Farm Parcel Extraction in High Resolution Remote Sensing Image Based on Hierarchical Spectrum and Shape Features

Bangyu Li (✉ bangyu.li@ia.ac.cn)

Institute of Automation Chinese Academy of Sciences <https://orcid.org/0000-0002-9887-3242>

Research Article

Keywords: Farm parcel, Segmentation, Remote sensing images, Shape, TransUNe

Posted Date: January 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1218962/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Farm Parcel extraction in high resolution remote sensing image based on hierarchical spectrum and shape features

Bangyu Li^{1,2*}, Yirong Wang², Liang Qi^{2†}, Zhu Lu² and Yongqi Xu²

Abstract

Background: Land-use classification schemes typically address both land use and land cover. Vectorized data extracted from farm parcel segmentation provides important cadastral data for the formulation and management of climate change policies. It also provides important basic data for research on pest control in large areas, crop yield forecasts, and crop varieties classification. It can also be used for the assessment of compensation for damages related to extreme weather events by the agricultural insurance department. Firstly, we investigate the effectiveness of an automated image segmentation method based on TransUNet architecture to enable that automate the task of farm parcel delineation that originally relied on high labor costs. Then, post-processing by vectoring binary segmentation image, which the area and regularity parameter to adjust the accuracy of segmentation, can get a more optimized image segmentation result.

Results: The results on the existing data show that the automatic segmentation system we proposed is a method that can effectively divide various types of agricultural land. The system was trained and evaluated using 94780 images. The performance parameters obtained showed that the accuracy rate reached 83.31%, the recall rate reached 82.13%, the F1-S rate was 80.37%, the total accuracy rate was 82.23%, and iou was 80.39%. At the same times, without losing too much accuracy, we train and test the model with 3m resolution image, which has the advantage of processing speed than 0.8m resolution. Therefore, our proposed method can be effectively applied to the task of extraction of agricultural land, which is better and more efficient than most manual annotations.

Conclusions: We have demonstrated the effectiveness of strategy using a TransUNet architecture and post-processing by vectoring binary segmentation for farm parcel extraction in high remote sensing images. The success of our approach is also a demonstration of feasibility of the deep learning to participate in and improve agricultural production activities, which is important for achieving scientific management of agricultural production.

Keywords: Farm parcel; Segmentation; Remote sensing images; Shape; TransUNet

*Correspondence: bangyu.li@ia.ac.cn

¹Institute of Automation, Chinese Academic Sciences, Beijing, China

²School of Electronics and Information, Jiangsu University of Science and Technology, 212003 Zhenjiang, China

Full list of author information is available at the end of the article

[†]Equal contributor

Introduction

In recent years, with the continuous growth of population and the decrease of arable land resources, modern agricultural production is developing towards intensive and refined direction, and people's demand for dynamic, large-scale,

timely and rapid arable land spatial information has become increasingly urgent. Since remote sensing technology can repeatedly obtain large-scale farmland information data, it is increasingly applied to land cover information extraction [1].

For the classification of remote sensing images, the traditional method is to use statistical methods to extract low-level features, including distance [2], K-nearest neighbor [3], maximum likelihood [4] and logistic regression [5] classifier. However, with the rapid development of aerospace, sensor and computer technologies, high-resolution remote sensing images are increasingly applied to land use classification [6]. But, spectral confusion exists in high-resolution remote sensing images, which reduces the effectiveness of traditional classification methods based on low-level features [7]. Low-level features encoded by typical pixel-based or object-based methods, such as spectrum, texture and geometric attributes, become basically invalid [8–10].

So more complex features and descriptors are needed to capture the semantics of the scene. An approach [11] based on middle-level feature modeling is developed on the basis of low-level feature method. Three types of mid-level feature extraction methods describe image semantics, namely the bag-of-visual-words (BoVW), latent Dirichlet allocation (LDA), and machine learning models. In practical application, the SIFT-BOVW method has been successfully applied to remote sensing land use image classification [12]. However, in practical applications, the performance of BoVW-based methods relies on the extraction of handcrafted local features [13]. LDA modeling methods rely on K-means clustering to produce a visual dictionary. Therefore, these are not the most suitable segmentation methods for high-resolution remote sensing images [14]. Machine learning models independently perform data expression and feature extraction [15] and discard the extracted feature mode [16, 17] according to pre-determined rules. Therefore, machine learning can often achieve better classification

results when processing complex hyperspectral remote sensing images.

