agronomy

MDPI

Article

Image Segmentation-Based Oilseed Rape Row Detection for Infield Navigation of Agri-Robot

Guoxu Li [1,2], Feixiang Le [1], Shuning Si [1,2], Longfei Cui [1,*] and Xinyu Xue [1,*]

[1] Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China; 82101225599@caas.cn (G.L.); lefeixiang@caas.cn (F.L.); 82101202108@caas.cn (S.S.)
[2] Graduate School of Chinese Academy of Agricultural Sciences, Beijing 100081, China
* Correspondence: cuilongfei@caas.cn (L.C.); xuexinyu@caas.cn (X.X.); Tel.: +86-15366093007 (L.C.); +86-25-8434-6244 (X.X.)

Abstract: The segmentation and extraction of oilseed rape crop rows are crucial steps in visual navigation line extraction. Agricultural autonomous navigation robots face challenges in path recognition in field environments due to factors such as complex crop backgrounds and varying light intensities, resulting in poor segmentation and slow detection of navigation lines in oilseed rape crops. Therefore, this paper proposes VC-UNet, a lightweight semantic segmentation model that enhances the U-Net model. Specifically, VGG16 replaces the original backbone feature extraction network of U-Net, Convolutional Block Attention Module (CBAM) are integrated at the upsampling stage to enhance focus on segmentation targets. Furthermore, channel pruning of network convolution layers is employed to optimize and accelerate the model. The crop row trapezoidal ROI regions are delineated using end-to-end vertical projection methods with serialized region thresholds. Then, the centerline of oilseed rape crop rows is fitted using the least squares method. Experimental results demonstrate an average accuracy of 94.11% for the model and an image processing speed of 24.47 fps/s. After transfer learning for soybean and maize crop rows, the average accuracy reaches 91.57%, indicating strong model robustness. The average yaw angle deviation of navigation line extraction is 3.76°, with a pixel average offset of 6.13 pixels. Single image transmission time is 0.009 s, ensuring real-time detection of navigation lines. This study provides upper-level technical support for the deployment of agricultural robots in field trials.

Keywords: visual navigation; agricultural robot; path recognition; oilseed rape; image segmentation

Citation: Li, G.; Le, F.; Si, S.; Cui, L.; Xue, X. Image Segmentation-Based Oilseed Rape Row Detection for Infield Navigation of Agri-Robot. *Agronomy* **2024**, *14*, 1886. https://doi.org/10.3390/agronomy14091886

Academic Editors: Xiangjun Zou, Yunchao Tang, Junfeng Gao, Liang Gong, Simon van Mourik and Ya Xiong

Received: 29 July 2024
Revised: 22 August 2024
Accepted: 22 August 2024
Published: 23 August 2024

1. Introduction

China is a major global producer and consumer of oilseed rape, with the area covered by oilseed rape and its production being among the world's largest [1]. Rapeseed oil produced by oilseed rape, as an important edible vegetable oil in China, only accounts for 40% of the total vegetable oil [1,2]. The domestic edible vegetable oil industry continues to rely on imported products, resulting in a significant imbalance between supply and demand. Accordingly, increasing the yield and oil production of oilseed rape [3] has been a major concern for many agriculturalists. Autonomous navigation technology [4], as a core technology of agricultural machinery intelligence, not only reduces the agricultural labor force needed, but also improves the quality and efficiency of agricultural machinery in field operation, which in turn improves the yield of oilseed rape. Currently, intelligent agricultural machinery navigation is mostly based on global navigation satellite system (GNSS) and machine vision navigation [5–7]. While satellite navigation systems can obtain the location information of agricultural machinery operation, they are costly to use and maintain. Additionally, they are susceptible to signal loss due to weather and environmental factors. In contrast, visual navigation is a low-cost alternative that acquires comprehensive information. Furthermore, it employs image processing to recognize the navigation path

between crop rows, thereby reducing the phenomenon of seedling pressure during robot field walking [8]. Consequently, visual navigation is becoming a research hotspot in the field of robot navigation.

Currently, the traditional algorithms based on machine vision navigation are typically only effective in specific crop types and environmental conditions. They are sensitive to factors like shadow occlusion and lighting changes [9], which can degrade performance in complex field scenes. Such algorithms have certain limitations as they have poor generalization ability when coping with new scenes or new tasks and need to be manually adjusted and optimized. To meet the needs of the field operation environment, the improved method of feature extraction from images based on the traditional algorithm of machine vision has been gradually applied. Some scholars have carried out relevant research with the help of this method. Utstumo et al. [10] conducted research based on visual surveying on image color space division green channel thresholding segmentation carrot rows, which were converted to straight line output crop row navigation lines by Hough transform threshold detection. Andrew English et al. [11] introduced a vision-based texture tracking method simulating an overhead view to extract image textures and offsets for predicting crop-specific details, guiding the robot's heading relative to the crop row. Radcliffe et al. [12] proposed a methodology for segmenting orchard canopies and skies in machine vision systems. This approach guides the route by segmenting the center of mass of the object with an error of 2.13 cm.

Deep learning semantic segmentation methods are widely used in fields such as geological exploration [13], autonomous driving [14–16], and medical detection [17]. In recent years, this approach has been applied in more areas, particularly in agricultural visual inspection, where machine vision navigation is transitioning from conventional computational methods to deep learning neural networks. Presently, mainstream deep learning methods in visual navigation primarily focus on image target detection and region segmentation. Gong et al. [18] used a neural network for infrared image target detection to identify maize seedlings, achieving an average error of 4.85 cm between the fitted navigation line and the midpoint of the reference detection frame's center. Ju et al. [19] proposed an enhanced methodology for rice seedling identification using YOLOv5s, which entails integrating the model's backbone network structure with MobileViTv3, substituting the loss function with WIoU loss to enhance rice seedling identification and subsequently fitting the navigation line with the least squares method. Bah et al. [20] employed a quantitative comparison of traditional methods based on Convolutional Neural Network (SegNet) combined with CNN-based Hough Transform algorithms. Their findings demonstrated the efficacy of this approach in detecting diverse types of crop rows. De Silva et al. [21] introduced a methodology to generate segmentation masks from predicted sugar beet datasets using U-Net. They subsequently extracted crop row centerline navigation using the triangular scanning method, which adapts better to field conditions. Adhikari et al. [22] employed the ES-Net neural network to train on paddy rows. They introduced a sliding window algorithm to locate the pixels corresponding to the paddy lines, subsequently extracting the two primary rows to fit the centerline.

Previous studies have shown that deep learning-based neural network methods have achieved significant advancements in crop row detection [23] and have matured in various visual scene tasks within agriculture. Nevertheless, the existing crop row recognition and detection methods based on deep learning still exhibit certain shortcomings. The high precision of the deep learning semantic segmentation algorithm requires a significant amount of computational resources and is time consuming. Balancing algorithm precision with computational speed is crucial. Most current research focuses on recognizing navigation paths in single crop row scenes, which lacks scene detection universality. The deep learning semantic segmentation model for navigation path recognition in crop row scenes is challenging to explain due to the complexity of the environment.

To address the existing problems, this study adopted a lightweight semantic segmentation model for VC-UNet based on U-Net improvement. The architecture is simple

and able to cope with the segmentation task with irregular boundaries of oilseed rape crop rows. The rapeseed crop row dataset containing light, terrain, and shadow data was constructed according to the complex environment of rapeseed crop field. Transfer learning was performed on soybean-corn composite crops for the model segmentation results, and the soybean-corn dataset was used to validate the prediction effect of the VC-UNet model. We cropped trapezoidal ROI regions from the model-predicted canola crop rows. The end-to-end vertical projection method was then applied to obtain the image threshold segmentation information based on the extracted boundary features. Finally, the navigation centerline was fitted by the least squares method through the extracted positions of the two crop rows.

2. Materials and Methods

The specific process of navigation line extraction is shown in Figure 1. The visual navigation line detection of rape crop rows proposed in this study consists of two parts: image semantic segmentation and navigation line extraction. The crop row pixel features were extracted from the camera-captured rape images based on the proposed VC-UNet model. The end-to-end vertical projection algorithm was used for the prediction of the results to locate the crop rows for extraction to the navigation line, and the final result was then returned for the projection of the results to the original image.

