Detection and threshold-adaptive segmentation of farmland residual plastic film images based on CBAM-DBNet

Lijian Xiong[1,2], Can Hu[2,3*], Xufeng Wang[2,3], Hongbiao Wang[2,3], Xiuying Tang[1], Xingwang Wang[2,3]

(1. *College of Engineering, China Agricultural University, Beijing 100083, China*;
2. *Modern Agricultural Engineering Key Laboratory at the Universities of Education Department of Xinjiang Uygur Autonomous Region, Tarim University, Alaer 843300, China*;
3. *College of Mechanical and Electrical Engineering, Tarim University, Alaer 843300, China*)

Abstract: Robust, accurate, and fast monitoring of residual plastic film (RPF) pollution in farmlands has great significance. Based on CBAM-DBNet, this study proposed a threshold-adaptive joint framework for identifying the RPF on farmland surfaces and estimating its coverage rate. UAV imaging was used to gather images of the RPF from several locations with various soil backgrounds. RPFs were manually labeled, and the degree of RPF pollution was defined based on the RPF coverage rate. Combining differentiable binarization network (DBNet) with the convolutional block attention module (CBAM), whose feature extraction module was improved. A dynamic adaptive binarization threshold formula was defined for segmenting the RPF's approximate binary map. Regarding the RPF image detection branch, the CBAM-DBNet exhibited a precision (*P*) value of 85.81%, a recall (*R*) value of 82.69%, and an F1-score (F1) value of 84.22%, which was 1.09 percentage points higher than the DBNet in the comprehensive index F1 value. For the RPF image segmentation branch, using CBAM-DBNet to segment the RPF image combined with an adaptive binarization threshold formula. Subsequently, the mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute error (MAE) of the prediction of RPF's coverage rate were 0.276, 0.366, and 0.605, respectively, outperforming the DBNet and the Iterative Threshold method. This study provides a theoretical reference for the further development of evaluation technology for RPF pollution based on UAV imaging.

Keywords: binarization threshold adaptive, residual plastic film, object detection, image segmentation, UAV remote sensing
DOI: 10.25165/j.ijabe.20241705.8069

Citation: Xiong L J, Hu C, Wang X F, Wang H B, Tang X Y, Wang X W. Detection and threshold-adaptive segmentation of farmland residual plastic film images based on CBAM-DBNet. Int J Agric & Biol Eng, 2024; 17(5): 231–238.

1 Introduction

The plastic film can increase temperature and moisture retention and promote the growth of crops in arid areas. The application of plastic film technology has increased crop yield and water utilization by 45.5% and 58.0%, respectively[1]. Nevertheless, the mulch that has not been fully recovered can lead to residual plastic film (RPF) being left in the farmland, which can damage the ecology of the farmland and endanger the health of the soil[2,3]. In areas with long-term plastic film mulch, mechanical recycling was often used to reduce the amount of RPF in the soil[4]. Due to the RPF's small size, irregular shape, and color close to the soil, it is difficult to identify. Rapid and accurate monitoring of RPF's coverage rate is crucial for enhancing the efficiency of mechanized recycling and implementing control of RPF pollution in agricultural soil, among other things[5].

The traditional assessment method for RPF pollution is manual investigation[6,7]. Field workers collected RPF through field investigation, sampling, and subsequent washing, drying, and weighing to gain a preliminary understanding of the distribution characteristics of residual film contamination in the study area[8]. However, the manual sampling method for residual film pollution monitoring in these studies is labor-intensive and low-efficiency.

UAV remote sensing technology has many advantages, such as high operation efficiency, good mobility, high spatial resolution, etc. In recent years, there has been a growing interest among scholars in focusing on RPF pollutant monitoring using UAV remote sensing technology[9-12]. Most of these studies use traditional image segmentation methods such as manual threshold segmentation, iterative threshold segmentation, large-law segmentation, and so on. However, these methods require complex preprocessing and feature parameter extraction. In addition, these RPF images had a larger size, good continuity, and low fragmentation.

With the continued popularity of deep learning, UAV imaging combined with convolutional neural networks (CNN) has been increasingly widely used in agriculture[13-15]. Meanwhile, other researchers have investigated RPF detection and identification using UAV images[16-19]. For example, Zhang et al.[16] improved a classical object detection framework (Faster RCNN) to recognize the RPF in the farmland; Zhai et al.[17] proposed an evaluation method of the RPF's coverage rate using pixel block and machine learning models in order to effectively recognize the RPF in the cotton field; Qiu et al.[19] built a model for segmenting residual film from the cotton field image based on the three deep learning frameworks LinkNet, FCN, and DeepLabv3. These methods above take full advantage of

Received date: 2023-02-07 **Accepted date:** 2023-07-14
Biographies: **Lijian Xiong**, PhD candidate, research interest: agricultural informatization and UAV remote sensing, Email: xiong_lijian@163.com; **Xufeng Wang**, Professor, research interest: agricultural informatization and agricultural mechanization, Email, wxf@taru.edu.cn; **Hongbiao Wang**, Professor, research interest: agricultural mechanization, Email: 2289882252@qq.com; **Xiuying Tang**, Professor, research interest: agricultural informatization and agricultural mechanization, Email: txying@cau.edu.cn; **Xingwang Wang**, Postgraduate, research interest: UAV remote sensing for agriculture, Email: 10757212236@stumail.taru.edu.cn.

***Corresponding author: Can Hu**, Associate Professor, research interest: agricultural informatization. College of Mechanicaland Electrical Engineering, Tarim University, Alaer 843300, Xinjiang, China. Tel: +86-15770065003, Email: 120140004@taru.edu.cn.