Commonly used machine learning methods include sparse coding [18], neural network [19], support vector machine [20][21] and Deep learning [22]. Among them, convolutional network has shown good performance in semantic segmentation [23–25]. In [26], the author trained a one-dimensional CNN containing input layer, convolution layer, maximum pooling layer, full connection layer and output layer to classify hyperspectral images directly in the spectral domain. Makantasis et al. [27] used a two-dimensional CNN to encode spectral and spatial information. Zhao and Du [28] proposed a classification framework based on spectral spatial features, which combined dimensionality reduction algorithm based on local discriminant embedding and two-dimensional CNN. Paoletti et al. [29] also developed a 3D CNN network capable of simultaneously processing spectral and spatial features of hyperspectral images. The authors of [30] propose an unsupervised convolutional network for learning spectral-spatial features using sparse learning to estimate the network weights in a greedy layerwise fashion instead of end-to-end learning. Mou et al. proposed a network structure for unsupervised spectral spatial feature learning of hyperspectral images, called total residual convolution - deconvolution network [31, 32]. In [33], the author proposed a RNN model for hyperspectral image classification based on a new activation function and an improved gate cycle element, which can effectively analyze hyperspectral pixels as sequence data and then determine information categories through network reasoning. Zhao and Du [34] proposed a multi-scale convolutional neural network for learning depth features of spatial relations, Pyramid structure is constructed according to the image to present spatial characteristics of different scales. In 2016, Lin et al. proposed the Refine NET [35] model for recognizing natural images. The Refine NET model is a universal multi-path optimization network, which explicitly takes the advantage of

all available information during the entire down-sampling process and uses remote residual connection to achieve high-resolution prediction.

In addition, some different methods have achieved some results in the classification of high resolution remote sensing images. GEO-BIA is a method that provides a high spatial resolution image analysis using spectral, spatial, textural and topological features [36]. The basic unit of analysis is an image object rather than a single pixel, so this method intends to bypass the problem of artificial square cells as used in per-pixel method[37]. Liu Weifeng proposed Multiview hessian regularization for image annotation [38] and Multiview hessian discriminative sparse coding for image annotation [39], and conducted a large number of experiments on PASCAL VOC '07 data set to verify the effectiveness of the methods. Lidar point clouds have been used in many semantic segmentation applications to provide additional dimensional information [40]. F.L. Luo et al. [41] proposed hyperspectral image feature learning using spatial hypergraph discriminant analysis. SAE[42] and DBN[43] were also used to extract the hierarchical features of the spectral domain. Tao et al. [44] used sparse SAEs to learn effective feature representations from input data. Then, the learned features are input into linear support vector machine for hyperspectral data classification. In [45], the author proposed a remote sensing image segmentation method using both spectral and texture information. Yao et al. [46] proposed a remote sensing image segmentation method based on adaptive clustering ensemble learning. Shen et al. [47] proposed a two-group particle swarm optimization algorithm to improve the performance of remote sensing image segmentation. A multi-layer fusion model for adaptive segmentation and change detection of optical remote sensing image sequences is proposed[48]. In [49] and [50], a multiple Feature Pyramid Network (MFPN) framework is proposed, which utilizes multi-level semantic features of high-resolution remote sensing images to implement effective feature pyramids and cus-

tomized pyramid pooling modules. Deep convolutional neural networks (CNNs) have been successfully applied to various tasks including image recognition [51–53], object detection[54–57], and object tracking[58–61], and achieved good results.

We investigated the application of using high-resolution remote sensing satellite images to delineate farm parcel areas. Firstly, the ArcGIS software to manually annotate the target image is used to obtain the training samples. Secondly, pre-process by augmented images is achieved in data set, and the modified TransUNet network is trained to build the models in different resolution 0.8m and 3.2m respectively. Finally, post-processing techniques by tuning parameter in the shape features are used to further improve the recognition results, and testing image of different resolutions are used on different models to perform effectiveness so that a better recognition result image can be obtained.

Materials And Method

Data Description

We propose TransUNet to process the recognition of farm parcel. It has the both advantages of U-Net and Transformer. On one hand, with the combination of U-Net to enhance finer details by recovering localized spatial information. On the other hand, the Transformer to enable precise localization[62].

The Gaofen-2 satellite has two sensors in different bands with different resolutions, and panchromatic wavebands have one band ranging from 0.4 to 0.9 and multi-wavebands have four bands ranging from 0.45 to 0.89. A 0.8 resolution remote sensing image can be obtained by combining two sensor images. In the case of accuracy, we did not fuse into imagery in different resolution. The process of land parcel extraction was accelerated with different resolutions. The main parameters of Gaofen-2 satellite are shown in Table1.

Firstly, all farm parcel types are extracted as a class. Typical parcels include Polder, Duo farmland, Dam farmland and etc as shown in Fig.1. The types of parcels are described as:

Table 1 Main indicators of Gaofen-2 satellite

Sensor	Band number	Spectral range	Spatial resolution	Width
Panchromatic camera	1	0.4-0.9 μm	1 m	45km
	2	0.45-0.52 μm	3.2 m	
Multispectral camera	3	0.52-0.59 μm	3.2 m	
	4	0.63-0.69 μm	3.2 m	
	5	0.77-0.89 μm	3.2 m	

- **Polder:** Farmland reclaimed by embankments along the river, near the sea, or lake-side areas. Ancient Chinese farmers invented polder to rebuilding low-lying land and transform the lake into the ground.
- **Duo farmland:** Stacked high fields formed by the excavation of deep net-shaped trenches or small rivers in low-humid areas along lakes or river networks in southern China. Its terrain is high, the drainage is good, and the soil is fertile and loose. It is suitable for planting various dry crops, especially for the production of melons and vegetables.
- **Dam farmland:** It is farmland created by people damming in a ravine and blocking the soil washed down from the mountain.
- **Strip farmland:** Strip farmland is a field surrounded by agricultural ditches and

windbreak forest area. It is the basic unit of land use for farming, irrigation, and soil improvement.

