remote sensing

MDPI

Article

WTS: A Weakly towards Strongly Supervised Learning Framework for Remote Sensing Land Cover Classification Using Segmentation Models

Wei Zhang [1,2], Ping Tang [1], Thomas Corpetti [3] and Lijun Zhao [1,*]

1 National Engineering Laboratory for Satellite Remote Sensing Applications (NELRS), Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; zhangwei@aircas.ac.cn (W.Z.); tangping@aircas.ac.cn (P.T.)
2 University of Chinese Academy of Sciences, Beijing 100049, China
3 CNRS, UMR 6554 LETG COSTEL, 35000 Rennes, France; thomas.corpetti@univ-rennes2.fr
* Correspondence: zhaolj201934@aircas.ac.cn; Tel.: +86-010-64855178

Abstract: Land cover classification is one of the most fundamental tasks in the field of remote sensing. In recent years, fully supervised fully convolutional network (FCN)-based semantic segmentation models have achieved state-of-the-art performance in the semantic segmentation task. However, creating pixel-level annotations is prohibitively expensive and laborious, especially when dealing with remote sensing images. Weakly supervised learning methods from weakly labeled annotations can overcome this difficulty to some extent and achieve impressive segmentation results, but results are limited in accuracy. Inspired by point supervision and the traditional segmentation method of seeded region growing (SRG) algorithm, a weakly towards strongly (WTS) supervised learning framework is proposed in this study for remote sensing land cover classification to handle the absence of well-labeled and abundant pixel-level annotations when using segmentation models. In this framework, only several points with true class labels are required as the training set, which are much less expensive to acquire compared with pixel-level annotations through field survey or visual interpretation using high-resolution images. Firstly, they are used to train a Support Vector Machine (SVM) classifier. Once fully trained, the SVM is used to generate the initial seeded pixel-level training set, in which only the pixels with high confidence are assigned with class labels whereas others are unlabeled. They are used to weakly train the segmentation model. Then, the seeded region growing module and fully connected Conditional Random Fields (CRFs) are used to iteratively update the seeded pixel-level training set for progressively increasing pixel-level supervision of the segmentation model. Sentinel-2 remote sensing images are used to validate the proposed framework, and SVM is selected for comparison. In addition, FROM-GLC10 global land cover map is used as training reference to directly train the segmentation model. Experimental results show that the proposed framework outperforms other methods and can be highly recommended for land cover classification tasks when the pixel-level labeled datasets are insufficient by using segmentation models.

Keywords: land cover classification; convolutional neural network; segmentation model; weakly supervised; seeded region growing; fully connected Conditional Random Fields

Citation: Zhang, W.; Tang, P.; Corpetti, T.; Zhao, L. WTS: A Weakly towards Strongly Supervised Learning Framework for Remote Sensing Land Cover Classification Using Segmentation Models. *Remote Sens.* **2021**, *13*, 394. https://doi.org/10.3390/rs13030394

Academic Editor: Chiman Kwan
Received: 21 December 2020
Accepted: 20 January 2021
Published: 23 January 2021

1. Introduction

Land cover classification of remote sensing images plays an incredibly important role in the study of ecological environment change, disaster recovery, urban planning or precision agriculture [1,2]. With the development of remote sensing technology, we have access to massive remote sensing databases that no manual method could handle, such as the USGS (United States Geological Survey) Earth Explorer, ESA (European Space Agency) Sentinel Mission or CHEOS (China High-resolution Earth Observation System). Therefore,

developing reliable and efficient methods for automatic land cover classification of these images is of prime importance.

A variety of algorithms have been introduced for land cover classification, including support vector machine (SVM) [3], random forest [4] and artificial neural networks [5]. These methods have achieved better classification results, but today they fail to reach the state-of-the-art performance, mainly because of their limited representation capability compared to feature learning approaches. In recent years, deep learning models, and especially convolutional neural networks (CNNs), have lead feature learning into a new era in computer vision [6–8]. Numerous attempts have been made to introduce CNNs into the field of remote sensing such as scene classification [9–11], object extraction [12,13] and change detection [14,15] as well as land cover classification [16–19]. These kind of methods improve the land cover classification results. However, they ignore the relationship betweenpatches, and boundary and outline distortions always exist among land covers of classification results [20]. Meanwhile, the problems of redundant re-computation [21] and low efficiency result in difficulty in efficient land cover classification.

The emergence of fully convolutional network (FCN) [22] overcomes the problems of CNN-based methods and has become a powerful and promising scheme in the field of semantic segmentation. Based on FCN, many semantic segmentation models have been proposed for remote sensing classification in recent years and have obtained sate-of-the-art performances compared with traditional methods [23–26]. Different from the image-level training samples for CNN-based methods in which one single class label is assigned to the whole input image, the training samples of segmentation models for the dense classification task are pixel-level, in which the input image has the same size as the reference data, and each pixel in the input image should be assigned to a class label. Collecting large-scale accurate pixel-level annotation becomes time-consuming and typically requires substantial financial investments.Even though there are abundant well-annotated datasets such as ISPRS benchmark [27], DeepGlobe [28], SEN12MS [29] and GID [30], which can provide great convenience for the research of land cover classification. When mapping a new type of remote sensing images in real-world applications, these annotations are out of operation. Fortunately, the existing numerous large-scale or even global land cover maps can provide adequate reference data and make it possible for land cover classification using segmentation models. For example, Isikdogan [31] used the global land cover facility (GLCF) [32] product as reference data to train a deep fully convolutional neural network model for water body mapping, and it performed significantly better than traditional approaches. In the work of Scepanovic [33], coordination of information on the environment (CORINE) land cover data was used as reference for land cover mapping with sentinel-1 SAR (Synthetic Aperture Radar) imagery. Chantharaj [34] used the dataset from Geo-Informatics and Space Technology Development Agency as true labels to train segmentation models for the classification of Landsat-8 remote sensing images. However, the resolutions of those maps typically range from 30 m to 1000 m per pixel [35], which may be not consistent with the classified images, and the classification system in practical applications may be different from that of those maps. In addition, a large number of noisy labels are potentially available in those products, which will have negative influences on the classification performance. Many methods, such as formulating robust loss function [36], adding a noising layer in the neural network to learning the noise distribution [37] or "co-teaching" robust training paradigm [38], have been proposed for robust learning from noisy data, but they are mainly concentrated on the machine learning research. Literature related to robust learning from noisy data for remote sensing land cover classification using segmentation models is still rather scarce. Two robust loss functions are proposed to deal with omission noise and registration noise in [39] for road detection from aerial images. However, they are not suitable for dealing with more complex noise distribution in the land cover classification. All these hinder the wide range of applications when using existing land cover maps as reference data for land cover classification.

Recently, weakly supervised learning has become a promising direction due to its need for only weakly labeled or even unlabeled data, which can be easily collected in large amounts and significantly reduce manual labeling. Many forms of weakly supervision are explored in the machine learning community, such as image-level labels [40], point-level [41], bounding box [42], scribbles [43], etc. Inspired by these techniques in machine learning, many image-level weakly supervised methods have been introduced into remote sensing classification research due to its significantly less annotation effort. In paper [44], the authors used the mainstream weakly supervised semantic segmentation methodology developed in natural scene images to map satellite images. They, however, achieved poor performance, and more work is needed for developing alternative methodologies to generalize them to satellite images. Considering the difference between computer vision datasets and remote sensing ones, a weakly supervised feature-fusion network was proposed in [45] for binary segmentation of remote sensing images and achieved comparable results to fully supervised methods only using image-level annotations. A hierarchical weakly supervised learning method was designed in [46] for pixel-level semantic residential area extraction in remote sensing images based on image-level labels, and results showed the superiority of the proposed method. Due to the absence of localization information, image-level supervised learning can hardly reach the performance of fully supervised methods. At the same time, the image-level labels also are needed to determine the presence or absence of classes in every training sample, which are still time-consuming, especially for remote sensing images.

