Research Article

Depth Semantic Segmentation of Tobacco Planting Areas from Unmanned Aerial Vehicle Remote Sensing Images in Plateau Mountains

Liang Huang,^{1,2} Xuequn Wu ^{1,2}, Qiuzhi Peng,^{1,2} and Xueqin Yu³

¹Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, China

²Surveying and Mapping Geo-Informatics Technology Research Center on Plateau Mountains of Yunnan Higher Education, Kunming 650093, China

³Kunming Surveying and Mapping Institute, Kunming 650051, China

Correspondence should be addressed to Xuequn Wu; wuxuequn520@163.com

Received 2 November 2020; Revised 27 January 2021; Accepted 9 February 2021; Published 1 March 2021

Academic Editor: Jose S. Camara

Copyright © 2021 Liang Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The tobacco in plateau mountains has the characteristics of fragmented planting, uneven growth, and mixed/interplanting of crops. It is difficult to extract effective features using an object-oriented image analysis method to accurately extract tobacco planting areas. To this end, the advantage of deep learning features self-learning is relied on in this paper. An accurate extraction method of tobacco planting areas based on a deep semantic segmentation model from the unmanned aerial vehicle (UAV) remote sensing images in plateau mountains is proposed in this paper. Firstly, the tobacco semantic segmentation dataset is established using Labelme. Four deep semantic segmentation models of DeeplabV3+, PSPNet, SegNet, and U-Net are used to train the sample data in the dataset. Among them, in order to reduce the model training time, the MobileNet series of lightweight networks are used to replace the original backbone networks of the four network models. Finally, the predictive images are semantically segmented by trained networks, and the mean Intersection over Union (mIoU) is used to evaluate the accuracy. The experimental results show that, using DeeplabV3+, PSPNet, SegNet, and U-Net to perform semantic segmentation on 71 scene prediction images, the mIoU obtained is 0.9436, 0.9118, 0.9392, and 0.9473, respectively, and the accuracy of semantic segmentation is high. The feasibility of the deep semantic segmentation method for extracting tobacco planting surface from UAV remote sensing images has been verified, and the research method can provide a reference for subsequent automatic extraction of tobacco planting areas.

1. Introduction

Tobacco is a crop of high economic value, which plays a significant role in the national financial accumulation and part of local economic development. In China, Yunnan Province is the main concentrated tobacco-producing area. In 2018, the total planted area and totaled output of tobacco accounted for 38.97% and 37.69% of the country. It is the largest tobacco production base in China [1]. But at the same time, the tobacco planting process has a high risk and is vulnerable to natural disasters and pests. Therefore, a timely grasp of tobacco spatial distribution, planting area, growth, yield and disaster loss, and other pieces of information is of great significance for achieving accurate tobacco

management, accurate production estimation, and assisting government decision-making. Among them, the rapid and accurate extraction of the tobacco planting area is an important prerequisite for fine tobacco management.

Due to the wide area of tobacco planting and the large distribution range, manual surveys are less efficient and susceptible to errors caused by human factors [2]. The emergence and development of remote sensing technology have made up for the shortcomings of manual surveys, and remote sensing technology has become the main technical means for monitoring tobacco planting area. The technology and methods of using remote sensing images to monitor tobacco planting area have made great progress in the past ten years. Data sources range from medium and low spatial

resolution satellite remote sensing images (such as Landsat, HJ-1) to high spatial resolution satellite remote sensing images (such as SPOT-5, China-Brazil Earth Resources Satellite 02B, ZY-1 02C, ZY-3) [3–7] and from an optical image to synthetic aperture radar [8]; a platform ranges from high-altitude satellite remote sensing to low-altitude UAV remote sensing [9–11]; monitoring methods range from statistical methods based on pixel features [5] to object-oriented methods [6, 7, 9]; monitoring content ranges from area to individual tobacco plant [12, 13].

