Research on Field Soybean Weed Identification Based on an Improved UNet Model Combined With a Channel Attention Mechanism

Helong Yu[1], Zhibo Men[1], Chunguang Bi[1]* and Huanjun Liu[2]*

[1]College of Information Technology, Jilin Agricultural University, Changchun, China, [2]Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China

Aiming at the problem that it is difficult to identify two types of weeds, grass weeds and broadleaf weeds, in complex field environments, this paper proposes a semantic segmentation method with an improved UNet structure and an embedded channel attention mechanism SE module. First, to eliminate the semantic gap between low-dimensional semantic features and high-dimensional semantic features, the UNet model structure is modified according to the characteristics of different types of weeds, and the feature maps after the first five down sampling tasks are restored to the same original image through the deconvolution layer. Hence, the final feature map used for prediction is obtained by the fusion of the upsampling feature map and the feature maps containing more low-dimensional semantic information in the first five layers. In addition, ResNet34 is used as the backbone network, and the channel attention mechanism SE module is embedded to improve useful features. The channel weight is determined, noise is suppressed, soybean and grass weeds are identified, and broadleaf weeds are extracted through digital image morphological processing, and segmented images of soybean plants, grass weeds and broadleaf weeds are generated. Moreover, compared with the standard semantic segmentation models, FCN, UNet, and SegNet, the experimental results show that the overall performance of the model in this paper is the best. The average intersection ratio and average pixel recognition rate in a complex field environment are 0.9282 and 96.11%, respectively. On the basis of weed classification, the identified weeds are further refined into two types of weeds to provide technical support for intelligent precision variable weed spraying.

Keywords: semantic segmentation, weed recognition, feature fusion, channel attention mechanism, improved UNet model

INTRODUCTION

Weeds are one of the main reasons for crop yield and quality decline (Hamuda et al., 2016). They compete with crops in a field (such as soybeans) for sunlight, water, nutrients, and living space (Hongbo et al., 2020). At present, the main cleaning and control method for weeds is to spray herbicides on a large area (Ma et al., 2011), but this method not only causes much pesticide

waste but also damages the ecological environment (Rodrigo et al., 2014), affects the quality of crops, and threatens human health (Yue et al., 2016). Weeds can be divided into two types, grass weeds and broadleaf weeds, according to the shape of their leaves (Javier Herrera et al., 2014). If different herbicides are applied to specific types of weeds, the weed control effect will be better. Grass weeds are monocotyledonous plants, their embryos have one cotyledon, their leaves are usually narrow and long, they have parallel veins, they do not have petioles, their leaf sheaths are open, and they have ligules (Li et al., 2015). Broadleaf weeds are also called dicotyledonous weeds. Their embryos have two cotyledons, they are herbaceous or woody, and they have reticulated veins and wide leaves (Hong-jun et al., 2009). Therefore, the rapid and accurate identification of different types of weeds is very important for the subsequent precise variable weed spraying. At present, machine vision recognition has been widely used in weed identification because of its advantages of fast and easy operation and noncontact and nondestructive target detection abilities (Wang et al., 2019). Machine vision recognition methods are mainly based on the different characteristic information of weeds and crops in images, such as color (Wenhua et al., 2009; Han et al., 2016), shape (Li et al., 2010), and texture (Hongyan and Jixing, 2014; Bakhshipour et al., 2017), to realize the identification of weeds and crops. Based on traditional machine learning algorithms, due to the limitations that arise when using a single feature in the process of weed identification, and to further improve the robustness of identification, it is necessary to fuse different features for identification (Ahmed et al., 2012; He et al., 2013; Zhao and Wei, 2014; Wang and Li, 2016). Although the above research has achieved certain results in weed identification, these methods rely on the manual selection of features and can only identify weeds satisfactorily under specific circumstances; they cannot identify weeds in complex and changeable field environments. Efficient extraction methods are less practical in addressing the problem of weed identification. In recent years, with the development of deep learning, convolutional neural networks have been used to automatically extract the deep features of weeds from images because they do not need to rely on a designer's experience to select features, which has become a research hotspot. There are two main methods for target recognition after extracting image features based on deep convolutional neural networks. One is to draw a rectangular detection frame around a recognized target, and the other is to perform pixel-level classification on a recognized target (Asad and Bais, 2019). In the target detection method of drawing a rectangular detection frame around a recognition target (Peng et al., 2019; Meng et al., 2020; Jin et al., 2021), since the detection frame also contains a large area of background in addition to the target weeds, it is impossible to accurately distinguish the two types of weeds in the image, which affects the subsequent precise variable weed spraying. In view of the above problems, this paper uses semantic segmentation to study weed identification. Semantic segmentation predicts the category of each pixel in an image according to the prelabeled category (Jiaxing and Yujie, 2021) with an improved UNet structure and embeds the channel attention squeeze and excitation (SE) module to perform pixel-level segmentation on grass weeds and soybeans in an image. After identifying the

grassy weeds and soybean pixel categories, other green pixels in the image are extracted by the digital image morphological processing method to extract broadleaf weeds to achieve the goal of simultaneously identifying two types of weeds, which is the next step. Intelligent equipment provides a reference for weed identification in variable pesticide spraying.

TEST DATA AND PREPROCESSING

Image Acquisition

The soybean and weed image data used in the experiment were collected from the soybean experimental field of Jilin Agricultural University. Through manual intervention at an early stage, the images collected showed the growth status of soybeans; the dataset included images with soybeans alone, soybeans and grassy weeds, and soybeans and broadleaf weeds. Leafy weeds are associated with soybeans, grasses, and broadleaf weeds. The collection time was in mid-June 2021, and the time period was from 9:00 to 15:00. In order to ensure that the weed dataset can fully reflect the complexity of the natural environment, we selected different weather conditions such as sunny and cloudy days for image acquisition. The device used was a Huawei mate30 mobile phone positioned vertical to the ground at 60 cm above the ground, and the resolution was $3,000 \times 4,000$ pixels. The images are in JPG format.

Image Preprocessing

The weed data of the object of this study were collected from a natural environment in a field. The background of a collected image mainly includes soil, stones, straw, and plant residue. To suppress the complexity of the field environment, the identification target and the background are segmented, and only the plants in the image are extracted. The part of interest is retained, thereby increasing the accuracy and efficiency of the weed identification model. The ultragreen feature algorithm (2g-rb) is selected to increase the weight of the green channel in the image, improve the contrast with the nongreen background part, and extract the green crop and weed information (Zhao et al., 2009). First, 2g-rb is used to obtain the grayscale image and histogram, and then the Otsu method (OTSU) is used to binarize the grayscale image of 2g-rb to obtain a binary image. This is used as a mask to obtain a green crop image, as shown in **Figure 1**.