Inspired by the point supervision [41] and selecting RoI (Region of Interest) as training set for training traditional machine learning methods, points with true class labels are selected as the training set in this paper for remote sensing land cover classification using semantic segmentation methods. They can be more easily acquired through field survey or visual interpretation using high-resolution images compared with pixel-level annotations and image-level datasets. A weakly towards strongly (WTS) supervised learning framework is proposed to better exploit these labeled points for remote sensing image classification. In short, to describe the proposed framework, a points training set is first used to generate the initial seeded pixel-level training set using Support Vector Machine (SVM). Then, the initial seeded training set is used to train the segmentation model. Once fully trained, the seeded region growing (SRG) [47] module and the fully connected Conditional Random Field (CRF) are used to progressively update these seeded training sets. Alternatively, the processes of training the segmentation model and updating seeded training set are performed for progressively refining pixel-level supervision of the segmentation model. Figure 1 presents the dynamic evolution of one training sample in seeded training set of the WTS framework. In summary, the superiority of the proposed WTS framework is indicated by the following:

1. Easy implementation. As is well known, a large annotated dataset is indispensable for deep learning research. In this study, pixel-level annotations are required for training semantic segmentation models, which is prohibitively expensive and laborious, especially in the field of remote sensing. However, only several point samples with true class labels as training set are needed in the proposed WTS framework. They can be easily acquired through field survey or visual interpretation using high-resolution images, which makes the land cover classification easy to implement when using segmentation models.
2. High flexibility. Because of the absence of abundant well-annotated datasets, using current large-scale or global land cover classification products as reference data is a reliable solution. However, the land cover classification system is fixed in these products, and some classes are not included in them when facing some practical applications. In the proposed WTS framework, we can select the training samples according to the pre-defined classification system, which can improve the flexibility of our framework.

3. High accuracy. In the generation of the initial seeded pixel-level training set using SVM, only pixels with high confidence are assigned with class labels, and then they are used to train the segmentation model. Furthermore, the SRG module and the fully connected CRF are used to progressively update training set for gradually optimizing their quality. All these make our framework achieve excellent classification performance.

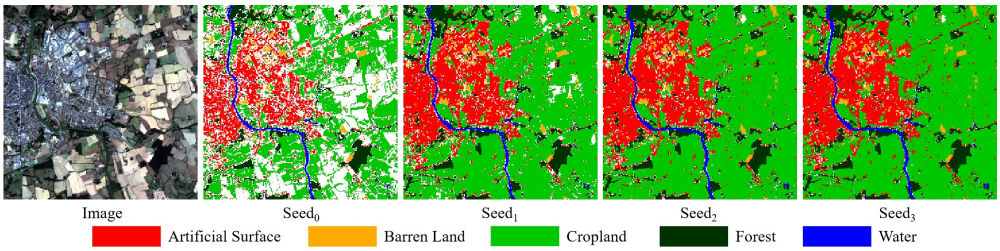


Figure 1. The dynamic evolution of one training sample in several iterations of the proposed weakly towards strongly (WTS) framework. It can be found that the quality of the training sample is gradually improved with the optimization process. ($Seed_i$ denotes the seeded training set of i iteration, while $Seed_0$ denotes the initial seeded training set. The white areas represent the unlabeled points).

The rest of this paper is structured as follows. The study area and experimental data are described in Section 2. Section 3 illustrates the proposed WTS framework in detail. Section 4 presents the experimental setup and the comparison of classification results. The influences of the experimental setting on classification results are analyzed in Section 5. Finally, Section 6 provides the conclusion and the future work.

2. Materials

2.1. Study Area and Remote Sensing Data

The study area is located in the region of northwest France. To cover the study area, two Sentinel-2B level-2A remote sensing images on 19 September 2019 were selected as the experimental data. As shown in Figure 2, the study area was divided into three parts for training, validation and testing of land cover classification methods. Sentinel-2B is one of two Sentinel-2 satellites and carries a multispectral instrument (MSI) with 13 spectral channels in the visible, near infrared (VNIR) and short wave infrared spectral range (SWIR) at 10 m, 20 m and 60 m spatial resolution. Table 1 describes the detailed parameters of the Sentinel-2B bands used in this study. The bands with 10 m spatial resolution are all re-sampled into 20 m using bilinear interpolation for consistency with the other bands.

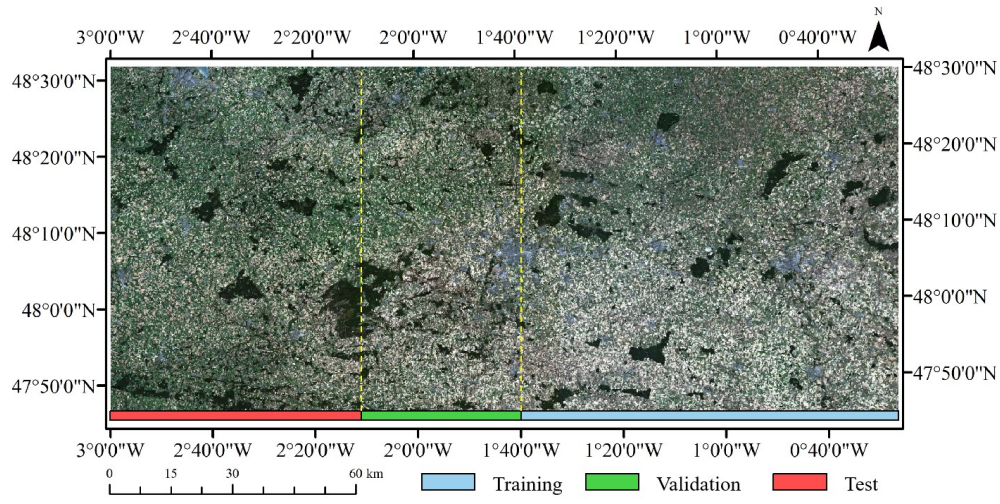


Figure 2. Overview of the study area and location of areas for training, validation and testing. (The base is Sentinel-2B band 4,3,2 true color composite image).

Table 1. Technical specification of Sentinel-2B bands used in this study.

Band Number	Spectral Region	Central Wavelength (nm)	Bandwidth (nm)	Spatial Resolution (m)
2	Blue	492.1	98	10
3	Green	559	46	10
4	Red	665	39	10
5	Vegetation Red Edge	703.8	20	20
6	Vegetation Red Edge	739.1	18	20
7	Vegetation Red Edge	779.7	28	20
8	NIR	833	133	10
8a	Vegetation Red Edge	864	32	20
11	SWIR	1610.4	141	20
12	SWIR	2185.7	238	20

2.2. Points Training Set

In this study, five classes including artificial surface, barren land, cropland, forest and water were defined as the land cover classification scheme. A total of 16,844 points were assigned with true class labels as training set by visual interpretation and viewing the high-resolution images from Google Earth. Detailed descriptions of training samples are illustrated in Table 2.

Table 2. Land cover classification scheme and training set size.

Class	Short Description	Number of Points	Area (km^2)
Artificial Surface	Artificial covers such as urban areas, rural cottages and roads.	3332	1.3328
Barren Land	Surface vegetation is hardly observable, such as urban areas with little constructed material, bare mines and beaches.	3309	1.3236
Cropland	Human planted land that generally has regular distribution patterns including cultivated land and fallow land.	3726	1.4904
Forest	Trees observable in the landscape, such as broadleaf forest, needleleaf forest and shrubland	3467	1.3868
Water	Water bodies such as rivers, lakes, reservoirs and ponds.	3010	1.2040

3. Methodology

In this section, the details of the proposed weakly towards strongly supervised learning framework are given. Firstly, we introduce general steps of the WTS framework. Then the initial seed generation, segmentation model, seeded loss, fully connected CRF and seeded region growing in the WTS framework are described in detail.

The overview of the proposed weakly towards strongly supervised learning framework is illustrated in Figure 3, and general steps can be described as follows.

(1) Initial seed generation: Use points training set (as described in Section 2.2) to generate the initial seeded pixel-level data set (denoted as $seed_0$) including training set and validation set using SVM, in which only confident points are treated as seed points.
(2) Train the segmentation model: Use $seed_i$ ($seed_0$ when firstly training) to train the segmentation model, and seeded loss is used to update the model parameters.
(3) Update seed: Take images of $seed_i$ as the input of the fully trained segmentation model from Procedure (2) to produce the probability maps, then the fully connected CRF and SRG are used to update $seed_i$ to get the updated $seed_{i+1}$ based on the input images and output probability maps.
(4) Iterate until convergence: Treat $seed_{i+1}$ as a new data set to iterate Procedures (2) and (3) until seed points within the data set no longer change.
(5) Classification stage: Use the final trained segmentation model to classify test images to get the classification results.

