Training SegNet for Cropland Classification of High Resolution Remote Sensing Images

Zhenrong Du
China Agricultural University
No. 17 Tsing Hua East Road, HaiDian District
Beijing, P. R. China
Iris_dudu@126.com

Jianyu Yang
China Agricultural University
No. 17 Tsing Hua East Road, HaiDian District
Beijing, P. R. China
ycjyyang@cau.edu.cn

Weiming Huang
GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University
Sölvegatan 12, 223 62
Lund, Sweden
weiming.huang@nateko.lu.se

Cong Ou
China Agricultural University
No. 17 Tsing Hua East Road, Haidian District
Beijing, P.R.China
oucong@cau.edu.cn

Abstract

China's rapid urbanization entails increasing importance of cropland protection, and automatic mapping through information extraction from high resolution remote-sensing (RS) images is a powerful tool towards this goal. RS images information extraction pertains the feature classification, which is a long-standing research issue in the RS community. The emerging deep learning technique, which is an effective method to automatically discover relevant contextual features, is a promising means for better image classification. In this study, we exploit the deep learning technique to classify and extract cropland in high resolution RS images. Specifically, we use the deep learning framework Caffe to construct a platform for sampling, training, testing and classifying to extract and map cropland based on SegNet. Leveraging the overlapped sampling technique proposed in this paper, we obtain more training samples with limited labeled data and achieve better training results. The results manifest that the proposed approach can efficiently obtain acceptable accuracy (OA = 0.98, Kappa = 0.93) in the study of cropland classification of the study area, and the approach performs better in urban areas where trees or bush could easily be misclassified as cropland. Furthermore, the proposed approach is highly scalable for the variety of crop types in cropland. Overall, the proposed approach can train a precise and effective model that is capable of adequately describing the small, irregular fields of smallholder agriculture and handling the great level of detail in high-resolution image.

Keywords: Deep learning, RS images classification, Cropland, Overlapped Sampling

1 Introduction

Mapping cropland from remote-sensing (RS) images is an effective measure to protecting cropland in the rapidly urbanized era in China. Recent technologies have significantly increased the resolution of available RS images (2m spatial resolution and higher), which provides the necessary detail to observe smallholder agriculture. Nonetheless, it also brings challenges to RS community in smart image interpretation for cropland. There is a vast literature on setting the thresholds to map cropland automatically, but the approaches usually consider the spectrum of every individual pixel to assign it to a certain class. Alternatively, more advanced techniques combine information from a few neighboring pixels to enhance the mapping performance, often referred to as spectral-spatial classification. These approaches rely on the separability of the different classes based on the spectrum of a single pixel or of some neighboring pixels. However, smallholder agricultural fields are small and irregularly shaped, and all these methods only consider the spectral features of the pixel and its neighborhoods. So we argue that a more thorough understanding of the context, such as the shape of objects, is required to aid the mapping process.

Therefore, Convolutional neural networks (CNN) (Lecun, Bottou, Bengio, & Haffner, 1998) are gaining attention due to their capability to automatically discover relevant contextual features in classification problems. CNNs consist of a stack of learned convolution filters that extract hierarchical contextual image features, and are a popular form of deep learning networks. They have already outperformed other approaches in various domains, such as digit recognition (Schmidhuber, Meier, & Ciresan, 2012) and natural image categorization (Krizhevsky, Sutskever, & Hinton, 2012).

Among different CNN models, SegNet is designed to be an efficient architecture for pixel-wise semantic segmentation (Badrinarayanan, Kendall, & Cipolla, 2017). In this study, we utilize SegNet architecture since it provides a good balance between accuracy and computational costs. Moreover, an overlapped sampling method is proposed to expand the training dataset. Using SegNet and overlapped sampling, this study develops a methodology to finish pixel-wise semantic segmentation and map the cropland from RS images automatically.

1.1 Related Work

In this section, we review mapping cropland methods and the use of CNNs in semantic segmentation of RS data.

