

## Article

# Convolutional Neural Network-Based Remote Sensing Images Segmentation Method for Extracting Winter Wheat Spatial Distribution

Chengming Zhang <sup>1,2,\*</sup>, Shuai Gao <sup>3,\*</sup>, Xiaoxia Yang <sup>1,2</sup>, Feng Li <sup>4</sup>, Maorui Yue <sup>5</sup>, Yingjuan Han <sup>6</sup>, Hui Zhao <sup>6</sup>, Ya'nan Zhang <sup>1</sup> and Keqi Fan <sup>1</sup>

<sup>1</sup> College of Information Science and Engineering, Shandong Agricultural University, 61 Daizong Road, Taian 271000, China; yangxx@sda.u.edu.cn (X.Y.); zyn980113@hotmail.com (Y.Z.); fkg980810@hotmail.com (K.F.)

<sup>2</sup> Shandong Technology and Engineering Center for Digital Agriculture, 61 Daizong Road, Taian 271000, China

<sup>3</sup> Chinese Academy of Sciences, Institute of Remote Sensing and Digital Earth, 9 Dengzhuangnan Road, Beijing 100094, China

<sup>4</sup> Shandong Climate Center, Mountain Road, Jinan 250001, China; lifengsd@outlook.com

<sup>5</sup> Taian Agriculture Bureau, Naihe Road, Taian 271000, China; yuemaoruita@outlook.com

<sup>6</sup> Key Laboratory for Meteorological Disaster Monitoring and Early Warning and Risk Management of Characteristic Agriculture in Arid Regions, CMA, 71 Xinchangxi Road, Yinchuan 750002, China; yjhan\_nx@outlook.com (Y.H.); zhaohui\_cau@outlook.com (H.Z.)

\* Correspondence: chming@sda.u.edu.cn (C.Z.); gaoshuai@radi.ac.cn (S.G.); Tel.: +86-139-5382-3659 (C.Z.); +86-010-64806258 (S.G.)

Received: 26 September 2018; Accepted: 16 October 2018; Published: 19 October 2018



**Featured Application:** In Gaofen-2 images, it is difficult to accurately extract winter wheat spatial distribution using traditional methods. Because our approach can better solve this problem, it has played an important role in agricultural surveys and improved the efficiency of agricultural surveys. Our approach has been utilized by the Department of Agriculture and the Meteorological Bureau of Shandong Province, China.

**Abstract:** When extracting winter wheat spatial distribution by using convolutional neural network (CNN) from Gaofen-2 (GF-2) remote sensing images, accurate identification of edge pixel is the key to improving the result accuracy. In this paper, an approach for extracting accurate winter wheat spatial distribution based on CNN is proposed. A hybrid structure convolutional neural network (HSCNN) was first constructed, which consists of two independent sub-networks of different depths. The deeper sub-network was used to extract the pixels present in the interior of the winter wheat field, whereas the shallower sub-network extracts the pixels at the edge of the field. The model was trained by classification-based learning and used in image segmentation for obtaining the distribution of winter wheat. Experiments were performed on 39 GF-2 images of Shandong province captured during 2017–2018, with SegNet and DeepLab as comparison models. As shown by the results, the average accuracy of SegNet, DeepLab, and HSCNN was 0.765, 0.853, and 0.912, respectively. HSCNN was equally as accurate as DeepLab and superior to SegNet for identifying interior pixels, and its identification of the edge pixels was significantly better than the two comparison models, which showed the superiority of HSCNN in the identification of winter wheat spatial distribution.

**Keywords:** remote sensing image segmentation; convolutional neural networks; Gaofen-2; hybrid structure convolutional neural networks; winter wheat spatial distribution; classification-based learning

## 1. Introduction

Winter wheat is the most important food crop in China, comprising 21.38% of the gross cropped area of the domestic food crops in 2017 according to the data released by the National Bureau of Statistics, with its output accounting for 21.00% of the total food crop production [1]. For national food security, the Chinese government has assigned a minimum area of arable land in each region that needs to be safeguarded (the “red line”) [2]. Timely and accurate acquisition of the size and spatial distribution of winter wheat fields assists the relevant government departments in guiding the farming activities, estimating the yield, and adjusting the agricultural structure for ensuring food security [3].

Remote sensing is capable of imaging and large-area monitoring, making it a good data source for rapid and accurate extraction of winter wheat planting information. Researchers have successfully extracted winter wheat spatial distribution information from MODIS (moderate-resolution imaging spectroradiometer) and ETM/TM (enhanced thematic mapper plus/thematic mapper), achieving accuracies of 85.5% and 89.1%, respectively [4,5]. This exhibits the advantage of remote sensing in this application. However, owing to limitations in the spatial resolution of the data source, the spatial resolution of the extraction results is also rather coarse and unable to satisfy the requirement of the application [6–10]. With the development of high-resolution remote sensing satellites, a crop planting area can be monitored more accurately using the corresponding images as the data source [11,12]. The winter wheat cultivation information is extracted from the remote-sensing images captured by Gaofen-1 of the Gaofen series of Chinese satellites, yielding satisfactory results, with maximum accuracy reaching about 89% [13–18]. Most researchers still use traditional methods, such as decision trees and textures features. These methods can only take advantage of low-level features, which make it easy to make mistakes in identifying pixels at the edge of winter wheat planting area.

Image segmentation has been successfully used in the processing of camera images and applied by researchers to high-resolution remote sensing images, achieving significantly more accurate classification by a pixel-by-pixel segmentation [19–21]. Feature extraction is the key step in remote sensing image segmentation. In high-resolution remote sensing images, as the spectral difference between the same type of objects is increased, and between different types of objects is diminished, the former has more probability of exhibiting different spectral properties, whereas the latter tends to be spectrally similar, which makes feature extraction increasingly difficult [22,23]. Traditional methods including k-nearest neighbors and maximum entropy can only identify low-level image features such as color, shape, and texture. They are not capable of visually providing a semantic description. This hinders the extraction of higher-level features and limits the use of these methods in the segmentation of high-resolution remote sensing images [24,25].

With the development of machine learning, algorithms such as neural networks (NNs) [26] and support vector machine (SVM) [27,28] are being used in the segmentation of high-resolution images [29–31]. In some studies, when compared with traditional statistical methods and object-oriented methods, machine learning algorithms yielded better image segmentation results [32, 33]. Both SVM and NNs are shallow-learning algorithms [34–36], which do not express complex functions well owing to the limitations in their network structure. Therefore, these models cannot adapt to the continuously increasing complexity caused by the increasing sample size and diversity [37,38].

Progress in deep learning has facilitated solving these problems by using deep neural networks (DNNs) [39–42]. As an important branch of deep learning, a convolutional NN (CNN) is widely used with visual data because of its excellent feature learning ability [43–45]. A CNN is a deep learning network, composed of several layers, capable of nonlinear mapping. Its strength in learning is exemplified by the good image segmentation results achieved [46–52]. Further, the capacity of many large CNNs can be scaled according to the size of the training data and complexity and processing ability of the model, and their performance in image segmentation has improved significantly [53–60].

A fully convolutional network (FCN) is a deep learning network for image segmentation, which was proposed in 2015. Taking advantage of convolution computation in its feature organization and extraction abilities, an FCN realizes pixel-by-pixel segmentation of camera images

by constructing a multi-layer convolutional structure and setting appropriate deconvolutional layers [61–63]. Accordingly, a series of convolution-based segmentation models has been developed including SegNet [64], UNet [65], DeepLab [66], multi-scale FCN [67], and ReSeg [68]. Of these models, SegNet and UNet are clearly structured, and it is easy to understand the convolution structure of the model. The processing speed is fast. DeepLab uses a method called “Atrous Convolution”, which has a strong advantage in processing detailed images. multi-scale FCN is designed to address the huge scale gap between different classes of targets, i.e., sea/land and ships. ReSeg exploits the local generic features extracted by Convolutional Neural Networks and the capacity of Recurrent Neural Networks (RNN) to retrieve distant dependencies. Each model has its own strengths and is adept at dealing with certain image types.

In the work of extracting the spatial distribution of crops with high GF-1 as the data source, in addition to methods such as decision trees, textures features, and maximum entropy, research has also been carried out using deep learning. However, most of these studies directly use the existing deep learning model as a tool, and seldom consider the influence of characteristics difference of edge pixels and inner pixels in the crop planting area are large.

On board the Gaofen-2 satellite is a panchromatic camera with a spatial resolution of 1 m, and a multi-spectral camera with a spatial resolution of 4 m, which provides ideal data for extracting winter wheat plantation information. Before the application of a CNN to GF-2 remote-sensing images for this purpose, trial extraction is performed with classical network architectures (such as SegNet) where misidentified pixels are categorized, of which approximately 90% are found at the edge of the crop field. Further analysis indicates the structure of the convolutional layer as the source of this problem. The outcome produced by operating the convolution kernel in the pixel block is treated as the eigenvalue of the central pixel of the pixel block. As such, for the pixels at the edge, 50% of the pixels involved in each convolution are from negative samples, whereas, for the pixels at the corner, this number is 75% or higher. This results in a significant difference between the eigenvalues of the pixels at these locations and those at the center of the image, and an increase in the probability of the recognition results being placed in a wrong category. To avoid these problems, a new method is herein proposed for the extraction of the winter wheat field information from the GF-2 remote sensing images. The main procedures are as follows.

