VEHICLE MASSIVE IMAGE DATA FILTERING AND USELESS IMAGE REUSE BASED ON FARMLAND BACKGROUND ANALYSIS
/
基于农田背景分析的车载大数据图像过滤与无效图像再利用

Hanlu JIANG, Fengzhu WANG, Gaoyong XING, Yangchun LIU, Weipeng ZHANG, Liming ZHOU
State Key Laboratory of Agricultural Equipment Technology, Chinese Academy of Agricultural Mechanization Science Group Co., Ltd. Beijing 100083, China;
Tel: +86-(010)64882659; E-mail: liuyangchun@caams.org.cn
Correspondent author: Yangchun Liu
DOI: https://doi.org/10.35633/inmateh-74-20

Keywords: *Image filtering; Background recognition; Effective farmland segmentation*

ABSTRACT
The real-time images captured by agricultural machinery on-board monitoring equipment have complex backgrounds and different shooting angles. Especially for straw monitoring tasks, the utilization rate of images is relatively low. This paper presents a novel image classification and effective region segmentation method for straw returning in agriculture, leveraging semantic segmentation to enhance the efficiency of agricultural data analysis. The study addresses the challenges of manual straw cover analysis by proposing an automated approach to select images that meet monitoring standards. The methodology employs an encoder-decoder structure model, enriched with residual units, multi-scale convolution, and attention mechanisms. This model classifies images by calculating the pixel proportions of various scene categories and segments farmland areas to be inspected by incorporating distance information. The model's design is tailored to handle the complex and variable natural environments typical of vehicular monitoring scenarios, where semantic object boundaries can be fuzzy. The experimental results demonstrate that the proposed method achieves an overall sample classification accuracy of 93% for straw returning image classification and an 85.37% accuracy in dividing areas to be inspected. The method outperforms several mainstream semantic segmentation models, providing a more accurate and efficient means of processing agricultural monitoring images. The integration of distance information proves particularly beneficial in distinguishing the farmland areas under inspection, leading to clearer segmentation and more reliable data for agricultural decision-making. In conclusion, the study contributes to the field of agricultural intelligence by offering a robust method for image analysis that can be applied to optimize the use of straw return monitoring data.

摘要
农机车载监控设备拍摄的实时图像背景复杂，特别是对于秸秆定量检测任务，图像的利用率相对较低。本文提出了一种基于语义分割的图像高效分类和背景区域分割方法，通过计算各种场景类别的像素比例对图像进行分类，筛选出符合秸秆检测要求的图像，再并结合距离信息进一步分割秸秆检测区域，提高车载图像数据分析利用效率与秸秆检测准确度，以此来解决农田大数据杂乱，利用率低的问题。该方法采用编码器-解码器模型结构，融合了轻量型残差单元、多尺度卷积和注意力机制，在保证分割边界清晰的情况下降低模型参数。从不同模型的对比结果和可视化处理结果可以表明：该方法对秸秆图像分类的总体样本分类准确率为93%，对待检区域的划分准确率为 85.37%，该模型对背景类别中农田的分割效果更好，空间位置更加准确。因此，本文的研究方法能提高秸秆信息化监测手段的应用效率，为农田监测图像的处理提供了一种更准确、更高效的方法。

INTRODUCTION
In recent years, with the aim of protecting the precious and rare black soil resources on Earth, there has been a particular emphasis on promoting conservation tillage techniques, such as straw mulch cover and no-tillage *(Yu et al., 2015)*, supported by information technology-enabled remote monitoring to enhance agricultural production. Currently, the monitoring terminals for conservation tillage have achieved the basic collection, transmission, and storage of agricultural big data, obtaining vast amounts of data on no-tillage machinery operation conditions and field straw vehicle monitoring images. However, for calculating the straw cover rate, data still require manual organization and human analysis, resulting in limited data application depth and low model technical content.

Therefore, it is imperative to strengthen the application research of data analysis technology in the context of conservation tillage, efficiently utilize straw cover images, and automatically select straw quantitative detection images that meet monitoring standards *(Wu et al., 2017)*.

Recent research findings indicate that in order to improve detection accuracy, image semantic segmentation methods are transitioning from probabilistic graphical models to deep learning models *(Habas et al., 2010; Chung et al., 2010; Liu et al., 2010; Long et al., 2015; Chen et al., 2018; Litjens et al., 2017; Qi et al., 2017; Jiang et al., 2020)*. For instance, Zhou proposed a method that combines fully convolutional networks with conditional random fields for image semantic segmentation, merging probabilistic graphs with deep learning while maintaining computational sensitivity and global consistency *(Zhou et al., 2016)*. Ronneberger introduced a U-Net structure that effectively utilizes annotated samples through data augmentation, achieving end-to-end model training with fewer samples *(Ronneberger et al., 2015)*. Liu Kaidong employed U-Net as the backbone network and introduced a model based on R2U-Net and compact dilated convolution for segmenting and recognizing target muscle areas *(Liu et al., 2020)*. Badrinarayanan presented a fully convolutional neural network architecture called SegNet with an encoder-decoder structure *(Badrinarayanan et al., 2017)*. The encoder is used for object information parsing, while the decoder employs upsampling to restore the reduced feature maps to their original size. Lin introduced a universal multi-path optimization network, RefineNet, this model utilizes remote residual connections to capture all available information during the downsampling process, directly capturing high-level semantic features *(Lin et al., 2017)*. Zhao Shida applied generative adversarial networks and transfer learning to train ICNet, enabling fine segmentation of sheep skeletons' spine, ribs, and neck under different lighting conditions *(Zhao et al., 2021)*. Fang Peng integrated attention mechanisms, residual networks, and feature pyramids to extract image features, using a region generation network for broiler target classification, segmentation, and contour extraction *(Fang et al., 2021)*. Liu Wenya applied semantic segmentation to remote sensing images, proposing an automated extraction architecture based on DeepLabv3+ that can eliminate interference from field pixels and accurately extract urban green spaces *(Liu et al., 2021)*.

