Deep Extraction of Cropland Parcels from Very High-Resolution Remotely Sensed Imagery

Liegang Xia
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
xialg@zjut.edu.cn

Jiancheng Luo
Institute of Remote Sensing and Digital Earth
Chinese Academy of Science
Beijing, China
luojc@radi.ac.cn

Yingwei Sun
Institute of Remote Sensing and Digital Earth
Chinese Academy of Science
Beijing, China
wfsunyingwei@yeah.net

Haiping Yang
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
yanghp@zjut.edu.cn

Abstract— Extracting cropland parcels from high resolution remote sensing images is a basic task for precision agriculture and other fields. Object based image analysis rely heavily on segmentation methods and can't satisfy the parcels' requisition in most situation. Inspired by the recent remarkable improvement on image understanding with deep learning, we propose a deep-edge guided method for cropland parcels extraction. Focus on the boundaries of these parcels, hard edge and soft edge are extracted respectively with U-Net and RCF model. Then all edges with the land type of cropland are constructed into parcels. At last accurate cropland-parcels are achieved.

Keywords—cropland parcels, high-resolution remote sensing, deep learning, semantic segmentation

I. Introduction

In southern China, which is comprised mostly of smallholder farms, fragmentary cropland and complicated farming system make it is hard to accurately statistic agricultural fields or rational plan the agricultural production[1, 2]. With the development of very high resolution (VHR) remote sensing, it is possible to monitor cropland by parcels, i.e. farms with agricultural fields of $\leqslant 30 \times 30$ m in dimension[3]. That is of great significance for precision agricultural, farmland protection, and land use and land cover change(LUCC)[4]. However, few studies have involved the extraction of cropland parcels in VHR images. Image segmentation and classification are two important methods to get cropland parcels from VHR images automatically[5]. In the framework of Object Based Image Analysis (OBIA)[6, 7], researchers focus on improving segmentation or objects proposal[8, 9], designing the most effective features and selecting classification method[10], combining multi-source information[11]. However, limited by the various interferences in the image (e.g. atmospheric conditions, illumination conditions, and soil moistures), the current extraction accuracy are still unsatisfactory. The more complex and fragmental of the farmland, the higher of the accuracy need to achieve. The contradiction between detail parcels (e.g. <1,000 square meters) and high precision is hard to balance in traditional framework.

Recently deep learning methods have dramatically improved the state-of-the-art in visual object recognition, image understanding[12]. This data-driven method allows the machine to automatically learn and mine features of interested objects from datasets[13]. In light of these methods, edge or object detection, semantic segmentation, scene understanding in VHR images are significantly improved[14]. These inspired us to reconsider the extraction task of high resolution remote sensing without OBIA.

In this study, we focus on cropland parcel itself, and divide the cropland boundaries into hard edges and soft edges. The former ones are stable and always covered by road or river (Fig. 1-a), always partition the big field farmland. The latter ones are changeful and thin so that are presented as lines in images, always partition different types of cropland (Fig. 1-b). By rational use of deep learning methods, we propose the deep extraction method for cropland parcels. By accurate positioning both edges and semantic recognize cropland areas, every parcel can be extracted from VHR images automatically. This shows great potential for large scale and precise agricultural application.

(a) Hard Edge (b) Soft Edge

Fig.1 Sample edges of cropland. (a) Hard edge are stable and region-like, (b) Soft edges are changeful and line-like. First row are field photos and second row are VHR images.

Our contributions include the following: (1) By considering the two types of boundaries of cropland parcels separately, different extraction methods can be used in the whole

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB0503600, in part by the National Natural Science Foundation of China under Grant 41701472 and Grant 41601437.

978-1-5386-5038-7/18/$31.00 ©2018 IEEE

extraction framework. (2) Deep edge which comparable to manual delineation are extracted for geometric construction of parcels. (3) Semantic segmentation methods are coupled for accurately discriminating the farmland from other land types.