Because Yunnan Province belongs to a low-latitude plateau area, tobacco planting area is small, spatial distribution is scattered, and tobacco and other crops are generally mixed or intercropped [3, 4]. Therefore, it is difficult to accurately extract planting area by using low and medium spatial resolution remote sensing images, and small plots are easy to miss detection. However, it is difficult to ensure that images can be obtained in specific regions and specific phenophase by using high spatial resolution satellite remote sensing images. UAV remote sensing has become the main means of tobacco planting area monitoring because of its flexible and high spatial resolution. Object-oriented image analysis and deep learning are the main classification methods of high spatial resolution remote sensing images. However, image segmentation and feature extraction restrict the development of object-oriented image analysis methods. At present, the deep semantic segmentation method has been widely used in the agricultural field and has achieved gratifying results. For example, a large-scale crop mapping method using multitemporal dual-polarization SAR data was proposed in the literature [14], and the U-Net was used to predict the different crop types. In literature [15], the convolutional neural network (CNN) and the Hough transform were used to detect crop rows in images taken by a UAV. Literature [16] used the deep learning framework TensorFlow to construct a platform for sampling, training, testing, and classifying to extract and map crop area based on DeeplabV3+. In conclusion, to realize the accurate extraction of tobacco planting area in the plateau of Yunnan Province, the planting area of tobacco was extracted accurately by using four deep semantic segmentation models, DeeplabV3+ [17], PSPNet [18], SegNet [19], and U-Net [20]. At the same time, to reduce the training cost, the MobileNet series [21] networks to replace the backbone networks of four deep networks.

2. Data and Methods

2.1. Overview of the Study Area. The study area is located in Xiyang County of Yi Nationality, Jinning District, Kunming City ($24^{\circ}23'N \sim 24^{\circ}33'N$, $102^{\circ}11'E \sim 102^{\circ}22'E$), as shown in Figure 1. The township covers an area of 160.32 km^2 , with complex terrain, and the difference between the highest and lowest elevations is 1223 m, belonging to a unique three-

dimensional climate. Tobacco planting is the pillar industry of the town.

2.2. Data Acquisition and Preprocessing. The low-altitude remote sensing platform used for data acquisition is phantom 4 RTK UAV, and the camera is fc6310r_8.8_5472 $\times$ 3648. To obtain the local tobacco data, several flight belts were designed, and one of them was used as case data for processing and analysis. The case data is Lvxi Village, Xiyang Country of Yi Nationality. The aerial photography time is July 29, 2020. The route planning is shown in Figure 2, and the data thumbnail is shown in Figure 3. The spatial resolution of the image is 0.027 m, and the coverage area is 0.1984 km^2 . The coordinate system is UTM zone 48 and Northern Hemisphere transverse Mercator WGS 84.

The original UAV remote sensing image is processed by PIE-UAV V6.0 for image matching, image alignment, camera optimization, orthophoto correction, image blending, and mosaic to generate a digital orthophoto map. The image size is 19439×22081 pixels.

2.3. Production of Tobacco Semantic Segmentation Dataset. In this paper, the original DOM image was cut into 1280×720 pixels in the batch, and the images without tobacco cover were deleted. The remaining 238 images containing tobacco were included. Labelme, an image marking tool, is used to label the tobacco single category manually, as shown in Figure 4.

2.4. Semantic Segmentation of Tobacco. Due to the great differences in tobacco growth (Figure 5(a)), planting area (Figure 5(b)), planting density (Figure 5(c)), and planting environment (Figures 5(d)–5(f)), it is difficult to find ideal features for high-precision extraction of tobacco from UAV remote sensing images by using an object-oriented method. Because of the advantages of self-learning features, deep learning can not only learn simple features but also learn more abstract features. Therefore, this paper uses the method of deep semantic segmentation to extract tobacco from UAV remote sensing images.

At present, there are many network models for deep semantic segmentation, including fully supervised learning image semantic segmentation methods and weakly supervised image semantic segmentation methods. However, the performance of most weak supervised methods still lags behind that of full supervised methods [22]. Therefore, this paper adopts fully supervised image semantic segmentation methods. In this paper, four network models, DeeplabV3+, PSPNet, SegNet, and U-Net, are used to the semantic segment of tobacco in UAV remote sensing images. To greatly reduce the network training time under the premise that the prediction accuracy is not affected, this paper uses the

FIGURE 1: Geographical situation of the study area.