1. First, a CNN consisting of two independent sub-networks of different depths is established. The deep and shallow sub-networks are trained to be sensitive only to the pixels at the interior and edge of a winter wheat planting field, respectively, and only these pixels are extracted. This model is named as a Hybrid Structure Convolutional Neural Network (HSCNN).
2. A classification algorithm is adopted in the model training. For initial training of the sub-network used for the edge pixel extraction, edge pixels are considered as positive samples, with the pixels at other locations being treated as negative samples. The inner pixels are then designated as positive samples, with the pixels at other locations as negative samples, for training the sub-network used for the inner pixel extraction. After the successful completion of the training, the neural network is able to extract the winter wheat field from the GF-2 images accurately.
3. Finally, a GF-2 image is segmented by the trained model. Because SegNet and DeepLab are classic semantic segmentation models of images, and, the working principles of these two models are very similar to our work, we choose these two models as the comparison model, and a comparison is performed with them to evaluate the accuracy of the segmentation results.

## 2. Data Sources and Methods

### 2.1. Data Sources

#### 2.1.1. Study Region

The whole study region is Shandong province, China. Shandong is located along the eastern coast of China (in the lower stream of the Yellow river), within  $34^{\circ}22' \text{ N}$ – $38^{\circ}24' \text{ N}$  and  $114^{\circ}47.5' \text{ E}$ – $122^{\circ}42' \text{ E}$ . It measures 721.03 km from east to west, and 437.28 km from north to south. The land area of the province is 155,800 km<sup>2</sup>, of which 14.59% is mountainous, 5.56% is water (such as lakes), 15.98% is forest, and 53.82% is cultivated land. The annual total planting area of crops in the province is approximately 162 million mu. The main food crops of this region are wheat and maize. In 2016, the wheat planting area was 57.45405 million mu, and in 2017 it was 57.6435 million mu [69].

In this paper, we used the ground data and remote sensing data of Feicheng county, Ningyang county and Zhangqiu county, Shandong province. The three counties are similar in topography, all relatively flat, which can eliminate the influence of topographic fluctuations on the experimental results.

#### 2.1.2. Ground-Based Data

For manufacturing sample to train our model, we conducted a field survey in Feicheng county, Ningyang county and Zhangqiu county in 2017 and 2018, and obtained the land use data of 369 sample points, among which 257 were winter wheat sample points and 112 were bare land. The survey results include the time, location and type of land use.

#### 2.1.3. Remote Sensing Data

We selected 39 GF-2 remote sensing image, size of each image is  $7300 \times 6900$ . Of these images, 15 were captured on 17 February 2017, 11 were captured on 21 March 2018 and 13 were captured on 12 April 2018. We select images from different periods to increase the anti-interference abilities of the HSCNN. These remote sensing data cover Feicheng county, Ningyang county and Zhangqiu county, and are matched with ground investigation time. At the same time, the selected remote sensing data have fewer clouds and better clarity.

The Environment for Visualizing Images (ENVI) software was used for preprocessing the tasks, including fusion of panchromatic spectrum and multispectral band to obtain 1-m spatial resolution multispectral data, and the contrast stretch to generate a color-enhanced color composite image.

### 2.2. Network Architecture of Our Method

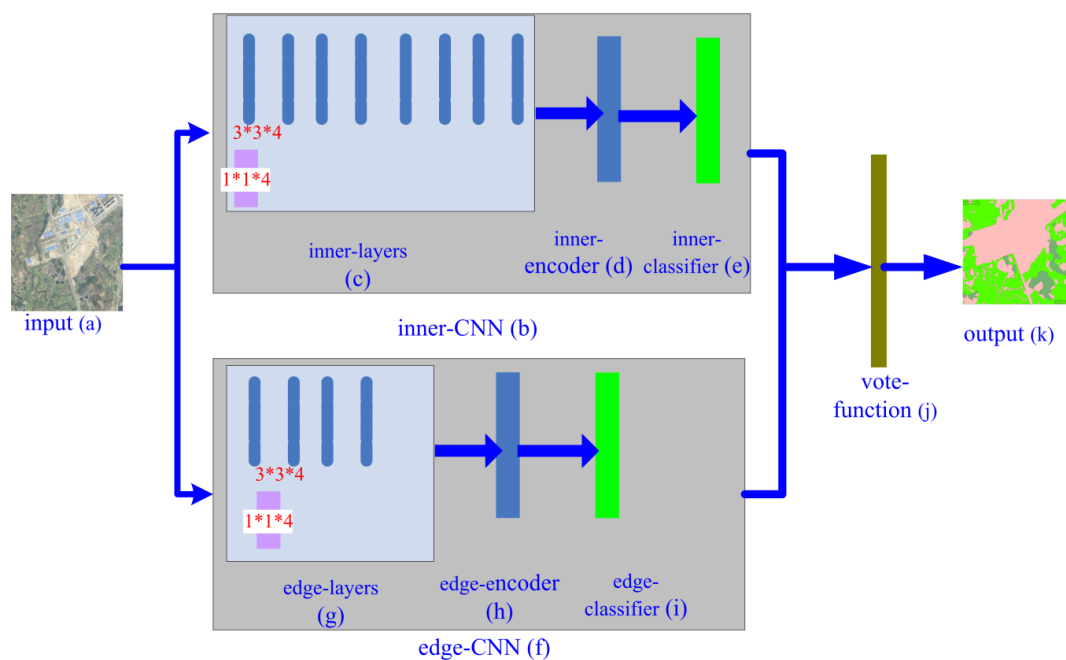
The HSCNN model is divided into five functional groups of components, input (a), inner-CNN (b), edge-CNN (f), vote function (j), and output (k), as shown in Figure 1. Both the edge-CNN and inner-CNN have convolution layers, an encoder layer, and a classifier layer. In the training stage, the inputs are original images and artificial classification labels. In the classification stage, the inputs are original GF-2 images, output is a single-band file, and content of each pixel in the output is the category number of the corresponding original image pixel. The HSCNN indicates the winter wheat area using category number 100, and category number 200 distinguishes other land use. The reason for adopting the two numbers is to fit with the coding value table we are working on to obtain detailed land use information.

#### 2.2.1. Inner-Layers and Edge-Layers

The operational characteristics of the pixel block-based convolution for image segmentation are described in Section 1, in addition to the effect of the pixel block location on the convolution results. Based on this analysis, two convolution sub-structures of different depths are setup for the feature extraction of the winter wheat field. The deep convolution sub-network is used to extract the features of the pixels in the interior of the winter wheat plantation, shown as inner-layers (c) in Figure 1.

The shallower sub-network is used to extract the features of the pixels at the edge of the winter wheat plantation, shown as edge-layers (g) in Figure 1. The benefits of this design are discussed in Section 4 based on the experimental results.

In our approach, an inner pixel refers to the pixel that only contains winter wheat pixels in the pixel block when convolution operation is carried out with the pixel as the center pixel. An edge pixel refers to the pixel that contains winter wheat pixels and other pixels when computing the feature of the pixel.



**Figure 1.** Network architecture of the Hybrid Structure Convolutional Neural Network (HSCNN): (a) input; (b) inner-CNN; (c) inner-layers; (d) inner-encoder; (e) inner-classifier; (f) edge-CNN; (g) edge-layers; (h) edge-encoder; (i) edge-classifier; (j) vote function; (k) output.

All kernels of the HSCNN take the form,  $w \times h \times c$ , where  $w$  is the width,  $h$  is the height, and  $c$  is the number of channels of a kernel. Two types of kernels are used in the first convolutional layers of inner-layers (c) and edge-layers (g). For one type  $w$  and  $h$  are set to 1, and for the other type the values are set to 3. In both cases,  $c$  is set to 4 because the data in the four multi-spectral bands of GF-2 are used. Kernels of the form  $1 \times 1 \times 4$  are used to extract the features of the pixels. The generated feature map is used instantaneously as the input of the encoder, and does not participate in the subsequent convolution. Convolution kernels of the form  $3 \times 3 \times 4$  are used to extract the spatial relation between the pixels and generate the spatial semantics by multi-level convolution.

After the operation of first convolution layer on the original image, we obtain a feature map which has only one channel. Because the input of convolution layer is the feature map calculated by the previous convolution layer, so the  $w$  and  $h$  values of the kernels used in all other convolutional layers are set to 3, and  $c$  is 1 from the second layer. To extract more features from the edge pixels of the crop field, the number of kernels used in each convolutional layer of edge-layers (g) is twice that used in the corresponding layer of inner-layers (c).

In the HSCNN, each convolution layer has only one activation layer attached, and there is no pool layer. Accordingly, the convolution result of each pixel block can be used directly as the feature of its central pixel, without the need to determine the position of the pixel that the feature corresponds to through deconvolution. As such, the HSCNN does not utilize a deconvolutional layer. This reduces the extent of computation and positioning error of the deconvolution, thereby improving the accuracy of the segmentation.

### 2.2.2. Inner-Encoder and Edge-Encoder

The inner-encoder and edge-encoder are used to encode the eigenvector extracted by the convolution layers on the pixel, ensuring that the classifier can establish the relationship between the eigenvector and pixel type. In the HSCNN model, the inner- and edge-encoders are both  $2 \times n$  matrices, where  $n$  is the length of the eigenvector.