- **Dry land:** Dry land refers to fields that do not store water on the ground. Its irrigation mainly relies on natural precipitation. Dryland is generally planted with xerophytic crops such as wheat, corn, and cotton.
- **Paddy Fields:** Paddy fields are arable land that has been flooded and used to grow semi-aquatic crops, especially rice and taro. Paddy farming is still the main form of modern rice cultivation.
- **Platform farmland:** A field shaped like a platform above the ground and surrounded by ditches. The platform farmland crops are mainly grain, cotton, vegetables and other crops. And it is a land improvement project to eliminate waterlogging and alkalinity.
- **Terrace:** In agriculture, a terrace is a sloping flat surface that is cut into a series of continuous receding flat surfaces or platforms, similar to steps, for the purpose of farming more efficiently. Terraced fields are

widely used in the cultivation of rice, wheat and barley in East Asia, South Asia and Southeast Asia.

Data Generation

There are many labeling methods for deep learning data set samples, which are mainly operated by Labelme[63], RSLabel and ArcGIS software[64]. The purpose of labeling is to classify the background and various plots in the sample and label them with their corresponding labels to facilitate model learning. The marking method is as follows:

- **Labelme:** Labelme is not convenient for managing remote sensing image annotation in multi-channel iamges and is not easy to modify in annotation. The labeling of farm parcels needs to be modified frequently due to undefined attributes and scales.
- **ArcGIS:** ArcGIS uses vector polygons by shpfiles to label remote sensing images which easily modify polygons in the area of farm parcels. The biggest advantage of using ArcGIS labeling is that we only need to manually label high-resolution samples to generate binarized labels for samples with 3.2m resolutions which can be mapped to the 0.8m resolution image. Annotation by the polygons can be used for images with different resolutions at one time as shown in Fig.2.

The green vector polygons are the marked farm parcel, and the rest are the unlabeled areas. After the labeling is completed, assign the attribute values of the farm parcel labels uniformly and use the rasterization function of ArcGIS to rasterize the vector labels to generate a binarized label file, and finally use ArcGIS itself or Python to binarize the entire image. The image label is cut into the original image and the annotated image corresponding to a fixed size at the same time, as shown in Fig.3.

Methods

The process of Farm Parcel Segmentation system includes three main stages: 1) Preprocessing, 2) semantic segmentation 3) post-processing, as shown in Fig.4.

Figure 3 Example of the original image of the sample after cropping and the corresponding annotated image

Figure 4 The flow chart of farm parcel extraction

Data Augmentation

As the neural network deepens, the parameters that need to be learned will also increase. When the data set is small, too many parameters will fit all the characteristics of the data set, rather than the commonality between the data, which will not meet the needs of the generalization ability required by the network model. In order to obtain more data and reflect the commonality of more data sets, we need to make minor changes to the existing data sets. This small change can not only reflect the common ground of the data set, but also make the neural network think that these are completely different images. When our neural network is trained with additional data sets, it can interpret remote sensing images of different locations, different light conditions, and different proportions.

The main method of data augmentation is to process samples such as flip, rotation, scale, crop, translation, and Gaussian noise. Its purpose is to increase the amount of training data, prevent over-fitting, and improve the generalization ability of the model, increase noise data to

Figure 2 Using Vector polygons to generate annotated binarized label images for different resolutions (Left: 3.2m resolution, Right: 0.8m resolution)

Figure 5 Data augmentation with mask by different techniques^[1]

improve the robustness of the model. The following data is rotated and flipped, and white noise, color enhancement and Gaussian noise are added, as shown in Fig.5 (only the label part is shown). The labels of the original image, Gaussian noise, tone enhancement and white noise are the same.

Segmentation For Transfromer and Unet

Based on the characteristics of remote sensing images and semantic segmentation technology, the TransUNet network architecture is proposed as shown in Fig.6 [62]. The overall network structure is divided into three parts: Downsample layer based on Resnet[51]; Transformer Layer based on attention mechanism with residual structure; Upsample Layer based on full convolution (FCN)[65].

Downsample Layer

This model uses the classic Resnet[51] structure as the down-sampling network layer. The down-sampling network mainly extracts image features, and the extracted feature information participates in the information fusion of the Upsample Layer (the embedded information output by Attention is used as another fusion information). Among them, only the first output feature layer uses maximum pooling, and the other output feature layers use the FCN structure. Downsample Layer is composed of basic residual basic Block, each Basic Block is composed of cross-layer residual block and non-cross-layer residual block. The cross-layer residual block mainly implements the extraction of information between feature maps of inconsistent sizes between layers, while the non-cross-layer residual block implements the extraction of information between feature maps of the same size.