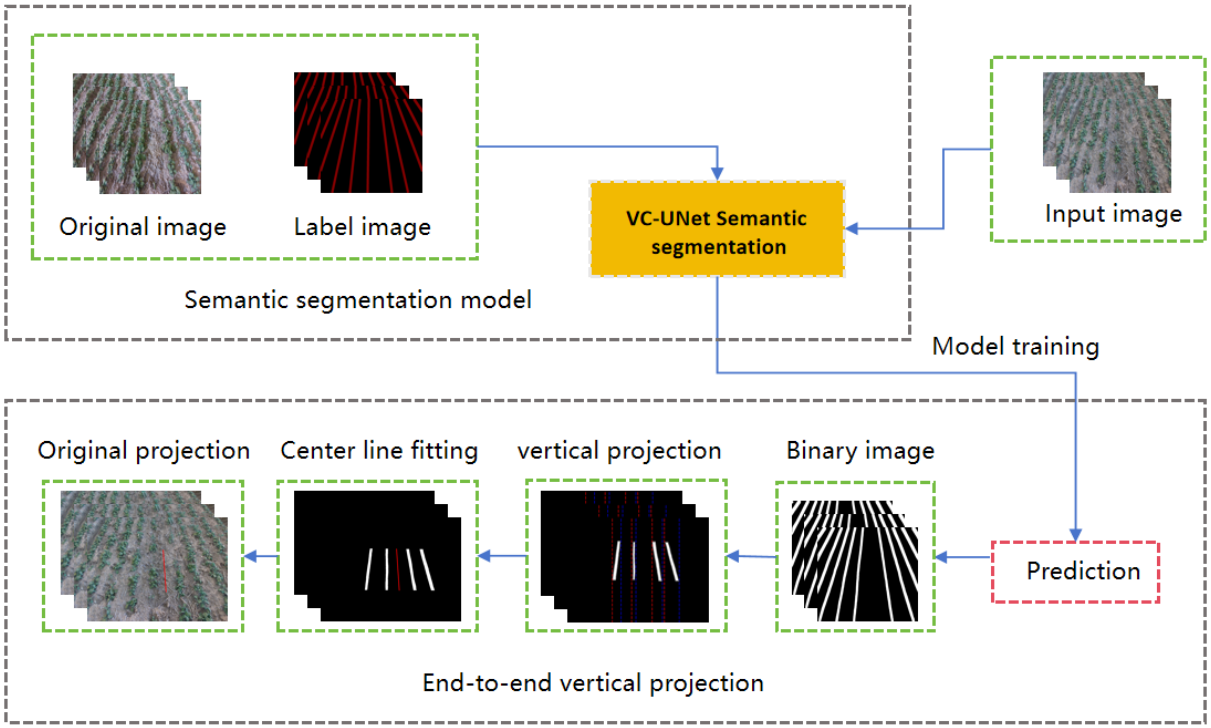

Figure 1. Navigation line extraction process method.

The model training and testing in this study was performed using the Pytorch (version 2.0.1) framework based on Windows 11 and the Anaconda environment. The experimental platform was configured with an Intel (R) Core (TM) i7-12700H processor (CPU, Intel, Santa Clara, CA, USA), an NVIDIA GeForce RTX 3080Ti graphics processing unit (GPU, NVIDIA, Santa Clara, CA, USA), and 32 GB of memory (RAM). The training process followed a predefined ratio of training to validation sets established during data acquisition, involving two categories (including oilseed rape crop rows and background). The platform performed image extraction of collected video streaming files of oilseed rape dataset. The image input network resolution was standardized from 1280 × 720 pixels to 512 × 512 pixels to align with the backbone network's input feature layer.

2.1. Data Collection and Labeling

The multifunctional field management robot, which was independently developed by the Nanjing Agricultural Mechanization Research Institute (NAMRI), was selected as the image acquisition platform for data collection, as shown in Figure 2. It is equipped with two working modes: manual remote control and autonomous navigation. The depth camera (D435i, Intel RealSense, Santa Clara, CA, USA) was employed as the visual sensor during data collection. It was installed in front of the platform at an angle of 60° to the vertical direction, with the camera capturing images of the rape at a speed of 30 frames per second with a resolution of 1280 × 720 pixels. The acquisition platform moved across and along the crop rows, and the camera continuously captured oilseed rape images to be saved as a video stream. Subsequent video frame extraction was performed to extract the oilseed rape images.

Figure 2. Agricultural robotics oilseed rape data collection platform.

The image collection location of the oilseed rape dataset was the demonstration base of the National Key Project on Oilseed Rape Production in Yancheng City, Jiangsu Province. A total of 3500 pieces of data were collected in December 2023 and February 2024, respectively, for the environments with different weather light levels, shadow shading, and mixed topographies of the oilseed rape crop rows. These environments were chosen to represent the variety of situations typically encountered in oilseed rape data collection. The data collection was conducted in six categories, as shown in Table 1.

Table 1. Oilseed rape crop type data set.

Serial Number	Crop Type	Image Resolution	Image Quantity
A	Strong light	1280 × 720	800
B	Low light	1280 × 720	700
C	Normal light	1280 × 720	600
D	Horizontal crop ridge	1280 × 720	450
E	Vertical crop ridge	1280 × 720	550
F	Shadow masking	1280 × 720	400

The oilseed rape crop row dataset was divided into a training set and a validation set in a 9:1 ratio. The training set consisted of 3150 sheets, while the validation set comprised 350 sheets. To ensure consistency between the segmented dataset and the original image, LabelMe (version 3.16.2) software was employed to annotate each crop row in the rape image. The labeling method utilized was line labels, which indicated the start and end points of the rape rows. The dots were then connected into a line in accordance with the desired segmentation target. The labeled files were stored in JSON format, as illustrated in Figure 3, detailing various environments within the oilseed rape crop row dataset. Through the batch conversion method, the labeled files were converted to PNG format images.

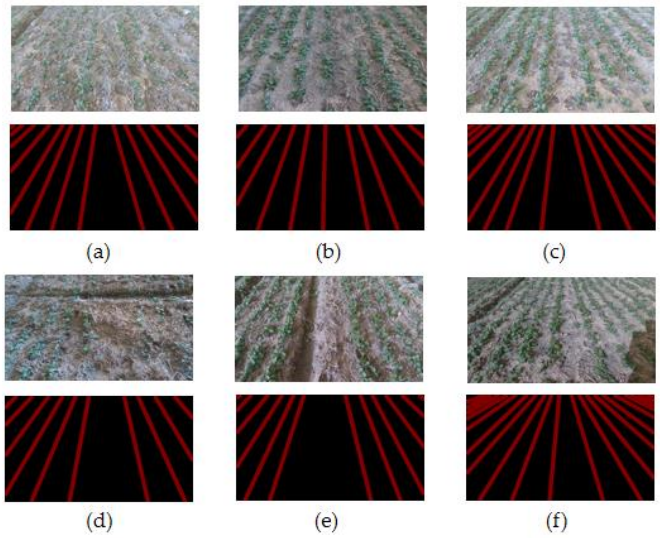

Figure 3. Labeling of oilseed rape crop row datasets in different environments: (**a**) Strong light; (**b**) Low light; (**c**) Normal light; (**d**) Horizontal crop ridge; (**e**) Vertical crop ridge; (**f**) Shadow masking.

2.2. Data Augmentation

The objective of data augmentation in the context of oilseed rape crop row image data is to increase the amount of new data, thereby enhancing the robustness and generalization ability of the model, mitigating overfitting, and circumventing the issue of sample imbalance. To expand the dataset and increase the diversity of the oilseed rape dataset, a series of image processing techniques were applied to the images, as shown in Figure 4. These include horizontal mirroring, angular rotation, random cropping, vertical mirroring, contrast adjustment, brightness change, perspective distortion, and random masking.

