# Film identification method based on improved deeplabv3+ for full-film double-ditch corn seedbed

Fei Dai[1†], Xiangzhou Li[1†], Ruijie Shi[1], Fengwei Zhang[1*], Wuyun Zhao[1], Wenjuan Guo[2*]

(1. *College of Mechanical and Electrical Engineering, Gansu Agricultural University, Lanzhou 730070, China*;
2. *School of Cyber Security, Gansu University of Political Science and Law, Lanzhou 730070, China*)

**Abstract:** This study aimed to investigate the task demand of intelligent unmanned fertilizer application in seedling stage of corn planted in full-film double-ditch seedbed, a film identification method based on improved DeepLabv3+ identification method for full-film double-ditch corn seedbed was proposed. The differences in performance indicators of the original Deeplabv3+ network taking Xception as the backbone network and the network model that replaced three lightweight backbone networks, MobileNetV2, MobileNetV3 and GhostNet were tested. At the same time, the network models, classical semantic segmentation was introduced to PSPNet and UNet for comparative test. The *MIoU* of DeepLabv3+ network model that replaced its backbone network increased by 5.01%, and *FPS* improved by 206% compared with original network, and the model size reduced by 90.3%. The three DeepLabv3+ models after replacing the backbone network were further compressed, and the two-layer expansion convolution with low expansion rate in ASPP was deleted, and the common convolution after feature fusion was replaced by the depthwise separable convolution to obtain a lightweight network model. After testing the improved network model, it was found that the average decline of precision indicators was only 0.17%, *FPS* raised to 66.5, with an average increase of 25.5%, and the size of the model was compressed to 10.53 MB. Test results showed that, the improved model showed excellent performance, and could provide important technology and method support for the research and development of intelligent topdressing and field management on full-film double-ditch corn seedbed during seedling stage.

**Keywords:** fertilizer application, intelligent unmanned machine, full-film double-ditch corn seedbed, film identification method, deep learning, semantic segmentation, DeepLabv3+

**DOI:** 10.25165/j.ijabe.20231605.8288

**Citation:** Dai F, Li X Z, Shi R J, Zhang F W, Zhao W Y, Guo W J. Film identification method based on improved deeplabv3+ for full-film double-ditch corn seedbed. Int J Agric & Biol Eng, 2023; 16(5): 165–172.

## 1 Introduction

The agronomic technologies for full-film double-ditch corn seedbed in the northwest arid area of China have good effects on rainfall collection, moisture conservation, drought resistance and yield increase. It is one of the key dry farming technologies promoted by Gansu Province and a milestone in the development of dry farming[1]. In order to further effectively improve the intelligent whole process mechanization level of full-film double-ditch corn seedbed, and ensure the intelligent control and monitoring in the key operation links, such as ridge forming and film mulching, seeding on film, topdressing under film and residual film recycling[2,3]. It is urgently needed to develop all kinds of intelligent agricultural machinery and equipment adapting to small and scattered plots in hilly area in loess plateau and interaction between seedbed and mulching films. Since such equipment should have highly reliable and adaptable software and hardware, with low cost and easy operation so that it can sense the environment, identify operational object and even make decisions, which are the key points in equipment research and development[4,5].

In recent years, more and more deep learning methods have been applied in agriculture, and semantic segmentation based on deep learning has been widely applied in remote sensing image processing, road recognition, pest identification, and navigation for its pixel-level classification ability[6-8]. The visual identification technology based on deep learning has good robustness in identifying objects with rich textures and complex structures. At present, a large number of recognition tasks have been completed using algorithms related to deep learning[9,10]. With the rapid development of deep learning, especially the deep convolutional neural networks, there are newly emerged network models that apply deep convolutional neural networks in semantic segmentation. Although these network models have achieved good effects in semantic segmentation, they have not been optimized according to specific tasks, thus the model size, quantity of parameters and identification speed cannot satisfy the requirements of specific tasks. Therefore, it is necessary to optimize the original network models. Among them, Sun et al. used deep learning method to identify greenhouses and film covered farmland in UAV aerial images, and obtained higher accuracy and speed compared with traditional classification methods based on pixels and objects[11]. Mu et al.[12] introduced convolutional attention mechanism into the DeepLabv3+ network and replaced the normal revolution in the ASPP module with depthwise separable convolution, and realized

**Received date:** 2023-04-05 **Accepted date:** 2023-08-15

**Biographies: Fei Dai**, PhD, Professor, research interest: design of agricultural mechanization equipment, Email: daifei@gsau.edu.cn; **Xiangzhou Li**, PhD candidate, research interest: computer simulation, Email: 572395298@qq.com; **Ruijie Shi**, Lecturer, research interest: agricultural mechanization engineering, Email: 1139230110@qq.com; **Wuyun Zhao**, PhD, Professor, research interest: farm machine and mechanical reliability, Email: zhaowy@gsau.edu.cn.

†These authors contributed equally to this study.

***Corresponding author: Fengwei Zhang**, PhD, Professor, research interest: farm machine and mechanical reliability. College of Mechanical and Electrical Engineering, Gansu Agricultural University, Lanzhou 730070, China. Tel: +86-451-7631207, Email: zhangfw@gsau.edu.cn; **Wenjuan Guo**, Lecturer, research interest: computer simulation. School of Cyber Security, Gansu University of Political Science and Law, Lanzhou 730070, China. Tel: +86-451-7601487, Email: 565105996@qq.com.

higher identification precision of lodged rice compared with traditional machine learning method. Based on the problems of poor real-time performance and poor universality of interridge navigation path identification, Rao et al.[13] tested the improved Unet model based on small sample datasets of cotton, corn and sugarcane. Yang et al.[14,15] applied CNN to identify corn rhizome for the planning of subsequent paths. Meng et al. used improved MobileNetV2 network to identify non-structured field pavements. However, there are little research on identification of seedbed with mulched films with deep learning methods.