The remainder of this paper is organized as follows. Section II introduces the proposed method in detail, including deep training, predicting and post-processing. Section III is the experiment results and discussion. Section IV provides a summary of the proposed method.

II. METHOD

Our method can be divided into three major stages(Fig.2). The first stage is sample preparing and training. The second stage is edge and cropland predicting. The last stage is post-processing for complete cropland parcels.

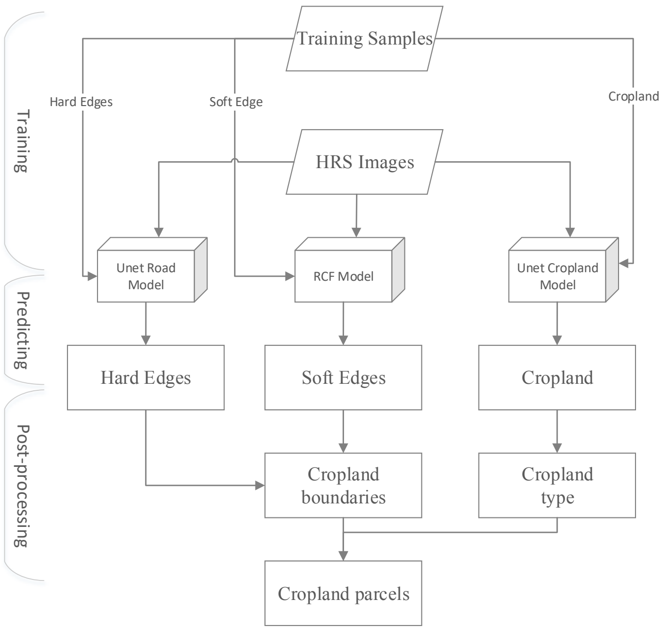

Fig.2 Flowchart for deep extraction of cropland parcels

A. *Training*

1) *Sample prepared*

In the training process, every HRS image have three corresponding labels: hard edges, soft edges and cropland area. Suppose X_i is one input image, the first hard edge labels is a channel with value 1 or 0 which denotes the pixel is or is not hard edge(specially the land type is road, river). The second soft edge labels is a channel with value 1 or 0 which denotes the pixel is or is not the edge of any cropland parcel. The last cropland label denotes all pixels with the land type of cropland. As shown in Fig.3, these labels has different proportion so it's hard to train them in one task. Soft edges label is line-like while hard edge and cropland label are region-like. As shown in Fig.2, we train three tasks with U-Net and RCF models.