CNN to automatically extract deep image features[20]. However, the robustness of image segmentation has not been fully considered. Due to the different colors and shapes of the RPF and the different soil backgrounds in the image, the contrast between the RPF and the background was different. Burgos-Artizzu et al.[21] pointed out that automatic adjustment of thresholds is crucial for robust segmentation across different backgrounds. Besides, few studies consider both RPF image detection and segmentation in the field of RPF pollution monitoring. Hence, the purpose of this study was to build a threshold-adaptive joint framework for RPF image detection and segmentation. The main contributions of this study are as follows:

1) Combining differentiable binarization network (DBNet)[22], an object detection network with a differentiable binarization method, with the convolutional block attention module (CBAM)[23], which has spatial and channel attention mechanisms, adapts DBNet to the detection of RPF images; 2) A threshold-adaptive joint framework for RPF image detection and segmentation based on CBAM-DBNet was proposed; 3) A dynamic adaptive binarization threshold formula was defined for segmenting the RPF's approximate binary map; 4) Image annotation datasets for RPF were constructed in a real environment; 5) Experiments demonstrated the effectiveness of the proposed framework for RPF image detection and segmentation.

2 Materials and methods

2.1 Data acquisition

Figure 1 shows the RPF image acquisition process. Images of RPF were taken at the farms near Alaer, and Shihezi, Xinjiang, China. The images were taken in April 2021, when the farmland had after-spring tillage and the surface RPF were evenly distributed. The images taken during this period facilitate normalization processing. The RPF images were collected using a UAV with a 12-megapixel effective camera developed by Dajiang (Shenzhen company). The camera shot from the top down at an angle of 90° and a height of 5 m from the ground. 2120 images were obtained from different soil backgrounds, and 237 images with duplicate or unclear objects were removed. A total of 1883 valid images were obtained. The obtained data were analyzed in three steps. First, the obtained image datasets were randomly divided into two groups: 1708 images in the training set and 175 images in the test set.

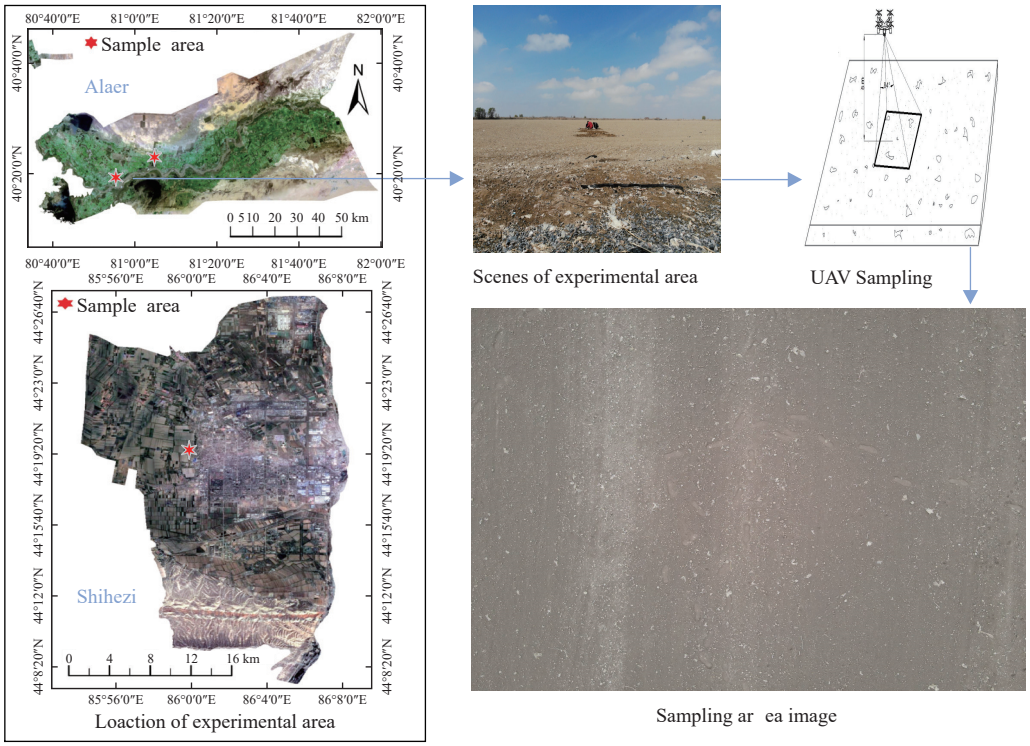


Figure 1 RPF image acquisition

Second, the image dataset was annotated using the tool "Sprite Labeling Assistant Colabeler" (http://www.jinglingbiaozhu.com/). After running the tool, the target samples of the RPF in each image were annotated; thereafter, a TXT file containing the target type and coordinate information was generated for training on the dataset. The detailed data operation steps are shown in Figure 2.

Third, the ratio of the sum of the manually labeled RPF area to the image area was calculated, and this was used as the label value for the RPF's coverage rate. S_i is the area of the ith labeled box of the image, S_{img} is the total area of the image. The coverage rate TR is defined as:

$$TR = \frac{\sum_{i=1}^{n} S_i}{S_{img}} \quad (1)$$

Figure 3 displays the quantity of RPF-labeled boxes and the RPF coverage rate for each image. The minimum value of the number of RPF-labeled boxes in the dataset is 4, the maximum value is 101, and the average value is 38. The minimum value of the percentage of RPF in the image is 0.18%, the maximum value is 5.39%, and the average value is 1.02%. The dataset is characterized by a large number of RPF in each image but a small percentage of the RPF area, which also presents challenges to the subsequent detection of RPF and coverage rate calculation.

2.2 Overview of DBNet

The RPF images have different backgrounds due to illumination, soil color, etc. The traditional image segmentation method based on global thresholds cannot match a more appropriate binarization threshold for each image, so it has an inherent accuracy bottleneck. Liao et al.[22] proposed a differentiable binarization

object detection neural network (DBNet). They used the differentiable binarization formula instead of the standard binarization formula and introduced the binarization operation into the neural network for joint optimization.