Therefore, the study on the identification method of corn seedbed with full-film double-ditch based on the improved DeepLabv3+ is a good attempt, since it can help implement quality evaluation, precision seeding, precision fertilization under the film, efficient residual film recycling for full-film double-ridge corn seedbed in the later stage, and offers reference for the design and manufacturing of whole-process intelligent mechanized agricultural equipment for full-film double-ridge corn seedbed and field operation.

## 2 Materials and methods

### 2.1 Data collection and processing

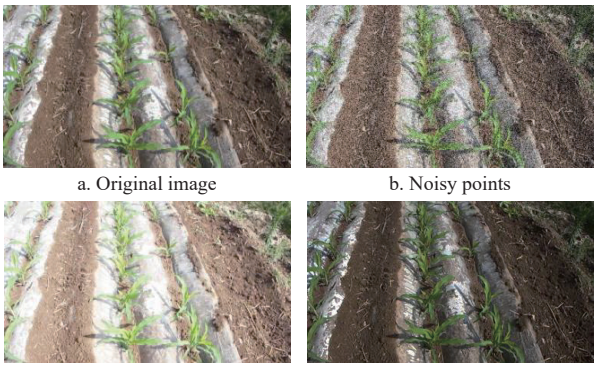
The images of the full-film double-ridge corn seedbed used in this study were acquired in the test field of Gansu Taohe Tractor Manufacturing Co., Ltd, and films on corn seedlings at seedling stage were taken as the object for film identification. The HUAWEI Mate20 Pro mobilephone was used for data collection. During the acquisition process, the automatic mode of the mobile phone camera was chosen, and the AI image enhancement function was turned off, and the image resolution was set to 1280×720 pixels to reduce the time spent in later image processing. The image acquisition process simulates the visual angle of the camera installed on the agricultural machinery. The height of the mobile phone from the shooting object was 1 m, and 45° and 90° were adopted respectively to expand the data set.

During the data acquisition process, the diversity of data samples should be emphasized. Rich data samples can not only improve the universality of feature extraction of network models, but also improve the adaptability of the generated network model in the real environment. Due to the limitations of various factors when collecting image data, the environment, lighting and other conditions of the images of full-film double-ditch corn seedbed collected in this study are relatively simple, so it is necessary to preprocess the image data (data enhancement) before labeling the data set to improve the diversity of data samples.

The commonly used data enhancement methods include: flipping, rotation, clipping, scaling, shifting, Gaussian noise and color transformation[16]. Due to the huge scale of image data sets collected in this study, it is not necessary to use a large number of data enhancement methods to amplify the data set. Only two methods of Gaussian noise and random brightness are used to randomly process some images, so as to increase the diversity of data samples under different clarity and different brightness, and simulate the operation conditions of the intelligent operation platform in different environments and weather conditions in the later stage. The data enhancement results are shown in Figure 1.

### 2.2 Generated data set

3600 images were preliminarily screened out from 6400 images collected for the production of target detection data sets, and 1500 images were screened out for the production of semantic segmentation data sets. At the same time, 100 images were extracted from the screened images to replace the original images after data enhancement. From the 1500 preliminarily screened images, another 1205 available images containing the target to be segmented were screened, and the target information in the images was labeled to generate a film dataset in PASCAL VOC format. A total of 1084 images were generated from the training set and verification set, 121 images were generated from the test set, and 975 images were generated from the training set and 109 images were generated from the verification set at a ratio of 9:1.



a. Original image b. Noisy points

c. Brightening d. Darkening

Figure 1 Data enhancement results

EIseg (Efficient Interactive Segmentation) is an interactive segmentation software developed by Baidu based on the PaddlePaddle platform. The PaddlePaddle version used in this study is 2.1.2; the EIseg version is 0.3.0. EIseg officially provides four kinds of pre-trained network models. In this study, a high-precision general scene network model based on COCO+LVIS training was used for image annotation. The data set format generated by EIseg is a json file consistent with Labelme or MS COCO format. In this study, json file was generated, and the PASCAL VOC dataset conversion script was used to convert the json file that recorded the coordinate information of the marking points in the target area into a gray scale image, thus completing the generation of the semantic segmentation dataset for the films (Figure 2).

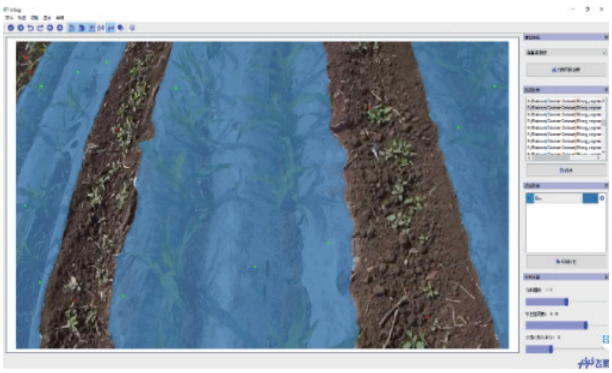


Figure 2 Elseg annotation process

### 2.3 DeepLabv3+

In order to solve the problem of spatial insensitivity caused by the using Deep Convolutional Neural Networks for feature extraction and resolution degradation caused by the underground image sampling operation in the semantic segmentation tasks, Chen et al.[17] proposed the DeepLabV1 semantic segmentation network model in 2014. In comparison with DeepLabV1, Chen et al.[18] proposed DeepLabV2 in 2017 to replace the backbone network in V1 from VGG to ResNet, and added the Atros Spatial Pyramid Pooling (ASPP) module after the backbone network. The ASPP

module increases the receptive field of the network by using four expansion convolutions with different expansion rates in parallel, thus solving the problem of multi-scale target identification.

