (GF1), Gaofen-2 (GF2), and ZiYuan-3 (ZY3)), make it possible for cropland mapping with high level of thematic detail to resolve small farmland parcels. However, the rich information in VHSR images brings high intraclass differences and low interclass diversities (Bruzzone and Carlin, 2006), which makes mapping cropland over large areas somewhat challenging. Furthermore, cropland is a complicated and varying land cover class, and hence cropland spectral, texture, and shape characteristics in VHSR imagery vary over space and time, resulting in a considerable internal variability. Consequently, optimal cropland extraction models obtained from annotated images generally scale poorly with insufficient performance on new images from different sensors, geo-locations, or imaging conditions, which hampers their large-scale application. Therefore, it is imperative to develop a generalized and valid cropland mapping method scalable over large VHSR image spatial extents.

Many classification methods have been adopted for cropland mapping, including maximum likelihood (Abou EL-Magd and Tanton, 2003), nearest neighbors (Samaniego and Schulz, 2009), decision tree (De Fries et al., 1998; Pittman et al., 2010), random forest (Duro et al., 2012), neural networks (Liu et al., 2005), support vector machines (Mathur and Foody, 2008), etc. These have been applied either pixel or object based (Duro et al., 2012; Long et al., 2013). However, most current methods can extract only low or middle level features from the original data, which is not robust enough for cropland classification due to considerable intra-class variance in the large spatial cover. Most features were designed based on domain-specific knowledge and hence do not scale well over space and time (Kaiser et al., 2017), resulting in repeated manual annotation to construct training samples as geographic locations or imaging conditions changes. This is extremely labor-intensive, time-consuming, and inefficient for large-scale cropland mapping. Thus, there is urgent demand for class-specific high-level feature representation with spatio-temporal transferability to monitor croplands at VHSR over large areas.

Deep convolutional neural networks (CNNs) (LeCun et al., 2015) have recently gained much attention in the machine learning field due to their ability to automatically learn representative and discriminative features in a hierarchical manner from the training set. Consequently, CNNs have been introduced to geoscience and remote sensing (RS) communities for RS big data analysis (Zhang et al., 2016; Zhu et al., 2017) with promising potential for high resolution remote sensing imagery, including such as object detection (Cheng et al., 2016; Ševo and Avramović, 2016), road network extraction (Wang et al., 2015), and land cover/use classification (Zhang et al., 2018). CNNs comprise a stack of convolution or pooling layers which can cover wide contexts and extract deep abstract features from the original remote sensing imagery, providing strong description and generalization to tackle large-scale problems (Zhu et al., 2017). Besides, many studied have investigated CNN transferability of deep features for big data challenges in remote sensing (Marmanis et al., 2016; Jean et al., 2016). For example, Lyu et al. (2018) applied deep information learning to Landsat data for long-term annual mapping of urban areas, effectively and efficiently detecting urban change in massive remote sensing data sets with limited training data, demonstrating strong generalization and transferability of deep features.

Therefore, this paper takes these CNN advantages and proposes a generalized deep CNN approach to address the difficulty of mapping croplands over very large areas using VHSR images. Specifically, the CNN is based on the pyramid scene parsing network (PSPNet) (Zhao et al., 2017), an end-to-end pixel wise deep CNN architecture for scene parsing. However, downsampling operation in deep CNN is required to capture the long-range spatial dependencies that help recognize the classes but lose location information. Although the generated low-resolution high-level feature maps will be subsequently upsampled to the input resolution (Long et al., 2015; Chen et al., 2018), the boundary information of the objects cannot be determined precisely because the upsampling reconstruct the presence and rough position of the objects rather than their shapes. These issues can be addressed in two flavors: (i) methods that use multi-resolution segmentation (MRS) as the post-processing step to delineate object boundaries (Sun et al., 2018); (ii) methods that combine low-level local features with high-level features to produce fine predictions (Maggiori et al., 2017). For the first approach, the segmentation is time-consuming and difficult, which is not conducive to operational applications in large-scale practice. Thus, we developed an modified PSPNet (MPSPNet) which improves on the original PSPNet by integrating low-level and high-level information to produce fine-grained classification with precise boundaries, providing a viable approach to learn long range context knowledge with high-level information, and capture small scale spatial variations with low-level feature representation regarding edge detectors. We evaluated the proposed method performance and accuracy on four study areas with large spatial extent and various complex agricultural landscapes in China.

The remainder of this paper is organized as follows. Section 2 presents the materials and process employed, and Section 3 describes the network architecture and implementation details for the proposed method. Section 4 and 5 present experimental results and discussion, and Section 6 summarizes and concludes the paper.

2. Materials and process

2.1. Study areas

We chose four study areas across China from north to south, spanning diverse terrain landscapes, climatic regions and agriculture systems with different crops and crop phonological characteristics, covering almost one tenth of China in total (more than 940,000 km2). The general characteristics for each study area were as follows.

- Study area 1: Heilongjiang Province is situated in the north-eastern part of China, covering a total area of 473,000 km^2 (Table 1). It is China's largest commodity grain base with flat and wide terrain, conducive to large-scale crop plantings. The area has cold temperate and temperate continental monsoon climate with full sunshine, abundant rainfall, and fertile land, making it suitable for various crops, including spring wheat, spring maize, rice, and soybean. Spring wheat is sown in mid-April and harvested in mid-August; whereas spring maize, rice, and soybean are sown from the end of

Table 1
Characteristics of the four study areas: coverage, climate and main crops.

Study area	Coverage (km2)	Climate	Main crops
Heilongjiang	473,000	Temperate continental monsoon and cold temperate	Spring wheat, spring maize, rice and soybean
Hebei	188,800	Temperate continental monsoon	Winter wheat and summer crops such as maize, rice, soybean as well as cotton
Zhejiang	101,800	Subtropical monsoon	Winter wheat and summer crops such as rice, soybean, maize along with cotton
Guangdong	179,700	Subtropical monsoon	Early rice, late rice, peanut and sugarcane

April to early May and harvested in early October.

- Study area 2: Hebei Province is situated in the North China Plain, the most important agricultural production base in China and it covers 188,800 km^2 (Table 1). The landscape is complex, including plains, plateaus, mountains, hills, basins. It has temperate continental monsoon climate with approximately 484.5 mm annual precipitation concentrated in summer. The typical crop calendar starts from early October to middle or late June of the next year for winter crops such as winter wheat, and from April to October for spring and summer crops including maize, rice, soybean, and cotton.
- Study area 3: Zhejiang province is located on the southeast coast of China with covering 101,800 km^2 (Table 1). Its territory is mostly hilly and mountainous with consequentially fragmented agricultural landscape with humid, sub-tropical monsoon climate. Main crop types include winter wheat, rice, soybean, maize, and cotton. Winter crops are planted from October and harvested in May, whereas summer crops are sown from the end of March and harvested in October at the latest.
- Study area 4: Guangdong province is situated in the southernmost part of mainland China with covering 179,700 km^2 (Table 1). Terrain is largely mountainous, with only approximately 23.7% flatlands. The climatic zone is subtropical with annual rainfall of approximately 1777-mm concentrated from April to September. Major crops include early and late rice, peanut, and sugarcane. Early rice is sown from early February with harvest starting in mid-August. Late rice is sown in early July and harvested in November, peanut calendar is from mid-March to late July, and sugarcane from mid-May to late November.

2.2. Remote sensing data collection and pre-processing

We collected available high quality VHSR satellite images (GF1, GF2, and ZY3) with slight cloud coverage (less than 10%) to cover the whole study areas from the Land Satellite Remote Sensing Application Center (LASAC), Ministry of Natural Resources of the People's Republic of China. The data source was mostly GF1 images, supplemented by GF2 and ZY3 images. GF1, GF2 and ZY3 satellites were launched by China National Space Administration on 26 April 2013, 19 August 2014, and 9 January 2012, respectively. GF-1 provides two panchromatic and multispectral (PMS) and four wide field of view (WFV) multispectral cameras. We chose PMS images for mapping cropland due to their high spatial resolution and rich spectral information. GF-2 includes two PMS cameras, providing supplementary data whereas GF1 images do not overlap. ZY3 includes three line array panchromatic cameras (TLC) and a multi-spectral (MS) camera, where the TLC cameras include a nadir (NAD) and two oblique cameras viewing forward (FWD) and backward (BWD) with fixed ± 22° inclinations. We selected MS and NAD images to define accurate cropland footprints. Table 2 shows detailed specifications for the GF1, GF2 and ZY3 satellites.

For pre-processing, the selected remote sensed images were first orthorectificated and projected onto Albers conical equal area projection. MS images were registered to corresponding PAN images using polynomial warping with automatically generated tie points, providing registration error < 1 pixel. We normalized the images by re-quantifying band responses to 8 bits to compensate for systematic biases between the various sensors. Images were then processed as composites with red, green, and blue bands at very high resolution through fusing MS images and the corresponding PAN (NAD) images based on the Gram Schmidt transformation (Chavez et al., 1991). Finally, the composites were resampled to uniform 2 m spatial resolution.

Table 3 summarizes the GF1, GF2, and ZY3 composites employed. We used almost 1000 remote sensing composites from 2016 to generate training samples and more than 2000 target composites (mainly from 2017) for cropland mapping. Considering the burden of annotating

Table 2
Specifications for GF1, GF2 and ZY3 satellites.

	GF1	GF2	ZY3
Launch time	26th April 2013	19th August 2014	9th January 2012
Orbit altitude (km)	645	631	506
Orbit type	Sun-synchronous	Sun-synchronous	Sun-synchronous
Global observation cycle	41 days	69 days	59 days
Repeat observation cycle	PMS: 4 days WFV: 2 days	5 days	5 days
Swath width (km)	PMS: 60 WFV: 800	45	PAN-FWD: 52 PAN-BWD: 52 PAN-NAD: 51 MS: 51
Spatial resolution (m)	PMS-PAN: 2 PMS-MS: 8 WFV: 16	PAN: 1 MS: 4	PAN-FWD: 3.5 PAN-BWD: 3.5 PAN-NAD: 2.1 MS: 5.8
Wavelength (nm)	PAN: 450–900 Blue: 450–520 Green: 520–590 Red: 630–690 Infrared: 770–890	PAN: 450–900 Blue: 450–520 Green: 520–590 Red: 630–690 Infrared: 770–890	PAN: 500–800 Blue: 450–520 Green: 520–590 Red: 630–690 Infrared: 770–890

(PMS = panchromatic and multispectral, WFV = wide field of view, PAN = panchromatic bands, MS = multispectral bands, FWD = forward, BWD = backward, and NAD = nadir).

Table 3
The number of scenes of the GF1, GF2 and ZY3 composites used in this study. Adopted composites of training data are from 2016 while those of target data are mostly from 2017.

Study area	Training data				Target data			
	GF1	GF2	ZY3	Total	GF1	GF2	ZY3	Total
Heilongjiang	292	73	25	390	365	132	277	774
Hebei	235	0	0	235	735	0	0	735
Zhejiang	125	24	0	149	123	157	0	280
Guangdong	137	57	22	216	227	101	45	373

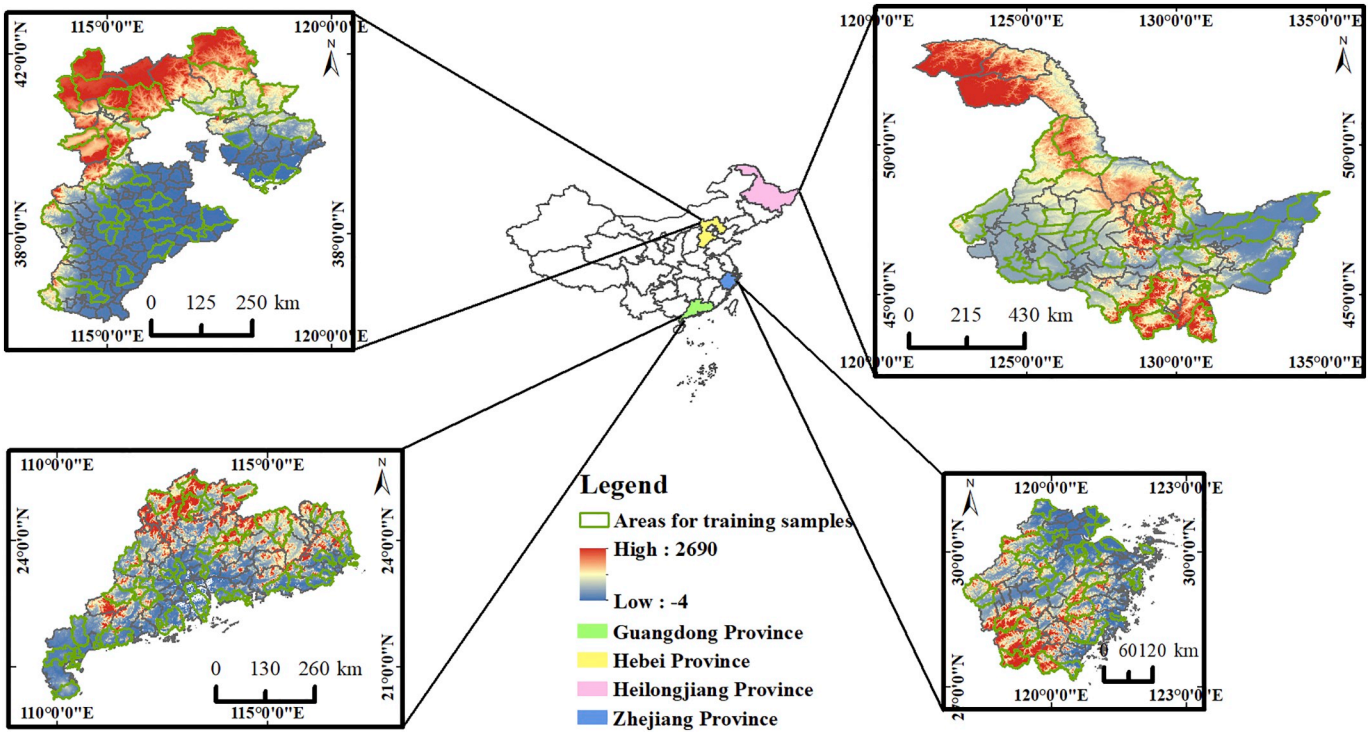

Fig. 1. Digital Elevation Model (DEM) Maps for four study areas overlaid with county boundaries. Also displayed are the selected areas for establishing cropland layer datasets and subsequently collecting training samples (Green lines). The select of these areas was based on the criteria to ensure representation across agricultural zones with different landscape characteristics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

label masks, we determined typical zones for each study area, comprising approximately 1/3 the entire region, to form the training set, as shown in Fig. 1, distributing the selected area locations across the study areas to ensure good representation for agricultural zones with different landscape characteristics. Fig. 2 indicates the temporal distribution of the data. Training data from each study area included all seasons to ensure representativeness, and acquisition time of the target data also covered almost the whole of 2017. Notably the available target composites from 2017 for Hebei, Zhejiang, and Guangdong provinces did not cover the entire area, hence, we incorporated composites of 2016 (distinct from those used in the training set) as complementary data.