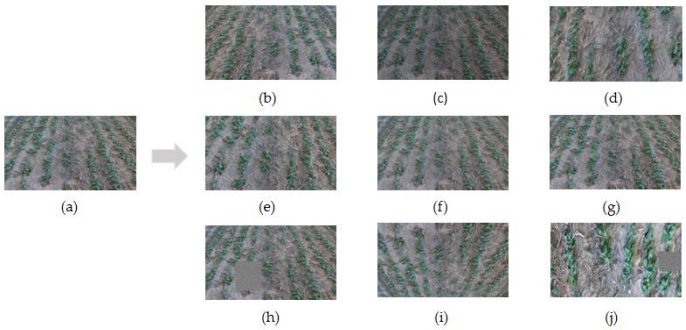

Figure 4. Data augmentation: (**a**) Original image; (**b**) Horizontal mirroring; (**c**) Random brightness; (**d**) Random cropping; (**e**) Clockwise rotation of 25°; (**f**) Contrast adjustment; (**g**) Perspective distortion; (**h**) Random occlusion; (**i**) Vertical mirroring; (**j**) Combined transformation.

2.3. Construction of Semantic Segmentation Model

The majority of traditional semantic segmentation model architectures are comprised of an encoder-decoder network [24]. The encoder transmits the image to be converted into a pre-trained backbone network with high-level semantic attributes through a series of convolution, pooling, and other operations. The decoder then returns the low-resolution attributes that have been passed in by the encoder to the high-resolution pixel space, thereby enhancing the network with dense features. The most used semantic segmentation network models are U-Net, Pspnet [25], and Deeplab V3+ [26], known for their high accuracy in segmentation tasks. However, these models are computationally intensive and lack sufficient explicitness to handle the segmentation details.

In this study, an improved U-Net semantic segmentation model was used to address the above problem. The feature extraction part of U-Net is similar to the structure of VGG16 network, as shown in Figure 5. The structure of the neural network is simplified by replacing the VGG16 model in the main feature extraction network part, thereby accelerating the convergence speed of the model and reducing the training time. The maximum pooling layer in the fifth convolution of the model, the fully connected layer and the maximum pooling structure in the model, were removed by cropping. The cropped model was constructed from 13 convolutional layers with convolutional kernels of size 3 × 3, step size 1, and fill pixel 1, and four maximal pooling layers of size 2 × 2, step size 2, and fill pixel 1, as well as the ReLU activation function. The VGG16 model employs small convolutional kernels of identical size to encourage parameter sharing, thereby reducing parameter count. The incorporation of scale dithering as a pre-training model has been shown to facilitate the acceleration of training and enhance the accuracy of the model, when compared to the original U-Net model.

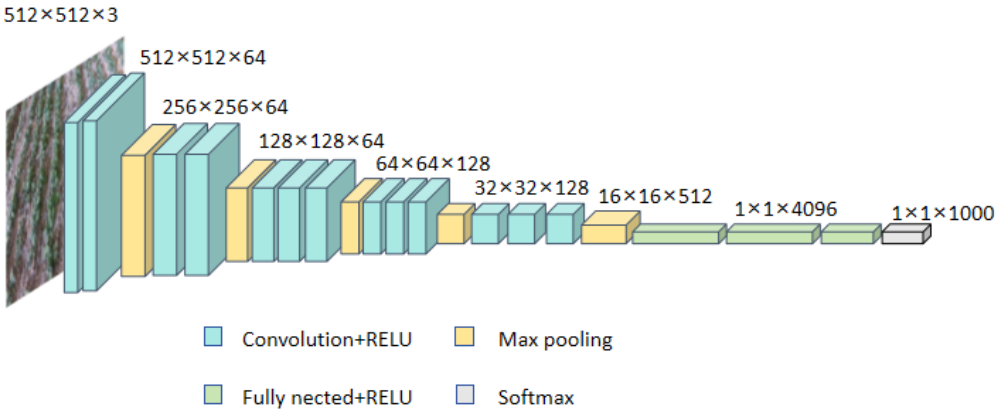

Figure 5. Improved VGG16 network model structure.

The enhanced feature extraction network comprises four upsampled transposed convolutional modules, followed by four CBAM (Convolutional Block Attention Module) attention mechanisms and four jump connection layers. The network comprises eight convolutional layers with a size of 3 × 3, a step size of 1, and a fill pixel of 0. The segmentation target in this study was predominantly a single category (oilseed rape crop rows), exhibiting low category complexity and a greater prevalence of rapeseed rows within a single category. Therefore, the CBAM attention mechanism could be used to better increase the attention to the segmentation of single-category crop rows. The CBAM attention mechanism is shown in Figure 6 and is divided into two distinct components: the channel attention mechanism and the spatial attention mechanism. In addition to the reduction in computational parameters, the integration of these parameters into the network architecture was also facilitated through the use of switches. The channel attention mechanism performed average pooling and maximum pooling on the input feature layer, and the results of the two processes were multiplied by the activation function with the original input features to output a weighted channel feature map. The spatial attention

mechanism involved the computation of the maximum and average values of the feature points on the input feature layer. These values were then convolved to produce the spatial feature map, which was subsequently multiplied by the weighted channel feature map to generate the final output feature map.

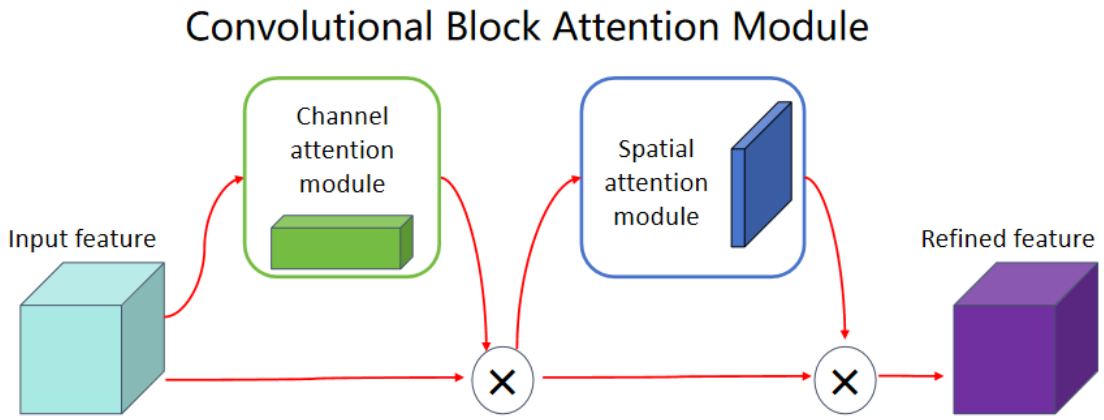

Figure 6. CBAM (Convolutional Block Attention Module) module structure.

The classical neural network U-Net [27] used for image segmentation tasks is widely applicable. However, the task of recognizing a single category of oilseed rape crop row segmentation is relatively simple and requires pruning to reduce the computational effort while ensuring the network structure [28]. In this study, the number of channels in the convolutional layer was pruned using channel pruning in the replaced VGG backbone feature extraction network. After pruning, the number of channels in the feature layer of the backbone network matched that of the augmented network. As shown in Figure 7, the scaling factors in the BN layer were linked to the channel number during the pruning process. Subsequently, sparse regularization was applied to these factors to automatically eliminate unimportant channels. The channels with smaller scaling factors were located on the left side (yellow), and following pruning, only those with larger scaling factors remained (green), resulting in a more compact network model on the right side. Finally, the model underwent fine-tuning to preserve the properties of the original network architecture, ensuring proper training with a slight increase in accuracy.

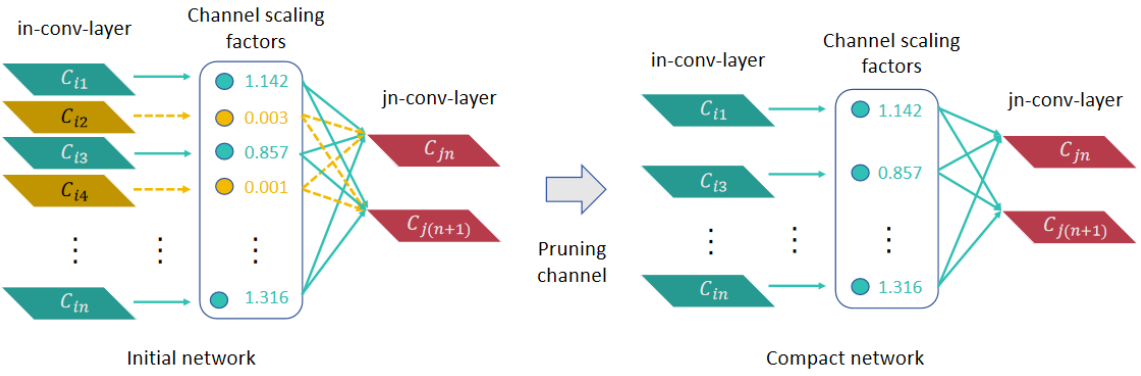

Figure 7. Schematic diagram of model channel pruning principle.