Although the above methods have solved the problem of detection accuracy in specific fields, there are still shortcomings in the domain of agricultural machinery onboard monitoring. Image quality is the key to straw detection, but it is often overlooked.

In view of the above problems, this paper addresses the impact of changes in the angle of in-vehicle cameras on the proportion of non-farmland backgrounds within the visual range and the visual distance on the division of areas to be inspected. It introduces a method for sample classification and the division of areas to be inspected based on semantic segmentation. Building upon the encoder-decoder structure, the paper incorporates key techniques such as residual units, multi-scale convolution, and attention mechanisms. It determines image categories by calculating the proportion of various background pixels in the image and simultaneously defines distance levels for the spatial distribution of farmland, establishing clear boundaries for the division of areas to be inspected in farmland.

MATERIALS AND METHODS

Problem analysis

The prerequisite for the application of straw return monitoring data is that the proportion of the fields to be inspected within the image range is large, and the distribution of straw is clear. It is common to retrofit agricultural machinery with rear-mounted cameras to make them intelligent. However, the stability and consistency of image quality are often compromised due to variations in the installation position and angle of the cameras.

Images captured by monocular cameras mostly follow a long-tail distribution, with areas closer to the camera having higher resolution, making it difficult to clearly observe the distribution of straw in distant pixels *(Sreepada et al., 2020)*. When the angle of view is too low, the machinery occupies the lower and middle parts of the image, obscuring the effective monitoring area of the field. When the angle of view is too high, the image may display a prominent skyline, the agricultural field area is reduced, and distant field objects appear blurred. Both of these image types fail to meet the requirements for straw monitoring.

As shown in Fig. 1, the left image illustrates the schematic of the manual method for measuring straw cover percentage using a rope. It uses a 50-meter rope divided into 250 collection points on average, it is placed along the diagonal of the plot to be tested, and the percentage of collection points where straw is present is calculated to determine the straw cover percentage of that plot. According to the area calculation formula, this method is most efficient when the plot is a square.

The 250 points can represent the collection area composed of 31,250 points. These collection points can be analogized to image resolution. Given that the YB-DB015A onboard camera has a spatial resolution of 320×240, it ensures that the agricultural background in the image is not less than 40%. As shown in the two images on the right, the intersection point of the camera's optical axis with the ground is always located at the center of the image. Therefore, as long as the center distance "c" exceeds 15 meters, it ensures that the pixel proportion of the areas to be inspected in the image is not less than 40%. Thus, based on the correspondence between the horizon vanishing point, the field of view angle "φ," and the pitch angle "σ," the maximum proportion of sky background at this point is 30%, and the machinery occupies no more than 10%.

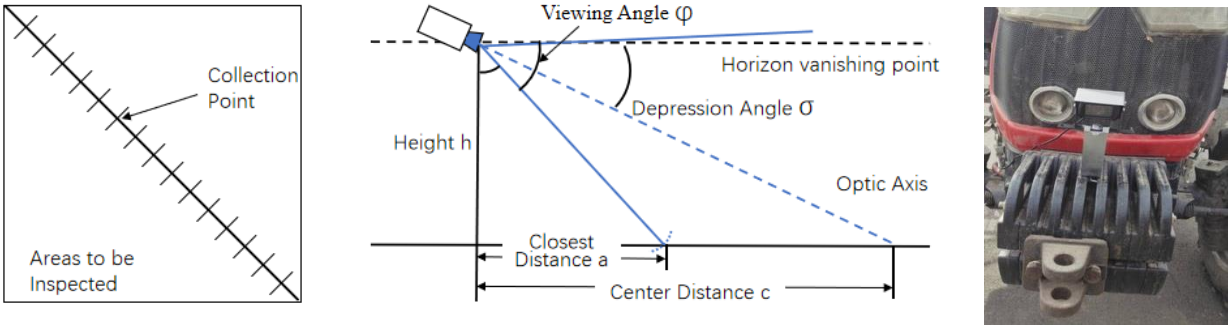


Fig. 1 - Pull rope method and camera imaging principle

Therefore, the image classification standards in this article are different from traditional images, it does not classify images based on scene categories but rather relies on combinations of pixel proportions in different scenes, images are divided into four categories *(Xiao et al., 2015; Liang et al., 2020; Miao et al., 2015)*. Pixel-level semantic segmentation method perform both image classification and segmentation tasks simultaneously *(Minaee et al., 2020; Jadhavo et al., 2018)*.