2.3. Cropland layer dataset

A large number of training samples were required to construct deep

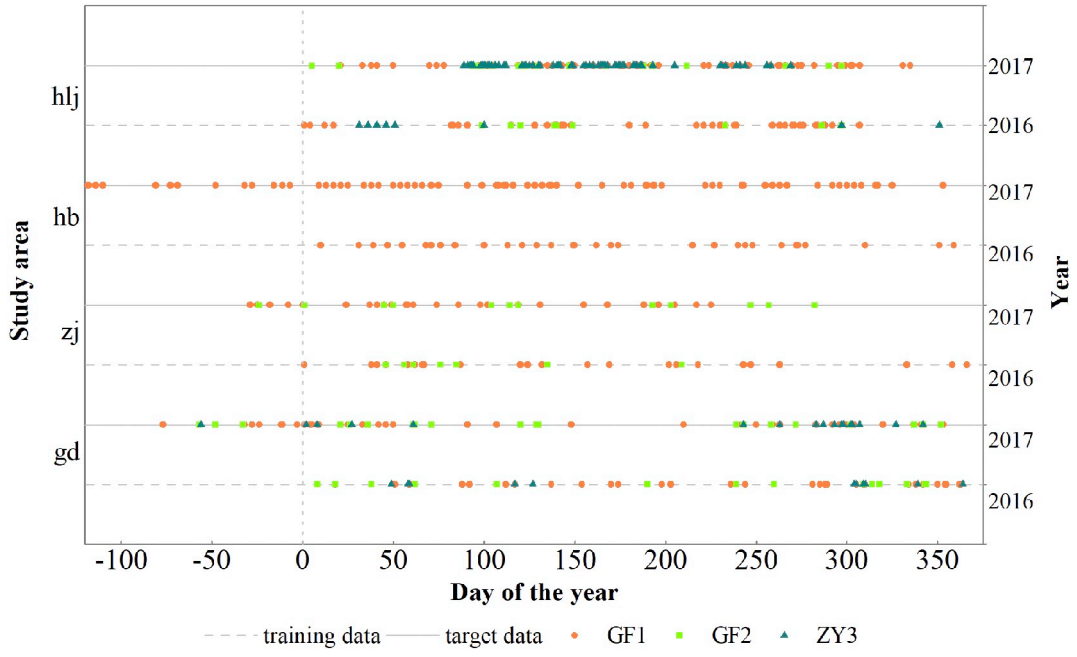

Fig. 2. Temporal distributions of adopted VHSR composites for Heilongjiang (hlj), Hebei (hb), Zhejiang (zj) and Guangdong (gd) study areas.

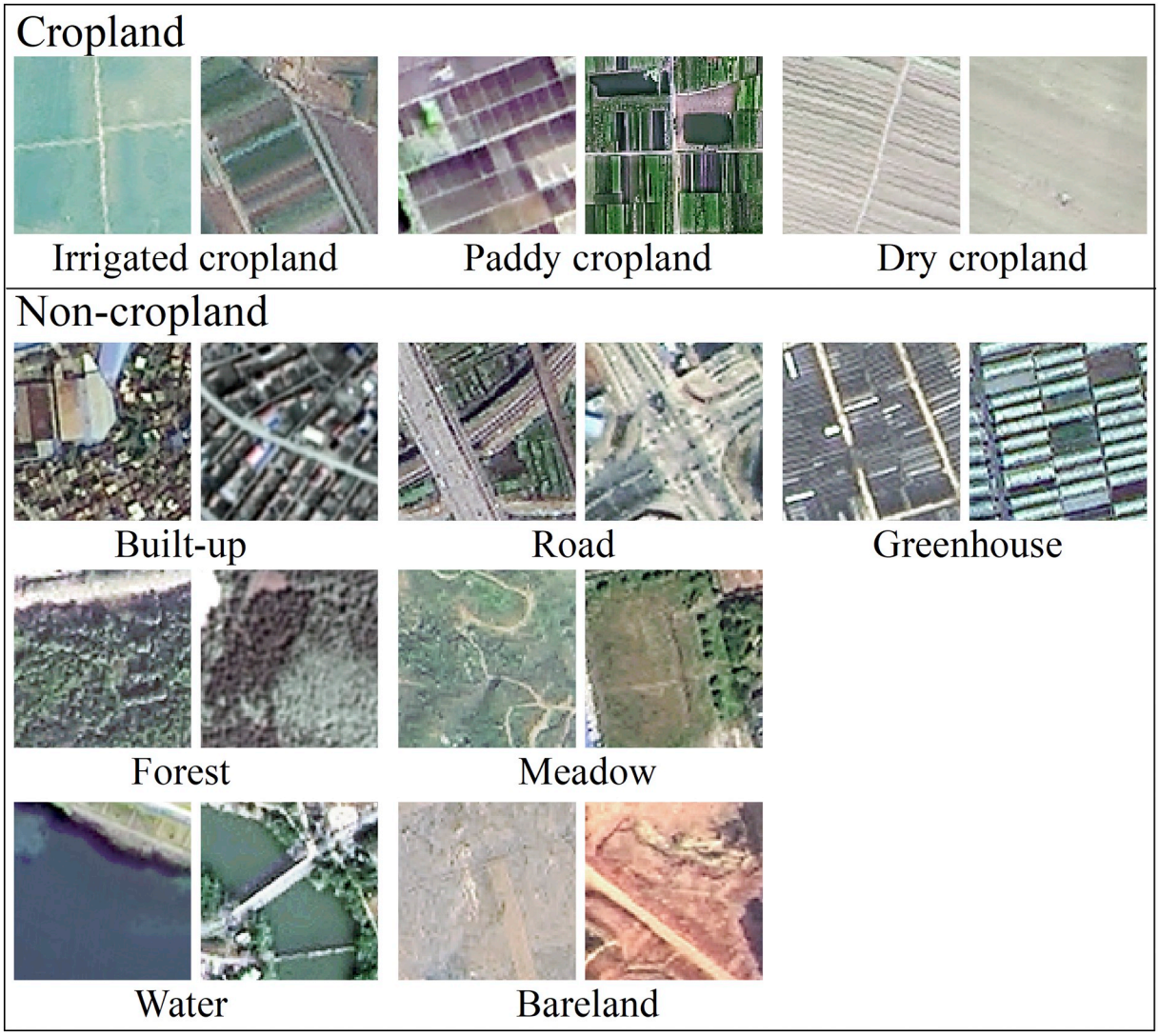

Fig. 3. Representative exemplars of cropland and non-cropland.

learning algorithms for extracting cropland cover from VHSR images. Cropland layer datasets for the four regions were established through visual inspection of composites with 2 m resolution from 2016. Fig. 3 demonstrates the typical examples of the cropland and non-cropland. GF1, GF2, and ZY3 composites were employed from Heilongjiang and Guangdong provinces to construct cropland datasets, whereas GF1 and GF2 composites were employed for Zhejiang, and GF1 composites for Hebei provinces, as shown in Table 3. We identified cropland objects by manually delineating field boundaries from the VHRS images, producing 539,459; 437,397; 594,608; and 256,088 cropland parcels with digitized cropland boundaries for Heilongjiang, Hebei, Zhejiang, and Guangdong provinces, respectively, to provide the data foundation for collecting training samples. We used in situ identification data collected by the National Bureau of Statistics of China to compare with the derived cropland datasets, and accuracy for each study area exceeded 95%.

3. Methodology

In a large-scale setting, the traditional classification approaches for cropland extraction require repeated manual annotation, model training and cropland classification every time the geo-locations, the sensor characteristics or the imaging conditions change, which are time-consuming and inefficient. In the proposed framework, we exploited CNNs with strong generalization ability to extract cropland from VHSR satellite imagery without the above-mentioned repeated operations. It includes the following steps.

1. Collect training samples for all seasons from cropland datasets (Section 3.2);
2. Train the model for cropland classification through feeding the training samples into the proposed MPSPNet to learn representative and discriminative feature representations; and

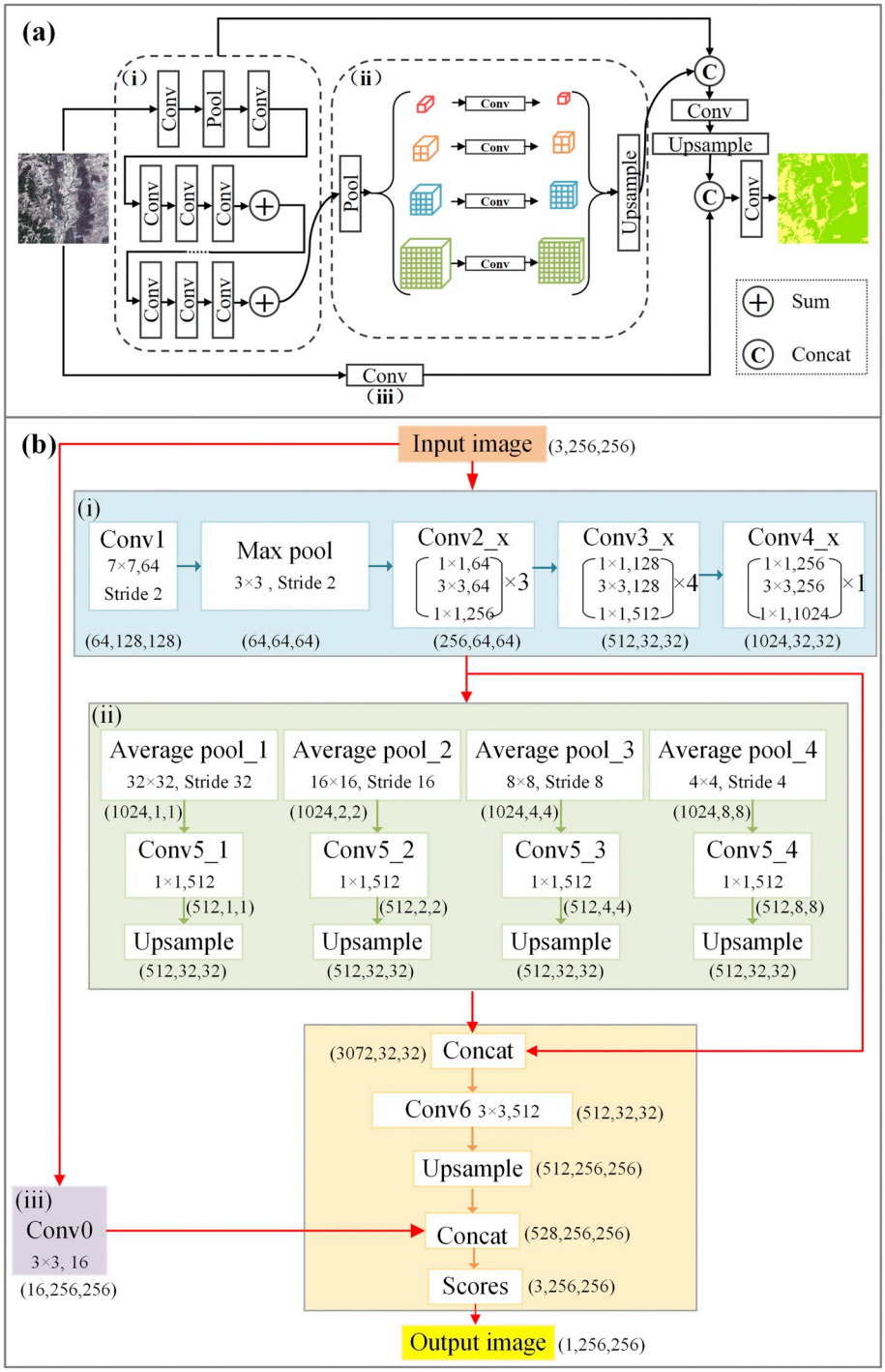

Fig. 4. The proposed MPSPNet architecture, which added an extra skip connection with low-level features (iii) to the original PSPNet. (i) extracts deep and abstract feature maps for an input image using a ResNet with the dilated convolution; (ii) is a spatial pyramid pooling module to capture multiscale feature representations, and aggregate local and global context information into high-level feature maps; (iii) obtains low-level features, which are combined with the high-level features for final precise pixel-wise classification. (a) is the Conceptual illustration of MPSPNet. (b) is the detailed design of MPSPNet. Feature maps with a three-dimensional structure per layer: dimensions are written in brackets, where the first number indicates the amount of feature maps channels, second and third represent spatial dimensions.

3. Classify target images to generate complete cropland layers using the trained model.

3.1. Proposed MPSPNet architecture

The proposed MPSPNet is an adaption of PSPNet architecture, integrating some extra shallow level information with edge features into the high-level features to preserve fine local details. Fig. 4 shows the proposed network for pixel-based classification of VHSR remote sensing images, incorporating three core structures: ResNet (He et al., 2015) with dilated convolution (Yu and Koltun, 2016), spatial pyramid average pooling module, and fusion of low-level and high-level information. ResNet with dilated convolution strategy is applied to extract deep and abstract feature maps for the input image, where eight residual blocks are employed and the dilation factors of the 3 × 3 dilated convolutions in the last residual block is set to 2. A spatial pyramid average pooling module is subsequently followed to harvest multiscale feature representations, aggregating local and global context information into the final deep feature maps. Finally, fusion of low-level and high-level information is exploited to improve object localization accuracy for precise pixel-based classification. To avoid low-level features overwhelming final feature importance, we chose high-level feature map resolution to be 16-fold larger than low level feature maps, i.e., 512 vs 16.

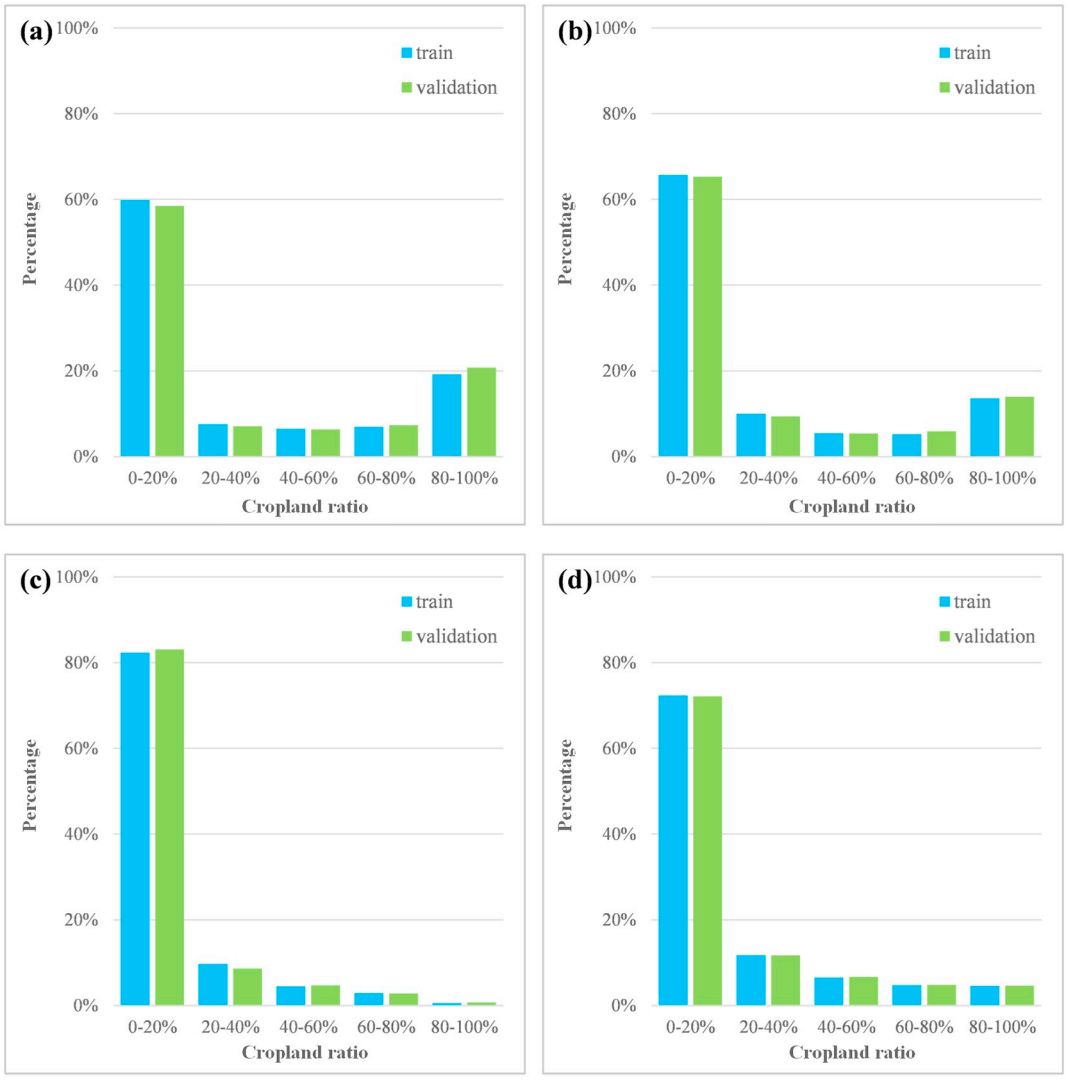

Fig. 5. Class distributions in training and validation set for (a) Heilongjiang, (b) Hebei, (c) Zhejiang and (d) Guangdong. The x-axis (cropland ratio) represents the proportion of cropland in each sample block with a spatial dimension of 256 × 256 pixels while y-axis (percentage) represents the proportion of the sample blocks with certain cropland ratio in training or validation set. Validation and training set have similar class distributions for each study area.