Before the emergence of deep networks, the best performing methods for mapping cropland mostly relied on hand engineered features classifying pixels independently. Typically, a patch is fed into a classifier e.g. Support Vector Machine (Yang, Everitt, & Murden, 2011), decision tree(Otukei & Blaschke, 2010; Xiong et al., 2017) or artificial neural network (Tseng, Chen, Hwang, & Shen, 2008) to predict the class probabilities of the center pixel. However, these methods need to set thresholds and feature design artificially. Besides, they are often influenced by the mixed pixel problem (Nagendra & Rocchini, 2008) and limited to spectral features, lacking a more thorough understanding of the context, which inspired us to use CNNs to overcome the problem.

CNNs, which learn the representative and discriminative features in a hierarchical manner from the data, have recently become a hot research topic in the machine-learning area and have been introduced into the geoscience and RS community for RS classifications (Zhang, Zhang, & Du, 2016). Zhou et al. (2017) employed CNN architecture as a deep feature extractor for high-resolution RS image retrieval (HRRSIR). Lagkvist et al. (2016) presented a novel RS imagery classification method based on CNNs for five classes (vegetation, ground, road, building, and water), which outperformed the existing classification approaches in the classification accuracy. Wang et al. (2015) used a CNN structure with three layers and Finite State Machine (FSM) for road network extraction for long-term path planning. Different network architectures, which are widely used in the field of computer science, are also compared when they are used in semantic segmentation of RS data (Scott, England, Starms, Marcum, & Davis, 2017). To accelerate the training stage, large pre-trained neural networks (Marmanis, Datcu, Esch, & Stilla, 2016) have been investigated for classifying RS images into a large set of diverse land-use classes, and have achieved promising results, significantly increasing the best stated performance through a simple and computationally efficient end-to-end approach.

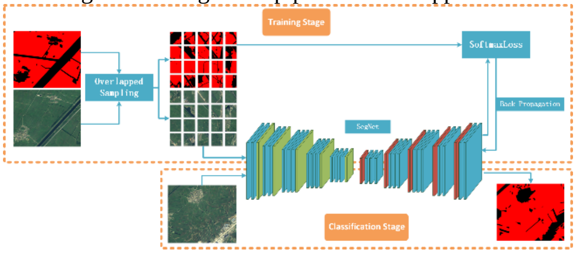
1.2 Contributions

Our contributions in this work lie on the use of SegNet for RS images cropland mapping, where we show that SegNet classifiers can be profoundly suitable for our mapping task. To the best of our knowledge, this is the first time to apply SegNet into cropland classification. Furthermore, in the training stage we propose an overlapped sampling method to get more training samples with limited labeled data. This produces better performances in classification accuracies compared with other methods, especially when there are different crop types in crop fields. Moreover, the approach shows its superiority in urban areas where trees or bush can easily be misclassified as cropland.

2 Proposed Method

Alike other supervised classification, our approach generally has two stages (Figure 1): the training stage and the classification stage. In the training stage, image-label pairs, with pixel-class correspondence, are input into the SegNet network as training samples. The error between predicted class labels and ground truth (GT) labels is calculated and back-propagated through the network using the chain rule, and then the parameters of the SegNet network are updated using the gradient descent method. In the classification stage, the trained SegNet network is performed on an input image to generate a class prediction. The details of the training stage and classification stage are presented in Sections2.2 and 2.3, respectively.

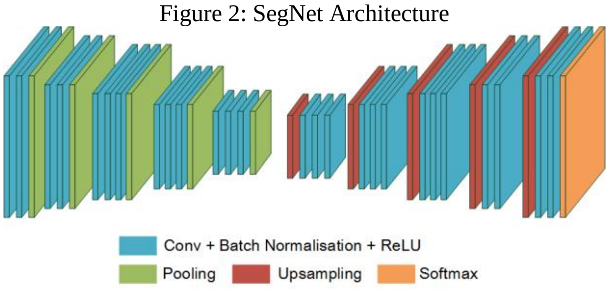
Figure 1: The general pipeline of our approach.