The DeepLabV3 network model proposed by LC Chen et al. in 2017 improved the ASPP module on the basis of DeepLabV2[19]. Compared with the ASPP module of the DeepLabV2 network model, the new module only retained three expansion convolution branches, adding 1×1 common convolution branch and global average pooled branch, then feature fusion was performed on the output of the five branches, and finally, 1×1 convolution was used to adjust the number of channels. After improvement, the MIoU measured by DeepLabV3 network model on PASCAL VOC2012 dataset was 86.7%, 6% higher than that measured by DeepLabV2.

The original network of DeepLabv3+ is shown in Figure 3. DeepLabv3+ uses the "Encoder Decoder" architecture, which is commonly used in semantic segmentation tasks. DeepLabv3+ uses the whole DeepLabV3 that replaced the backbone network as the "coding" part. The shallow features of the backbone network are extracted. After the deep features are processed by the ASPP module, the shallow features and deep features are fused in the "decoding" part. Finally, the prediction results are obtained by 4 times of bilinear sampling. The "Encoder Decoder" structure enhances the ability of image edge segmentation and improves the segmentation accuracy of target pixels[20].

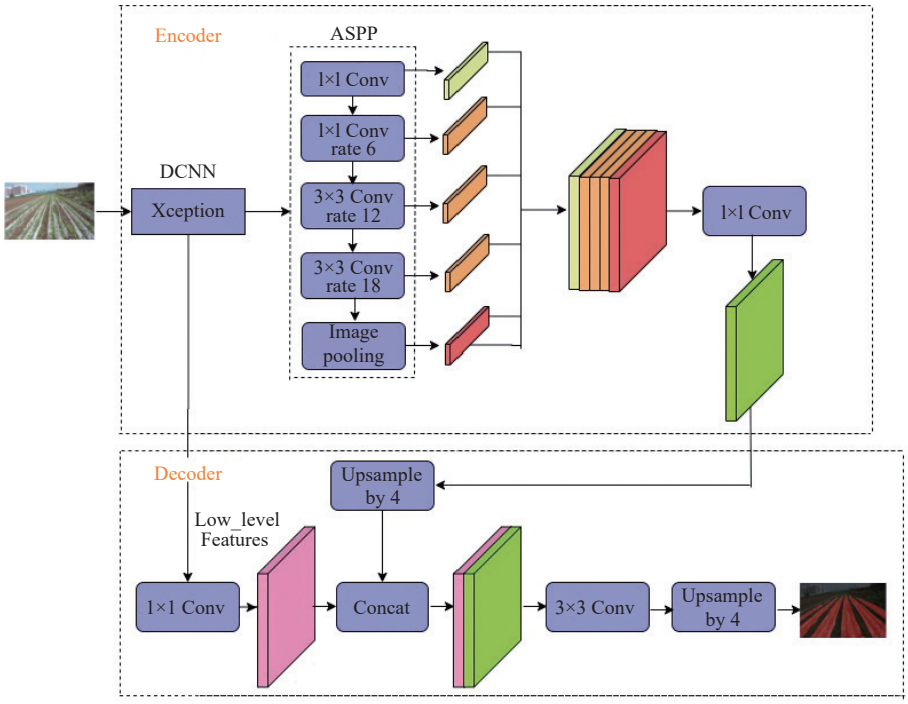



Figure 3 Original network structure of DeepLabv3+

## 2.4 Improvement of the network model

The DeepLabv3+ network model with Xception as the backbone network has the problems of large number of calculation parameters, large model size, and redundant backbone network. Since this research aims to achieve the operation of the network model with the minimum hardware cost, the original DeepLabv3+ network model cannot meet the requirements of further developing an intelligent topdressing platform for full-film double-ditch corn seedbed, and the DeepLabv3+ network model needs to be improved. It is expected to reduce the number of model parameters, improve the network identification speed and further improve the real-time performance of network model detection when the accuracy and other evaluation indicators are basically unchanged[21-23].

The backbone network using the deep CNN model is mainly composed of a series of complex operations such as convolution, pooling and activation function, and the hierarchical structure is complex, which contributes most of the computation of the entire network model. In order to deploy deep learning applications on mobile devices, various lightweight network models have been proposed and applied to tasks. Xception is a lightweight network developed from the Inception network structure. The original DeepLabv3+network model also uses this network model. However, its release time is too early and the difference between Xception and the current mainstream lightweight network models, especially the difference in parameter quantity and model file size is too large. In this study, the Xception backbone network of the original DeepLabv3+ network model was replaced by MobileNetV2, MobileNetV3, GhostNet network models in order to reduce the number of parameters and improve the network performance[24-26].