3.2. Collection of the training samples

We generated the training sample set using the cropland dataset and corresponding remote images from 2016. The training dataset should contain as many samples as possible from as many seasons as possible to represent different cropland status types, such as bare land resulting from previous harvest or soil preparation, low vegetation cover during early crop growth, high vegetation cover during flourishing crop growth, etc. We then cropped all samples into 256 × 256 pixel images with 25% overlap. Thus, we produced 969,922; 397,146; 203,152; and 345,718 sample blocks for Heilongjiang, Hebei, Zhejiang, and Guangdong study areas, respectively.

3.3. Model training

A validation set is essential to prevent overfitting when training the network architecture. Therefore, we randomly took 20% of the training sample blocks as the validation set. As shown in Fig. 5, the class distributions in training and validation set are similar for each study area. We used the Adam (Kingma and Ba, 2015) algorithm as for gradient descent optimization with batch size = 2, initial learning rate $\alpha = 0.0001$, and total number of epochs = 10 to fully learn generic feature representations through backpropagation. Learning rate decreased iteratively based on a declining factor gamma = 0.1 after every three epochs. We also set weight decay as 0.00001 for avoiding overfitting and momentum as 0.9 to regularize learning. We determined above-mentioned hyper parameters by referring to Kingma and Ba (2015). To accelerate training, we initialized network weights from a pre-trained model on ImageNet, which learned low level image features, such as edges and corners in early network layers. Weights corresponding to the best validation loss were used for cropland mapping of 2017. It should be noted that different models were trained for each

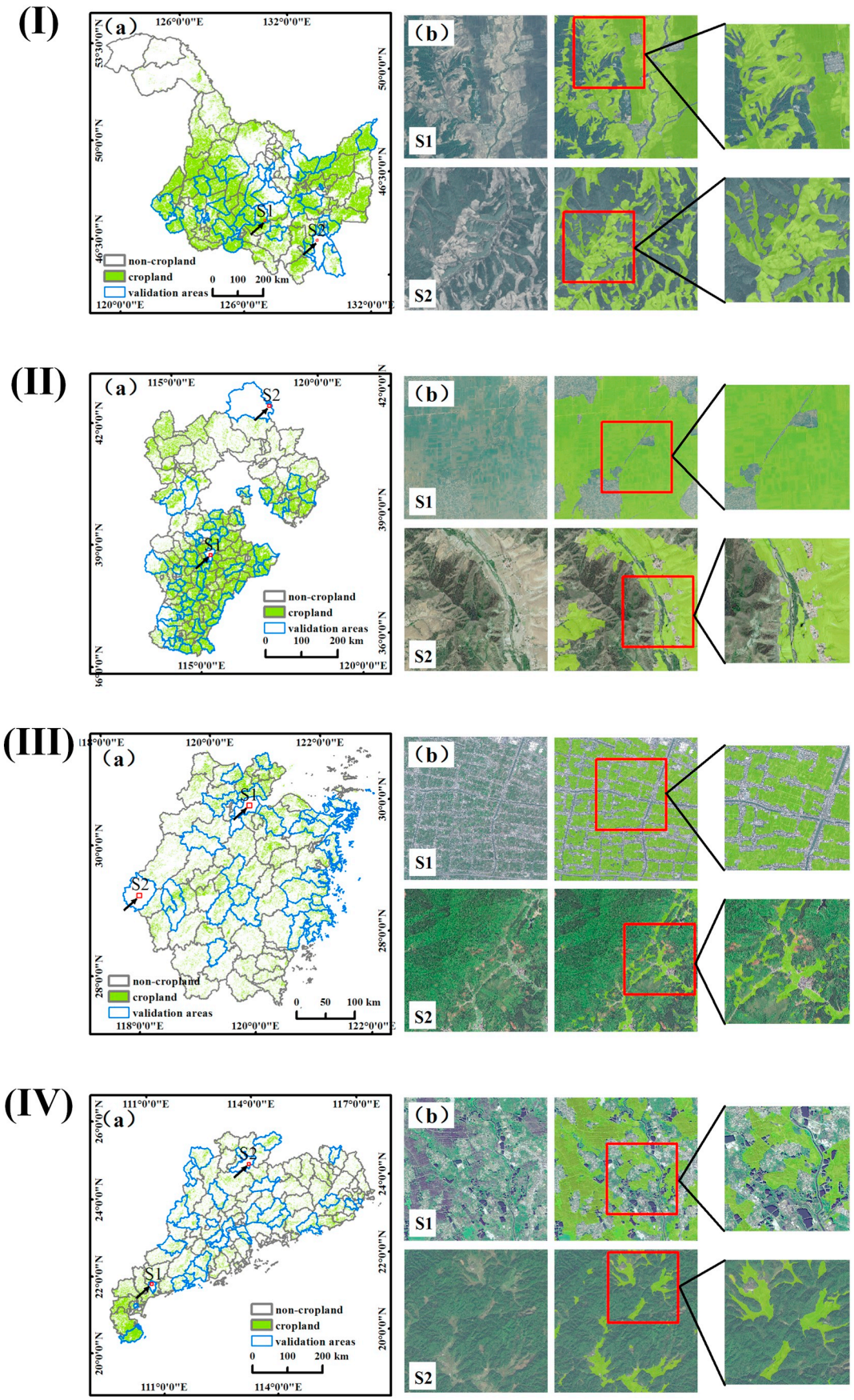

Fig. 6. Cropland mapping results for (I) Heilongjiang, (II) Hebei, (III) Zhejiang and (IV) Guangdong study areas in 2017: (a) cropland maps overlaid with validation areas. (b) 4× 4 km typical subsets marked in each study area as S1 and S2, depicting examples of mapping results.

Table 4
Accuracy indices of the classifications: Overall Accuracy (OA), kappa, Producers' accuracy (PA) and Users' Accuracy (UA). The overall accuracies range from 89.99% to 92.31% and the kappa exceeded 0.7 for the four study areas. (HLJ) Heilongjiang, (HB) Hebei, (ZJ) Zhejiang and (GD) Guangdong.

	OA (%)	Kappa	Cropland		Non-cropland	
			PA (%)	UA (%)	PA (%)	UA (%)
HLJ	92.31	0.844	92.70	93.56	91.83	90.75
HB	91.00	0.820	90.63	91.15	91.36	90.86
ZJ	89.99	0.700	79.74	73.25	92.60	94.73
GD	90.56	0.744	77.35	84.12	95.04	92.52

study area, since each study area presents different cropland characteristics. Fig. 14 provided an example of the loss evolution during training process for Guangdong.

3.4. Accuracy assessment

To evaluate the proposed method, we selected several typical zones with various landscape characteristics to generate validation samples in every study area, taken from an in-situ data survey by the National Bureau of Statistics of China. These in situ data sets were obtained through field survey in 2017 by different site managers as vector files, incorporating investigated object geometric shapes. In total, validation dataset contained 4056 cropland patches and 7675 non-cropland patches (comprising built-up, forest, meadow, water, and bare land) for each study area. We adopted the Kappa coefficient; and overall (OA), producer and user accuracy evaluation metrics to access classification performance.

4. Results

4.1. Cropland mapping results

Fig. 6 shows the final cropland maps for the four study areas, with several typical subsets for different landscapes in study areas from plain areas and mountain zones shown in Fig. 6(b) to show more details, displaying the original images and results. The proposed algorithm generally achieved good visual results in both northern areas with large average field size, and southern areas with fragmented and heterogeneous landscapes. More importantly, these areas could be effectively characterized regardless of whether or not plants were growing in the cropland areas. Large cropland areas were distinguished from others, such as built-up, rivers, roads and forests for plain areas (s1 in Heilongjiang and Hebei provinces), and boundaries for fragmented cropland with regular distribution can be precisely located for Zhejiang province. Moreover the irregularly distributed cropland areas were extracted and forest with similar spectral features to cropland was well discriminated (Guangdong, s1). In mountain areas, small villages and large forest areas could be filtered out (area s2 in Heilongjiang, Zhejiang, and Guangdong provinces). For Hebei, cropland without plants (fallow) was effectively detected, whereas bare land on the hill was suppressed. The proposed MPSPNet successfully identified croplands with various texture and spectral characteristics, whereas traditional methods tended to require dividing the croplands into different types that were subsequently merged into one category. Thus, the proposed approach would be suitable to map VHSR cropland areas over large areas effectively and automatically.

4.2. Statistical accuracy assessment

Table 4 summarizes quantitative accuracy assessment for the proposed approach. The MPSPNet performed well for all four provinces. Heilongjiang and Hebei (Northern China) achieved better performance due to larger average field size. Heilongjiang Province achieved highest accuracy (OA = 92.31%), but all the other provinces also had good performances (89.99% to 91.00%). Kappa exceeded 0.7 for all study areas, reaching 0.844 for Heilongjiang, indicating high cropland identification accuracy for the proposed MPSPNet. The proposed method accuracy was worst for the Zhejiang study area (OA = 89.99%, Kappa = 0.700), but still sufficient for fragmented landscape area.

We divided validation zones into mountainous, plain, and plain and mountain mixed areas according to terrain and the quantitative assessment results of each validation zone are further analyzed with main evaluation metrics shown in Fig. 7. Fig. 7(a) is choropleth maps of accuracies for each validation zone, and Fig. 7(b) is boxplots of main evaluation metrics for different landscape types. The detailed accuracies of each validation zone are provided in Table A1 in the appendix. Overall, similar to the above outcomes, Heilongjiang and Hebei provinces achieved have higher accuracy than Guangdong and Zhejiang provinces, and additionally accuracy for plains areas was higher than for mountain areas for all study areas.

More concretely, Average OA and kappa for plain areas were slightly higher than for mountainous areas and mixed areas in Heilongjiang Province. Average omission error was least for mixed areas, followed by plain and mountainous areas successively. And average commission error for plain area is lower than that for the other two types of areas. For all zones in plain areas, 65% of them have an overall accuracy of higher than 90% while the ratio is 30% in mountain areas and 66.7% in mixed areas. Kappa values in plain areas were maintained around 0.760 ± 0.11, with higher variability across all zones in mountain and mixed areas (0.680 and 0.630 minimum and 0.821 and 0.842 maximum, respectively). Producer accuracies in mixed areas showed little fluctuation, ranging 88.79% to 94.82%, whereas large differences occur for other terrain types, ranging 76.52% to 96.54% in plain areas, and 76.17% to 95.87% in mountain areas. User accuracies of all validation zones in plain areas exceed 90% except Daoli (86.51%) and Yanshou (87.78%). User accuracies in mixed areas exceeded 90% (90.46% to 95.32%) and were slightly lower for mountain areas (82.67% to 94.40%).

Cropland map of Hebei Province exhibited satisfactory performance similar to Heilongjiang Province, with mean evaluation metrics for plain areas higher than for the other two area types. Mean overall accuracy, kappa, and producer accuracy for mountain areas were extremely close to those for mixed areas, whereas average user accuracy was 8.38% lower. In plain areas, 64% of validation zones had overall accuracy exceeding 90% (maximum = 96.96%). And kappa exhibited relatively high variability (0.622–0.935), with more than 70% of validation zones producer accuracy and almost 90% user accuracy exceeding 90%. Overall accuracies for the two validation zones in mountain areas were 88.97% and 91.16%, respectively, but user accuracies were low, revealing commission errors. Overall accuracies for mixed areas ranged from 85.52% to 94.55% with 40% of zones exceeding 90%. Producer and user accuracies were well balanced (73.4% to 92.2% and 72.41% to 88.27%, respectively).