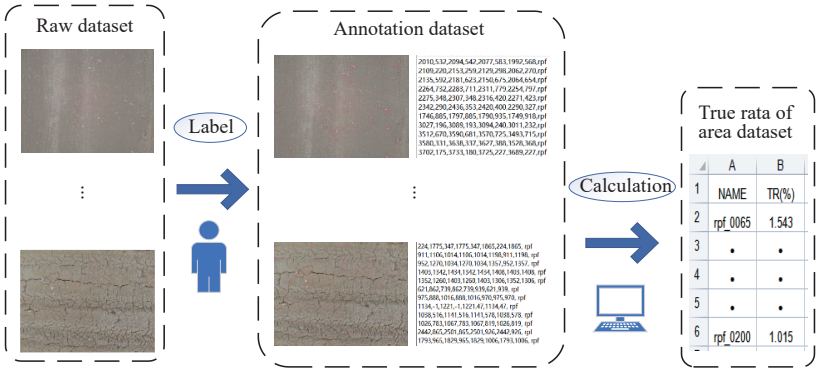


Figure 2 RPF images annotation

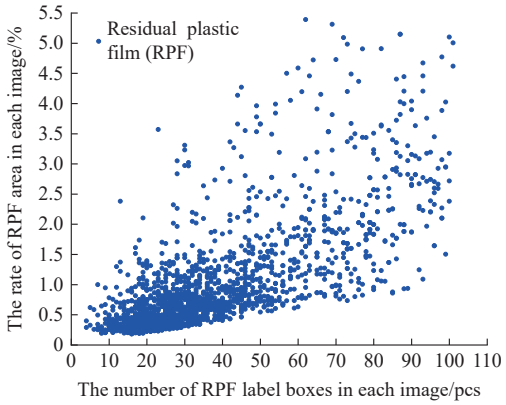


Figure 3 Annotated object and anchor box statistics.

DBNet first fed the image to the feature pyramid network (FPN)[24]. Then, the pyramid features were upsampled to the same scale to generate a feature map (*F*-map). Thirdly, the *F*-map was used to predict the probability map (*P*-map) and the threshold map (*T*-map). Finally, the approximate binary map ($\hat{B}-map$) was calculated by *P*-map and *F*-map. During the training period, the *P*-map, *T*-map, and $\hat{B}-map$ were supervised. In the inference period, the bounding boxes can be easily obtained from the approximate binary map or the probability map using a box formulation module.

2.2.1 Differentiable Binarization

The standard binarization formula is not differentiable. The differentiable binarization formula was proposed[22], and then the binarization operation was put into a partitioned network for joint optimization. The approximate binary map is represented by $\widehat{B_{i,j}}$, the probability map by $P_{i,j}$ the threshold map by $T_{i,j}$, and the amplifying factor by *K*. The differentiable binarization is defined as:

$$\widehat{B_{i,j}} = \frac{1}{1+\mathrm{e}^{-K\left(P_{i,j}-T_{i,j}\right)}} \tag{2}$$

2.2.2 Loss function

The loss function *L* can be expressed as a weighted sum of the loss for the probability map L_s, the loss for the binary map L_b, and the loss for the threshold map L_t. The loss calculation formula is as follows:

$$L = L_s + \alpha \times L_b + \beta \times L_t \tag{3}$$

In Equations (3), α and β are set to 1.0 and 10.0, respectively. Both L_s and L_b use binary cross-entropy loss. Online hard example mining is also employed to address the imbalance of positive and negative examples. The proportion of positive and negative samples is adjusted to 1:3[25].

2.3 Proposed the threshold-adaptive joint framework for RPF image detection and segmentation

In this study, Figure 4 shows the threshold-adaptive joint framework for RPF image detection and segmentation, which was based on CBAM-DBNet. The procedure was as follows:

1) The feature map was extracted. A classical feature extraction network was the residual network (ResNet)[26]. The CBAM was introduced to improve the feature extraction network of the DBNet since the RPF object was tiny and the features were not evident;

2) The feature map was spliced. An FPN structure was employed to fuse the feature maps of various sizes. Poor resolution but high semantics are features of small-scale feature maps, while high resolution but low semantics are characteristics of large-scale feature maps[27]. As a result of the lateral connections between the feature maps of the various layers, the feature maps include extensive positional and semantic information on the remaining membrane. The *F*-map was created by splicing together the feature maps of the four scales that have been upsampled to 1/4 size;

3) The $\hat{B}-map$ was created. The *F*-map was used to construct the *P*-map and the *T*-map, which were then combined using the differentiable binarization technique to produce the $\hat{B}-map$;

4) The RPF image was detected. The RPF's $\hat{B}-map$ and the box formulation module were used to generate the bounding box, and the RPF detection results were obtained;

5) The RPF image was segmented. The RPF's $\hat{B}-map$ was segmented into a binary image by using the adaptive threshold formula. RPF's coverage rate was calculated by combining the coverage rate calculation formulas.

2.3.1 Improved feature extraction module

There was some phenotypic similarity between the RPF and disturbances such as soil clods and leaves. Due to the RPF photos' small size, the original feature extraction module was unable to extract deeper features from them[28]. Therefore, CBAM was included in the feature extraction module in this study. The modules are plug-and-play, allowing the network to concentrate on the crucial areas of the feature graph and provide features with greater semantic information. CBAM was added to the first and last convolutional layers of ResNet. Figure 5 shows the CBAM's fundamental structure: *F* is the input feature mapping; M_c is the channel attention module; F'_c is the feature mapping for channel attention refinement; F' is the feature map of the fusion of *F* and F'_c; M_s is a spatial attention module; F'_s is the feature mapping of spatial attention refinement; F'' is the feature mapping for the final refinement of CBAM.

