

Article

# Urban Green Plastic Cover Mapping Based on VHR Remote Sensing Images and a Deep Semi-Supervised Learning Framework

Jiantao Liu <sup>1</sup>, Quanlong Feng <sup>2,\*</sup>, Ying Wang <sup>3,4</sup>, Bayartungalag Batsaikhan <sup>5</sup>, Jianhua Gong <sup>6</sup>, Yi Li <sup>6</sup>, Chunting Liu <sup>1</sup> and Yin Ma <sup>7</sup>

<sup>1</sup> School of Surveying and Geo-Informatics, Shandong Jianzhu University, Jinan 250101, China; liujiantao18@sdjzu.edu.cn (J.L.); liuct19952020@126.com (C.L.)

<sup>2</sup> College of Land Science and Technology, China Agricultural University, Beijing 100083, China

<sup>3</sup> Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China; ywang@iue.ac.cn

<sup>4</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>5</sup> Institute of Geography and Geoecology, Mongolian Academy of Sciences, Ulaanbaatar 15170, Mongolia; bayartungalag\_b@mas.ac.mn

<sup>6</sup> National Engineering Research Center for Geoinformatics, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; gongjh@aircas.ac.cn (J.G.); liyi@aircas.ac.cn (Y.L.)

<sup>7</sup> School of Information Engineering, China University of Geosciences (Beijing), Beijing 100083, China; 3004200008@cugb.edu.cn

\* Correspondence: fengql@cau.edu.cn; Tel.: +86-10-6273-7554

Received: 10 August 2020; Accepted: 31 August 2020; Published: 2 September 2020



**Abstract:** With the rapid process of both urban sprawl and urban renewal, large numbers of old buildings have been demolished in China, leading to wide spread construction sites, which could cause severe dust contamination. To alleviate the accompanied dust pollution, green plastic mulch has been widely used by local governments of China. Therefore, timely and accurate mapping of urban green plastic covered regions is of great significance to both urban environmental management and the understanding of urban growth status. However, the complex spatial patterns of the urban landscape make it challenging to accurately identify these areas of green plastic cover. To tackle this issue, we propose a deep semi-supervised learning framework for green plastic cover mapping using very high resolution (VHR) remote sensing imagery. Specifically, a multi-scale deformable convolution neural network (CNN) was exploited to learn representative and discriminative features under complex urban landscapes. Afterwards, a semi-supervised learning strategy was proposed to integrate the limited labeled data and massive unlabeled data for model co-training. Experimental results indicate that the proposed method could accurately identify green plastic-covered regions in Jinan with an overall accuracy (OA) of 91.63%. An ablation study indicated that, compared with supervised learning, the semi-supervised learning strategy in this study could increase the OA by 6.38%. Moreover, the multi-scale deformable CNN outperforms several classic CNN models in the computer vision field. The proposed method is the first attempt to map urban green plastic-covered regions based on deep learning, which could serve as a baseline and useful reference for future research.

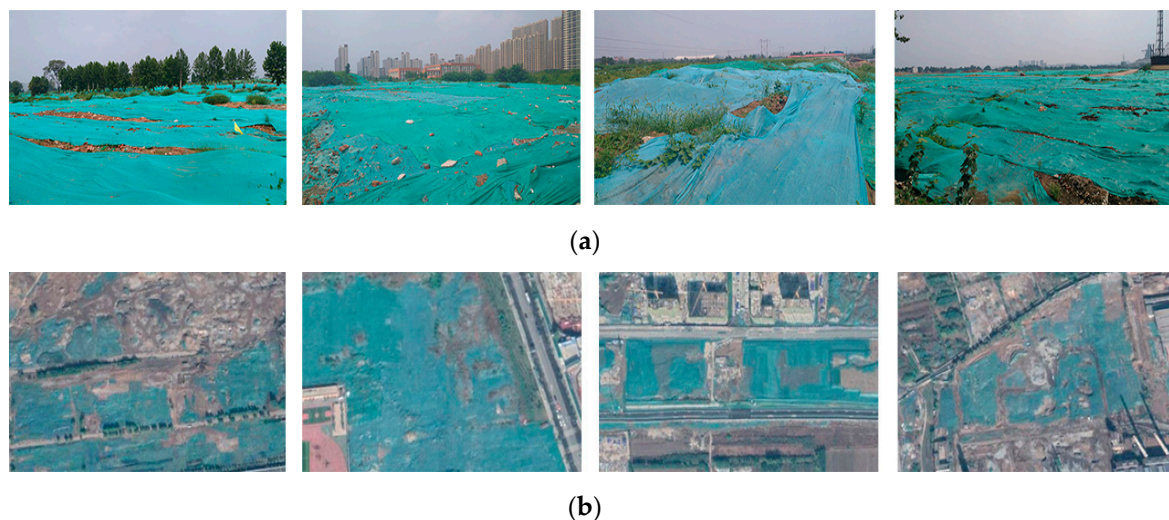
**Keywords:** green plastic cover; semi-supervised learning; deep learning; urban land cover mapping

## 1. Introduction

Nowadays, urban renewal has been widely performed around the globe, which could effectively relieve the shortage of urban land resources and improve urban land use efficiency [1–3]. For instance,

urban renewal in China has led to a large-scale demolition of old, low-density urban areas and urban villages over the past few decades [2]. During the renewal process, construction sites can be a source of huge amounts of dust, which could easily be transferred to the air and water nearby, leading to severe environmental pollution.

To alleviate the accompanied dust contamination, plastic mulch has been widely utilized by local governments in China (Figure 1). Moreover, the plastic mulch is always green, making it appear environmentally friendly. Actually, green plastic mulch is commonly made from polyethylene. Most urban renewal projects in China use the same green plastic mulch to alleviate dust contamination. After the construction process, the plastic mulch can be recycled at relevant chemical plants. Due to the stringent environmental protection regulations in China, green plastic mulch has been a must in urban renewal projects, offering an opportunity to accurately identify construction sites during urban sprawl and renewal. Therefore, it is of great significance to monitor and detect these green plastic covers (GPC), which could provide the spatial distribution of construction sites. Moreover, the detection of GPC could also help the environmental protection department with the precise control of construction dusts. However, as far as we know, there is still no report on GPC detection in the remote sensing field; therefore, we are highly motivated to propose an accurate classification method for GPC based on deep learning (DL) from VHR remotely sensed imagery.



**Figure 1.** Green plastic covered regions: (a) On-site photos; (b) VHR remote sensing images.

The accurate classification of GPC is challenging for the following reasons. Firstly, the complex urban landscapes lead to a high variability of the spatial patterns of GPC. Secondly, the limited labeled data of GPC could lead to overfitting of the deep learning-based classification model. To tackle these issues, we first exploited a multi-scale deformable CNN to account for the scale and shape variability of GPC. Afterwards, we integrated unlabeled GPC samples with labeled data into a semi-supervised learning framework to increase the model's generalization capability.

Actually, urban green plastic cover could be viewed as a specific urban land cover category. Due to its synoptic view and cost-effectiveness, remote sensing has been widely utilized for urban land use and land cover (LULC) mapping [4–6]. Traditional methods mainly focused on the visual inspection and vectorization from VHR remotely sensed imagery. However, this is both time and labor-intensive. Therefore, how to develop an automatic urban LULC classification method has become a hot research topic [7–9]. Early studies [10–15] mainly combined hand-crafted features (i.e., spectral indices, texture features) with machine learning classifiers to automatically extract a specific urban LULC type. For example, Shao et al. [10] performed the extraction of urban impervious surface based on random forest (RF) from GaoFen-1 and Sentinel-1A imagery. Yin et al. [11] applied both sub-pixel and super-pixel based methods for characterizing urban green space in Haidian District, Beijing. In our previous

studies, we also adopted random forest and texture analysis for urban vegetation mapping [12] and urban inundated regions extraction [13] from unmanned aerial vehicle (UAV) remote sensing data.

Meanwhile, there are still no relevant studies on urban green plastic cover mapping from remotely sensed data. Similar research mainly consists of the detection of construction sites and urban landfill. Yu et al. [16] proposed an unsupervised learning method for the classification of buildings under construction from multi-temporal UAV data. Silvestri et al. [17] utilized maximum likelihood classifier (MLC) and IKONOS images to recognize the uncontrolled urban landfills. Considering that no published studies focus on green plastic cover classification, this paper could be the first attempt to solve this important and challenging issue.