Dataset Construction

To enrich the training set of image data, better extract the features of the image, and prevent the built model from overfitting, data enhancement was performed on the 700 images collected in this experiment through operations such as rotations and color transformations. After data enhancement, the number of samples is expanded by five times for a total of 4,200 images. A total of 3,600 images were randomly selected for the training set, 400 images were selected for the validation set, and 200 images were selected for the test set. The labelme tool was used to manually label the image and label the identified objects, such as soybeans and grass weeds, as shown
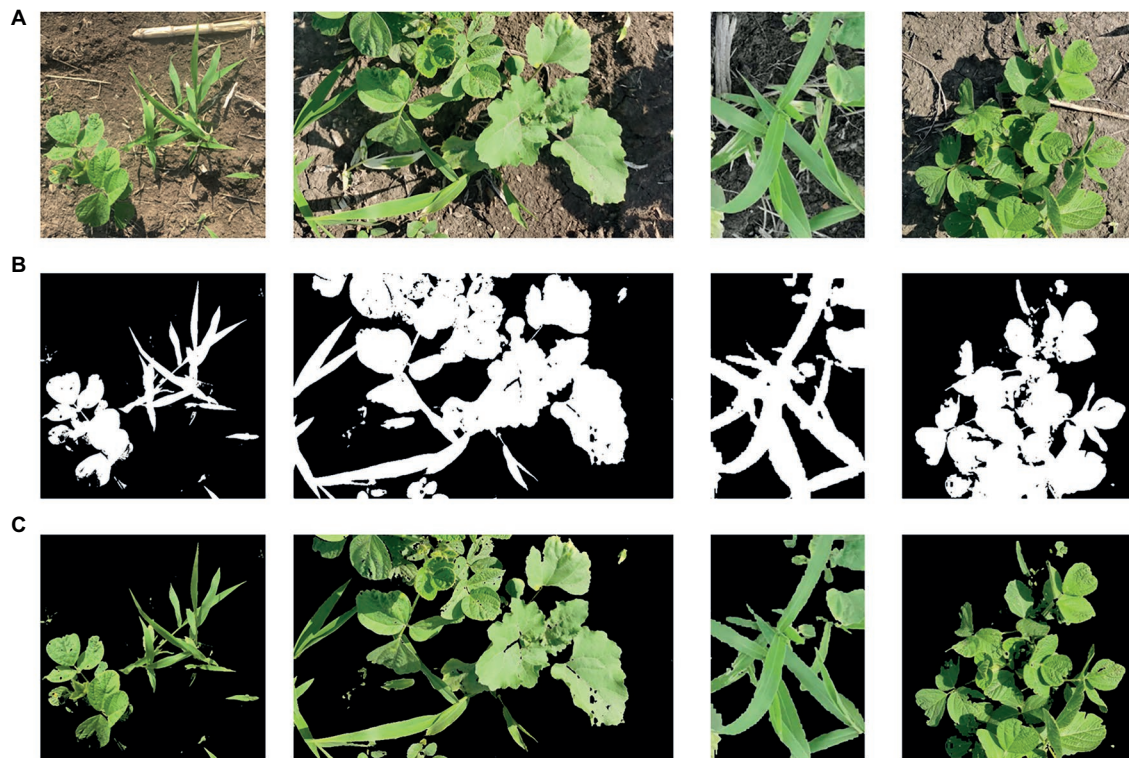
FIGURE 1 | Hyper green feature algorithm **(A)** RGB original image, **(B)** Binary image, and **(C)** Final image.

FIGURE 2 | Original image annotation.

in **Figure 2**. The background (black) has an RGB value of [0,0,0], soybeans (green) have an RGB value of [0,128,0], and grass weeds (red) have an RGB value of [128,0,0].

SEMANTIC SEGMENTATION MODEL CONSTRUCTION

Original UNet Model

UNet was proposed by Ronneberger et al. (2015) at the 2015 MICCAI conference; it consists of an encoder (downsampling)-decoder (upsampling) structure, and the encoder stage consists of two 3×3 convolutional layers, a 2×2 maximum pooling layer is formed, and the activation function is a rectified linear unit (ReLU) function. A total of four downsampling operations are performed. After each pooling operation, the size of the feature map is reduced to half of the original size, and the number of channels is doubled. The decoder stage and the encoder part correspond to a total of four upsampling iterations through a 2×2 deconvolution layer (a transposed matrix). Each time the size of the upsampling feature map is doubled, the number of channels is halved. Different from other semantic segmentation networks, UNet combines the feature map obtained in the encoding stage with the feature map obtained in the decoding stage through skip connections to form a thicker feature map. UNet uses skip connections in the same stage. The final features are fused with more shallow features to retain more detailed information and make the segmentation results more refined. The traditional UNet model structure is shown in **Figure 3**.

UNet Model Improvement

The visualization of shallow and deep feature maps is achieved through convolutional neural networks, as shown in **Figure 4**. We selected Resnet34 to extract the feature map of weeds and realize visualization. Some feature maps selected from different layers of visualization can summarize a little rule. The shallow network extracts texture and detail features, and the deep network extracts features such as contour and shape.

FIGURE 3 | Original UNet model structure.
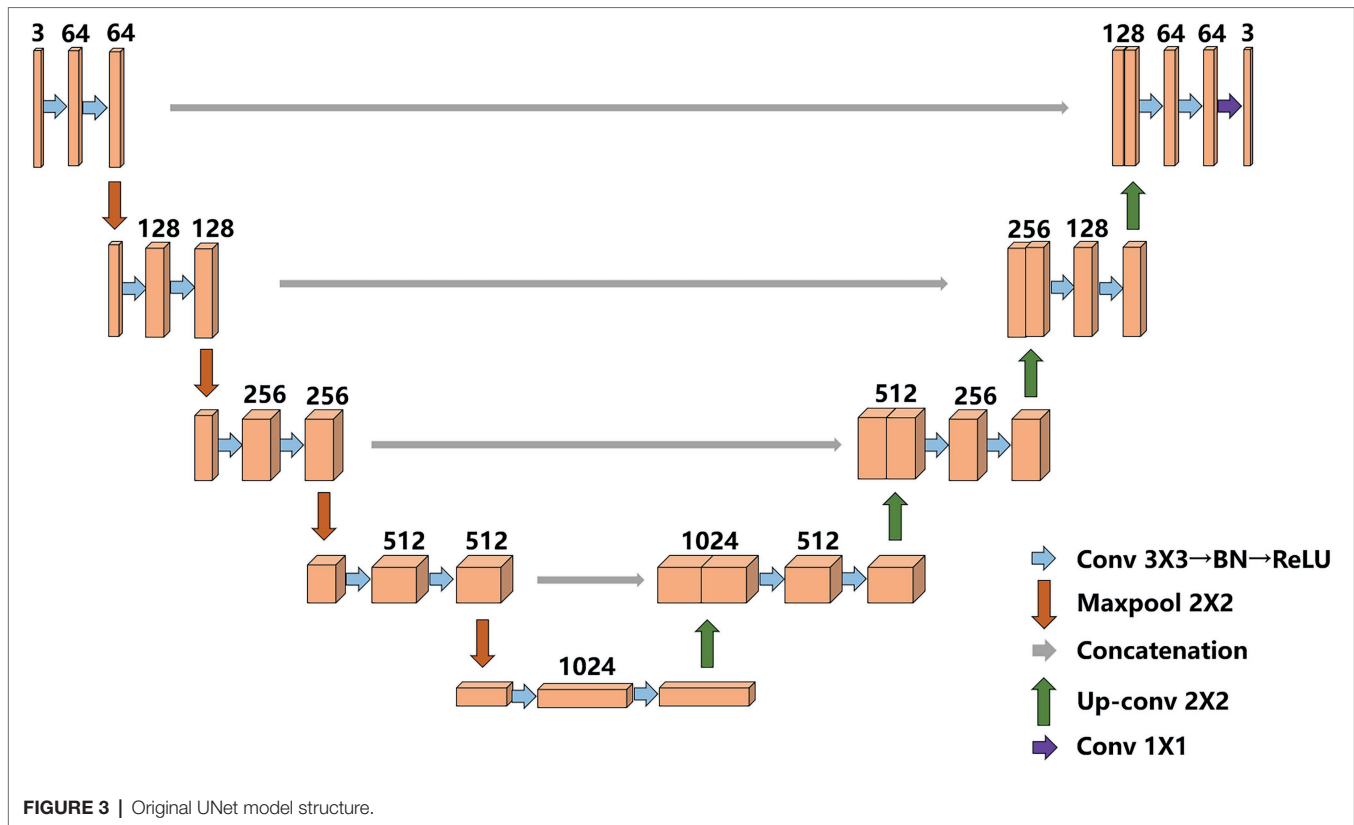
Shallow networks contain more features and are also capable of extracting key features, such as the elongated shape of grass weeds. Relatively speaking, the deeper the number of layers, the more representative the extracted features, and as the number of layers increases, the resolution of the image becomes smaller and smaller. Since the soybeans and weeds to be segmented are similar in color, the main difference is in their shape and leaf area. The low-dimensional semantic information contained in the shallow feature map obtained in the encoding stage is the same as the high-dimensional semantic information contained in the deep feature map obtained in the decoding stage. There is a semantic gap when information is combined through skip connections. The final feature map used for prediction will lose some low-dimensional feature information, and the loss of these low-dimensional feature information will affect the segmentation accuracy of the weed boundary.

To eliminate the semantic gap, after the original image feature map is extracted by UNet network, our model builds a shallow feature fusion structure. After downsampling of original image, the feature map is upsampled by a 2×2 deconvolution layer (transposed matrix) for eight/six/four/ two times, and the size of feature maps are restored to the original size, the feature maps are operated by Conv1×1, respectively, to realize cross-channel information interaction and integration, and reduce the latitude of the number of channels in the feature map. Then, the obtained feature map is concat connected with the final feature map. Finally, the

feature map used for prediction is upsampled, and the contain features with more low-dimensional information. The shallow layers of feature maps are used because they contain more low-dimensional semantic information, such as shape and texture. The multilayer feature maps are combined because they have different receptive fields and can adapt to targets in different areas. The improved UNet model is shown in **Figure 5**.