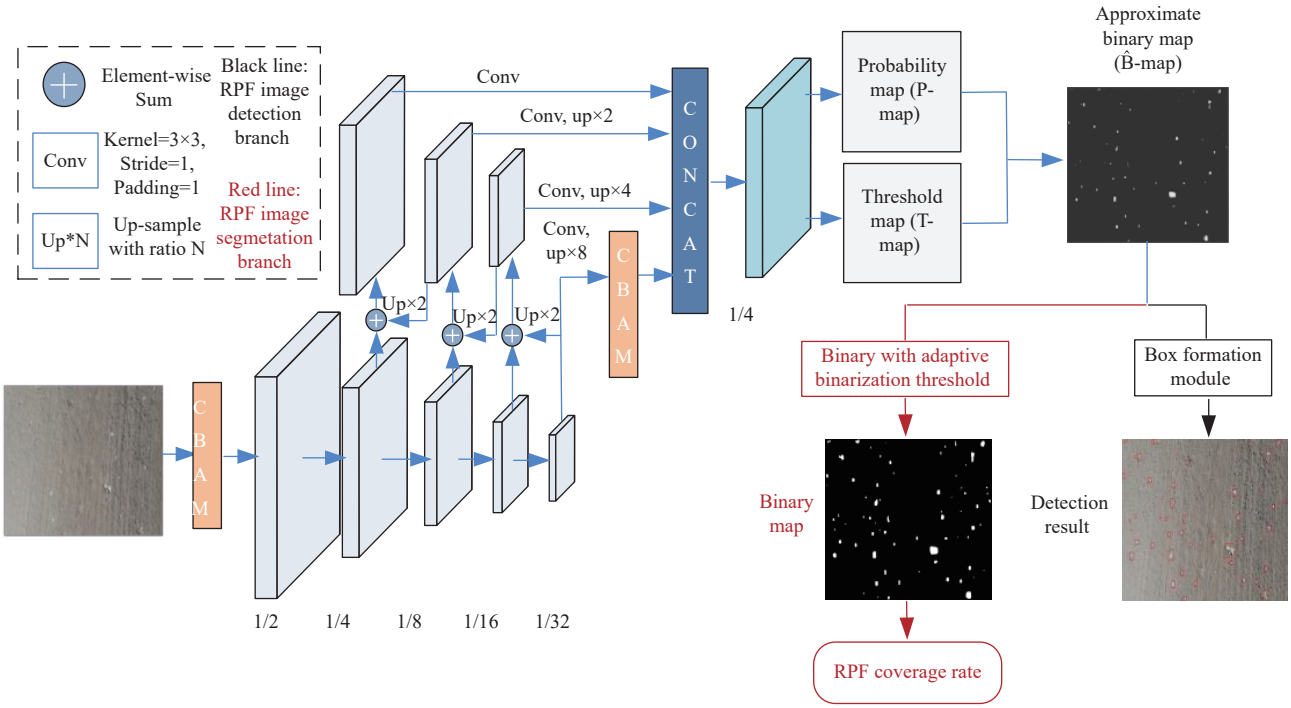


Figure 4 A threshold-adaptive joint framework for RPF image detection and segmentation

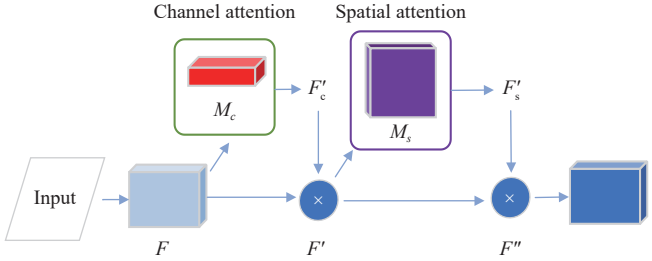


Figure 5 Convolutional Attention Module

The CBAM is mainly composed of two parts: the channel attention module and the spatial attention module. Relevant studies demonstrated that improving the channel attention module's features first before connecting it to the spatial attention module in series produced the best results. After the channel attention refinement, feature mapping will be more effective in extracting the contour features of the RPF. The average pooling and maximum pooling operations are performed on the multilayer perceptron to aggregate the spatial information of the feature mapping, generating two different spatial context descriptors F^c_{avg} and F^c_{max}. σ is the sigmoid activation function, and W_0 and W_1 are the weights of the shared network. For the input feature mapping $F=R^{C\times H\times W}$, F'_c is calculated as follows:

$$F'_c = \sigma\left(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1\left(W_0\left(F^c_{max}\right)\right)\right) \quad (4)$$

The spatial attention module focuses on the spatial relationship between features and is a supplement to channel attention. After spatial attention refinement, feature mapping will pay more attention to the location information of the RPF. σ is the sigmoid activation function, $f^{7\times7}$ represents the 7×7 convolution operation, and F^s_{avg} and F^s_{max} represent the channel-based global maximum pooling and global average pooling, respectively. The formula for F'_s is as follows:

$$F'_s = \sigma\left(f^{7\times7}\left(F^s_{avg}\cdot F^s_{max}\right)\right) \quad (5)$$

2.3.2 RPF image detection branch

The RPF's $\hat{B}-map$ was used in this study in the RPF detection branch to produce the RPF detection boxes. The generation of the detection boxes consisted of three steps: 1) the RPF's approximate binary map was binarized with a constant threshold to obtain a binary map; 2) the binary map's connected domains were obtained; 3) the shrunken text area was expanded back again by expansion according to the offset factor D' of the Vatti clipping algorithm to obtain the final detection boxes[29]. D' is calculated by the following equation:

$$D' = \frac{A'\times r'}{L'} \quad (6)$$

where, A' is the area of the shrinking polygon; L' is the perimeter of the shrinking polygon; r' is empirically set to 1.5.