The 5-layer parallel ASPP structure of the original DeepLabv3+ network model is improved from the 4-layer parallel structure of the DeepLabV2 network model, and the 3-layer expansion convolution expansion rate of the structure is 6, 12, and 18 respectively. The expansion convolution with different expansion rates has different effects on the extraction of target features. The larger the expansion rate, the stronger the network's ability to extract features of large targets. The full-film double-ditch film targets with concentrated semantic segmentation data set used in this study account for a large proportion in the whole image with distinctive features, therefore, only the expansion convolution with an expansion rate of 18 was selected in this study, and the expansion convolution with expansion rate of 6 and 12 was deleted to reduce the parameter quantity and simplify the model. The improved ASPP structure was changed from a 5-layer parallel structure to a 3-layer parallel structure, which includes a 1×1 convolution layer, one expansion convolution layer and one global average pooling layer.

In the original DeepLabv3+ network model, the common 3×3 convolution should be used for feature extraction after fusion of shallow and deep features. Due to the large number of parameters introduced by the common convolution and the complex calculation process, the common convolution was replaced by the depthwise separable convolution to further reduce the number of parameters. Specifically, first, use the 1×1 convolution to adjust the channel of features; second, use the 3×3 depthwise separable convolution for feature extraction. The structure of the improved network model is shown in Figure 4.

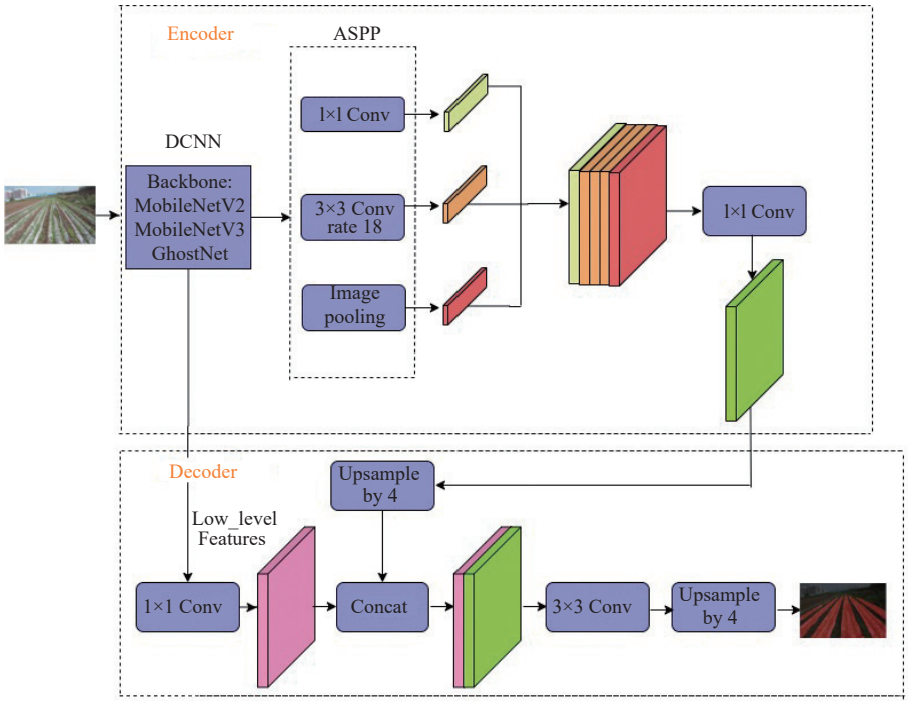



Figure 4 Improved network structure of DeepLabv3+

### 2.5 Test platform

The test environment CPU built in this study is AMD (R) Ryzen (TM) 75800X CPU @3.8 GHz (8-core 16 thread), it has a memory of 16 GB, a GPU of NVIDIA GeForce RTX 3070Ti 8 GB, hard disk combination of 512 GB SSD+1TB HDD, and the operating system is Windows10 Professional 20H2. Anaconda was used to create a virtual running environment; Anaconda version 1.10.0 was used in this research, conda version was 4.9.2, Python version was 3.8.5, PyTorch version was 1.10.2, CUDA Toolkit version was 11.3 and cuDNN version was 8.2.1, respectively.

### 2.6 Evaluation indicators

After training is completed, the performance of the model should be evaluated on the verification data set. In this study, the final semantic segmentation network model obtained from the training was evaluated from the following six aspects: Pixel Accuracy (*PA*), Mean Pixel Accuracy (*mPA*), Mean Intersection over Union (*MIoU*), detection speed, quantity of parameters and size of model files. At present, the mainstream semantic segmentation evaluation indicators are calculated based on the error between the prediction results of the network model and the pre-labeled content pixels of the validation data set. Suppose that the data set has a total of $k$+1 (a total of $k$ target categories, and the semantic segmentation network needs to include a background category by default) target categories, set $i$ as the background pixel of the target pixel $j$, and $P_{ii}$ as the total number of pixels that are themselves class $i$ pixels and are predicted as class $i$ pixels, $P_{ij}$ refers to the total number of pixels that are class $i$ pixels and are predicted to be class $j$ pixels, and $P_{ij}$ refers to the total number of pixels that are class $j$ pixels and are predicted to be class $i$ pixels. The following formula can be used to define the above precision evaluation indicators:

(1) Pixel precision: the ratio between predicted accurate pixel number and total image pixels, its calculation equation is as follows:

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}} \tag{1}$$

(2) Average pixel precision: the ratio between predicted accurate pixel number and total image pixels in each category, then get the mean value of all categories. Its calculation equation is as follows:

$$mPA = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \tag{2}$$

(3) Average regional contact degree: the ratio of the intersection and union of the number of predicted pixels in each category and the number of real pixels labeled in the data set, and then the average value of all categories is calculated. This indicator reflects the contact degree of predicted pixels and labeled pixels in each category. The calculation equation is as follows:

$$mIoU = \frac{\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}}}{k+1} \tag{3}$$

## 3 Results and discussion

### 3.1 Model training

The self-built full-film double-ditch film data set was used to train the original DeepLabv3+, the network model that only replaced backbone network, the network that replaced backbone network, compressed ASPP structure and replaced the convolution

function.