Despite fragmented landscape and small parcel size distribution, Zhejiang and Guangdong results achieved 89.99% and 90.56% overall accuracies, respectively. Overall average accuracies in plain areas for Zhejiang and Guangdong were 88.02% (± 4%) and 90.62% (± 6%), respectively. Approximately 50% of Guangdong validation zones

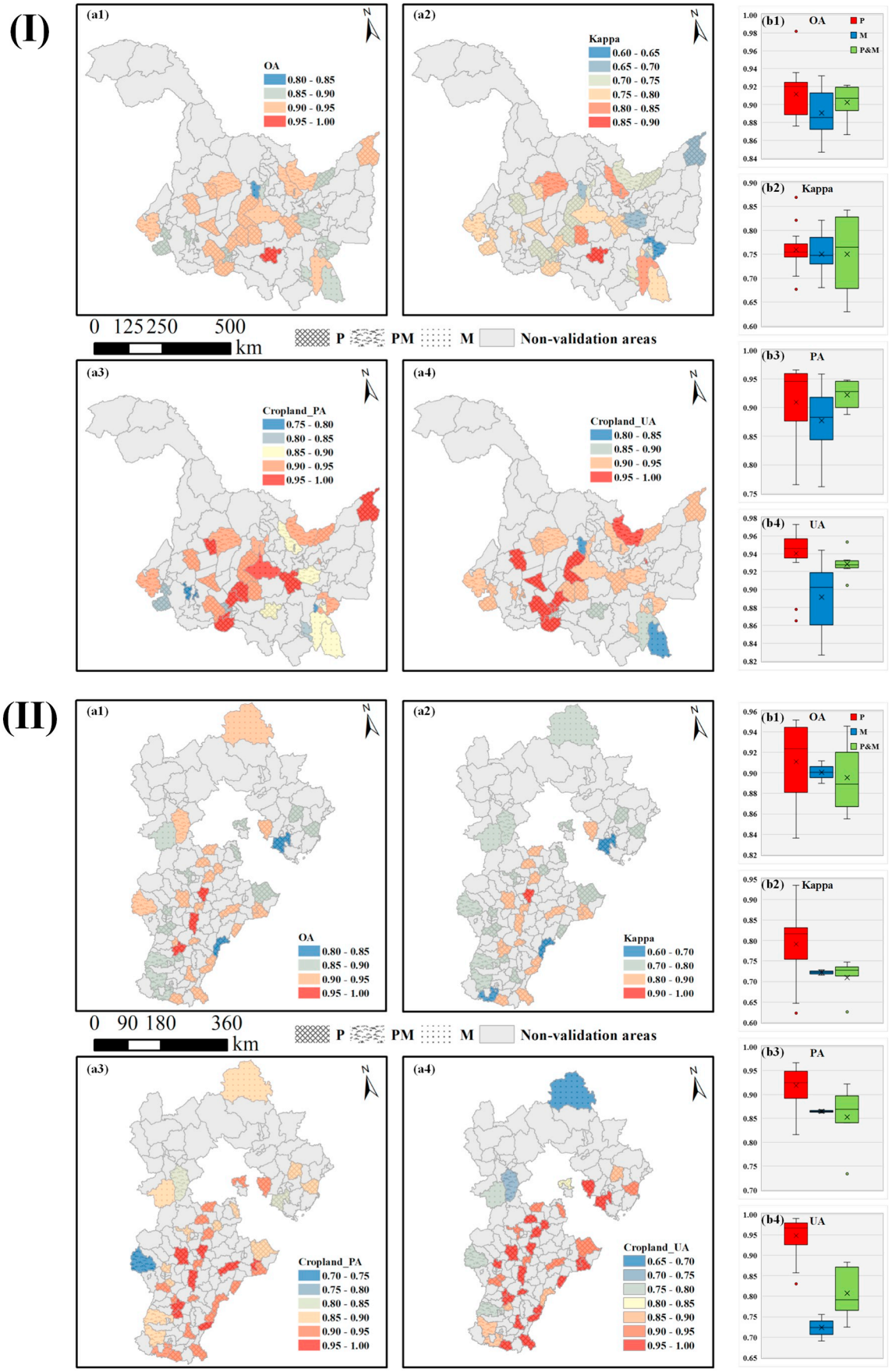

Fig. 7. Statistical accuracy assessment results for (I) Heilongjiang, (II) Hebei, (III) Zhejiang and (IV) Guangdong in 2017. The group (a) is choropleth maps of main evaluation metrics for each validation zone. The group (b) is boxplots of main evaluation metrics in different terrain landscape, where the mark "×" presents the location of mean value. Plain areas (P), Mountain zones (M), Plain and mountain mixed zones (PM).

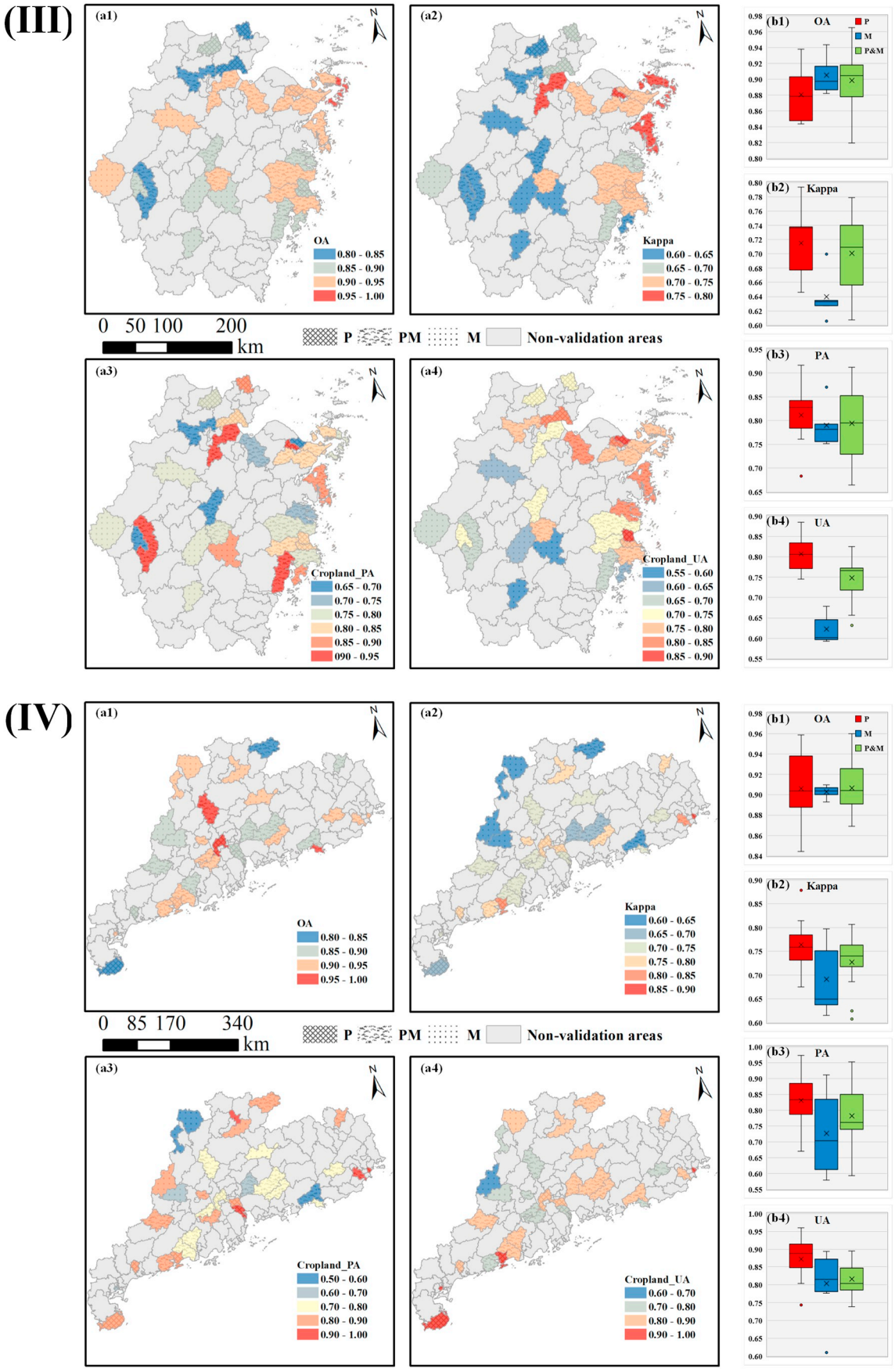

Fig. 7. (*continued*)

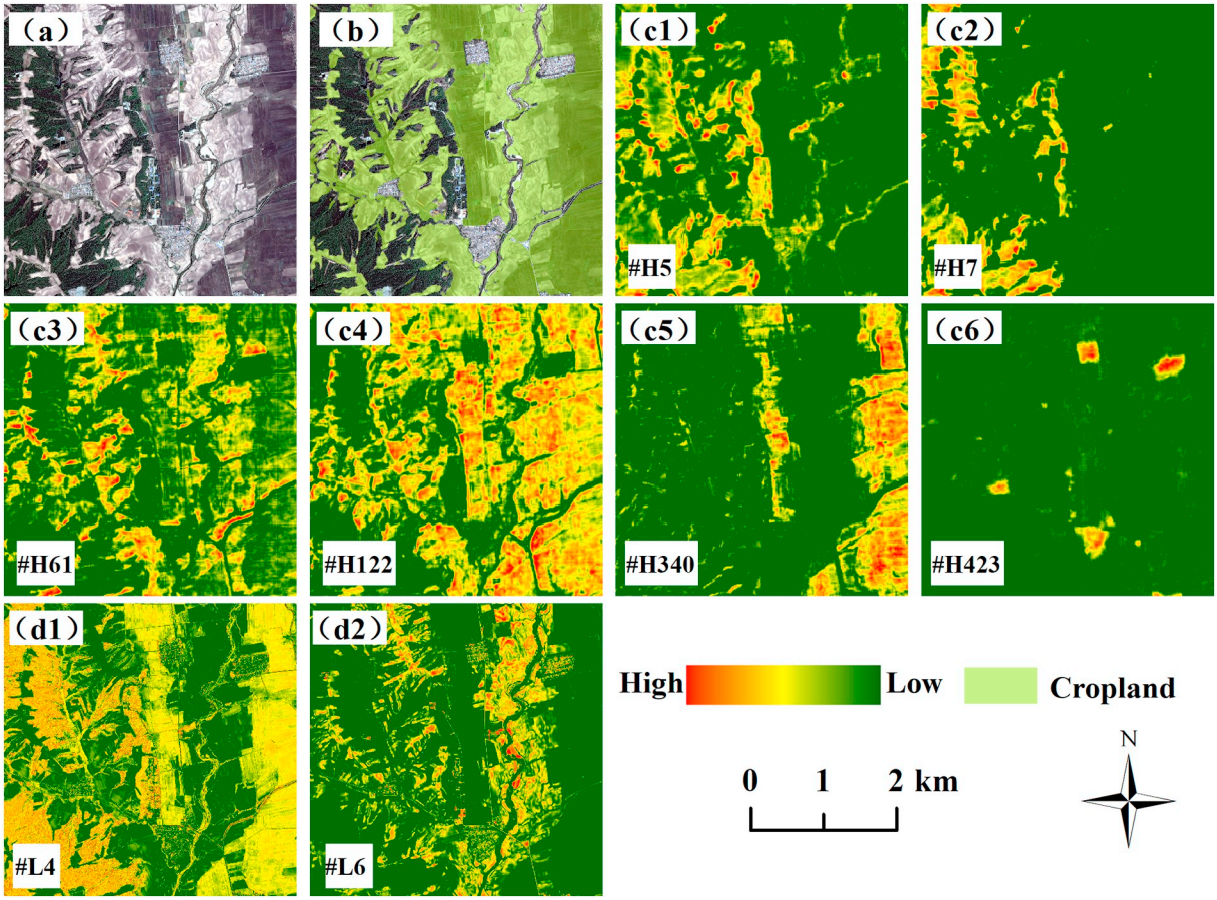

Fig. 8. Learned high-level and low-level feature maps: (a) original image, (b) classification result, (c1)-(c6) 6 of 512 deep high-level feature maps, (d1) and (d2) 2 of 16 low-level feature maps.

exhibited overall accuracy exceeding 90%, whereas only two out of seven (Zhenhai and Jiangbei) in Zhejiang achieved this (91.12% and 93.8%, respectively). All plain validation zones in Guangdong exhibited kappa between 0.700 and 0.900 except Xuwen (0.675), whereas Zhejiang plain zones exhibited kappa between 0.646 and 0.794. Average producer accuracy for Guangdong was 83.05% (± 16%) with Longhu and Xiashan exhibiting highest and lowest producer accuracy (97.29% and 67.07%), respectively. Average produce accuracy for plain areas in Zhejiang was 81.14% (± 13%) with Jiangbei and Zhenhai zones exhibiting highest and lowest accuracy (91.64% and 68.33%), respectively. In contrast to producer accuracies, user accuracies for plain areas were relatively smaller for Guangdong and Zhejiang: mean user accuracies were 87.22% (± 10%) and 80.69% (± 6%), respectively. In mountain areas, Guangdong showed better performance than Zhejiang, with mean kappa 0.691(0.616–0.797) and 0.640 (0.605–0.700), respectively. For mixed terrain zones, approximately 60% of validation zones in Zhejiang and Guangdong had overall accuracy exceeding 90%. In conclusion, the proposed method exhibited strong ability to describe cropland location and extent with high accuracy.

5. Discussion

5.1. Feature representations for croplands

We examined shallow and deep feature representations to better understand how the proposed MPSPNet model extracted cropland, using Heilongjiang S1 (see Fig. 6) as an example. Two of the 16 learned low level features and six of 512 high-level features were selected to depict two-scale characteristics. Fig. 8 shows the high-level feature map of #H122, confirming this filter was responsive to cropland cover, whereas #H5 obstructs the representation of this class. Features #H61 and #H340 tend to highlight different types of cropland land cover, with #H61 corresponding to bright white cropland land whereas #H340 corresponds to dark purple areas. The proposed network not only extracted deep high level features distinguish between cropland and non-cropland, but also separated forest (#H7) and buildings (#H423). Thus, even though the model was not explicitly set to segregate such features, it learned these features to promote cropland identification. This aspect will be beneficial for transfer learning to detect other target objections, such as building, etc. In contrast to the high-level abstract features, low-level features trend to extract local detailed boundary information. For example #L4 (Fig. 8) identifies dark objects and #L6 responds more to bright objects. Thus, deep high-level features can present spatial contextual information which help infer classes whereas shallow low-level features learned edge information to characterize object boundaries. Combining these aspects contributes to improved classifications.

To further explore high- and low-level feature roles in end-to-end pixel wise classification, we took a horizontal transect from S1 of Heilongjiang. Fig. 9(b1) shows it is almost impossible to discriminate between cropland and non-cropland using only spectral features,

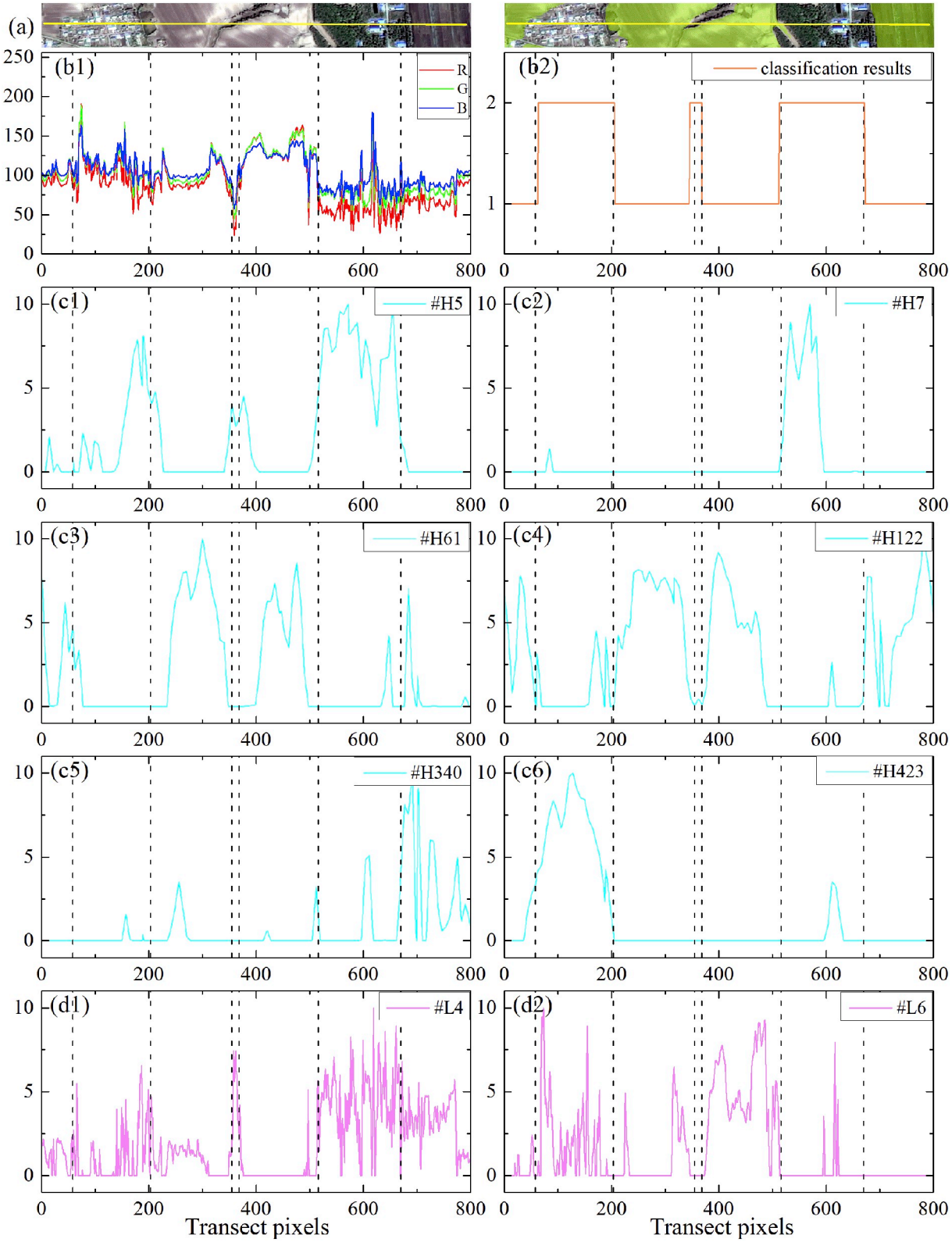

Fig. 9. Horizontal transects for S1 from Heilongjiang. Dashed lines indicate the boundary between cropland and non- cropland through visual interpretation. (a) original image, (b1) pixel values for R, G, B bands, (b2) classification results (1 = cropland, 2 = non-cropland), (c1)-(c6) pixel values of high-level features, and (d1)-(d2) low-level features. Note that data from low-level and high-level features were all normalized into the same range for visualizing.