Transformer Layer

Mainly composed of the following parts: local feature information segmentation; feature information Embedding Layer; Multi-Head Attention Layer; Feed-forward Network Layer. The local feature information segmentation is mainly

to segment the output features of the last layer of Downsample Layer. Embedding Layer includes Patch Embedding and Position Embedding, the two are combined for flattening processing, as the input of Attention Layer. Attention self-attention mechanism mainly focuses on the connection between each encoded embedding and other embeddings, and calculates related self-attention. The embedding of each input is multiplied by the weight of the Multi-Head Attention Layer to calculate the three vectors of the corresponding embedding: query vector (q), key vector (k) and value vector (v). Calculate the self-attention vector of the current embedding and other embeddings through each key vector and query vector, and express them in the form of probability (weight), multiply the probability by the value vector and add them to get the output of the attention layer embedding from that location. Multi-Head Attention deals with the same embedding in the horizontal direction. The Feed-forward Network Layer is mainly to normalize the Attention output on the weight layer. Innovations of Transformer Layer: 1. Transformer Layer Embedding uses convolution instead of linear processing; 2. In Attention attention, query vector (q), key vector (k), value vector (v), change q , k or v to q' , k' , v , minimum local attention to the same feature embedding (Pay attention to specific details, such as the connection between the edge of the plots and the center); 3. Add an attention mechanism to the feature information output by each Downsample Layer, replace the direct output as the input source of fusion information.

Upsample Layer

Upsample Layer mainly includes: Transformer Layer output layer inverse flattening; Upsample Layer convolution; feature fusion; Segmentation-Head output layer. Inverse flattening mainly restores the embedded features of the Transformer Layer output layer before Embedding, and arranges the new feature information learned by Attention according to the position before Attention learning. Feature fusion is to fuse the

^[1]A,I,J and K are same in mask image

information after the Upsample Layer convolution with the feature information of the Down-sample Layer at the same latitude(fuse a total of 5 scales). The Seg-Head output layer outputs the fused and convolved information according to categories.

Upsample Layer innovation point: Only one convolution is performed to ensure that the fused feature information will not cause gradient problems due to the increase of the number of layers.

Post-processing

After the semantic segmentation network is processed, the resultant image often has pixel-level misrecognition, including voids and noise regions, vector polygons are not smooth and trivial small areas, etc. Therefore, it is necessary to design a flexible post-processing scheme to improve the recognition accuracy for farm parcel. The post-processing scheme generally includes three main steps: morphological analysis, vectorization operation, and data fusion, as shown in Fig.7. Post-processing is an important mean-

accurate land recognition effect. Use the following three methods to operate on the identified binarized image.

Morphological operations: expansion, corrosion, opening and closing operations, hit and miss transformation, TOP-HAT transformation, black hat transformation, etc. Combining smoothing and opening operations to eliminate small objects in the image. Corrosion and then expansion combined with closing operations are used to fill small holes inside the target.

Vectorization operation: The area is judged on the recognition result, and small areas are removed. Then adjust the vector polygons of each farm parcel to merge and delete the interference area.

Fusion: The vector polygons data generated by the binarization map after morphological operation is fused with the vector data generated directly from the original data, and the neighboring vectors are judged by geographic information, and the recognition results are optimized through vector attributes(area, regularity, etc.) The original image and the result of recognition

s for the secondary correction of the recognition results and the practical application of engineering. When the processing results of the recognition algorithm are not ideal, the post-processing technology is used to make subsequent adjustments to the recognition results to obtain a more

are shown in Fig.8 (left: original image, right: image after deep learning recognition), and the result of morphological processing after deep learning recognition is shown in Fig.9. It can be seen that morphological processing is difficult to control the expansion coefficient. In an image with a resolution of 3.2m, an expansion of one pixel will affect the actual situation of 3.2 meters, which will cause the edge of the farm parcel to

be misidentified. (The natural boundary of the

farmland is defined at 2 meters, and the remote sensing image has a resolution of 3.2m and 0.8m). Therefore, in this scheme, the area parameter is first used to filter the misrecognition of trivial areas as defined:

$$A = \sigma * P \quad (1)$$

Among them, A is the area, σ is the spatial resolution and P is the number of pixels. Filter small area vector polygons by setting a threshold, as shown in Fig.10(B). Then the area vector is smoothed. The threshold of the smoothing scale can be set smaller, because the definition of the edge of the plots is greater than 2 meters, and the buffer distance is set to 0.5 meters, the definition of regularity is defined:

$$\beta = (A / \min AreaRect(A)) \quad (2)$$

$\beta \in (0,1)$, if it is close to 1, it proves that the extracted target is relatively regular, and if the value is small, it proves that the extraction is incomplete or the recognition ability is insufficient. An iterative calculation based on the extracted area is proposed here. If the set regularity threshold is met, the area will be filled and smoothed until the set threshold is met. The operation of the degree of regularity can use the dynamic sorting method to arrange them

from large to small. In order to improve efficiency, first use the structure of GDAL to establish a regularity field linked list, calculate the area, and then calculate the regularity and add it to the field linked list, and put the largest regularity on the upper level through quick sorting. The regularization operation is performed on the part of the linked list whose regularity is greater than the threshold, without the need to perform a global traversal operation on the entire vector polygons data, which can increase the post-processing speed. Fig.10 has shown the results of vector polygons data processing by processing some specific areas. A, B, C, and D in Fig.10 show various results of post-processing by vectorization technology when the recognition effect is not ideal.