It should be noted that the aforementioned studies mainly rely on hand-crafted features and machine learning approaches for urban LULC classification. However, the design of hand-crafted features relies heavily on domain expertise, which might lead to inability to discover high-level and discriminative features from remote sensing images. On the other hand, deep learning has a strong ability to extract representative multi-level features from original data instead of empirical feature design and can work in an end-to-end manner, which has led to impressive performance in the computer vision field [18–22], such as in image classification [18], object detection [19], and semantic segmentation [22]. More recently, deep learning, especially deep CNN, has also been successfully applied in numerous remote sensing applications [23–29]. For instance, Huang et al. [23] proposed a semi-transfer deep CNN for urban land use mapping, based on VHR WorldView-2 imagery, and achieved an accuracy of 91.25%. Zhang et al. [24] proposed an object-based CNN for urban land use classification and achieved excellent classification accuracy and computational efficiency. Dong et al. [25] exploited a hybrid approach of random forest and CNN for subtropical forest mapping, and their results indicated that the developed model could lead to an improvement in information extraction. In our previous studies [30], we modified a two-branch CNN for urban land use mapping and found that the proposed CNN model outperforms traditional machine learning algorithms such as MLC, RF, and support vector machine (SVM). Moreover, we extended the above model to a multi-branch version for the fusion of multi-sensor and multi-temporal Sentinel-1/2 imagery [31]. All of the above studies demonstrated that CNN could provide an effective tool for remote sensing image classification. Therefore, in this study, we exploited a novel multi-scale deformable CNN to learn high-level and representative features for green plastic cover classification.

There is no denying that great improvements have been made in urban LULC mapping from remote sensing images through deep learning. However, deep learning works in an exhaustive data-driven manner, and a large number of labeled samples need to be fed into a DL model to avoid overfitting. Meanwhile, it should be noted that labeling enormous training samples is both labor-extensive and time-consuming, especially in the remote sensing and geoscience fields. Therefore, how to integrate the limited labeled samples with massive unlabeled data to improve the model's generalization capability is a key question. Semi-supervised learning precisely provides an effective tool to tackle this issue. He et al. [32] proposed generative adversarial network (GAN)-based, semi-supervised learning to classify hyperspectral images (HSI), while the unlabeled samples were from the GAN's generator. Fang et al. [33] also utilized a semi-supervised learning strategy based on several sample selection methods for HSI classification. Inspired by these studies, we also introduced a semi-supervised learning framework for the classification of urban green plastic covers based on limited well-annotated samples.

To sum up, the contributions of this study are as follows:

- (1) For the first time, we developed a deep learning method for urban green plastic cover mapping from VHR remote sensing data, which could provide an effective tool for construction site monitoring and environmental protection.
- (2) We exploited a multi-scale deformable CNN to tackle the variability of land object's scales and shapes under complex urban landscapes.

- (3) We integrated the limited labeled samples with massive unlabeled data into a semi-supervised learning framework to increase the generalization capability of the classification model for green plastic covers.

## 2. Study Area and Dataset

### 2.1. Study Area

The study area (Figure 2) is the urban built-up regions of Jinan City, which is the provincial capital of Shandong Province, China. It includes parts of Licheng District, Lixia District, Tianqiao District, Huaiyin District, Shizhong District, and Changqing District, with an approximate area of 1015 km<sup>2</sup>.

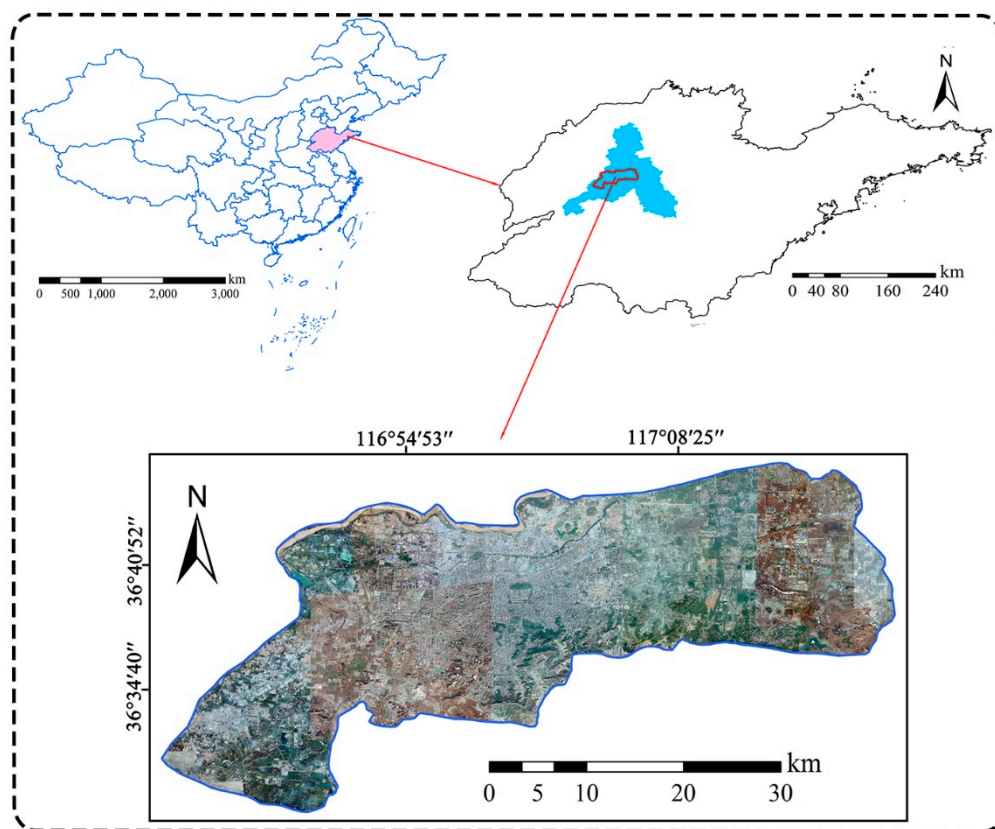


Figure 2. Study area.

Jinan City lies in the midwest of Shandong Province, on the eastern edge of the North China Plain. It is characterized by a temperate, semi-humid continental monsoon with an annual average temperature of 13.8 °C, an average frost-free period of 178 days, and an annual average rainfall of approximately 685 mm. Recently, Jinan has witnessed rapid urban sprawl and renewal. Numerous villages on the fringe of urban areas have been demolished, and some old buildings in the urban areas have been reconstructed. Most of these renewal regions are covered by green plastic mulch.

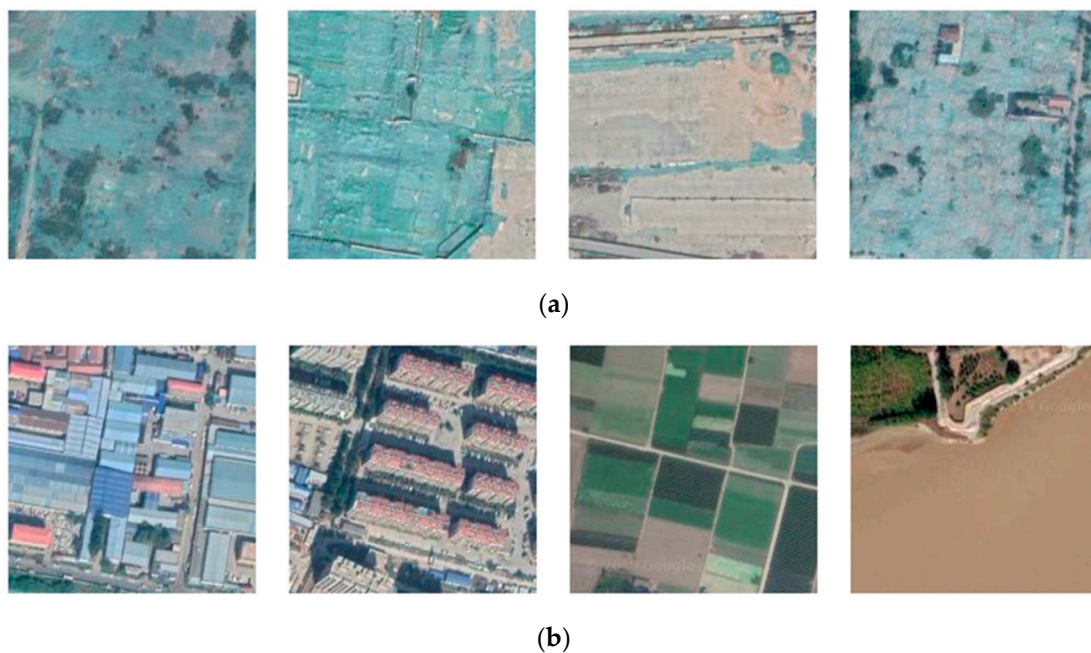
### 2.2. Dataset

Considering the widespread usage and data availability, the remotely sensed data from the Google Earth (GE) platform [34] were adopted. Specifically, the image was from the GE history database (obtained in 2019) and had a spatial resolution of about 1.19 m/pixel. Actually, the corresponding remote sensing imagery was mainly provided by Maxar (namely, DigitalGlobe company, Westminster, CO, USA). The optical sensors included WorldView-2, WorldView-3, and WorldView-4. Although the WorldView series could provide multi-spectral observations, the data provided by the Google



Earth platform have only three bands (namely, red, green, and blue, RGB). Moreover, the Google Earth platform only provides data at an 8-bit radiometric resolution.