Backbone Network

Convolutional neural networks (Ronneberger et al., 2015; Jingjing et al., 2020; Zhang et al., 2020) have been widely used in image classification, target detection, semantic segmentation, and other tasks. The depth of a convolutional neural network has a great influence on subsequent recognition and classification. However, as the network depth is gradually increased, the phenomenon of gradient disappearance becomes increasingly obvious, and the network training performance appears to degrade. To solve the above problems, this paper adopts a ResNet (He et al., 2016) model as the backbone network to extract image features and introduces the residual block structure. Compared with ResNet18, ResNet34 is a deeper layer network and can learn more about identifying target features; it also has fewer parameters than ResNet50. Therefore, this paper selects ResNet34 as the backbone extraction network. **Figure 6** shows the basic residual block structure of ResNet34 and the embedded SE module.
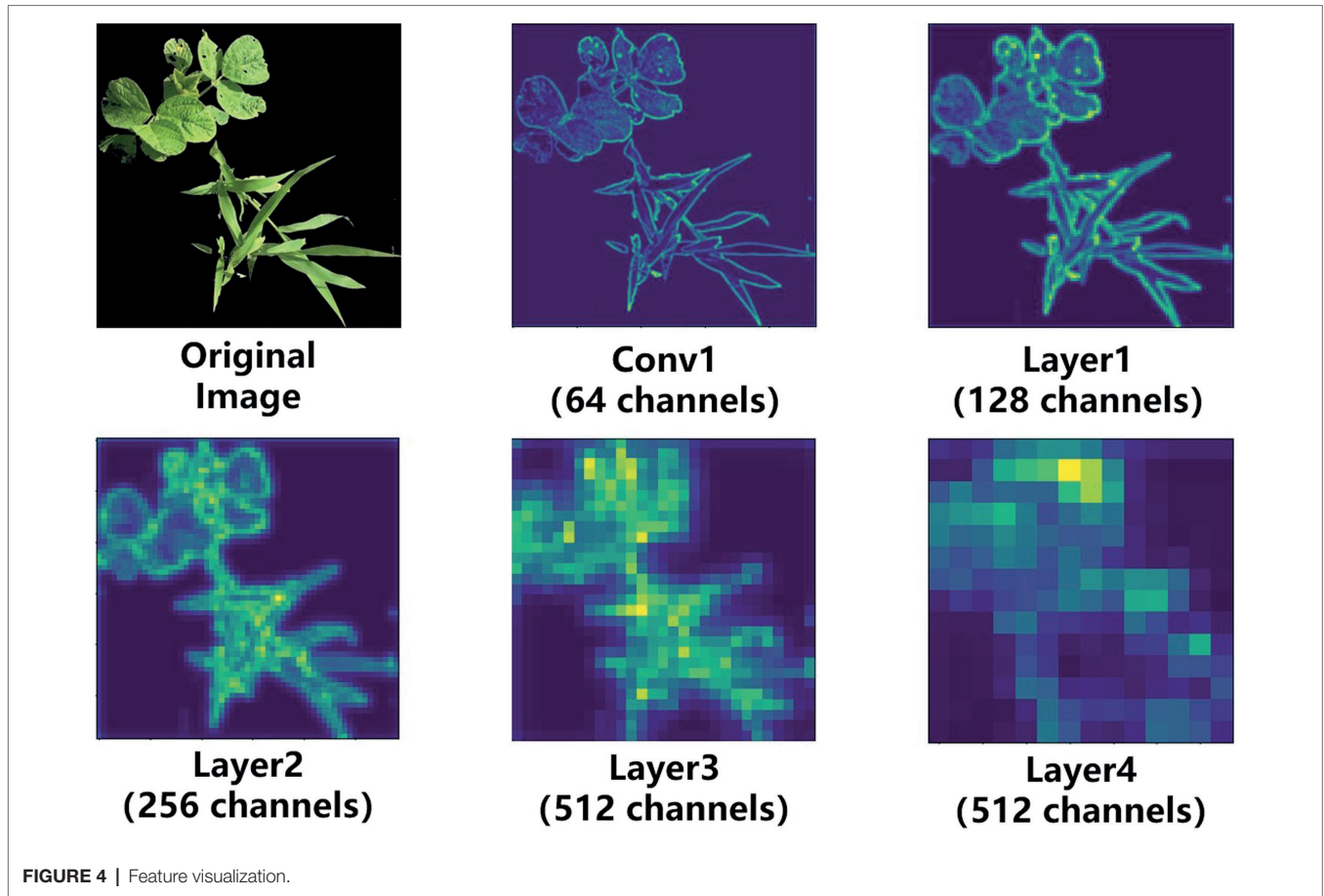
FIGURE 4 | Feature visualization.

Attention Mechanisms

Since the data in this paper are collected in a complex natural field environment, there is considerable noise in the feature extraction process. To overcome the propagation of noise information in the feature channel, the attention mechanism SE module is introduced into the residual module (Hu et al., 2018). The SE module can explicitly model the interdependence between feature channels, thereby adjusting the weight of each channel; suppressing the weights of feature channels that are not related to the recognition target, such as noise channels; and enhancing the useful feature weights. A schematic diagram of the SE module structure is shown in **Figure 7** and includes a squeeze module (squeeze) and an excitation module (excitation).

First, through the squeeze operation, the feature map U is input, the size is W×H, the number of channels is C, the average pooling algorithm is selected so that it has a global receptive field, and the input of W×H×C is generated as 1. The channel descriptor of ×1×C is used for the output, and the extrusion formula is

$$Z_C = F_{sq}(U_c) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} U_c(i,j) \tag{1}$$

where Z_C is the output feature map, F_sq is the extrusion function, and U_c is the input feature map. Then, the excitation (excitation) operation is performed, and the formula is

$$s = F_{ex}(z_C, W) = \sigma(g(z,W)) = \sigma(W_2 \delta(W_1 z_C)) \tag{2}$$

where S is the channel weight parameter, F_{ex} is the excitation function, σ is the ReLU activation function, and δ is the sigmoid activation function. First, through a fully connected layer operation, $W_1 \in R^{\frac{c}{r} \times c}$, where r is a compression parameter, which is used to reduce the number of channels, the number of parameters, and the complexity of the model. Then, the ReLU activation function for nonlinearization. Then, through a fully connected layer operation, $W_2 \in R^{c \times \frac{c}{r}}$, the number of channels is restored. Finally, the channel weight coefficient is calculated by the sigmoid activation function and multiplied by the original feature to realize the recalibration of the feature. The formula is

$$\widetilde{x_c} = F_{scale}(u_c, s_c) = s_c u_c \tag{3}$$

where $\widetilde{x_c}$ is the adjusted C-th channel feature, and F_{scale} is the product function of the C-th channel weight parameter s_c and the input feature map u_c.

