remote sensing

MDPI

Article

Intelligent Mapping of Urban Forests from High-Resolution Remotely Sensed Imagery Using Object-Based U-Net-DenseNet-Coupled Network

Shaobai He [1,2,3], Huaqiang Du [1,2,3,*], Guomo Zhou [1,2,3], Xuejian Li [1,2,3], Fangjie Mao [1,2,3], Di'en Zhu [4], Yanxin Xu [1,2,3], Meng Zhang [1,2,3], Zihao Huang [1,2,3], Hua Liu [1,2,3] and Xin Luo [1,2,3]

1 State Key Laboratory of Subtropical Silviculture, Zhejiang A & F University, Hangzhou 311300, China; 2018103241005@stu.zafu.edu.cn (S.H.); zhougm@zafu.edu.cn (G.Z.); 2017303661004@stu.zafu.edu.cn (X.L.); maofj@zafu.edu.cn (F.M.); xuyanxin@stu.zafu.edu.cn (Y.X.); 2017103242004@stu.zafu.edu.cn (M.Z.); 2018103241008@stu.zafu.edu.cn (Z.H.); 2018103242003@stu.zafu.edu.cn (H.L.); 2019602041026@stu.zafu.edu.cn (X.L.)

2 Key Laboratory of Carbon Cycling in Forest Ecosystems and Carbon Sequestration of Zhejiang Province, Zhejiang A & F University, Hangzhou 311300, China

3 School of Environmental and Resources Science, Zhejiang A & F University, Hangzhou 311300, China

4 The College of Forestry, Beijing Forestry University, Beijing 100083, China; 2015116021018@stu.zafu.edu.cn

* Correspondence: duhuaqiang@zafu.edu.cn

Received: 26 October 2020; Accepted: 25 November 2020; Published: 30 November 2020

Abstract: The application of deep learning techniques, especially deep convolutional neural networks (DCNNs), in the intelligent mapping of very high spatial resolution (VHSR) remote sensing images has drawn much attention in the remote sensing community. However, the fragmented distribution of urban land use types and the complex structure of urban forests bring about a variety of challenges for urban land use mapping and the extraction of urban forests. Based on the DCNN algorithm, this study proposes a novel object-based U-net-DenseNet-coupled network (OUDN) method to realize urban land use mapping and the accurate extraction of urban forests. The proposed OUDN has three parts: the first part involves the coupling of the improved U-net and DenseNet architectures; then, the network is trained according to the labeled data sets, and the land use information in the study area is classified; the final part fuses the object boundary information obtained by object-based multiresolution segmentation into the classification layer, and a voting method is applied to optimize the classification results. The results show that (1) the classification results of the OUDN algorithm are better than those of U-net and DenseNet, and the average classification accuracy is 92.9%, an increase in approximately 3%; (2) for the U-net-DenseNet-coupled network (UDN) and OUDN, the urban forest extraction accuracies are higher than those of U-net and DenseNet, and the OUDN effectively alleviates the classification error caused by the fragmentation of urban distribution by combining object-based multiresolution segmentation features, making the overall accuracy (OA) of urban land use classification and the extraction accuracy of urban forests superior to those of the UDN algorithm; (3) based on the Spe-Texture (the spectral features combined with the texture features), the OA of the OUDN in the extraction of urban land use categories can reach 93.8%, thereby the algorithm achieved the accurate discrimination of different land use types, especially urban forests (99.7%). Therefore, this study provides a reference for feature setting for the mapping of urban land use information from VHSR imagery.

Keywords: urban forests; OUDN algorithm; deep learning; object-based; high spatial resolution remote sensing

1. Introduction

Urban land use mapping and the information extraction of urban forest resources are significant, yet challenging tasks in the field of remote sensing and have great value for urban environment monitoring, planning, and designing [1–3]. In addition, smart cities are now an irreversible trend in urban development in the world, and urban forests constitute "vital," "green," and indispensable infrastructure in cities. Therefore, the intelligent mapping of urban forest resources from remote sensing data is an essential component of smart city construction.

Over the past few decades, multispectral (such as the Thematic Mapper (TM)) [4–7], hyperspectral, and LiDAR [8–10] techniques have played important roles in the monitoring of urban forest resources. Currently, with the rapid development of modern remote sensing technologies, a very large amount of VHSR remotely sensed imagery (such as WorldView-3) is commercially available, creating new opportunities for the accurate extraction of urban forests at a very detailed level [11–13]. The application of VHSR images in urban forest resource monitoring has attracted increasing attention because of the rich and fine properties in these images. However, the ground objects of VHSR images are highly complex and confusing. For one thing, numerous land use types (such as Agricultural Land and grassland) have the same spectrum and texture characteristics [14], resulting in strong homogeneity in different categories [15], that is, the phenomenon of "same spectrum with different objects." For another, rich detailed information gives similar objects (such as building composed of different construction materials) strong heterogeneity in the spectral and structural properties [16], resulting in the phenomenon of "same object with different spectra". In addition, traditional statistical classification methods encounter these problems in the extraction of urban forests from VHSR remote sensing images. Additionally, urban forests with fragmented distributions are composed of scattered trees, street trees, and urban park forest vegetation. This also creates very large challenges for urban land use classification and the accurate mapping of urban forests [17].

Object-based classification first aggregates adjacent pixels with similar spectral and texture properties into complementary and overlapping objects through the image segmentation method to achieve image classification, and the processing units are converted from conventional pixels to image objects [18]. This classification method is based on homogeneous objects. In addition to applying the spectral information of images, this method fully exploits spatial features such as geometric shapes and texture details. The essence of object-based classification is to break through the limitations of traditional pixel-based classification and reduce the phenomena of the "same object with different spectra" and the "salt-and-pepper" phenomenon caused by distribution fragmentation. Therefore, object-based classification methods often yield better results than traditional pixel-based classification methods [19]. Recently, the combination of object-based and machine learning (ML) is widely used to detect features in the forest such as damage detection, landslide detection, and insect-infested forests [20–23]. In terms of ML, deep learning (DL) uses a large amount of data to train the model and can simulate and learn high-level features [24], making deep learning a new popular topic in the current research on the intelligent extraction of VHSR remote sensing information [15,25,26].

For DL, DCNNs and semantic segmentation algorithms are widely used in the classification of VHSR images, providing algorithmic support for accurate classification and facilitating great progress [27–38]. Among them, DCNNs are the core algorithms for the development of deep learning [39]. These networks learn abstract features through multiple layers of convolutions, conduct network training and learning, and finally, classify and predict images. DenseNet is a classic convolutional neural network framework [40]. This network can extract abstract features while combining the information features of all previous layers, so it has been widely applied in the classification of remote sensing images [41–44]. However, this network has some problems such as the limited extraction of abstract features. Semantic segmentation places higher requirements on the architectural design of convolutional networks, classifying each pixel in the image into a corresponding category, that is, achieving pixel-level classification. A typical representation of semantic segmentation is U-net [45], which combines upsampling with downsampling. U-net can not only

extract deeper features but also achieve accurate classification [46,47]. Therefore, U-net and DenseNet can be integrated to address the problem of the limited extraction of abstract features in DenseNet, and this combination may facilitate more accurate extraction from VHSR images.

In summary, object-based multiresolution segmentation offers obvious advantages in dealing with the problems of "same object with different spectra" and the "salt-and-pepper" phenomenon caused by distribution fragmentation [48–53], and deep learning is an important method for the intelligent mapping of VHSR remote sensing images. Consequently, this research proposes the novel classification method of the object-based U-net-DenseNet-coupled network (OUDN) to realize the intelligent and accurate extraction of urban land use and urban forest resources. This study takes subregion of the Yuhang District of Hangzhou City as the study area, with WorldView-3 images as the data source. First, the DenseNet and U-net network architectures are integrated; then, the network is trained according to the labeled data sets, and land use classification results are obtained based on the trained model. Finally, the object boundaries derived by object-based multiresolution segmentation are combined with the classification results of deep learning to optimize the classification results with the majority voting method.

2. Materials and Methods

2.1. Study Area

In this research, a subregion of the Yuhang District (YH) of Hangzhou, in Zhejiang Province in Southeast China, was chosen as the study area (Figure 1). WorldView-3 images of the study area were captured on 28 October 2018. The images contain four multispectral bands (red, green, blue, and near infrared (NIR)) with a spatial resolution of 2 m and a panchromatic band with a spatial resolution of 0.5 m. According to the USGS land cover classification system [54] and the FROM-GLC10 [55,56], the land use categories were divided into six classes, including Forest, Built-up, Agricultural Land, Grassland, Barren Land, and Water. As shown in Figure 1b, due to shadows in VHSR image, this study added a class of Others, including Shadow of trees and buildings. The detailed descriptions of each land use class and its corresponding subclasses are listed in Table 1.

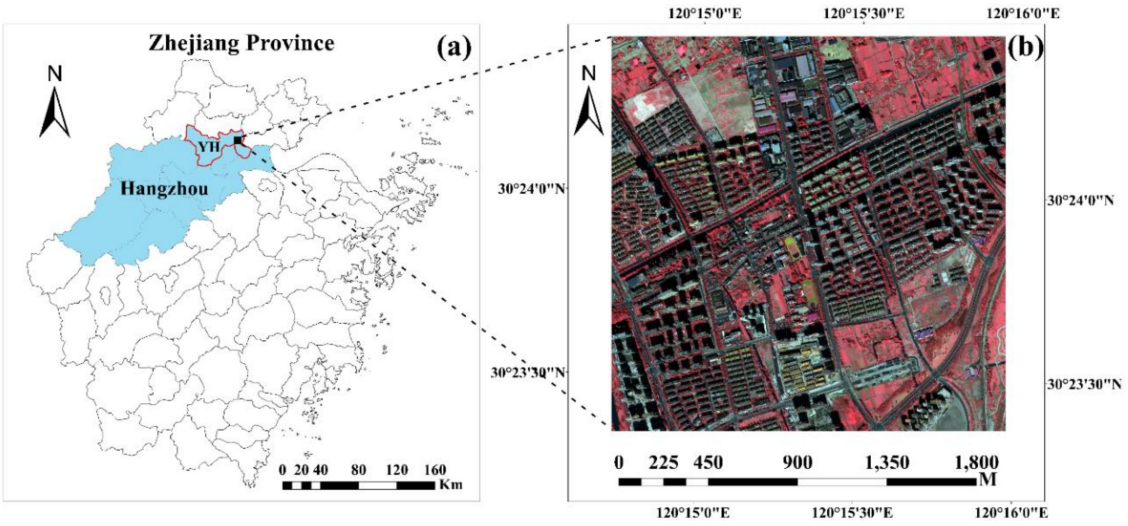

Figure 1. Location of the study area: (**a**) Zhejiang Province, and the blue polygons represent Hangzhou, (**b**) the subregion of the Yuhang District (YH) of Hangzhou.

Table 1. The urban land use classes and the corresponding subclass components.