Experimental Data

To illustrate and verify the farmland background recognition and segmentation method proposed in this study, 1700 images were collected from straw application field monitored by image grabbing in Da'an City, Siping, Jilin province from April 15 through June 20, 2021. The conservation tillage types in Jilin province were Minimum-Tillage and No-Tillage of straw land covering. In order to meet the needs of straw monitoring, the monitoring terminals which are about more than 50 are installed on No-tillage planters, and the images were collected on sunny or daytime days at different times and environments. A YB-DB015A camera with a resolution of 320×240 was used and placed on front of the cab roof and initially tilted down 60 degrees toward the ground to get the straw image in front of the planter.

Fig. 2 - No-tillage layer in Da'an City

Based on the pixel proportions of different categories in the agricultural scene, sample images are divided into four classes (C1: high angle; C2: low angle; C3: standard image; C4: to-be-processed image). Each class has 365, 284, 325, and 726 images, respectively. As shown in Fig. 3, C1 images have a sky proportion greater than 30%, indicating a high angle that does not meet monitoring requirements.

C2 images have a machinery proportion exceeding 10%, indicating a low angle that does not meet monitoring requirements. C3 images have an agricultural field proportion of approximately 90%, representing standard images for straw cover monitoring. C4 images have a sky proportion less than 30% and a machinery proportion less than 10%, indicating to-be-processed images that can meet monitoring requirements through the division of areas to be inspected in the field.

(A) C1: High angle

(B) C2: Low angle

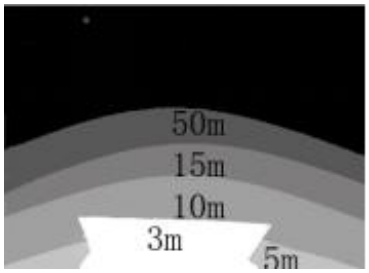
(C) C3:Standard image

(D) C4:To be processed

Fig. 3 - Typical images of four categories

All images were marked in LABELME and saved in red, green, and yellow in JSON format. In order to get label data, JSON was converted to single channel image. Moreover, for the fuzzy farmland in the distance (unable to distinguish straw distribution), distance information was used as supplementary information to segment effective farmland. The label picture and distance hierarchical map was showed in Fig. 4.

(A) Label image



(B) Distance hierarchical map

Fig. 4 - Annotated images of samples

Furthermore, the hardware setup used in this study consisted of a Precision T7920 tower workstation equipped with two Intel Xeon SP series processors and an Nvidia Quadro P2200 graphics card. The operating system utilized was Ubuntu Linux 18.04, and the development software included PyCharm and the PyTorch framework.

Model structure analysis

The design of model is an encoder-decoder structure, including a lightweight and efficient feature extraction module and a decoding module based on multi-scale convolution and ordinary convolution as shown in Fig. 5. In order to extract the background features in filed, the encode module consists of three layers. Thus, the image was reduced to 80×60 pixels but increased to 128 features in the end by convolution and down sampling. The decode module includes multiple channels to analyze the semantic information between multi-scale features, wherefore the background label and distance level of the pixel are determined.

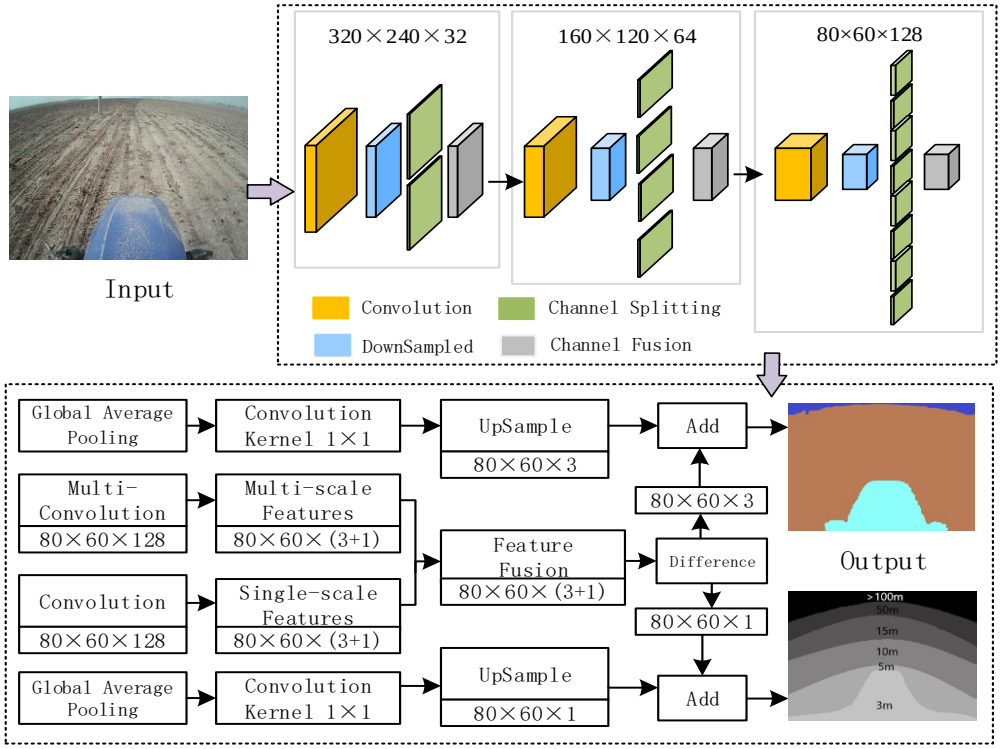


Fig. 5 - General structure diagram of the model

Small paraments encoder

In order to facilitate real time computation in filed, reducing the number of parameters and the calculation process are the key to model construction. In this study, with the increase of the number of encode layer, convolution kernel size and channel number play a major role. Assuming the optimization unit is not considered, the number of parameters is described as follows:

$$N_p \sim O\left(\sum_{l=1}^{D} K_{lx} \cdot K_{ly} \cdot C_{l-1} \cdot C_l\right) \quad (1)$$

where, C_{l-1} is the number of input channel; C_l is the number of output channel; K_{ly} and K_{lx} are the size of convolution kernel.

Then, the amount of calculation can be expressed as:

$$N_c \sim O\left(\sum_{l=1}^{D} K_{lx} \cdot K_{ly} \cdot M_{lx} \cdot M_{ly} \cdot C_{l-1} \cdot C_l\right) \quad (2)$$

where, M_{lx} and M_{ly} are the size of feature map in layer l.

Suppose the parameters of hybrid dilated convolution are not considered, the size of convolution kernel is 3×3, and the input channels are divided into two branches. Methods after improvement, the product of channels is reduced by 4 times, and convolution kernel is reduced by 1.5 times. Therefore, after the two channels are added, the parameters and the calculation are reduced to one third of the former one.

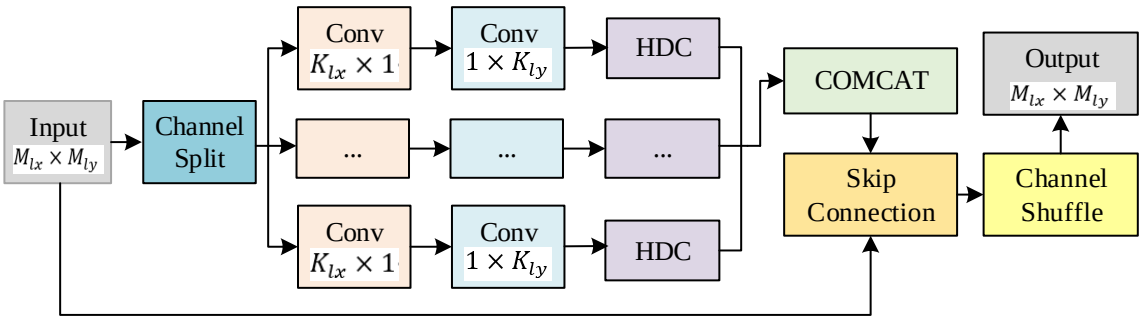


Fig. 6 - Residual network structure of block convolution

Additionally, in pixel-level semantic segmentation tasks, local information around the prediction points has a significant impact on predicting categories.

To expand the receptive field without sacrificing resolution, a Hybrid Dilated Convolution (HDC) is used to enhance local information, addressing the issue of information loss when restoring the image to its original size in the decoding module through upsampling and pixel addition, as shown in Fig. 7. Finally, after concatenating the channels of multiple branches, the original feature maps and the residuals extracted from each branch are fused in the form of skip connections to restore the original channel count. Due to the lack of correlation in feature dimensions caused by channel branches, it is necessary to shuffle the channel order of feature maps before entering a feature extraction module to enhance the global correlation of feature maps.

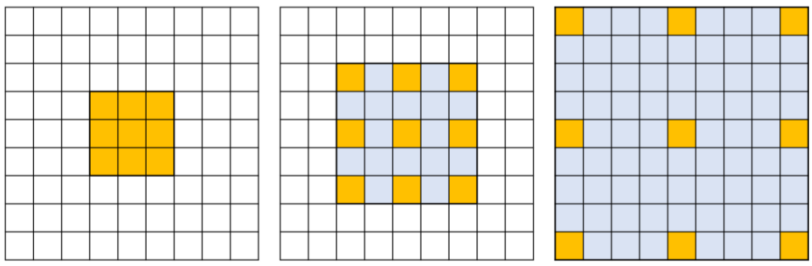


Fig. 7 - Dilation convolution with 3×3 kernel size

Multi-scale convolution decoder

Multi-scale convolution algorithm is a common method to extract different layers of semantic information. The output of the encoder is convoluted in three different scales, for which 7×7, 5×5, and 3×3 convolution kernels were selected, respectively. The convolution kernel with large scale can capture the association between far away pixels, while the convolution kernel with small scale can extract the detail of near small targets. It is easier to analyze near machine and distant sky and farmland by combining convolution kernels of different sizes. Therefore, the size of the feature map is reduced in the initial process. As shown in Fig. 8, because the semantic segmentation task requires that the output size should be the same as the original image, it is necessary to design an upsample to recover the feature map. Subsequently, the same size features are fused by convolution with 1×1 convolution kernel. After 3 times of upsampling and 2 times of feature fusion, the output is the same size as the input feature map and semantic channel.