FIGURE 2: Design of flight belt.

lightweight MobileNet series model [21] to replace the original backbone network of the four network models. Among them, the DeeplabV3+ network adopts the MobileNetV2 model, and the other three networks use the MobileNetV1 model; the structure of the four network models is shown in Figure 6.

2.4.1. Network Training. In order to verify the computing efficiency and efficiency of the lightweight backbone network model, a medium configuration hardware device is selected on the processing platform. The specific configuration is as follows: Intel Core i7-8700 four-core processor, NVIDIA GTX1070, 8G GDDR5 video memory, and 16G

FIGURE 3: Thumbnail of case data.

FIGURE 4: Continued.

FIGURE 4: Dataset of semantic segmentation: (a) partial cropping data in dataset; (b) corresponding label data.

FIGURE 5: Tobacco in different scenes.

DDR4 memory. The training environment of DeeplabV3+, PSPNet, SegNet, and U-Net is TensorFlow GPU version 1.13.1 and Keras version 2.1.5. Because of the small dataset in this paper, the ratio of training data (including

verification data) and prediction data is 7:3, 167 training images (151 training images and 16 verification images) and 71 predicted images. The training steps and skills are as follows:

- (1) Setting parameter: Defining the number of classes (NCLASSES), learning rate (LR), Batch Size (BS), and Epoch. The images are divided into the tobacco image and nontobacco image, so NCLASSES=2 is defined. In order to verify the impact of other parameters on time efficiency and accuracy, based on comprehensive consideration of the experimental platform and data set, taking the U-Net network as an example, the BS values are set as 2, 4, 6, and 8, the LR values are set as 1×10^{-2} , 1×10^{-3} and 1×10^{-4} , the Epoch values are 40, 50, and 60, respectively. Table 1 shows the time efficiency and accuracy under different values of the three parameters. According to Table 1, LR = 1×10^{-2} , BS = 4, and Epoch = 50 are selected in this paper.
- (2) Downloading the weight file: The MobileNetV1 and MobileNetV2 weight files are downloaded at <https://github.com/fchollet/deep-learning-models/releases>.
- (3) Disrupting the training data randomly: When a CNN is trained, although the training data is fixed, due to the minibatch training mechanism, the training data set can be shuffled randomly before each Epoch of the model training. Such processing not only will increase the rate of model convergence but also can slightly improve the model's prediction results on the test set.
- (4) Batch Normalization (BN) [23]: To speed up the convergence speed of model training, in the model training, the network activation is normalized by minibatch, so that the mean value of the result is 0 and the variance is 1.
- (5) Selecting optimizer: In order to reduce training time and computing resources, optimization algorithms that make the model converge faster are needed. The Adam [24] training optimizer is selected in this paper. The LR of each parameter is adjusted dynamically by the first-order moment estimation and second-order moment estimation of the gradient in Adam [24], so the parameter update is more stable.

Figure 7 shows the training accuracy and test accuracy of the four kinds of networks. It can be seen from Figure 6 that the training accuracy and test accuracy of the four kinds of networks have a small gap, and the semantic segmentation performance is good.

3. Results and Discussion

The training parameters are obtained by network training, and it is used for semantic segmentation of tobacco planting area from 71 predicted images in the prediction function. In order to verify the accuracy of semantic segmentation, mIoU [25] and mPA [25] are used to evaluate the overall accuracy of 71 scene images; Precision [26], Recall [26], F1 [27], IoU [25], and PA [25] are used as evaluation indicators to quantitatively evaluate semantic segmentation accuracy of each scene image.

Assume that there are $k + 1$ classes ($0, \dots, k$) in the data set, and 0 usually represents the background. The calculation formula of each indicator is as follows:

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{TP + FP + FN}, \quad (1)$$

$$mPA = \frac{1}{k+1} \sum_{i=0}^k \frac{TP + TN}{TP + TN + FP + FN}, \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (4)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (5)$$

$$IoU = \frac{TP}{TP + FP + FN}, \quad (6)$$

$$PA = \frac{TP + TN}{TP + TN + FP + FN}. \quad (7)$$

In formulas (1)~(7), i represents the i -th class; TP is True Positive; FP is False Positive; FN is False Negative; TN is True Negative.