2.1 Network Architecture

A number of network architectures for semantic segmentation have been proposed in the last few years, e.g., FCN (Long, Shelhamer, & Darrell, 2015), DeepLab (Chen, Papandreou, Kokkinos, Murphy, & Yuille, 2017). In this study, we choose the SegNet architecture (Figure 2), which is designed to be an efficient architecture for pixel-wise semantic segmentation. It is primarily motivated by road scene understanding applications which require the capabilities of modelling appearance, shape and understanding the spatial-relationship (context) between different classes. At the same time, it provides a good balance between accuracy and computational cost. Moreover, SegNet's symmetrical architecture and its use of the pooling/upsampling combination is very effective for precise re-localisation of features, which is intuitively crucial for RS data.

Figure 2: SegNet Architecture



2.2 Network Training

Unlike the traditional computer vision images, RS images often have larger coverage and size, which makes it difficult to be trained as a whole. Therefore, before training, we need to split the labeled RS images into small parts. As RS images labeled with ground truth are limited, we propose an overlapped sampling method (Figure 4) rather than the general sampling procedure (Figure 3) to expand the training dataset. The motivation for this approach is as follows: image augmentation, such as random rotation, shifts, shear, flips, and so forth, is usually applied to boost the performance of deep networks. For RS images, taking advantage of the idea of shift and shear for image augmentation, it is convenient to get more samples by overlapped sampling on an integral image. The expanded training dataset, which is organized by Image-Ground truth (GT) label pairs, is then input into SegNet as training samples. The Softmax function is performed on the output feature map generated by the network to predict the class distribution. Then the softmax loss is calculated and back-propagated, and finally the network parameters are updated using Stochastic Gradient Descent (SGD) with momentum.

Figure 3: General Sampling Procedure (Here we take RS images as an example, GT-labels' sampling method is the same.)

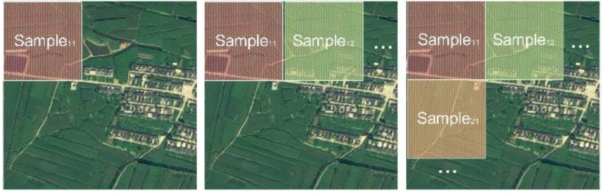
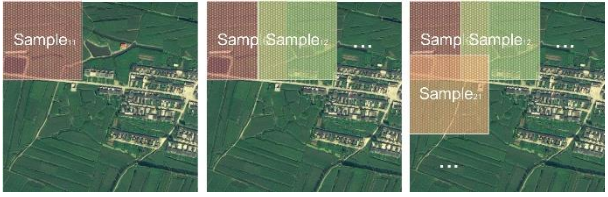


Figure 4: Overlapped Sampling Method (Here we take RS images as an example, GT-labels' sampling method is the same.)



2.3 Classification Using the Trained Network

High resolution RS images are often too large to be processed in only one pass through a CNN. Given current GPU memory limitations, we split our images into small patches with a simple sliding window. It is then possible to process arbitrary large images in a linear time. For the overlapped part of the image predicted, we average the multiple predictions to obtain the final classification for overlapping pixels. This smoothes the predictions along the borders of each patch and removes potential discontinuities.