FIGURE 5 | Improved UNet.
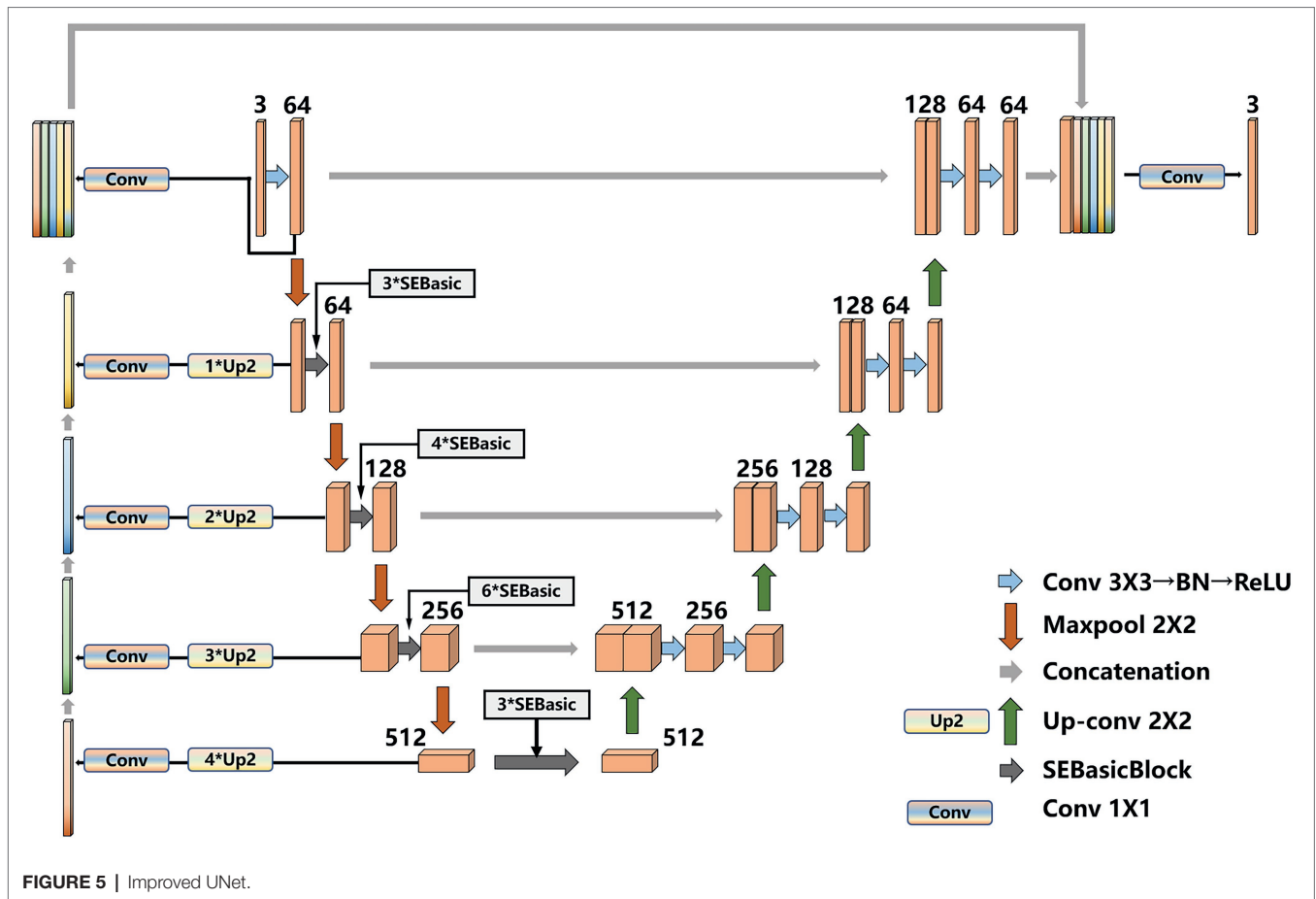
Test Platform

The test platform is equipped with an Intel® Core™ i7-9800X CPU clocked at 3.8 GHz, with eight cores, 16 threads, 64 GB memory, and an NVIDIA RTX2080Ti GPU with 11 GB video memory. A Windows 10 operating system, the Python programming language, a Pycharm-integrated development environment, and a PyTorch deep learning framework are applied.

Data Augmentation During Training

Due to the limited number of training samples, dataset enhancement is also required during training. During training, the original image and label are randomly cropped, rotated, blurred, and randomly moved by the RGB channel, as shown in the figure, and the obtained image is adjusted to the network input size of 512×512 pixels and sent to the network for training. It can adapt to images and recognition targets of different scale transformations, which is due to the increases generalization ability of the network model (**Figure 8**).

Model Training

Deep learning model training requires a large amount of sample data to make the model perform better. The amount of weed data collected in this paper is small, and it takes considerable
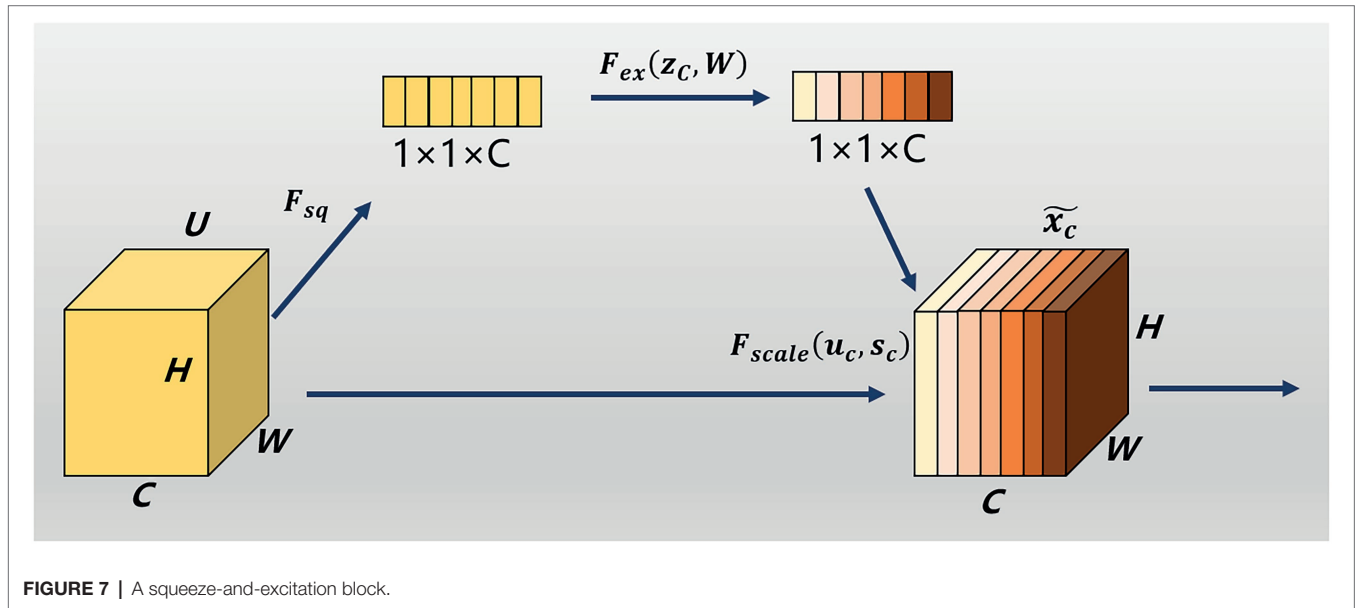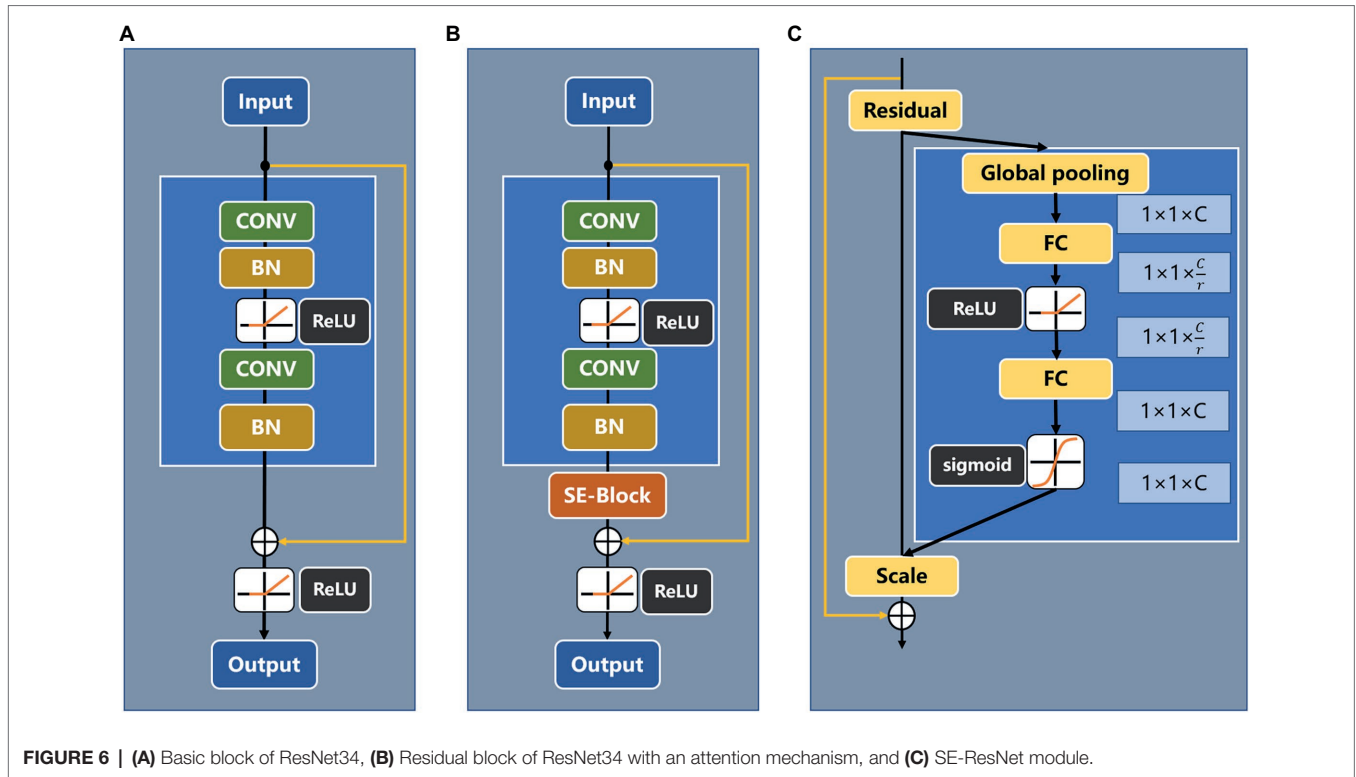
time and effort to label the data. To improve the generalization ability of the network, the transfer learning method is adopted. To train the model, the backbone network trains the pretrained ResNet34 model on the ImageNet dataset as an initialization parameter. The number of model training epochs is 500 epochs, and the batch training size is 8. The learning rate is dynamically adjusted. The initial learning rate is 0.01. As the number of iterations increases, the learning rate gradually decreases, with a minimum of 0.001. The momentum factor is 0.9, the Adam optimization algorithm is used, and the cross-entropy loss function is used.