Deeplabv3+, PSPNet, SegNet, and U-Net networks are used to semantically segment 71 scene prediction images; the mIoU and mPA obtained are shown in Table 2. It can be seen from Table 2 that, in the two indicators of mIoU and mPA, the results obtained by using the four networks are better than 90%. The results show that deep learning has a very good performance for semantic segmentation of tobacco planting areas. Among them, the U-Net network has the highest mIoU and mPA, and the semantic segmentation results have the best overall performance; the PSPNet network has the lowest overall prediction accuracy. The overall prediction accuracy of U-Net network, Deeplabv3+, and SegNet network has a small difference.

Due to the large number of predicted images, it cannot be fully displayed. This paper selects six images as example data, as shown in Figures 8(a)–8(c) and 9(a)–9(c). Among them, the tobacco in Figure 8(a) grows well and is densely planted; in Figure 8(b), some areas of tobacco are exposed too much and there are crops similar to the tobacco spectrum; in Figure 8(c), the planting density of tobacco is sparse and the growth is different; in Figure 9(a), there are roads, tobacco, and weeds; in Figure 9(b), there are buildings, roads, tobacco, and weeds, and some tobacco planting areas are small; in Figure 9(c), there are large areas of crops similar to tobacco spectrum.

It can be seen from Figures 8 and 9 and Table 3 that the four kinds of networks used in the data of example 1 have obtained good semantic segmentation results. Among them, the U-Net network has the highest scores in the five evaluation indicators, but there are also some ridges misdetected as tobacco planting areas. The PSPNet network is

FIGURE 6: Continued.

FIGURE 6: Four kinds of deep learning network structure. (a) DeeplabV3+. (b) PSPNet. (c) SegNet. (d) U-Net.

TABLE 1: Time efficiency and accuracy under different values of three parameters.

Parameter	Parameter value	Time (s)	Training accuracy	Test accuracy
LR	LR = 1×10^{-2} ; BS = 4; Epochs = 50	443.000	0.9762	0.9427
	LR = 1×10^{-3} ; BS = 4; Epochs = 50	438.491	0.9615	0.9119
	LR = 1×10^{-4} ; BS = 4; Epochs = 50	450.245	0.9182	0.8996
BS	LR = 1×10^{-2} ; BS = 2; Epochs = 50	518.000	0.9661	0.9342
	LR = 1×10^{-2} ; BS = 4; Epochs = 50	443.000	0.9762	0.9427
	LR = 1×10^{-2} ; BS = 6; Epochs = 50	428.673	0.9695	0.9235
	LR = 1×10^{-2} ; BS = 8; Epochs = 50	397.276	0.9555	0.9287
Epoch	LR = 1×10^{-2} ; BS = 4; Epochs = 40	358.656	0.9667	0.9312
	LR = 1×10^{-2} ; BS = 4; Epochs = 50	443.000	0.9762	0.9427
	LR = 1×10^{-2} ; BS = 4; Epochs = 60	549.154	0.9654	0.9303

TABLE 2: The accuracy evaluation indicators of tobacco and nontobacco.

Class	Network	mIoU	mPA
Tobacco	DeeplabV3+	94.36	96.09
	PSPNet	91.18	94.41
	SegNet	93.92	96.08
	U-Net	94.73	96.54
Nontobacco	DeeplabV3+	78.26	91.88
	PSPNet	67.56	84.28
	SegNet	76.38	89.75
	U-Net	79.23	91.49

particularly obvious than the other three networks, and the IoU index in which is also the lowest. In the data of example 2, no network scores the highest in all five indicators. The PSPNet network scores are relatively high and balanced; it also has good performance for a scene with uneven exposure. From the visual comparison, it can be seen that there is

false detection in Deeplabv3+, SegNet, and U-Net networks; some crops close to the tobacco spectrum are mistakenly detected as tobacco. Although the problem does not appear in the PSPNet network, some tobacco is missing in the PSPNet network. In the data of example 3, none of the networks achieved the highest scores in all five indicators. In all four networks, some sparse areas of tobacco growth are missed, so the semantic segmentation of uneven growing areas needs to be further strengthened. In the data of example 4, the segmentation results of the four kinds of network semantics are good. The U-Net network scores the highest among the five evaluation indicators; in the Deeplabv3+ and PSPNet networks, the small tobacco planting area in the lower right corner was missed. In the data of example 5, the U-Net network scores the highest among the five evaluation indicators, but the performance of precision, IoU, and PA is lower than that of the first four scenes. The four kinds of networks have different degrees of missed detection, and U-Net has the smallest missed detection area.