The size of the image was  $35,976 \times 63,055$  pixels, corresponding to about  $43 \times 75$  km<sup>2</sup> (Figure 2). The classification scheme in this study included two types: Green plastic cover (GPC) and non-GPC. Both the training and testing samples belong to image patches with a size of  $224 \text{ pixel} \times 224 \text{ pixel}$ . Actually, the size of  $224 \text{ pixel} \times 224 \text{ pixel}$  has been a standard image patch size in the computer vision (CV) field, where the popular convolutional neural networks (e.g., ResNet, DenseNet) take a  $224 \times 224$  image patch and output a predicted label. Therefore, to be comparable with these CV models, we also used this setting in this study. Furthermore, as the spatial resolution is about 1.2 m/pixel, the  $224 \times 224$  image patch corresponds to  $268 \times 268$  m<sup>2</sup>. Under this context, the image patch could cover a scene that is not too big or too small for the task of plastic covered region detection. Figure 3 illustrates several samples of each land cover type.



**Figure 3.** Samples of each land cover category: (a) Green plastic cover (GPC); (b) non-GPC.

In order to describe the material composition of GPC in detail, we downloaded Sentinel-2 L2A data acquired on 28 August 2019 from the European Space Agency (ESA) and delineated the spectral reflectance signature of GPC (Figure 4) using bands 2-8 (Visible/Near Infrared), band 8a (Near Infrared), and band 11-12 (Shortwave Infrared). They indicated that the spectral reflection signature of green plastic cover is similar to that of built-up or bare land, which leads to spectral confusion in image classification (Figure 4), especially for RGB images with only three bands, as in our experiment.

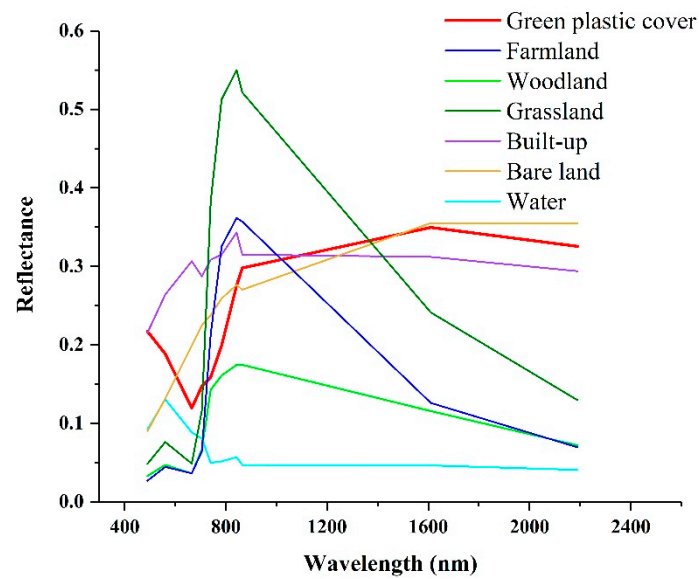


Figure 4. Comparison of spectral characteristics between GPC and other land cover types.

### 3. Methods

#### 3.1. Overview of the Proposed Model

Figure 5 illustrates the overview of the proposed method for green plastic cover mapping. The input is an image patch with 224 rows and 224 columns, and the final result is a predicted land cover class. More specifically, the proposed method consists of two components: (1) Feature extraction based on a deep CNN; and (2) semi-supervised learning that integrates both labeled and unlabeled data. As for the former, we exploited a multi-scale deformable CNN to learn representative spatial features under complex urban landscapes. For the latter, the trained CNN was first utilized to endow the unlabeled data with a pseudo label. Afterwards, the most confident data were selected through top- $k$  ranking and added to the training set to retrain the CNN model.

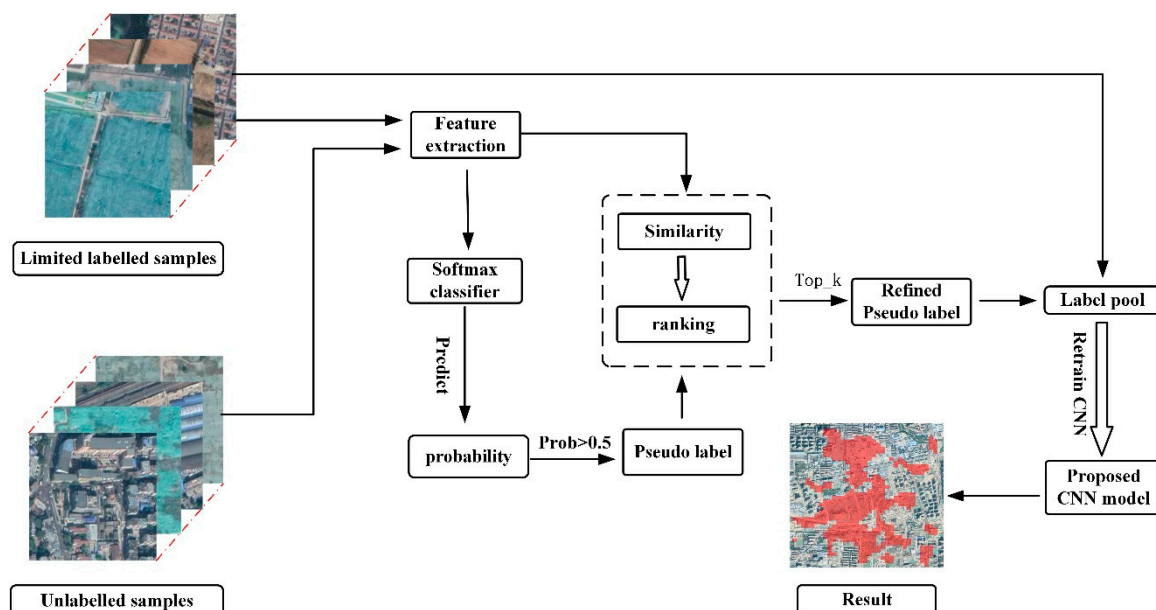
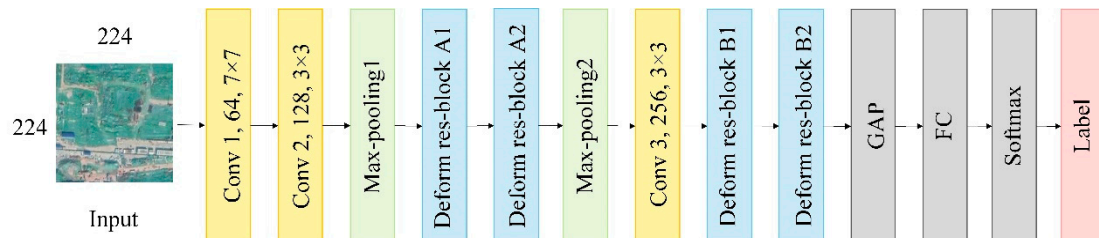


Figure 5. Overview of the proposed method for green plastic cover mapping.

### 3.2. Multi-Scale Deformable CNN for Feature Representation

Figure 6 and Table 1 shows the detailed structure of the multi-scale deformable CNN for deep feature representation. Specifically, it includes several convolutional layers, max pooling layers, and deformable multi-scale residual blocks. Meanwhile, to obtain the final classification result, a global average pooling (GAP), a fully connected (FC) layer, and a Softmax layer were cascaded.