The backbone feature extraction network up-sampling process stacks each of the five initial effective feature layers, followed by feature fusion after adding CBAM (Convolutional Block Attention Module) attention mechanism directly after twice up-sampling. Ultimately, the number of convolutional layer channels between the input and output layers of the entire network was pruned to obtain the feature layer with the same height

and width as the input image. The number of convolutional layer channels at the pruning site became 64 and 128. The structure of the obtained VC-UNet model is shown in Figure 8.

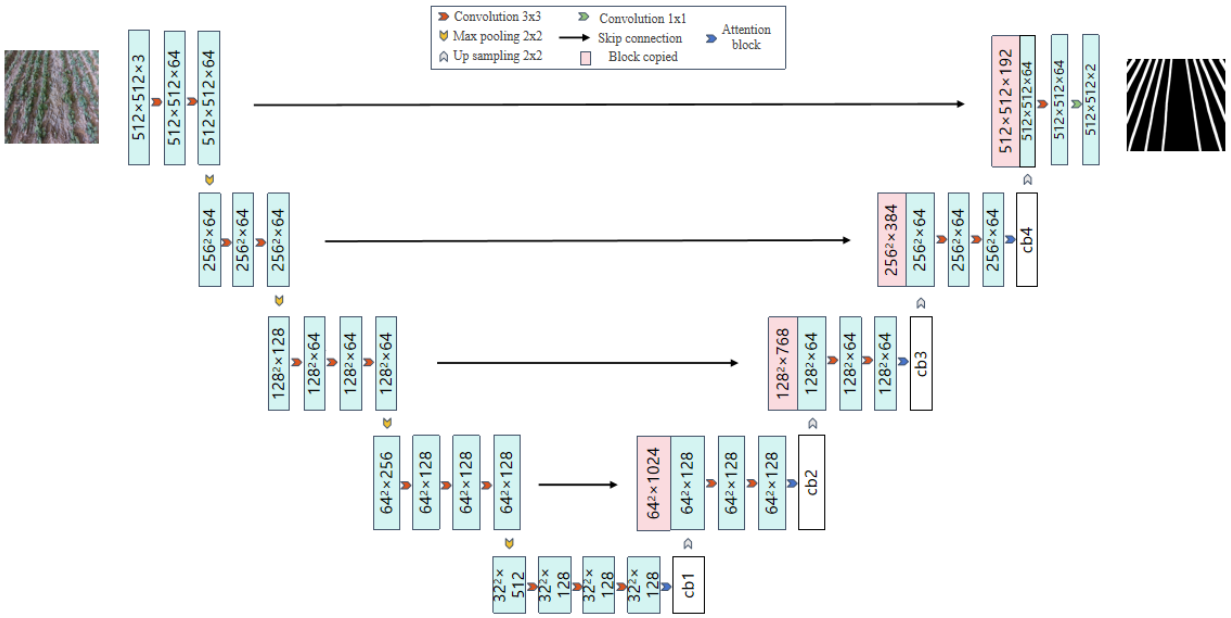

Figure 8. VC-UNet network model structure.

2.4. Transfer Learning

Transfer learning, as a deep learning method, involves acquiring knowledge to facilitate learning a new task by leveraging existing knowledge from a related, previously learned task [29]. The domain where the knowledge has been learned is called the source domain and the domain where the learning is to be transferred is called the target domain. As shown in Figure 9, the source domain for the transfer learning process study was the oilseed rape crop row dataset feature space. The target domain was the crop row dataset feature space of soybean corn compound planting (category 2). The samples based on the training weights of the source domain oilseed rape dataset were transferred to the target domain using a direct push learning method.

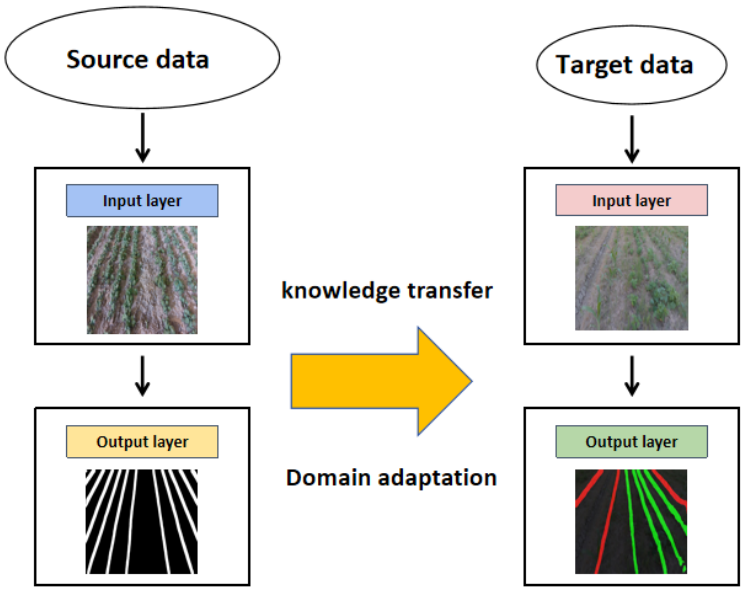

Figure 9. Schematic diagram of transfer learning.

Based on the oilseed rape sample dataset in the process of knowledge migration to soybean corn composite planting crops to maintain domain adaptive, fine-tuned the model on the target domain. The number of soybean corn composite planting dataset was only 800, which was relatively small compared to the number of source domain dataset. Therefore, to prevent overfitting during the training process, an L2 regularization constraint (weight decay) was employed.

2.5. Navigation Line Extraction

2.5.1. Target Row Feature Extraction

The field management robot has a wheelbase of 1.7 m, and the inter-row spacing of the rape crop rows is 0.4 m. The operation travels across four rape crop rows. Unlike most that only extract the nearest two crop rows, this study focused on determining crop information from the two rows adjacent to both sides of the wheel edge among the four rows. Multiple inclined rape crop rows were acquired in the camera mounted field of view, and training was accomplished by using a semantic segmentation model. The predicted mask image needed to extract ROI regions for background and crop rows based on the number of rows spanned by the robot. Subsequently, trapezoidal ROI regions were defined on the binary mask image based on the size of the pixel values. The image detection accuracy was improved by limiting the range of unknown edges in the crop rows, which accelerated the processing efficiency.

The ROI region was intercepted to obtain four binary mask crop lines to satisfy the wheel spacing requirement during the driving process of oilseed rape fields. Distinguishing the background and crop row lines within the mask map based on image thresholding required the use of an end-to-end vertical projection method to delineate the position of each line within the region based on the threshold size. The crop line detection process was executed sequentially from left to right. It first changed from a black pixel threshold to a white pixel threshold indicating a shift from the background to the rape crop row. The red vertical projection line was detected to the leftmost endpoint. Then, the white pixel threshold was changed back to the black pixel threshold, the first rape crop row was detected, and the blue vertical projection line was extracted to the rightmost endpoint. In the next process the second, third, and fourth rows were detected in the same order as the previous method. Finally, the right-to-left traversal loop verified the previous crop line extraction position. The process is shown in Algorithm 1.