Evaluation Indicators

To prove the effectiveness of the semantic segmentation method, it is necessary to strictly evaluate the semantic segmentation method. The evaluation indicators used in this paper mainly include the mean pixel accuracy (mPA) and the mean intersection over union (mIoU).

The mPA metric refers to the average pixel accuracy of each category, and the calculation formula is

$$mPA = \frac{1}{3}\left(\frac{TP}{TP+FP} + \frac{TN}{TN+FN} \right) \qquad (4)$$

FIGURE 6 | **(A)** Basic block of ResNet34, **(B)** Residual block of ResNet34 with an attention mechanism, and **(C)** SE-ResNet module.

FIGURE 7 | A squeeze-and-excitation block.

The mIoU ratio refers to the ratio of the intersection and union of the numbers of pixels in the predicted area and the real area. mIoU is the most commonly used metric due to its simplicity and strong representation and is used by most researchers. to report results. The calculation formula is

$$mIoU = \frac{1}{3}\left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FN + FP} \right) \quad (5)$$

where true positive (TP) represents a true example; that is, in the prediction of category i, TP is the number of pixels predicted as category i. False positive (FP) denotes the number of pixels that are mispredicted as background or the number of pixels of other classes predicted as class i. True negative (TN) denotes the number of pixels that are correctly predicted as background or other category pixels. False negative (FN) denotes pixels that are mispredicted as class i, background, or other class pixels.

FIGURE 8 | Data augmentation during training.
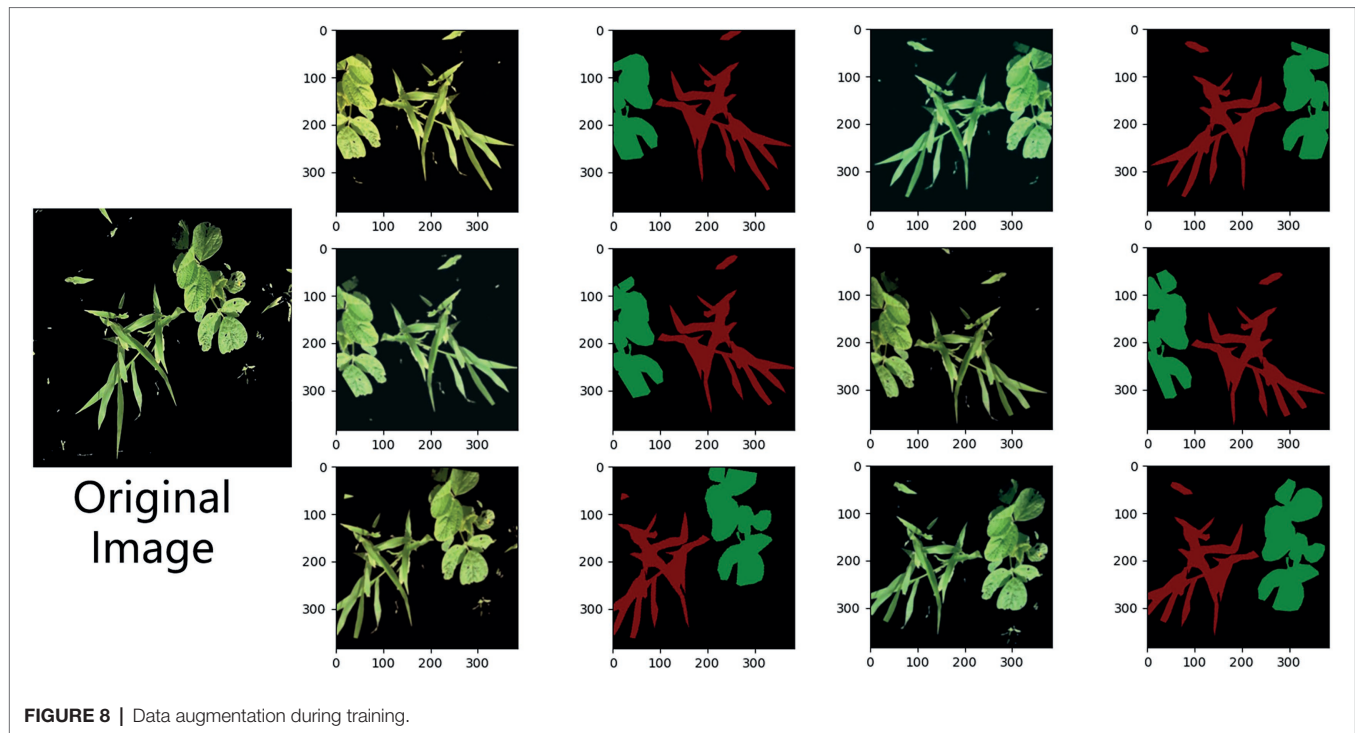
Broadleaf Weed Extraction

Based on the improved UNet model, to realize the identification of soybean and grass weeds, this paper extracts broadleaf weeds through a digital image morphological processing method (Wang et al., 2021). First, the grayscale image obtained by the 2g-rb algorithm is used, and then OTSU is used to binarize the grayscale image of 2g-rb to obtain a binary image. Then, the semantic segmentation model is used to identify the soybeans. After the expansion and inversion of the grassy weed prediction map, the binary image obtained from the original image is added to delete the grassy weeds and soybean parts predicted by the semantic segmentation model, and the remainder are broadleaf weeds. Moreover, the connected domain area filtering operation is performed on the obtained broadleaf weeds to remove small-area connected domains to prevent noise interference, and the broadleaf weeds are marked as (yellow) RGB = [128,128,0]. Finally, the obtained broadleaf weeds, the grass weeds, and the soybeans predicted by semantic segmentation are added to generate the final weed prediction image. The flowchart is shown in **Figure 9**.

RESULTS AND ANALYSIS

Under the same test platform conditions, the FCN, SegNet, UNet, and improved UNet semantic segmentation models were established, and the weed training set was used for training and testing. The loss function change curve of each network architecture is shown in the figure. The training is divided into two stages, namely, the freezing stage and the unfreezing stage. The first 50 epochs freeze the training parameters, and the feature extraction network does not change and occupies a small amount of video memory. Only network fine-tuning is performed. The last 450 epochs unfreeze the training parameters. At this time, the backbone of the model is not frozen, the memory occupied is large, and all the parameters of the network are changed. It can be seen from the figure that each model can achieve a good training process. The changes in the loss function values are basically the same. The loss decreases rapidly in the early stage of training, decreases steadily and slightly in the middle stage, and basically tends to stability in the later stage, and the network converges. At the final iteration, the improved UNet has the smallest loss value of 0.109 (**Figure 10**).

Table 1 shows that the intersection-over-union (IoU) value of the background is the largest among all categories, and the IoU values of soybeans are higher than those of grassy weeds.