**Figure 6.** Structure of the multi-scale deformable CNN.

**Table 1.** Detailed parameters of the multi-scale deformable CNN.

Name	Input	Output	Kernel Size	Filter Number	Stride
Input	$224 \times 224 \times 3$	—	—	—	—
Conv 1	$224 \times 224 \times 3$	$109 \times 109 \times 64$	$7 \times 7$	64	2
Conv 2	$109 \times 109 \times 64$	$54 \times 54 \times 128$	$3 \times 3$	128	2
Max-pooling1	$54 \times 54 \times 128$	$27 \times 27 \times 128$	—	—	2
Deform res-block A1	$27 \times 27 \times 128$	$27 \times 27 \times 128$	—	—	—
Deform res-block A2	$27 \times 27 \times 128$	$27 \times 27 \times 128$	—	—	—
Max-pooling2	$27 \times 27 \times 128$	$13 \times 13 \times 128$	—	—	2
Conv 3	$13 \times 13 \times 128$	$6 \times 6 \times 256$	$3 \times 3$	256	2
Deform res-block B1	$6 \times 6 \times 256$	$6 \times 6 \times 256$	—	—	—
Deform res-block B2	$6 \times 6 \times 256$	$6 \times 6 \times 256$	—	—	—
GAP	$6 \times 6 \times 256$	$1 \times 1 \times 256$	$6 \times 6$	—	—
FC	256	128	—	—	—
Softmax	128	2	—	—	—

In this study, both deformable convolutions and multi-scale residual blocks were introduced into the deep CNN model for better feature representation. Through deformable convolution, the receptive field and sampling locations were trained to be adaptive to the shapes and scales of land objects, which was beneficial for extracting highly discriminative features. Meanwhile, a multi-scale residual block could extract hierarchical, multi-scale features and improve gradient flow at the same time. In addition, the integration of deformable convolutions into the multi-scale residual block could combine the merits of both modules, increasing the feature adaptability to the complex spatial patterns of urban landscapes. Figure 7 illustrates the detailed parameters of deformable multi-scale residual blocks.

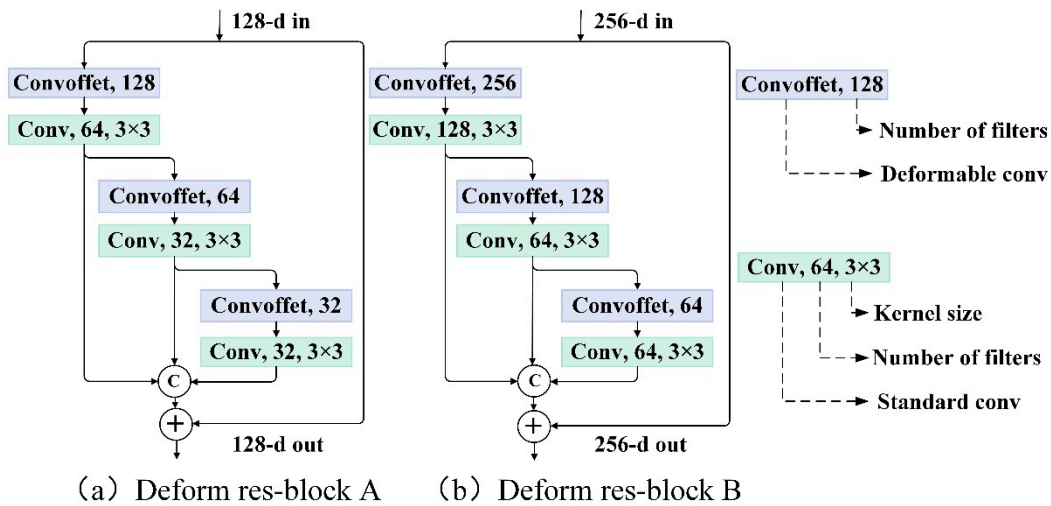


Figure 7. Structure of the deformable multi-scale residual blocks.

Moreover, in our previous study [31], the multi-scale deformable CNN was proposed for spatial feature learning in a coastal wetland landscape, and showed good performance. Therefore, we also introduced it in this study when considering the spatial heterogeneity of complex urban scenarios. More details of the above model can be found in [31].

### 3.3. Samples Selection for Semi-Supervised Learning

The data-driven nature of deep learning calls for a massive number of high-quality labeled samples to maintain the model's generalization capability. However, in the field of remote sensing and geoscience, manually labeling sufficient samples is infeasible due to both the high labor intensity and the low efficiency. Semi-supervised learning, on the other hand, aims to learn from both labeled and unlabeled data, providing a favorable strategy to address the insufficient training data issue, and can achieve satisfactory accuracy with the mining of a massive number of unlabeled samples. Therefore, we resorted to deep semi-supervised learning and proposed a two-step strategy to select the most confident unlabeled samples for model retraining.

Before the description of the two-step strategy for unlabeled samples selection, we first introduce the details of the labeled data. To begin with, we annotated 700 samples for each category, including both GPC and non-GPC, to construct the initial labeled pool. The labeled samples were randomly divided into two parts: 300 for the training set and 400 for the testing set. Meanwhile, 90% of the training set was employed to train the CNN, while the remaining 10% were used as a validation set to evaluate the performance during training.

The proposed two-step strategy for semi-supervised learning was as follows. In the first step, the trained CNN was used to predict samples from the unlabeled pool to derive the posterior probability. Only the unlabeled samples with a probability exceeding 0.5 would be selected and assigned with a predicted category (namely, pseudo-labeled samples). However, these pseudo-labeled samples may be unreliable. If we directly added all these samples into the labeled pool to retrain the CNN model, the performance would not always increase due to additional noise.

To ensure the reliability of the pseudo-labeled samples, we introduced a second step for unlabeled data selection. We calculated the similarities between each pseudo-labeled sample and all labeled samples, which are measured by the Euclidean distance:

$$s(u_i, l_j) = \|f(u_i) - f(l_j)\|_2^{1/2}$$

where  $u_i$  and  $l_j$  denote the  $i$ -th unlabeled and  $j$ -th labeled sample, respectively;  $s(\cdot)$  represents the similarity metric; and  $f(\cdot)$  stands for the deep feature expression. Afterwards, we sorted the labeled



pool by descending order of the above similarities. If the top- $k$  training samples have the same category as the pseudo-labeled sample, then this pseudo-labeled sample was regarded as reliable and could be added to the labeled pool for CNN retraining [29]. In addition, we analyzed the impact of value  $k$  in top- $k$  on GPC classification; the results are shown in Section 4.4.

### 3.4. Details of Network Training

Although the number of training samples could be increased by means of semi-supervised learning, we still adopted the data augmentation technique to further boost the generalization capability and decrease the risk of overfitting. Specifically, all the initial labeled samples were rotated 90, 180, or 270° and flipped up and down.

All the weights of the proposed CNN model were initialized with *He* normalization [35], and all biases were initially set to 0. For optimizing weights and biases to improve classification performance, an Adam optimizer [36] was used with an initial learning rate of  $10^{-4}$ . An early-stopping technique was adopted to select the best model. Cross-entropy loss [37] was adopted, whose expression is as follows:

$$L = \sum_i^N y_i \log(\hat{y}_i)$$

where  $L$  denotes cross-entropy loss;  $\hat{y}_i$  stands for the probability predicted by the model;  $y_i$  denotes the ground truth; and  $N$  refers to the number of classes.

The training procedure included the following steps:

- (1) Firstly, the backbone, i.e., the multi-scale deformable CNN was trained using only the initial labeled data.
- (2) Next, the backbone was utilized to predict the unlabeled datasets, and only the samples that passed the two-step selection process would be added to the labeled pool with pseudo labels.
- (3) The backbone was retrained with samples from the new labeled pool.

In addition, the deep learning library used was TensorFlow [38]. The entire semi-supervised learning framework was conducted on the Ubuntu 18.04 operating system with Intel Xeon(R) Gold 5118 CPU and NVIDIA TITAN V with 12 GB memory.