FIGURE 7: Training accuracy and test accuracy of the four networks.

FIGURE 8: Continued.

FIGURE 8: Semantic segmentation results of four kinds of networks: (a) sample data 1; (b) sample data 2; (c) sample data 3; (d) label data of sample data 1; (e) label data of sample data 2; (f) label data of sample data 3; (g) DeeplabV3+ result of sample data 1; (h) DeeplabV3+ result of sample data 2; (i) DeeplabV3+ result of sample data 3; (j) PSPNet result of sample data 1; (k) PSPNet result of sample data 2; (l) PSPNet result of sample data 3; (m) SegNet result of sample data 1; (n) SegNet result of sample data 2; (o) SegNet result of sample data 3; (p) U-Net result of sample data 1; (q) U-Net result of sample data 2; (r) U-Net result of sample data 3.

FIGURE 9: Continued.

FIGURE 9: Continued.

FIGURE 9: Semantic segmentation results of four kinds of networks: (a) sample data 4; (b) sample data 5; (c) sample data 6; (d) label data of sample data 4; (e) label data of sample data 5; (f) label data of sample data 6; (g) DeeplabV3+ result of sample data 4; (h) DeeplabV3+ result of sample data 5; (i) DeeplabV3+ result of sample data 6; (j) PSPNet result of sample data 4; (k) PSPNet result of sample data 5; (l) PSPNet result of sample data 6; (m) SegNet result of sample data 4; (n) SegNet result of sample data 5; (o) SegNet result of sample data 6; (p) U-Net result of sample data 4; (q) U-Net result of sample data 5; (r) U-Net result of sample data 6.

In the data of example 6, the Deeplabv3+ network scored the highest among the three indicators of Recall, F1, and IoU. It can also be seen from Figure 9(i) that the results obtained by the Deeplabv3+ network are more consistent with the labeled image. Some ridges are marked as tobacco planting areas in the PSPNet network, and some other crops are marked as tobacco planting areas in SegNet and U-Net network, but the false detection areas are small.

Combining Figures 8 and 9 and Table 3, it can be seen that good semantic segmentation results can be obtained for tobacco planting areas in different scenes by using the four kinds of deep learning networks. But different networks still have certain differences in performance for different scenarios. There are a total of 30 scores for 6 scenes of images and 5 evaluation indicators, among which U-Net has 18 highest scores, DeeplabV3+ has 6 highest scores, SegNet has 4 highest scores, and PSPNet has 2 highest scores. Therefore, U-Net and Deeplabv3+ are better than SegNet and PSPNet for small sample tobacco planting areas data set. However, the dependence of the U-Net network on devices is lower than that of the DeeplabV3+ network, and the operation efficiency is higher. It can be seen from Table 4 that the prediction time and training time of the U-Net network are less than those of the DeeplabV3+ network, especially the training time.

4. Conclusions and Discussion

4.1. Conclusions. This paper mainly discusses the application potential of the deep semantic segmentation method on automatic extraction of tobacco planting areas in plateau mountains. Using four deep semantic segmentation methods of DeeplabV3+, PSPNet, SegNet, and U-Net, 151 images are trained, 16 images are verified, and 71 images are predicted. The experimental results show that, compared with the traditional object-oriented image analysis method, the deep semantic segmentation method does not need feature selection and optimization and has higher automation and better universality. At the same time, compared with the four networks, the performance of the U-Net

TABLE 3: The accuracy evaluation indicators of 6 sample data.