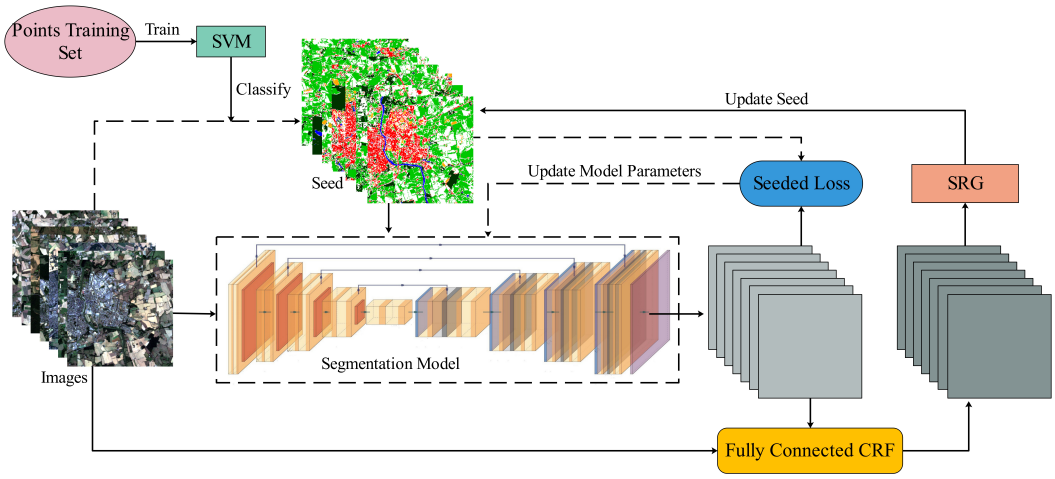


Figure 3. Overview of the proposed weakly towards strongly supervised learning framework.

3.1. Initial Seed Generation Using Points Training Set

As patches are required for fully convolutional semantic-segmentation trainings, single training points can not be used as is. To deal with this, SVM is selected to transform the points training set into a "patch-patch" pixel-level data set, which is defined as the initial seed in this paper. The procedures of the initial seed generation are shown in Figure 4. Firstly, the points training set is used to train SVM. Then, patch images with size of 256×256 are clipped from training/validation images and fed into the fully trained SVM to get the class probability maps, which are computed based on the isotonic regression. In order to ensure the diversity of training samples, patch images are clipped by two ways: clipping by sliding the patch window with no overlap and clipping randomly in training/validation area. Finally, a probability threshold is defined for filtering pixels of the output class probability maps to get the initial seed. If the maximum class probability is higher than the threshold, the pixel is defined as seed point and is assigned as the corresponding class label; otherwise the pixel is treated as the unlabeled point. Note that the patch images with no seed points are not considered. The probability threshold determines the sparsity and quality of seed points and is a vital hyper-parameter in this study. Its influence on the classification results will be analyzed in Section 5.1. To sum up, 10,000 and 2500 "patch-patch" samples are generated separately as the initial training seed and initial validation seed from the training area and validation area. In order to get a robust classification result, the initial seed generation is repeated five times in parallel to get five different training/validation sets. The accuracy evaluation results in this paper are obtained by averaging the results of five parallel experiments. In addition, as we know, the diversity of training samples is one of the most important factors that influence classification results. The study area in this paper is relatively small, and spectral distribution differences in the study area are not obvious. Thus, clipping training samples only from the training area is enough to ensure the diversity of training samples. This is different from dealing with large territories because of the big spectral distribution differences among different areas. In this case, the selecting and clipping operations of patch images should be evenly distributed over the whole large study area.

3.2. Semantic Segmentation Model

In order to achieve dense prediction, FCN [22] was proposed, which is a modification of the CNN architecture and has made promising improvements in the performance of semantic segmentation. In the FCN, all fully connected layers are replaced by convolutional layers. This modification enables the model to take inputs of any arbitrary size and produce corresponding-sized output instead of a single label with efficient inference and learning. FCN is the pioneering work of semantic segmentation, which defines a general framework for dense pixel-wise prediction. Based on the FCN, various semantic segmen-

tation models have been designed to improve segmentation performance in recent years, such as SegNet [48], DeepLab [49], U-Net [50] and FC-DenseNet [51]. All of them aim to extract and combine multi-scale context information or enhance feature discriminability for implementing precise segmentation. Considering the popularity in the remote sensing field, U-Net is selected as the segmentation model to be studied in this paper, and the architecture of U-Net is shown in Figure 5.

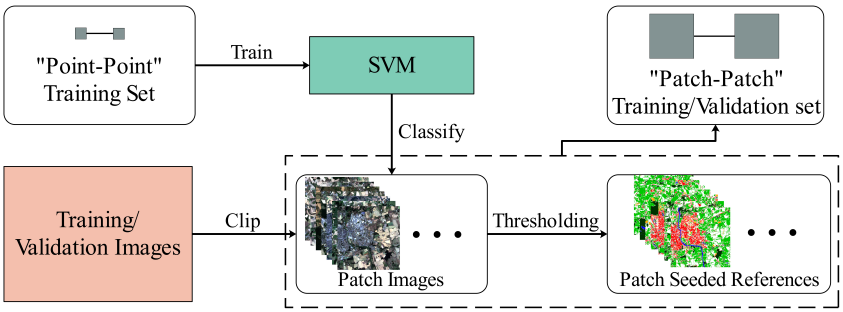


Figure 4. Procedures of the initial seed generation.

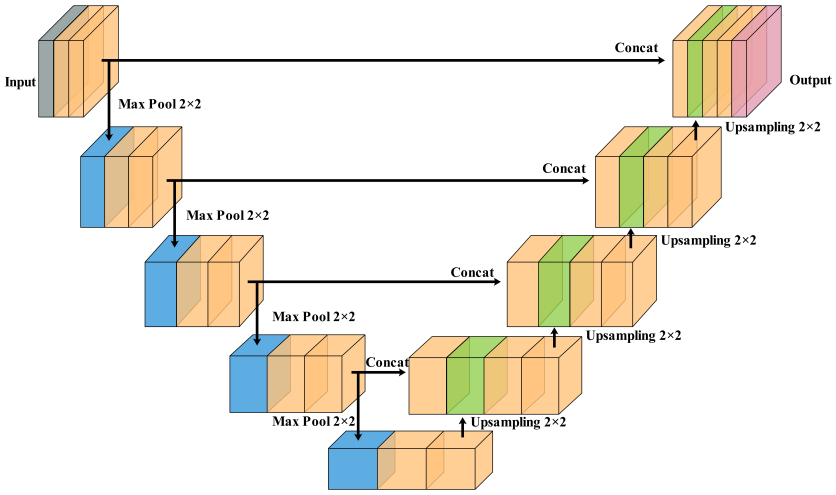


Figure 5. The architecture of U-Net.

The U-Net stems from the FCN model but was modified in a way that it yields better segmentation in medical images. As shown in Figure 5, this architecture is symmetric and consists of three sections—encoder, bottleneck and decoder—which gives it the U-shaped network. The encoder converts the input image into compact representation by many contraction blocks. The bottleneck plays a role of the bond between encoder and decoder. The core of this architecture lies in the decoder, which recovers the representation to a pixel-wise classification output with the same size as the input image. Similar to encoder, it also consists of several expansion blocks. In addition, the skip connections are applied between the encoder and the decoder to provide local information to the global high-level features while upsampling. It is worth noting that the cropping operation in original U-Net was not used in this study.

3.3. Seeded Loss

Because many pixels in the seeded training set are unlabeled, the seeded loss [52] is used to guide the weakly supervised learning of segmentation models, for only matching the seed points while ignoring the rest pixels of the image. The seeded loss could be defined

as a cross-entropy between the seeded annotations and the probability maps generated by the segmentation model, and the formula is as follows.

$$l_{seeded} = -\frac{1}{\sum_{c\in C}|S_c|}\sum_{c\in C}\sum_{u\in S_c}\log p_{u,c} \quad (1)$$

where C is the class set used in this study, S_c is a location set of seed points of class c and $p_{u,c}$ is the probability value of the pixel of class c at position u.