3 Experiments

3.1 Experimental Setup

Our training dataset is collected from WorldView-1 (true color fusion images with 0.5 meter resolution) of Fengnan, Hebei, China. The images were taken on 20 July 2011. The reason of choosing images obtained in July is that the crop is in its growing season, which makes it easier to extract crop fields from other sorts of land cover via semantic segmentation. We manually labeled two slices (the size of both is 4341 * 3669) of the whole image at the pixel level as GT label data. One of the two slices is used for sampling, and the other for deploying. In our training dataset, there are a total of 203,000 pairs of samples (size 128 * 128). So for each pixel in RS images, there exists a pixel-class correspondence with it. We used 200,000 images for training, and the remaining 3,000 images for testing. A "step" policy is used for learning rate adjustment (gamma = 0.1, step_size = 20, 000). The max iteration in our training is 200,000. In the training procedure, we feed the samples into the network in batches, and each batch contains 10 images. In addition, we use the deep learning framework *Caffe* to construct a platform for all the work including sampling, training, testing and classifying to extract and map cropland.

Accuracy assessment is based on the pixel-based classification evaluation method. By obtaining the final mapping result and calculating the confusion matrix, we can obtain the overall accuracy (OA), kappa coefficient and F1-Score for each class. Additionally, for comparison with the proposed method, we use three different methods, i.e., artificial neural networks (ANN), fully convolutional networks (FCN) (Mnih, 2013) with overlapped samples, SegNet without overlapped samples, trained with the same training samples to map crop fields.

3.2 Results

We adopt our trained model on the other slice of the high resolution RS images for the classification. The image is the testing image that is not involved in training. Figure 5 is the illustration of the results and the comparison.

We employ overall accuracy, F1-Score, and Kappa coefficient as the indicators to evaluate our approach. These indexes are calculated from the confusion matrix C, where the overall accuracy is calculated as

$$\sum_i C_{ii} \, / \sum_i \sum_j C_{ij} \qquad (1)$$

where *i*, *j* represent the row and column number of the confusion matrix, respectively. Overall accuracy denotes the proportion of the pixels that are correctly classified, and the F1-Score is computed as

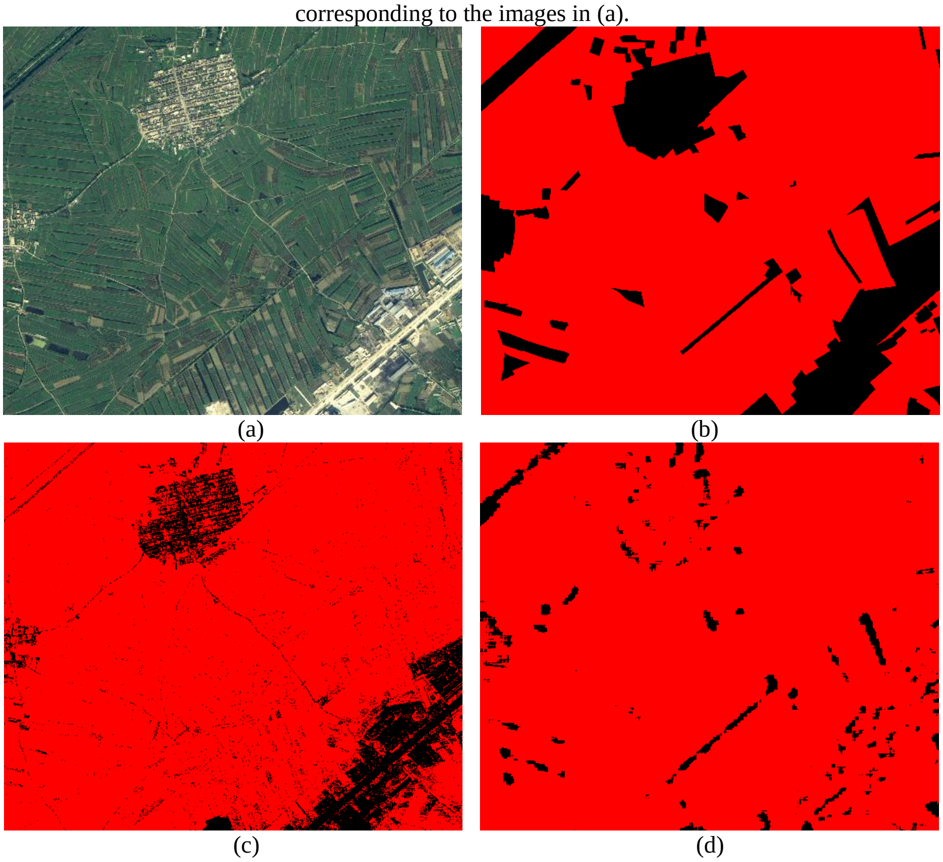
$$F_1 = 2 * \frac{precision*recall}{precision+recall} \qquad (2)$$