whereas Fig. 9(b2) shows classification results from the proposed approach have strong agreement with visual results, indicating excellent performance. Similar to Fig. 8, #H122 shows active response to cropland whereas #H5 shows an opposite response. However, these high level features lose spatial precision after the convolution stack, and only poorly locate object boundaries. On the other hand, while low level features (e.g. #L4 and #L6) were relatively insensitive to abstract semantic information, they show good response to boundary information.

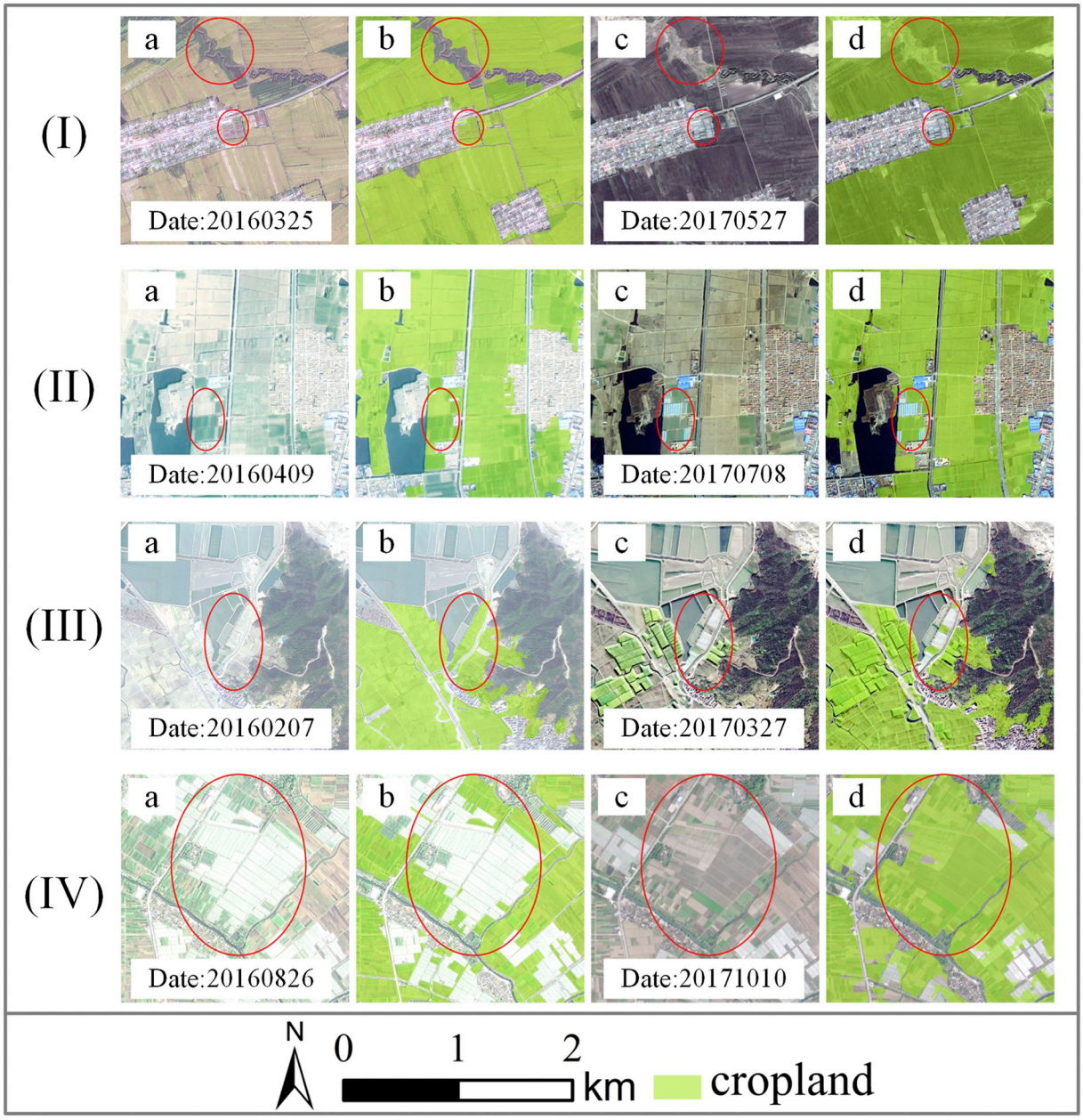

Fig. 10. 2 × 2 km zoom-in regions in (I) Heilongjiang, (II) Hebei, (III) Zhejiang and (IV) Guangdong, displaying examples of classification performance concerning temporal transfer: (a) the original image in 2016, (b) the ground reference in 2016, (c) the original image in 2017, (d) the cropland mapping result in 2017. The red circles denote the land cover type changes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Thus, combining high- and low-level features was an effective way to realize cropland type recognition and precise localization.

5.2. Assessment of spatio-temporal transferability

One fundamental issue for large area cropland mapping remains the model transferability. Spatio-temporal transferability is the model temporal and spatial generalization capability for classifying unseen data, which means applying a trained classification model derived from certain locations or time periods to other places or periods efficiently (Phalke and Özdoğan, 2018). To validate temporal transferability of the proposed method, we applied a classification model gained from training areas in 2016 to map cropland extent of the identical areas (same as training areas) in 2017 and investigated the extraction of cropland over the four study areas, especially when spectral characteristic variation or land cover type change occurs. The spatial transferability was tested over different areas (independent from training areas) with similar cropland landscape to the training areas in 2017.

We evaluated temporal transfer from the qualitative and

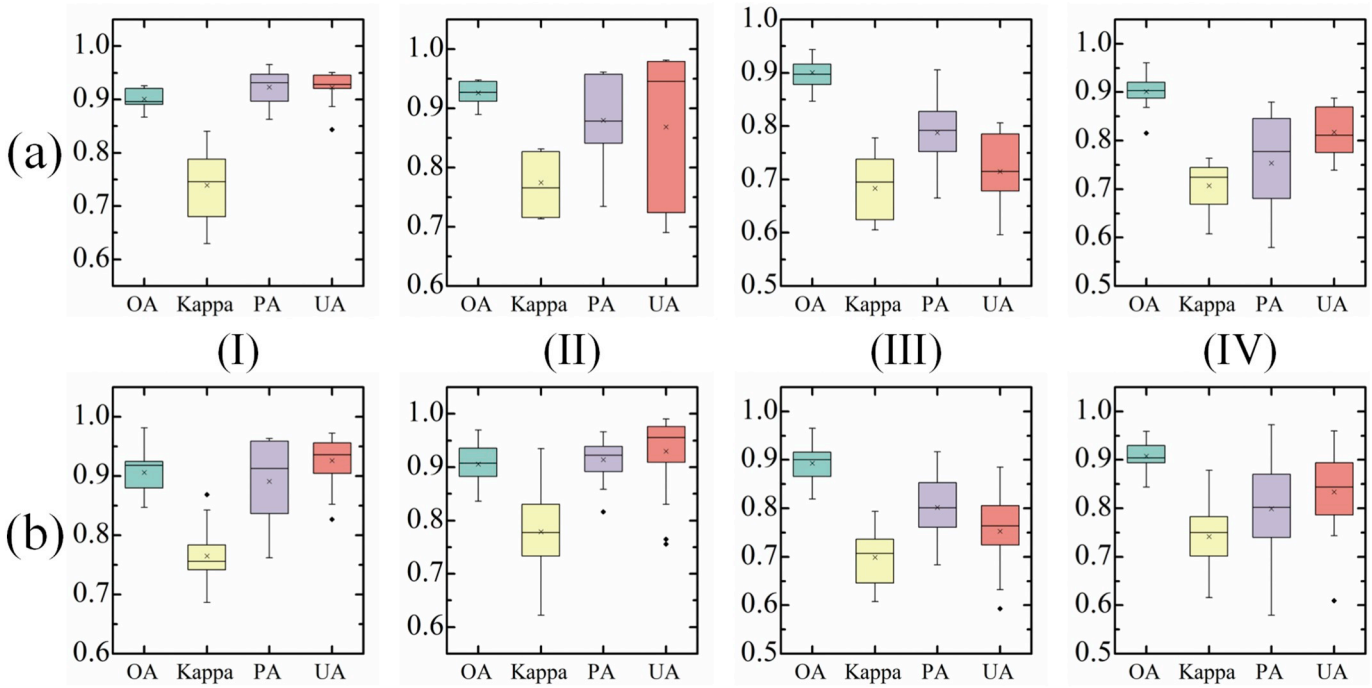

Fig. 11. The quantitative evaluation regarding temporal and spatial transferability with boxplots of four evaluation metrics over the validation zones for (I) Heilongjiang, (II) Hebei, (III) Zhejiang and (IV) Guangdong, where the mark "×" denotes the location of mean value. (a) is the result of temporal transfer and (b) represents the result of spatial transfer. Note that the validation zones in terms of temporal transfer are the ones (in Fig. 7) within the training areas (in Fig. 1) while the validation zones with regard to spatial transfer are outside the training areas.

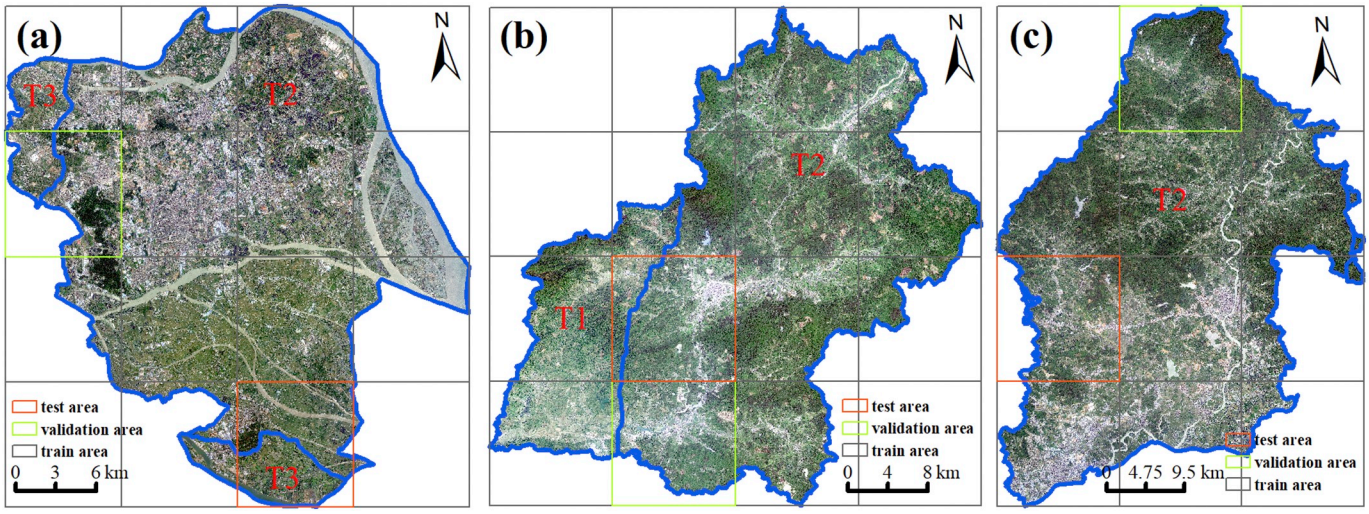

Fig. 12. Three study areas with different terrain landscapes. Each study area was split into a training, a validation, and a test set in a grid-wise random manner. (a) Panyu, plain area. (b) Fogang, Mountainous area. (c) Zengcheng, Plain and mountain mixed area. T1, T2 and T3 mean the different image acquisition times: February 20, May 16 and December 7 of the year 2016, respectively.

quantitative perspectives, respectively. As presented by Fig. 10, whether significant variations of spectral information exist or not between training images and target images, the cropland can be accurately identified. In particular, cropland can also be correctly distinguished where the land cover categories change (red circles in Fig. 10). Fig. 11(a) illustrates the boxplots of four important accuracy evaluation indicators over the validation zones (in Fig. 7) within the training areas (in Fig. 1) for every study area in a quantitative accuracy evaluation manner. Synoptic view, for Heilongjiang and Hebei, the average kappa exceeded 0.73, ranging 0.630 to 0.840 and 0.714 to 0.831, respectively. Satisfactory performances were also obtained in Zhejiang and Guangdong with relatively fragmented landscape, whose kappa values were maintained around 0.683 ± 0.1 and 0.707 ± 0.1, respectively. The trained model can be applied effectively in other periods because the proposed MPSPNet is able to yield the hierarchical and deep abstract semantic features on the basis of the training data from all seasons (as shown in Fig. 2) and the cropland can be stably described by such features without being affected by changes in spectral, texture, size and