Let  $X$  denote the eigenvector of the pixel,  $W$  denote the encoder matrix, and  $R$  the encoded vector result. The encoding calculation is displayed in Equation (1).

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \end{bmatrix} \times \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (1)$$

where each row of matrix  $w$  represents a fitting function for a particular type of pixel,  $b_1$  and  $b_2$  are the respective biases, and the corresponding component of  $r$  is the encoded value of eigenvector  $x$  on that pixel type. The inner- and edge-encoders are trained separately.

### 2.2.3. Inner-Classifier and Edge-Classifier

For each pixel, the inner-classifier converts its vector of the encoded values given by the inner-encoder into a probability distribution over a set of classes, and classifies the pixel as an inner pixel or a non-inner pixel of the winter wheat plantation based on the location of the component with the highest probability. Similarly, the edge-classifier distinguishes between the edge and non-edge pixels of the winter wheat field using the vectors of the encoded values generated on the pixels by the edge-encoder.

In reference to the classic softmax classifier [60–65], Equation (2) is used here to convert vector  $r$  of the encoded values to vector  $p$  of the class probabilities for each pixel.

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \frac{e^{r_1}}{e^{r_1} + e^{r_2}} \\ \frac{e^{r_2}}{e^{r_1} + e^{r_2}} \end{bmatrix} \quad (2)$$

After the transformation, the index of  $\max(p_1 \text{ and } p_2)$  is taken as the predicted category of the pixel. For the inner-classifier, index numbers of 1 and 0 are assigned to the inner pixels of the winter wheat field and other pixels, respectively. Accordingly, index numbers of 1 and 0 are assigned to the edge pixels of the winter wheat field and other pixels, respectively.

### 2.2.4. Vote-Function

The vote-function determines the category number of a pixel given by the inner-classifier and edge-classifier and writes it to the output file. As described in the beginning of Section 2.2, the HSCNN indicates the winter wheat area using category number 100 and other land uses using category number 200. The category number of a pixel is calculated in Equation (3).

$$o = \begin{cases} 100 & p_{\text{inner}} = 1 \text{ or } p_{\text{edge}} = 1 \\ 200 & p_{\text{inner}} = 0 \text{ and } p_{\text{edge}} = 0 \end{cases} \quad (3)$$

where  $o$  represents the final category number of a pixel and  $p_{\text{inner}}$  and  $p_{\text{edge}}$  are the outputs of the inner-classifier and edge-classifier, respectively.

## 2.3. HSCNN Training

We manually labeled all images at the pixel level as ground truth (GT) label data. In other words, for each image, there exists a  $7300 \times 6900$  label map, having a pixel-class (row-col indexed) correspondence with it. We used 36 images for training, and the remaining 3 images for testing. The GF-2 images and their corresponding artificial classification labels will be input to the HSCNN as training samples.



The training process includes error calculation, error back propagation and weight update. This process is iterated until the difference becomes smaller than the predetermined threshold.

We calculated the errors between the predicted classification label and manual classification label by the chain rule. The chain rule, the derivative rule in calculus, is used to find the derivative of a complex function, which is a common way to do the derivative calculation of calculus. The derivative of a composite function is the product of the derivatives of this finite number of functions at the corresponding point, as a chain. Then, the errors are back-propagated through the network. The backward propagation algorithm is a kind of training and learning method in deep learning, which can spread the error of the output layer backward to realize weight adjustment, adjust the weight between each node in the deep network, and achieve the goal that the sample tag output from the network is consistent with the actual tag. We use gradient descent method to update HSCNN parameters. Gradient descent method is the most commonly used optimization method. The idea is to use the negative gradient direction of the current position as the search direction, because that direction is the fastest descending direction of the current position.

### 2.3.1. Sample Labeling

We use the ENVI software for labeling and designing a preprocessor to build the labels. The process of artificial labeling is as follows:

1. The region-of-interest (RoI) tool in the ENVI software is used to select the winter wheat regions and other regions in the image. Then, the map locations of the pixels in each region are output to different files based on the category.
2. A band is added to the image file by the preprocessor as a mask band. The spatial resolution, size, and other parameters of the mask band are the same as the original image. Then, the category number of each pixel is written to the mask band according to the map location of the pixel previously output. We manually label all the images at the pixel level. Thus, for each image, there exists a  $7300 \times 6900$  label map, with a row-column-indexed pixel-class correspondence.
3. The pixels marked as winter wheat are further categorized as edge pixels and inner pixels. Based on the parameters given above, the inner-layers comprise of eight convolutional layers, each with a  $3 \times 3$  (length  $\times$  width) convolution kernel. Therefore, the feature extraction from pixel  $s$  involves a  $9 \times 9$  pixel block centered at  $s$  in the calculation. As defined in Section 2.2.1, the winter wheat pixels are divided sequentially into edge pixels and inner pixels. For training class by class, we use temporary code 160 to denote edge pixels and 170 to denote inner pixels in the mask band.

Figure 2 shows an example of an image-label pair.

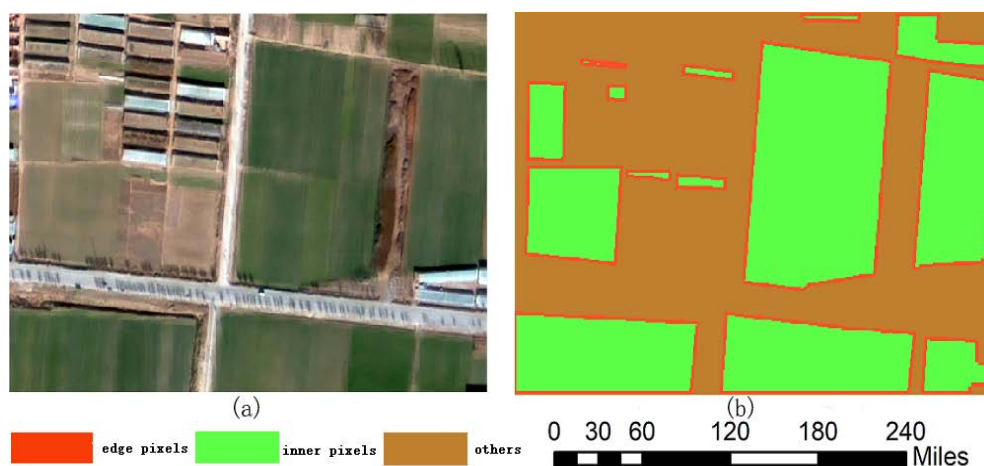


Figure 2. Image-label pair example: (a) original image; and (b) labels.

### 2.3.2. Loss Function

In our method, new loss functions are defined for the inner-CNN and edge-CNN, which still use the cross entropy as the basic element for the calculation, as expressed in Equation (4).

$$H(p, q) = - \sum_{i=1}^2 q_i \log(p_i) \quad (4)$$

where  $p$  and  $q$  are, respectively the predicted and actual probability distribution, and  $i$  is index of a component in the probability distribution. On this basis, the loss function of the inner-CNN is defined as

$$\text{loss} = - \frac{1}{m} \sum_m \sum_{i=1}^2 q_i \log(p_i) \quad (5)$$

When computing loss of inner-CNN,  $m$  is obtained by subtracting the number of edge pixels of the winter wheat field from the total number of samples, and when computing loss of edge-CNN,  $m$  is obtained by subtracting the number of inner pixels of the winter wheat field from the total number of samples.

### 2.3.3. Model Training

Images from two different periods were selected as the training data. We selected images from different periods for increasing the anti-interference abilities of the HSCNN and mitigating the complications such as the change in the seasons, and thus enhancing applicability. The training stage proceeded through the following steps:

1. Image-label pairs are input into the HSCNN as training samples. Network parameters are initialized.
2. Forward propagation is performed on the sample images.
3. The [loss]\_inner is calculated and back propagated to the inner-CNN, whereas the [loss]\_edge is calculated and back propagated to the edge-CNN.
4. The network parameters are updated using the stochastic gradient descent (SGD) [41,48] with momentum.
5. Steps (2)–(4) are iterated until both [loss]\_inner and [loss]\_edge are less than the predetermined threshold values.

The training yields two sub-networks, an inner-CNN and edge-CNN. The former can accurately extract the inner pixels of the winter wheat plantation from the sweet GF-2 remote sensing images, whereas the latter allows the best possible distinction between the edge pixels of the winter wheat planting region and other pixels.

In our training, the SGD method with momentum was used for parameter updates, which is illustrated in the following expression:

$$W^{(n+1)} = W^{(n)} - \Delta W^{(n+1)} \quad (6)$$

where  $W^{(n)}$  denote the old parameters,  $W^{(n+1)}$  denote the new parameters, and  $\Delta W^{(n+1)}$  is the increment in the current iteration, which is a combination of the old parameters, gradient, and historical increment, i.e.,

$$\Delta W^{(n+1)} = \vartheta \left( d_w \cdot W^{(n)} + \frac{\partial J(W)}{\partial W^{(n)}} \right) + m \cdot \Delta W^{(n)} \quad (7)$$

where  $J(W)$  is the loss function,  $\vartheta$  is the learning rate for step length control,  $d_w$  denotes the weight decay, and  $m$  denotes the momentum.