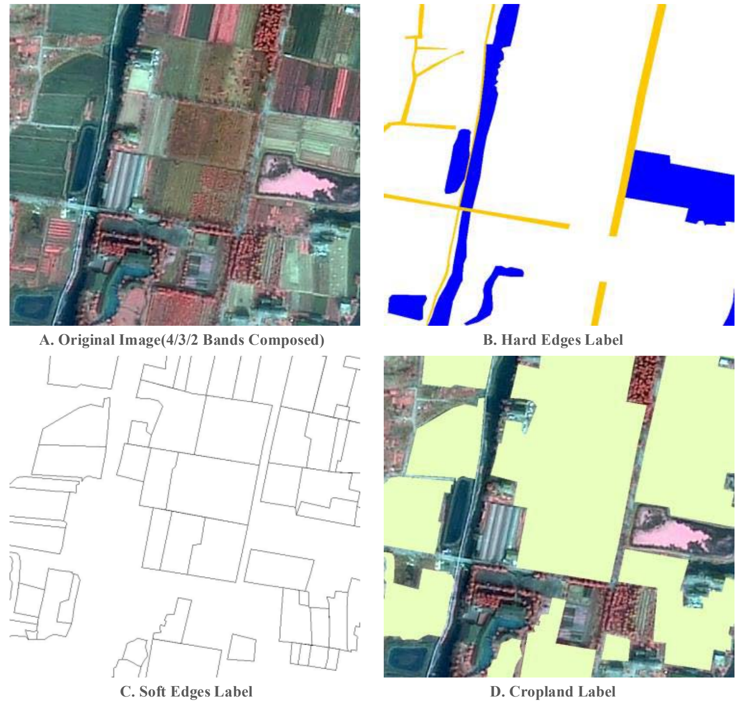

Fig.3. Training samples

2) *Deep edge Models*

RCF model was used to detect soft edges[15]. Improved from HED model[16], five stage of convolution layer groups were combined together to exactly extract the scale-invariant edge features. In every stage loss Y_{side} could be computed by comparing predicted edge and second channel of Y. All stages' outputs were weighted fused with

$$Y_{fuse} = \sum_{m=1}^{5} h_m Y_{side}^{m} \tag{1}$$

where h_m is the fuse weight of each scale which also learned from network.

While useful information captured by each convolution layer becomes coarser with its receptive field size increasing. The whole network can learn multiscale, including low-level and object-level, information that is helpful to edge detection. The fusion loss L_{fuse} can be computed by compare with label and the training objective is minimizing the following function via standard (back-propagation) stochastic gradient descent:

$$(W,h)^{*} = \operatorname{argmin}\left(L_{fuse} + \sum_{m=1}^{5} \alpha_m L_{side}^{m}\right) \tag{2}$$

where α_m is loss weight of each scale, W and h are the training results.

In our application edges of every cropland parcel was needed. Theoretically we could detect all edges with this model. As shown in Figure 5-c, all edges were extracted. But in VHR image cropland may be confused with adjacent parcels. A small edge error may result corresponding parcels corrupted, especially near the hard edges. Thus hard edges were significant for improving the overall accuracy. To those region-like edges, semantic segmentation is a better choice.

3) *Semantic Segmentation Model*

U-Net model was used to detect region-like hard edges and cropland types[17]. The design of u-shaped architecture allows the network to propagate context information to higher resolution layers (Fig. 4). Thus we may get a more precise output. In our application binary classification was needed. So we change the final layer' feature vector from number of classes into one channel. The energy function is computed by a pixel-wise soft-max over the final feature map combined with the jaccard loss function. At last the model can predict the probability of each pixel is our preferred class.