Data	Network	Precision	Recall	F1	IoU	PA
Sample data 1	DeeplabV3+	96.79	98.79	97.78	95.66	96.79
	PSPNet	96.81	96.75	96.78	93.77	96.81
	SegNet	96.21	99.08	97.62	95.36	96.21
	U-Net	96.84	99.11	97.96	96.00	96.84
Sample data 2	DeeplabV3+	97.96	95.43	96.68	93.57	97.96
	PSPNet	96.78	99.38	98.06	96.20	96.78
	SegNet	97.42	85.54	91.09	83.65	97.42
	U-Net	96.61	80.94	88.08	78.70	96.61
Sample data 3	DeeplabV3+	89.43	96.80	92.97	86.86	89.43
	PSPNet	85.36	94.50	89.70	81.32	85.36
	SegNet	93.28	95.96	94.60	89.76	93.28
	U-Net	93.75	95.27	94.50	89.58	93.75
Sample data 4	DeeplabV3+	91.43	98.97	95.05	90.57	91.43
	PSPNet	93.37	97.43	95.35	91.12	93.37
	SegNet	96.34	97.28	96.81	93.82	96.34
	U-Net	97.25	97.70	97.48	95.08	97.25
Sample data 5	DeeplabV3+	73.85	97.27	83.96	72.35	73.85
	PSPNet	80.05	95.83	87.23	77.35	80.05
	SegNet	81.13	97.41	88.53	79.42	81.13
	U-Net	88.08	98.05	92.79	85.56	88.08
Sample data 6	DeeplabV3+	94.82	99.06	96.89	93.97	94.82
	PSPNet	94.74	97.27	95.99	92.29	94.74
	SegNet	95.59	96.75	96.17	92.62	95.95
	U-Net	95.94	97.82	96.87	93.93	95.94

TABLE 4: Prediction time of four kinds of networks.

Networks	Prediction time (s)	Training time (s)
DeeplabV3+	13.437	1016.998
PSPNet	10.179	478.716
SegNet	10.597	390.346
U-Net	9.651	443.000

network in tobacco semantic segmentation under a small sample set is better than other networks, and the equipment requirements are not too high, which is convenient for the promotion of deep semantic segmentation method in tobacco planting areas extraction.

4.2. Discussion. The advantages of deep learning in tobacco planting area extraction have been effectively verified, but there are still some problems to be further studied:

- (1) The results showed that the ridges between tobacco planting areas are mistakenly detected. It is worth further study whether the extraction accuracy can be improved by adding field boundary information.
- (2) Weeds or crops with similar spectral characteristics are easy to be mistakenly detected. It is worthy of further attempt to eliminate the same and different spectra by using the “shape-spectrum” joint feature.
- (3) Semantic segmentation of multiple crops needs to be verified. The image is only divided into tobacco and nontobacco for semantic segmentation in this paper, which can weaken the problem of sample imbalance. If a variety of crops are to be semantically segmented at the same time, the applicability of the proposed method needs to be verified.
- (4) The planting areas with poor growth or no harvest are easy to be missed. Morphological operation is an effective method to deal with holes. It is a feasible choice to use the morphological method to deal with the problem of missing detection caused by poor growth area.
- (5) There are many hyperparameter settings in deep learning. Different parameters will have a certain impact on the subsequent time efficiency and accuracy. How to choose the best parameters can still be a relatively difficult task.
- (6) The extraction effect of tobacco planting areas under different spatial resolution needs to be further verified, because the flying height of the UAV determines the spatial resolution of the image. For areas with a large drop, there may be differences in tobacco features obtained at the same height. The effect of the terrain drop on the semantic segmentation effect of tobacco needs further research.

Data Availability

The original data have not been made publicly available, but it can be used for scientific research. Other researchers can send emails to the first author if needed.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China, Grants nos. 41961039 and 41961053, and the Applied Basic Research Programs of Science and

Technology Department of Yunnan Province, Grant no. 2018FB078.