3.4. Fully Connected CRF

In the training phase of the segmentation model, the seeded loss was used to optimize the prediction, resulting in high accuracy in the seed points but low confidence in other regions. To this end, the fully connected CRF [53] was firstly used in the phase of updating seed to optimize the output probability maps of the segmentation model. Fully connected CRF is a graphical model and has been successfully used in the semantic segmentation task due to its qualitative and quantitative performance to improve localization. Suppose that the x is the class assignment for pixels, the following energy function is employed in the fully connected CRF model:

$$E(x) = \sum_i \psi_u(x_i)\sum_{ij}\psi_p(x_i, x_j) \quad (2)$$

The $\psi_u(x_i)$ is the unary potential and is computed as $\psi_u(x_i) = -\log p_{x_i}$, where p_{x_i} is the class probability at pixel i of the segmentation model output. Function $\psi_p(x_i, x_j)$ is the pairwise potential and has the form $\psi_p(x_i, x_j) = \mu(x_i, x_j)\sum_{m=1}^{K}\omega_m k_m(f_i, f_j)$, where $\mu(x_i, x_j) = 1$ if $i \neq j$, otherwise $\mu(x_i, x_j) = 0$. Each pixel in the image is fully connected with others no matter how far from each other to build the pairwise term. Parameter ω_m is the weighted parameter, and k_m stands for the Gaussian kernel, which depends on the features (f_i, f_j) of pixel i and pixel j. Parameter K is the number of Gaussian kernels. Notably, the bilateral kernel is adopted, which is defined in terms of the spectral vectors I_i and I_j and positions p_i and p_j:

$$k(f_i, f_j) = -\omega_1 \exp\left(\frac{|p_i - p_j|^2}{2\sigma_\alpha^2} + \frac{|I_i - I_j|^2}{2\sigma_\beta^2}\right) - \omega_2 \exp\left(\frac{|p_i - p_j|^2}{2\sigma_\gamma^2}\right) \quad (3)$$

where the first kernel depends on both pixel positions and spectral vectors, and the second kernel only depends on the pixel positions; σ_β, σ_ω and σ_γ are hyper parameters and control the scale of the Gaussian kernels. In this study, the unary potentials are computed based on the probability maps of the segmentation model, while the original image pixels are used to infer pairwise potentials. The fully connected CRF model is amenable to efficient approximate probabilistic inference. The influence of fully connected CRF on the classification results will be analyzed in Section 5.2 to validate its importance in the proposed framework.

3.5. Seeded Region Growing (SRG)

In initial seed generation, only the pixels with high confidence are defined as the initial seed points, and they are relatively sparse. To have a denser supervision of segmentation model for better classification performance, the unlabeled pixels should be grown based on the seed points to generate more dense pixel-level annotations. A classical segmentation algorithm named Seeded Region Growing is adopted to formulate this problem after the process of fully connected CRF. The basis of seed points growing is the pixels in the small homogeneous regions should have the same class.

In SRG, the initial seed points are firstly selected based on some simple criteria such as color, texture and intensity. In this study, we used SVM to generate the initial seed points, which was described in Section 3.1. Once placed, the regions are grown from adjacent

unlabeled points of these seed points based on the similarity criterion. The following similarity criterion was used to determine whether the unlabeled point should be merged into the special region or not, which is based on the output probability maps of fully connected CRF.

$$P(p_{u,c}, \theta_c) = \begin{cases} TRUE & p_{u,c} > \theta_c \quad (c = \underset{\acute{c}}{\operatorname{argmax}}\, p_{u,\acute{c}}) \\ FALSE & otherwise \end{cases} \tag{4}$$

where the $P(p_{u,c}, \theta_c)$ is the similarity criterion; $p_{u,c}$ is the probability value of class c at position u of probability maps; θ_c is the probability threshold of class c. In practice, the same threshold was set for all classes. θ_c is set as 0.95 initially, then is added by 0.002 per iteration.

Once the similarity criterion is defined, the probability maps and seed points are fed into SRG for growing regions. SRG is an iterative algorithm for visiting each class. At the iteration of class c, we visit every pixel in the S_c and compute the $P(p_{u,c}, \theta_c)$ of its 8-connected neighbor pixels. Then, a new set of labeled pixels are generated and they are appended to S_c. After that, the new S_c is revisited, and S_c is updated again until the S_c is changeless. Once all classes are iterated, the SRG is stopped and new seed points are obtained, which will be used to train the segmentation model.

4. Results and Analysis

4.1. Experimental Setup

4.1.1. Implement Details

The Keras deep learning framework was used to implement all experiments. The ResNet50 [6] architecture was used as the backbone to build U-Net, which was initialized using the Gaussian distribution function in the initial training of WTS. U-Nets of other iterations in WTS were initialized by using the fully trained model's parameters of the last iteration. Adaptive moment estimation (Adam) algorithm was selected to optimize all models. The batch size was set as 10. All segmentation models were trained until the training loss converged. All implements were evaluated on the Windows 7 operating system with one 3.6 GHz 8-core i7-4790 CPU and 32GB memory. A NVIDIA GTX 1070 GPU was used to accelerate computing. In addition, SVM was selected as the compared method. The LIBSVM [54] was used to implement it. The radial basis function (RBF) was set as the kernel function, and the hyper-parameters of SVM were optimized by using cross-validation.

4.1.2. Evaluation Metrics

Overall accuracy (OA), kappa coefficient, precision, recall, F_1 score and intersection over union (IoU) were used to assess the quantitative classification performance. All of them can be computed by calculating the confusion matrix, which is an informative table that can allow a direct visualization of the performance on each class and can be used for analyzing the errors and confusions between different classes easily. OA is defined as the number of correctly classified pixels divided by total test pixels, which is the most intuitive measure to reveal the classification performance of all test pixels. Kappa coefficient is thought to be a more robust measure than a simple percent agreement calculation because it takes into account the possibility of the agreement occurring by chance. Precision is the ratio of correctly predicted pixels to the total predicted pixels, and recall is the ratio of correctly predicted pixels to all pixels in the actual label. The F_1 is the weighted average of precision and recall. IoU measure is the proportion of intersection among the predicted pixels and true pixels over their union. F_1 and IoU are all effective metrics for evaluating categorical accuracy. The formulas of them are as follows.

$$\text{Precision}_i = \frac{\mathrm{N}_{ii}}{\sum_{j=1}^{C} \mathrm{N}_{ji}} \tag{5}$$

$$\text{Recall}_i = \frac{N_{ii}}{\sum_{j=1}^{C} N_{ij}} \tag{6}$$

$$F_{1i} = \frac{2 \times \text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{7}$$

$$\text{IoU}_i = \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i - \text{Precision}_i \times \text{Recall}_i} \tag{8}$$

$$OA = \frac{\sum_{i=1}^{C} N_{ii}}{N} \tag{9}$$

$$\text{Kappa} = \frac{OA - p_e}{1 - p_e} \tag{10}$$

$$p_e = \frac{\sum_{i=1}^{C} (N_{ii} \times \sum_{j=1}^{C} N_{ij})}{N \times N} \tag{11}$$

where precision_i is the precision of class i, recall_i is the recall of class i, F_{1i} is the F_1 of class i, IoU_i is the IoU of class i, N_{ij} is the number of pixels that have class i but be classified into class j, C is the total number of classes and $C = 5$ in this study, N is the total number of test pixels. All these metrics are in the range of 0 to 1, except for Kappa with a range of -1 to 1. A higher value indicates a better classification performance.

4.1.3. Test Set

For accuracy evaluation, though it is better to use all pixels in the test area, the assignment of the true class label to each pixel is a complicated task. The grid point sampling is an alternative method since it can ensure the spatial distribution of testing points is uniform. However, it may lead to a serious class imbalance, and classes with small proportions may be not selected when some classes account for the most area (such as the cropland in the study area). Therefore, in this study, thousands of points with true class labels were selected manually from the test area to evaluate the classification results. For a fair evaluation, the following rules were followed when selecting testing points (taking the cropland as an example). First, croplands in many parts of the test area should be selected, not only focusing on a small part, to ensure a uniform spatial distribution. Second, all types of croplands, including not only cultivated farmland but fallow farmland, should be considered. Finally, more points should be selected at the border between cropland and other land covers than the inside homogeneous cropland region. The true class label of each point is defined by visual interpretation and viewing the high-resolution images from Google Earth. The number of each class in the test set is as follows: artificial surface, 4600; barren land, 2000; cropland, 5000; forest, 4000; water, 2000.