2.3.3 RPF image segmentation branch

In the RPF image segmentation branch, the RPF's $\hat{B}-map$ generated by the prediction model. The processing was as follows: 1) the binarization threshold value was obtained by an adaptive-threshold formula; 2) the RPF's binary map was achieved by the adaptive threshold; 3) the RPF's coverage rate was computed. The details are as follows:

Adaptive-threshold formula. Each pixel value of the $\hat{B}-map$ was stored as an RPF's approximate binary matrix G; g_{ij} is the approximate binary value of the ith row and jth column of the approximate binary image; $g_{ij}\in[0,1]$; h and w are the height and width of the approximate binary map, respectively. The approximate binary matrix G is defined as:

$$\boldsymbol{G} = \begin{bmatrix} g_{11} & \cdots & g_{1w} \\ \cdots & g_{ij} & \cdots \\ g_{h1} & \cdots & g_{hw} \end{bmatrix} \quad (7)$$

The adaptive-binarization threshold bn is defined as Equation (9).

$$\text{bn} = \frac{\sum_i^m\left(\sum_j^n g_{ij}\right)}{h\cdot w}\times 255 \quad (8)$$

So far, the binarization threshold for each image has been

obtained adaptively. The RPF's approximate binary map was segmented into a binary map according to the bn, which was then used for the subsequent calculation of the RPF's coverage rate.

The RPF's coverage rate calculation. The RPF's coverage rate, i.e., the ratio of the total number of the RPF pixels to the total number of pixels in the image. n_{rpf} is the number of RPF pixels, i.e., the number of pixels with pixel values greater than bn. The RPF's coverage rate (PR) calculation is defined as:

$$\text{PR} = \frac{n_{\text{rpf}}}{h \cdot w} \tag{9}$$

2.4 Evaluation methods

In this section, the effectiveness of the proposed method was evaluated in terms of both RPF image detection and RPF image segmentation.

RPF detection belongs to the category of object detection. Therefore, the commonly used precision, recall, and F1-score (F1) were chosen as the evaluation indexes for RPF detection. Recall (R) indicates the ability of the model to find all relevant samples without missing any. Precision (P) indicates how well the model performs in correctly identifying relevant samples while minimizing false positives. The F1 is the harmonic mean of precision and recall, which provides a single metric that balances both precision and recall. R, P, and F1 are defined as follows:

$$R = \frac{T_p}{T_p + F_N} \times 100\% \tag{10}$$

$$P = \frac{T_p}{T_p + F_p} \times 100\% \tag{11}$$

$$\text{F1} = \frac{2 \times P \times R}{P + R} \tag{12}$$

where, T_P represents the case in which the model predicts the RPF and it is the RPF, F_N represents the case in which the model predicts the RPF but it is not the RPF, and F_P represents the case in which the model predicts the RPF but it is not the RPF.

The mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE) were often used as the evaluation index of the predicted model[30-32]. The MAPE is often used in practice because of its very intuitive interpretation in terms of relative error[30]. The RMSE indicates the sample standard deviation of the difference between the predicted and observed values (known as the residual)[31]. The MAE indicates the mean of the absolute error between the predicted and observed values[32]. TR_i is the coverage rate label value of the ith image and PR_i is the coverage rate prediction value of the ith image. N is the number of images. The smaller the values of MAPE, RMSE, and MAE, the higher the model prediction accuracy. The MAPE is defined as follows:

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{n} \left| \frac{\text{PR}_i - \text{TR}_i}{\text{TR}_i} \right| \tag{13}$$

The RMSE is defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{n} \left(\text{PR}_i - \text{TR}_i\right)^2} \tag{14}$$

The MAE is defined as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{n} \left|\text{PR}_i - \text{TR}_i\right| \tag{15}$$

2.5 Experimental software and hardware conditions

The hardware and software conditions in the experiments are listed in Table 1, and GPU acceleration was used in the training.

Table 1 Experimental settings

Configuration	Parameter
CPU	Intel(R) Core(TM) i9-10900K
GPU	NVIDIA GeForce RTX3060
Memory	12 GB
Operating system	Windows 10
Development enviroment	PyCharm Community Edition 2021
Programming language	Python3.6.2
Accelerated environment	CUDA 10.1;CUDNN 7.6.5
Library	Torch 1.9.0;TorchVision 0.10.0; Opencv 4.5.2;

In this study, the Adam optimizer was used to train the model. The training batch size was set to 8, and the Poly strategy was used to dynamically adjust the learning rate. The learning rate of each iteration was $l_r = l_r\left(1 - \frac{\text{iter}}{\text{maxiter}}\right)^{\text{power}}$, where the initial learning rate was set to 0.0001. The training batch epochs were set to 300.

3 Results and discussion

3.1 RPF image detection experiment

In the experiment of RPF image detection, this paper compared the DBNet with the CBAM-DBNet. The performance index comparison of the two methods is listed in Table 2. It can be seen that the improved method has increased the precision by 3.02 percentage points, slightly decreased the recall by 0.79 percentage points, and increased the F1 value by 1.09 percentage points. After the convolutional attention module was added to the feature extraction module, the network paid more attention to the deep features, which enabled the network to detect the RPF objects more accurately. However, some false positive samples were also generated, which made the recall slightly lower.