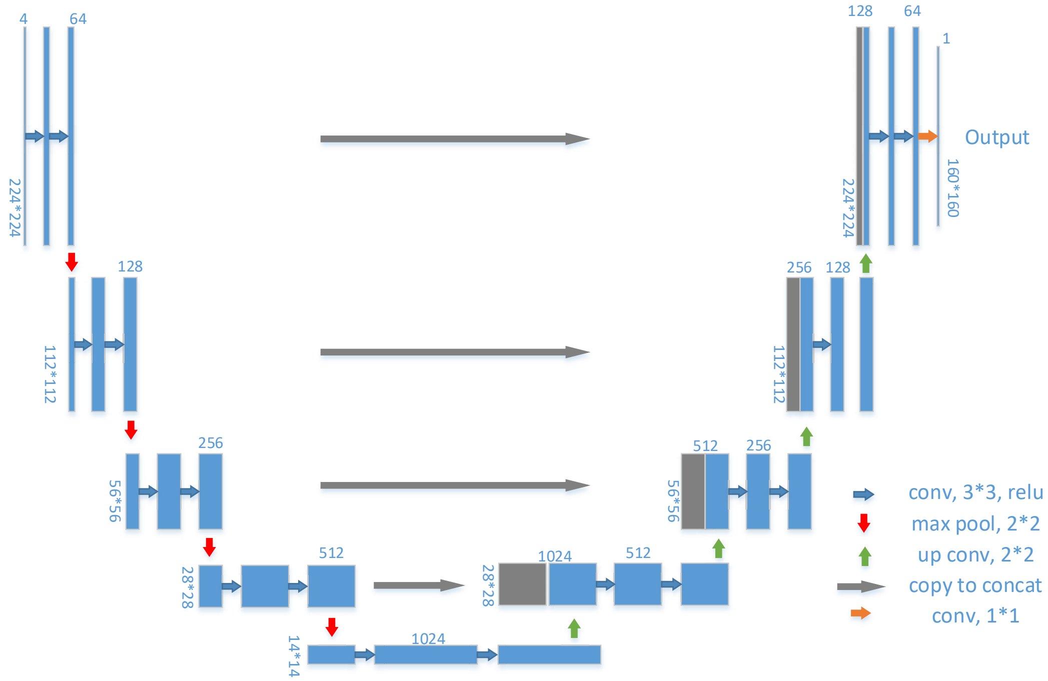

Fig. 4. U-net architecture. Each blue box corresponds to a H*W*C feature map. C is denoted on top of the box. H*W is provided at the lower left edge of the box. Gray boxes represent copied feature maps for concat. The arrows denote the different operations.

III. Results And Discussion

A. Data Used

The proposed method was tested in Zongyang, Anhui Province of East China. The images were acquired from GF-2 satellite with a pansharpened resolution of 0.8m.

B. Results

With the methods described in Section 2, we extracted cropland parcels from GF-2 images. A part of results is shown in Fig. 5. Because of the heterogeneous and fragmental landscape, about half a million parcels were extracted in this county with area of 1,808 square kilometres.

C. Comparison

The Traditional OBIA method first segment the image, then design kinds of cropland features on object, at last distinguish each object is or is not cropland. Take eCognition as example, the first step of segmentation is so critical that the last parcels depends on these objects[18]. Fig. 6 shows the quantitative comparison of eCognition's and our result. Here we set region IoU to evaluate the availability of each extracted cropland parcel.

As a supervised method, the higher IoU means it match the reference parcel more (Eq. 3). To evaluate the overall accuracy we first calculate IoU corresponding to ground truth for each detected parcel, and then performs statistics according to different IoU threshold. Obviously the higher the threshold set, the less number or area of parcels reserved (Eq. 4). At last, mean IoU counts parcels with IoU bigger than 0.5.

$$\text{IoU} = \frac{Gt \cap Dp}{Gt \cup Dp} \tag{3}$$

$$\text{recall} = \frac{Area_{IoU>thresh}}{Area_{crop}} \tag{4}$$

where Dp denotes detected parcels, Gt denotes the ground truth, and thresh denotes the IoU threshold.

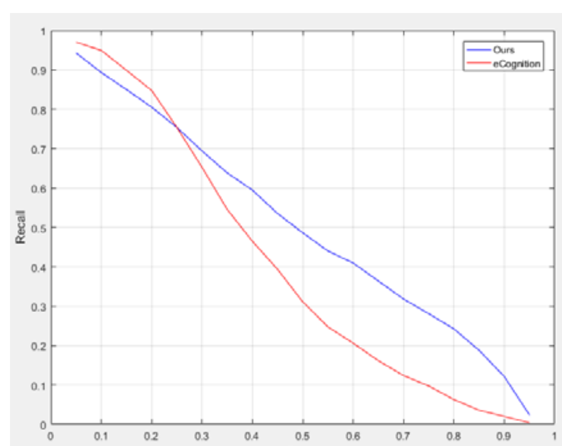

Fig. 6. Comparison of recall in different IoU threshold (here higher threshold indicates more precise parcels and higher recall indicates larger area of parcels truly detected)