Algorithm 1. End-to-end vertical projection detection of crop rows

Input: Binary thresholded original image
1: Scan through all pixel positions in the image, define trapezoidal ROI region
2: Define function to compute vertical projection of the image, returning vertical projection values
3: Find cropping row ranges based on projection values and threshold
4: Initialize an empty list to store detected cropping row ranges
5: Set starting point marker to None
6: Loop through each pixel value in the vertical projection:
7: if current pixel value is greater than the specified threshold and starting point is not marked:
8: mark starting point i
9: else if current pixel value is less than or equal to threshold and starting point is marked:
10: add range composed of starting point and previous pixel value to cropping row range list
11: reset starting point to None to find next cropping row range
12: Handle the last cropping row range:
13: if starting point is not reset to None, add range composed of starting point and last two pixel values from projection histogram to cropping row range list
14: Return detected cropping row range list
15: End

The vertical projection peaks correspond to the cropping rows, which are numbered from left to right. The left endpoints and right endpoints of the first and fourth cropping rows were recorded in the detection and labeled as the four feature points a, b, c, and d in the order of detection, as shown in Figure 10.

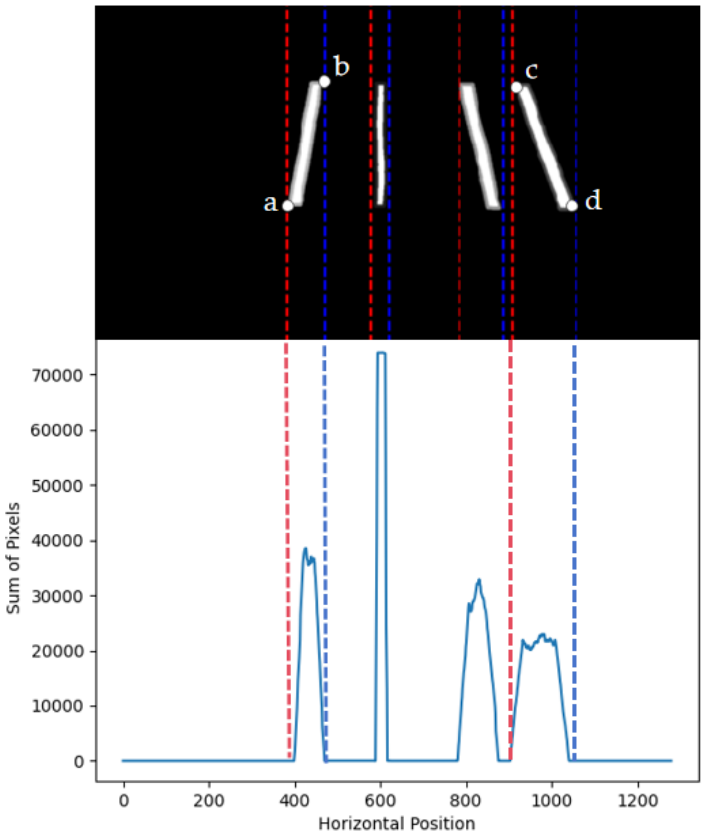

Figure 10. End-to-end vertical projection to determine target row endpoints.

2.5.2. Navigational Line Fitting

The usual methods for fitting navigation lines are the Hough transform [30], random sampling method [31], and least squares method [32]. The Hough transform algorithm is robust to noise and local variations in the image and can also detect curved crop rows in the field. However, the parameter selection and computational complexity is high, which affects the detection speed. The random sampling method of detection is not limited to a specific shape, and has flexibility for linear extraction in different scenes. However, it is suitable for scenes with numerous noise points, making it susceptible to noise sensitivity, potentially reducing extraction accuracy.

The labeling lines of the segmentation targets in this study were made according to the line connecting the start point to the end point. The least squares method is simple and intuitive, with fewer computational parameters, and is more applicable. Based on the trapezoidal region of the upper edge of the two endpoints (b, c) and the bottom edge of the two endpoints (a, d) of the midpoint, we extracted the coordinates of the two points p_m and p_n. The linear fitting formula y = kx + b was used to fit a straight line. The result of the navigation line fitting is shown in Figure 11. The extraction process is shown in Algorithm 2.

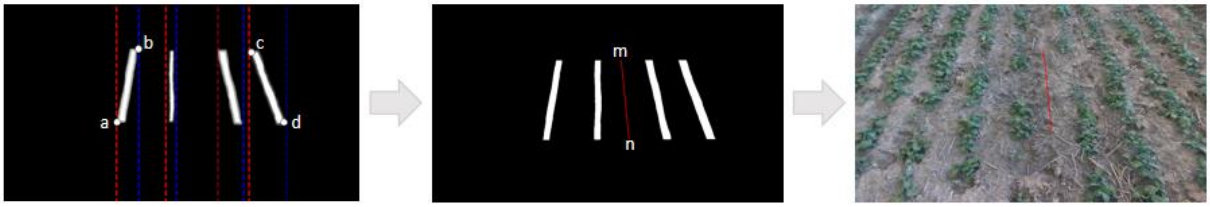

Figure 11. Result of fitted navigation line for oilseed rape crop rows.

Algorithm 2. Navigation Centerline Fitting

Input: Vertical projection located reference lines 1 and 4
1: Sort cropping row start and end points based on vertical projection peaks
2: Define left endpoint of the first cropping row $p_a = (x_a, y_n)$
3: Define right endpoint of the first cropping row $p_b = (x_b, y_m)$
4: Define left endpoint of the fourth cropping row $p_c = (x_c, y_m)$
5: Define right endpoint of the fourth cropping row $p_d = (x_d, y_n)$
6: Use least squares method to fit line connecting midpoint of top and bottom endpoints
$p_m = ((p_b[x_b] + p_c[x_c])//2, y_m)$
$p_n = ((p_a[x_a] + p_d[x_d])//2, y_n)$
7: Calculate linear coefficients k and b of the fitted line
8: Convert the slope to radians
angle radians = arctan(k)
9: Convert the angle from radians to degrees
angle degrees = angle radians × (180/pi)
10: Return angle degrees
11: Check if the angle is less than or equal to 10 degrees
12: if −10 <= angle degrees <= 10
return angle degrees
13: else
return null
14: end if
15: End

3. Results

3.1. Network Model Training

To enhance the VC-UNet model's learning of labeling information from the oilseed rape crop row dataset, adjustments to dataset-related parameters were made during model training.

Dice Loss and Focal Loss loss functions were used during model training to solve the problem of having a small number of training samples and an imbalance between positive and negative samples of training categories. Meanwhile, GPU NVIDIA CUDA (version 11.8) and CPU processor were enabled to accelerate the training speed of the model. The learning rate of the Adam tunable optimizer was set to 0.0001, the batch size of the freeze phase was set to eight, and the batch size of the unfreeze phase was set to four. The whole process was only slightly adjusted on the feature extraction network. The training process consisted of 300 iterations, with weights saved every 10 cycles. After pruning and training, the size of the model weights file was reduced to 1/9 of the original, achieving model lightweighting.

The loss function curves during the training process of U-Net, Pspnet, Deeplab V3+, and VC-UNet are shown in Figure 12. The cosine decay method was used, and the curve direction showed a decreasing trend with the increase of the number of iterations. The number of iterations in the freezing phase of training was 80, and the loss function decreased more obviously from the unfreezing phase, and then tended to be smooth. When the number of iterations reached 250, the curve direction of each model basically tended to

be parallel. There was no overfitting phenomenon in the training process, indicating that the accuracy of the training and validation sets increased steadily.