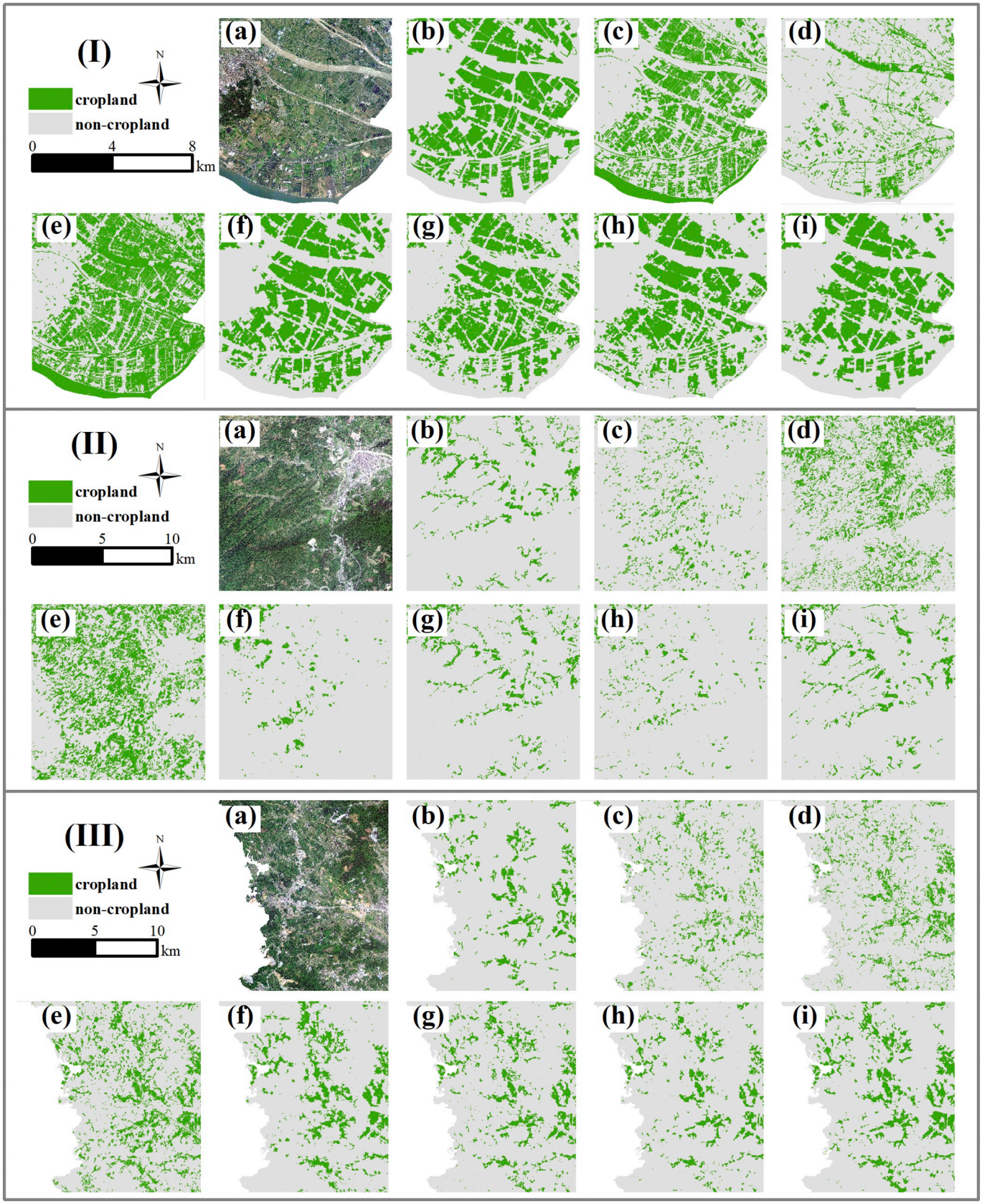

Fig. 13. Cropland mapping results for (I) Plain area, (II) Mountainous area and (III) Plain and mountain mixed area: (a) The original image, (b) Ground reference, (c) result of OBIA-RF_P, (d) result of OBIA-RF_M, (e) result of OBIA-RF_{PM}, (f) result of $MPSPNet_P$, (g) result of $MPSPNet_M$, (h) result of $MPSPNet_{PM}$, and (i) result of $MPSPNet_W$. Specifically, OBIA-RF_P or $MPSPNet_P$ represents the model trained on the training set from plain area, OBIA-RF_M or $MPSPNet_M$ represents the model from mountainous area, OBIA-RF_{PM} or $MPSPNet_{PM}$ represents the model from mixed area, and $MPSPNet_W$ represents the model used for Guangdong in Section 4.

Table 5
Accuracy comparison of different classification models under different terrain landscapes.

Terrain types	Evaluation Metrics	OBIA-RF$_P$	OBIA-RF$_M$	OBIA-RF$_{PM}$	MPSPNet$_P$	MPSPNet$_M$	MPSPNet$_{PM}$	MPSPNet$_W$
P	OA(%)	72.15	52.19	64.77	**90.48**	79.90	80.83	85.27
	kappa	0.433	−0.045	0.308	**0.807**	0.584	0.603	0.695
	PA(%)	67.24	14.55	74.88	**90.30**	69.74	70.19	74.92
	UA(%)	68.33	37.48	57.33	**88.14**	81.48	83.29	89.56
M	OA(%)	83.70	79.32	67.42	89.47	**95.42**	91.79	94.51
	kappa	−0.009	0.279	0.135	0.246	0.672	0.369	**0.679**
	PA(%)	7.44	66.17	61.70	23.73	68.16	29.79	**70.85**
	UA(%)	8.52	26.35	16.76	40.31	**72.31**	64.24	71.01
PM	OA(%)	88.21	89.15	84.64	89.11	93.49	**95.11**	92.57
	kappa	0.358	0.479	0.432	0.528	0.693	**0.747**	0.673
	PA(%)	41.54	61.01	77.05	74.44	83.76	80.64	**88.95**
	UA(%)	43.23	48.42	38.28	48.55	64.50	**74.53**	59.65

Plain areas (P), Mountain zones (M), Plain and mountain mixed zones (PM). The bold font highlights the greatest classification accuracy per row.

shape characteristics. Admittedly, these qualitative and quantitative assessments demonstrated the robust temporal transferability of the proposed method in cropland mapping applications.

Similarly, qualitative and quantitative aspects were both chosen as evaluation measurement to investigate the spatial transferability. The typical subsets in Fig. 6(b) (such as s2 in Heilongjiang, s1 in Hebei and Zhejiang as well as s1 and s2 in Guangdong) revealed the proposed method performed well in delineating spatial distributions of various croplands. For quantitative assessment, Fig. 11(b) exhibits the boxplots of primary evaluation indicators over the validation zones (in Fig. 7) outside the training areas (in Fig. 1). Generally, the four study areas implemented relatively high accuracies and their average kappa values

Table 6
Accuracy assessment amongst MPSPNet-Resnet, PSPNet-Resnet, MPSPNet-GoogLeNet and PSPNet- GoogLeNet for Guangdong.

Model	OA(%)	Kappa	PA	UA
MPSPNet-Resnet	**90.56**	0.744	**77.35**	84.12
MPSPNet- GoogLeNet	90.49	**0.746**	75.59	**86.34**
PSPNet-Resnet	89.01	0.717	76.70	84.01
PSPNet- GoogLeNet	88.78	0.712	74.67	84.94

The bold font highlights the greatest classification accuracy per column.

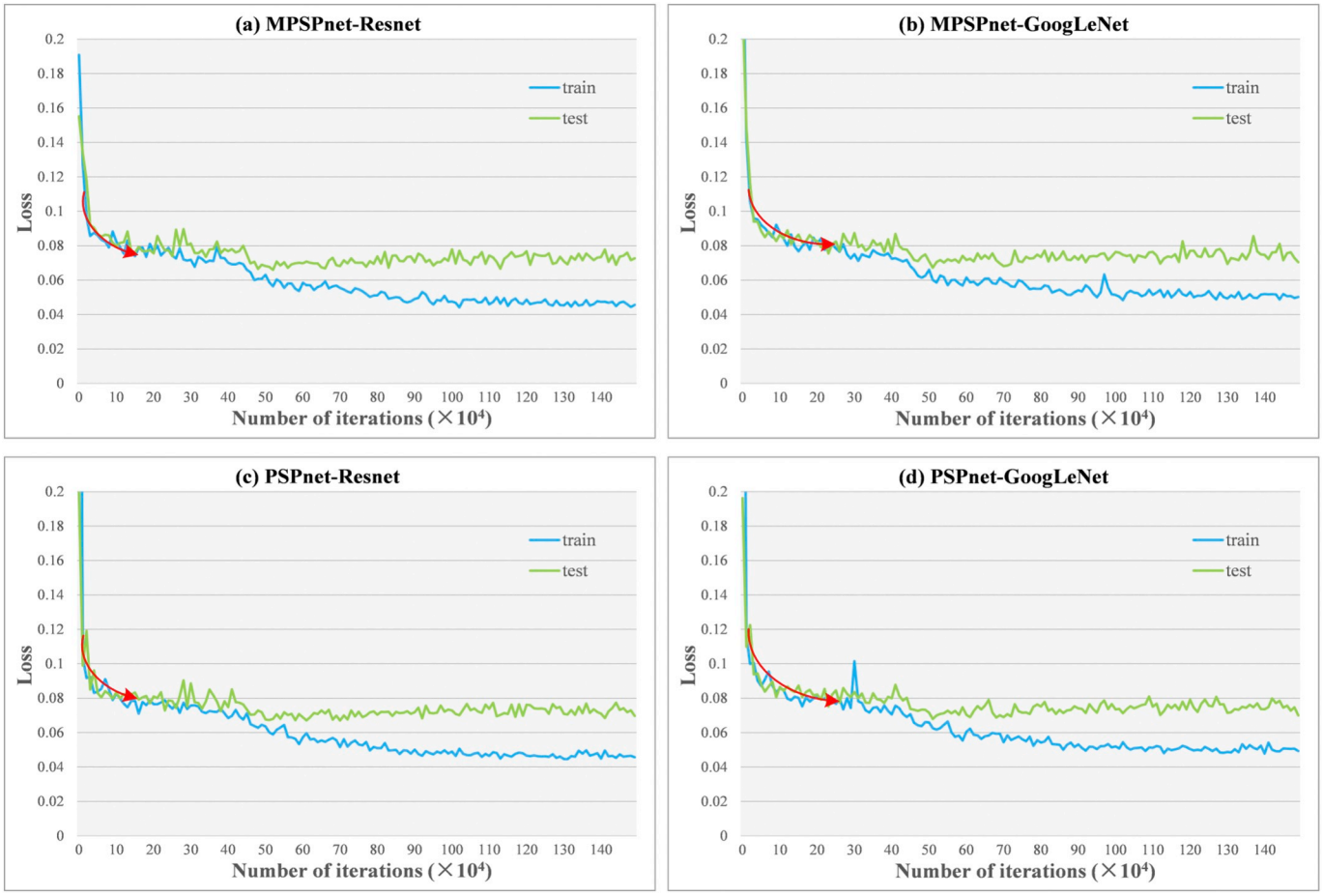

Fig. 14. Loss curves of the different CNN models during the training process. ResNet model provides faster convergence at the early stage than GoogLeNet model.

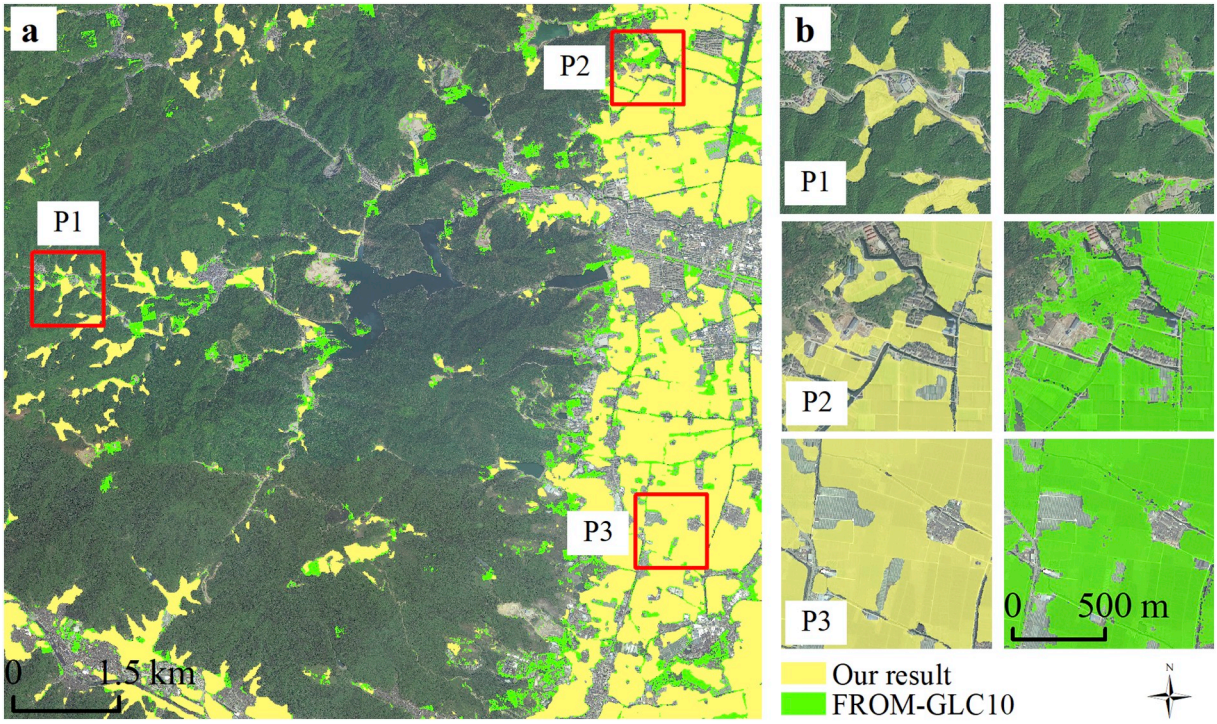

Fig. 15. Comparison between our results and FROM-GLC10: (a) true colour image from Yinzhou in Zhejiang overlaid with FROM-GLC10 and our result from bottom to top; (b) close view for the marked patches from (a).

exceeded 0.700 except Zhejiang (0.699). As aforementioned, the proposed method could learn stable feature representations of cropland, thereby a trained model has the strong transferability in space and it can be utilized in other places with similar cropland landscape to the training areas. Both qualitative and quantitative assessments confirmed the strong generalization ability of the proposed method in spatial transfer. Notably, we merely discuss the spatial transferability of the proposed method within a limited large-scale area (at the provincial or regional level) with similar cropland landscape in this paper. Spatial transfer at national or global scales will remain to be explored in future study.

5.3. Comparison with traditional object-based classification method

To evaluate the classification performance of the proposed MPSPNet method, it was comprehensively compared with the traditional object-based classification method. We selected the representative Object-based image analysis with Random forest (OBIA-RF) method, where spectral features (mean and standard deviation), texture (grey-level co-occurrence matrix) and geometry (e.g. shape index, perimeter-area ratio) were fed into a parameterized RF for classification. Considering the complexity and difficulty of object-based classification for large-scale cropland extraction, we replaced the whole province with selected

Table 7
Average accuracies of our results and FROM-GLC10 under different topographical landscape characteristics for each study area. (HLJ) Heilongjiang, (HB) Hebei, (ZJ) Zhejiang and (GD) Guangdong.