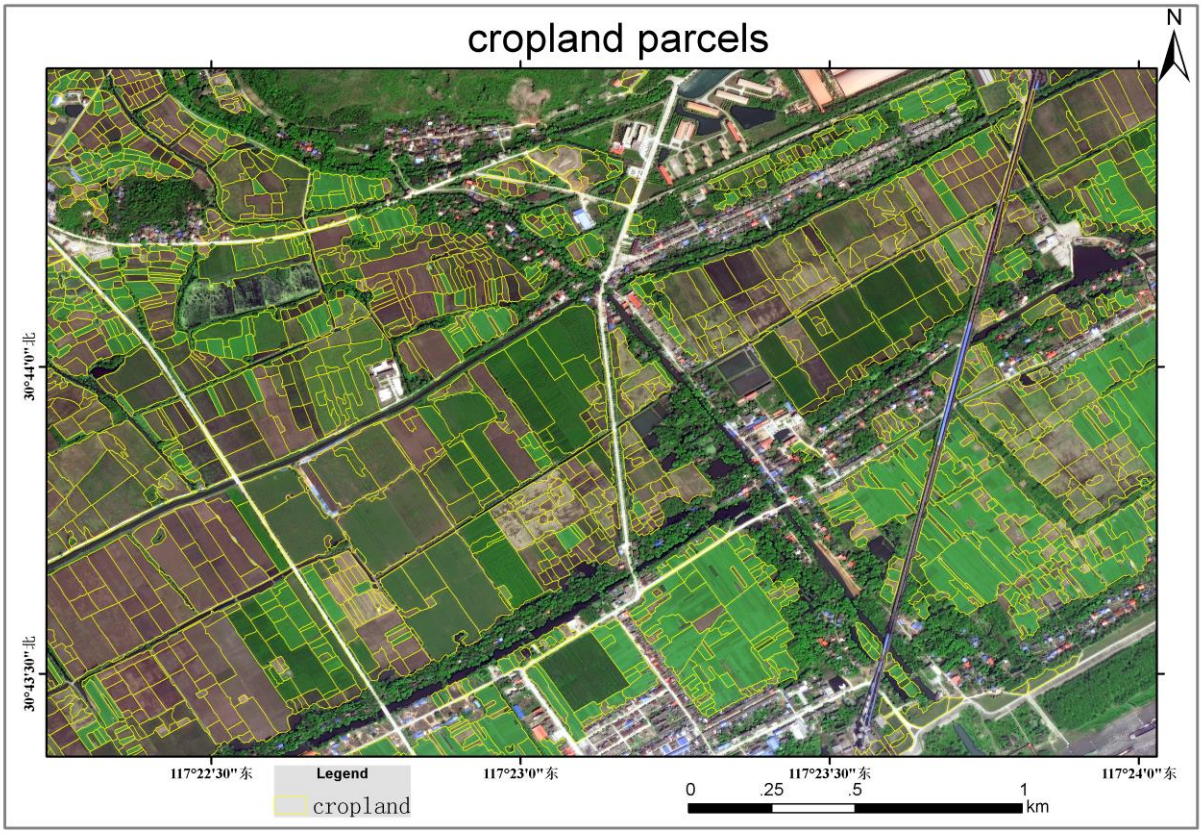

Fig.5 result of cropland parcels

D. Discussion

In our framework, edge features learned from natural images can be adapted to VHR images. Semantic features learned from this village can also be adapted to other villages of same county. This transfer learning ability make it's possible to extract accurate parcels in large scale area which is exactly the desire of modern agriculture application. As mentioned in 2.2, we did not train soft edge model on our dataset and the accuracy was acceptable. We can train a more adapted model based on our samples to achieve better soft edges. But on the other hand this model may not so adapted to other region. Although the current semantic models have a good generalized performance, but to remote sensing images with regional heterogeneity that is not enough. So larger area and more detailed experiments may help to understand and balance this contradiction better.

IV. Conclusion

In order to accurately map smallholder cropped area over a large region, we propose this deep extraction method. With hard edges and soft edges extracted respectively we shaped the parcels accurately. With semantic segmentation we further filter out the cropland parcels. This deep learning based framework showed great potential to balance the large-scale and fragmental demands in agricultural remote sensing. We perform IoU-recall curve to address how much detected parcels could be used in GIS map. The application of the method on Zongyang County shows great advantage when we compare with traditional segmentation-classification methods. As the development of deep learning methods and larger region test, the framework can produce more precise cropland parcels in more efficient way.