Land Use Classes	Subclass Components
Forest	Deciduous Forest, Evergreen Forest
Build-up	Residential, Commercial and Services, Industrial, Transportation
Agricultural Land	Cropland, Nurseries, Other Agricultural Land
Grassland	Nature Grassland, Managed Grassland
Barren Land	Dry Salt Flats, Sandy Areas other than Beaches, Bare Exposed Rock
Water	Streams, River, Pond
Others	Shadow of trees and buildings

2.2. Data Processing

The image preprocessing including radiation correction and atmosphere correction was first performed using ENVI 5.3. Then, this label maps of the actual land use categories were made by eCognition software based on the results of the field survey combined with the method of visual interpretation. Due to the limitations on the size of the processed images from the GPU as well as to obtain more training images and to better extract image features, this study used the overlapping cropping method (Figure 2) to segment the images in the sample set into 4761 subimage blocks using 128 × 128 pixel windows for the minibatch training of the DL algorithms.

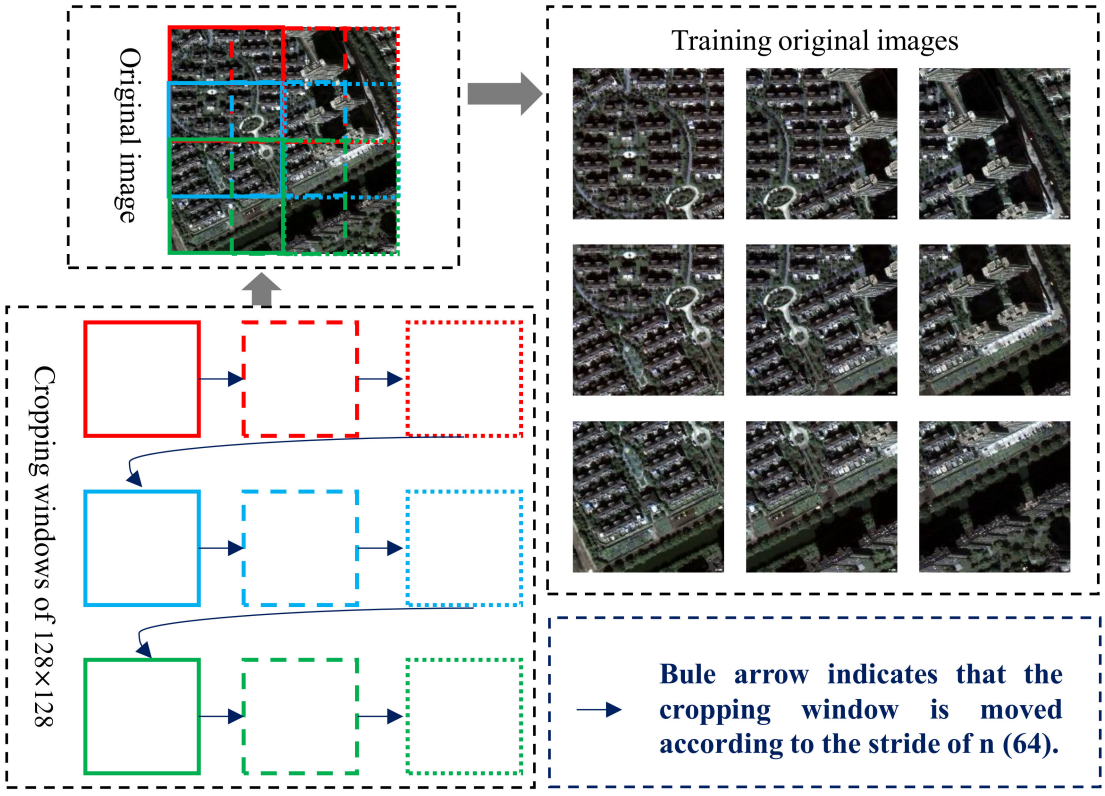

Figure 2. The overlapping cropping method for training the deep learning (DL) network. The size of the cropping windows is set to 128 × 128 pixels, where n is defined as half of 128.

2.3. Feature Setting

In this study, the classification features are divided into three groups: (1) the original R, G, B, and NIR bands, namely, the spectral features (Spe), the spectral features combined with the vegetation index features (Spe-Index), and the spectral features combined with the texture features (Spe-Texture). Based on these three groups of features, the performance of the OUDN algorithm

in the mapping of urban land use and urban forest information is evaluated. Descriptions of the spectral features, vegetation index, and texture are given in Table 2. The texture features based on the gray-level co-occurrence matrix (GLCM) [57] include mean, variance, entropy, angular second moment, homogeneity, contrast, dissimilarity, and correlation [58–60] with different calculation windows (3 × 3, 5 × 5, 7 × 7, 9 × 9, 11 × 11 and 13 × 13) [61].

Table 2. All the involved features are listed in detail in this paper, including original bands of WorldView-3 data, vegetation indices, and texture features based on the gray-level co-occurrence matrix (GLCM).

Feature Types	Feature Names	Details	Remarks
Original bands	Blue band (B) Green band (G) Red band (R) Near infrared band (NIR)	450–510 nm 510–580 nm 630–690 nm 770–1040 nm	WorldView-3 data
Vegetation indices	Difference vegetation index (DVI) Ratio vegetation index (RVI) Normalized difference vegetation index (NDVI) Optimized soil adjusted vegetation index (OSAVI) Soil adjusted vegetation index (SAVI) Triangular vegetation index (TVI)	NIR − R NIR/R (NIR − R)/(NIR + R) (NIR − R)/(NIR + R + X) (NIR − R) (1 + L)/(NIR + R + L) 0.5 [120 (NIR − G) − 200 (R − G)]	X take value for 0.16 L take value for 0.5 [62]
Texture features based on the gray-level co-occurrence matrix (GLCM)	Mean (ME) Variance (VA) Entropy (EN) Angular second moment (SE) Homogeneity (HO) Contrast (CON) Dissimilarity (DI) Correlation (COR)	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} iP(i,j)$ $\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-mean)^2P(i,j)$ $-\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}P(i,j)\log(P(i,j))$ $\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}P(i,j)^2$ $\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\frac{P(i,j)}{1+(i-j)^2}$ $\sum_{\lvert i-j\rvert=0}^{N-1}\lvert i-j\rvert^2\left\{\sum_{i=1}^{N}\sum_{j=1}^{N}P(i,j)\right\}$ $\sum_{\lvert i-j\rvert=0}^{N-1}\lvert i-j\rvert\left\{\sum_{i=1}^{N}\sum_{j=1}^{N}P(i,j)\right\}$ $\frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}P(i,j)^2-\mu_x\mu_y}{\sigma_x\sigma_y}$	$P(i,j)=V(i,j)/\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}V(i,j)$ $V(i,j)$ is the ith row of the jth column in the Nth moving window $u_x=\sum_{j=0}^{N-1}j\sum_{i=0}^{N-1}P(i,j)$ $u_y=\sum_{i=0}^{N-1}i\sum_{j=0}^{N-1}P(i,j)$ $\sigma_x=\sum_{j=0}^{N-1}(j-u_i)^2\sum_{i=0}^{N-1}P(i,j)$ $\sigma_y=\sum_{i=0}^{N-1}(i-u_j)^2\sum_{j=0}^{N-1}P(i,j)$

2.4. Methodology

The DenseNet architecture takes the output of all the previous layers as input and combines the previous information features to extract abstract features that are fairly limited. U-net performs deep feature extraction on the basis of the previous layer. Therefore, this study first improves the U-net and DenseNet networks and deeply couples them into the U-net-DenseNet-coupled network (UDN). Then, this network is combined with object-based multiresolution segmentation methods to construct the OUDN algorithm for intelligent and accurate extraction of urban land use and urban forest resources from VHSR images. The following introduces the DL algorithms in detail based on a brief introduction of DCNNs.

2.4.1. Brief Introduction of CNNs

Convolutional neural networks (CNNs) are the core algorithms of DL in the field of computer vision (CV) applications (such as image recognition) because of their ability to obtain hierarchically abstract representations with local operations [63]. This network structure was first inspired by biological vision mechanisms. There are four key ideas behind CNNs that take advantage of the properties of natural signals: local connections, shared weights, pooling, and the use of many layers [24], which are fully utilized.

As shown in Figure 3, the CNN structure consists of four basic processing layers: the convolution layer (Conv), nonlinear activation layer (such as ReLU), normalization layer (such as batch normalization (BN)), and pooling layer (Pooling) [63,64]. The first few layers are composed of two types of layers:

convolutional layers and pooling layers. The units in a convolutional layer are organized in feature maps, within which each unit is connected to local patches in the feature maps of the previous layer through a set of weights called a filter bank, and all units in a feature map share the same filter bank. Different feature maps in every layer use different filter banks, so different features can be learned. The result of this local weighted sum is then passed through a nonlinear activation function such as a ReLU, and the output results are pooled and nonlinearly processed through normalization (such as BN). In addition, nonlinear activation and nonlinear normalization are nonlinear blocks of processing that leads to a bigger boost in model training, so they play a significant role in CNN architecture. After multiple convolutions (combining a convolutional layer and a pooling layer is called a convolution), the results are flattened as the input of the fully connected layer, namely, the artificial neural network (ANN). Thus, the prediction result is finally obtained. Specifically, the major operations performed in the CNNs can be summarized by Equations (1)–(5):

$$S^{[l]} = pool_p\left(\varphi\left(S^{[l-1]} * W^{[l]} + b^{[l]}\right)\right) \tag{1}$$

$$\varphi(Z) = R = \begin{cases} Z; & if\ Z \geq 0 \\ 0; & Z < 0 \end{cases} \tag{2}$$

$$\mu = \frac{1}{m}\sum_{i=1}^{m} R^{(i)} \tag{3}$$

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}\left(R^{(i)} - \mu\right)^2 \tag{4}$$

$$R_{norm}^{(i)} = \frac{R^{(i)} - \mu}{\sqrt{\sigma^2 + \varepsilon}}, \tag{5}$$

where $S^{[l]}$ indicates the feature map at the lth layer [25], $S^{[l-1]}$ denotes the input feature map to the lth layer, and $W^{[l]}$ and $b^{[l]}$ represent the weights and biases of the layer, respectively, that convolve the input feature map through linear convolution $*$. These steps are often followed by a max-pooling operation with $p \times p$ window size ($pool_p$) to aggregate the statistics of the features within specific regions, which forms the output feature map $S^{[l]}$. The $\varphi(Z)$, R, indicates the nonlinearity function outside the convolution layer and corrects the convolution result of each layer, Z denotes the result of the convolution operation by calculating $S^{[l-1]} * W^{[l]} + b^{[l]}$, m represents the batch size (the number of samples required for a single training iteration), μ represents the mean, σ^2 represents the variance, ε is a constant set to keep the value stable to prevent $\sqrt{\sigma^2 + \varepsilon}$ from being 0, and $R_{norm}^{(i)}$ is the normalized value.