#### 2.4. Segmentation Using the Trained Network

After successful training, the HSCNN can be used to segment an input imagery pixel-by-pixel. According to our design, the output is written in a new band. The benefit of this design avoids damaging the original file.

### 3. Experiments and Results

The data used in the experiment are presented in Section 2.1. In this section, the models used for comparison are described in Section 3.1, and the experimental results and assessment of accuracy are given in Section 3.2.

#### 3.1. Comparison Model

Feature selection is the basis of remote sensing image segmentation. At present, there are mainly two methods based on artificial feature selection and machine learning. Haralick et al. (1973) put forward the method of gray level co-occurrence matrix, which is a classical artificial selection feature method. This method is mainly used to select image texture features. Since the texture is formed by repeated alternating changes of gray distribution in image space, so there is a certain gray-scale relationship between two separate pixels away certain distance, Haralick et al. described this correlation through a matrix [70]. Based on the artificial selection feature, only limited, shallow features can generally be selected. The feature selection based on machine learning can fully explore the deep feature and spatial semantic feature of the image. SegNet and DeepLab are classic semantic segmentation models of images, which have achieved very good results in the processing of images. Moreover, the working principles of these two models are very similar to our work, so we choose these two models as the comparison model, which can better reflect the advantages of our model in feature extraction. A comparative experiment was conducted using the methods established in the published literature.

##### 3.1.1. SegNet

For the SegNet model, we directly employed the structure proposed by Badrinarayanan et al. [64], which consists of an encoder, a decoder, and a classifier. The encoder uses the first 13 convolutional layers of the VGG16 network, each having its corresponding decoder layer, totaling 13 decoder layers. The last decoder generates a multi-channel feature map as the input to the classifier, which outputs a probability vector of length K, where K is the number of classes. The final predicted category corresponds to the class having maximum probability at each pixel. In terms of training, SegNet can be trained end-to-end using SGD.

##### 3.1.2. DeepLab

For DeepLab, we directly employed the DeepLab v3 model proposed by Liang-Chieh Chen et al. [66]. DeepLab was also developed based on the VGG network. To ensure that the output size would not be too small without excessive padding, DeepLab changes the stride of the pool4 and pool5 layers of the VGG network from the original 2 to 1, plus 1 padding. To compensate for the effect of the stride change on the receptive field, DeepLab uses a convolution method called “atrous convolution” to ensure that the receptive field after pooling remains unchanged and the output is more refined. Finally, DeepLab incorporates a fully connected conditional random field (CRF) model to refine the segmentation boundary.

#### 3.2. Results and Result Comparison

In the comparative experiment, we applied our trained model to three GF-2 images for segmentation. These images were only used for testing and not involved in training. Figure 3 illustrates the results obtained from the comparison methods and proposed method. In Figure 3, the first column

illustrates the results of Experiment 1, the second column illustrates the results of Experiment 2 and the third column illustrates the results of Experiment 3.

Tables 1–3 are confusion matrices c for the segmentation results of SegNet model, DeepLab model, and HSCNN model, respectively. Each row of the confusion matrix represents the proportion taken by the actual category, and each column represents the proportion taken by the predicted category. As can be seen from the tables, our method achieves better classification results. In the example above, the proportion of “winter wheat” wrongly categorized as “background” is on average 0.069, and the proportion of “background” wrongly classified as “winter wheat” is on average 0.019, resulting in an overall accuracy of 91.2%.

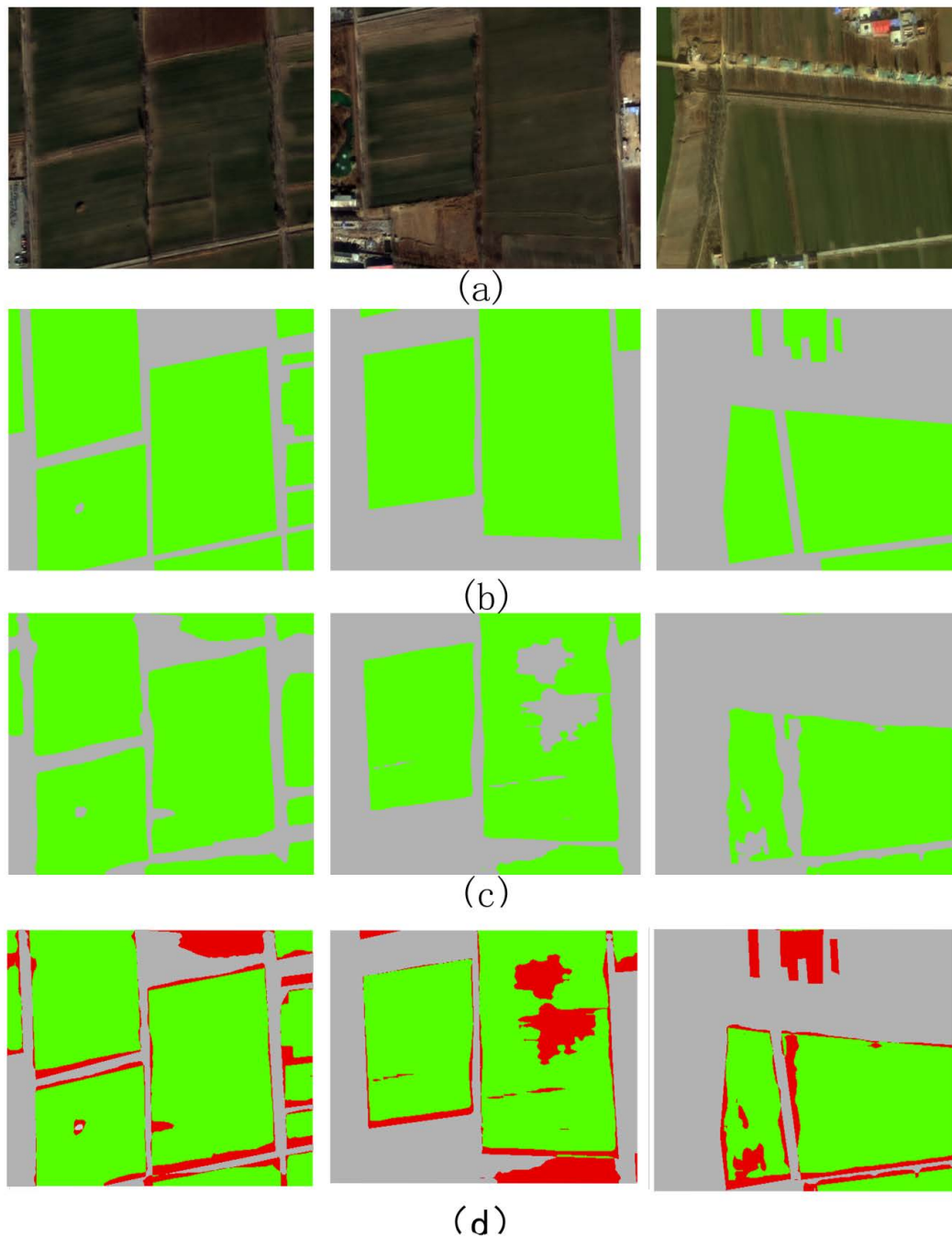
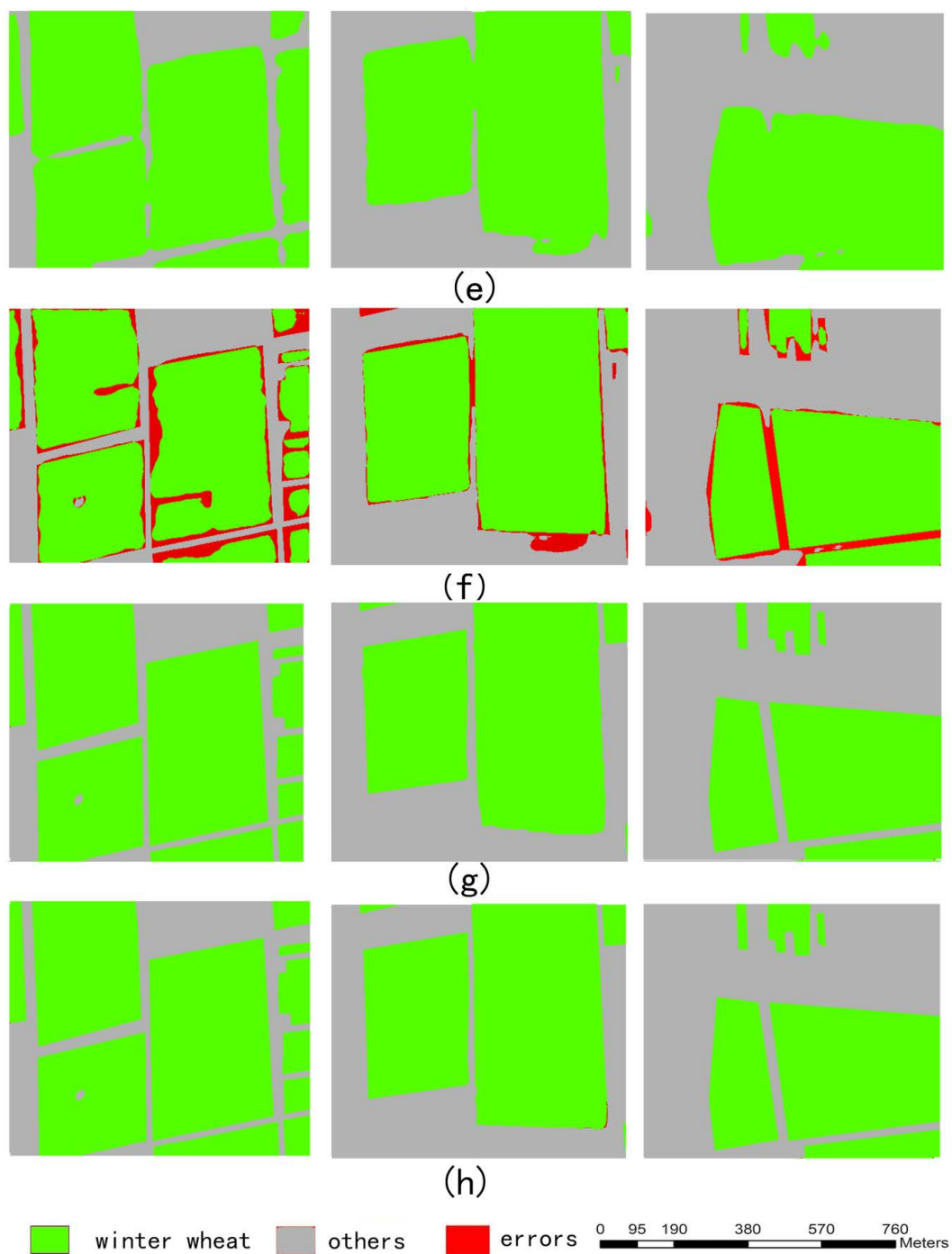