In complex field environments, grass weeds are relatively dense and overlap, making it difficult to distinguish grass weeds. All indicators of the traditional UNet model are better than those of FCN and SegNet because UNet uses skip connections to fuse low-dimensional semantic features extracted in the encoding stage with high-dimensional semantic features to obtain multiscale features and achieve better segmentation results. The model in this paper is superior to other semantic segmentation methods in terms of the intersection ratio, average intersection ratio, and average accuracy. Compared with the traditional UNet model, the average intersection ratio is 2.31 percentage points higher, and the average accuracy rate is 2.69 percentage points higher. Compared with other semantic segmentation methods,
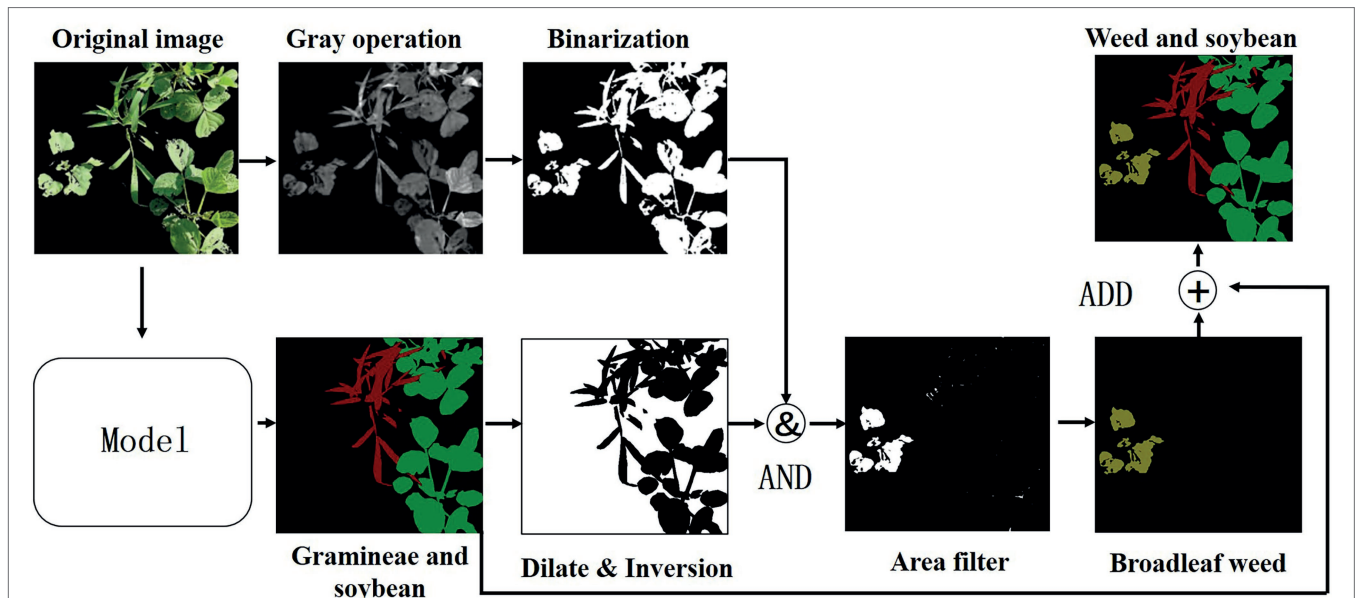
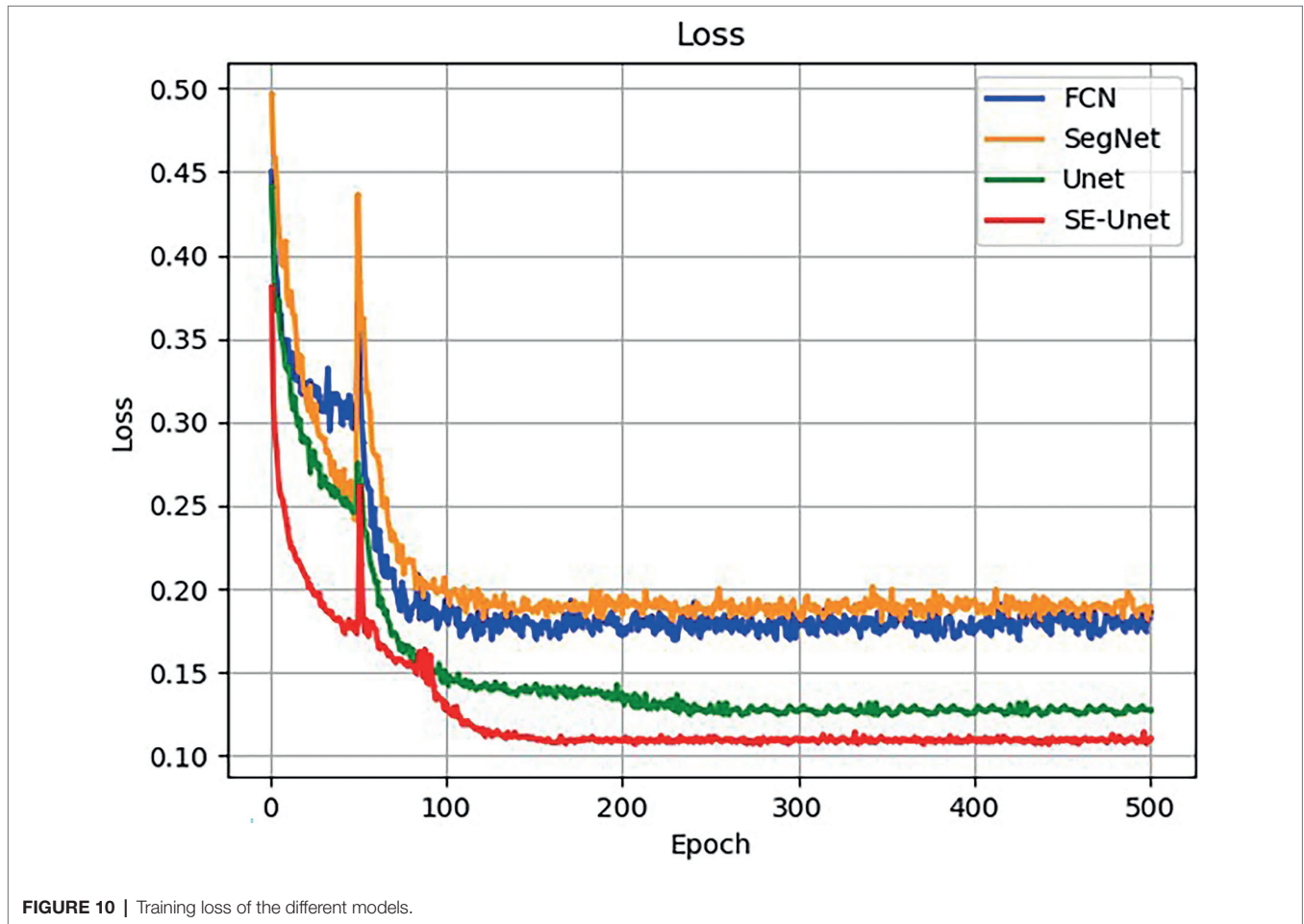**FIGURE 9 |** Broadleaf weed extraction process.

FIGURE 10 | Training loss of the different models.

it has a significant improvement. The confusion matrix of our module is shown in **Figure 11**.

To verify the effectiveness of the method proposed in this paper for weed semantic segmentation, different models are used to test the semantic segmentation results on the same set of images, and the segmentation effect is shown in **Figure 12**.