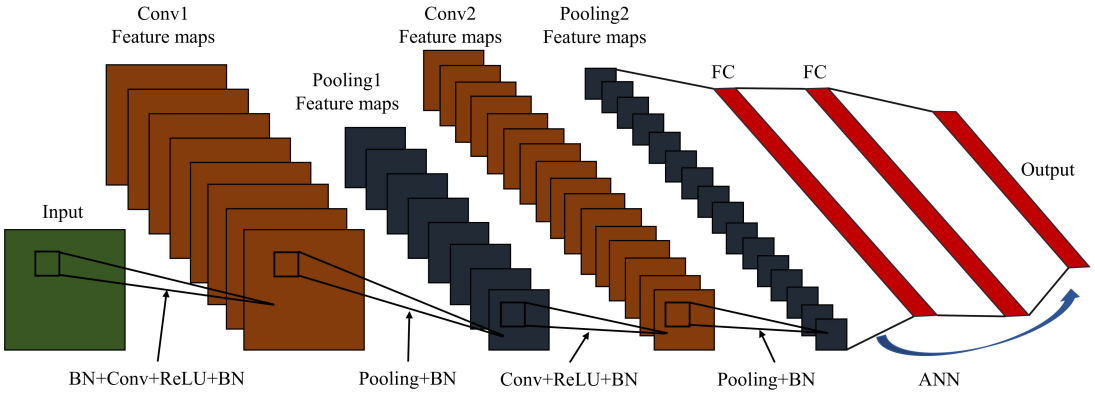

Figure 3. The classical structure of convolutional neural networks (CNNs). Batch normalization (BN) is a technique for accelerating network training by reducing the offset of internal covariates.

2.4.2. DL Algorithms

Improved DenseNet (D): DenseNet is based on ResNet [65], and its most important characteristic is that the feature maps of all previous networks are used as input for each layer of the network. Additionally, the feature maps are used as input by the following network layer, so the problem of gradient disappearance can be alleviated and the number of parameters can be reduced. The improved DenseNet network structure in this study is shown in Figure 4. Figure 4a is the complete structure, which adopts 3 Dense Blocks and 2 Translation layers. Before the first Dense Block, two convolutions are used. In this study, the bottle layer (1×1 convolution) in the Translation layer is converted to a 3×3 convolution operation, followed by an upsampling layer and finally the prediction result. The specific Dense Block structure is shown in Figure 4b and summarized by Equation (6):

$$X_{\updownarrow} = H_{\updownarrow}\left(\left[X_0\,,\ X_1\,,\ldots,\ X_{\updownarrow-1}\right]\right), \tag{6}$$

where $\left[X_0\,,\ X_1\,,\ldots,\ X_{\updownarrow-1}\right]$ denotes the feature maps with layers of X_0, X_1 … , $X_{\updownarrow-1}$ and $H_{\updownarrow}\left(\left[X_0\,,\ X_1\,,\ldots,\ X_{\updownarrow-1}\right]\right)$ indicates that the $\updownarrow$ layer takes all feature maps of the previous layers (X_0, X_1 … , $X_{\updownarrow-1}$) as input. In this study, all the convolution operations in the Dense Block use 3×3 convolution kernels, and the number of output feature maps (K) in each layer is set to 32.

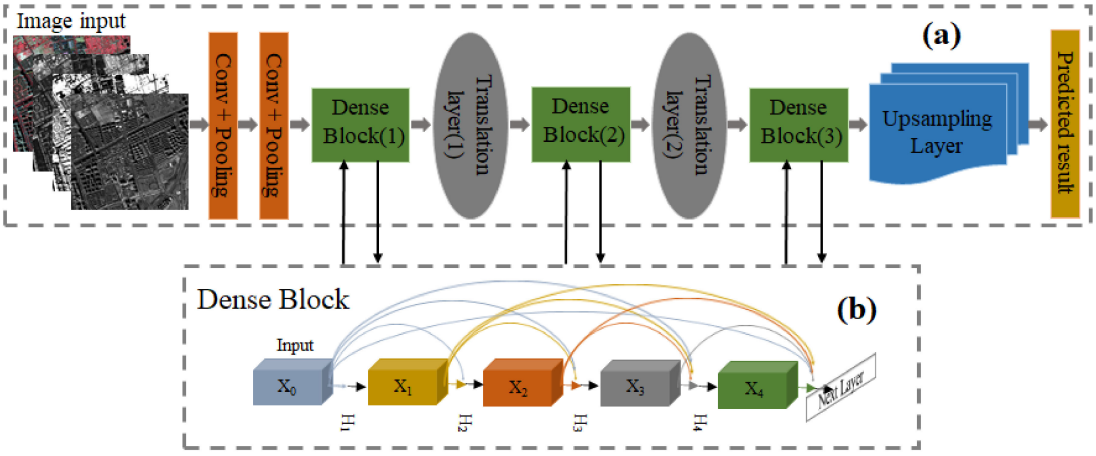

Figure 4. The improved DenseNet structure composed of three dense blocks: (**a**) the complete structure and (**b**) the Dense Block composed of five feature map layers.

Improved U-net (U): U-net is an improved fully convolutional network (FCN) [66]. This network has attracted extensive attention because of its clear structure and excellent performance on small data sets. U-net is divided into a contracting path (to effectively capture contextual information) and an expansive path (to achieve a more precise position for the pixel boundary). Considering the characteristics of urban land use categories and the rich details of WorldView-3 images, the improved structure in this study mainly increases the number of network layers to 11 layers, and each layer increases the convolution operations, thereby obtaining increasingly abstract features. The network is constructed around convolution filters to obtain images with different resolutions, so the structural features of the image can be detected on different scales. More importantly, BN is performed before the convolutional layer and pooling layer, and the details are shown in Figure 5.

(1) The left half of the bottom layer is the contracting path. With the input of a 128×128 image, each layer uses three 3×3 convolution operations. After each convolution, followed by the ReLU activation function, max-pooling with a step of 2 is applied for downsampling. In each downsampling stage, the number of feature channels is doubled. Five downsamplings are applied, followed by two 3×3 convolutions in the bottom layer of the network architecture.

The size of the feature maps is eventually reduced to 4 × 4 pixels, and the number of feature map channels is 1024.

(2) The right half of the network, that is, the expansive path, mainly restores the feature information of the original image. First, a deconvolution kernel with a size of 2 × 2 is used to perform upsampling. In this process, the number of the feature map channels is halved, while the feature maps of the symmetrical position generated by the downsampling and the upsampling are merged; then, three 3 × 3 convolution operations are performed on the merged features, and the above operations are repeated until the image is restored to the size of input image; ultimately, four 3 × 3 and one 1 × 1 convolution operations and a Softmax activation function are used to complete the category prediction of each pixel in the image. The Softmax activation function is defined as Equations (7):

$$p_k(X) = \frac{\exp(a_k(X))}{\left(\sum_{k'=1}^{K} \exp(a_{k'}(X))\right)}, \tag{7}$$

where $a_k(X)$ represents the activation value of the kth channel at the position of pixel X. K indicates the number of categories, and $p_k(X)$ denotes the function with the approximate maximum probability. If $a_k(X)$ is the largest activation value in the kth channel, $p_k(X)$ is approximately equal to 1; in contrast, $p_k(X)$ is approximately equal to zero for other k values.

UDN: The detailed coupling process of the improved U-net and DenseNet is shown in Figure 6. (a) The first two layers use the same convolutional layer and pooling layer to obtain abstract feature maps; (b) then, the feature maps obtained by the above operations are input into the Combining Block structure to realize the coupling of the convolution results from the two structures. After two convolution operations are performed on the coupling result, max-pooling is used to perform downsampling, followed by two Combining Block operations; (c) after the downsampling, two convolutions are performed on the coupling result to obtain 1024 feature maps of 4 × 4; (d) the smallest feature maps (4 × 4 × 1024) are restored to the size of the original image after 5 upsamplings; (e) finally, the classification result is output based on the front feature maps through the 1 × 1 convolution operations and the Softmax function.

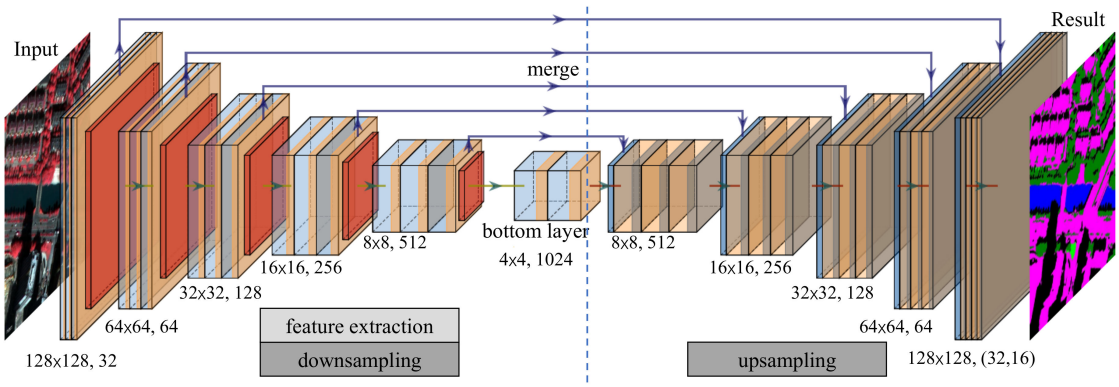

Figure 5. The improved U-net structure is composed of eleven convolution layers.

OUDN: The boundary information of the categories is the basis of the accurate classification of VHSR images. In this study, the OUDN algorithm combines the category objects obtained by object-based multiresolution segmentation [18] with the classification results of the UDN algorithm to constrain and optimize the classification results. Four multispectral bands (red, green, blue, and near infrared) together with vegetation indices and texture features, useful for differentiating urban land use objects with complex information, are incorporated as multiple input data sources for the image segmentation using eCognition software. Then, all the image objects are transformed into GIS vector polygons with distinctive geometric shapes, which are combined with the classification results

of the UDN algorithm. Based on the Spatial Analysis Tools of ArcGIS, the category with the largest statistics is taken as the category of the object by counting the number of pixels in each object and using the majority voting method. Thereby, final classification results of the OUDN algorithm are obtained. The segmentation scale directly affects the boundary accuracy of the categories. Therefore, according to the selection method of the optimal segmentation scale [67], this study gains the segmentation results by setting different segmentation scales, and determines the final segmentation scale 50.

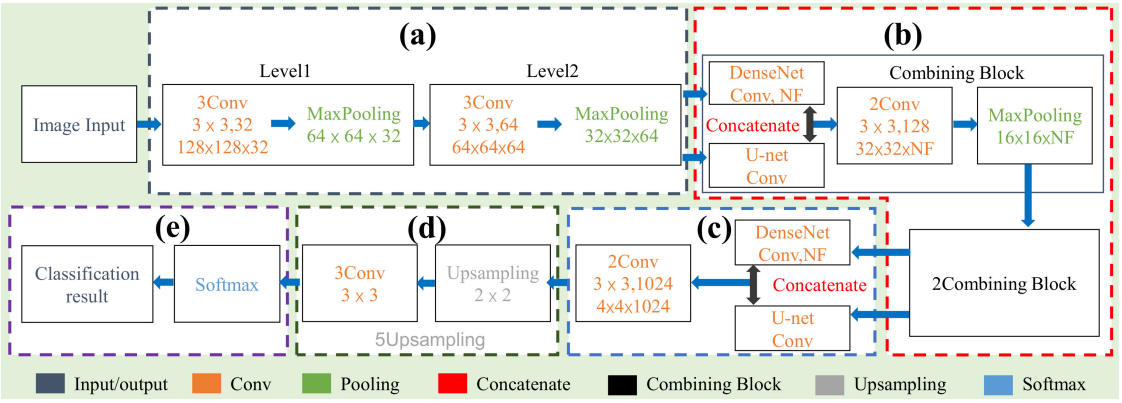

Figure 6. The network structure of the UDN algorithm, where NF represents the number of convolutional filters. (**a**) The first two layers (Level1 and Level2) including convolutional layers and pooling layers; (**b**) the coupling of U-net and DenseNet algorithms; (**c**) the bottom layer of the network; (**d**) Upsampling layers; (**e**) predicted classification result.