		Our result				FROM-GLC10			
		OA (%)	Kappa	PA (%)	UA (%)	OA (%)	Kappa	PA (%)	UA (%)
HLJ	Plain	91.15	0.76	90.95	94.03	81.96	0.54	86.45	82.68
	Mountain	89.06	0.75	87.75	89.14	71.17	0.39	85.68	48.98
	Mixed	90.23	0.75	92.24	92.82	79.05	0.54	90.37	77.86
HB	Plain	91.22	0.79	91.88	94.76	85.57	0.63	83.73	97.02
	Mountain	90.07	0.72	86.48	72.31	80.09	0.42	63.33	54.94
	Mixed	89.53	0.71	85.27	80.69	80.92	0.52	63.60	87.69
ZJ	Plain	88.02	0.72	81.14	80.69	79.83	0.55	62.90	81.70
	Mountain	90.50	0.64	79.03	62.31	85.16	0.40	62.98	40.26
	Mixed	89.82	0.70	79.43	74.79	82.09	0.52	60.27	69.79
GD	Plain	90.62	0.76	83.05	87.22	71.79	0.39	53.63	82.23
	Mountain	90.30	0.69	72.7	80.34	75.62	0.37	44.13	68.65
	Mixed	90.65	0.73	78.19	81.68	68.80	0.33	41.48	83.79

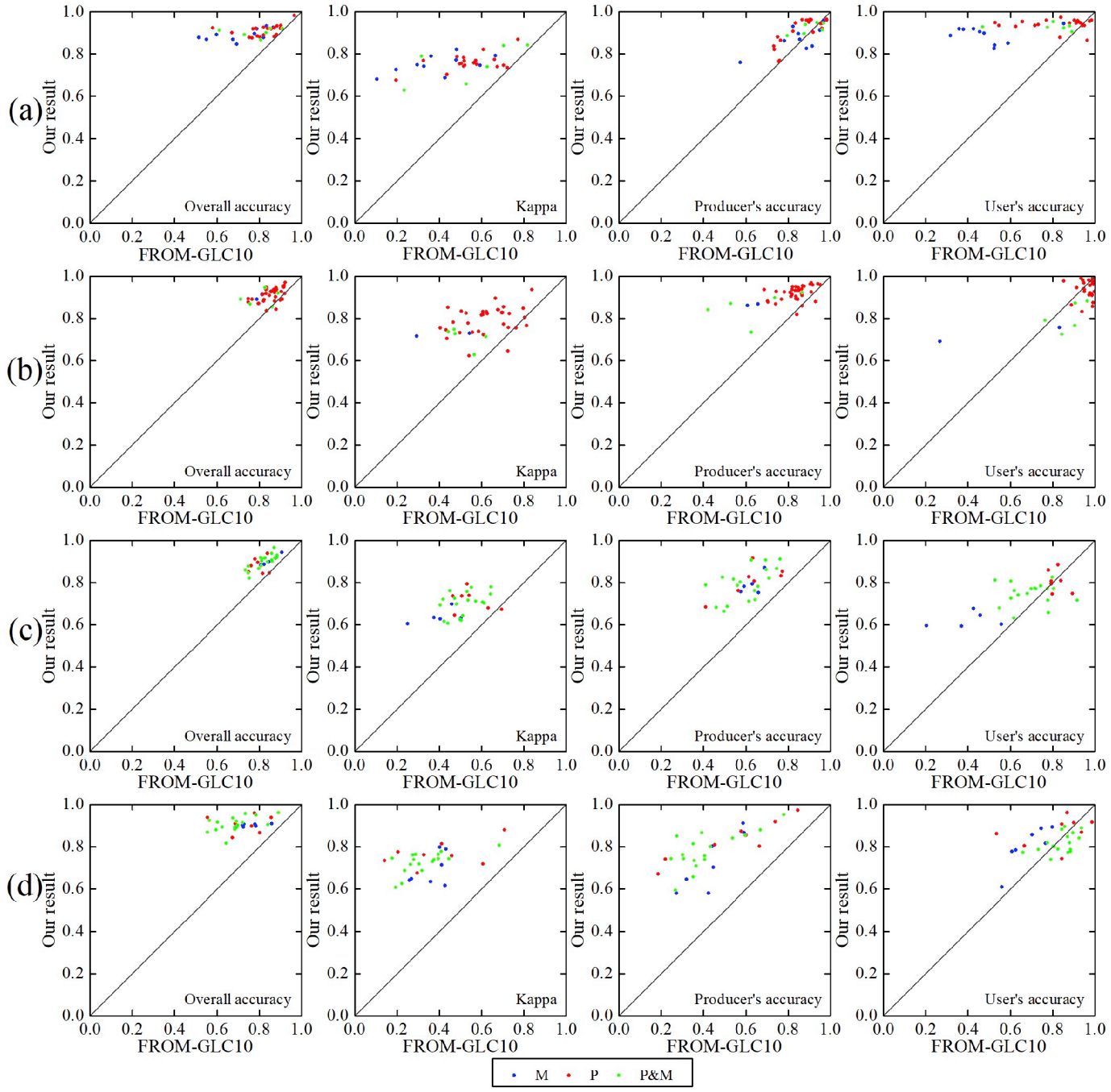

Fig. 16. Comparison of the accuracies between our result (y-axis) and FROM-GLC10 (x-axis) under different topographical landscapes in the four study areas. (a) Heilongjiang, (b) Hebei, (c) Zhejiang and (d) Guangdong.

three county-level study areas (Fig. 12) with different terrain types in Guangdong Province to conduct the experiment. We utilized the grid as the basic unit to segment every study area into 16 or 12 sub-regions and randomly split them into a training, a validation and a test set (Fig. 12). The training set was used to train the model. The validation set was used for determining the optimal model parameters. And the test set were exploited for final accuracy evaluation. In addition to the model trained on the training set from the same study area, the models derived from the other two study areas were leveraged for cropland extraction to compare the transferability and generalization ability of different methods. For convenience, we define the obtained model from its own study area as inner-area model and the one trained from the other study areas as inter-area model. Simultaneously, the MPSPNet model used for the whole Guangdong province in Section 4 is also compared with the above models.

we compared the performance of MPSPNet and OBIA-RF methods from the respects of qualitative and quantitative (as displayed in Fig. 13 and Table 5). It should be noted that taking plain area as an example, OBIA-RF_P and $MPSPNet_P$ are inner-area model for its test set whereas OBIA-RF_M, $MPSPNet_M$, OBIA-RF_{PM} and $MPSPNet_{PM}$ are inter-area model. In general, Regardless of MPSPNet or OBIA-RF, the inner-area model had more accurate classification performance than the other two

inter-area model. And MPSPNet is significantly superior to OBIA-RF, corroborating the advantages of CNN in extracting cropland using VHSR remote sensed images. According to the results of inter-area model, the more undesirable visual effects and misclassifications were observed for OBIA_RF approach in contrast to MPSPNet. In particular, the inter-area model of MPSPNet achieved a higher accuracy compared with inner-area model of OBIA-RF. This shows that MPSPNet possesses the strong spatio-temporal transferability and generalization capabilities for data variation. Amongst all MPSPNet models, the model used for the whole province obtained the closest performance to inner-area model and realized better classification results than inter-area models, which indicated that MPSPNet can generalize various characteristics of cropland through large amount of training samples and demonstrated the applicability of the proposed MPSPNet to large-scale cropland classification based on VHSR images.

5.4. Effects of CNN models

Taking Guangdong as an example, we tested the effects of the different CNN models. In addition to original PSPNet, GoogLeNet with eight inception modules (Szegedy et al., 2015) was introduced to replace Resnet, thereby four CNN models including MPSPNet-Resnet (proposed in the study), PSPNet-Resnet, MPSPNet-GoogLeNet and PSPNet-GoogLeNet were compared. It should be noted that we applied the hyper parameters setting as mentioned in Section 3.3 to training the four CNN models.

The loss curves during the training process and accuracy assessment of the different CNN models were presented in Fig. 14 and Table 6. Clearly, the accuracies of Resnet was not substantially superior to the GoogLeNet. Nevertheless, the ResNet can ease the optimization by converging faster at the early stage. Regardless of Resnet or GoogLeNet, the results of the MPSPNet is better than the original PSPNet, which demonstrates it is an effective strategy to combine low-level local features with high-level features to generate fine classification.

5.5. Comparison with current cropland products

We compared our results with current cropland maps generated from FROM-GLC10, i.e., 10 m resolution global land cover maps with ten classes, for 2017 from Sentinel-2 data (Gong et al., 2019). Fig. 15 shows visual differences between the two products for Yinzhou of Zhejiang as an example. As shown in Fig. 15(a), the FROM-GLC10 and our result are superimposed on the satellite imagery from bottom to top, respectively. The proposed approach produced broadly similar distribution patterns of cropland with FROM-GLC10, but significantly more accurate cropland characteristics with finer spatial details. In particular the proposed approach not only extracted more details within plain zones, but also detected scattered small cropland parcels in mountain areas, due to higher resolution input images. For example, Fig. 15(b) p2 shows the proposed approach distinguished cropland from others for plain areas, whereas some forest lands were misclassified as cropland by FROM-GLC10. Similarly, Fig. 15(b) p1 shows that fragmented small cropland areas omitted by FROM-GLC10 can be successfully extracted by the proposed approach in mountain areas.

Table 7 compares average accuracy for the proposed approach with FROM-GLC10 for all validation zones in every study area. The proposed approach consistently exhibited higher average accuracy in plain, mountain and mixed zones. Heilongjiang and Hebei exhibited considerable OA, kappa, and PA improvements for mountain zones, implying much less omission error. Areas with fragmented landscapes, such as Zhejiang and Guangdong, exhibited significantly improved accuracies for all terrain landscape zones. Fig. 16 demonstrates the scatterplots of the four accuracy metrics under different landscapes for four study areas, by marking our results in the vertical axis and FROM-GLC10 in the horizontal axis, respectively. Most points situate above the 1:1 diagonal line, denoting that significant advantage from the proposed approach, i.e., higher accuracy than FROM-GLC10 in a majority of the validation zones. Given that FROM-GLC10 is existing large-scale cropland product with the finest spatial resolution, we could not confirm the proposed approach was definitely superior to FROM-GLC10 because they were derived from very different data sources and hence inconsistency could be due to different spatial resolutions. However, the results verified the proposed approach effectiveness to automatically extract large-scale croplands with high precision at 2 m spatial resolution for operational application.

5.6. Future research

The proposed method achieved good performance and efficiency for large-scale cropland classification based on VHSR remote sensed images. However, accuracy imbalances occurred between mountain and plain zone. It may be insufficient to extract cropland for the whole area without considering terrain landscapes. Therefore, future study will separate mountain and plain areas for individual model training and cropland mapping. Additionally, this study used the mean of visual interpretation to annotate samples and hence ensure training data accuracy, but this would be ineffective in large-scale settings. This issue might be addressed by using existing land cover information to extract the training samples in future research.

6. Conclusion

This study developed a novel approach based deep CNNs to automatically identify cropland at very high resolution over large areas for operational agriculture monitoring. We selected four provinces incorporating diverse agricultural systems (Heilongjiang, Hebei, Zhejiang, and Guangdong Provinces) as study areas with images acquired by Gaofen-1 (GF1), Gaofen-2 (GF2) and ZiYuan-3 (ZY3) satellites. A manually labeled training dataset was extracted for each study area based on observations from 2016 to train the network, and applied to cropland identification for similar images from 2017. Cropland mapping results for the four study areas achieved the overall accuracies ranging from 89.99% to 92.31% and confirmed unsupervised learned features based on deep learning exhibited strong generalization ability, hence croplands with various shapes, sizes along with texture and spectral characteristics could be successfully identified on VHSR images. We investigated how the proposed MPSPNet model learns and understands input images on different layers through features visualization, and in particular confirmed the proposed strategy to combine low-level and high-level features provided an efficient and accurate approach for cropland mapping, capturing edge information related to object boundaries and long range spatial dependencies to identify classification classes, respectively. The temporal transfer and spatial transfer assessments demonstrated the practical potential and applicability of the proposed approach for large-scale cropland mapping. The contrast to the traditional object-based classification method and other CNN models corroborated the advantages of MPSPNet. We further compared the proposed approach with current cropland maps produced from FROM-GLC10, confirming higher accuracy for different terrains (plain, mountain and mixed zones) from the proposed approach in each study area. Thus, the proposed MPSPNet provided an effective and promising approach for operational agriculture monitoring over large areas at very high resolution.

Table A1
Accuracies of our results for each validation zone. Plain areas (P), Mountain zones (M), Plain and mountain mixed zones (PM).

Study area	Terrain type	Validation zone	OA (%)	Kappa	Cropland_PA (%)	Cropland_UA (%)
Heilongjiang	P	Bayan	92.46	0.742	96.26	94.58
	P	Daoli	88.26	0.740	96.17	86.51
	P	Fuyuan	92.51	0.677	96.54	94.83
	P	Hulan	93.17	0.766	96.08	95.62
	P	Kedong	92.02	0.784	95.92	93.55
	P	Longfeng	89.00	0.757	76.52	93.04
	P	Mulan	92.10	0.822	94.69	93.51
	P	Qinggang	91.80	0.773	92.14	97.25
	P	Qingan	92.09	0.747	94.45	95.74
	P	Ranghulu	88.47	0.754	76.83	93.53
	P	Shuangcheng	93.57	0.771	96.36	95.91
	P	Sifangtai	90.19	0.769	90.70	95.28
	P	Songbei	87.57	0.746	83.68	95.85
	P	Suibin	89.14	0.735	90.28	94.56
	P	Tailai	87.72	0.756	82.17	94.00
	P	Xingan	87.98	0.751	86.56	93.44
	P	Yanshou	98.15	0.869	88.00	87.78
	P	Yian	92.18	0.704	94.78	95.96
	P	Yilan	92.59	0.788	95.91	94.53
	P	Zhaodong	91.99	0.745	94.95	95.09
	Average_P		**91.15**	**0.760**	**90.95**	**94.03**
	M	Cuiluan	84.67	0.687	91.24	82.67
	M	Dongan	87.97	0.747	83.78	85.19
	M	Dongning	89.59	0.772	86.22	84.33
	M	Hengshan	86.88	0.726	89.63	88.67
	M	Lishu	87.84	0.680	93.11	90.60
	M	Mashan	93.18	0.790	76.17	91.73
	M	Mulimg	91.83	0.821	86.98	89.85
	M	Tieli	92.56	0.793	95.87	94.40
	M	Wumahe	89.09	0.748	91.99	92.04
	M	Yangming	87.01	0.741	82.53	91.91
	Average_M		**89.06**	**0.751**	**87.75**	**89.14**
	PM	Beian	92.14	0.842	94.82	90.46
	PM	Dongshan	92.13	0.840	88.79	93.32
	PM	Huanan	86.65	0.658	89.52	92.35
	PM	Jidong	89.07	0.630	94.01	92.66
	PM	Longjiang	91.23	0.790	94.73	92.81
	PM	Luobei	90.16	0.740	91.58	95.32
	Average_PM		**90.23**	**0.750**	**92.24**	**92.82**
Hebei	P	Anci	87.65	0.745	87.61	91.45
	P	Anping	92.58	0.840	94.38	93.98
	P	Anxin	91.67	0.803	92.51	95.56
	P	Baixiang	94.67	0.894	92.65	96.67
	P	Botou	92.12	0.828	95.08	92.75
	P	Changli	86.97	0.733	86.70	90.88
	P	Daming	94.59	0.851	94.85	98.10
	P	Dingxing	92.70	0.827	91.64	98.14
	P	Dingzhou	94.77	0.823	95.75	97.89
	P	Fengnan	84.21	0.647	81.63	96.04
	P	Gucheng	83.63	0.622	90.47	85.69
	P	Huanghua	88.94	0.766	87.86	94.56
	P	Li	96.96	0.935	96.17	99.01
	P	Linxi	94.55	0.824	95.28	98.02
	P	Linzhang	93.52	0.834	93.57	97.73
	P	Longyao	95.16	0.823	96.62	97.60
	P	Luancheng	88.24	0.756	89.23	91.02
	P	Mancheng	89.23	0.780	87.99	93.45
	P	Mengcun	94.40	0.831	96.10	96.81
	P	Qianan	86.98	0.739	88.94	86.38
	P	Qinghe	93.69	0.830	92.97	98.84
	P	Quzhou	90.70	0.703	90.39	98.44
	P	Ren	88.65	0.755	95.11	87.40
	P	Sanhe	88.89	0.777	93.45	83.01
	P	Wuqiao	91.83	0.755	91.42	98.51
	P	Xinhe	92.74	0.847	90.89	97.47
	P	Xinji	95.63	0.854	95.90	98.79
	P	Xiong	90.14	0.774	89.15	96.69
	P	Yanshan	93.70	0.842	93.87	97.53
	P	Yuanshi	88.86	0.731	92.24	92.01
	P	Yutian	92.54	0.816	93.21	96.44
	P	Zhengding	88.06	0.723	85.84	97.61
	P	Zhuozhou	91.19	0.818	92.46	92.56
	Average_P		**91.22**	**0.792**	**91.88**	**94.76**
	M	Weichang	91.16	0.715	86.82	69.05