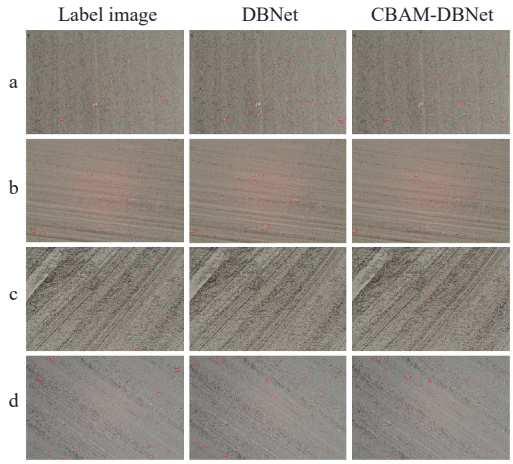
Table 2 Recall, Precision, and F1-score on different methods including DBNet and CBAM-DBNet

Method	Recall	Precision	F1-score
DBNet	83.48%	82.79%	83.13%
CBAM-DBNet	82.69%	85.81%	84.22%

A comparison of the detection results of the two methods is shown in Figure 6, with a total of four original images selected with different soil backgrounds. Overall, both methods can detect most of the RPF. In particular, from Figures 6a-6d, after the CBAM was added to the feature extraction module, the Improved DBNet focused more on deeper features and detected more small objects.

For further comparison, take a portion (1/4) of an image from the test set as an example for observation. After the original image is labeled, there are 16 RPF labeling boxes in Figure 7. The DBNet correctly detected 6 RPFs, but 10 RPFs were missed. Using the CBAM-DBNet, 12 RPFs were detected correctly, but 2 RPFs were detected incorrectly, and 2 RPFs were missed. Compared to Figure 7 (Label Img) and Figure 7 (DBNet), the CBAM-DBNet detected more small objects while improving the detection accuracy. The RPF, however, has also been missed or falsely detected using the aforementioned two methods. At a flight height of 5 m and a camera field of view of 84°, one image represents an area of 38.91 m^2. The size of the RPF was concentrated in tens to hundreds of square centimeters, accounting for a very small proportion of the image[18]. Additionally, a few RPFs were detected

incorrectly or were missing altogether because they were too close to the soil's color and texture.



Note: a, b, c, and d with different background conditions.

Figure 6 Comparison of the detection results on DBNet and CBAM-DBNet for four sets of RPF images

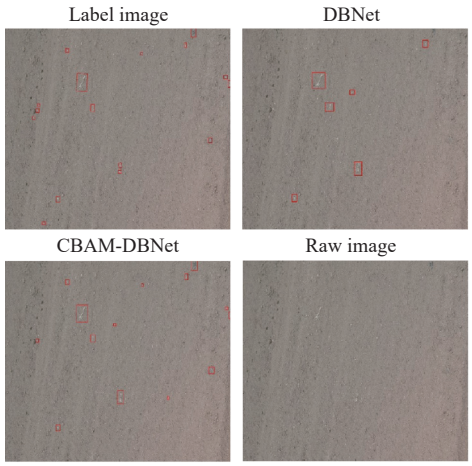


Figure 7 Detailed comparison of the RPF detection results on DBNet and CBAM-DBNet

3.2 RPF image segmentation experiment

The Iterative Threshold method is a common method for calculating the RPF coverage rate[9-11]. The Iterative Threshold method was applied to the test set. The three indicators (MAPE, RMSE, and MAE) used to evaluate the model have the optimal solution under the condition of traversing the global threshold. As shown in Figure 8, the optimal values of the three indicators correspond to a threshold range between 198 and 205 after testing. When the binarization threshold was 204, the MAPE reached the minimum value of 0.817. When the binarization threshold was 199, the RMSE reached the minimum value of 1.132. When the binarization threshold was 201, the MAE reached the minimum value of 0.831. Since these three indicators are evaluated from different perspectives, the optimal solution corresponds to different thresholds[32].

The two methods based on the DBNet and the CBAM-DBNet may offer an adaptive binarization threshold depending on the actual circumstances of each image. This is shown in Figure 9. The DBNet had 80 images with a binarization threshold of 28, 90 images with a threshold of 29, and 5 images with a binarization threshold of 30; the CBAM-DBNet had 134 images with a binarization threshold of 21, 36 images with a binarization threshold of 22, and 5 images with a binarization threshold of 23.

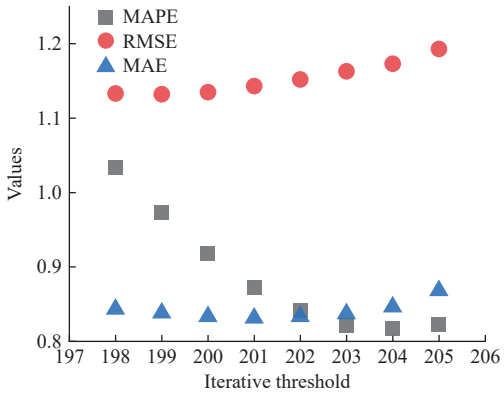


Figure 8 Three model evaluation indicators corresponding to the method iterative threshold

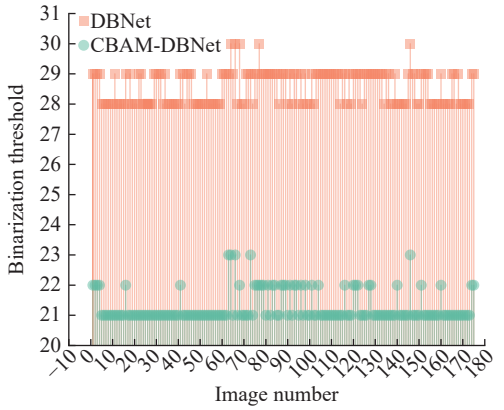


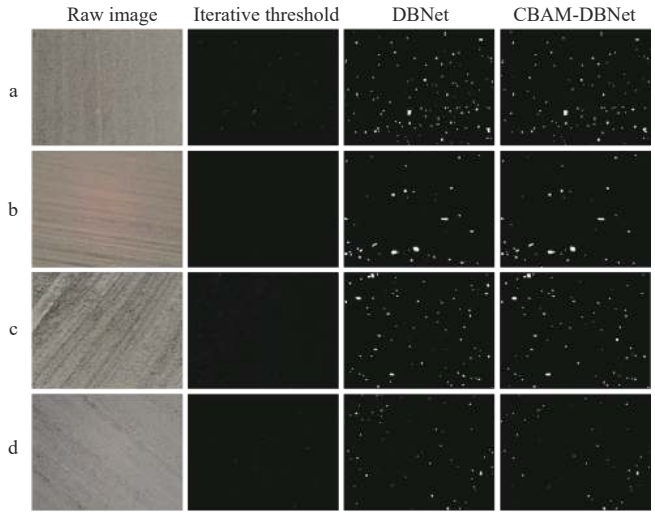
Figure 9 Binary threshold of each image on the test set based on DBNet and CBAM-DBNet