Figure 3. Cont.



**Figure 3.** Segmentation results for Gaofen-2 (GF-2) images: (a) original images; (b) ground truth, (c) results of SegNet corresponding to the images in (a); (d) errors of SegNet; (e) results of DeepLab; (f) errors of DeepLab; (g) results of HSCNN; and (h) errors of HSCNN.

**Table 1.** Confusion matrix of the SegNet approach for Figure 4.

Experiment	GT/Predicted	Winter Wheat	Others
Experiment-1	winter wheat	0.621	0.129
	Others	0.087	0.163
Experiment-2	winter wheat	0.387	0.153
	Others	0.084	0.376
Experiment-3	winter wheat	0.217	0.123
	Others	0.129	0.531

**Table 2.** Confusion matrix of the DeepLab approach for Figure 4.

Experiment	GT/Predicted	Winter Wheat	Others
Experiment-1	winter wheat	0.653	0.107
	Others	0.055	0.185
Experiment-2	winter wheat	0.432	0.086
	Others	0.039	0.443
Experiment-3	winter wheat	0.301	0.108
	Others	0.045	0.546

**Table 3.** Confusion matrix of our HSCNN approach for Figure 4.

Experiment	GT/Predicted	Winter Wheat	Others
Experiment-1	winter wheat	0.681	0.075
	Others	0.027	0.217
Experiment-2	winter wheat	0.458	0.062
	Others	0.013	0.467
Experiment-3	winter wheat	0.329	0.071
	Others	0.017	0.583

Accuracy, precision, recall, and the Kappa coefficient were used to evaluate the models. These indices are calculated via mixing matrix c.

Accuracy is the ratio of the number of correctly classified samples to the total number of samples, and is given in this case by the following equation:

$$\text{Accuracy} = \frac{\sum_{i=1}^2 C_{ii}}{\sum_{i=1}^2 \sum_{j=1}^2 C_{ij}} \quad (8)$$

Here,  $C_{ii}$  denotes the number of correctly classified samples, and  $C_{ij}$  denotes the number of samples of class  $i$  misidentified as class  $j$ .

Precision denotes the average proportion of pixels correctly classified to one class from the total retrieved pixels. Precision is calculated as:

$$\text{Precision} = \frac{1}{2} \sum_i C_{ii} / \sum_j C_{ij} \quad (9)$$

Recall represents the average proportion of pixels that are correctly classified in relation to the actual total pixels of a given class. Recall is computed as:

$$\text{Recall} = \frac{1}{2} \sum_i C_{ii} / \sum_i C_{ij} \quad (10)$$


The Kappa coefficient measures the consistency of the predicted classes with artificial labels. The Kappa coefficient is computed as:

$$\text{Kappa} = \frac{\frac{\sum_{i=1}^2 C_{ii}}{\sum_{i=1}^2 \sum_{j=1}^2 C_{ij}} - \frac{\sum_{i=1}^2 C_{ii} \sum_{j=1}^2 C_{ij}}{(\sum_{i=1}^2 \sum_{j=1}^2 C_{ij})^2}}{1 - \frac{\sum_{i=1}^2 C_{ii} \sum_{j=1}^2 C_{ij}}{(\sum_{i=1}^2 \sum_{j=1}^2 C_{ij})^2}} \quad (11)$$

Equations (8)–(11) use the definitions given in Reference [18] and are modified according to our actual situation. The minimum accepted precision is 89% according to practical application.

The indicator values are listed in Table 4.

**Table 4.** Comparison of the approaches using SegNet, DeepLab, and HSCNN.

Approach	Index	Experiment-1	Experiment-2	Experiment-3	Average
SegNet	Accuracy	0.784	0.763	0.748	0.765
	Precision	0.740	0.767	0.721	0.743
	Recall	0.718	0.766	0.720	0.734
	Kappa	0.579	0.617	0.564	0.586
DeepLab	Accuracy	0.838	0.875	0.847	0.853
	Precision	0.815	0.877	0.830	0.840
	Recall	0.778	0.877	0.852	0.836
	Kappa	0.665	0.778	0.716	0.720
 HSCNN	Accuracy	0.898	0.925	0.912	0.912
	Precision	0.895	0.927	0.897	0.906
	Recall	0.853	0.927	0.921	0.900
	Kappa	0.776	0.860	0.826	0.821

#### 4. Analysis and Discussion

From the experimental results in Section 3, it is clear that our method significantly improves the accuracy of winter wheat extraction. In this section, the advantages of our model are discussed in terms of the differences between the remote sensing images and camera images. This is followed by more specific comparisons with SegNet and DeepLab. Finally, the role of our model in the classification of land uses by remote sensing is discussed briefly.

##### 4.1. Advantages of the HSCNN Model

CNNs have achieved significant success in camera image segmentation, which has motivated researchers to apply them to remote sensing images. The HSCNN model proposed here is developed based on a previous work followed by a further in-depth analysis of the fundamental difference between camera images and remote sensing images. Thus, it possesses clear advantages compared with the traditional practice of the straightforward application of camera image segmentation model to remote sensing images.

Camera images and remote sensing images essentially differ in information representation. Owing to their advantages in shooting distance and the pixel quality of the camera, camera images are superior in terms of the rich details they contain, such that one object is formed by multiple pixels. Thus, the color of a pixel reflects the information at a certain point on an object but not the spatial relation between the pixels, which is found and expressed only by convolution. The nature of convolution is to represent the spatial correlation between the pixels by constructing a complex fitting function by operating on the pixel value of a pixel block. Particularly because it makes good use of the essential characteristics of camera images, deep convolution is extremely successful in camera image processing.

In remote sensing images, particularly for crop fields, a pixel generally contains multiple objects. For example, generally in GF-2 images  $1 \text{ m}^2$  of ground is covered by a pixel, which contains 600–700 winter wheat plants. A pixel embodies the color information of the plants and the spatial information between them. However, at the edge of a winter wheat field, the region covered by a pixel is often a mixture of the winter wheat and bare land or winter wheat and other geographical objects, with varying percentages of winter wheat in the space. In this view, the information contained in a pixel at the edge region is significantly different from that at the interior. These two types of pixels can even be regarded as two different types of objects.

Based on the above analyses, the HSCNN network architecture is designed with a complete consideration of the properties of the remote sensing image and extraction target, and it makes good use of the characteristics of the winter wheat field captured in the GF-2 remote sensing images. The strengths of this model are exhibited in the following three aspects:

1. Considering the significant difference between the pixels of the interior and edge region of the winter wheat plantation (during extraction), these two regions are treated as two subclasses. Accordingly, the features of the inner pixels are more focused, which facilitates the model training. Two sub-networks with different depths are then designed with respect to the characteristics of the two subclasses. The deep sub-network extracts the pixels at the interior, whereas the shallower sub-network extracts those at the edge. This scheme reduces the effect of the non-winter wheat pixels on the features and improves the stability of the model for edge pixel extraction.
2. Two types of kernels are used in the first convolutional layer. The  $1 \times 1 \times 4$  kernels are used to extract the feature of the pixels, and the  $3 \times 3 \times 4$  kernels are used to extract the spatial relation between pixels. This design takes advantage of the ability of convolution for extracting higher-level spatial semantics and for obtaining the rich pixel information contained in the remote sensing images.
3. Our model does not utilize pooling, instead the convolution result is taken as the eigenvalue of the central pixel of the pixel block. In the application of the convolutional network to image classification, the basic target (sample) for the classification is the entire image. Thus, pooling can produce the main features of the feature map and reduce the amount of subsequent computation. Although the information on the accurate position of the features is lost during this process, their relative positions are nevertheless retained, which ensures the normal operation of the subsequent computational steps. However, in image segmentation, the basic target (sample) for the classification is an individual pixel, whose exact location must be mapped by the eigenvalue. Therefore, the major advantage of our model is its ability to preserve the spatial location of the eigenvalue, which makes it possible to remove the deconvolution adopted by the traditional FCN. Accordingly, the amount of computation is reduced. Further, the loss in precision due to positioning error is reduced, as the accurate position of the eigenvalue is kept.