4.2. Experimental Results and Analysis

4.2.1. Results of WTS and Compared Methods

Classification results of WTS are obtained from the eight iterations in this section. The probability threshold to generate initial seed was set as 0.7. SVM was selected to be compared with WTS due to its popularity and efficient performance in remote sensing classification applications. The training set of SVM was the same as WTS. Moreover, global land cover map FROM-GLC10 [55] was used as reference data to train U-Net. The corresponding classification results were also compared in this section. FROM-GLC10 is acquired based on 10 m resolution Sentinel-2 data and achieved an overall accuracy of 72.76% at global scale. It was down-sampled to 20 m resolution to be consistent with images used in this study. In FROM-GLC10, cropland, forest, grassland, shrubland, wetland, water, tundra, impervious surface, barren land and snow/ice were used as the classification system. Tundra and snow/ice were not included in the study area. In order to keep

consistent with the classification system in this study as close as possible, and by analyzing the class definition of two classification systems, the following merging rules of classes in FROM-GLC10 were followed to get the final reference data: cropland and grassland were merged as cropland; forest and shrubland were merged as forest; wetland and water were merged as water; impervious surface was treated as artificial surface. OA, kappa coefficient, F_1 and IoU of all classes of all methods are gathered in Table 3. Classification results of two representative areas are shown in Figure 6 for a better visual interpretation and analysis.

Table 3. OA, kappa coefficient, F_1 and IoU of all classes of WTS and compared methods. (The optimal results are marked in bold. F_1 and IoU are separated with symbol "/", and the former stands for F_1 while the latter stands for IoU).

	Artificial Surface	Barren Land	Cropland	Forest	Water	OA	Kappa
SVM	0.7879 /0.6501	0.6492 /0.4806	0.7391 /0.5861	0.8689 /0.7681	0.9480 /0.9012	0.7965	0.7377
U-Net(FROM-GLC10)	0.5513 /0.3801	0.0070 /0.0035	0.7365 /0.5829	0.8602 /0.7547	0.8928 /0.8064	0.6943	0.5919
WTS	**0.8103 /0.6810**	**0.7100 /0.5503**	**0.7793 /0.6384**	**0.8847 /0.7932**	**0.9623 /0.9273**	**0.8252**	**0.7738**

From Table 3, it can be observed that WTS obtained the best results on all metrics. WTS achieved OA of 82.52% and outperformed SVM by approximately 3%, which is a considerable accuracy improvement on the land cover classification of remote sensing. The U-Net that uses FROM-GLC10 as reference data obtained the worst result and achieved OA of merely 69.43%, which is almost 10% lower than SVM. This is due to a number of factors such as imaging time inconsistency between Sentinel-2B images and FROM-GLC10(2017), classification system inconsistency and incorrectly labeled information in FROM-GLC10. Thus, using the current land cover map can solve the problem of insufficiency of reference data when using segmentation models, but it has limitations when meeting practical applications. As for the categorical accuracy analysis, U-Net also obtained the worst results on all classes except for cropland, which is a little higher than SVM. WTS increased the F_1 by more than 6% than SVM on barren land. Barren land was the hardest class to identify among all classes in our study, which is always confused with artificial surface and cropland. This is because the existence of buildings with high brightness among the artificial surface and fallow farmlands among the cropland, which all have similar spectral values with barren land. Due to different definitions of barren land in our study and FROM-GLC10 and a small percent of barren land on the training set, U-Net only obtained 0.070 F_1 on the barren land. Moreover, 2.24%, 4.02%, 1.58% and 1.43% F_1 improvements were achieved by WTS than SVM for artificial surface, cropland, forest and water, separately. All these demonstrate the effectiveness of the proposed WTS framework on the land cover classification. This good performance benefits not only from the ability of learning multi-scale features of segmentation models, but also the constant seed updates based on iterative process by SRG and fully connected CRF that can progressively optimize the segmentation model.

As for qualitative comparison, the classification results in Figure 6 show that there was more salt and pepper noise in the classification result of SVM, while the results of U-Net and WTS looked more compact and continuous. This is because the segmentation model had a large receptive field and could not only use the spectral information but also multi-scale features of the neighborhood field. U-Net achieved bad results on artificial surfaces. This is mainly because of many incorrectly labeled points in FROM-GLC10, which had a significant negative impact on the results. In addition, as shown in purple circles marked in Figure 6 for SVM, many fallow farmlands among the cropland were misclassified as barren land. This misclassification existed in the results of WTS, but has been greatly reduced, while U-Net could avoid this misclassification well. At the same time, some croplands were also confused with artificial surfaces for SVM (shown in the

yellow circles). All these confusions were due to the limitation of expression of the spectral value. For the barren land shown in the white circle, SVM and WTS could extract well, while U-Net misclassified them as artificial surfaces, which is caused by the class definition difference. In our study, bare mines were treated as barren land, whereas they belong to the impervious surface in FROM-GLC10. As for water and forest, all methods had great performances via visualization interpretation.

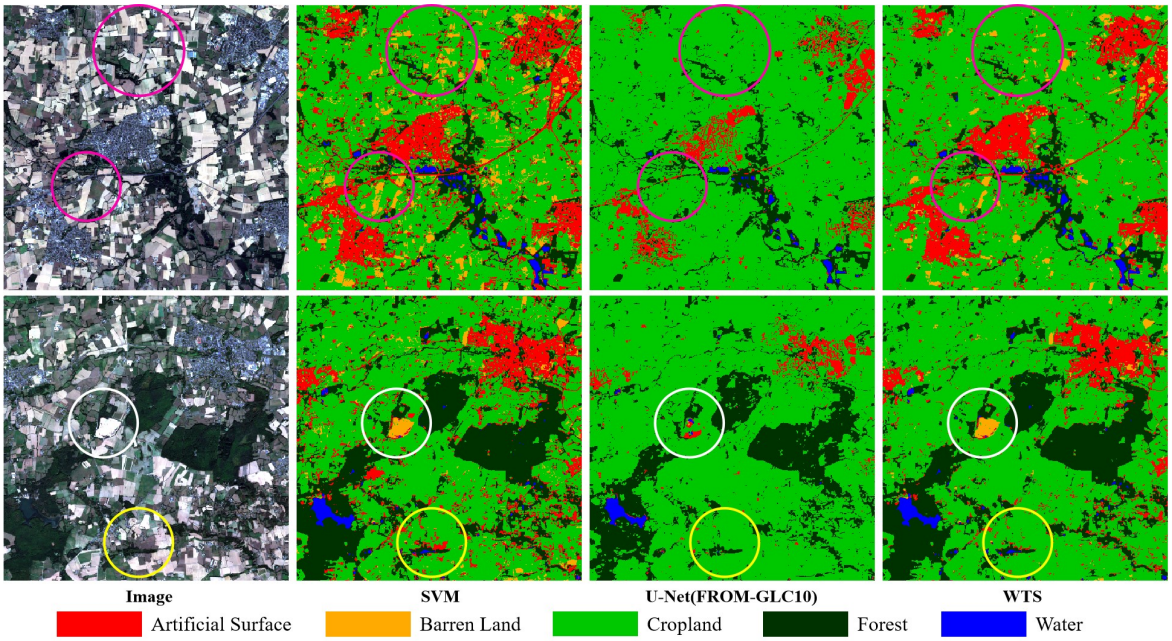
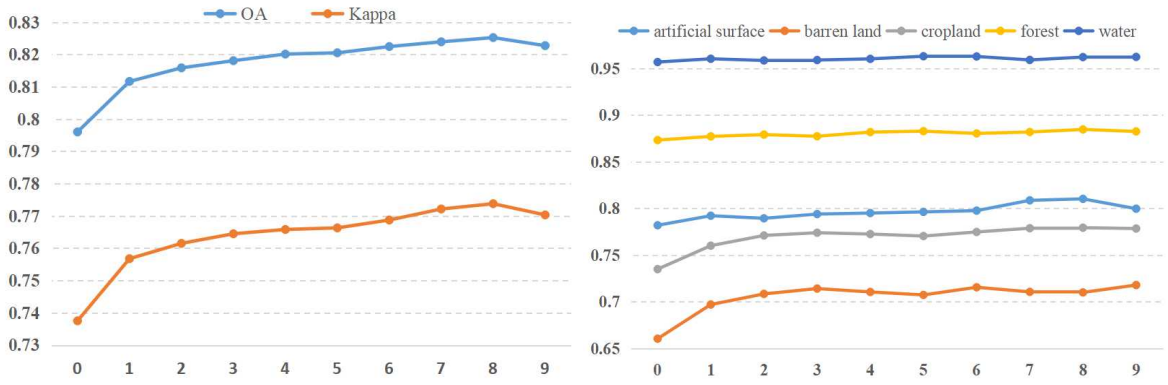


Figure 6. Classification results visual comparison of SVM and compared methods.