Finally, the template for training the minibatch neural network based on the above algorithms in this research is shown in Algorithm 1 [68]. The network uses the loss function of categorical cross entropy and the adaptive optimization algorithm of Adam. Additionally, the number of iterations is set to 50, and the learning rate (lr) is set to 0.0001. In each iteration, b images are sampled to compute the gradients, and then the network parameters are updated. The training of the network stops after K passes through the data set.

Algorithm 1 Train a neural network with the minibatch Adam optimization algorithm.

initialize (*net*)
for *epoch* = 1, . . . , *K* **do**
  for *batch* = 1, . . . , # *images*/*b* do
    images ← uniformly sample *batch* − *size* images
    X, *y* ← preprocess(images)
    z ← forward (net, *X*)
    l ← loss (*z*, *y*)
    lr, *grad* ← background (*l*)
    update (*net*, *lr*, *grad*)
  end for
end for

2.5. Experiment Design

The flowchart of steps is shown in Figure 7. The WorldView-3 image with 15.872 × 15.872 pixels is first preprocessed by image fusion, radiometric calibration, and atmospheric correction (Figure 7a). According to this preprocessed image, a 3968 × 3968 pixel subimage with various categories is cropped for model prediction, and other representative subimages are cropped as the sample set including training set and validation set for model training; then, labeled maps are made based on the sample set, followed by image cropping (Figure 7b); the cropped original images and the corresponding labeled

maps are used to train the DL models (Figure 7c); the image with 3968 × 3968 pixels is classified by the trained model (Figure 7d); finally, objects of multiresolution segmentation are applied to optimize the classification results of the UDN algorithm to obtain the classification results of the OUDN, followed by detailed comparisons of the results from all algorithms including U, D, UDN, and OUDN (Figure 7e).

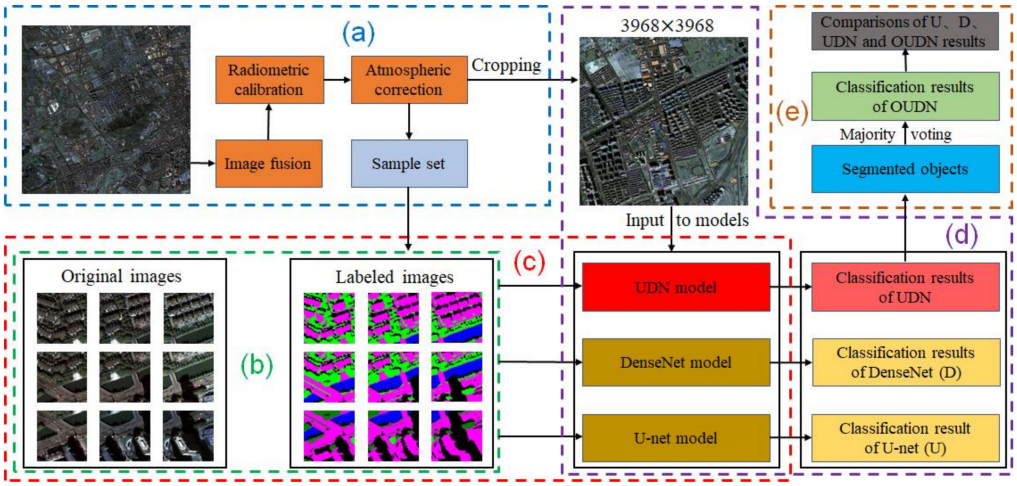

Figure 7. A flowchart of the experimental method in this paper, including five major steps: (**a**) image preprocessing; (**b**) image labeling and image cropping; (**c**) model training; (**d**) model prediction; (**e**) object-based optimization of the UDN results and comparisons of the results from all algorithms.

3. Results and Analysis

The tests of the proposed OUDN algorithm were presented in this section, and the classification results are compared with those of UDN, improved U-net (U), and improved DenseNet (D). To evaluate the proposed algorithm, the classification results in this study were assessed with the overall accuracy (OA), kappa coefficient (Kappa), producer accuracy (PA), and user accuracy (UA) [69]. The detailed results and analysis of the model training and classification results are clarified as follows.

3.1. Training Results of U, D and UDN Algorithms

There were a total of 4761 image blocks with 128 × 128 pixels in the sample set. Additionally, 3984 of these blocks were selected for the training, and the remaining blocks were used for the validation. Then, the cropped original image blocks and the corresponding labeled maps were used to train the minibatch network model according to the template of Algorithm 1. Based on the three feature groups of Spe, Spe-Index, and Spe-Texture, the overall model accuracies including training accuracy (TA) and validation accuracy (VA) of the U, D, and UDN algorithms were demonstrated in Table 3. In all feature combinations, the UDN algorithm obtained the highest training accuracies (98.1%, 98%, and 98.4%). However, for the U and U algorithms, the training accuracies of the Spe-Texture were the lowest (96.3% and 96%) compared with those of the Spe and Spe-Index. The UDN algorithm achieved the highest model accuracies (TA of 98.4% and VA of 93.8%, respectively) based on Spe-Texture.

Table 3. The overall training and validation accuracies of the improved U-net (U), improved DenseNet (D), and U-net-DenseNet-coupled network (UDN) algorithms based on the three feature groups of Spe, Spe-Index, and Spe-Texture. The algorithms in this table did not include object-based U-net-DenseNet-coupled network (OUDN), since the OUDN algorithm was based on UDN algorithm to optimize classification results.

Feature	TA			VA		
	U	D	UDN	U	D	UDN
Spe	0.975	0.971	0.981	0.914	0.923	0.936
Spe-Index	0.969	0.977	0.980	0.935	0.916	0.920
Spe-Texture	0.963	0.960	0.984	0.927	0.929	0.938

3.2. *Classification Results*

3.2.1. Classification Results Based on Four Algorithms

The classification accuracies of the U, D, UDN, and OUDN algorithms on the three feature groups of Spe, Spe-Index, and Spe-Texture are demonstrated in Tables 4–6, respectively. In general, among the three feature combinations, the U and D algorithms yielded the lowest OA and Kappa, followed by UDN; in contrast, the OUDN algorithm achieved the highest OA (92.3%, 92.6%, and 93.8%) and Kappa (0.910, 0.914, and 0.928). The average accuracy of the OUDN algorithm was much higher (approximately 3%) than those of the U and D algorithms. As shown in Table 4, the UDN algorithm obtained better accuracies for Agricultural Land and Grassland than the U and D algorithms. For example, the PA values of Agricultural Land were 89%, 88%, and 90.3% for the U, D, and UDN algorithms, respectively, and the PA values of Grassland were 64.3%, 73%, and 74%, respectively. Compared with those of the UDN algorithm, the OUDN algorithm obtained better PA values for Agricultural Land, Grassland, Barren Land, and Water. Table 5 shows that the PA of Agricultural Land of the UDN algorithm was 5% and 3.3% higher than those of the U and D algorithms, respectively. In addition, the OUDN algorithm mainly yielded improvements in the PA values of Forest, Built-up, Agricultural Land, Grassland, and Barren Land. As shown in Table 6, the UDN algorithm yielded higher accuracies for Forest, Built-up, Grassland, Barren Land, and Water than the U and D algorithms, and in particular, the PA value of Grassland was significantly higher, by 15.6% and 17.3%, respectively. Meanwhile, the OUDN algorithm yielded the accuracies superior to those of the UDN algorithm in some categories. In summary, the OUDN algorithm obtained high extraction accuracies for urban land use types, and coupling object-based segmentation effectively addressed the fragmentation problem of classification with high-resolution images, thereby improving the image classification accuracy. Therefore, the OUDN algorithm offered great advantages for urban land-cover classification.

Table 4. The classification accuracies of the U, D, UDN, and OUDN algorithms based on the Spe, including the accuracies (user accuracy (UA) and producer accuracy (PA)) of every class, overall accuracy (OA) and kappa coefficient (Kappa).

Algorithms	OA	Kappa		Forest	Build-UP	Agricultural Land	Grassland	Barren Land	Water	Others
U	0.903	0.887	UA	0.834	0.892	0.756	0.951	0.963	0.997	0.990
			PA	0.990	0.963	0.890	0.643	0.873	0.987	0.973
D	0.905	0.889	UA	0.863	0.855	0.781	0.920	0.975	0.997	0.993
			PA	0.990	0.980	0.880	0.730	0.787	0.990	0.983
UDN	0.920	0.907	UA	0.908	0.910	0.755	0.961	0.973	1.000	0.987
			PA	0.990	0.980	0.903	0.740	0.853	0.987	0.993
OUDN	0.923	0.910	UA	0.911	0.909	0.767	0.958	0.974	0.993	0.993
			PA	0.990	0.970	0.910	0.753	0.857	0.993	0.990

Table 5. The classification accuracies of the U, D, UDN, and OUDN algorithms based on the Spe-Index, including the accuracies (UA and PA) of every class, OA and Kappa.

Algorithms	OA	Kappa		Forest	Build-Up	Agricultural Land	Grassland	Barren Land	Water	Others
U	0.913	0.899	UA	0.878	0.872	0.809	0.930	0.977	1.000	0.958
			PA	0.983	0.977	0.873	0.757	0.867	0.950	0.987
D	0.917	0.903	UA	0.870	0.857	0.842	0.927	0.977	0.997	0.980
			PA	0.983	0.980	0.890	0.760	0.850	0.973	0.983
UDN	0.923	0.910	UA	0.891	0.892	0.817	0.957	0.978	0.997	0.958
			PA	0.983	0.963	0.923	0.750	0.883	0.960	0.997
OUDN	0.926	0.914	UA	0.892	0.901	0.822	0.974	0.982	0.997	0.955
			PA	0.987	0.973	0.937	0.753	0.887	0.957	0.990

Table 6. The classification accuracies of the U, D, UDN, and OUDN algorithms based on the Spe-Texture, including the accuracies (UA and PA) of every class, OA and Kappa.

Algorithms	OA	Kappa		Forest	Build-Up	Agricultural Land	Grassland	Barren Land	Water	Others
U	0.898	0.881	UA	0.864	0.840	0.787	0.914	0.955	1.000	0.961
			PA	0.993	0.963	0.860	0.677	0.857	0.947	0.987
D	0.897	0.879	UA	0.897	0.824	0.750	0.943	0.980	0.993	0.971
			PA	0.987	0.970	0.930	0.660	0.797	0.943	0.990
UDN	0.932	0.921	UA	0.857	0.873	0.913	0.954	0.985	1.000	0.970
			PA	0.997	0.983	0.877	0.833	0.873	0.977	0.983
OUDN	0.938	0.928	UA	0.877	0.866	0.932	0.970	0.985	1.000	0.967
			PA	0.997	0.987	0.913	0.853	0.857	0.973	0.987

The classification maps of the four algorithms based on the Spe, Spe-Index, and Spe-Texture are presented in Figures 8–10, respectively, with the correct or incorrect classification results marked in black or red circles, respectively. In general, the classification results of the UDN and OUDN algorithms were better than those of the other methods, and there was no obvious "salt-and-pepper" effect in the classification results of the four algorithms. However, due to the splicing in the U, D, and UDN algorithms, the ground object boundary exhibited discontinuities, whereas the proposed OUDN algorithm addressed this problem to a certain extent.