Acknowledgment

The authors thank the Imagesky Corp. in Suzhou for supporting the GF-2 data and samples.

References

[1] N. Xia et al., "Demarcation of Prime Farmland Protection Areas around a Metropolis Based on High-Resolution Satellite Imagery," Scientific Reports, Article vol. 6, p. 37634, 12/21/online 2016.

[2] H. Yin, A. V. Prishchepov, T. Kuemmerle, B. Bleyhl, J. Buchner, and V. C. Radeloff, "Mapping agricultural land abandonment from spatial and temporal

segmentation of Landsat time series," Remote Sensing of Environment, vol. 210, pp. 12-24, 2018/06/01/ 2018.
[3] J. L. McCarty, C. S. R. Neigh, M. L. Carroll, and M. R. Wooten, "Extracting smallholder cropped area in Tigray, Ethiopia with wall-to-wall sub-meter WorldView and moderate resolution Landsat 8 imagery," Remote Sensing of Environment, vol. 202, pp. 142-151, 2017/12/01/ 2017.
[4] P. He, J. Li, and X. Wang, "Wheat harvest schedule model for agricultural machinery cooperatives considering fragmental farmlands," Computers and Electronics in Agriculture, vol. 145, pp. 226-234, 2018/02/01/ 2018.
[5] T. Sohl, J. Dornbierer, S. Wika, K. Sayler, and R. Quenzer, "Parcels versus pixels: modeling agricultural land use across broad geographic regions using parcel-based field boundaries," Journal of Land Use Science, vol. 12, no. 4, pp. 197-217, 2017.
[6] T. Blaschke, "Object based image analysis for remote sensing," ISPRS journal of photogrammetry and remote sensing, vol. 65, no. 1, pp. 2-16, 2010.
[7] M. Belgiu and O. Csillik, "Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis," Remote sensing of environment, vol. 204, pp. 509-523, 2018.
[8] R. Gaetano, G. Masi, G. Poggi, L. Verdoliva, and G. Scarpa, "Marker-controlled watershed-based segmentation of multiresolution remote sensing images," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 6, pp. 2987-3004, 2015.
[9] J. Pont-Tuset, P. Arbeláez, J. T. Barron, F. Marques, and J. Malik, "Multiscale Combinatorial Grouping for Image Segmentation and Object Proposal Generation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 1, pp. 128-140, 2017.
[10] D. G. Goodin, K. L. Anibas, and M. Bezymennyi, "Mapping land cover and land use from object-based classification: an example from a complex agricultural landscape," International Journal of Remote Sensing, vol. 36, no. 18, pp. 4702-4723, 2015/09/17 2015.
[11] T. Esch, A. Metz, M. Marconcini, and M. Keil, "Combined use of multi-seasonal high and medium resolution satellite imagery for parcel-related mapping of cropland and grassland," International Journal of Applied Earth Observation and Geoinformation, vol. 28, pp. 230-237, 2014/05/01/ 2014.
[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-44, May 28 2015.
[13] A. Sharma, X. Liu, X. Yang, and D. Shi, "A patch-based convolutional neural network for remote sensing image classification," Neural Networks, vol. 95, pp. 19-28, 2017/11/01/ 2017.
[14] Y. Xu, L. Wu, Z. Xie, and Z. Chen, "Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters," Remote Sensing, vol. 10, no. 1, p. 144, 2018.
[15] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, and X. Bai, "Richer convolutional features for edge detection," in Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, 2017, pp. 5872-5881: IEEE.
[16] S. Xie and Z. Tu, "Holistically-nested edge detection," International Journal of Computer Vision, pp. 1-16, 2017.
[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, 2015, pp. 234-241: Springer.
[18] H. Costa, G. M. Foody, and D. S. Boyd, "Supervised methods of image segmentation accuracy assessment in land cover mapping," Remote Sensing of Environment, vol. 205, pp. 338-351, 2018.