4.2.2. Results in Different Iterations of WTS

As illustrated in the methodology section, WTS is an iterative process to progressively update the training set and optimize the segmentation model. In order to demonstrate its progressive optimization on classification performance, classification results in different iterations of WTS are compared in this section; 0.7 was set for the probability threshold to generate initial seed in WTS. Figure 7 shows the OA, kappa coefficient and F_1 of all classes at different iterations of WTS. Classification results of one selected area in different iterations are shown in Figure 8.



(**a**) OA and Kappa. (**b**) F_1 of all classes.

Figure 7. OA, Kappa coefficient and F_1 of all classes at different iterations of WTS.

For the overall classification performance, OA and kappa coefficient constantly increased with the advance of the optimization process of WTS and gradually tended to be stable in the later stage. The accuracy improvement was obvious in the front stage, and gradually decreased. This demonstrates that WTS can continuously optimize the seeded training set in the iterative process, which can be observed in Figure 1. As a result, the classification performance can keep getting better. For the category accuracy, the F_1 of water and forest were basically not affected and were relatively stable. This is because these two land covers are more homogeneous than others, and they are easy to be distinguished even using only the spectral information. So in the initial seed generation phase using SVM, most pixels of water and forest have been assigned as initial seed points given the 0.7 probability threshold. Thus, the seed update of the these two land covers can not change too much. The hardest distinguished land cover, barren land, achieved the biggest accuracy improvement of almost 0.05 on F_1, further illustrating the effectiveness of WTS on updating the training set. The accuracy improvements of artificial surface and cropland were placed in the middle. Via visual interpretation in Figure 8, some fallow farmlands were misclassified as barren land (shown in the purple circle marked areas) in the initial iteration. With the iterative process, these misclassifications were gradually reduced, which is consistent with the quantitative evaluation results. The classification performances of water and forest were stable.

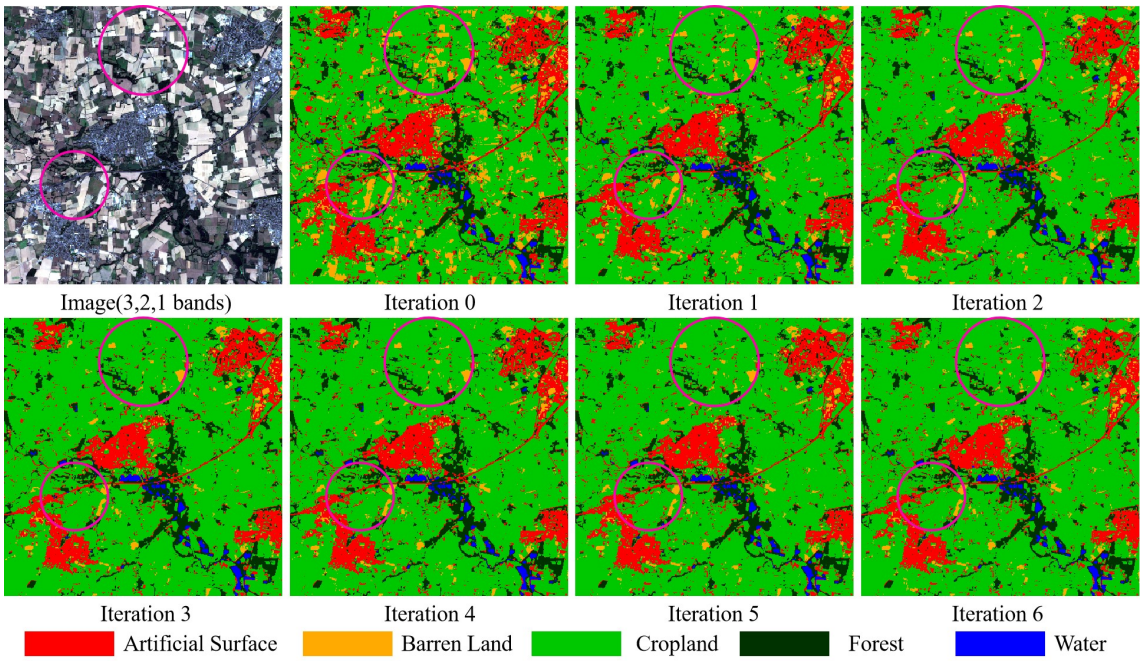


Figure 8. Classification results evolution in different iterations of WTS. (Iteration0 stands for the result that using the initial seed as training set).

5. Discussion

In this section, the influences of probability of threshold to generate initial seed and using the fully connected CRF on classification results are studied and discussed.

5.1. Influence of Probability Threshold to Generate Initial Seed on Classification Result

The probability threshold to generate the initial seed is a vital hyper-parameter in the proposed WTS framework. It controls the sparsity and quality of the initial seed. Thus, the probability threshold was set as [0.5, 0.6, 0.7, 0.8, 0.9] to evaluate its influence on classification results. Percents of all land covers and unlabeled points in the initial training set based on different probability thresholds are listed in Table 4. When the probability threshold increases, fewer pixels will be assigned as seed points but with higher quality. Otherwise, the number of initial seed points is increasing, but the quality is going to get

worse. This is because many misclassified pixels by SVM are also treated as seed points, which will result in a negative effect on the classification performance.

Table 4. Percents (%) of all land covers and unlabeled points in the initial training set based on different probability thresholds.

Probability Threshold	Unlabeled	Artificial Surface	Barren Land	Cropland	Forest	Water
0.5	6.14	8.63	3.62	61.55	19.06	1.01
0.6	18.70	6.20	2.22	55.67	16.26	0.95
0.7	31.32	4.10	0.95	49.39	13.35	0.90
0.8	45.20	2.06	0.39	41.37	10.14	0.84
0.9	63.32	0.14	0.17	28.99	6.64	0.74

Overall accuracies of classification results based on different probability thresholds for generating initial seed are illustrated in Figure 9. It can be observed that the probability threshold of 0.7 achieved the best classification performance. The worst performance belonged to 0.9. The seed points had high confidence when given probability threshold of 0.9. However, as illustrated in Table 4, more than 63% pixels were unlabeled, and artificial surface and barren land all only accounted for less than 0.2%. These limited training datasets may not guarantee adequate training of the deep classification model. At the same time, the extreme imbalance class distribution will further make a negative effect on the classification result. The accuracy increased when going to the next iteration. This may due to the SRG algorithm that updates the initial seed and leads to more points that can be trained in the deep classification model. However, accuracy started to decrease when the iteration increased and was even lower than the initial iteration. When the probability threshold was set as 0.5, only 6.14% pixels were unlabeled, and the classification accuracy was close to SVM. The accuracy improved in the later iteration, but it soon was stable. The 0.6 and 0.8 indicate similar good classification performances, but still worse than using probability threshold of 0.7. Comprehensively, 0.7 was the optimal probability threshold that could balance the sparsity and quality of training set well.

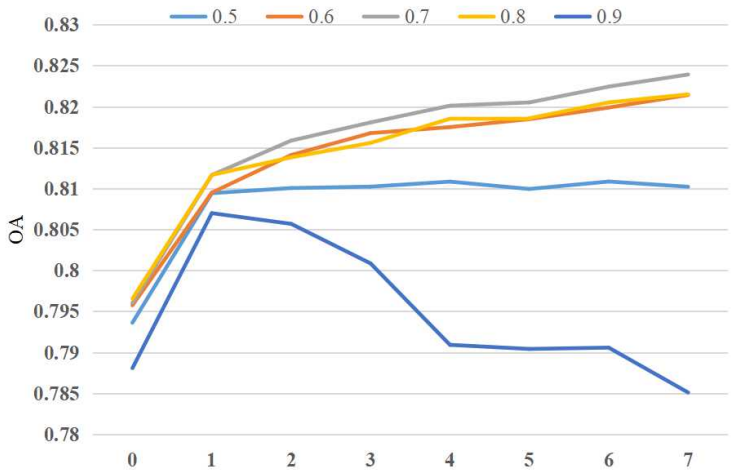


Figure 9. Overall accuracies of classification results based on different probability thresholds for generating the initial seed.

5.2. Influence of Fully Connected CRF on Classification Results

In the procedure of updating seed in the proposed WTS, seeded region growing and fully connected CRF were used. This is no doubt because the seeded region growing algorithm plays the most important role on improving the training set. However, fully connected CRF is also an indispensable module in WTS. In order to verify its validity on the

classification result, a comparative experiment of WTS with and without fully connected CRF was conducted in this section. The probability threshold to generate initial seed was set as 0.7. Figure 10 shows the accuracy comparison of classification results.