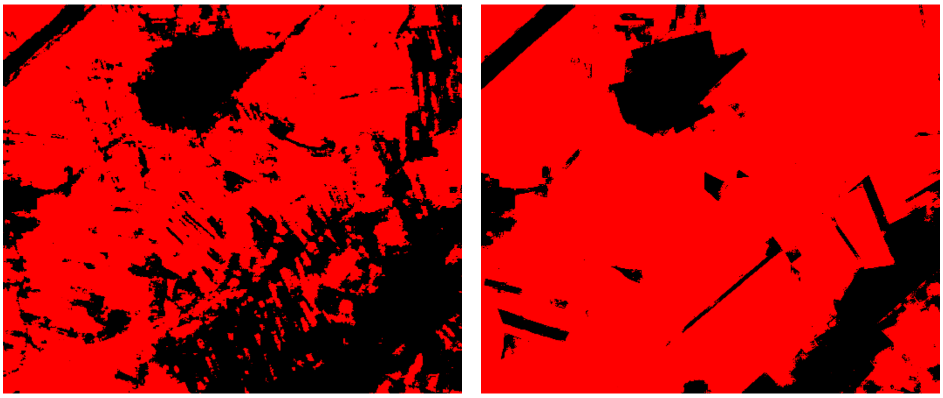
where precision is the number of correct positive results divided by the number of all positive results returned by the classifier, and recall is the number of correct positive results divided by the number of all relevant samples, that represents the harmonic average of the precision and recall, and the Kappa coefficient measures the consistency of the predicted classes with the GT classes. The comparisons between our approach and other three methods are listed in Table1.

Table 1: Comparison between approaches using artificial neural networks, FCN, SegNet without overlapped samples, and our approach.

Index	ANN	FCN	SegNet without Overlapped Samples	Our Approach
Overall Accuracy	87.73%	73.57%	74.83%	97.85%
F1-Score (Cropland)	0.93	0.83	0.82	0.99
F1-Score (Non-Cropland)	0.57	0.44	0.55	0.94
Kappa Coefficient	0.50	0.28	0.41	0.93

Figure 5: Classification result where cropland is red and non-cropland is black. (a) Original images; (b) GT labels corresponding to the images in (a); (c–e) Results of the artificial neural networks classification, FCN with overlapped samples, and SegNet without overlapped samples corresponding to the images in (a), respectively; (f) Our classification results corresponding to the images in (a).





(e)

(f)

3.3 Analysis

The statistics in Table 1 show our approach obtains the best performance compared with the others. Approaches using FCN and SegNet without overlapped samples achieve similar overall accuracy, and the ANN approach performs better than those two.

For ANN approach, Figure 6 shows that the result is not satisfactory in urban areas where trees or bush could easily be misclassified as cropland. And for FCN approach, serious reduction of the resolution is result from pooling operations. The output has lost many valuable detail information. When using SegNet without overlapped samples, the F1-Score for “non-cropland” is 0.44. That means more than half of the pixels are wrongly classified. It is easy to see in the Figure 7 that the low accuracy is caused by different crop types in crop fields. Compared with other approaches, our approach, which takes advantage of SegNet and overlapped sampling in the training stage, outperforms them in terms of accurateness, detail preserving and scalability. Therefore, the classification accuracy is highly improved.

Figure 6: Result produced by ANN. Yellow squares in the figure mark out the apparent mistakes.