#### 4.2. Comparison with SegNet and Analysis

SegNet is founded on the FCN model. Its main strength lies in the search and extraction of the rich details of an image by deep convolution, and it is very distinct when extracting target objects with relatively few pixels. If the target objects contain only a few pixels or even one pixel, the deep convolution does not generate more details and may introduce more noise owing to the expanded field of view, affecting the determination of the pixel type.

In the remote sensing images of GF-2, the edge and interior of the winter wheat plantation are very different in composition, which makes it more difficult for SegNet to locate the common features, because of its structure containing a single convolutional network. In comparison, the HSCNN is equipped with two sub-networks of different depths and is adaptable to the characteristics of the edge and interior. It also uses two different sizes of kernel, which are capable of uncovering the spatial relation between the pixels, and the information embedded in the pixels.



As shown in Figure 3, the segmentation results of HSCNN and SegNet are nearly identical for the interior of the winter wheat field. SegNet, however, produces prominent errors at the edge of the field, while HSCNN does not.

Both HSCNN and SegNet use classifiers to generate the probability distribution of the classes, and consider the class with the maximum probability (max) in the distribution as the type to which the pixel belongs. Clearly, a larger difference between the max and background implies a higher separability of the pixels and more reliable results. The probability differences given by the HSCNN and SegNet model for the inner wheat and edge wheat classes are presented in Figures 4 and 5, respectively. It is clear in Figure 4 that HSCNN and SegNet lead to significant probability differences for many pixels in the interior, which demonstrates the high separability of this region and the strength of CNN. In the probability distribution in Figure 5, fewer pixels are noted as having large probability differences in both the HSCNN and SegNet; nevertheless, the number is maintained at a quite high level for the HSCNN, whereas SegNet exhibits a reduced performance.

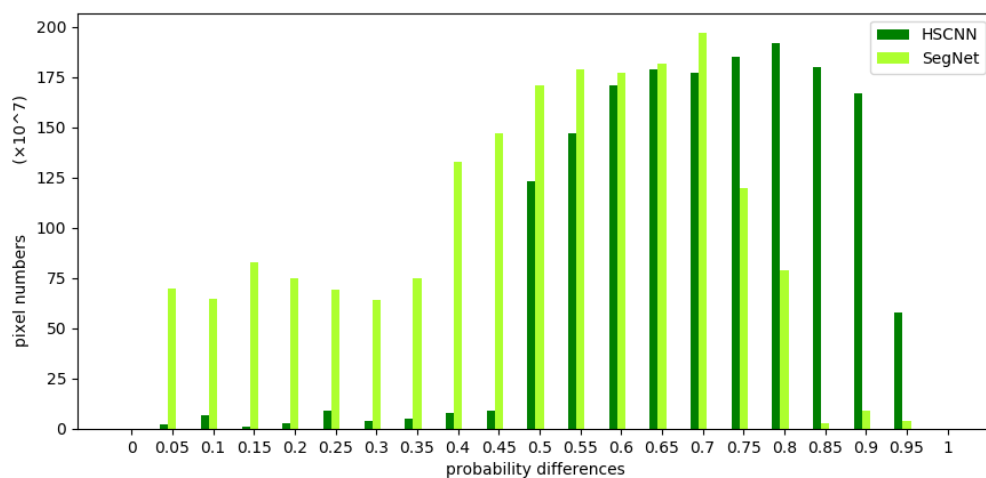


Figure 4. Distribution of the probability differences for the inner wheat pixels.

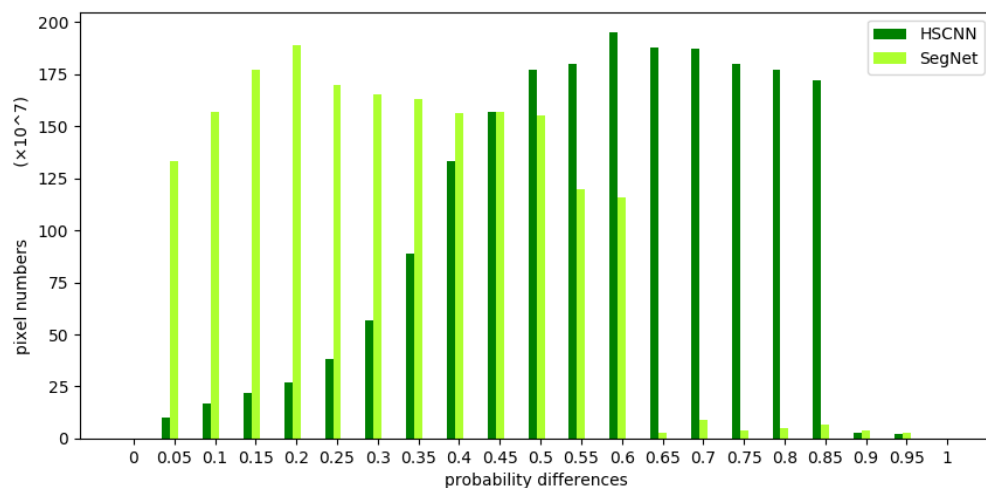


Figure 5. Distribution of the probability differences for the edge wheat pixels.

#### 4.3. Comparison with DeepLab and Analysis

Compared with the FCN and SegNet, DeepLab has significant improvements in two aspects: (1) the deconvolution; and (2) the refinement of the boundary of the segmentation result by fully connected CRFs. These two improvements are beneficial for the segmentation of individual objects covering numerous pixels. Based on the published literature, DeepLab displays a higher segmentation

accuracy at the boundary than the FCN and SegNet, because it better utilizes the detailed information contained in the image and the large-scale spatial correlation between the pixels. However, in its application to winter wheat identification, the strength of DeepLab is not fully realized, because the details within a pixel block of the winter wheat plantation do not change significantly. Therefore, less information is available to the model, and the spatial correlation within the farmlands and woods is not strong over large regions.

As mentioned in Section 4.2, the HSCNN completely utilizes the characteristics of the pixels and the spatial relation between them. Therefore, it is well adapted to the data characteristics of the winter wheat plantation. Further, it effectively avoids the deficiencies of DeepLab and ensures the accuracy of segmentation.

As in Section 4.2, the probability differences between the HSCNN and DeepLab models in the inner wheat and edge wheat class are displayed in Figures 6 and 7, respectively. It is clear in Figure 6 that both the models produce large probability differences for many pixels in the interior. In the probability distribution of Figure 7, a considerable number of pixels still display large probability differences after the HSCNN processing, whereas DeepLab shows a much poorer performance (even lower than SegNet), proving again the notion that the atrous convolution is not suitable for farmlands.

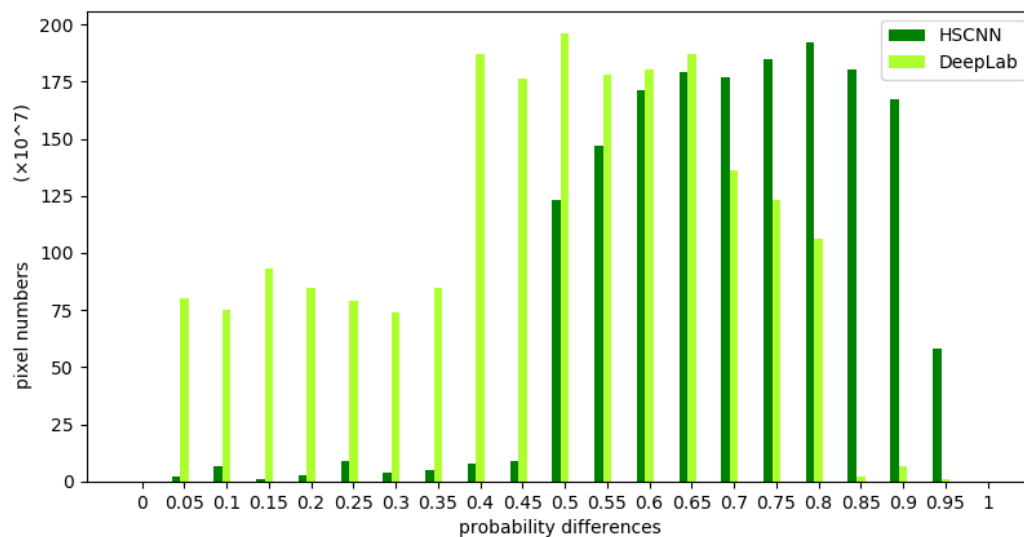


Figure 6. Distribution of the probability differences for the inner wheat pixels.