References

- [1] Department of Rural Social Economic Investigation, National Bureau of Statistics, *China Rural Statistical Yearbook*, China Statistics Press, Beijing, China, 2019.
- [2] J. Tao, G. Shen, Y. Xu, and H. Liang, “Prospect of applying remote sensing to tobacco planting monitoring and management,” *Acta Tabacaria Sinica*, vol. 21, no. 2, pp. 111–116, 2015.
- [3] M. Wu, Q. Cui, L. Zhang, and N. Zhao, “Tobacco field monitoring and classification method study in mountainous area,” *Remote Sensing Technology and Application*, vol. 23, no. 3, pp. 305–309, 2008.
- [4] G. Peng, L. Deng, W. Cui, T. Ming, and W. Shen, “Remote sensing monitoring of tobacco field based on phenological characteristics and time series image—a case study of Chengjiang county, Yunnan province, China,” *Chinese Geographical Science*, vol. 19, no. 2, pp. 186–193, 2009.
- [5] Z. Wang, Y. Chen, J. Mo, and X. Wang, “Recognition of flue-cured tobacco crop based on spectral characteristics extracted from HJ-1 remote sensing images,” *Tobacco Science & Technology*, vol. 1, pp. 72–76, 2014.
- [6] T. Li, “The research of extracted tobacco planting area based on object-oriented classification method,” Master thesis, Sichuan Agricultural University, Chengdu, China, 2013.
- [7] M. Liu, “Study on yield estimate model of relay-cropping tobacco in hilly regions based on ZY-3 remote sensing images,” Master thesis, Shandong Agricultural University, Tai’an, China, 2016.
- [8] Z. Zhou, *Application of Synthetic Aperture Radar in Mountain Agriculture - Tobacco Planting Monitoring*, Science Press, Beijing, China, 2017.
- [9] M. Dong, “Extraction of tobacco planting areas from UAV remote sensing imagery by object-oriented classification method,” *Science of Surveying and Mapping*, vol. 39, no. 9, pp. 87–90, 2014.
- [10] X. Zhu, G. Xiao, P. Wen, J. Zhang, and C. Hou, “Mapping tobacco fields using UAV RGB images,” *Sensors*, vol. 19, no. 8, p. 1791, 2019.
- [11] J. Chen, P. Liu, G. Huang, Z. Li, and J. Liu, “Information extraction of tobacco planting area based on unmanned aerial vehicle remote sensing images,” *Hunan Agricultural Sciences*, vol. 1, pp. 96–99, 2018.
- [12] Z. Fan, J. Lu, M. Gong, H. Xie, and E. D. Goodman, “Automatic tobacco plant detection in UAV images via deep neural networks,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 876–887, 2018.
- [13] J. Fu, “Statistical method and experiment of tobacco plant numbers in mountainous areas based on UAV image,” Master thesis, Guizhou University, Guiyang, China, 2019.
- [14] S. Wei, H. Zhang, C. Wang, Y. Wang, and L. Xu, “Multi-temporal SAR data large-scale crop mapping based on U-net model,” *Remote Sensing*, vol. 11, no. 1, p. 68, 2019.
- [15] M. D. Bah, A. Hafiane, and R. Canals, “CRowNet: deep network for crop row detection in UAV images,” *IEEE Access*, vol. 8, pp. 5189–5200, 2020.
- [16] Z. Du, J. Yang, C. Ou, and T. Zhang, “Smallholder crop area mapped with a semantic segmentation deep learning method,” *Remote Sensing*, vol. 11, no. 7, p. 888, 2019.

- [17] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the 15th European Conference on Computer Vision, ECCV*, pp. 833–851, Munich, Germany, September 2018.
- [18] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 6230–6239, Honolulu, HI, USA, July 2017.
- [19] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI*, pp. 234–241, Munich, Germany, October 2015.
- [21] A. G. Howard, "MobileNets: efficient convolutional neural networks for mobile vision applications," 2017, <https://arxiv.org/abs/1704.04861>.
- [22] X. Tian, L. Wang, and Q. Ding, "Review of image semantic segmentation based on deep learning," *Journal of Software*, vol. 30, no. 2, pp. 440–468, 2019.
- [23] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning, ICML*, pp. 448–456, Lille, France, July 2015.
- [24] D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations, ICLR*, pp. 1–15, San Diego, CA, USA, May 2015.
- [25] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 3431–3440, IEEE Computer Society, Boston, MA, USA, June 2015.
- [26] L. Huang, B. X. Yao, P. D. Peng, A. P. Ren, and Y. Xia, "Superpixel segmentation method of high resolution remote sensing images based on hierarchical clustering," *Journal of Infrared and Millimeter Waves*, vol. 39, no. 2, pp. 263–272, 2020.
- [27] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9976–9992, 2019.