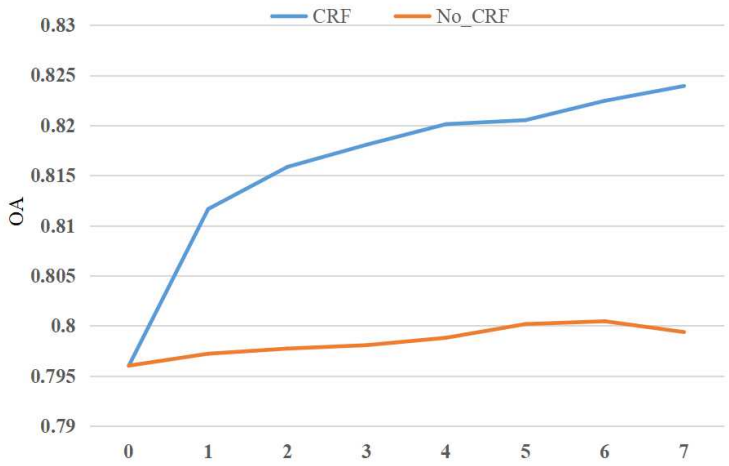


Figure 10. Overall accuracies of classification results of WTS with and without fully connected CRF.

From Figure 10, it can be found that the accuracy increased very slowly when not using fully connected CRF. This is because the segmentation model falls into a state of "self-deception". It is a state that can be understood as follows: when not using fully connected CRF, the output probability map learned by the segmentation model is directly fed into the SRG module to update training set. Then, the updated training set will be further used to train the segmentation model. This means the segmentation model is always trained by its self-learned knowledge. This will make it difficult to update the parameters of the model. Therefore, the accuracy is almost unchanged. However, the fully connected CRF will help the segmentation model to escape such a "self-deception" state as it can optimize the output probability map based on the corresponding image. Therefore, the fully connected CRF is a vital component in the proposed WTS framework. Figure 11 shows one training sample evolution of WTS with and without fully connected CRF, which can also verify the effectiveness of fully connected CRF in visual interpretation.

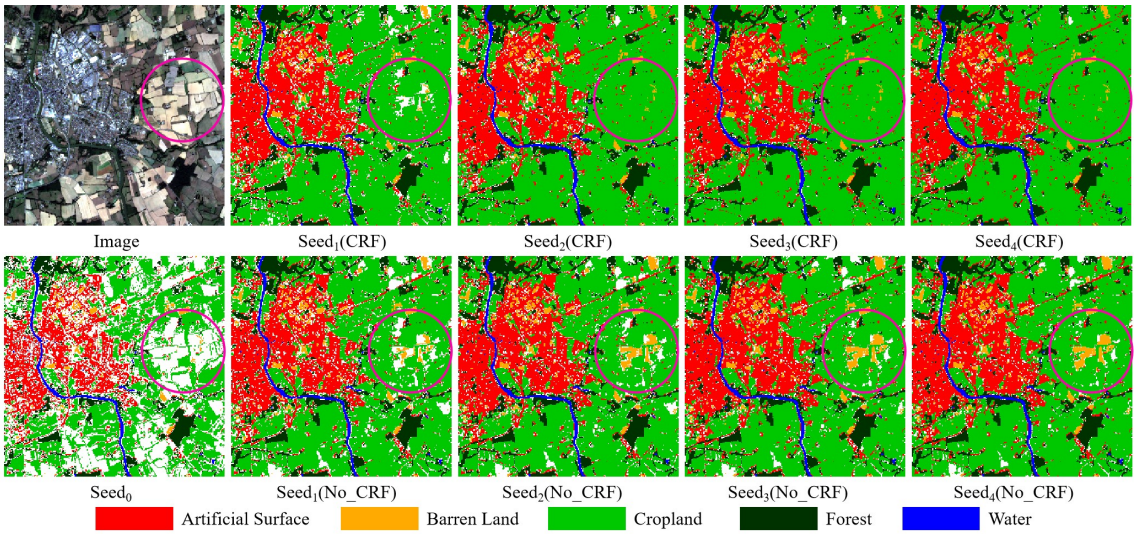


Figure 11. One training sample evolution comparison of WTS with and without fully connected CRF. ("CRF" means using fully connected CRF in WTS, and "No_CRF" means not using using fully connected CRF in WTS. The white areas represent the unlabeled points).

6. Conclusions and Future Work

In order to deal with the insufficiency of pixel-level annotations for training semantic segmentation models, a weakly towards strongly (WTS) supervised learning framework is proposed in this study for remote sensing land cover classification, which is inspired by the weakly supervised learning method and seeded region growing traditional segmentation algorithm. In the proposed framework, only several "point-point" style training samples are required to generate the initial "patch-patch" seeded training set using SVM for training segmentation models. Compared with pixel-level annotations, they are much less expensive to be acquired. Then, the fully connected CRF and SRG modules are used to gradually update the training set, which can progressively improve the pixel-level supervision of segmentation models. The superiority of the proposed WTS framework has been verified on Sentinel-2 remote sensing images. Experimental results show that the proposed WTS framework is superior to SVM and the method of U-Netthat uses global land cover map FROM-GLC10 as training reference data. WTS is a reliable and effective method for land cover classification using segmentation models when the pixel-level labeled datasets are insufficient. SVM is not a unique way to generate the initial seed; other classifiers such as neural network and random forest can also be used. Analyzing current land cover classification products and treating the class consistent points as the initial seed points is also an advisable way. In future work, these works will be studied and compared to further improve the quality of the initial seed. The U-Net segmentation model used in this paper is also not unalterable, which just provides a benchmark and can be improved and replaced by other segmentation models. In addition, the effectiveness of different bands of remote sensing images on the classification results will be analyzed in future work for providing more valuable information on land cover classification.

Author Contributions: Funding acquisition, L.Z. and P.T.; Investigation, W.Z.; Methodology, W.Z. and L.Z.; Project administration, P.T.; Supervision, P.T. and T.C.; Validation, W.Z.; Visualization, W.Z.; Writing—original draft, W.Z.; Writing—review & editing, L.Z., P.T. and T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under grant 41701397 and grant 41971396.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors are grateful to the anonymous reviewers for their careful assessment, valuable comments and suggestions that improved the quality of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shi, H.; Chen, L.; Bi, F.; Chen, H.; Yu, Y. Accurate Urban Area Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1948–1952. [CrossRef]
2. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [CrossRef]
3. Mountrakis, G.; Im, G.; Ogole, C. Support Vector Machines in Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]
4. Phan, T.N.; Kuch, V.; Lehnert, L.W. Land Cover Classification Using Google Earth Engine and Random Forest Classifier—The Role of Image Composition. *Remote Sens.* **2020**, *12*, 2411. [CrossRef]
5. Zhou, L.; Yang, X. Training Algorithm Performance for Image Classification by Neural Networks. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 945–951. [CrossRef]
6. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
7. Girshick, R.B.; 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, CVPR 2014, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]