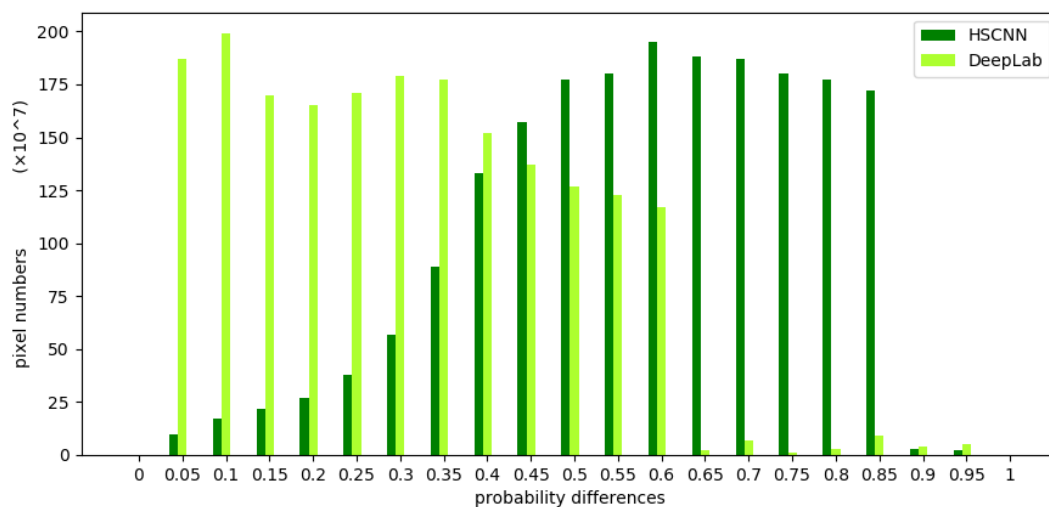


Figure 7. Distribution of the probability differences for the edge wheat pixels.

#### 4.4. Benefits of Using the Proposed Approach to Classify Land Use

Accurate land use classification is of tremendous importance in scientific research and agriculture with the use of remote sensing data as an increasingly common practice for this purpose. Based on a CNN and taking complete advantage of the convolution in feature extraction, the design of the CNN architecture adapting to the features of the remote-sensing images is the key in land use classification by this method.

We have taken the feature difference between the edge pixel and the inner pixel in the white wheat planting area into full consideration, this significantly improve the extraction accuracy of the edge pixel. Compared with earlier research, the model presented in this paper has the following advantages.

Firstly, two types of kernels were used in the convolution of the model, which allowed the full utilization of the strength of the convolution in the extraction of spatial semantics and made appropriate use of the rich information contained in the pixels of the remote sensing images, thus achieving a more accurate segmentation.

Secondly, pooling layers were not used in the model. Although the speed of the feature aggregation was consequently reduced, the information of the exact location to which an eigenvalue corresponds was retained, thereby effectively mitigating the loss in the accuracy due to the positioning error of the deconvolution and improving the overall effect of the segmentation.

The model presented in this paper provides a solution for the edge extraction problem or the segmentation of the winter wheat plantation using GF-2 images. It has an important role to play and enhances the efficiency of the agricultural survey. This model has been utilized by the Department of Agriculture and the Meteorological Bureau of Shandong Province, China.

#### 5. Conclusions

This paper presents a novel approach for the extraction of the winter wheat distribution from GF-2 remote-sensing images. Compared with the two typical deep learning-based approaches, the extraction accuracy is obviously improved. Our approach combines the segmentation and classification stages, taking the accuracy as the only constraint, and achieves high quality classification in an end-to-end way. The GT classes of ground objects are taken as the supervised information that guides both the feature extraction and the region generation. Taking into account the significant differences between pixel and edge pixel in the planting area, different convolution structures were used to extract the feature of edge and interior pixels, focusing on the common features in the two subclasses for more effective model training, and obtained a high resolution class prediction.

Our model is still limited in many aspects, and further improvements could be made in the following two areas: (1) The current encoder uses a relatively simple regression algorithm for encoding; thus, a regression that can express the complex relationship between the eigenvalues needs to be explored. (2) A new pooling method, which allows for expedited feature aggregation without the loss of the spatial information of the eigenvalues, should be established. We will continue our work in the future to improve the current model and obtain better classification performance.

**Author Contributions:** C.Z. wrote the manuscript; C.Z. and S.G. presented the direction of this study and designed the experiments; X.Y. and F.L. carried out the experiments and analyzed the results; M.Y. and Y.Z. carried out ground investigation and preprocessing; and Y.H., H.Z. and K.F. carried out sample labeling.

**Funding:** This work was funded by National Key R&D Program of China, grant number 2017YFA0603004; Science Foundation of Shandong, grant numbers ZR2017MD018 and ZR2016DP01; the National Science Foundation of China, grant number 41471299; and Open research project of Key Laboratory on meteorological disaster monitoring, early warning and risk management in characteristic agricultural areas of arid area, grant numbers CAMF-201701 and CAMF-201803.

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

## References

1. Announcement of the National Statistics Bureau on Grain Output in 2017. Available online: [http://www.gov.cn/xinwen/2017-12/08/content\\_5245284.htm](http://www.gov.cn/xinwen/2017-12/08/content_5245284.htm) (accessed on 8 December 2017).
2. Wang, L.M.; Liu, J.; Yao, B.M.; Ji, F.H.; Yang, F.G. Area change monitoring of winter wheat based on relationship analysis of GF-1 NDVI among different years. *Trans. CSAE* **2018**, *34*, 184–191. [\[CrossRef\]](#)
3. He, H.; Zhu, X.F.; Pan, Y.Z.; Zhu, W.Q.; Zhang, J.S.; Jia, B. Study on scale issues in measurement of winter wheat plant area by remote sensing. *J. Remote Sens.* **2008**, *1*, 168–176.
4. Zhang, J.H.; Feng, L.L.; Yao, F.M. Improved maize cultivated area estimation over a large scale combining MODIS-EVI time series data and crop phenological information. *ISPRS J. Photogramm. Remote Sens.* **2014**, *94*, 102–113. [\[CrossRef\]](#)
5. Wu, M.Q.; Wang, C.Y.; Niu, Z. Mapping paddy field in large areas, based on time series multi-sensors data. *Trans. CSAE* **2010**, *26*, 240–244.
6. Xu, Q.Y.; Yang, G.J.; Long, H.L.; Wang, C.C.; Li, X.C.; Huang, D.C. Crop information identification based on MODIS NDVI time-series data. *Trans. CSAE* **2014**, *30*, 134–144.
7. Becker-Reshef, I.; Vermote, E.; Lindeman, M.; Justice, C. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* **2010**, *114*, 1312–1323. [\[CrossRef\]](#)
8. Zhang, J.G.; Li, X.W.; Wu, Y.L. Object oriented estimation of winter wheat planting area using remote sensing data. *Trans. CSAE* **2008**, *24*. [\[CrossRef\]](#)
9. Zhu, C.M.; Luo, J.C.; Shen, Z.F.; Chen, X. Winter wheat planting area extraction using multi-temporal remote sensing data based on filed parcel characteristic. *Trans. CSAE* **2011**, *27*, 94–99.
10. Lu, L.L.; Guo, H.D. Extraction method of winter wheat phenology from time series of SPOT/VEGETATION data. *Trans. CSAE* **2009**, *25*, 174–179.
11. Jha, A.; Nain, A.S.; Ranjan, R. Wheat acreage estimation using remote sensing in tarai region of Uttarakhand. *Vegetos* **2013**, *26*, 105–111. [\[CrossRef\]](#)
12. Wu, M.Q.; Yang, L.C.; Yu, B.; Wang, Y.; Zhao, X.; Niu, Z.; Wang, C.Y. Mapping crops acreages based on remote sensing and sampling investigation by multivariate probability proportional to size. *Trans. CSAE* **2014**, *30*, 146–152.
13. You, J.; Pei, Z.Y.; Wang, F.; Wu, Q.; Guo, L. Area extraction of winter wheat at county scale based on modified multivariate texture and GF-1 satellite images. *Trans. CSAE* **2016**, *32*, 131–139. [\[CrossRef\]](#)
14. Wang, L.M.; Liu, J.; Yang, F.G.; Fu, C.H.; Teng, F.; Gao, J. Early recognition of winter wheat area based on GF-1 satellite. *Trans. CSAE* **2015**, *31*, 194–201. [\[CrossRef\]](#)
15. Ma, S.J.; Yi, X.S.; You, J.; Guo, L.; Lou, J. Winter wheat cultivated area estimation and implementation evaluation of grain direct subsidy policy based on GF-1 imagery. *Trans. CSAE* **2016**, *32*, 169–174. [\[CrossRef\]](#)
16. Wang, L.M.; Liu, J.; Yang, L.B.; Yang, F.G.; Teng, F.; Wang, X.L. Remote sensing monitoring winter wheat area based on weighted NDVI index. *Trans. CSAE* **2016**, *32*, 127–135. [\[CrossRef\]](#)
17. Wu, M.Q.; Huang, W.J.; Niu, Z.; Wang, Y.; Wang, C.Y.; Li, W.; Hao, P.Y.; Yu, B. Fine crop mapping by combining high spectral and high spatial resolution remote sensing data in complex heterogeneous areas. *Comput. Electron. Agric.* **2017**, *139*, 1–9. [\[CrossRef\]](#)
18. Fu, G.; Liu, C.J.; Zhou, R.; Sun, T.; Zhang, Q.J. Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sens.* **2017**, *9*, 498. [\[CrossRef\]](#)
19. Liu, Y.D.; Cui, R.X. Segmentation of Winter Wheat Canopy Image Based on Visual Spectral and Random Forest Algorithm. *Spectrosc. Spect. Anal.* **2015**, *35*, 3480–3485.
20. Dong, Z.P.; Wang, M.; Li, D.R. A High Resolution Remote Sensing Image Segmentation Method by Combining Superpixels with Minimum Spanning Tree. *Acta Geod. Cartogr. Sin.* **2017**, *46*, 734–742. [\[CrossRef\]](#)
21. Basaeed, E.; Bhaskar, H.; Al-Mualla, M. Supervised remote sensing image segmentation using boosted convolutional neural networks. *Knowl.-Based Syst.* **2016**, *99*, 19–27. [\[CrossRef\]](#)
22. Liu, D.W.; Han, L.; Han, X.Y. High Spatial Resolution Remote Sensing Image Classification Based on Deep Learning. *Acta Opt. Sin.* **2016**, *36*, 0428001. [\[CrossRef\]](#)
23. Luo, B.; Zhang, L.P. Robust autodual morphological profiles for the classification of high-resolution satellite images. *IEEE Trans. Geosci. Remote* **2014**, *52*, 1451–1462. [\[CrossRef\]](#)