### 3.5. Accuracy Assessments

After the classification model was trained, a total of 400 testing samples were utilized to calculate the overall accuracy and confusion matrix. The following metrics were also calculated: Producer accuracy (PA), user accuracy (UA), and Kappa coefficient. Meanwhile, visual evaluation was also involved to check for obvious classification errors. In general, visual inspection is a subjective evaluation method that determines whether the classification result is good or not through comparing the green plastic cover mapping results with high-resolution images from Google Earth. Since the green plastic mulch could be identified by eye on Google Earth, we used the visual interpreted images as the “gold standard.” Moreover, we conducted field surveys in several places in Jinan to make sure that the interpreted images were correct.

We also conducted an ablation study to justify the performance of the semi-supervised learning strategy. Furthermore, a comparison with several commonly used CNN models in the computer vision field was performed to evaluate the effectiveness of the multi-scale deformable CNN in this paper.

## 4. Results and Discussion

### 4.1. Classification Results of GPC

After the semi-supervised learning procedure, the trained best model was utilized to classify the entire VHR remote sensing imagery. A sliding window of  $224 \times 224$  was adopted for green plastic

cover prediction. Figure 8 displays the spatial distribution of the GPC prediction results. It could be observed that the green plastic covered regions were mainly located in the eastern part of Jinan, indicating that Jinan has been experiencing urban renewal towards the east. The above remote sensing monitoring results are in accordance with Jinan's urban planning, which verifies the effectiveness of the proposed method in discovering key information on urban renewal.

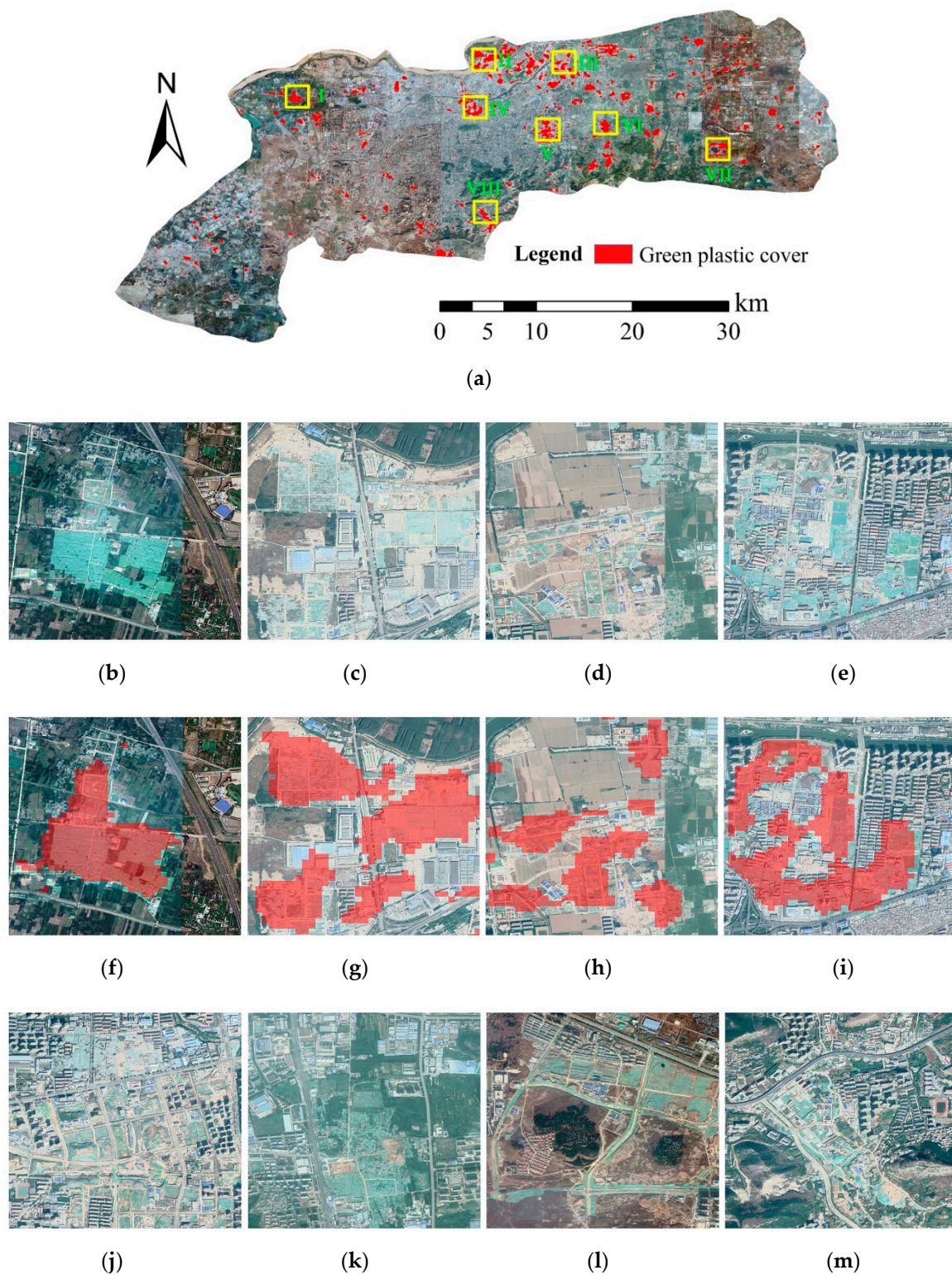
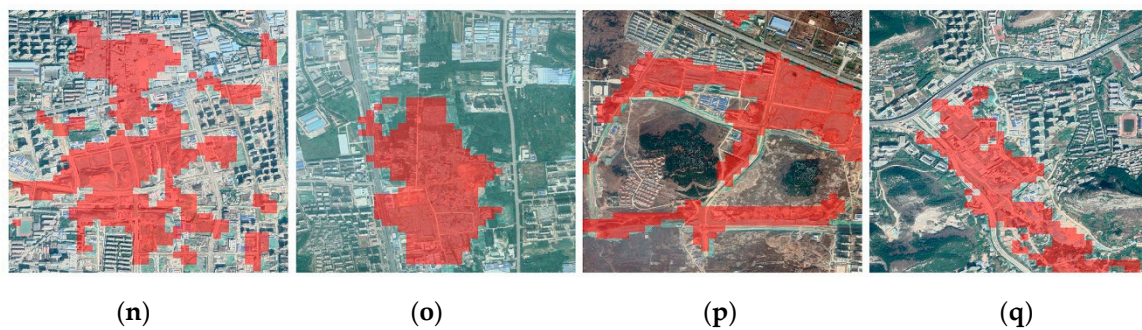


Figure 8. Cont.





**Figure 8.** Mapping results of green plastic cover for (a) the whole study area; (f) sub-region I; (g) sub-region II; (h) sub-region III; (i) sub-region IV; (n) sub-region V; (o) sub-region VI; (p) sub-region VII; and (q) sub-region VIII. True color images for (b) sub-region I; (c) sub-region II; (d) sub-region III; (e) sub-region IV; (j) sub-region V; (k) sub-region VI; (l) sub-region VII; and (m) sub-region VIII.

Figure 8 also illustrates several parts of the original remote sensing imagery and the GPC prediction results. From the sub-regions, it could be seen that the urban landscape is rather complex, with a high spatial heterogeneity. Therefore, the accurate detection of GPC is a challenging task. However, careful visual inspection indicates that the extraction results of green plastic-covered areas have good consistency with the ground truth, justifying the robustness of our proposed method.

#### 4.2. Accuracy Assessment Results

Section 4.1 mainly evaluates the classification results qualitatively from a visual inspection. To further justify the performance, this section adopts a confusion matrix calculated from the testing set to quantitatively evaluate the accuracy of urban green plastic cover mapping. The number of testing samples is 400 for each class. Table 2 lists the accuracy assessment results.

**Table 2.** Confusion matrix.

Class	Ground Truth Test Data		
	GPC	Non-GPC	UA (%)
GPC	374	41	90.12
Non-GPC	26	359	93.25
PA (%)	93.50	89.75	
OA (%)	91.63	Kappa	0.8325

Note: GPC: Green plastic cover; PA: Producer accuracy; UA: User accuracy; OA: Overall accuracy.

Table 2 indicates that the overall accuracy reached 91.63% and the Kappa index reached 0.8325, indicating that the proposed method achieved an excellent performance in urban green plastic cover mapping from VHR remote sensing data. Meanwhile, since we viewed the GPC identification as a remote sensing scene classification task, the patch-based classification and sliding window strategy would result in a serrated boundary, which would lead to extra errors when calculating the total areas of GPC. To tackle this issue, we would like to exploit semantic segmentation methods such as UNet [39] and DeepLab series [40] in future studies to retrieve the exact boundaries of GPC. However, it should be noted that semantic segmentation methods need to vectorize the GPC for training data preparation, which calls for more labor than our proposed method. In this situation, the proposed method could be viewed as a fast, cost-effective yet still reliable way to detect GPC, especially when considering the compromise between workload and accuracy.