8. Ren, S.; He, K.; Girshick, R.B.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
9. Zhao, L.; Zhang, W.; Tang, P. Analysis of the Inter-Dataset Representation Ability of Deep Features for High Spatial Resolution Remote Sensing Image Scene Classification. *Multimed. Tools Appl.* **2019**, *78*, 9667–9689. [CrossRef]
10. Zhang, W.; Tang, P.; Zhao, L. Remote Sensing Image Scene Classification Using CNN-CapsNet. *Remote Sens.* **2019**, *11*, 494. [CrossRef]
11. Ma, D.; Tang, P.; Zhao, L. SiftingGAN: Generating and Sifting Labeled Samples to Improve the Remote Sensing Image Scene Classification Baseline In Vitro. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1046–1050. [CrossRef]
12. Wang, C.; Bai, X.; Wang, S.; Zhou, J.; Ren, P. Multiscale Visual Attention Networks for Object Detection in VHR Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 310–314. [CrossRef]
13. Chen, C.; Gong, W.; Chen, Y.; Li, W. Object Detection in Remote Sensing Images Based on A Scene-Contextual Feature Pyramid Network. *Remote Sens.* **2019**, *11*, 339. [CrossRef]
14. Wang, Q.; Zhang, X.; Chen, G.; Dai, F.; Gong, Y.; Zhu, K. Change Detection Based on Faster R-CNN for High-Resolution Remote Sensing Images. *Remote Sens. Lett.* **2018**, *9*, 923–932. [CrossRef]
15. Ji, S.; Shen, Y.; Lu, M.; Zhang, Y. Building Instance Change Detection from Large-Scale Aerial Images using Convolutional Neural Networks and Simulated Samples. *Remote Sens.* **2019**, *11*, 1343. [CrossRef]
16. Mnih, V. *Machine Learning for Aerial Image Labeling*; University of Toronto: Toronto, ON, Canada, 2013.
17. MahdianPari, M.; Salehi, B.; Rezaee, M.; Mohammadimanesh, F.; Zhang, Y. Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. *Remote Sens.* **2018**, *10*, 1119. [CrossRef]
18. Liu, T.; Abd-Elrahman, A.; Morton, J.; Wilhelm, V.L. Comparing Fully Convolutional Networks, Random forest, Support Vector Machine, and Patch-Based Deep Convolutional Neural Networks for Object-Based Wetland Mapping Using Images From Small Unmanned Aircraft System. *GISci. Remote Sens.* **2018**, *55*, 243–264. doi:10.1080/15481603.2018.1426091. [CrossRef]
19. Kwan, C.; Ayhan, B.; Budavari, B.; Lu, Y.; Perez, D.; Li, J.; Bernabe, S.; Plaza, A. Deep Learning for Land Cover Classification Using Only a Few Bands. *Remote Sens.* **2020**, *12*, 2000.10.3390/rs12122000. [CrossRef]
20. Pan, X.; Zhao, J. A Central-Point-Enhanced Convolutional Neural Network for High-Resolution Remote-Sensing Image Classification. *Int. J. Remote Sens.* **2017**, *38*, 6554–6581. [CrossRef]
21. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 645–657. [CrossRef]
22. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [CrossRef]
23. Persello, C.; Stein, A. Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2325–2329. [CrossRef]
24. Wang, H.; Wang, Y.; Zhang, Q.; Xiang, S.; Pan, C. Gated Convolutional Neural Network for Semantic Segmentation in High-Resolution Images. *Remote Sens.* **2017**, *9*, 446. [CrossRef]
25. Sun, W.; Wang, R. Fully Convolutional Networks for Semantic Segmentation of Very High Resolution Remotely Sensed Images Combined with DSM. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 474–478. [CrossRef]
26. Fu, G.; Liu, C.; Zhou, R.; Sun, T.; Zhang, Q. Classification for High Resolution Remote Sensing Imagery Using A Fully Convolutional Network. *Remote Sens.* **2017**, *9*, 498. [CrossRef]
27. Gerke, M. *Use of The Stair Vision Library Within The ISPRS 2D Semantic Labeling Benchmark (Vaihingen)*; ResearcheGate: Berlin, Germany, 2014. [CrossRef]
28. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raska, R. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 172–17209.
29. Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. SEN12MS–A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *IV-2/W7*, 153–160. [CrossRef]
30. Xin-Yi, T.; Gui-Song, X.; Qikai, L.; Huanfeng, S.; Shengyang, L.; Shucheng, Y.; Liangpei, Z. Learning Transferable Deep Models for Land-Use Classification with High-Resolution Remote Sensing Images. *arXiv* **2018**, arXiv:1807.05713.
31. Isikdogan, F.; Bovik, A.C.; Passalacqua, P. Surface Water Mapping by Deep Learning. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2017**, *10*, 4909–4918. [CrossRef]
32. Feng, M.; Sexton, J.O.; Channan, S.; Townshend, J.R. A Global, High-Resolution (30-m) Inland Water Body Dataset for 2000: First Results of A Topographic-Spectral Classification Algorithm. *Int. J. Digit. Earth* **2016**, *9*, 113–133. [CrossRef]
33. Scepanovic, S.; Antropov, O.; Laurila, P.; Ignatenko, V.; Praks, J. Wide-Area Land Cover Mapping with Sentinel-1 Imagery using Deep Learning Semantic Segmentation Models. *arXiv* **2019**, arXiv:1912.05067.
34. Chantharaj, S.; Pornratthanapong, K.; Chitsinpchayakun, P.; Panboonyuen, T.; Vateekul, P.; Lawavirojwong, S.; Srestasathiern, P.; Jitkajornwanich, K. Semantic Segmentation on Medium-Resolution Satellite Images Using Deep Convolutional Networks with Remote Sensing Derived Indices. In Proceedings of the 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhonpathom, Thailand, 11–13 July 2018; pp. 1–6.

35. Grekousis, G.; Mountrakis, G.; Kavouras, M. An Overview of 21 Global and 43 Regional Land-Cover Mapping Products. *Int. J. Remote Sens.* **2015**, *36*, 5309–5335. [CrossRef]
36. Ghosh, A.; Kumar, H.; Sastry, P. Robust Loss Functions under Label Noise for Deep Neural Networks. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
37. Sukhbaatar, S.; Bruna, J.; Paluri, M.; Bourdev, L.; Fergus, R. Training Convolutional Networks with Noisy Labels. *arXiv* **2014**, arXiv:1406.2080.
38. Han, B.; Yao, Q.; Yu, X.; Niu, G.; Xu, M.; Hu, W.; Tsang, I.; Sugiyama, M. Co-teaching: Robust Training of Deep Neural Networks with Extremely Noisy Labels. In Proceedings of the 2018 Neural Information Processing Systems, Montreal, QC, Canada, 2–8 December 2018; pp. 8535–8545.
39. Mnih, V.; Hinton, G.E. Learning to Label Aerial Images from Noisy Data. In Proceedings of the 29th International conference on machine learning (ICML-12), Edinburgh, Scotland, 26 June–1 July 2012; pp. 567–574.
40. Papandreou, G.; Chen, L.; Murphy, K.P.; Yuille, A.L. Weakly-and Semi-Supervised Learning of A Deep Convolutional Network for Semantic Image Segmentation. In Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1742–1750.
41. Bearman, A.; Russakovsky, O.; Ferrari, V.; Fei-Fei, L. What's the Point: Semantic Segmentation with Point Supervision. In Proceedings of the 2016 European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 549–565.
42. Dai, J.; He, K.; Sun, J. Boxsup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation. In Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1635–1643.
43. Lin, D.; Dai, J.; Jia, J.; He, K.; Sun, J. Scribblesup: Scribble-Supervised Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 3159–3167.
44. Chan, L.; Hosseini, M.S.; Plataniotis, K.N. A Comprehensive Analysis of Weakly-Supervised Semantic Segmentation in Different Image Domains. *arXiv* **2019**, arXiv:1912.11186.
45. Fu, K.; Lu, W.; Diao, W.; Yan, M.; Sun, H.; Zhang, Y.; Sun, X. WSF-NET: Weakly Supervised Feature-Fusion Network for Binary Segmentation in Remote Sensing Image. *Remote Sens.* **2018**, *10*, 1970. [CrossRef]
46. Zhang, L.; Ma, J.; Lv, X.; Chen, D. Hierarchical Weakly Supervised Learning for Residential Area Semantic Segmentation in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 117–121. [CrossRef]
47. Adams, R.; Bischof, L. Seeded Region Growing. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 641–647. [CrossRef]
48. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
49. 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. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef] [PubMed]
50. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015—18th International Conference Munich, Germany, 5–9 October 2015; Proceedings, Part III; pp. 234–241. [CrossRef]
51. Simon, J.; Michal, D.; David, V.; Adriana, R.; Yoshua, B. The One Hundred Layers Tiramisu: Fully Convolutional Densenets for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.
52. Kolesnikov, A.; Lampert, C.H. Seed, Expand and Constrain: Three Principles for Weakly-Supervised Image Segmentation. In Proceedings of the 2016 European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 695–711.
53. Krähenbühl, P.; Koltun, V. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In Proceedings of the 2011 Neural Information Processing Systems, Granada, Spain, 12–17 December 2011; pp. 109–117.
54. Chih-Wei, H.; Chih-Chung, C.; Chih-Jen, L. A Practical Guide to Support Vector Classification. *BJU Int.* **2008**, *101*, 1396–1400.
55. Gong, P.; Liu, H.; Zhang, M.; Li, C.; Wang, J.; Huang, H.; Clinton, N.; Ji, L.; Li, W.; Bai, Y.; et al. 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.* **2019**, *64*, 370–373. [CrossRef]