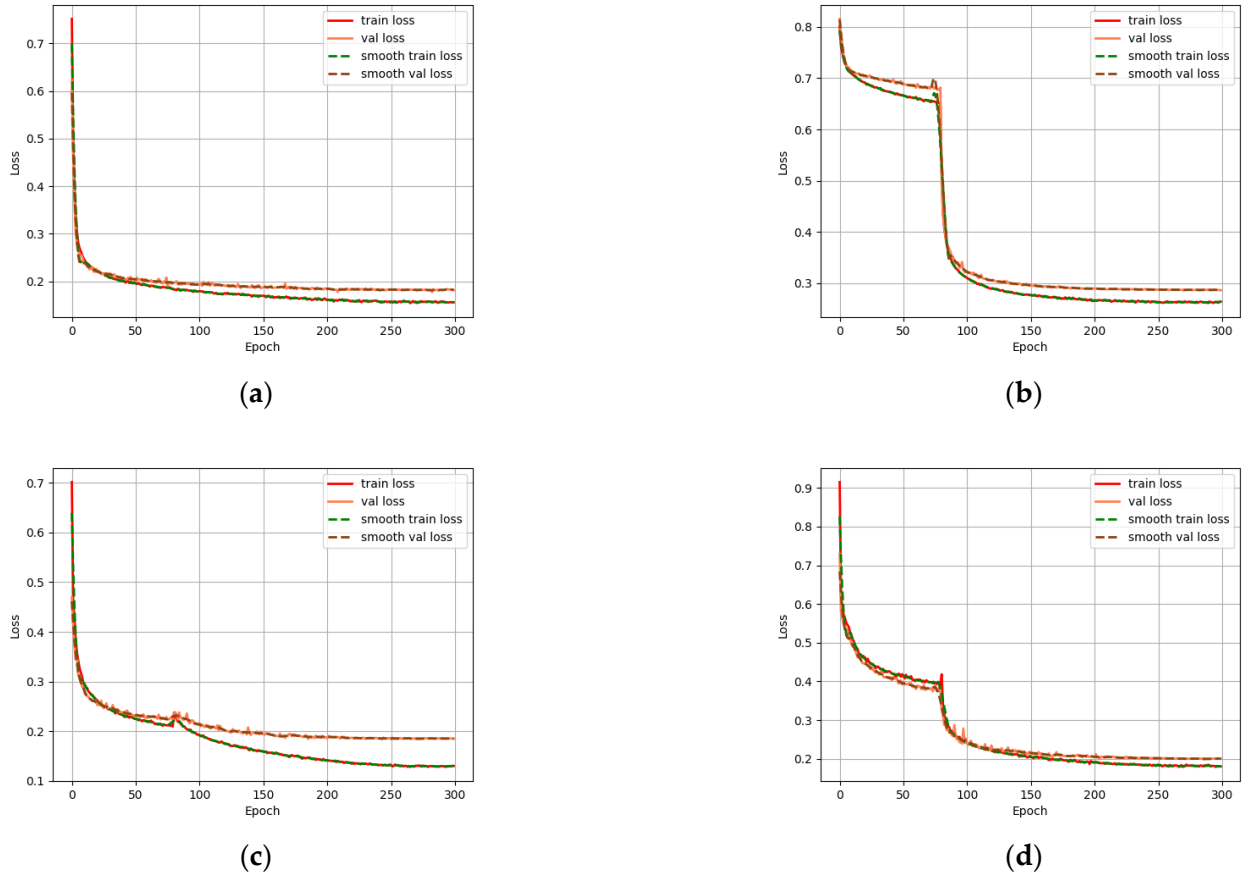

Figure 12. Training loss function curves for each model: (**a**) Deeplab v3+ model; (**b**) Pspnet model; (**c**) U-Net model; (**d**) VC-UNet model(ours).

3.2. Segmentation Accuracy Test

The VC-UNet model was experimentally controlled with the semantic segmentation network models U-Net, Pspnet, and Deeplab V3+, respectively, to verify the effect of recognizing the crop rows of oilseed rape. The mean pixel accuracy (MPA) and the mean intersection over union (MIOU), which are the evaluation indices of semantic segmentation models, were used to evaluate the accuracy of crop row segmentation. The calculations are shown in Equations (1) and (2).

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

$$\mathrm{MIOU} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{2}$$

where p_{ii} is the number of correctly predicted pixels, p_{ij} is the number of pixels with true value of i but predicted as j, p_{ji} is the number of pixels with true value of j and predicted as i, and k is the category value.

With the same number of training model parameters, the above three models were selected for comparison with the model studied in this paper through performance evaluation metrics such as MPA, MIOU, and detection speed. Table 2 shows the performance parameters of each network model.

Table 2. Comparison of model performance parameters.

Semantic Segmentation Model	MPA/%	MIOU/%	Detection Speed/fps·s^{-1}
U-Net	90.15	87.3	13.62
Pspnet	89.78	83.2	11.58
Deeplab V3+	92.42	86.57	15.02
VC-UNet	94.11	90.97	24.47

The VC-UNet model achieved an MPA of 94.11% and MIOU of 90.97%, while significantly increasing the processing speed to 24.47 fps/s. It outperformed U-Net, Pspnet, and Deeplab V3+. The majority of scene segmentation and parsing frameworks within Pspnet are based on the FCN network, which was deficient in terms of fineness and had the lowest detection accuracy. Our training parameters were comparable to those seen in the other three model metrics.

The soybean corn composite planting crop row dataset has increased crop type compared to the oilseed rape dataset. There is also a certain amount of weed interference in the soybean corn field environment, which can vary in accuracy compared to the canola crop row dataset. As shown in Table 3, for each model, accuracy metrics for soybean corn composite planting crops were obtained after transfer learning. The accuracy of models trained with transfer learning exceeded those without. According to the table data, the VC-UNet model showed superior transfer learning effects compared to Pspnet, Deeplab V3+, and U-Net. The MPA after transfer learning exceeded 90%, indicating that the VC-UNet model exhibits enhanced reliability in crop row segmentation training.

Table 3. Comparison of model transfer learning MPA results.

Crop Category	U-Net	Pspnet	Deeplab V3+	VC-UNet
Soybean corn (transfer learning)	87.02	83.65	88.74	91.57
Soybean corn (no transfer learning)	80.97	75.21	83.58	89.43

To evaluate the impact of the training model on crop row segmentation in oilseed rape, we extracted 100 images from each of the six categories in the aforementioned oilseed rape dataset to validate the prediction results of crop row segmentation in the field. The effect of crop row category segmentation in an oilseed rape field is shown in Figure 13. The crop row extraction results of the VC-UNet model under the three light intensity variations were better in Figure 13a–c. The U-Net, Pspnet, and Deeplab V3+ models exhibited minor deficiencies in breakpoint handling, smoothness, and edge processing, respectively. The strip crop ridges in Figure 13d and e indicated no effect on the extraction effect of crop row segmentation in oilseed rape. There was a shadow occlusion in Figure 13f, and the models lacked the effect on the processing of crop rows that were closely connected at the edges of the image. The region with indistinct boundaries at the edges had a negligible impact on the extraction of the ROI region of the navigation line. Furthermore, it fulfilled the processing requirements for smooth articulation in the segmentation of discontinuous crop row regions.

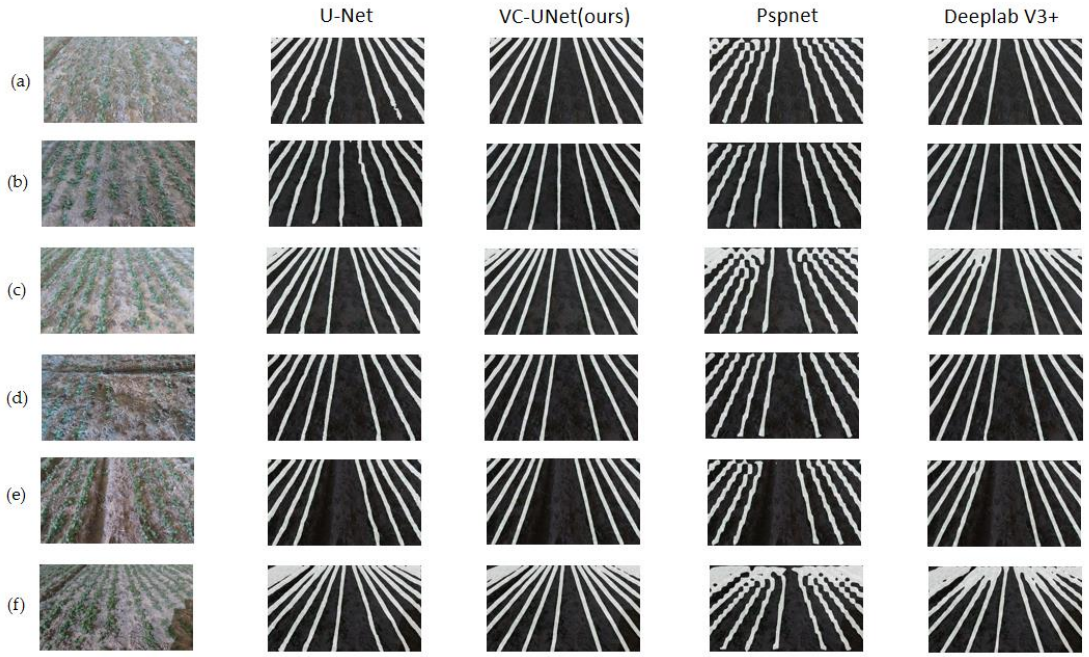

Figure 13. Segmentation effect of each model in oilseed rape crop row: (**a**–**c**) Different lighting environment; (**d**,**e**) Crop ridge environment; (**f**) Shadow masking environment.