In training the original versions of the networks, in order to ensure the consistency of the initial conditions of the network models, the size of the input images in the networks should be set to 512×512 pixels, and the undistorted Resize method was used to process the input images. In training the original networks, pre-training weight training was not used on the whole DeepLabv3+ model, the official pre-training weight was applied only on the part of backbone networks. The number of iterations of network training was set to 100 epochs, since pre-training weight was applied on backbone networks, therefore, freezing training was applied to speed up the rate of convergence of the network models. Freezing training was applied on the first 50 epochs. When the backbone network was Xception, due to limitations of equipment performance, the batchsize was set to 4. When the backbone network was other network models, the batchsize was set to 8. The initial learning rate was set to $5\times10^{-4}$, at this time, the backbone part of the network model was frozen, and only the weight of the part other than the backbone network was adjusted slightly, which only required a small hardware cost. Freezing training was also applied on the last 50 epochs, at this time, the backbone part of the network model was frozen, and the weight parameters of the whole network changed, requiring larger hardware costs. When the backbone networks were other network models, the batchsize was set to 4, and the initial learning rate was set to $5\times10^{-5}$. The factor of momentum was set to 0.9, and the weight attenuation value was set to $5\times10^{-4}$. In order to screen out the optimal model weight files in the later stage, each epoch was saved into the model weight file in the training process. The loss curves of the four backbone networks are shown in Figure 5.

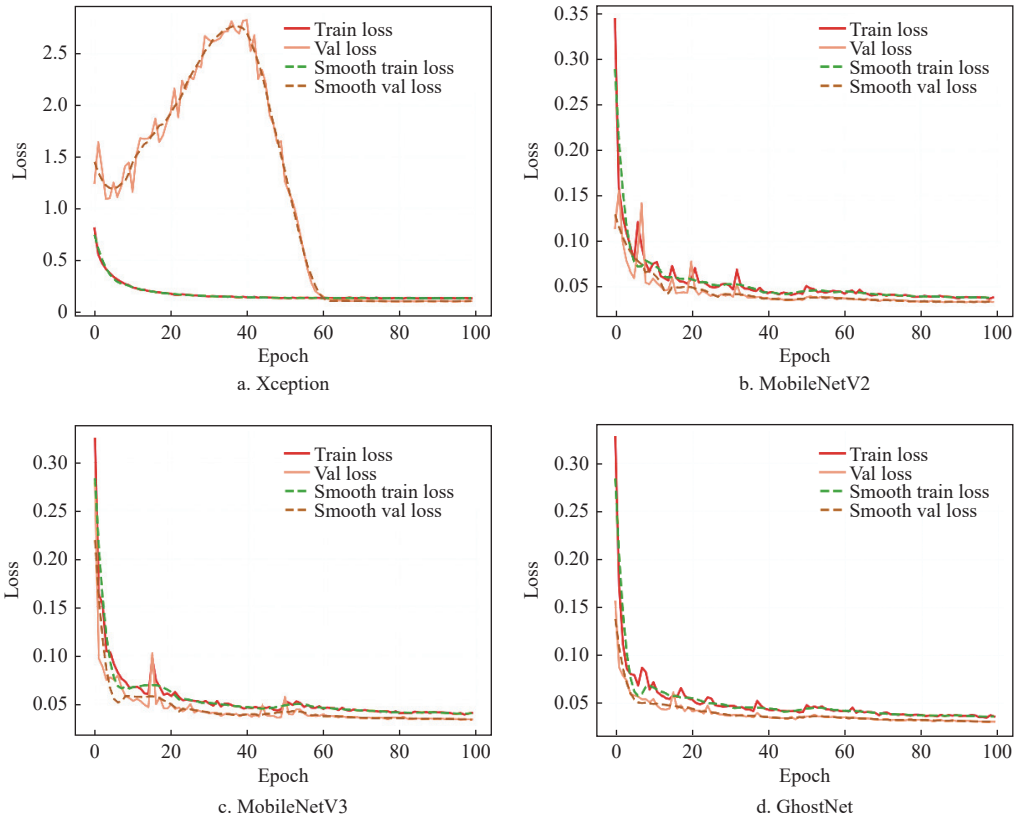



Figure 5 Loss curves of the four backbone networks

In training the improved DeepLabv3+ network model, since there were no Xception backbone networks that require a great deal of hardware costs, pre-training weight was not applied on all network models in the training process, all of them were trained from the start. The parameter setting in the training process was consistent with the original network training.

### 3.2 Results and analysis

Performance comparison of backbone networks in the original model are listed in Table 1.