#### 4.3. Impact of Semi-Supervised Learning on GPC Classification

To justify the contribution of semi-supervised learning on GPC classification, we conducted an ablation study. Specifically, only the initial 270 labeled samples for each class (GPC and non-GPC)

were utilized to train the classification model. The accuracy was evaluated using the same testing set as that of Section 4.2. The new confusion matrix is as follows.

Table 3 indicates that when using only limited labeled data, the classification performance is inferior to that of semi-supervised learning. The OA only reaches 85.25%, a decrease of 6.38%, while the Kappa index dropped from 0.8325 to 0.7050, a decrease of 0.1275. Therefore, the introduction of semi-supervised learning could improve the classification performance. This is mainly due to the capability of semi-supervised learning to effectively mine the massive unlabeled data. The two-step selection strategy of pseudo labeled data in this study could ensure that the most confidential ones are added to the labeled pool.

**Table 3.** Confusion matrix based on the initial training set.

Class	Ground Truth Test Data		
	GPC	Non-GPC	UA (%)
GPC	352	70	83.41
Non-GPC	48	330	87.30
PA (%)	88.00	82.50	
OA (%)	85.25	Kappa	0.7050

Note: GPC: Green plastic cover; PA: Producer accuracy; UA: User accuracy; OA: Overall accuracy.

#### 4.4. Impact of $k$ in Top- $k$ on GPC Classification

In this section, we analyze the impact of  $k$  in top- $k$  on GPC classification. A series of  $k$  from 45 to 270 with a step of 45 were considered. Due to the fact that the number of GPC samples is much less than that of non-GPC in the unlabeled pool, we first selected a number of  $M$  GPC samples; afterwards, the same number of  $M$  non-GPC samples was also selected. The accuracy assessment results are shown in Table 4.

**Table 4.**  $k$  vs. classification accuracy.

Top- $k$	Pseudo GPC Label	Pseudo Non-GPC Label	OA (%)
45	1246	1246	88.79
90	1029	1029	91.63
135	888	888	89.34
180	791	791	88.51
225	156	156	86.92
270	0	0	85.25

Table 4 indicates that the number of candidate pseudo labeled samples progressively decreased with the increase of  $k$  in top- $k$ . This is understandable since the higher the value of  $k$ , the higher the confidence threshold of these pseudo-labeled samples. When the value of  $k$  is too high, there would be no pseudo labeled samples that would satisfy the selection strategy.

Table 4 also indicates that the GPC classification accuracy is the highest when  $k$  equals 90. This might be due to a compromise between the additional information gain and the introduced noise. When  $k$  is less than the optimal value (90 in this study), there would be more pseudo-labeled samples added into the labeled pool. However, more noise would also be introduced. Meanwhile, when  $k$  increases beyond the optimal value, both the number of pseudo-labeled samples and the accompanying information gain would decrease, leading to a reduction in the classification performance.

#### 4.5. Comparison with Classic CNN Models

To further justify the effectiveness of the proposed model, several classic CNN models in the computer vision field were adopted for comparison, such as VGG [41], ResNet [42], and DenseNet [43].



It should be noted that all the above models were trained using the same semi-supervised learning strategy and evaluated on the same testing set. The comparison results are listed in Table 5.

**Table 5.** Comparison with classical CNN models.

CNN Model	OA (%)	Kappa
VGG	85.87	0.7175
ResNet	86.88	0.7375
DenseNet	89.62	0.7925
Proposed CNN model	91.63	0.8325

Table 5 indicates that the proposed CNN model (multi-scale deformable CNN) achieved the highest accuracy among the four deep learning models. More specifically, VGG had a relatively lower OA (85.87%) in comparison with ResNet (86.88%) and DenseNet (89.62%). This is mainly because VGG utilized a simple cascade of convolutional layers in building its network architecture [41], and has difficulty in extracting highly representative features. Meanwhile, ResNet avoided the gradient vanishing issue in the process of error back-propagation due to the introduction of residual learning and skip connection, which led to higher accuracy. As for DenseNet, its network architecture contained more skip connections for aggregating features and had the best performance. However, in this paper the multi-scale deformable CNN outperformed all the classical CNN models. This could be because the proposed CNN has better adaptability when considering the shape and scale variations of complex urban landscapes.

Furthermore, we compared the above CNN models without the semi-supervised learning strategy, i.e., only the initial limited labeled samples were utilized. The comparison results are in Table 6.

**Table 6.** Comparison with classical CNN models using only limited samples.

CNN Model	OA (%)	Kappa
VGG	83.62	0.6725
ResNet	82.50	0.6500
DenseNet	83.38	0.6675
Proposed CNN model	85.25	0.7050

Similar to Table 5, Table 6 indicates that the proposed CNN model outperformed other backbone networks with an OA of 85.25% and a Kappa index of 0.7050. Therefore, the effectiveness of the proposed CNN in GPC classification was further verified under the condition of limited labeled samples.

#### 4.6. Comparison with Sentinel-2 Data

Since the successful implementation of the European Copernicus program initiated by the European Space Agency (ESA), Sentinel-2 multi-spectral data are now open-access and free to the public, providing new insights for remote sensing applications, such as coastal land cover classification [31], crop mapping [44], and urban areas monitoring [45]. To further justify the performance of the proposed method, we utilized the proposed CNN to detect GPC from Sentinel-2 data. Specifically, the Sentinel-2 L2A data were acquired on 28 August 2019. A total of 10 bands were used in the experiment, including bands 2–4 (10 m), bands 5–7 (20 m), band 8 (10 m), band8a (20 m), and bands 11–12 (20 m). Meanwhile, bands with a 20 m spatial resolution were resampled to 10 m using the SNAP software developed by ESA. Since the image patch of GE data used is  $224 \times 224$  with a spatial resolution of 1.19 m/pixel, to maintain comparability, the image patch of Sentinel-2 data was set to  $27 \times 27$ . Moreover, the same training and testing dataset were used to train and evaluate the model. The accuracy comparison results are listed in Table 7.

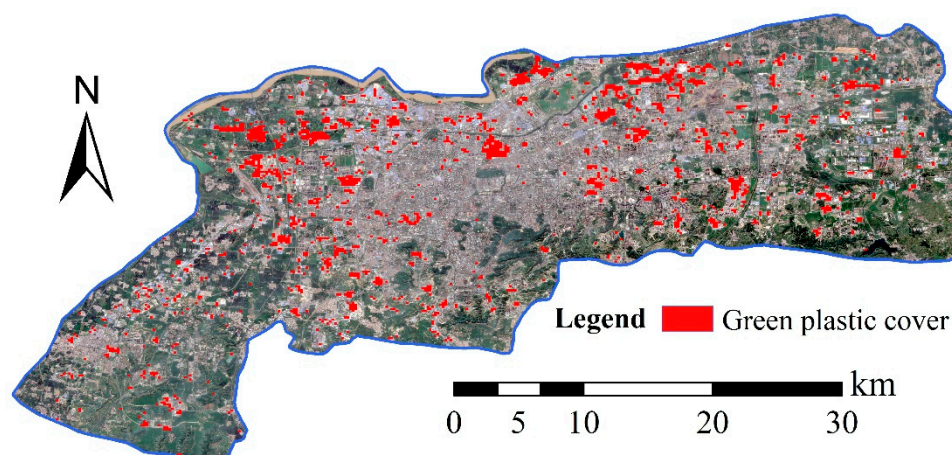
**Table 7.** Comparison with Sentinel-2 data.