Classification maps of different algorithms based on Spe: Based on the Spe, the proposed method in this paper better identified the ground classes that are difficult to distinguish, including Built-up, Barren Land, Agricultural Land, and Grassland. However, the recognition effect of the U and D algorithms was undesirable. As shown in Figure 8, the U and D algorithms confused Built-up and Barren Land (red circle (1)), while the UDN and OUDN algorithms correctly distinguished them (black circle (1)); for the U algorithm, Built-up was misclassified as Barren Land (red circle (2)), while the other algorithms accurately identified these classes (black circle (2)); the D algorithm did not identify Barren Land (red circle (3)), in contrast, the recognition effect of the other methods was favorable (black circle (3)); for the U and D algorithms, Grassland was misclassified as Agricultural Land (red circle (4)), while other algorithms precisely distinguished them (black circle (4)); the four algorithms mistakenly classified some Agricultural Land as Grassland and confused them (red circle (5)).

Classification maps of different algorithms based on Spe-Index: Based on the Spe-Index, the proposed method in this paper better recognized Built-up, Barren Land, Agricultural Land, and Grassland. However, the recognition effect of the U and D algorithms was poor. As demonstrated by Figure 9, U and D algorithms confused Built-up and Barren Land (red circle (1)), whereas the UDN and OUDN algorithms correctly distinguished them (black circle (1)); the U algorithm incorrectly identified Barren Land (red circle (2)), while the classification results of other algorithms were superior (black circle (2)); the U and D algorithms mistakenly classified Barren Land as Agricultural Land (red circle (3)), in contrast, the UDN and OUDN better identified them (black circle (3)); for all four algorithms, some Agricultural Land was misclassified as Grassland (red circle (4)).

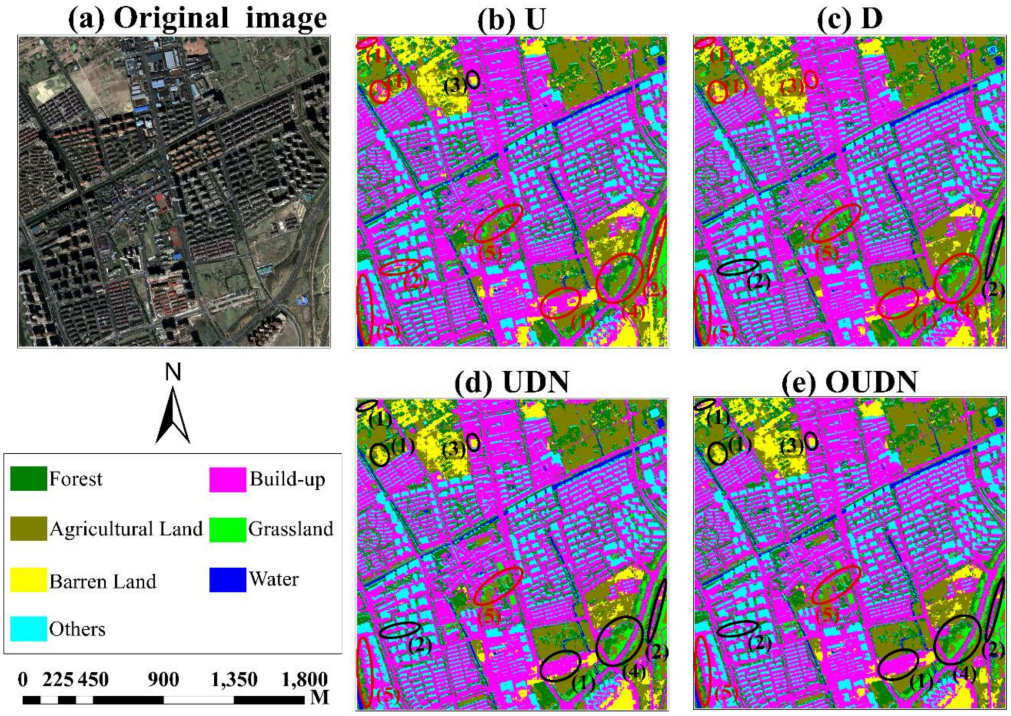

Figure 8. (**a**) Original image; (**b**) the classification map of the U algorithm based on the Spe; (**c**) the classification map of the D algorithm based on the Spe; (**d**) the classification map of the UDN algorithm based on the Spe; (**e**) the classification map of the OUDN algorithm based on the Spe; and the red and black circles denote incorrect and correct classifications, respectively.

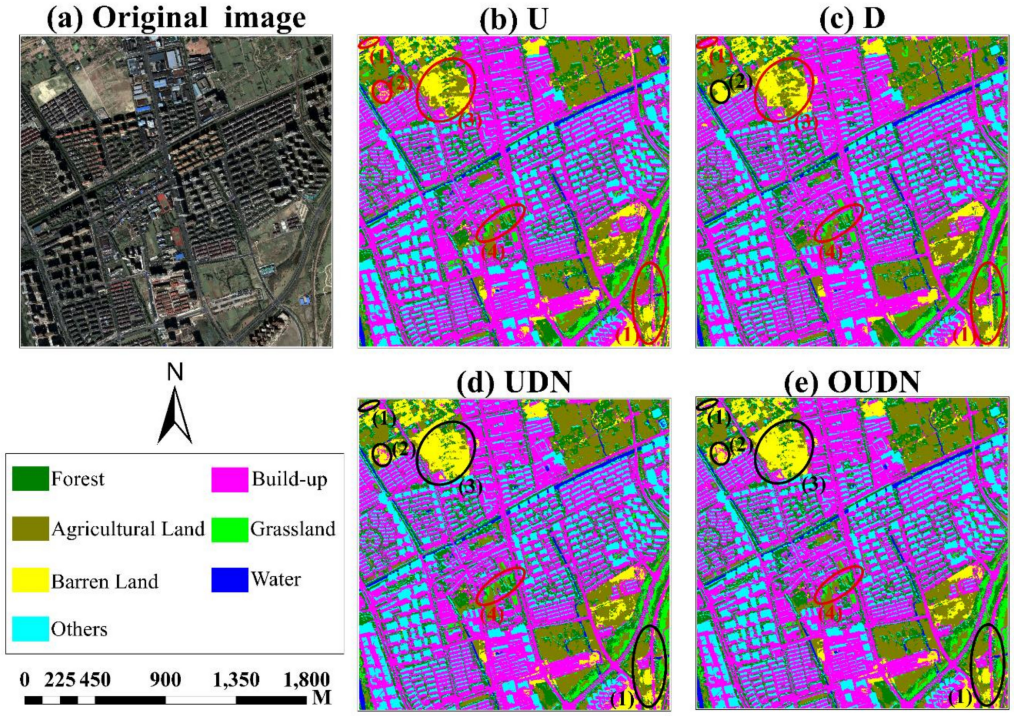

Figure 9. (**a**) Original image; (**b**) the classification map of the U algorithm based on the Spe-Index; (**c**) the classification map of the D algorithm based on the Spe-Index; (**d**) the classification map of the UDN algorithm based on the Spe-Index; (**e**) the classification map of the OUDN algorithm based on the Spe-Index; and the red and black circles denote incorrect and correct classifications, respectively.

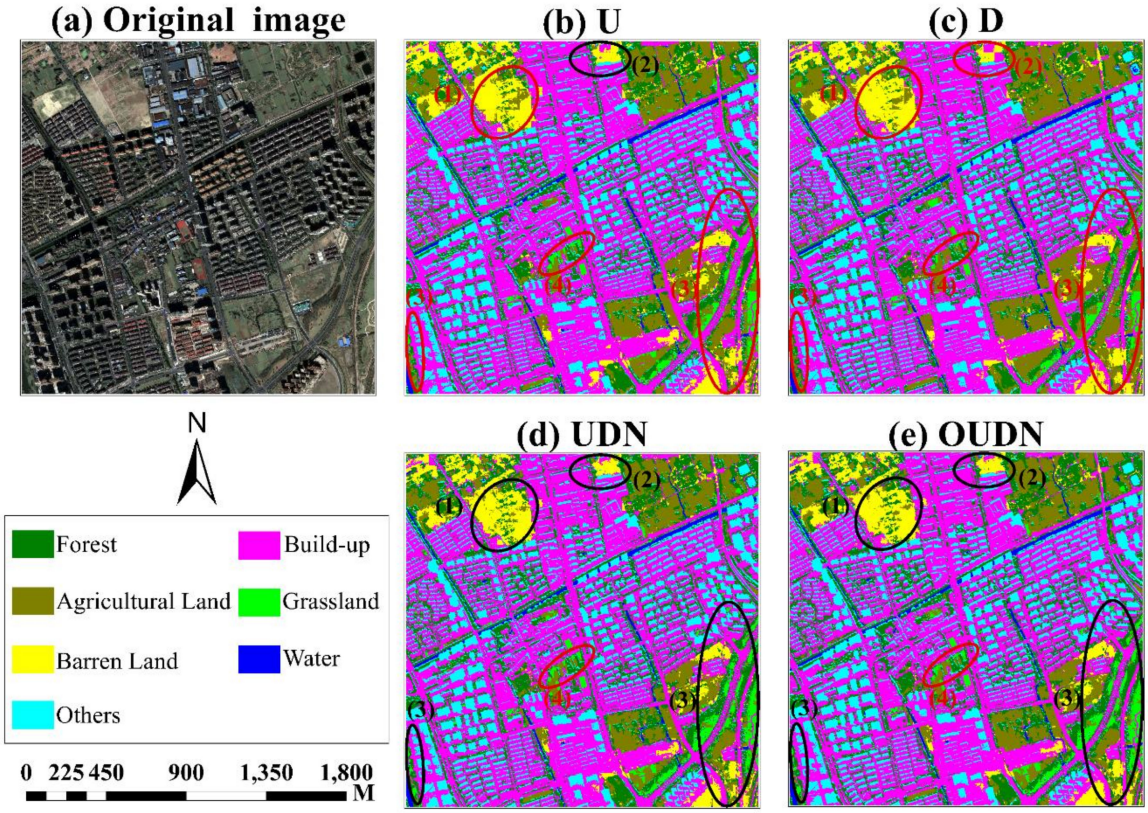

Figure 10. (**a**) Original image; (**b**) the classification map of the U algorithm based on the Spe-Texture; (**c**) the classification map of the D algorithm based on the Spe-Texture; (**d**) the classification map of the UDN algorithm based on the Spe-Texture; (**e**) the classification map of the OUDN algorithm based on the Spe-Texture; and the red and black circles denote incorrect and correct classifications, respectively.