It can be seen from the comparison in Table 1 that the DeepLabv3+network model with the original Xception as the backbone network had poor performance, with the model size reaching 208.72 MB, which is the maximum result among all test models. Although the detection accuracy of PSPNet and UNet as contrast has been improved, the final model file generated was still too large to meet the lightweight deployment requirements of the network model in this study. After replacing the backbone network with the MobileNet series and GhostNet lightweight network, the performance of the network model was significantly improved, the number of parameters were reduced by an order of magnitude, and the minimum model file generated was 20.28 MB (GhostNet is the backbone network), which is 90.3% less than the original network, obviously smaller than other contrast networks. Compared with the results before replacing the backbone network, the *MIoU* of the network model was increased from 92.06% to 97.07% (GhostNet is the backbone network), showing a significant increase of 5.01%. The maximum *FPS* value was 56.47 (MobilenetV2 is the backbone network), which is 38.07 higher than the original network, showing a lifting range of 206% and indicating that the network model after replacing the backbone network had excellent performance.

Comparison of performance of improved backbone networks are listed in Table 2.

**Table 1 Performance comparison of backbone networks in the original model**

| Backbone/ Network | *MIoU*/% | *mPA*/% | *PA*/% | *FPS* | Number of parameters | Model size/ MB |
|---|---|---|---|---|---|---|
| Xception | 92.06 | 96.02 | 96.00 | 18.40 | $5.47\times10^7$ | 208.72 |
| PSPNet | 95.86 | 97.87 | 97.96 | 27.06 | $4.67\times10^7$ | 178.21 |
| UNet | 97.03 | 98.44 | 98.55 | 17.22 | $4.39\times10^7$ | 167.60 |
| MobileNetV2 | 96.88 | 98.34 | 98.47 | 56.47 | $5.81\times10^6$ | 22.19 |
| MobileNetV3 | 96.88 | 98.34 | 98.48 | 53.25 | $4.83\times10^6$ | 18.45 |
| GhostNet | 97.07 | 98.48 | 98.57 | 45.33 | $5.31\times10^6$ | 20.28 |

**Table 2 Comparison of performance of improved backbone networks**

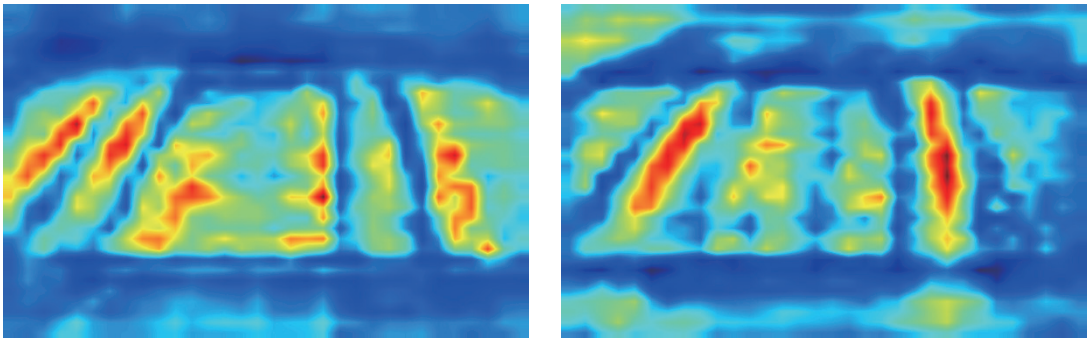
| Backbone | *MIoU*/% | *mPA*/% | *PA*/% | *FPS* | Number of parameters | Model size/ MB |
|---|---|---|---|---|---|---|
| MobileNetV2 | 96.76 | 98.28 | 98.42 | 57.70 | $3.00\times10^6$ | 11.49 |
| MobileNetV3 | 96.71 | 98.26 | 98.39 | 56.50 | $2.76\times10^6$ | 10.53 |
| GhostNet | 96.83 | 98.33 | 98.45 | 66.50 | $5.31\times10^6$ | 12.36 |

It can be seen from the comparison in Table 2 that after compressing and improving DeepLabv3+ network model, the three evaluation indicators of the network model *MIoU*, *mPA* and *PA* using the three lightweight backbone networks decreased, with an average decrease of 0.17%, which is not obvious and still within the acceptable range. *FPS* of the three network models were improved, with an average increase of 25.5%. The model file size decreased significantly, especially for the network model with MobileNetV3 as the backbone network, the model size decreased to 10.53 MB, showing an average decrease of 94.9% compared with the original network and 48% compared with the original network.

The feature map visualization can vividly show the effect of feature extraction of specific layers in the network, which is a commonly used auxiliary method in network modification. Visualization of the output of the last layer of the ASPP structure before and after the modification was completed to observe the difference of the feature extraction effect of the modified network.

In the process of visualization of the feature map, MobileNetV3 was taken as the backbone network. As shown in Figure 6, the feature extraction ability of the network after modification was reduced slightly. The extracted features at the edge of the images were not complete, and there were insignificant differences in the total target feature extraction effect before and after network extraction, thus it could satisfy the demand of research task. The results above showed that, the models replacing the backbone network and after compression and improvement showed high identification accuracy rate and rapid running speed, with smaller model size, thus they could satisfy the scenario requirements of full-film double-ditch films segmentation tasks.