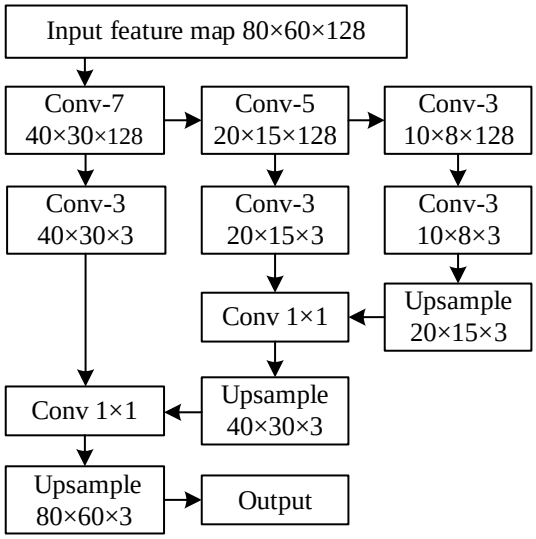


Fig. 8 - Structure of multi-scale convolution

Effective region segmentation based on distance information

Distance estimation and semantic segmentation are both of scene perception. Distance estimation describes the geometric relationship in space, while semantic information represents the entity meaning of different background of the scene. Therefore, the distance information is introduced into the decode module as the auxiliary information, both of which share context information to the purpose of segmenting effective farmland. The decoder consists of four branches, two of which are multiscale convolution module and normal convolution module. Except for the three semantic channels mentioned in the previous section, a distance channel is added.

After point multiplication, the output of the two modules is divided into semantic and distance parts. The other two branches complete global average pooling (GAP), convolution with 1×1 and up sampling in turn to retain the uniqueness of the two types of information. Finally, the semantic prediction results and distance classification results are obtained by adding the outputs of the two kinds of information.

RESULTS

Case Analysis Impact of Residual Networks on Training Accuracy

To address the increased difficulty in training deep networks, residual networks were introduced in this model to enhance the fitting capability of high-dimensional models, and improvements were made to the convolutional kernels and residual branches of the residual network, resulting in a three-fold reduction in the number of parameters. To verify the impact of introducing residual modules and improved residual modules on training accuracy, this section compares the training iteration accuracy curves for different models to highlight the advantages of residual modules. The experimental results are shown in Fig. 9. It can be observed from the curve variations that the model without residual modules exhibits lower training speed and iteration accuracy compared to the results with the introduction of residual modules. Additionally, the improved residual module, due to its significantly reduced number of parameters, has slightly lower iteration speed than the unimproved model. However, the final training accuracy of the model is essentially consistent with the unimproved version. Therefore, introducing residual networks in the model optimizes the training process and, after improvements, does not affect the model's training accuracy.

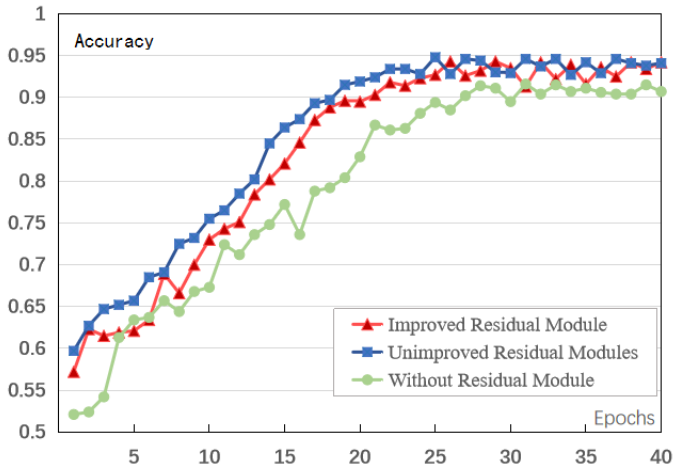


Fig. 9 - Iterative accuracy curve of sample training

Classification Results of Captured Image Samples

Among various semantic segmentation methods, this paper selected four representative networks for comparison: UNet with an encoder-decoder structure, RefineNet with residual optimization, PSPNet with pyramid pooling, and the prevailing DeepLabv3+. The model was trained using 1000 images, with 200 for validation and 500 for testing. From the classification results of the test images, it can be seen that the overall accuracy of all methods exceeds 90%, with this paper's method achieving the highest accuracy at 93%. The results clearly show that the three methods that utilize multi-scale convolution outperform UNet and RefineNet. This paper's method demonstrates greater accuracy in classifying C3 and C4, indicating that the model is more sensitive to pixels labeled as farmland. DeepLabv3+ performs well in all categories, particularly excelling in C2, suggesting that this model provides more accurate predictions for individual pixels.