Data Source	OA (%)	Kappa
S2	90.87	0.8175
GE	91.63	0.8325

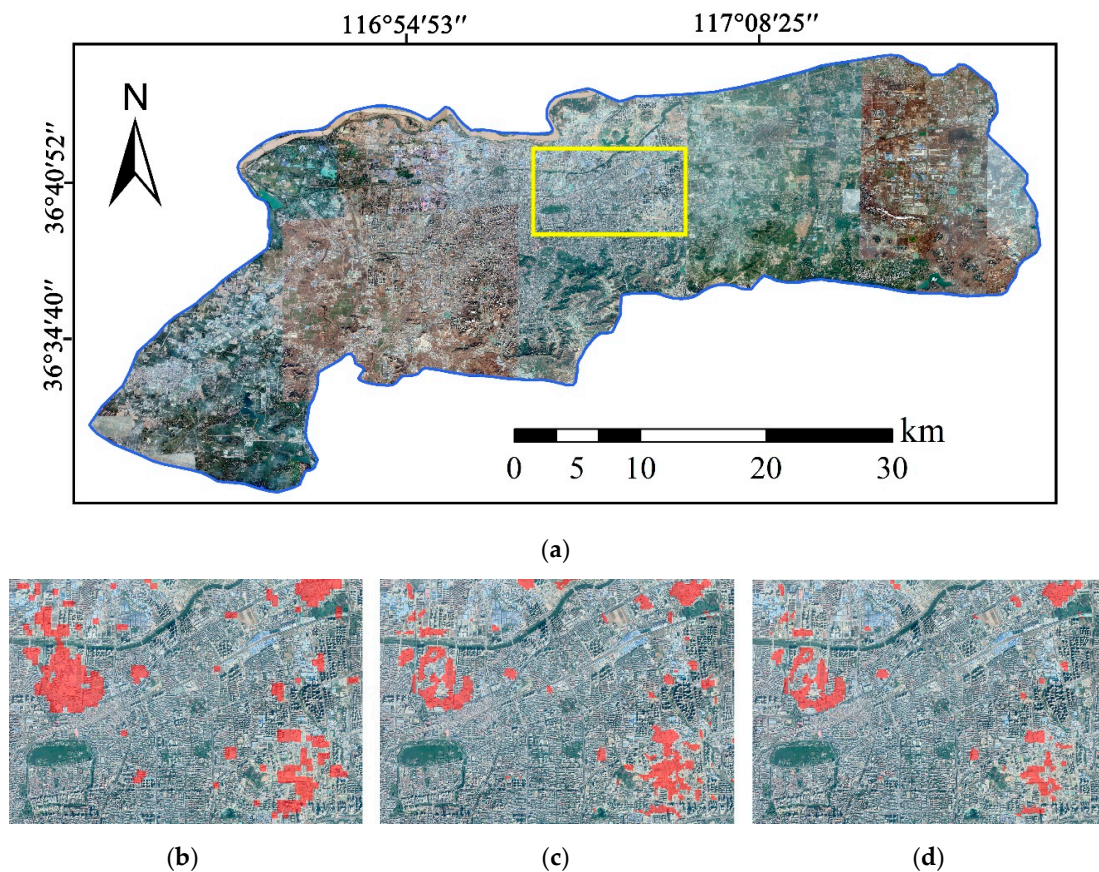
Note: GE: Google Earth data; S2: Sentinel-2 data.

Table 7 indicates that the proposed CNN could yield high performance for both Sentinel-2 and Google Earth data, with an OA of 91.63% and 90.87%, respectively. This further demonstrated that our proposed CNN model has a strong GPC identification ability for either Google Earth data or Sentinel-2 multi-spectral data as the network input.

Figure 9 displays the spatial distribution of the GPC prediction results using Sentinel-2 data. Through a comparison with the GPC prediction results using Google Earth data, it could be observed that the GPC prediction results using these two different images have similar spatial patterns.

**Figure 9.** Mapping results of green plastic cover using Sentinel-2 data.

Now that the GPC classification result from Sentinel-2 is available, it could be used to refine the result from GE data. It should be noted that the entire study region covers a large area of approximate 1015 km<sup>2</sup>, making it difficult to cover the whole region with single-date VHR imagery. Actually, the study area is covered by multi-date VHR datasets from GE. Meanwhile, most GPCs would be replaced by new buildings within a short period, therefore, if we refine the entire classification results from multi-date GE by single-date Sentinel-2, there would be errors from the mismatch of observation dates. In this section, we selected a subset of GE data, whose observation date (23 August 2019) is close to that of Sentinel-2 data (28 August 2019). We then applied a decision level fusion to merge the classification results of GE and Sentinel-2. Only the intersection of GE and Sentinel-2 classification results were maintained to increase the reliability of GPC recognition, which is shown in Figure 10.



**Figure 10.** (a) The whole study area; mapping result of green plastic cover using (b) Sentinel-2 data; (c) Google Earth data; (d) refined result.

#### 4.7. Comparison with Random Forest Classification

Random forest (RF), proposed by Breiman [46], has been widely utilized for land use/land cover mapping in the remote sensing field with improved classification accuracy [12,13,47,48]. To further justify the performance of the proposed CNN, it should be compared with RF. Therefore, RF was trained and tested with the same training and testing samples as the proposed method to maintain fairness. The accuracy comparison results are listed in Table 8.

**Table 8.** Comparison with random forest classification.

Data Source	Method	OA (%)	Kappa
GE	RF	83.87	0.6775
S2	RF	85.75	0.7150
GE	Proposed CNN	91.63	0.8325
S2	Proposed CNN	90.87	0.8175

Note: GE: Google Earth data; S2: Sentinel-2 data.

Table 8 indicates that, compared with RF classification, the proposed CNN could increase the OA by 7.76% and 5.88% for GE and S2 data, respectively. This is mainly because the CNN could extract high-level discriminative features compared with RF, which was beneficial for the improvement of classification accuracy.

## 5. Conclusions

This study proposed a deep semi-supervised learning framework for urban green plastic cover mapping from VHR remote sensing imagery. A multi-scale deformable CNN was exploited for

discriminative feature learning in the complex urban landscapes. A two-step sample selection strategy was proposed for semi-supervised learning to identify the most reliable sample from the unlabeled pool. Experiments and an ablation study were conducted to confirm the good performance of the proposed method.

The experimental results indicate that the proposed method could classify green plastic covered regions in Jinan with a high performance. An accuracy assessment showed that the overall accuracy (OA) was 91.63% and the Kappa index was 0.8325. Moreover, a careful visual inspection showed that most of the green plastic-covered areas could be correctly identified. An ablation study showed that the semi-supervised learning strategy could increase the OA by 6.38% compared with supervised learning, indicating that the mining of the most confidential unlabeled data could effectively improve the classification accuracy. Meanwhile, the comparison with several classic CNN models in the computer vision field showed that the multi-scale deformable CNN in this study yielded the highest accuracy, justifying its effectiveness for spatial feature learning in complex urban landscapes.

Moreover, this study is the first attempt to identify green plastic cover from VHR remote sensing data based on deep learning methods, which could provide a baseline for relevant studies. Although the proposed CNN is now utilized for urban plastic-covered region recognition, it could also be applied to other applications, such as remote sensing scene understanding. In future work, we will further justify the model's effectiveness and use semantic segmentation models to derive the exact boundaries of the green plastic covered regions.

**Author Contributions:** Methodology, Jiantao Liu and Quanlong Feng; validation, Jiantao Liu and Ying Wang; data curation, Jiantao Liu; writing—original draft preparation, Jiantao Liu and Quanlong Feng; writing—review and editing, Ying Wang, Bayartungalag Batsaikhan, Jianhua Gong, Yi Li, Chunting Liu, and Yin Ma. All authors have read and agreed to the published version of the manuscript.

**Funding:** We acknowledge grants from the National Key Research and Development Program of China (grant no. 2018YFE0122700) and the Doctoral Research Fund of Shandong Jianzhu University (XNBS1903).