Classification maps of the different algorithms based on Spe-Texture: Based on the Spe-Texture, the proposed method in this paper better identified each category, especially Grassland, yielding the best recognition result; nevertheless, the recognition effect of the U and D algorithms was worse. As shown in Figure 10, the U and D algorithms incorrectly classified much Barren Land as Agricultural Land (red circle (1)), whereas the UDN and OUDN algorithms identified these types better (black circle (1)); the D algorithm confused Built-up and Barren Land (red circle (2)), while the other algorithms better distinguished them (black circle (2)); the extraction effects for Grassland of the UND and OUDN algorithms (black circle (3)) were better than those of the U and D algorithms (red circle (3)); all the algorithms mistakenly classified some Agricultural Land as Grassland (red circle (4)).

3.2.2. Extraction Results of Urban Forests

This section focuses on the analysis of urban forest extraction based on the Spe, Spe-Index, and Spe-Texture with the four algorithms. As shown in Tables 4–6, the PA values of the urban forest information extraction for all algorithms were above 98%, which indicated that the DL algorithms used in this study offered obvious advantages in the extraction of urban forests. Additionally, for the OUDN algorithm, the average PA (99.1%) and UA (89.3%) of urban forest extraction were better than those of the other algorithms based on the three groups of features. This demonstrated that the OUDN algorithm exhibited fewer errors from urban forest leakage and misclassification errors between urban forests and other land use types.

The classification results for urban forests, including scattered trees and street trees, of the different algorithms based on the Spe-Texture are presented in Figure 11. In this study, two representative subregions (subset (1) and subset (2)) were selected for the analysis of the results of the different

algorithms, with the correct or incorrect classification results marked in black or blue circles, respectively. In general, the urban forest extraction effect of the OUDN algorithm was the best. According to the classification results of subset (1), the U and D algorithms mistakenly identified some street trees (blue circles), while UDN and OUDN better extracted these trees (black circles). As shown in the results of subset (2), the extraction results for some scattered trees of the U and D algorithms were not acceptable (blue circles); nevertheless, UDN and OUDN accurately distinguished them (black circles). Additionally, the U and D algorithms misclassified some Forest as Grassland and Built-up (blue circles), whereas UDN and OUDN correctly identified the urban forests (black circles).

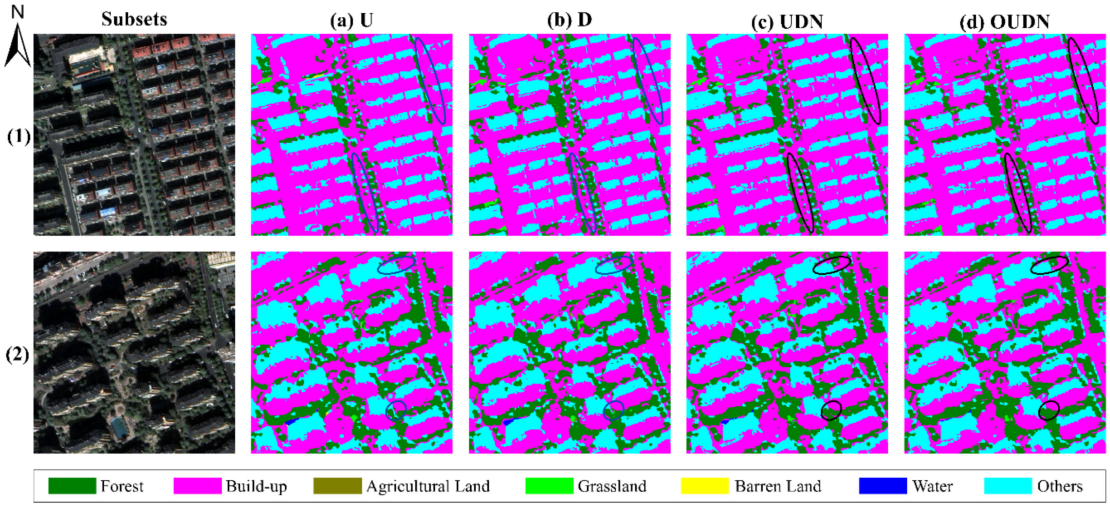

Figure 11. There are two subsets (Subset (**1**) and Subset (**2**)) dominated by urban forests. (**a**) the classification maps of the U algorithm in the subsets; (**b**) the classification maps of the D algorithm in the subsets; (**c**) the classification maps of the UDN algorithm in the subsets; and (**d**) the classification maps of the OUDN algorithm in the subsets.

3.2.3. Result Analysis

According to the classification results of the four algorithms on the Spe, Spe-Index, and Spe-Texture, a confusion matrix is constructed, which is shown in Figure 12. In general, regardless of the feature combinations, the classification accuracies of each algorithm for Forest, Built-up, Water, and Others were relatively high, and the recognition accuracy was above 95%. In particular, the classification accuracy of the Forest was above 98%, whereas the classification accuracies of the other categories varied greatly. As demonstrated by Figure 12, (1) based on the Spe, the extraction accuracies of the OUDN algorithm for Agricultural Land and Grassland were significantly superior to those of the U and D algorithms, whereas the U and D algorithms misclassified Agricultural Land and Grassland as Forest at a higher rate. Compared with that of the U algorithm, the OUDN algorithm yielded better Grassland classification accuracy (75%, an increase in 11%) while optimizing the extraction accuracy of UDN (74%). The D algorithm misclassified 15% of the Barren Land as Built-up, whereas only 8% was incorrectly predicted by the UDN and OUDN algorithms. Therefore, the OUDN algorithm offered obvious advantages in urban land-cover classification. (2) For the Spe-Index, compared with those of the U and D algorithms (87% and 89%, respectively), the OUDN algorithm yielded higher extraction accuracies of Agricultural Land (94%) and optimized the classification accuracy of UDN (92%). The U and D algorithms misclassified 12% of the Barren Land as Built-up, whereas only 11% and 10% were incorrectly predicted by the UDN and OUDN algorithms, so the OUDN algorithm captured the best classification effect. (3) For the Spe-Texture, the extraction accuracies of the UDN and OUDN algorithms for Grassland were very high (83% and 85%, respectively), and the accuracies were the highest among all the Grassland classification results. Compared with the classification accuracies of the U and D algorithms (68% and 66%), the accuracies of UDN and OUDN were 15–19%

higher. Figure 12 showed that the U and D algorithms misclassified 21% and 25% of the Grassland as Agricultural Land, respectively, whereas the misclassification rates of UDN and OUDN were fairly low (7% and 6%, respectively).

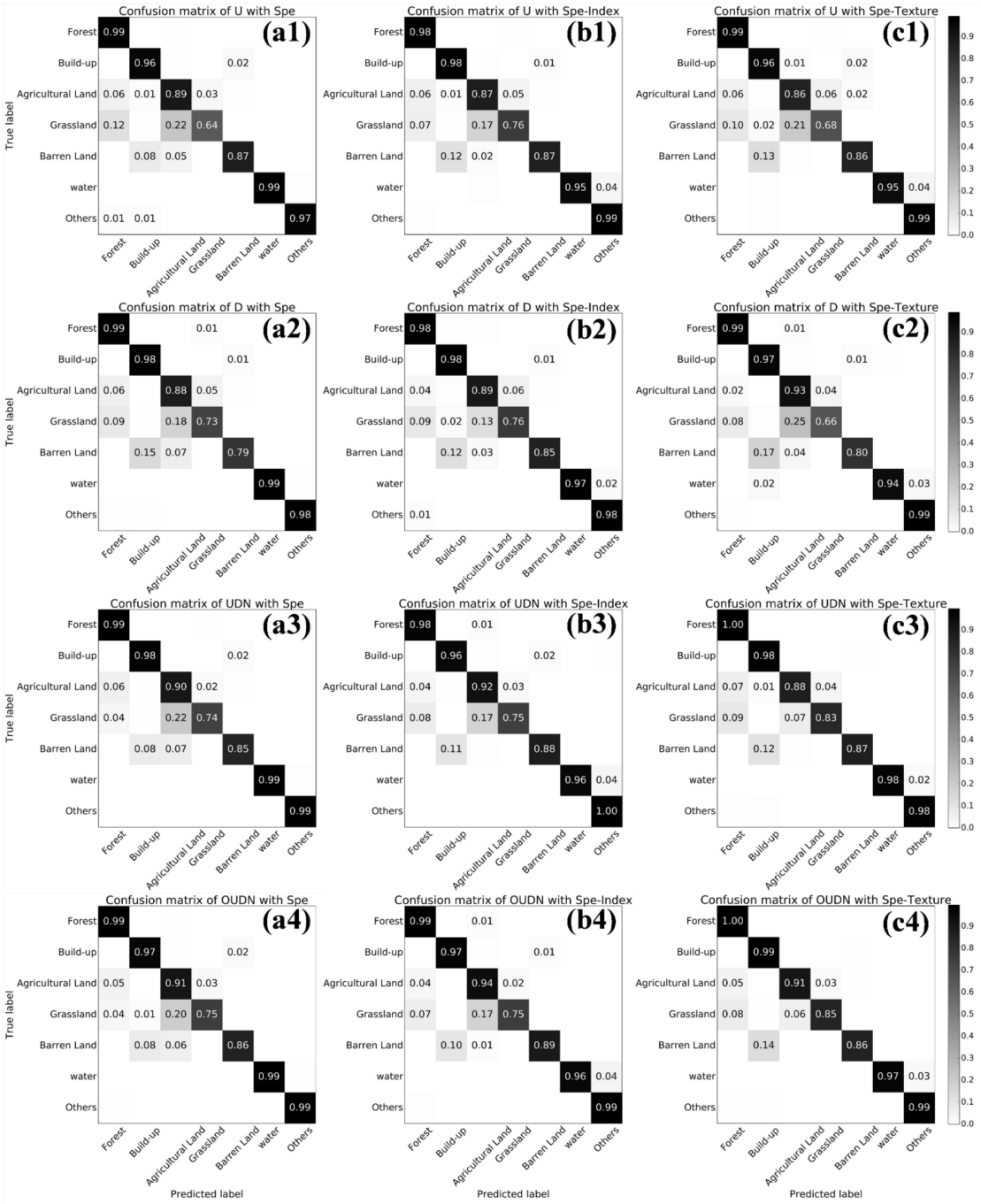

Figure 12. (**a1**) Confusion matrix of U algorithm based on Spe; (**b1**) Confusion matrix of U algorithm based on Spe-Index; (**c1**) Confusion matrix of U algorithm based on Spe-Texture; (**a2**) Confusion matrix of D algorithm based on Spe; (**b2**) Confusion matrix of D algorithm based on Spe-Index; (**c2**) Confusion matrix of D algorithm based on Spe-Texture; (**a3**) Confusion matrix of UDN algorithm based on Spe; (**b3**) Confusion matrix of UDN algorithm based on Spe-Index; (**c3**) Confusion matrix of UDN algorithm based on Spe-Texture; (**a4**) Confusion matrix of OUDN algorithm based on Spe; (**b4**) Confusion matrix of OUDN algorithm based on Spe-Index; (**c4**) Confusion matrix of OUDN algorithm based on Spe-Texture.

For urban forests, as demonstrated by Figure 12, (1) based on the Spe, the extraction accuracy of urban forests was 99% for each algorithm, however, these algorithms misclassified Agricultural Land and Grassland as Forest generally. Compared with those of the U and D algorithms, OUDN's rate of misclassification of Agricultural Land and Grassland as Forest was the lowest (5% and 4%); (2) based on the Spe-Index, the OUDN algorithm obtained the highest urban forest extraction accuracy (99%) and the lowest rate of Agricultural Land and Grassland misclassified as Forest (4% and 7%); (3) based on the Spe-Texture, the urban forest extraction accuracy of the OUDN algorithm was the highest (approximately 100%).