Through the segmentation effect map, it is found that the segmentation ability of FCN and SegNet is poor, the edges of the segmented weeds and soybean crops are not smooth enough, and the overlapping parts of the soybean crops and weeds are easily segmented into soybeans. The overall contour segmentation effect of UNet is good, but the details cannot be accurately identified, and misjudgment is prone to occur. The channel

attention mechanism SE module and the improved UNet structure designed in this paper can make the model focus more on the detailed segmentation of the overlapping parts of soybeans and weeds, resulting in smoother contours and smoother segmentation renderings of soybeans and weeds with greater detail.
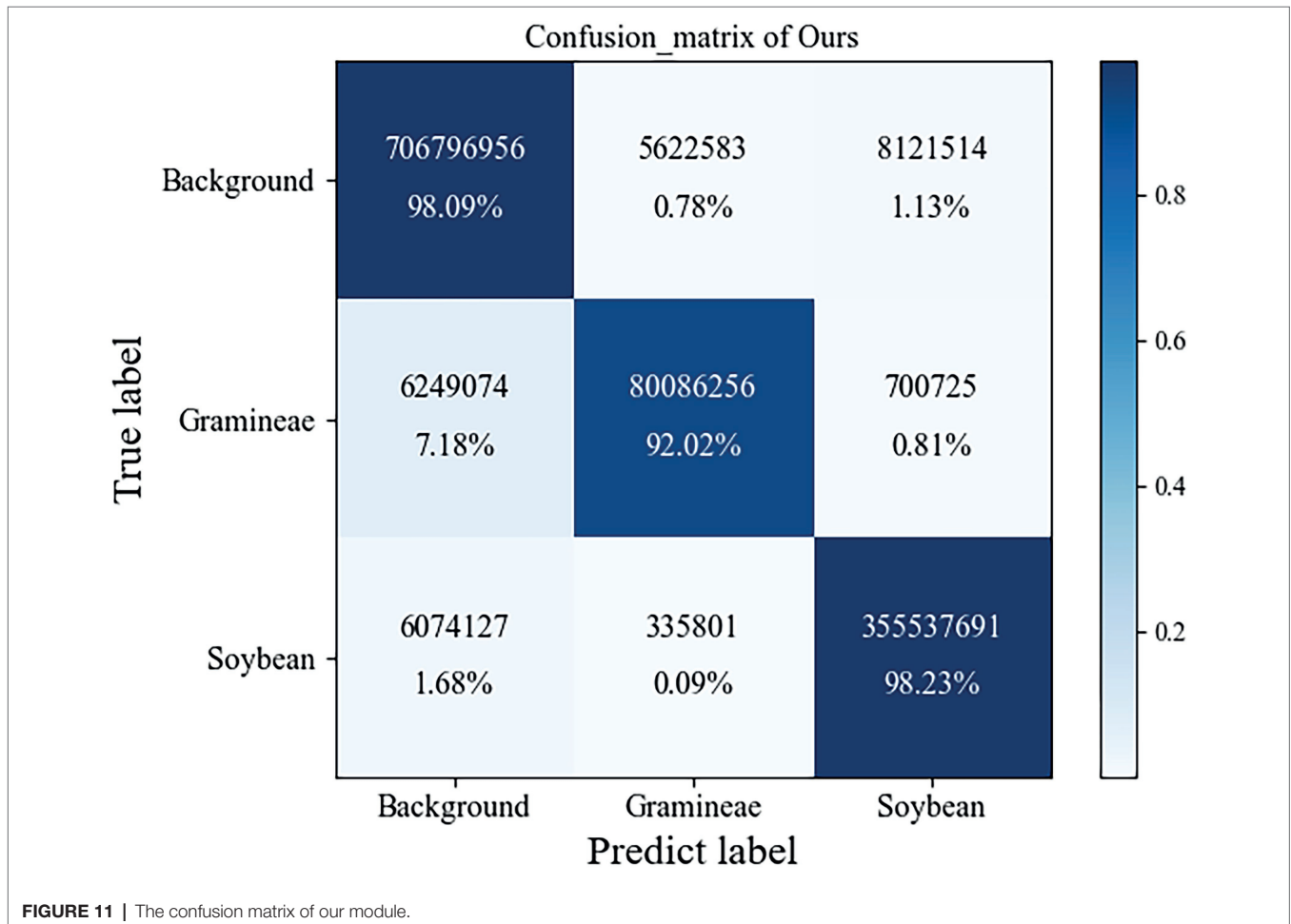
Through the above experiments, the model proposed in this paper is compared with the existing semantic segmentation models in the same experimental environment, which proves that the method in this paper has certain advantages in weed identification.

To evaluate the impact of the various components of the proposed method on the performance of the model, ablation experiments were designed on the soybean weed dataset. After comparing the performance of different models, it is found that UNet has the best segmentation effect. Therefore, UNet is selected as the basic network structure to evaluate the impact of the residual structure with SENet and the improved UNet structure on the performance of the model, as shown in **Table 2**.

Compared with the traditional UNet model, due to the addition of the channel attention mechanism SE module, the weight of each channel is adjusted, the noise and other feature channel weights that are not related to the recognition target are suppressed, and the useful feature weights are enhanced.

TABLE 1 | Performance comparison of the different models.

Model	IoU			mIoU	mPA%
	Grass	Soybean	Background		
FCN	68.31	84.58	88.67	80.52	88.18
SegNet	70.34	91.19	92.28	84.61	91.76
UNet	79.04	92.79	94.06	88.63	93.42
Ours	86.12	95.89	96.44	92.82	96.11

FIGURE 11 | The confusion matrix of our module.

FIGURE 12 | Comparison of the segmentation effects of the different models.
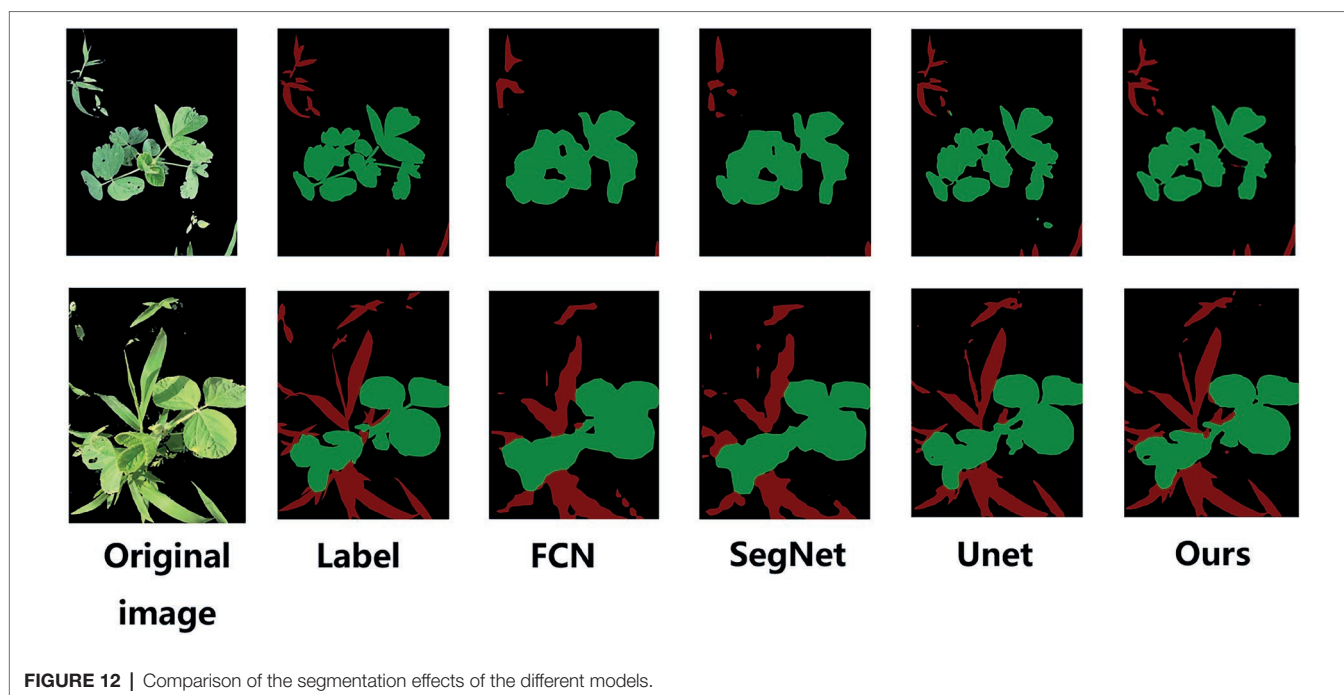
TABLE 2 | Performance comparison of the different models.

Model	mIoU%	mPA%
UNet	88.63	93.42
UNet + SE	89.56	94.31
UNet structure improvement	90.32	94.82
UNet + SE + structure improvement	92.82	96.11

After adding the channel attention mechanism SE module, the average intersection ratio and average accuracy are improved by 0.93 and 0.89 percentage points, respectively.