**Acknowledgments:** The authors would like to thank the Google Earth and European Space Agency for providing high spatial resolution image and Sentinel-2 data. Additionally, the authors would like to give special thanks to the anonymous reviewers and editors for their very useful comments and suggestions to help improve the quality of this paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Liu, Y.; Zhu, A.X.; Wang, J.; Li, W.; Hu, G.; Hu, Y. Land-use decision support in brownfield redevelopment for urban renewal based on crowdsourced data and a presence-and-background learning (PBL) method. *Land Use Policy* **2019**, *88*, 104188. [[CrossRef](#)]
2. Xia, C.; Zhang, A.; Yeh, A.G.-O. Shape-weighted landscape evolution index: An improved approach for simultaneously analyzing urban land expansion and redevelopment. *J. Clean. Prod.* **2020**, *244*, 118836. [[CrossRef](#)]
3. Yu, B.; Wang, J.; Li, J.; Zhang, J.; Lai, Y.; Xu, X. Prediction of large-scale demolition waste generation during urban renewal: A hybrid trilogy method. *Waste Manag.* **2019**, *89*, 1–9. [[CrossRef](#)] [[PubMed](#)]
4. Shackelford, A.K.; Davis, C.H. A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2354–2364. [[CrossRef](#)]
5. Zhou, W.; Troy, A. An object-oriented approach for analysing and characterizing urban landscape at the parcel level. *Int. J. Remote Sens.* **2008**, *29*, 3119–3135. [[CrossRef](#)]
6. Bhaskaran, S.; Paramananda, S.; Ramnarayan, M. Per-pixel and object-oriented classification methods for mapping urban features using Ikonos satellite data. *Appl. Geogr.* **2010**, *30*, 650–665. [[CrossRef](#)]
7. Myint, S.W.; Gober, P.; Brazel, A.; Grossman-Clarke, S.; Weng, Q. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sens. Environ.* **2011**, *115*, 1145–1161. [[CrossRef](#)]



8. Wang, H.; Wang, C.; Wu, H. Using GF-2 Imagery and the Conditional Random Field Model for Urban Forest Cover Mapping. *Remote Sens Lett.* **2016**, *7*, 378–387. [\[CrossRef\]](#)
9. Bialas, J.; Oommen, T.; Havens, T.C. Optimal segmentation of high spatial resolution images for the classification of buildings using random forests. *Int. J. Appl. Earth Obs.* **2019**, *82*, 101895. [\[CrossRef\]](#)
10. Shao, Z.; Fu, H.; Fu, P.; Yin, L. Mapping Urban Impervious Surface by Fusing Optical and SAR Data at the Decision Level. *Remote Sens.* **2016**, *8*, 945. [\[CrossRef\]](#)
11. Yin, W.; Yang, J. Sub-pixel vs. super-pixel-based greenspace mapping along the urban–rural gradient using high spatial resolution Gaofen-2 satellite imagery: A case study of Haidian District, Beijing, China. *Int. J. Remote Sens.* **2017**, *38*, 6386–6406. [\[CrossRef\]](#)
12. Feng, Q.; Liu, J.; Gong, J. UAV Remote Sensing for Urban Vegetation Mapping Using Random Forest and Texture Analysis. *Remote Sens.* **2015**, *7*, 1074–1094. [\[CrossRef\]](#)
13. Feng, Q.; Liu, J.; Gong, J. Urban Flood Mapping Based on Unmanned Aerial Vehicle Remote Sensing and Random Forest Classifier—A Case of Yuyao, China. *Water* **2015**, *7*, 1437–1455. [\[CrossRef\]](#)
14. Liu, J.; Li, P.; Wang, X. A new segmentation method for very high resolution imagery using spectral and morphological information. *ISPRS J. Photogramm. Remote Sens.* **2015**, *101*, 145–162. [\[CrossRef\]](#)
15. Ruiz Hernandez, I.E.; Shi, W. A Random Forests classification method for urban land-use mapping integrating spatial metrics and texture analysis. *Int. J. Remote Sens.* **2017**, *39*, 1175–1198. [\[CrossRef\]](#)
16. Yu, B.; Wang, L.; Niu, Z.; Tappert, M.C. Unsupervised building extraction using remote sensing data to detect changes in land use. In Proceedings of the Spie Asia-Pacific Remote Sensing, Land Surface Remote Sensing II, Beijing, China, 10–13 October 2014; Volume 9260, p. 926007.
17. Silvestri, S.; Omri, M. A method for the remote sensing identification of uncontrolled landfills: Formulation and validation. *Int. J. Remote Sens.* **2007**, *29*, 975–989. [\[CrossRef\]](#)
18. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
19. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 580–587.
20. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans. Pattern. Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [\[CrossRef\]](#)
21. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans. Pattern. Anal. Mach. Intell.* **2016**, *38*, 142–158. [\[CrossRef\]](#)
22. Lin, G.; Shen, C.; van den Hengel, A.; Reid, I. Efficient Piecewise Training of Deep Structured Models for Semantic Segmentation. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 July 2016; pp. 3194–3203.
23. Huang, B.; Zhao, B.; Song, Y. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sens. Environ.* **2018**, *214*, 73–86. [\[CrossRef\]](#)
24. Zhang, C.; Sargent, I.; Pan, X.; Li, H.; Gardiner, A.; Hare, J.; Atkinson, P.M. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sens. Environ.* **2018**, *216*, 57–70. [\[CrossRef\]](#)
25. Dong, L.; Du, H.; Mao, F.; Han, N.; Li, X.; Zhou, G.; Zhu, D.E.; Zheng, J.; Zhang, M.; Xing, L.; et al. Very High Resolution Remote Sensing Imagery Classification Using a Fusion of Random Forest and Deep Learning Technique—Subtropical Area for Example. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 113–128. [\[CrossRef\]](#)
26. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Trans. Geosci. Remote Sens Lett.* **2017**, *14*, 778–782. [\[CrossRef\]](#)
27. Alshehhi, R.; Marpu, P.R.; Woon, W.L.; Mura, M.D. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 139–149. [\[CrossRef\]](#)
28. Xu, X.; Li, W.; Ran, Q.; Du, Q.; Gao, L.; Zhang, B. Multisource Remote Sensing Data Classification Based on Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 937–949. [\[CrossRef\]](#)
29. Tong, X.-Y.; Xia, G.-S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [\[CrossRef\]](#)

30. Feng, Q.; Zhu, D.; Yang, J.; Li, B. Multisource Hyperspectral and LiDAR Data Fusion for Urban Land-Use Mapping based on a Modified Two-Branch Convolutional Neural Network. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 28. [CrossRef]
31. Feng, Q.; Yang, J.; Zhu, D.; Liu, J.; Guo, H.; Bayartungalag, B.; Li, B. Integrating Multitemporal Sentinel-1/2 Data for Coastal Land Cover Classification Using a Multibranch Convolutional Neural Network: A Case of the Yellow River Delta. *Remote Sens.* **2019**, *11*, 1006. [CrossRef]
32. He, Z.; Liu, H.; Wang, Y.; Hu, J. Generative Adversarial Networks-Based Semi-Supervised Learning for Hyperspectral Image Classification. *Remote Sens.* **2017**, *9*, 1042. [CrossRef]
33. Fang, B.; Li, Y.; Zhang, H.K.; Chan, J.C.W. Semi-Supervised Deep Learning Classification for Hyperspectral Image Based on Dual-Strategy Sample Selection. *Remote Sens.* **2018**, *10*, 574. [CrossRef]
34. Google Earth. Available online: <http://earth.google.com/> (accessed on 2 May 2020).
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. *arXiv* **2015**, arXiv:1502.01852.
36. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
37. Cox, D.R. The Regression Analysis of Binary Sequences. *J. R. Stat. Soc. Ser. B.* **1958**, *20*, 215–242. [CrossRef]
38. TensorFlow. Available online: <https://tensorflow.google.cn/> (accessed on 7 March 2020).
39. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
40. Chen, L.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal.* **2018**, *40*, 834–848. [CrossRef]
41. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
42. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
43. Huang, G.; Liu, Z.; Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
44. Wang, J.; Xiao, X.; Liu, L.; Wu, X.; Qin, Y.; Steiner, J.L.; Dong, J. Mapping sugarcane plantation dynamics in Guangxi, China, by time series Sentinel-1, Sentinel-2 and Landsat images. *Remote Sens. Environ.* **2020**, *247*, 111951. [CrossRef]
45. Lefebvre, A.; Sannier, C.; Corpetti, T. Monitoring Urban Areas with Sentinel-2A Data: Application to the Update of the Copernicus High Resolution Layer Imperviousness Degree. *Remote Sens.* **2016**, *8*, 606. [CrossRef]
46. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
47. Liu, J.; Feng, Q.; Gong, J.; Zhou, J.; Li, Y. Land-cover classification of the Yellow River Delta wetland based on multiple end-member spectral mixture analysis and a Random Forest classifier. *Int. J. Remote Sens.* **2016**, *37*, 1845–1867. [CrossRef]
48. Liu, J.; Feng, Q.; Gong, J.; Zhou, J.; Liang, J.; Li, Y. Winter wheat mapping using a random forest classifier combined with multi-temporal and multi-sensor data. *Int. J. Digit. Earth* **2017**, *11*, 783–802. [CrossRef]