Through the above analysis, it was concluded that (1) the classification results of the OUDN algorithm were significantly better than those of the other algorithms for confusing ground categories (such as Agricultural Land, Grassland, and Barren Land); (2) the accuracy of the UDN algorithm was improved through object constraints; (3) especially for Spe-Texture, the OUDN algorithm achieved the highest OA (93.8%), which was 4% and 4.1% higher than those of the U and D algorithms, respectively; (4) the UDN and OUDN algorithms had obvious advantages regarding the accurate extraction of urban forests, and they not only accurately extracted the street trees but also identified the scattered trees ignored by the U and D algorithms.

4. Discussion

The UDN and OUDN algorithms constructed by this study achieved higher accuracies in the extraction of urban land use information from VHSR imagery than the U and D algorithms. The UDN algorithm applied the coupling of the improved 11-layer U-net network and the improved DenseNet to train the network and realize prediction using the learned deep level features. With the advantages of both networks, the accurate extraction of urban land use and urban forests was ensured. Meanwhile, the UDN algorithm addressed the problems of common misclassifications and omissions in the classification process (Tables 4–6) and dealt with the confusion of Agricultural Land, Grassland, Barren Land and Built-up (Figure 12), thereby improving the classification accuracies of urban land use and urban forest. In all feature combinations, especially for the Spe-Texture, the classification accuracies of the UDN algorithm were 3.4% and 3.5% higher than those of the U and D algorithms, respectively. This study chose 50 as the optimal segmentation scale, and the phenomenon of misclassification with UDN was corrected by the constraints of the segmentation objects (Tables 4–6). The OUDN algorithm not only alleviated the distribution fragmentation of ground objects and the common "salt-and-pepper" phenomenon in the classification process but also dealt with the problem of discontinuous boundaries during the splicing process of the classification results of segmented image blocks (Figures 8–10). Compared with previous studies about classifications using U-net and DenseNet [41,46], this study fully combined the advantages of U-net and DenseNet network and achieved more high classification accuracies. Compared with previous object-based DL classification methods [49], in this study, object-based multiresolution segmentations were used to constrain and optimize the UND classification results and not used to participate in the UND classification. It is necessary to further study in this respect.

The overall classification accuracies (OA) of different features based on different algorithms are shown in Figure 13. (1) In terms of the UDN and OUDN algorithms, accuracies of the Spe-Texture were the highest (93.2% and 93.8%), followed by those of the Spe-Index (92.3% and 92.6%). As demonstrated by Figure 12, for Grassland, Built-up and Water, the classification accuracies of the Spe-Texture were significantly higher than those of the Spe-Index. For example, the Grassland accuracies of the Spe-Texture were 8% and 10% higher than those of the Spe-Index, and the Built-up accuracies of the Spe-Texture were 2% and 2% higher than those of the Spe-Index. It can be concluded from Table 3 that the TA and VA of Spe-Texture are higher. Thus, the classification results of the Spe-Texture were better than those of the Spe-Index and Spe. (2) In terms of the U and D algorithms, classification accuracies of the Spe-Texture were the lowest, 21%, 25% of the Grassland, and 13% and 17% of the Barren Land

were misclassified as Forest and Built-up, respectively. In contrast, the accuracies of Spe-Index were the highest.

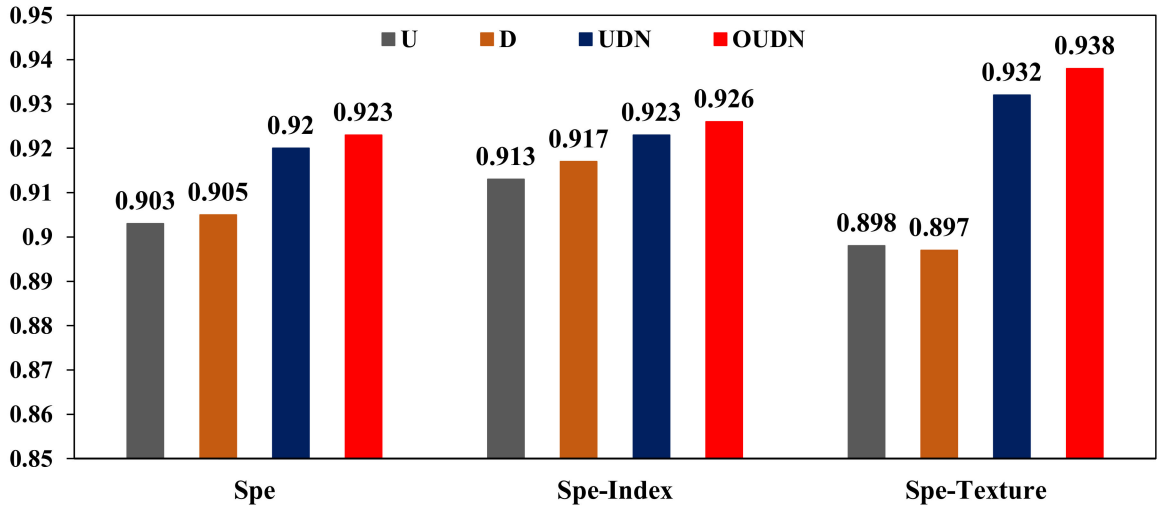

Figure 13. Overall classification accuracies (OA) of different features based on different algorithms.

For urban forests, after texture was added to the Spe, that is, based on the Spe-Texture, the UDN and OUDN algorithms achieved the highest classification accuracies (approximately 100%) for extracting the information of urban forests from VHSR imagery. Similarly, the U and D algorithms also offered relatively obvious advantages for extracting urban forest information based on the feature. As shown in Figure 12, (1) for the U algorithm, the Spe-Index yielded the lowest urban forest extraction accuracy. However, based on Spe-Texture, the accuracy is the highest (99%), and the ratio of Grassland misclassified as Forest was lower (10%) than that of Spe (12%), so the Spe-Texture offered advantages in the extraction of urban forests. (2) For the D algorithm, the urban forest extraction accuracy with the Spe-Texture, compared with those with the Spe and Spe-Index features, was the highest; meanwhile, the ratios of Agricultural Land and Grassland misclassified as Forest were the lowest (2% and 8%). (3) For the UDN and OUDN algorithms, although the urban forest extraction accuracy based on the Spe-Texture was the highest, the ratios of Agricultural Land and Grassland that were misclassified as Forest, compared with those of the other features, were the highest (5% and 8%), thereby resulting in confusion between urban forests and other land use categories.

5. Conclusions

Urban land use classification using VHSR remotely sensed imagery remains a challenging task due to the extreme difficulties in differentiating complex and confusing land use categories. This paper proposed a novel OUDN algorithm for the mapping of urban land use information from VHSR imagery, and the information of urban land use and urban forest resources was extracted accurately. The results showed that the OA of the UDN algorithm for urban land use classification was substantially higher than those of the U and D algorithms in terms of Spe, Spe-Index, and Spe-Texture. Object-based image analysis (OBIA) can address the problem of the "salt-and-pepper" effect encountered in VHSR image classification to a certain extent. Therefore, the OA of urban land use classification and the urban forest extraction accuracy were improved significantly based on the UDN algorithm combined with object-based multiresolution segmentation constraints, which indicated that the OUDN algorithm offered dramatic advantages in the extraction of urban land use information from VHSR imagery. The OA of spectral features combined with texture features (Spe-Texture) in the extraction of urban land use information was as high as 93.8% with the OUDN algorithm, and different land use classes were identified accurately. Especially for urban forests, the OUDN algorithm achieved the highest classification accuracy of 99.7%. Thus, this study provided a reference for the feature setting of urban

forest information extraction from VHSR imagery. However, for the OUDN algorithms, the ratios of Agricultural Land and Grassland misclassified as Forest were higher based on Spe-Texture, which led to confusion between urban forests and other categories. This issue will be further studied in future research.

Author Contributions: Conceptualization, H.D.; data curation, D.Z., M.Z., Z.H., H.L., and X.L. (Xin Luo); formal analysis, S.H., X.L. (Xuejian Li), F.M., and H.L.; investigation, S.H., D.Z., Y.X., M.Z., Z.H., and X.L. (Xin Luo); methodology, S.H.; software, S.H. and Z.H.; supervision, H.D. and F.M.; validation, G.Z.; visualization, X.L. (Xuejian Li) and Y.X.; writing—original draft, S.H.; writing—review and editing, H.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation (No. U1809208, 31670644, and 31901310), the State Key Laboratory of Subtropical Silviculture (No. ZY20180201), and the Zhejiang Provincial Collaborative Innovation Center for Bamboo Resources and High-efficiency Utilization (No. S2017011).

Acknowledgments: The authors gratefully acknowledge the supports of various foundations. The authors are grateful to the editor and anonymous reviewers whose comments have contributed to improving the quality.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Voltersen, M.; Berger, C.; Hese, S.; Schmullius, C. Object-based land cover mapping and comprehensive feature calculation for an automated derivation of urban structure types at block level. *Remote Sens. Environ.* **2014**, *154*, 192–201. [CrossRef]
2. Wu, C.; Zhang, L.; Du, B. Kernel Slow Feature Analysis for Scene Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2367–2384. [CrossRef]
3. Lin, J.; Kroll, C.N.; Nowak, D.J.; Greenfield, E.J. A review of urban forest modeling: Implications for management and future research. *Urban. For. Urban. Green.* **2019**, *43*, 126366. [CrossRef]
4. Ren, Z.; Zheng, H.; He, X.; Zhang, D.; Yu, X.; Shen, G. Spatial estimation of urban forest structures with Landsat TM data and field measurements. *Urban. For. Urban. Green.* **2015**, *14*, 336–344. [CrossRef]
5. Guang-Rong, S.; Wang, Z.; Liu, C.; Han, Y. Mapping aboveground biomass and carbon in Shanghai's urban forest using Landsat ETM+ and inventory data. *Urban. For. Urban. Green.* **2020**, *51*, 126655. [CrossRef]
6. Zhang, M.; Du, H.; Mao, F.; Zhou, G.; Li, X.; Dong, L.; Zheng, J.; Zhu, D.; Liu, H.; Huang, Z.; et al. Spatiotemporal Evolution of Urban Expansion Using Landsat Time Series Data and Assessment of Its Influences on Forests. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 64. [CrossRef]
7. Zhang, Y.; Shen, W.; Li, M.; Lv, Y. Assessing spatio-temporal changes in forest cover and fragmentation under urban expansion in Nanjing, eastern China, from long-term Landsat observations (1987 –2017). *Appl. Geogr.* **2020**, *117*, 102190. [CrossRef]
8. Alonzo, M.; McFadden, J.P.; Nowak, D.J.; Roberts, D.A. Mapping urban forest structure and function using hyperspectral imagery and lidar data. *Urban. For. Urban. Green.* **2016**, *17*, 135–147. [CrossRef]
9. Alonzo, M.; Bookhagen, B.; Roberts, D.A. Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sens. Environ.* **2014**, *148*, 70–83. [CrossRef]
10. Liu, L.; Coops, N.C.; Aven, N.W.; Pang, Y. Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. *Remote Sens. Environ.* **2017**, *200*, 170–182. [CrossRef]
11. Pu, R.; Landry, S. A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sens. Environ.* **2012**, *124*, 516–533. [CrossRef]
12. Pu, R.; Landry, S.M. Mapping urban tree species by integrating multi-seasonal high resolution pléiades satellite imagery with airborne LiDAR data. *Urban. For. Urban. Green.* **2020**, *53*, 126675. [CrossRef]
13. Puissant, A.; Rougier, S.; Stumpf, A. Object-oriented mapping of urban trees using Random Forest classifiers. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 235–245. [CrossRef]
14. Pan, G.; Qi, G.; Wu, Z.; Zhang, D.; Li, S. Land-Use Classification Using Taxi GPS Traces. *IEEE Trans. Intell. Transp. Syst.* **2012**, *14*, 113–123. [CrossRef]
15. Zhao, W.; Du, S. Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS J. Photogramm. Remote Sens.* **2016**, *113*, 155–165. [CrossRef]
16. Moser, G.; Serpico, S.B.; Benediktsson, J.A. Land-Cover Mapping by Markov Modeling of Spatial–Contextual Information in Very-High-Resolution Remote Sensing Images. *Proc. IEEE* **2012**, *101*, 631–651. [CrossRef]