(*continued on next page*)

Table A1 (*continued*)

Study area	Terrain type	Validation zone	OA (%)	Kappa	Cropland_PA (%)	Cropland_UA (%)
	M	Wei	88.97	0.729	86.13	75.57
	Average_M		**90.07**	**0.722**	**86.48**	**72.31**
	PM	Ci	85.52	0.626	92.20	88.27
	PM	Pingshan	92.02	0.714	73.40	79.14
	PM	Wuan	86.68	0.728	89.73	87.15
	PM	Xingtai	88.89	0.735	86.92	76.50
	PM	Zhuolu	94.55	0.747	84.10	72.41
	Average_PM		**89.53**	**0.710**	**85.27**	**80.69**
Zhejiang	P	Haining	84.33	0.682	83.14	80.93
	P	Jiangbei	91.12	0.794	91.64	80.59
	P	Jiaojiang	87.85	0.737	80.70	85.93
	P	Jiashan	84.51	0.673	85.34	74.89
	P	Luqiao	89.52	0.738	82.74	79.44
	P	Nanxun	85.03	0.646	76.11	74.55
	P	Zhenhai	93.80	0.736	68.33	88.50
	Average_P		**88.02**	**0.715**	**81.14**	**80.69**
	M	Jinyun	88.20	0.635	87.01	59.30
	M	Kaihua	94.34	0.700	79.22	67.84
	M	Tonglu	91.62	0.633	78.16	60.21
	M	Wuyi	88.65	0.627	75.59	64.54
	M	Yunhe	89.71	0.605	75.19	59.64
	Average_M		**90.50**	**0.640**	**79.03**	**62.31**
	PM	Beilun	91.05	0.726	80.12	76.32
	PM	Dinghai	93.83	0.757	81.58	77.25
	PM	Huangyan	91.84	0.722	82.53	72.40
	PM	Kecheng	85.89	0.607	68.29	71.70
	PM	Leqing	87.78	0.695	90.56	67.99
	PM	Linhai	91.60	0.711	78.28	74.14
	PM	Putuo	96.53	0.760	78.93	76.93
	PM	Qujiang	81.96	0.622	90.65	65.68
	PM	Sanmen	89.81	0.698	71.83	81.30
	PM	Shangyu	90.79	0.708	71.21	82.49
	PM	Wenling	90.15	0.716	78.63	77.22
	PM	Xiangshan	91.52	0.778	86.61	80.56
	PM	Xiaoshan	92.88	0.779	91.22	74.90
	PM	Yinzhou	91.97	0.745	81.99	77.15
	PM	Yiwu	87.84	0.624	68.65	71.50
	PM	Yongkang	90.06	0.703	76.10	77.02
	PM	Yuhang	84.64	0.615	66.46	78.56
	PM	Yuhuan	86.62	0.643	86.18	63.18
	Average_PM		**89.82**	**0.700**	**79.43**	**74.79**
Guangdong	P	Jiangcheng	90.90	0.815	87.26	91.70
	P	Longhu	93.93	0.878	97.29	91.41
	P	Maonan	89.47	0.761	81.00	86.88
	P	Nanhai	95.87	0.775	74.16	86.22
	P	Nansha	86.75	0.718	91.88	74.37
	P	Panyu	89.88	0.757	85.51	80.43
	P	Xiashan	93.75	0.736	67.07	90.73
	P	Xuwen	84.43	0.675	80.21	96.01
	Average_P		**90.62**	**0.764**	**83.05**	**87.22**
	M	Deqing	89.31	0.641	64.61	77.69
	M	Fengkai	89.86	0.633	80.22	60.95
	M	Lianshan	90.97	0.616	57.92	78.43
	M	Lianzhou	90.83	0.650	57.94	88.68
	M	Qujiang	90.17	0.797	86.75	89.34
	M	Xinfeng	90.56	0.714	70.34	85.72
	M	Zhenjiang	90.37	0.788	91.11	81.57
	Average_M		**90.30**	**0.691**	**72.70**	**80.34**
	PM	Boluo	88.02	0.686	73.51	80.04
	PM	Chaoyang	90.41	0.807	95.20	84.12
	PM	Chengqu	95.74	0.740	74.26	78.60
	PM	Dinghu	95.00	0.764	81.33	77.25
	PM	Enping	89.42	0.719	70.94	88.76
	PM	Gaoming	92.20	0.718	74.34	78.65
	PM	Haifeng	86.89	0.608	59.30	81.92
	PM	Heshan	92.57	0.746	85.00	73.93
	PM	Huicheng	93.67	0.765	76.16	84.73
	PM	Jiaoling	89.10	0.778	85.45	89.51
	PM	Jiexi	90.15	0.730	79.97	78.92
	PM	Luoding	88.14	0.739	84.15	81.86
	PM	Nanxiong	81.57	0.624	87.90	80.33
	PM	Qingxin	96.02	0.743	75.57	77.34
	PM	Yangdong	91.34	0.743	73.89	86.61
	PM	Yangxi	91.46	0.763	86.52	77.73
	PM	Zengcheng	89.33	0.687	65.76	88.22
	Average_PM		**90.65**	**0.727**	**78.19**	**81.68**

Funding

This work was supported by the National High Resolution Earth Observation System (The Civil Part) Technology Projects of China under Grant No. 11-Y20A16-9001-17/18 and the National Key Research and Development Program of China under Grant No. 2018YFC1504603.

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.

References

Abou EL-Magd, I., Tanton, T., 2003. Improvements in land use mapping for irrigated agriculture from satellite sensor data using a multi-stage maximum likelihood classification. Int. J. Remote Sens. 24 (21), 4197–4206.
Bartholomé, E., Belward, A.S., 2005. GLC2000: a new approach to global land cover mapping from earth observation data. Int. J. Remote Sens. 26, 1959–1977.
Biradar, C.M., Thenkabail, P.S., Noojipady, P., Li, Y., Dheeravath, V., Turral, H., et al., 2009. A global map of rainfed cropland areas (GMRCA) at the end of last millennium using remote sensing. Int. J. Appl. Earth Obs. Geoinf. 11 (2), 114–129.
Bontemps, S., Defourny, P., Bogaert, E.V., Arino, O., Kalogirou, V., Perez, J.R., 2011. GLOBCOVER 2009—Products Description and Validation Report. ESA, Paris, France.
Bruzzone, L., Carlin, L., 2006. A multilevel context-based system for classification of very high spatial resolution images. IEEE Trans. Geosci. Remote Sens. 44 (9), 2587–2600.
Chavez, P., Sides, S.C., Anderson, J.A., 1991. Comparison of three different methods to merge multiresolution and multispectral data-Landsat TM and SPOT panchromatic. Photogramm. Eng. Remote. Sens. 57, 295–303.
Chen, J., Ban, Y., China, Li S., 2014. Open access to earth land-cover map. Nature 514 (7523), 434.
Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., et al., 2015. Global land cover mapping at 30 m resolution: a POK-based operational approach. ISPRS J. Photogramm. Remote Sens. 103, 7–27.
Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV) 801–818.
Cheng, G., Zhou, P., Han, J., 2016. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Trans. Geosci. Remote Sens. 54 (12), 7405–7415.
De Fries, R.S., Hansen, M., Townshend, J.R.G., Sohlberg, R., 1998. Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. Int. J. Remote Sens. 19, 3141–3168.
Duro, D.C., Franklin, S.E., Dubé, 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 HGR imagery. Remote Sens. Environ. 118, 259–272.
Friedl, M.A., McIver, D.K., Hodges, J.C., Zhang, X.Y., Muchoney, D., Strahler, A.H., et al., 2002. Global land cover mapping from MODIS: algorithms and early results. Remote Sens. Environ. 83 (1–2), 287–302.
Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182.
Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., et al., 2015. Mapping global cropland and field size. Glob. Chang. Biol. 21, 1980–1992.
Gong, P., Wang, J., Yu, L., Zhao, Y., Liang, L., Niu, Z., et al., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. Int. J. Remote Sens. 34 (7), 2607–2654.
Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., et al., 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. Sci. Bull. 64, 370–373.
Griffiths, P., van der Linden, S., Kuemmerle, T., Hostert, P., 2013. A pixel-based Landsat compositing algorithm for large area land cover mapping. IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens. 6 (5), 2088–2101.
He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778.
Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A., Schmullius, C., 2008. Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. Remote Sens. Environ. 112, 2538–2556.
Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science 353, 790–794.
Justice, C.O., Becker-Reshef, I., 2007. Report from the Workshop on Developing a Strategy for Global Agricultural Monitoring in the Framework of Group on Earth Observations (Geo). UN FAO (July).
Kaiser, P., Wegner, J.D., Lucchi, A., Jaggi, M., Hofmann, T., Schindler, K., 2017. Learning aerial image segmentation from online maps. IEEE Trans. Geosci. Remote Sens. 55, 6054–6068.
Kingma, D.P., Ba, J., 2015. Adam: a method for stochastic optimization. In: International Conference on Learning Representations.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. LiDAR remote sensing for ecosystem studies. Biosci. 52 (1), 19–30.
Liu, J., Shao, G., Zhu, H., Liu, S., 2005. A neural network approach for enhancing information extraction from multispectral image data. Can. J. Remote. Sens. 31, 432–438.
Long, J.A., Lawrence, R.L., Greenwood, M.C., Marshall, L., Miller, P.R., 2013. Object-oriented crop classification using multitemporal ETM+ slc-off imagery and random forest. GISci. Remote Sens. 50 (4), 418–436.
Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440.
Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L.W.M.J., Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1km AVHRR data. Int. J. Remote Sens. 21 (6–7), 1303–1330.
Lu, M., Wu, W.B., Zhang, L., Liao, A.P., Peng, S., Tang, H.J., 2016. A comparative analysis of five global cropland datasets in China. Sci. China Earth Sci. 59, 2307–2317.
Lyu, H., Lu, H., Mou, L., Li, W., Wright, J., Li, X., et al., 2018. Long-term annual mapping of four cities on different continents by applying a deep information learning method to Landsat data. Remote Sens. 10, 471.
Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional neural networks for large-scale remote-sensing image classification. IEEE Trans. Geosci. Remote Sens. 55 (2), 645–657.
Marmanis, D., Datcu, M., Esch, T., Stilla, U., 2016. Deep learning earth observation classification using ImageNet Pretrained networks. IEEE Geosci. Remote Sens. Lett. 13, 105–109.
Massey, R., Sankey, T.T., Yadav, K., Congalton, R.G., Tilton, J.C., 2018. Integrating cloud-based workflows in continental-scale cropland extent classification. Remote Sens. Environ. 219, 162–179.
Mathur, A., Foody, G.M., 2008. Crop classification by support vector machine with intelligently selected training data for an operational application. Int. J. Remote Sens. 29 (8), 2227–2240.
Ozdogan, M., Woodcock, C.E., 2006. Resolution dependent errors in remote sensing of cultivated areas. Remote Sens. Environ. 103 (2), 203–217.
Phalke, A.R., Özdoğan, M., 2018. Large area cropland extent mapping with Landsat data and a generalized classifier. Remote Sens. Environ. 219, 180–195.
Pittman, K., Hansen, M.C., Becker-Reshef, I., Potapov, P.V., Justice, C.O., 2010. Estimating global cropland extent with multi-year MODIS data. Remote Sens. 2, 1844–1863.
Ramankutty, N., Evan, A.T., Monfreda, C., Foley, J.A., 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. Glob. Biogeochem. Cycles 22 (1).
Samaniego, L., Schulz, K., 2009. Supervised classification of agricultural land cover using a modified k-nn technique (mnn) and landsat remote sensing imagery. Remote Sens. 1 (4), 875–895.
See, L., Fritz, S., You, L., Ramankutty, N., Herrero, M., Justice, C., et al., 2015. Improved global cropland data as an essential ingredient for food security. Glob. Food Sec. 4, 37–45.
Ševo, I., Avramović, A., 2016. Convolutional neural network based automatic object detection on aerial images. IEEE Geosci. Remote Sens. Lett. 13 (5), 740–744.
Sun, Y., Zhang, X., Xin, Q., Huang, J., 2018. Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data. ISPRS J. Photogramm. Remote Sens. 143, 3–14.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9.
Thenkabail, P.S., Wu, Z., 2012. An automated cropland classification algorithm (ACCA) for Tajikistan by combining Landsat, MODIS, and secondary data. Remote Sens. 4 (10), 2890–2918.
Thenkabail, P.S., Biradar, C.M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Gangalakunta, O.R.P., Turral, H., Cai, X., et al., 2009. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. Int. J. Remote Sens. 30, 3679–3733.
Waldner, F., Canto, G.S., Defourny, P., 2015. Automated annual cropland mapping using knowledge-based temporal features. ISPRS J. Photogramm. Remote Sens. 110, 1–13.
Waldner, F., Fritz, S., Di Gregorio, A., Plotnikov, D., Bartalev, S., Kussul, N., et al., 2016. A unified cropland layer at 250 m for global agriculture monitoring. Datamation 1 (1), 3.
Wang, J., Song, J., Chen, M., Yang, Z., 2015. Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine. Int. J. Remote Sens. 36 (12), 3144–3169.
Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnelt, J., Congalton, R.G., et al., 2017. Automated cropland mapping of continental Africa using Google earth engine cloud computing. ISPRS J. Photogramm. Remote Sens. 126, 225–244.
Yu, F., Koltun, V., 2016. Multi-scale context aggregation by dilated convolutions. Int. Conf. Learn. Represent. 1–13.
Yu, L., Wang, J., Clinton, N., Xin, Q., Zhong, L., Chen, Y., Gong, P., 2013. FROM-GC: 30 m global cropland extent derived through multisource data integration. Int. J. Digital Earth 6, 521–533.
Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: a technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Mag. 4, 22–40.
Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2018. An object-based convolutional neural networks (OCNN) for urban land use classification. Remote Sens. Environ. 216, 57–70.
Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2881–2890.
Zhu, X.X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5 (4), 8–36.