3.3. Navigation Line Test Results

In this study, the least squares method was used to compare with both the traditional Hough transform and random sampling method to verify the accuracy of the line fitting algorithms. A random selection of 100 crop line images from each of the three methods was conducted, and the fitted lines with yaw angles greater than 10° were excluded to test the accuracy of navigation line extraction. As shown in Table 4, the results indicate that the fitting accuracy of the least squares method was better than that of the Hough transform and random sampling method. The average yaw angle was found to be 3.58°, with a navigation line detection accuracy of 92.43% after exclusion of yaw angle fluctuations.

Table 4. Comparison of navigation line fitting methods.

Method	Image Quantity	Average Yaw Angle/°	Detection Accuracy/%
Least squares method	100	3.58	92.43
Hough transform	100	7.88	86.55
Random sampling method	100	10.62	75.31

Three environments of low light, strong light, and normal light were selected as shown in Figure 14 to test the effect of navigation line extraction. Table 5 gives the parameter comparison of the least squares method for calculating the navigation centerline in different light intensity environments. The minimum average yaw angle was 2.91° in the balanced light environment and the maximum average yaw angle was 4.93° in the bright light environment. The average deviation of the yaw angle was 3.76° under different light conditions. The average pixel deviation was 4.98 pixels for a balanced lighting environment and 7.83 pixels for a strong lighting environment. The average pixel deviation under different lighting was 6.13 pixels, and the navigation line extraction accuracy was high. The average time consumed for a single image in the three different lighting environments was 0.009 s, which indicated that the image reception and processing speed of the navigation line parameter extraction met the demand of real-time navigation.

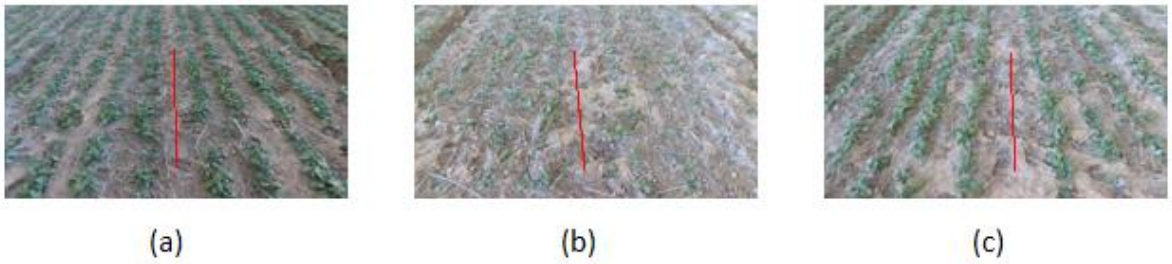

Figure 14. Least squares navigation line extraction results under different lighting: (**a**) Low light; (**b**) Strong light; (**c**) Normal light.

Table 5. Comparison of parameter extraction for navigation lines in different environments.

Extraction Environment	Average Pixel Deviation	Average Yaw Angle/°	Average Single Image Time-Consume/s
Low light	5.58	3.45	0.008
Strong light	7.83	4.93	0.013
Normal light	4.98	2.91	0.006
Average deviation	6.13	3.76	0.009

4. Discussion

In contrast to previous line fitting extraction algorithms, the least squares method is fully adapted to straight line extraction between rows of oilseed rape crops. After eliminating fluctuations, its detection accuracy improved by 5.88% and 17.12% compared to the Hough transform and random sampling methods, respectively. In addition, the results of average yaw angle and average pixel deviation under three different lighting environments showed that the navigation line extraction accuracy can meet the requirements of visual navigation of agricultural robots in the field. The end-to-end vertical projection method proposed in this paper took the shortest time to receive a single image under normal light, compared to strong and low light. This may be because both crop rows and background colors are darker in low light, making it difficult to differentiate them, while crop rows are difficult to identify in strong light due to noise. Yang et al. [32] used the U-Net model after replacing the backbone network with VGG16 with an MPA of 97.29% and a detection speed of 12.62 fps/s. In contrast, our improved semantic segmentation model extracted rape crop rows with a slightly lower MPA and a higher detection speed. This may be because model pruning ws carried out in this study, while the number of crop row strips in a single image was high, and the boundary of the edge region was not obvious. The pruning also improves the model detection speed. Meanwhile, the results after transfer learning show that the model can be adapted to other crop rows with similar traits. Some soybean corn plants develop faster during the seedling growth stage compared to the oilseed rape seedling stage, resulting in a slight difference between soybean corn plant height and oilseed rape plant height. For the soybean and corn crop row image of the target pixel closer to the location of the boundary recognition will have a small amount of residuals. Therefore, there will be some obstacles in the transfer learning process. However, the accuracy before and after transfer learning using the VC-UNet model is close to 90%. On the one hand, it is verified that the segmentation effect of soybean and corn crop rows is improved after transfer learning from the rape dataset. On the other hand, it also proves that the model has strong robustness and generalization ability and has better results for segmentation recognition of different crops.

In response to the soil topography factor, the oilseed rape dataset has striped crop ridges, a trait that is more prominent in the images. This trait interferes with crop row identification in oilseed rape, in particular, for vertical crop ridges, where some oilseed rape crop rows are planted very close to them. The trained segmentation results effectively exclude the unfavorable factors due to the presence of ridges in the soil.

Although our lightweight model was able to predict the crop rows of oilseed rape with high accuracy, it suffered from the drawback of a small number of datasets. In addition, the oilseed rape dataset only represented crop row characteristics at the seedling stage of oilseed rape, and other growth stages were not taken into account. Expansion of the oilseed rape field dataset should be considered in future research and further deployment of the methods of this study into agricultural robotic systems for field testing.

5. Conclusions

This study proposed a semantic segmentation algorithm model based on U-Net, which had been improved upon to create VC-UNet, for the purpose of recognizing and detecting rape crop rows. The reliability of the model was verified by model comparison and transfer learning, and the effective extraction of the navigation lines of rape crop rows under three different lighting environments was realized. The main conclusions were as follows:

(1) This study proposed a lightweight VC-UNet semantic segmentation model based on U-Net research. The original backbone feature extraction network of U-Net was replaced by VGG16 to accelerate network training. Furthermore, a convolutional block attention module (CBAM) was added to the up-sampling part of the feature extraction network to increase the attention to the segmentation target region. Finally, the number of channels in the convolutional layer of the network was pruned to achieve a lightweight model, with the memory consumption of the trained weights file reduced to 1/9 of the original. The average accuracy of the model is 94.11%, with a processing speed of 24.47 fps/s. The computational results were significantly better than the three network models of U-Net, Pspnet, and Deeplab V3+, confirming the strong robustness of the model.

(2) The training results of the rape crop row dataset were migrated to soybean corn composite planting crop rows through the introduction of transfer learning. The results demonstrated that the average accuracy of soybean and corn crop rows after transfer learning reaches 91.57%, exhibiting a superior segmentation effect. Consequently, this methodology may be applied to other categories of crop rows for visual navigation.

(3) An end-to-end vertical projection method was proposed based on the effect of crop row segmentation in oilseed rape in this study. The crop rows were sorted and localized to the navigation line extraction location by detecting the threshold information. Subsequently, the least squares method was employed to fit the navigation line, as it was determined to be the most accurate detection method. The parametric performance of the navigation line extraction yaw angle under three different lighting conditions was verified, with an average yaw angle of 3.76°, an average pixel offset of 6.13 pixels, and an average single image transmission time of 0.009 s. The results of the study showed the ability to meet the real-time and accuracy requirements of visual navigation for agricultural robots.