Figure 7 shows the comparison of segmentation results of the DeepLabv3+ network model using original backbone network and the three types of improved lightweight backbone networks. It can be obtained after comparing the segmentation results of *a*, *b*, *c* groups that, the segmentation results of network model using



a. Before modification b. After modification

Figure 6 Visualization of the ASPP structural feature map

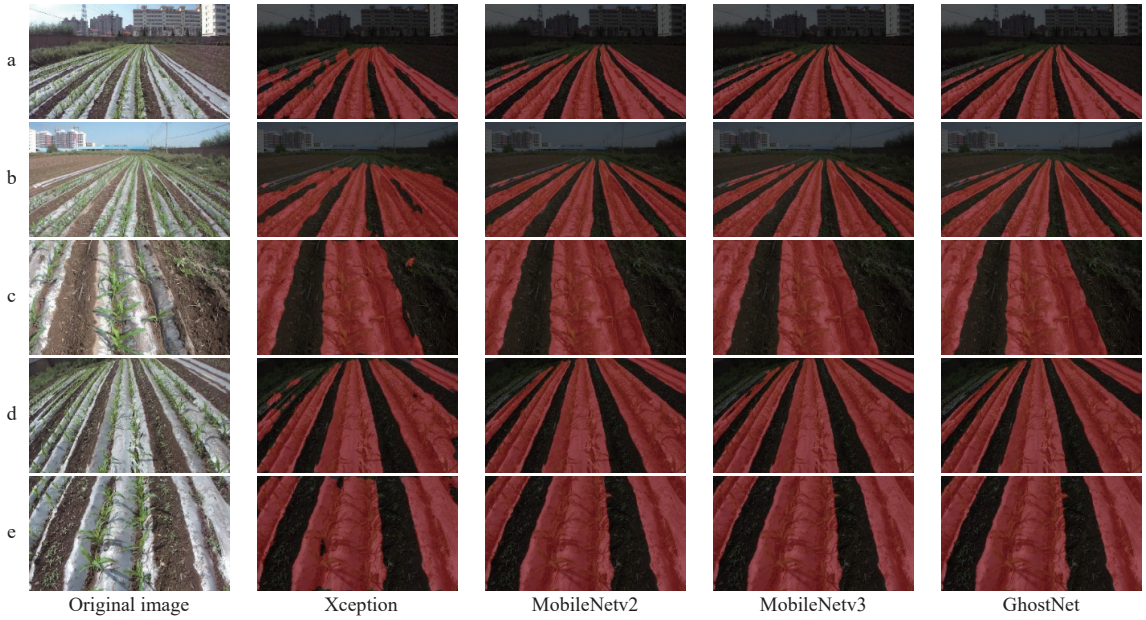



Figure 7 Segmentation results

Xception backbone network were not so ideal on the distant small targets and near-end inconspicuous big targets, also the background and the targets could not be effectively differentiated, showing that the feature extraction ability of the network model was not strong. The segmentation results of groups d and e on the seedbed films were damaged, showing insufficient generalization ability. The segmentation results of the network models using three lightweight backbone networks on five groups of images showed excellent performance. Comparing with the segmentation results of group a, it was found that the network model taking MobileNetV2 as the backbone network detected the unmarked targets at the utmost edge and showed optimal effect in detecting targets at far-end edge. However, its indicators in terms of precision degree were not optimal, and number of parameters and model size were both larger than the network model taking MobileNetV3 as the backbone network.

Since the *MIoU*, *mPA* and *PA* values of the improved network model were approximate to each other, comprehensively considering the three evaluation indicators, *FPS*, number of parameters and model size, as well as image segmentation results, the improved network model taking GhostNet as the backbone network showed optimal performance.

## 4 Conclusions

In this study, by taking the full-film double-ditch corn seedbed as research object, by using the semantic segmentation method based on deep learning, the performance of improved network models was evaluated. The conclusions were as follows:

1) The differences in performance indicators of the original Deeplabv3+ network taking Xception as the backbone network and the network models that replaced three lightweight backbone networks, MobileNetV2, MobileNetV3 and GhostNet were tested. Then the classical semantic segmentation network models PSPNet and UNet were introduced for comparative study. By pruning and compressing the ASPP structure and replacing the common convolution after feature fusion with the depthwise separable convolution, a lightweight network model was obtained.

2) Test results showed that, the *MIoU* of improved DeepLabv3+ network model increased from 92.06% to 97.07% (Taking GhostNet as the backbone network), an increase of 5.01%. The model size reduced from 208.72 MB to 20.28 MB, marking a decline of 90.3%. The *FPS* value increased from 18.4 to 56.47 (taking MobilenetV2 as the backbone network), showing an increase of 38.07 compared with the original network, which is a 206% increase. All the indicators of the model after replacing backbone network presented excellent performance. After compressing the ASPP structure, the *FPS* of the model enjoyed an average increase of 25%, and the minimum model size of 10.53 MB was obtained. Comprehensively considering the performance of each indicator, the improved model taking GhostNet as the backbone network showed optimal performance. The research results could offer technological and method support to subsequent operation platform development, such as intelligent topdressing, field management, for the full-film double-ditch corn seedbed, as well as the deployment of deep learning applied in intelligent agricultural machinery terminals like edge calculation equipment.

## Acknowledgements

The authors acknowledge that this work was financially supported by the National Natural Science Foundation of China (Grant No. 52065005; 52365029), Outstanding Youth Foundation of Gansu Province (Grant No. 20JR10RA560), China Postdoctoral Science Foundation (Grant No. 2021M700741).

## [References]

[1] Zhou L M, Jin S L, Liu C A, Xiong Y C, Si J T, Li X G, et al. Ridge-furrow and plastic-mulching tillage enhances maize-soil interactions: opportunities and challenges in a semiarid agroecosystem. Field Crops Research, 2012; 126: 181–188.