Experiments And Results

This part will introduce the data source, data production process, training environment and various algorithm performance comparison. First, we obtained the remote sensing image data of the GaoFen-2 satellite, and generated

66346 training sets, 18956 verification sets and 9478 test sets through manual annotation and enhancement processing, a total of 94780 remote sensing image data sets of 256x256 pixels. Then, use these data to train various algorithms and compare the training results of different algorithms.

Experimental Data

The research scope of the farm parcel identification in this paper has selected the areas containing more typical farm parcels and crops area in Zhejiang and Anhui, as shown in Fig.11. Anhui Province is located in the eastern part of mainland China, with 4.22 million hectares of arable land in the province, and Zhejiang Province is located on the southeast coast of China, with 1.98 million hectares of arable land in the province. These two regions have large crop planting areas, high yields, and relatively intensive crops. Therefore, their characteristics are complex and have strong representation. Studying them is of great significance for ensuring national food security, and it is also of guiding significance for the identification of farm parcel in other regions of China. The experimental data contains two resolutions of 0.8m and 3.2m, as well as four channels. Since different band combinations can highlight the different features of the farm parcel, we choose the standard false-color combination (red, green and blue are assigned to bands 4, 3 and 2 respectively) to label farm parcel. The vegetation in the standard false-color image is displayed in red, which can highlight the characteristics of vegetation, as shown in Fig.12, which is the most suitable band combination for make experimental data. After the label is completed, we use the python segmentation tool to refine and preprocess the original image and its label, and we get the experimental data. It contains a total of that Anhui remote sensing images with a size of 7300x6908 with two different resolutions, as well as two Zhejiang remote sensing images with a size of 27500x26760 and 6882x7329 with two different resolutions.

Because the original image is too large, we use the python program to crop the original image

Figure 12 An example of the standard false-color view labeling of the 0.8-meter resolution remote sensing image in Zhejiang

into 256x256 pixel data. After cutting, the four original images became 11229 Zhejiang remote sensing images with 0.8 meters resolution and 769 Zhejiang remote sensing images with 3.2 meters resolution, and 1542 remote sensing images of Anhui with a resolution of 0.8 meters and a resolution of 3.2 meters, a total of 13,540 remote sensing images of 256x256 pixels.

Enough training samples will make the network training effect better. When the training samples are insufficient, we can augment the data set. After flip, rotation, scale, crop, translation, Gaussian noise, etc., the total data set pictures are 94780 images. The training data set, validation data set, and test data set are divided according to the ratio of 7:2:1, which means that there are 66346 training sets, 18956 verification sets and 9478 test sets to participate in the training. In addition, in order to further verify the performance of our algorithm, we additionally prepared 2341 remote sensing images from the QuickBird satellite as a test set to verify the generalization ability of the segmentation model.

Experiments Results

All the experiment are based on PyTorch to achieve Deeplabv3 and Deeplabv3+ method-

Figure 11 Illustration of data augmentation.

TP = Correctly identified farmland plots   Recall rate = $TP / (TP + FN)$
 FP = Misrecognition of non-farm land plots   Accuracy = $TP / (TP + FP)$
 FN = Misidentification of farmland plots
 TN = Correctly identify non-farmland plots
 $F1-S = 2 * \text{Accuracy} * \text{Recall rate} / (\text{Accuracy} + \text{Recall rate})$
 $IOU = A \cap B / (A \cup B)$, A is the prediction result, B is the Groud Truth
 Total accuracy = $(TP + TN) / (FP + FN + TP + TN)$
 $FWIoU = [(TP + TN) / (FP + FN + TP + TN)] * [TP / (TP + FP + FN)]$

Figure 13 Algorithm indicator parameter description

s and TensorFlow to achieve our methods, running on Ubuntu 16.04 LTS 64 and Dual-processor Inter(R) Xeon, GPU Silver 4114 2.20GHz, memory 128G. The two nvidia's GTX1080Ti is used to accelerate the deep learning model.

Evaluations Of Different Resolution

The process of farm parcel Segmentation includes input image, block processing, semantic segmentation, Stitching, vectorization and output result of vectorization. Therefore, the segmentation performance indicators of remote sensing images should not only include the conventional performance parameters of the segmentation model, but the time-consuming and

recognition speed of each functional module is also very important. The experiment is divided into training data set, verification data set and test data set according to 7: 2: 1, batch-size is set to 32, epoch is set to 50, and the initial learning rate is set to 0.001 for performance testing. We ran our algorithm on two data set with different resolutions, and compared the performance of each stage in detail, as shown in Table2 below. It can be seen from the data that it takes significantly more time to process high-resolution images than to process low-resolution images, and the time-consuming for cropping, recognition, and stitching is more than ten times that of low-resolution images. From the perspective of model performance, the accuracy, recall, and intersection ratio of the algorithm in this paper on high-resolution images are also weaker than those on low-resolution images. For the problem of farm parcel identification, the performance of image segmentation is mainly evaluated according to the indicators in Fig.13.