On the test set, a comparison experiment comparing the three approaches was conducted. In Table 3, the specific indication values were displayed. The technique based on the differentiable binarization neural network including DBNet and CBAM-DBNet, outperforms the Iterative Threshold method in forecasting RPF's coverage rate. Regarding the indicator MAPE, both DBNet and CBAM-DBNet exhibited lower values compared to the Iterative Threshold (204) method, with a difference of 0.481 and 0.541, respectively. In terms of the RMSE indicator, methods DBNet and CBAM-DBNet showed lower values compared to the Iterative Threshold (199), with a difference of 0.709 and 0.766, respectively. Similarly, for the MAE indicator, both DBNet and CBAM-DBNet demonstrated lower values compared to the Iterative Threshold method (201), with a difference of 0.214 and 0.226, respectively. The differentiable binarization neural network incorporates the binarization threshold into network training to achieve joint optimization. The optimal binarization threshold for each image may be determined adaptively during the inference phase. Compared with the DBNet, the CBAM-DBNet decreased by 0.06, 0.057, and 0.012 in MAPE, RMSE, and MAE respectively, which proved the validity of the model improvement.

Table 3 MAPE on different methods include Iterative Threshold, DBNet, and CBAM-DBNet

Method	MAPE	RMSE	MAE
Iterative Threshold (199)	0.973	1.132	0.838
Iterative Threshold (201)	0.873	1.143	0.831
Iterative Threshold (204)	0.817	1.173	0.846
DBNet	0.336	0.423	0.617
CBAM-DBNet	0.276	0.366	0.605

For further comparison, there were four test images with different soil backgrounds. The binary mapping of the RPF is shown in Figure 10. For example, in sub-image a, the label value of the RPF's coverage rate of the original image was 1.85%; the Iterative Threshold method calculated a coverage rate of 0.72%; the DBNet calculated a coverage rate of 1.45%; and the CBAM-DBNet calculated a coverage rate of 1.74%. It could be seen that the effect based on CBAM-DBNet was the best. In sub-image b, the label value of the RPF's coverage rate of the original image was 1.1%. The coverage rate calculated based on the Iterative Threshold, DBNet, and CBAM-DBNet methods was 0.29%, 0.85%, and 1.12%, respectively. The best result is based on the CBAM-DBNet, which found more fine RPF in the image. The method based on Iterative Threshold was to binarize all images with the optimal threshold value selected for the test set as a whole and then calculate the coverage rate with this fixed threshold value, so the method could not adapt well to each different image, resulting in a large deviation of the calculated RPF's coverage rate compared with the labeled value.



Note: a, b, c, and d with different soil backgrounds.

Figure 10 Comparison of the binarization results on methods Iterative Threshold, DBNet, and CBAM-DBNet for four sets of RPF images

4 Conclusions

This study developed a threshold-adaptive joint framework based on CBAM-DBNet for residual plastic film (RPF) image detection and segmentation. By introducing DBNet into the field of RPF image processing, the binarization threshold could be optimized in conjunction with the network, allowing for the adaptive selection of an appropriate binarization threshold based on the image during the prediction stage. The main conclusions are as follows:

1) The integration of the convolutional block attention module (CBAM) into the differentiable binarization network (DBNet) obtained richer semantic information and extracted more adequate features. The detection precision of CBAM-DBNet was 3.02 percentage points higher than that of DBNet;

2) In the RPF image detection experiment, the P value of CBAM-DBNet was 85.81%, the value of R was 82.69%, and the F1 value was 84.22%, which was 1.09 percentage points higher than the original DBNet in the comprehensive index F1 value;

3) The binarization threshold of each image could be adaptively obtained based on the CBAM-DBNet;

4) In the RPF image segmentation experiment, the CBAM-DBNet was also better than methods DBNet and Iterative Threshold. Using the CBAM-DBNet to establish an RPF's coverage rate prediction model with MAPE, RMSE, and MAE of 0.276, 0.366, and 0.605, respectively;

5) The results indicated that robust, accurate, and fast prediction of RPF pollution can be achieved based on UAV imaging and the CBAM-DBNet.

In the future, we will collect multiple types of residual plastic film image data to improve the robustness and adaptability of the model.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (Grant No. 32060288); the National Natural Science Foundation of China (Grant No. 32160300); the Bingtuan Science and Technology Program (Grant No. 2019AB007); the Science and Technology Planning Project of the first division of Alaer city (Grant No. 2022XX06).

[References]

[1] Sun D B, Li H G, Wang E L, He W Q, Hao W P, Yan C R, et al. An overview of the use of plastic-film mulching in China to increase crop yield and water-use efficiency. National Science Review, 2020; 7(10): 1523–1526.

[2] Hu C, Lu B, Guo W S, Tang X Y, Wang X F, Xue Y H, et al. Distribution of microplastics in mulched soil in Xinjiang, China. Int J Agric & Biol Eng, 2021; 14(2): 196–204.

[3] Hu C, Wang X F, Chen X G, Tang X Y, Zhao Y, Yan C R, Current situation and control strategies of residual film pollution in Xinjiang. Transactions of the CSAE, 2019; 35(24): 223–234. (in Chinese)

[4] Hu C, Wang X F, Wang S G, Lu B, Guo W S, Liu C J, et al. Impact of agricultural residual plastic film on the growth and yield of drip-irrigated cotton in arid region of Xinjiang, China. Int J Agric & Biol Eng, 2020; 13(1): 160–169.