In the future, we will deploy the oilseed rape crop row segmentation model with the navigation line extraction method to the robot platform. We will establish communication with the robot's position information through the navigation line parameters, and then find suitable control algorithms for crop row path tracking to ensure that the robot travels along the crop rows. In the subsequent process, we will carry out field tests and conduct in-depth research on visual navigation and multi-sensor fusion algorithms to improve the robot's adaptability and operational efficiency in various environments.

Author Contributions: Conceptualization, G.L. and X.X.; methodology, G.L., X.X. and L.C.; software, G.L. and S.S.; validation, G.L. and L.C.; formal analysis, G.L., F.L. and S.S.; investigation, G.L., L.C. and F.L.; resources, X.X., L.C. and F.L.; data curation, G.L., L.C. and X.X.; writing—original draft preparation, G.L.; writing—review and editing, G.L. and L.C.; visualization, G.L. and S.S.; supervision, X.X. and L.C.; project administration, X.X. and L.C.; funding acquisition, X.X. and L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the National Key R&D Program of China (No. 2022YFD2000700); Innovation Program of Chinese Academy of Agricultural Sciences (No. CAAS-SAE-202301); the China Modern Agricultural Industrial Technology System (grant number CARS-12) and Jiangsu Province and Education Ministry Cosponsored Synergistic Innovation Center of Modern Agricultural Equipment Project (No. XTCX1004).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, R.; Feng, Z. Contribution rate of elements to China's rapeseed output growth per unit area. *Chin. J. Oil Crop Sci.* **2010**, *32*, 152–155.
2. Cheng, L.; Jie, H.; Bofeng, L.; Zhongchao, F.; Junpeng, L. Current situation, development difficulties and suggestions of Chinese rape industry. *J. China Agric. Univ.* **2017**, *22*, 203–210.
3. Rondanini, D.P.; Gomez, N.V.; Agosti, M.B.; Miralles, D.J. Global trends of rapeseed grain yield stability and rapeseed-to-wheat yield ratio in the last four decades. *Eur. J. Agron.* **2012**, *37*, 56–65. [CrossRef]
4. Xie, B.; Jin, Y.; Faheem, M.; Gao, W.; Liu, J.; Jiang, H.; Cai, L.; Li, Y. Research progress of autonomous navigation technology for multi-agricultural scenes. *Comput. Electron. Agric.* **2023**, *211*, 107963. [CrossRef]
5. Ball, D.; Upcroft, B.; Wyeth, G.; Corke, P.; English, A.; Ross, P.; Patten, T.; Fitch, R.; Sukkarieh, S.; Bate, A. Vision-based obstacle detection and navigation for an agricultural robot. *J. Field Rob.* **2016**, *33*, 1107–1130. [CrossRef]
6. Ma, Z.; Yin, C.; Du, X.; Zhao, L.; Lin, L.; Zhang, G.; Wu, C. Rice row tracking control of crawler tractor based on the satellite and visual integrated navigation. *Comput. Electron. Agric.* **2022**, *197*, 106935. [CrossRef]
7. Bakker, T.; van Asselt, K.; Bontsema, J.; Müller, J.; van Straten, G. Autonomous navigation using a robot platform in a sugar beet field. *Biosyst. Eng.* **2011**, *109*, 357–368. [CrossRef]
8. Wang, T.; Chen, B.; Zhang, Z.; Li, H.; Zhang, M. Applications of machine vision in agricultural robot navigation: A review. *Comput. Electron. Agric.* **2022**, *198*, 107085. [CrossRef]
9. Li, Y.; Wang, X.; Liu, D. 3D autonomous navigation line extraction for field roads based on binocular vision. *J. Sens.* **2019**, *2019*, 6832109. [CrossRef]
10. Utstumo, T.; Urdal, F.; Brevik, A.; Dørum, J.; Netland, J.; Overskeid, Ø.; Berge, T.W.; Gravdahl, J.T. Robotic in-row weed control in vegetables. *Comput. Electron. Agric.* **2018**, *154*, 36–45. [CrossRef]
11. English, A.; Ross, P.; Ball, D.; Corke, P. Vision based guidance for robot navigation in agriculture. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 1693–1698.
12. Radcliffe, J.; Cox, J.; Bulanon, D.M. Machine vision for orchard navigation. *Comput. Ind.* **2018**, *98*, 165–171. [CrossRef]
13. Ma, Z.; Mei, G. Deep learning for geological hazards analysis: Data, models, applications, and opportunities. *Earth Sci. Rev.* **2021**, *223*, 103858. [CrossRef]
14. Grigorescu, S.; Trasnea, B.; Cocias, T.; Macesanu, G. A survey of deep learning techniques for autonomous driving. *J. Field Rob.* **2020**, *37*, 362–386. [CrossRef]
15. Muhammad, K.; Ullah, A.; Lloret, J.; Del Ser, J.; de Albuquerque, V.H.C. Deep learning for safe autonomous driving: Current challenges and future directions. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 4316–4336. [CrossRef]
16. Xinyu, Z.; Hongbo, G.; Jianhui, Z.; Mo, Z. Overview of deep learning intelligent driving methods. *J. Tsinghua Univ. Sci. Technol.* **2018**, *58*, 438–444.
17. Ker, J.; Wang, L.; Rao, J.; Lim, T. Deep learning applications in medical image analysis. *IEEE Access* **2017**, *6*, 9375–9389. [CrossRef]
18. Gong, H.; Zhuang, W. An Improved Method for Extracting Inter-row Navigation Lines in Nighttime Maize Crops using YOLOv7-tiny. *IEEE Access* **2024**, *12*, 27444–27455. [CrossRef]
19. Ju, J.Y.; Chen, G.Q.; Lv, Z.Y.; Zhao, M.Y.; Sun, L.; Wang, Z.T.; Wang, J.F. Design and experiment of an adaptive cruise weeding robot for paddy fields based on improved YOLOv5. *Comput. Electron. Agric.* **2024**, *219*, 17. [CrossRef]
20. Bah, M.D.; Hafiane, A.; Canals, R. CRowNet: Deep Network for Crop Row Detection in UAV Images. *IEEE Access* **2020**, *8*, 5189–5200. [CrossRef]
21. De Silva, R.; Cielniak, G.; Wang, G.; Gao, J. Deep learning-based crop row detection for infield navigation of agri-robots. *J. Field Rob.* **2023**, *23*. [CrossRef]
22. Adhikari, S.P.; Kim, G.; Kim, H. Deep Neural Network-Based System for Autonomous Navigation in Paddy Field. *IEEE Access* **2020**, *8*, 71272–71278. [CrossRef]
23. Qingkuan, M.; Jie, H.; Ruicheng, Q.; Xiaodan, M.; Yongsheng, S.; Man, Z.; Gang, L. Crop Recognition and Navigation Line Detection in Natural Environment Based on Machine Vision. *Acta Opt. Sin.* **2014**, *34*, 180–186.
24. Mo, Y.; Wu, Y.; Yang, X.; Liu, F.; Liao, Y. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neural Comput.* **2022**, *493*, 626–646. [CrossRef]
25. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.

26. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
27. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access* **2021**, *9*, 82031–82057. [CrossRef]
28. Liu, Z.; Sun, M.; Zhou, T.; Huang, G.; Darrell, T. Rethinking the value of network pruning. *arXiv* **2018**, arXiv:1810.05270.
29. Bosilj, P.; Aptoula, E.; Duckett, T.; Cielniak, G. Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture. *J. Field Rob.* **2020**, *37*, 7–19. [CrossRef]
30. Fengrong, S.; Jiren, L. Fast Hough Transform Algorithm. *Chin. J. Comput.* **2001**, *24*, 1102–1109.
31. Xiao, K.; Xia, W.; Liang, C. Visual Navigation Path Extraction Algorithm in Orchard under Complex Background. *Trans. Chin. Soc. Agric. Mach.* **2023**, *54*, 197.
32. Yang, R.; Zhai, Y.; Zhang, J.; Zhang, H.; Tian, G.; Zhang, J.; Huang, P.; Li, L. Potato visual navigation line detection based on deep learning and feature midpoint adaptation. *Agriculture* **2022**, *12*, 1363. [CrossRef]