Table 1

Results (%) of image classification

Number	Model	C1	C2	C3	C4	Accuracy
1	UNet	90.24	82.89	93.61	91.30	90.24
2	RefineNet	89.43	85.52	91.48	92.75	90.62
3	PSPNet	91.05	86.83	94.68	92.75	91.81
4	DeepLabv3+	93.49	92.10	92.55	92.27	92.63
5	Ours	92.68	89.47	94.68	93.71	93.00

The confusion matrix for this paper's method in Fig. 10 shows that C2 is easily misclassified as C3, indicating that some machinery pixels are incorrectly identified as farmland, resulting in the classification results of C3 exceeding the number of samples. However, there are no misclassifications between C3 and C4, primarily because most images that meet the requirements for C4 arise due to different seed machine models, causing the camera to capture both the sky and machinery within the normal monitoring angle. The proportions of both scenarios are less than the requirements for C1 and C2. Consequently, this paper's method not only achieves higher accuracy, while providing a more precise representation of the C4 category, which is the focus of this paper, effectively addressing the image classification issue.

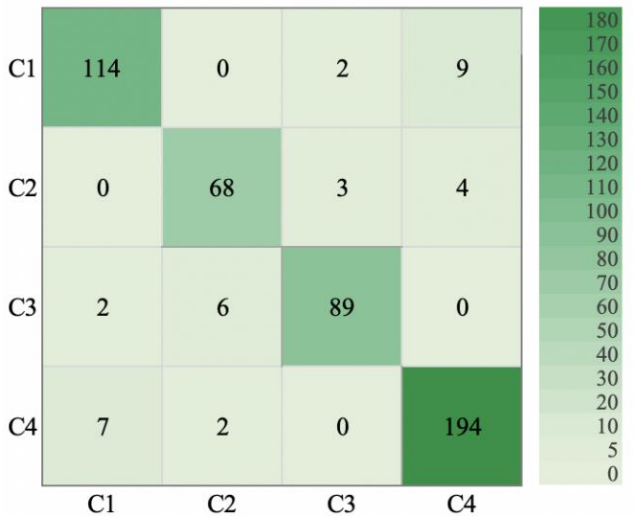


Fig. 10 - Confusion Matrix of our method

Background Segmentation and Areas to be Inspected Division Results

The background segmentation and areas to be inspected division results are presented in Table 2. Two sample images from classes C1 and C4, each containing three different types of plots, were selected for quantitative analysis of the model's performance. The C1 image was captured on a cloudy afternoon, while the C4 image was captured on a sunny morning. From the experimental results, it can be observed that the model proposed in this paper performed well in terms of total accuracy, intersection over union (IoU), and the number of parameters used. Particularly, compared to DeepLabv3+, one of the best-performing semantic segmentation models, although it did not have an advantage in single-pixel prediction accuracy, it achieved a roughly 20-fold reduction in model parameters, and achieved the best results in the segmentation of farmland areas to be inspected, with an accuracy of 85.37%. This is due to the residual modules maintaining high feature extraction capability even with reduced parameters, the entire encoder module exhibits higher encoding efficiency. Additionally, it indicates that the decoder, which combines multi-scale convolution and attention mechanisms, effectively highlights distance features. In the context of areas to be inspected division, it outperforms other semantic segmentation models significantly. Therefore, this method is the most suitable for the requirements of farmland areas to be inspected division.

Table 2

Results (%) of segmentation

Background Categories and Evaluation Methods	UNet	RefineNet	PSPNet	DeepLabv3+	Ours
Sky	92.24	94.54	94.46	96.24	94.26
Machine	80.26	82.64	85.41	86.29	86.37
Fram	77.50	77.12	74.95	88.10	86.34
Areas to be Inspected	63.24	69.34	72.24	73.31	85.37
Total Accuracy	76.04	86.61	85.26	96.02	94.54
IoU	73.42	80.45	79.79	83.25	84.96
Model Parameters(M)	7.76	14.35	37.30	41.00	2.10

Visualization of Farmland Background Segmentation Results

The visualization of background segmentation results for C1 samples is shown in Figure 10. Comparing the ground truth and predicted results for various methods in Fig. 11, it can be observed that most algorithms accurately capture the boundary between the sky and farmland in the scene images. However, UNet exhibits confusion in boundary distinction, as machinery with darker colors tends to be confused with the surrounding farmland, resulting in poorer segmentation results. RefineNet produces segmentation regions that are similar to the center region of the ground truth, but with overly smooth boundaries. PSPNet's segmentation is influenced by the category with more pixels, causing the center region to shift downward. However, it has a clear boundary. DeepLabv3+ demonstrates good segmentation results and is visually similar to this paper's method, but the boundaries are more distinct. This paper's method achieves more accurate segmentation of the central farmland region, benefiting from the inclusion of distance information in the network, which enhances the spatial characteristics of the farmland category. This ensures that the farmland area is not misclassified as other categories in the entire image, aligning better with the task requirements of this paper.