Evaluations Of Different Methods

In order to better evaluate the method in this article, we also used other methods to conduct

Table 2 Performance indicators for recognition at different resolutions (accuracy takes 2 decimal places)

time-consuming(s)	Resolution(m)	Crop	Identify	Stitching	Vectorization	Total time consuming		
	3.2	40	75	355	5+26	501		
	0.8	475	986	4950	1411+169	7991		
Model indicators(%)	Resolution(m)	Accuracy	Recall rate	F1-S	Total accuracy	IoU	mIoU	FWIoU
	3.2	90.19	89.75	90.95	89.97	87.76	85.23	86.34
	0.8	86.58	84.38	85.46	85.53	81.36	81.48	81.49

Table 3 Performance comparison of multiple algorithms

Algorithm name	Accuracy(%)	Recall rate(%)	F1-S(%)	Total accuracy(%)	IoU(%)	mIoU(%)	FWIoU(%)
TransUNet	83.31	82.13	80.37	82.23	80.39	79.23	80.59
Unet	78.25	79.30	78.65	79.30	82.12	80.35	81.89
Deeplabv3+	77.32	80.56	78.32	74.25	74.30	75.23	77.98
Deeplabv3	74.23	73.21	72.23	73.34	75.32	71.25	72.65

comparative experiments [66]^[2]. Use the above data set to compare with Deeplabv3, Unet and the more advanced Deeplabv3+, we conducted multiple experiments to verify the semantic segmentation effect, as shown in the following Table3 . During the experiment, the input image size, learning rate, epoch, batch-size, and experimental hardware environment of the above-mentioned network were controlled to be consistent, and data was not used for online amplification. As shown in Fig.14 , it is a comparison of the results of the algorithm test in this paper and the content of the label. The green framed content is the artificially labeled farm parcel, and the red framed content is the results of the farm parcel range identified by the algorithm in this paper. It can be found that the recognition results are accurate.

It can be seen from the above data that compared with the existing classic segmentation methods, the improved recognition network in this paper can effectively improve the accuracy of the original model for extracting remote sensing farm parcel features and the learning ability of edge details, which can provide an effective solution to the problem of farm parcel identification in high-resolution remote sensing images.

^[2]8 Segmentation Models, we achieve Unet, deeplabv3+ and pretrained Unet and pretrained Unet

Conclusion And Perspectives

In this study, we proposed TransUNet architecture to segment farm parcel areas and boundaries in high resolution remote sensing image. We built our own farm parcel data set and applied it to training various neural network models. The results show that our TransUNet model achieves an accuracy of 83.31% and performs better than other algorithms on other parameters. This shows that the algorithm we proposed can better identify farm parcels and may achieve better results in practical applications. Due to the particularity of remotely sensed plots objects, general recognition models cannot perform well when dealing with such objects, and cannot reach the performance and efficiency of the method in this paper. In addition, the data set used in this article is a small self-made data set with a small amount of data, which may be one of the reasons why the existing models cannot perform well. If the amount of data is further increased, the performance of the existing model and the performance of the method in this paper may be improved to a certain extent. Finally, in the recognition results, there is an inaccurate recognition of the edge of the plots. In addition to expanding the number of training sets, improving the accuracy of manual labeling may also directly improve the efficiency of training and the recognition effect.

We divided the GaoFen-2 satellite data into a training set and a verification set for testing,

and satisfactory results can be obtained. However, when we use the remote sensing image of the QuickBird satellite as a verification set to test the segmentation effect, we found that the effect is not good. As a future work, in order to improve the versatility in different scenarios, we will increase the data of the QuickBird satellite for training to further improve the generalization ability of the model.

Acknowledgements

We thank Hui Zhang and Daoliang Tan for their continued support and supervision and for the suggestions on augmentation techniques and the Strategies in this study.

Funding

Not applicable

Availability of data and materials

The authors are pleased to share analysed/raw data and remote sensing image upon reasonable request.

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable

Authors' contributions

BL and QL conceived and designed the study. YW and QL performed the TransUNet network and post-processing in segmentation image, YX achieved augmented image to form data set. YW, ZL and YX did the annotations and collaborated on the design of the study and realized the Comparative methods. BL, YW wrote the manuscript. All authors read and approved the final manuscript.

Author details

¹Institute of Automation, Chinese Academic Sciences, Beijing, China. ²School of Electronics and Information, Jiangsu University of Science and Technology, 212003 Zhenjiang, China.

References

1. Pu, R., Landry, S.: A comparative analysis of high spatial resolution ikonos and worldview-2 imagery for mapping urban tree species. *Remote Sensing of Environment* **124**, 516–533 (2012)
2. Tzeng, Y.-C.: Remote sensing images classification/data fusion using distance weighted multiple classifiers systems. In: 2006 Seventh International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'06), pp. 56–60 (2006). IEEE
3. Meng, Q., Cieszewski, C.J., Madden, M., Borders, B.E.: K nearest neighbor method for forest inventory using remote sensing data. *GIScience & Remote Sensing* **44**(2), 149–165 (2007)
4. Bruzzone, L., Prieto, D.F.: Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* **39**(2), 456–460 (2001)
5. Lee, S.: Application of logistic regression model and its validation for landslide susceptibility mapping using gis and remote sensing data. *International Journal of remote sensing* **26**(7), 1477–1491 (2005)
6. 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 Sensing* **7**(11), 14680–14707 (2015)
7. Paisitkriangkrai, S., Shen, C., van den Hengel, A.: Pedestrian detection with spatially pooled features and structured ensemble learning. *IEEE transactions on pattern analysis and machine intelligence* **38**(6), 1243–1257 (2015)
8. Pesaresi, M., Gerhardinger, A.: Improved textural built-up presence index for automatic recognition of human settlements in arid regions with scattered vegetation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **4**(1), 16–26 (2010)
9. Rizvi, I.A., Mohan, B.K.: Object-based image analysis of high-resolution satellite images using modified cloud basis function neural network and probabilistic relaxation labeling process. *IEEE Transactions on Geoscience and Remote Sensing* **49**(12), 4815–4820 (2011)
10. Gaetano, R., Masi, G., Poggi, G., Verdoliva, L., Scarpa, G.: Marker-controlled watershed-based segmentation of multiresolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* **53**(6), 2987–3004 (2014)
11. Bosch, A., Munoz, X., Marti, R.: Which is the best way to