[5] Zhao Y, Chen X G, Wen H J, Zheng X, Niu Q, Kang J M. Research status and prospect of control technology for residual plastic film pollution in farmland. Transactions of the CSAM, 2017; 48(6): 1–14. (in Chinese)

[6] Niu R K, Wang X F, Hu C, Hou S L, Lu B, Li J B. Analysis of the current situations of plastic films residue pollution of cotton field in Xinjiang Aksu Area. Xinjiang Agricultural Sciences, 2016; 53(2): 283–288. (in Chinese)

[7] Zhang D, Liu H B, Hu W L, Qin X H, Ma X W, Yan C R, et al. The status and distribution characteristics of residual mulching film in Xinjiang. Journal of Integrative Agriculture, 2016; 15(11): 2639–2646.

[8] He H J, Wang Z H, Guo L, Zheng X R, Zhang J Z, Li W H, et al. Distribution characteristics of residual film over a cotton field under long-term film mulching and drip irrigation in an oasis agroecosystem. Soil and Tillage Research, 2018; 180: 194–203.

[9] Liang C J, Wu X M, Wang F, Song Z J, Zhang F G. Research on recognition algorithm of field mulch film based on unmanned aerial vehicle. Acta Agriculturae Zhejiangensis, 2019; 31(6): 1005–1011. (in Chinese)

[10] Zhu X F, Li S B, Xiao G F. Method on extraction of area and distribution of plastic-mulched farmland based on UAV images. Transactions of the CSAE, 2019; 35(4): 106–113. (in Chinese)

[11] Wu X M, Liang C J, Zhang D B, Yu L H, Zhang F G. Identification method of plastic film residue based on UAV remote sensing images. Transactions of the CSAM, 2020; 51(8): 189–195. (in Chinese)

[12] Jiang S Q, Zhang H D, Hua Y J. Research on location of residual plastic film based on computer vision. Journal of Chinese Agricultural Mechanization, 2016; 37(11): 150–154. (in Chinese)

[13] Yang K, Liu H, Wang P, Meng Z J, Chen J P. Convolutional neural network-based automatic image recognition for agricultural machinery. Int J Agric & Biol Eng, 2018; 11(4): 200–206.

[14] Zhao J, Pan F J, Li Z M, Lan Y B, Lu L Q, Yang D J, et al. Detection of cotton waterlogging stress based on hyperspectral images and convolutional neural network. Int J Agric & Biol Eng, 2021; 14(2):

167–174.

[15] Ma Z, Wang Y, Zhang T S, Wang H G, Jia Y J, Gao R, et al. Maize leaf disease identification using deep transfer convolutional neural networks. Int J Agric & Biol Eng, 2022; 15(5): 187–195.

[16] Zhang X J, Huang S, Jin W, Yan J S, Shi Z L, Zhou X C, et al. Identification method of agricultural film residue based on improved faster R-CNN. Journal of Hunan University (Natural Sciences), 2021; 48(8): 161–168. (in Chinese)

[17] Zhai Z Q, Chen X G, Qiu F S, Meng Q J, Wang H Y, Zhang R Y. Detecting surface residual film coverage rate in pre-sowing cotton fields using pixel block and machine learning. Transactions of the CSAE, 2022; 38(6): 140–147. (in Chinese)

[18] Zhai Z Q, Chen X G, Zhang R Y, Qiu F S, Meng Q J, Yang J K, et al. Evaluation of residual plastic film pollution in pre-sowing cotton field using UAV imaging and semantic segmentation. Frontiers in Plant Science, 2022; 13: 991191.

[19] Qiu F S, Zhai Z Q, Li Y L, Yang J K, Wang H Y, Zhang R Y. UAV imaging and deep learning based method for predicting residual film in cotton field plough layer. Frontiers in Plant Science, 2022; 13: 1010474.

[20] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2012; 60(6): 84–90.

[21] Burgos-Artizzu X P, Ribeiro A, Guijarro M, Pajares G. Real-time image processing for crop/weed discrimination in maize fields. Computers and Electronics in Agriculture, 2011; 75(2): 337–346.

[22] Liao M H, Wan Z Y, Yao C, Chen K, Bai X. Real-time scene text detection with differentiable binarization. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2019; 34(7): 11474–11481.

[23] Woo S, Park J, Lee J-Y, Kweno I S. CBAM: Convolutional Block Attention Module. In: Proceedings of the European Conference on Computer Vision – ECCV 2018, 2018; pp.3-19. doi: 10.1007/978-3-030-01234-2_1.

[24] Lin T-Y, Dollar P, Girishick R, He K M, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu: IEEE, 2017; pp.936-944.

[25] Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas: IEEE, 2016; pp.761-769.

[26] He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas: IEEE, 2016; pp.770–778. doi:10.1109/CVPR.2016.90.

[27] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus: IEEE, 2014; pp.580–587. doi: 10.1109/cvpr.2014.81.

[28] Sun F, Tian X D. Lecture video automatic summarization system based on DBNet and Kalman filtering. Mathematical Problems in Engineering, 2022; 2022(1): 5303503.

[29] Vatti B R. A generic solution to polygon clipping. Communications of the ACM, 1992; 35(7): 56–63.

[30] De Myttenaere A, Golden B, Le Grand B, Rossi Fabrice. Mean absolute percentage error for regression models. Neurocomputing, 2016; 192: 38–48.

[31] Karunasingha D S K. Root mean square error or mean absolute error? Use their ratio as well. Information Sciences, 2022; 585: 609–629.

[32] Willmott C J, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 2005; 30(1): 79–82.