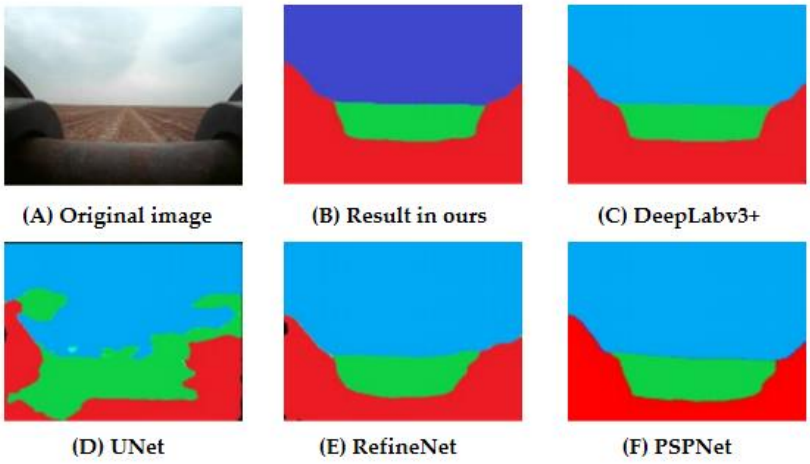


Fig. 11 - Farmland background segmentation image

Visualization of Areas to be Inspected Division Results

The visualization of areas to be inspected division results for C4 samples is shown in Fig. 12.

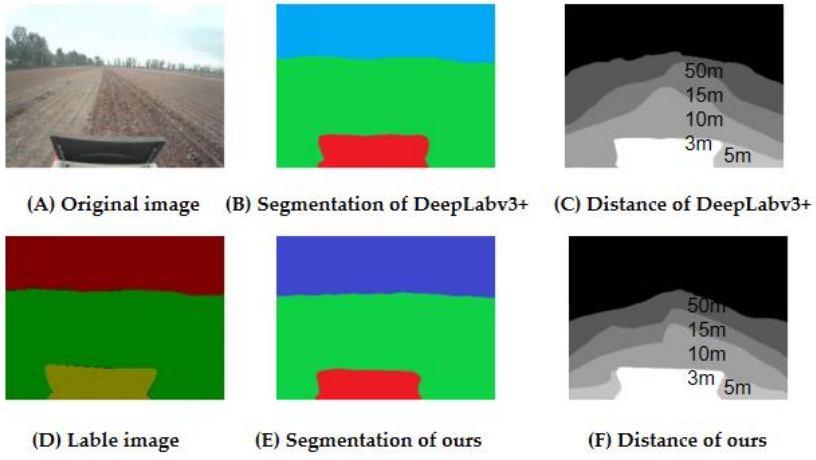


Fig. 12 - Effective region segmentation image

The results show that, in terms of background segmentation, both methods still exhibit no significant differences. This paper's method focuses more on the central location, while DeepLabv3+ shows more distinct boundaries. In the smoothed distance prediction map, DeepLabv3+ exhibits a more pronounced overflow phenomenon beyond the 10-meter level boundary. It also misclassifies a 5-meter area in the bottom left corner. The 50-meter boundary line aligns almost with the skyline, failing to separate farmland at a distance. This paper's method provides a clearer separation of areas for each level, without misclassification due to the lower proportion of samples. The position of the 50-meter level boundary is closer to the level map, allowing for better differentiation of distant information.

This paper uses the 15-meter level line as the boundary for areas to be inspected and obtains the division map of farmland areas to be inspected. From the map, it is evident that this paper's method produces a more accurate division of areas to be inspected. This is attributed to the decoding module, which includes both multi-scale convolution branches and a separate distance branch, allowing for better resolution of distance information during the division of areas to be inspected.

Experimental effect of straw cover detection

For the straw target detection task, the average deviation of various image regions after segmentation is generally more balanced and better than the results before segmentation. This is because after the division of the test area, there is no interference from the sky background and distant farmland in the sample image, making the test sample image closer to the standard image. However, in undivided image samples, the higher the camera angle, the greater the proportion of sky background in the image. In the white balance processing, the sky with a color closer to white replaces straw as the benchmark for color correction, rendering the color perception module for straw in the model ineffective. Therefore, without dividing the inspection area, when a large amount of straw is covered in the field, an angle that is too high will lead to a higher straw coverage rate, while when there is less straw covered, it will result in a lower straw coverage rate.

Table 3

Results of before and after effective region segmentation

Camera Angle	Before region segmentation		After region segmentation	
	Excessive straw	Fewer straw	Excessive straw	Fewer straw
Normal	0.974	1.062	0.974	1.062
Slightly high	1.095	0.892	0.965	1.054
Too high	1.134	0.851	0.963	1.087
Average Deviation	0.096		0.05	

CONCLUSIONS

(1) A novel approach for image classification and division of farmland areas to be inspected based on semantic segmentation is proposed. This method combines encoder-decoder structures, residual units, multi-scale convolution, and attention mechanisms. It classifies individual pixels in onboard captured images using semantic analysis, calculates the pixel proportions of various scene categories, and determines the image state. At the same time, it incorporates distance information into the segmentation of farmland areas to be inspected, making it possible to use a large number of images in the "to-be-processed" state as the calculation basis for quantitative analysis of straw returning.

(2) Experimental results indicate that, for onboard captured samples in various farmland scenes, the method proposed in this paper, based on semantic segmentation, achieves an overall sample classification accuracy of 93% for straw returning image classification and farmland areas to be inspected division, a background segmentation accuracy of 94.54%, and an areas to be inspected division accuracy of 85.37%. This method outperforms various mainstream semantic segmentation methods in straw monitoring applications, providing technical support for straw returning data processing and quantitative calculations.