24. Li, D.R.; Zhang, L.P.; Xia, G.S. Automatic Analysis and Mining of Remote Sensing Big Data. *Acta Geod. Cartogr. Sin.* **2014**, *43*, 1211–1216. [[CrossRef](#)]
25. Chan, T.H.; Jia, K.; Guo, S.H.; Lu, J.; Zeng, Z.; Ma, Y. PCANet: A Simple Deep Learning Baseline for Image Classification. *IEEE Trans. Image Process.* **2015**, *24*, 5017–5032. [[CrossRef](#)] [[PubMed](#)]
26. Mas, J.F.; Flores, J.J. The application of artificial neural networks to the analysis of remotely sensed data. *Int. J. Remote Sens.* **2008**, *29*, 617–663. [[CrossRef](#)]
27. Gustavo, C.V.; Bruzzone, L. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1351–1362.
28. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [[CrossRef](#)]
29. Pacifici, F.; Chini, M.; Emery, W.J. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sens. Environ.* **2009**, *113*, 1276–1292. [[CrossRef](#)]
30. Huang, X.; Zhang, L. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high resolution remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 257–272. [[CrossRef](#)]
31. Liu, C.; Hong, L.; Chen, J.; Chun, S.S.; Deng, M. Fusion of pixel-based and multi-scale region-based features for the classification of high-resolution remote sensing image. *J. Remote Sens.* **2015**, *19*, 228–239. [[CrossRef](#)]
32. Yuan, Y.; Lin, J.; Wang, Q. Hyperspectral Image Classification via Multitask Joint Sparse Representation and Stepwise MRF Optimization. *IEEE Trans. Cybern.* **2016**, *46*, 2966–2977. [[CrossRef](#)] [[PubMed](#)]
33. Xie, F.D.; Li, F.F.; Lei, C.K.; Ke, L.N. Representative Band Selection for Hyperspectral Image Classification. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 338. [[CrossRef](#)]
34. Bengio, Y. Learning deep architectures for AI. *Found. Trends Mach. Learn.* **2009**, *2*, 1–127. [[CrossRef](#)]
35. Larochelle, H.; Bengio, Y.; Louradour, J.; Lamblin, P. Exploring strategies for training deep neural networks. *J. Mach. Learn. Res.* **2009**, *10*, 1–40.
36. Jones, N. The learning machines. *Nature* **2014**, *505*, 146–148. [[CrossRef](#)] [[PubMed](#)]
37. Gao, Q.S.; Lim, S.S.; Jia, X.P. Hyperspectral Image Classification Using Convolutional Neural Networks and Multiple Feature Learning. *Remote Sens.* **2018**, *10*, 299. [[CrossRef](#)]
38. Dong, Y.; Liu, Y.N.; Lian, S.G. Automatic age estimation based on deep learning algorithm. *Neurocomputing* **2016**, *187*, 4–10. [[CrossRef](#)]
39. Taormina, R.; Chau, K.W. Data-driven input variable selection for rainfall–runoff modeling using binary-coded particle swarm optimization and Extreme Learning Machines. *J. Hydrol.* **2015**, *529*, 1617–1632. [[CrossRef](#)]
40. Liang, Z.; Shan, S.; Liu, X.; Wen, Y. Fuzzy prediction of AWJ turbulence characteristics by using typical multi-phase flow models. *Eng. Appl. Comput. Fluid Mech.* **2017**, *11*, 225–257. [[CrossRef](#)]
41. Bellary, S.A.I.; Adhav, R.; Siddique, M.H.; Chon, B.H.; Kenyery, F.; Samad, A. Application of computational fluid dynamics and surrogate-coupled evolutionary computing to enhance centrifugal-pump performance. *Eng. Appl. Comput. Fluid Mech.* **2016**, *10*, 171–181. [[CrossRef](#)]
42. Zhang, J.; Chau, K.W. Multilayer Ensemble Pruning via Novel Multi-sub-swarm Particle Swarm Optimization. *J. Univ. Comput. Sci.* **2009**, *15*, 840–858.
43. Wang, W.C.; Chau, K.W.; Xu, D.M.; Chen, X.Y. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resour. Manag.* **2015**, *29*, 2655–2675. [[CrossRef](#)]
44. Zhang, S.; Chau, K.W. Dimension reduction using semi-supervised locally linear embedding for vegetation leaf classification. *Emerg. Intell. Comput. Technol. Appl.* **2009**, *5754*, 948–955.
45. Wu, C.; Chau, K.; Fan, C. Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *J. Hydrol.* **2010**, *389*, 146–167. [[CrossRef](#)]
46. Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. *arXiv*, 2015; arXiv:1508.00092.
47. Hu, F.; Xia, G.S.; Hu, J.; Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [[CrossRef](#)]
48. Zhu, C.; Cheng, T. Research on geological hazard identification based on deep learning. In Proceedings of the 6th International Conference on Computer-Aided Design, Manufacturing, Modeling and Simulation, Busan, Korea, 14–15 April 2018. [[CrossRef](#)]

49. Wu, Z.; Zhang, Q. On combining spectral, textural and shape features for remote sensing image segmentation. *Acta Geod. Cartogr. Sin.* **2013**, *42*, 44–50.
50. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1520–1528.
51. Paisitkriangkrai, S.; Sherrah, J.; Janney, P.; Hengel, V.D. Effective semantic pixel labelling with convolutional networks and conditional random fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 36–43.
52. Papandreou, G.; Kokkinos, I.; Savalle, P.A. Untangling local and global deformations in deep convolutional networks for image classification and sliding window detection. *arXiv*, 2014; arXiv:1412.0296.
53. Badrinarayanan, V.; Handa, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *arXiv*, 2015; arXiv:1505.07293.
54. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv*, 2015; arXiv:1502.03167.
55. Liu, J.; Liu, B.; Lu, H. Detection guided deconvolutional network for hierarchical feature learning. *Pattern Recognit.* **2015**, *48*, 2645–2655. [[CrossRef](#)]
56. Volpi, M.; Ferrari, V. Semantic segmentation of urban scenes by learning local class interactions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
57. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv*, 2014; arXiv:1409.1556.
58. Mittal, A.; Hooda, R.; Sofat, S. LF-SegNet: A Fully Convolutional Encoder-Decoder Network for Segmenting Lung Fields from Chest Radiographs. *Wirel. Pers. Commun.* **2018**, *101*, 511–529. [[CrossRef](#)]
59. Kendall, A.; Badrinarayanan, V.; Cipolla, R. Bayesian SegNet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv*, 2015; arXiv:1511.02680.
60. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected Crfs. *arXiv*, 2014; arXiv:1412.7062.
61. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
62. Långkvist, M.; Kiselev, A.; Alirezaie, M.; Loutfi, A. Classification and segmentation of satellite orthoimagery using convolutional neural networks. *Remote Sens.* **2016**, *8*, 239. [[CrossRef](#)]
63. Dolz, J.; Desrosiers, C.; Ben, A. 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *NeuroImage* **2018**, *170*, 456–470. [[CrossRef](#)] [[PubMed](#)]
64. 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](#)]
65. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv*, 2015; arXiv:1505.04597.
66. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv*, 2016; arXiv:1606.00915.
67. Lin, H.N.; Shi, Z.W.; Zou, Z.X. Maritime Semantic Labeling of Optical Remote Sensing Images with Multi-Scale Fully Convolutional Network. *Remote Sens.* **2017**, *9*, 480–501. [[CrossRef](#)]
68. Visin, F.; Ciccone, M.; Romero, A.; Kastner, K.; Cho, K.; Bengio, Y.; Matteucci, M.; Courville, A. Reseg: A recurrent neural network-based model for semantic segmentation. *arXiv*, 2016; arXiv:1511.07053.
69. Statistical Yearbook of Shandong Province. Available online: <http://www.stats-sd.gov.cn/col/col6279/index.html> (accessed on 27 October 2017).
70. Haralick, R.M.; Shanmugam, K. Texture features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [[CrossRef](#)]