- organize/classify images by content? Image and vision computing **25**(6), 778–791 (2007)
12. Yang, Y., Newsam, S.: Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 270–279 (2010)
  13. Alkhawani, M., Elmogly, M., Elbakry, H.: Content-based image retrieval using local features descriptors and bag-of-visual words. International Journal of Advanced Computer Science and Applications **6**(9), 212–219 (2015)
  14. Fan, H., Wei, Q., Shu, D.-Q.: The method of extracting land classification information by hri-refinenet model. IEEE Access **8**, 599–610 (2019)
  15. Camps-Valls, G.: New machine-learning paradigm provides advantages for remote sensing. Spie Newsroom (2008)
  16. Tuia, D., Munoz-Mari, J., Gomez-Chova, L., Malo, J.: Graph matching for adaptation in remote sensing. IEEE Transactions on Geoscience and Remote Sensing **51**(1), 329–341 (2012)
  17. Lin, D., Fu, K., Wang, Y., Xu, G., Sun, X.: Marta gans: Unsupervised representation learning for remote sensing image classification. IEEE Geoscience and Remote Sensing Letters **14**(11), 2092–2096 (2017)
  18. Jiang, C., Zhang, H., Shen, H., Zhang, L.: Two-step sparse coding for the pan-sharpening of remote sensing images. IEEE journal of selected topics in applied earth observations and remote sensing **7**(5), 1792–1805 (2013)
  19. Yuan, H., Van Der Wiele, C.F., Khorram, S.: An automated artificial neural network system for land use/land cover classification from landsat tm imagery. Remote Sensing **1**(3), 243–265 (2009)
  20. Blanzieri, E., Melgani, F.: An adaptive svm nearest neighbor classifier for remotely sensed imagery. In: 2006 IEEE International Symposium on Geoscience and Remote Sensing, pp. 3931–3934 (2006). IEEE
  21. Dai, C.-g., Huang, X.-b., Dong, G.-j.: Support vector machine for classification of hyperspectral remote sensing imagery. In: Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), vol. 4, pp. 77–80 (2007). IEEE
  22. Zhang, L., Zhang, L., Du, B.: Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and Remote Sensing Magazine **4**(2), 22–40 (2016)
  23. Zhong, Z., Li, J., Cui, W., Jiang, H.: Fully convolutional networks for building and road extraction: Preliminary results. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1591–1594 (2016). IEEE
  24. Fu, G., Liu, C., Zhou, R., Sun, T., Zhang, Q.: Classification for high resolution remote sensing imagery using a fully convolutional network. Remote Sensing **9**(5), 498 (2017)
  25. Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P.: Convolutional neural networks for large-scale remote-sensing image classification. IEEE Transactions on geoscience and remote sensing **55**(2), 645–657 (2016)
  26. Hu, W., Huang, Y., Wei, L., Zhang, F., Li, H.: Deep convolutional neural networks for hyperspectral image classification. Journal of Sensors **2015** (2015)
  27. Makantasis, K., Karantzalos, K., Doulamis, A., Doulamis, N.: Deep supervised learning for hyperspectral data classification through convolutional neural networks. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4959–4962 (2015). IEEE
  28. Zhao, W., Du, S.: Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing **54**(8), 4544–4554 (2016)
  29. Paoletti, M.E., Haut, J.M., Plaza, J., Plaza, A.: A new deep convolutional neural network for fast hyperspectral image classification. ISPRS journal of photogrammetry and remote sensing **145**, 120–147 (2018)
  30. Romero, A., Gatta, C., Camps-Valls, G.: Unsupervised deep feature extraction for remote sensing image classification. IEEE Transactions on Geoscience and Remote Sensing **54**(3), 1349–1362 (2015)
  31. Mou, L., Ghamisi, P., Zhu, X.X.: Unsupervised spectral-spatial feature learning via deep residual conv-deconv network for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing **56**(1), 391–406 (2017)
  32. Mou, L., Ghamisi, P., Zhu, X.X.: Fully conv-deconv network for unsupervised spectral-spatial feature extraction of hyperspectral imagery via residual learning. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 5181–5184 (2017). IEEE
  33. Mou, L., Ghamisi, P., Zhu, X.X.: Deep recurrent neural networks for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing **55**(7), 3639–3655 (2017)
  34. Zhao, W., Du, S.: Learning multiscale and deep representations for classifying remotely sensed imagery. ISPRS Journal of Photogrammetry and Remote Sensing **113**, 155–165 (2016)
  35. Lin, G., Milan, A., Shen, C., Reid, I.: Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1925–1934 (2017)
  36. Lang, S.: Object-based image analysis for remote sensing applications: modeling reality – dealing with complexity. Springer Berlin Heidelberg (2008)
  37. Blaschke, T., Lang, S., Hay, G.J.: Geographic object-based image analysis (geobia): A new name for a new discipline. Springer Berlin Heidelberg **10.1007/978-3-540-77058-9**(Chapter 4), 75–89 (2008)
  38. Liu, W., Tao, D.: Multiview hessian regularization for image annotation. IEEE Transactions on Image Processing **22**(7), 2676–2687 (2013). doi:10.1109/TIP.2013.2255302
  39. Liu, W., Tao, D., Cheng, J., Tang, Y.: Multiview hessian discriminative sparse coding for image annotation. Computer Vision & Image Understanding **118**(jan.), 50–60 (2014)
  40. Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U.: Classification with an edge: Improving semantic image segmentation with boundary detection. Isprs Journal of Photogrammetry & Remote Sensing **135** (2017)
  41. Luo, F., Du, B., Zhang, L., Zhang, L., Tao, D.: Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image. IEEE Transactions on Cybernetics **49**(7), 2406–2419 (2019)
  42. Chen, Y., Lin, Z., Xing, Z., Gang, W., Gu, Y.: Deep learning-based classification of hyperspectral data. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing **7**(6), 2094–2107 (2017)