To eliminate the semantic gap in the improved UNet, the final feature map used for prediction is obtained by the fusion of upsampling and the feature maps that contain more low-dimensional semantic information in the first five layers. The ratio and average accuracy are improved by 1.69 and 1.40 percentage points, respectively, proving the effectiveness of the structure improvement. After adding the improved UNet structure to the SE module of the channel attention mechanism, the average intersection ratio and average accuracy are improved more significantly; they are increased by 4.19 and 2.69 percentage points, respectively. According to the characteristics of soybean shape and area, appropriate shallow features and deep features are selected for fusion to achieve the expected recognition effect. In summary, the improved UNet and embedded channel attention mechanism SE module are effective.

CONCLUSION

To solve the problem of identifying different types of weeds in a natural complex field environment and achieve the goal of spraying different types of herbicides with precise variables to prevent pesticide waste and drug residues, this paper proposes an improved UNet structure and embeds an SE module for soybean weeds in a natural field environment. A grass identification method is applied. Compared with the original UNet model, the overall performance is improved, the low-dimensional semantic features of weeds and soybeans are fully considered in the training samples, appropriate feature layers are selected for fusion and SE modules are embedded to improve the recognition performance; thus, the effectiveness of the method in this paper is proven. The effectiveness of identifying different types of weeds under complex natural conditions in a field provides a valuable reference for subsequent intelligent precision variable spraying and weeding.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

HY, CB, and HL contributed to the conception and design of the study. ZM preprocessed the data. ZM and HY wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Ahmed, F., Al-Mamun, H. A., Bari, A. S. M. H., Hossain, E., and Kwan, P. (2012). Classification of crops and weeds from digital images: A support vector machine approach. *Crop Prot.* 40, 98–104. doi: 10.1016/j.cropro.2012.04.024

Asad, M. H., and Bais, A. (2019). Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. *Info. Process. Agric.* 7, 534–545. doi: 10.1016/j.inpa.2019.12.002

Bakhshipour, A., Jafari, A., Nassiri, S. M., and Zare, D. (2017). Weed segmentation using texture features extracted from wavelet sub-images. *Biosyst. Eng.* 157, 1–12. doi: 10.1016/j.biosystemseng.2017.02.002

Hamuda, E., Glavin, M., and Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Comput. Electron. Agric.* 125, 184–199. doi: 10.1016/j.compag.2016.04.024

Han, D., Wu, P., Zhang, Q., Han, G., and Tong, F. (2016). Feature extraction and image recognition of typical grassland forage based on color moment. *Trans. Chin. Soc. Agric. Eng.* 32, 168–175. doi: 10.11975/j.issn.1002-6819.2016.23.023

He, D., Qiao, Y., Li, P., Gao, Z., Li, H., and Tang, J. (2013). Weed recognition based on SVM-DS multi-feature fusion. *Trans. Chin. Soc. Agric. Machine.* 44, 182–187. doi: 10.6041/j.issn.1000-1298.2013.02.034

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. arXiv [Preprint]. doi: 10.48550/arXiv.1512.03385

Hongbo, Y., Nudong, Z., and Cheng, M. (2020). Review of weeds recognition based on image processing. *Trans. Chin. Soc. Agric. Machine.* 51, 323–334. doi: 10.6041/j.issn.1000-1298.2020.S2.038

Hong-jun, Z., Si-bin, G. U. O., Wen-da, Z. H. U., and Yin-zhong, L. I. (2009). Effect of Clopyralid 75% SG on broad-leaf weed and yield of rape. *J. Huazhong Agric. Univ.* 28, 27–30. doi: 10.3321/j.issn:1000-2421.2009.01.008

Hongyan, W., and Jixing, L. (2014). Identifying corn weed based on texture features and optimized SVM. Hubei. *Agric. Sci.* 53, 3163–3169. doi: 10.3969/j.issn.0439-8114.2014.13.048

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks." in *31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 7132–7141.

Javier Herrera, P., Dorado, J., and Ribeiro, A. (2014). A novel approach for weed type classification based on shape descriptors and a fuzzy decision-making method. *Sensors* 14, 15304–15324. doi: 10.3390/s140815304

Jiaxing, S., and Yujie, L. (2021). Multi-feature fusion network for road scene semantic segmentation. *Comput. Electr. Eng.* 92:107155. doi: 10.1016/j.compeleceng.2021.107155

Jin, X., Che, J., and Chen, Y. (2021). Weed identification using deep learning and image processing in vegetable plantation. *IEEE Access* 9, 10940–10950. doi: 10.1109/ACCESS.2021.3050296

Jingjing, M., Lei, P., Lei, Y., and Jiang, X. (2020). Detection of black spot of rose based on hyperspectral imaging and convolutional neural network. *AgriEngineering* 2, 556–567. doi: 10.3390/agriengineering2040037

Li, G., Wang, X., Zhang, C., Huang, H., and Wei, S. (2015). Effects of rice-straw residues and herbicide clodinafop-propargyl on wheat and Gramineae weeds. *Acta Phytophylacica Sinica* 42, 130–137. doi: 10.1016/j.aoas.2013.01.003

Li, X., Zhu, W., Ji, B., Liu, B., and Ma, C. (2010). Shape feature selection and weed recognition based on image processing and ant colony optimization. *Trans. Chin. Soc. Agric. Eng.* 26, 178–182. doi: 10.3969/j.issn.1002-6819.2010.10.030

Ma, X., Qi, L., Liang, B., Tan, Z., and Zuo, Y. (2011). Present status and prospects of mechanical weeding equipment and technology in paddy field.

Trans. Chin. Soc. Agric. Eng. 27, 162–168. doi: 10.3969/j.issn.1002-6819.2011.06.030

Meng, Q., Zhang, M., Yang, X., Liu, Y., and Zhang, Z. (2020). Recognition of maize seedling and weed based on light weight convolution and feature fusion. *Trans. Chin. Soc. Agric. Machine.* 51, 238–245. doi: 10.6041/j.issn.1000-1298.2020.12.026

Peng, M., Xia, J., and Peng, H. (2019). Efficient recognition of cotton and weed in field based on faster R-CNN by integrating FPN. *Trans. Chin. Soc. Agric. Eng.* 35, 202–209. doi: 10.11975/j.issn.1002-6819.2019.20.025

Rodrigo, M. A., Oturan, N., and Oturan, M. A. (2014). Electrochemically assisted remediation of pesticides in soils and water: a review. *Chem. Rev.* 114, 8720–8745. doi: 10.1021/cr500077e

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation." in *18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI),* 234–241.

Wang, C., and Li, Z. (2016). Weed recognition using SVM model with fusion height and monocular image features. *Trans. Chin. Soc. Agric. Eng.* 32, 165–174. doi: 10.11975/j.issn.1002-6819.2016.15.023

Wang, C., Wu, X., Zhang, Y., and Wang, W. (2021). Recognition and segmentation of maize seedlings in field based on dual attention semantic segmentation network. *Trans. Chin. Soc. Agric. Eng.* 37, 211–221. doi: 10.11975/j.issn.1002-6819.2021.09.024

Wang, A., Zhang, W., and Wei, X. (2019). A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.* 158, 226–240. doi: 10.1016/j.compag.2019.02.005

Wenhua, M., Wang, H., and Zhao, B. (2009). Weed detection method based the centre color of corn seedling. *Trans. CSAE* 25, 161–164. doi: 10.3969/j.issn.1002-6819.2009.z2.030

Yue, Q., Li, J., Bing, Y., Zhenzhen, D., and Gang, F. (2016). Impact of herbicides on wild plant diversity in agro-ecosystems: a review. *Biodivers. Sci.* 24, 228–236. doi: 10.17520/biods.2015208

Zhang, S., Wang, Z., and Wang, Z. (2020). Method for image segmentation of cucumber disease leaves based on multi-scale fusion convolutional neural networks. *Trans. Chin. Soc. Agric. Eng.* 36, 149–157. doi: 10.11975/j.issn.1002-6819.2020.16.019

Zhao, B., Song, Z., Mao, W., Mao, E., and Zhang, X. (2009). Agriculture extra-green image segmentation based on particle swarm optimization and K-means clustering. *Trans. Chin. Soc. Agric. Machine.* 40, 166–169.

Zhao, P., and Wei, X. (2014). Weed recognition in agricultural field using multiple feature fusions. *Trans. Chin. Soc. Agric. Machine.* 45, 275–281. doi: 10.6041/j.issn.1000-1298.2014.03.045

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.