[2] Dai F, Zhao W Y, Zhang F W, Ma H J, Xin S L, Ma M Y. Research progress analysis of furrow sowing with whole plastic-film mulching on double ridges technology and machine in northwest rainfed area. Transactions of the CSAM, 2019; 50(5): 1–16. (in Chinese)

[3] Dai F, Guo W J, Song X F, Zhang Y, Shi R J, Wang F, Zhao W Y. Optimization of mechanized soil covering path based on the agronomic mode of full-film double-ditch with double-width filming. Int J Agric & Biol Eng, 2022; 15(1): 139–146.

[4] Luo X W, Liao J, Hu L, Zhou Z Y, Zhang Z G, Zang Y, et al. Research progress of intelligent agricultural machinery and practice of unmanned farm in China. Journal of South China Agricultural University, 2021; 42(6): 8–17. (in Chinese)

[5] Tang H, Xu C S, Wang Z M, Wang Q, Wang J W. Optimized design, monitoring system development and experiment for a long-belt finger-clip precision corn seed metering device. Frontiers in Plant Science, 2022; 13: 814747.

[6] Liu C L, Lin H Z, Li Y M, Gong L, Miao Z H. Analysis on status and development trend of intelligent control technology for agricultural equipment. Transactions of the CSAM, 2020; 51(1): 1–18. (in Chinese)

[7] Dai Y S, Zhong X C, Sun C M, Yang J, Liu T, Liu S P. Identification of fusarium head blight in wheat based on image processing and Deeplabv3+ model. Journal of Chinese Agricultural Mechanization, 2021; 42(9): 209–215. (in Chinese)

[8] Zhou J, He Y Q. Research progress on navigation path planning of agricultural machinery. Transactions of the CSAM, 2021; 52(9): 1–14. (in Chinese)

[9] Mu R H, Zeng X Q. A review of deep learning research. KSII Transactions on Internet and Information Systems, 2019; 13: 1738–1764.

[10] Wan S H, Goudos S. Faster R-CNN for multi-class fruit detection using a robotic vision system. Computers Networks, 2020; 168: 107036.

[11] Sun Z T, Zhu S N, Gao Z J, Gu M Y, Zhang G L, Zhang H M. Recognition of grape growing areas in multispectral images based on band enhanced DeepLabv3+. Transactions of the CSAE, 2022; 38(7): 229–236. (in Chinese)

[12] Mu T Y, Zhao W, Hu X Y, Li D. Rice lodging recognition method based on UAV remote sensing combined with the improved DeepLabV3+ model. Journal of China Agricultural University, 2022; 27(2): 143–154. (in Chinese)

[13] Rao X Q, Zhu Y H, Zhang Y N, Yang H T, Zhang X M, Lin Y Y, et al. Navigation path recognition between crop ridges based on semantic segmentation. Transactions of the Chinese Society of Agricultural Engineering, 2021; 37(20): 179–186. (in Chinese)

[14] Yang Y, Zhang Y L, Miao W, Zhang T, Chen L Q, Huang L L. Accurate identification and location of corn rhizome based on faster R-CNN. Transactions of the CSAM, 2018; 49(10): 46–53. (in Chinese)

[15] Meng Q K, Yang X X, Zhang M, Guan H O. Recognition of unstructured field road scene based on semantic segmentation model. Transactions of the Chinese Society of Agricultural Engineering, 2021; 37(22): 152–160. (in Chinese)

[16] Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning. Journal of big data, 2020; 6: 1–48.

[17] Chen L C, Papandreou G, Kokkinos L, Murphy K, Yuille A L. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018; 40: 834–848.

[18] Ji J, Li S T, Xiong J, Chen P, Miao Q G. Semantic image segmentation with propagating deep aggregation. IEEE Transactions on Instrumentation and Measurement, 2020; 69: 1.

[19] Li Z Y, Wang R, Zhang W, Hu F M, Meng L K. Multiscale features supported DeepLabV3 optimization scheme for accurate water semantic segmentation. IEEE Access, 2019; 7: 155787–155804.

[20] Fu H X, Meng D, Li W H, Wang Y C. Bridge Crack Semantic Segmentation Based on Improved Deeplabv3+. Journal of Marine Science

and Engineering, 2021; 9: 671.

[21] Zhang X F, Bian H N, Cai Y H, Zhang K Y, Li H. An improved tongue image segmentation algorithm based on Deeplabv3+ framework. IET Image Process, 2022; 16: 1473–1485.

[22] Hu Z F, Zhao J, Luo Y, Ou J F. Semantic SLAM based on improved DeepLabv3+ in dynamic scenarios. IEEE Access, 2022; 10: 21160–21168.

[23] Liao J, Chen M H, Zhang K, Zou Y, Zhang S, Zhu D Q. Segmentation of crop plant seedlings based on regional semantic and edge information fusion. Transactions of the CSAM, 2021; 52(12): 171–181. (in Chinese)

[24] Buiu C, Dănăilă V R, Răduţă C N. MobileNetV2 ensemble for cervical precancerous lesions classification. Processes, 2020; 8: 595.

[25] Yin X, Li W H, Li Z, Yi L L. Recognition of grape leaf diseases using MobileNetV3 and deep transfer learning. Int J Agric & Biol Eng, 2022; 15(3): 184–194.

[26] Yuan X G, Li D, Sun P, Wang G, Ma Y L. Real-time counting and height measurement of nursery seedlings based on Ghostnet-YoloV4 network and binocular vision technology. Forests, 2022; 13: 1459.