43. Chen, Y., Zhao, X., Jia: Spectral-spatial classification of hyperspectral data based on deep belief network. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* **8**(6), 2381–2392 (2015)
44. Tao, C., Pan, H., Li, Y., Zou, Z.: Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. *IEEE Geoscience & Remote Sensing Letters* **12**(12), 2438–2442 (2015)
45. Yuan, J., Wang, D.L., Li, R.: Remote sensing image segmentation by combining spectral and texture features. *IEEE Transactions on Geoscience and Remote Sensing* (2013)
46. Yao, T., Chang, L., Deng, Z., Liu, X., Liu, J.: Adaptive Ensemble Clustering for Image Segmentation in Remote Sensing. Springer, ??? (2017)
47. Liang, S., Huang, X., Fan, C.: Double-group particle swarm optimization and its application in remote sensing image segmentation. *Sensors* **18**(5), 1393 (2018)
48. Sziranyi, T., Shadaydeh, M.: Segmentation of remote sensing images using similarity-measure-based fusion-mrf model. *IEEE Geoscience & Remote Sensing Letters* **11**(9), 1544–1548 (2014)
49. Xun, G., Xian, S., Zhang, Y., Yan, M., Xu, G., Hao, S., Jiao, J., Fu, K.: An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network. *IEEE Access* **6**, 1–1 (2018)
50. Zhang, X., Zhu, K., Chen, G., Tan, X., Gong, Y.: Geospatial object detection on high resolution remote sensing imagery based on double multi-scale feature pyramid network. *Remote Sensing* **11**(7), 755 (2019)
51. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
52. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **25**, 1097–1105 (2012)
53. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
54. Uzket, B., Yeh, C., Ermon, S.: Efficient object detection in large images using deep reinforcement learning. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1824–1833 (2020)
55. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: *European Conference on Computer Vision*, pp. 740–755 (2014). Springer
56. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* **28**, 91–99 (2015)
57. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263–7271 (2017)
58. Kristan, M., Leonardis, A., Matas, J., Felsberg, M., Pflugfelder, R., Cehovin Zajc, L., Vojir, T., Hager, G., Lukežić, A., Eldesokey, A., et al.: The visual object tracking vot2017 challenge results. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 1949–1972 (2017)
59. Bertinetto, L., Valmadre, J., Henriques, J.F., Vedaldi, A., Torr, P.H.: Fully-convolutional siamese networks for object tracking. In: *European Conference on Computer Vision*, pp. 850–865 (2016). Springer
60. Uzket, B., Rangnekar, A., Hoffman, M.: Aerial vehicle tracking by adaptive fusion of hyperspectral likelihood maps. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 39–48 (2017)
61. Uzket, B., Seo, Y.: Enkcf: Ensemble of kernelized correlation filters for high-speed object tracking. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1133–1141 (2018). IEEE
62. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation (2021). 2102.04306
63. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: Labelme: a database and web-based tool for image annotation. *International journal of computer vision* **77**(1–3), 157–173 (2008)
64. Dumitru, C.O., Schwarz, G., Datcu, M.: Land cover semantic annotation derived from high-resolution sar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **9**(6), 2215–2232 (2016)
65. Dai, J., Li, Y., He, K., Sun, J.: R-fcn: Object detection via region-based fully convolutional networks. In: *Advances in Neural Information Processing Systems*, pp. 379–387 (2016)
66. The Canada Centre for Mapping and Earth Observation. <https://github.com/NRCan/geo-deep-learning>