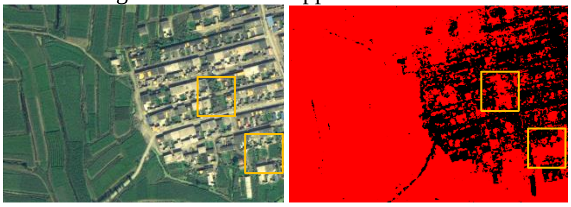
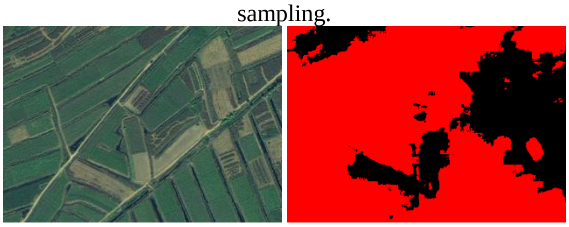


Figure 7: Result produced by SegNet without overlapped sampling.



4 Conclusion and Future Work

In this paper, we have proposed an overlapped sampling method, and utilized SegNet for mapping cropland in RS images. Through our proposed framework, we have achieved promising results using a simple and computationally efficient end-to-end approach.

In our future research, we will study the classification in a more detailed way, e.g., classification of land cover or crop types. Moreover, we will also investigate the potential of deep networks on a larger scale experiment, incorporating satellite data with greater spectral resolution and geographical variations.

References

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence, PP(99), 1-1.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell. doi: 10.1109/TPAMI.2017.2699184

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Communications of the Acm, 60(2), 2012.

Längkvist, M., Kiselev, A., Alirezaie, M., & Loutfi, A. (2016). Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks. Remote Sensing, 8(4), 329.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Marmanis, D., Datcu, M., Esch, T., & Stilla, U. (2016). Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks. IEEE Geoscience & Remote Sensing Letters, 13(1), 105-109.

Mnih, V. (2013). Machine learning for aerial image labeling. University of Toronto (Canada).

Nagendra, H., & Rocchini, D. (2008). High resolution satellite imagery for tropical biodiversity studies: the devil is in the detail. Biodiversity & Conservation, 17(14), 3431-3442.

Otukei, J. R., & Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. International Journal of Applied Earth Observation & Geoinformation, 12(1), S27-S31.

Schmidhuber, J., Meier, U., & Ciresan, D. (2012). Multi-column deep neural networks for image classification. Paper presented at the Computer Vision and Pattern Recognition.

Scott, G. J., England, M. R., Starms, W. A., Marcum, R. A., & Davis, C. H. (2017). Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery. IEEE Geoscience & Remote Sensing Letters, PP(99), 1-5.

Tseng, M. H., Chen, S. J., Hwang, G. H., & Shen, M. Y. (2008). A genetic algorithm rule-based approach for land-cover classification. Isprs Journal of Photogrammetry & Remote Sensing, 63(2), 202-212.

Wang, J., Song, J., Chen, M., & Yang, Z. (2015). Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine. International Journal of Remote Sensing, 36(12), 3144-3169.

Xiong, J., Thenkabail, P. S., Gumma, M. K., Teluguntla, P., Poehnelt, J., Congalton, R. G., . . . Thau, D. (2017). Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. Isprs Journal of Photogrammetry & Remote Sensing, 126, 225-244.

Yang, C., Everitt, J. H., & Murden, D. (2011). Evaluating high resolution SPOT 5 satellite imagery for crop identification. Computers & Electronics in Agriculture, 75(2), 347-354.

Zhang, L., Zhang, L., & Du, B. (2016). Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. IEEE Geoscience & Remote Sensing Magazine, 4(2), 22-40.

Zhou, W., Newsam, S., Li, C., & Shao, Z. (2017). Learning Low Dimensional Convolutional Neural Networks for High-Resolution Remote Sensing Image Retrieval. Remote Sensing, 9(5), 489.