17. 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]
18. Han, N.; Du, H.; Zhou, G.; Xu, X.; Ge, H.; Liu, L.; Gao, G.; Sun, S. Exploring the synergistic use of multi-scale image object metrics for land-use/land-cover mapping using an object-based approach. *Int. J. Remote Sens.* **2015**, *36*, 3544–3562. [CrossRef]
19. Sun, X.; Du, H.; Han, N.; Zhou, G.; Lu, D.; Ge, H.; Xu, X.; Liu, L. Synergistic use of Landsat TM and SPOT5 imagery for object-based forest classification. *J. Appl. Remote Sens.* **2014**, *8*, 083550. [CrossRef]
20. Hamdi, Z.M.; Brandmeier, M.; Straub, C. Forest Damage Assessment Using Deep Learning on High Resolution Remote Sensing Data. *Remote Sens.* **2019**, *11*, 1976. [CrossRef]
21. Shirvani, Z.; Abdi, O.; Buchroithner, M.F. A Synergetic Analysis of Sentinel-1 and -2 for Mapping Historical Landslides Using Object-Oriented Random Forest in the Hyrcanian Forests. *Remote Sens.* **2019**, *11*, 2300. [CrossRef]
22. Stubbings, P.; Peskett, J.; Rowe, F.; Arribas-Bel, D. A Hierarchical Urban Forest Index Using Street-Level Imagery and Deep Learning. *Remote Sens.* **2019**, *11*, 1395. [CrossRef]
23. Abdi, O. Climate-Triggered Insect Defoliators and Forest Fires Using Multitemporal Landsat and TerraClimate Data in NE Iran: An Application of GEOBIA TreeNet and Panel Data Analysis. *Sensors* **2019**, *19*, 3965. [CrossRef] [PubMed]
24. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
25. Romero, A.; Gatta, C.; Camps-Valls, G. Unsupervised Deep Feature Extraction for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 1349–1362. [CrossRef]
26. Dong, L.; Du, H.; Mao, F.; Han, N.; Li, X.; Zhou, G.; Zhu, D.; Zheng, J.; Zhang, M.; Xing, L.; et al. Very High Resolution Remote Sensing Imagery Classification Using a Fusion of Random Forest and Deep Learning Technique—Subtropical Area for Example. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *13*, 113–128. [CrossRef]
27. Zhao, W.; Du, S.; Wang, Q.; Emery, W. Contextually guided very-high-resolution imagery classification with semantic segments. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 48–60. [CrossRef]
28. Sherrah, J. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. *arXiv* **2016**, arXiv:1606.02585.
29. Liu, Y.; Fan, B.; Wang, L.; Bai, J.; Xiang, S.; Pan, C. Semantic labeling in very high resolution images via a self-cascaded convolutional neural network. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 78–95. [CrossRef]
30. Sun, Y.; Zhang, X.; Xin, Q.; Huang, J. Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 3–14. [CrossRef]
31. Chen, G.; Zhang, X.; Wang, Q.; Dai, F.; Gong, Y.; Zhu, K. Symmetrical Dense-Shortcut Deep Fully Convolutional Networks for Semantic Segmentation of Very-High-Resolution Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1633–1644. [CrossRef]
32. Chen, K.; Weinmann, M.; Sun, X.; Yan, M.; Hinz, S.; Jutzi, B. Semantic Segmentation of Aerial Imagery via Multi-Scale Shuffling Convolutional Neural Networks with Deep Supervision. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *4*, 29–36. [CrossRef]
33. Huang, B.; Zhao, B.; Song, Y. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sens. Environ.* **2018**, *214*, 73–86. [CrossRef]
34. Xu, J.; Feng, G.; Zhao, T.; Sun, X.; Zhu, M. Remote sensing image classification based on semi-supervised adaptive interval type-2 fuzzy c-means algorithm. *Comput. Geosci.* **2019**, *131*, 132–143. [CrossRef]
35. Mi, L.; Chen, Z. Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 140–152. [CrossRef]
36. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. High-resolution semantic labeling with convolutional neural networks. *arXiv* **2016**, arXiv:1611.01962.
37. Liu, Y.; Piramanayagam, S.; Monteiro, S.T.; Saber, E. Dense Semantic Labeling of Very-High-Resolution Aerial Imagery and LiDAR with Fully-Convolutional Neural Networks and Higher-Order CRFs. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1561–1570.
38. Zhang, B.; Zhao, L.; Zhang, X. Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images. *Remote Sens. Environ.* **2020**, *247*, 111938. [CrossRef]

39. Dumoulin, V.; Visin, F. A guide to convolution arithmetic for deep learning. *arXiv* **2016**, arXiv:1603.07285.
40. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
41. Wang, W.; Dou, S.; Jiang, Z.; Sun, L. A Fast Dense Spectral–Spatial Convolution Network Framework for Hyperspectral Images Classification. *Remote Sens.* **2018**, *10*, 1068. [CrossRef]
42. Li, R.; Duan, C. LiteDenseNet: A Lightweight Network for Hyperspectral Image Classification. *arXiv* **2020**, arXiv:2004.08112.
43. Bai, Y.; Zhang, Q.; Lu, Z.; Zhang, Y. SSDC-DenseNet: A Cost-Effective End-to-End Spectral-Spatial Dual-Channel Dense Network for Hyperspectral Image Classification. *IEEE Access* **2019**, *7*, 84876–84889. [CrossRef]
44. Li, G.; Zhang, C.; Lei, R.; Zhang, X.; Ye, Z.; Li, X. Hyperspectral remote sensing image classification using three-dimensional-squeeze-and-excitation-DenseNet (3D-SE-DenseNet). *Remote Sens. Lett.* **2020**, *11*, 195–203. [CrossRef]
45. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015.
46. Flood, N.; Watson, F.; Collett, L. Using a U-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across Queensland, Australia. *Int. J. Appl. Earth Obs. Geoinfor.* **2019**, *82*, 101897. [CrossRef]
47. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 94–114. [CrossRef]
48. Zhao, W.; Du, S.; Emery, W.J. Object-Based Convolutional Neural Network for High-Resolution Imagery Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3386–3396. [CrossRef]
49. Zhang, C.; Sargent, I.; Pan, X.; Li, H.; Gardiner, A.; Hare, J.S.; Atkinson, P.M. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sens. Environ.* **2018**, *216*, 57–70. [CrossRef]
50. Liu, T.; Abd-Elrahman, A. Deep convolutional neural network training enrichment using multi-view object-based analysis of Unmanned Aerial systems imagery for wetlands classification. *ISPRS J. Photogramm. Remote Sens.* **2018**, *139*, 154–170. [CrossRef]
51. Martins, V.S.; Kaleita, A.L.; Gelder, B.K.; Da Silveira, H.L.; Abe, C.A. Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 56–73. [CrossRef]
52. Zhang, C.; Yue, P.; Tapete, D.; Shangguan, B.; Wang, M.; Wu, Z. A multi-level context-guided classification method with object-based convolutional neural network for land cover classification using very high resolution remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *88*, 102086. [CrossRef]
53. Tong, X.-Y.; Xia, G.-S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [CrossRef]
54. Anderson, J.R. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*; USGS Professional Paper, No. 964; US Government Printing Office: Washington, DC, USA, 1976. [CrossRef]
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]
56. Sun, J.; Wang, H.; Song, Z.; Lu, J.; Meng, P.; Qin, S. Mapping Essential Urban Land Use Categories in Nanjing by Integrating Multi-Source Big Data. *Remote Sens.* **2020**, *12*, 2386. [CrossRef]
57. Bharati, M.H.; Liu, J.; MacGregor, J.F. Image texture analysis: Methods and comparisons. *Chemom. Intell. Lab. Syst.* **2004**, *72*, 57–71. [CrossRef]
58. Li, Y.; Han, N.; Li, X.; Du, H.; Mao, F.; Cui, L.; Liu, T.; Xing, L. Spatiotemporal Estimation of Bamboo Forest Aboveground Carbon Storage Based on Landsat Data in Zhejiang, China. *Remote Sens.* **2018**, *10*, 898. [CrossRef]
59. Fatiha, B.; Abdelkader, A.; Latifa, H.; Mohamed, E. Spatio Temporal Analysis of Vegetation by Vegetation Indices from Multi-dates Satellite Images: Application to a Semi Arid Area in Algeria. *Energy Procedia* **2013**, *36*, 667–675. [CrossRef]

60. Taddeo, S.; Dronova, I.; Depsky, N. Spectral vegetation indices of wetland greenness: Responses to vegetation structure, composition, and spatial distribution. *Remote Sens. Environ.* **2019**, *234*, 111467. [CrossRef]
61. Zhang, M.; Du, H.; Zhou, G.; Li, X.; Mao, F.; Dong, L.; Zheng, J.; Liu, H.; Huang, Z.; He, S. Estimating Forest Aboveground Carbon Storage in Hang-Jia-Hu Using Landsat TM/OLI Data and Random Forest Model. *Forests* **2019**, *10*, 1004. [CrossRef]
62. Ren, H.; Zhou, G.; Zhang, F. Using negative soil adjustment factor in soil-adjusted vegetation index (SAVI) for aboveground living biomass estimation in arid grasslands. *Remote Sens. Environ.* **2018**, *209*, 439–445. [CrossRef]
63. Hadji, I.; Wildes, R.P. What do we understand about convolutional networks? *arXiv* **2018**, arXiv:1803.08834.
64. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
65. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
66. 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.
67. Yin, R.; Shi, R.; Li, J. Automatic Selection of Optimal Segmentation Scale of High-resolution Remote Sensing Images. *J. Geo-inf. Sci.* **2013**, *15*, 902–910. [CrossRef]
68. He, T.; Zhang, Z.; Zhang, H.; Zhang, Z.; Xie, J.; Li, M. Bag of Tricks for Image Classification with Convolutional Neural Networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 558–567.
69. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [CrossRef]