(3) Compared with the results, it has been proven that the method proposed in this paper performs better in segmenting farmland in background categories and has more accurate spatial positions for vehicle captured samples in different agricultural scenes. This method integrates distance information to divide the image waiting area, which enables a large number of pending state images to serve as the calculation basis for quantitative analysis of straw returning to the field, effectively improving the application efficiency of straw monitoring images

ACKNOWLEDGEMENT

The work was sponsored by the National Key R&D Program Project of China (2022YFD2000300).

REFERENCES

[1] Badrinarayanan V, Kendall A, Cipolla R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, USA.

[2] Chen L C, Papandreou G, Kokkinos I. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 40, pp. 834-848, United States.

[3] Chung C, Chiu, MinY. (2010). A Robust Object Segmentation System Using a Probability-Based Background Extraction Algorithm. *IEEE Transactions on Circuits & Systems for Video Technology*, Vol. 20, pp. 518-528, United States.

[4] Fang P, Hao H, Li T. (2021). Instance Segmentation of Broiler Image Based on Attention Mechanism and Deformable Convolution (基于注意力机制和可变形卷积的肉鸡图像实例分割). *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 52, pp. 257-265, Beijing/China.

[5] Habas P A, Kim K, Corbett-Detig J M. (2010). A spatiotemporal atlas of MR intensity, tissue probability and shape of the fetal brain with application to segmentation. *Neuroimage*, Vol. 53, pp. 460-470, United States.

[6] Jadhav J K, Singh R P. (2018). Automatic semantic segmentation and classification of remote sensing data for agriculture. *Mathematical Models in Engineering*, Vol. 4, pp. 112-137, United States.

[7] Jiang H, Zhang C, Zhang Z. (2020). Detection Method of Corn Weed Based on Mask R-CNN (基于 Mask R-CNN 的玉米田间杂草检测方法). *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 51, pp. 227-235+254, Beijing/China.

[8] Liang X, Li Y, Zhou Y. (2020). Study on the abandonment of sloping farmland in Fengjie County, *Three Gorges Reservoir Area, a mountainous area in China. Land Use Policy*, Vol. 97, pp. 104760, United Kingdom.

[9] Lin G, Milan A, Shen C. (2017). RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE*, pp. 5168-5177, United States.

[10] Litjens G, Kooi T, Bejnordi B E. (2017). A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, Vol. 42, pp. 60-88, Netherlands.

[11] Liu B, Cheng H D, Huang J. (2010). *Probability density difference-based active contour for ultrasound image segmentation*. *Pattern Recognition*, Vol. 43, pp. 2028-2042, United Kingdom.

[12] Liu K, Xie B, Zhai Z. (2020). Target Muscle region Recognition in Ovine Hind Leg Segmentation Based on R2U-Net and Atrous Convolution Algorithm (基于 R2U-Net 和空洞卷积的羊后腿分割目标肌肉区识别). *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 51, pp. 507-514, Beijing/China.

[13] Liu W, Yue A, Ji J. (2020). Urban green space extraction from GF-2 remote sensing image based on DeepLabv3 + semantic segmentation model (基于 DeepLabv3+语义分割模型的 GF-2 影像城市绿地提取). *Remote Sensing for Land and Resources*, Vol. 32, pp. 120-129, Beijing/China.

[14] Long J, Shelhamer E, Darrell T. (2015). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, pp. 640-65, United States.

[15] Miao R, Tang J, Chen. (2015). Classification of farmland images based on color features. *Journal of visual communication & image representation*, pp. 138-146, United States.

[16] Minaee S, Boykov Y, Porikli F. (2020). Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Software Engineering*, United States.

[17] Qi C R, Su H, Mo K. (2017). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 77-85, United States.

[18] Ronneberger O, Fischer P, Brox T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing*, Vol. 9351, pp. 234-241, Germany.

[19] Sreepada R, Patra B. (2020). Mitigating long tail effect in recommendations using few shot learning technique. *Expert Systems with Application*, Vol. 140, pp. 112887.1-112887.17, United States.

[20] Wu J, Ye C, Sheng V S. (2017). Active learning with label correlation exploration for multi-label image classification. *IET Computer Vision*, Vol. 11, pp. 577-584, United Kingdom.

[21] Xiao D, Jiang, Qi G. (2015). Extraction of farmland classification based on multi-temporal remote sensing data. *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 31, pp. 145-150, Beijing/China.

[22] Yu H, Liang X, Zhang Y. (2015). Effects of Different Straw Returning Modes on the Soil Microorganism and Enzyme Activity in Corn Field. *Journal of Agricultural Resources and Environment*, Vol. 32, pp. 305-311, Tianjin/China.

[23] Zhao S, Wang S, Bai Y. (2021). Real-time Semantic Segmentation of Sheep Skeleton Image Based on Generative Adversarial Network and ICNet (基于生成对抗网络与 ICNet 的羊骨架图像实时语义分割). *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 2, pp. 329-339+380, Beijing/China.

[24] Zhou H, Zhang J, Lei J. (2016). Image semantic segmentation based on FCN-CRF model. *2016 International Conference on Image, Vision and Computing (ICIVC)*, Vol. 8, pp. 9-14, United Kingdom.