ARTICLE IN PRESS

Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE xxx (xxxx) xxx

journal homepage: www.keaipublishing.com/en/journals/information-processing-in-agriculture/

Semantic segmentation of agricultural images: A survey

Zifei Luo [a,b], *Wenzhu Yang* [a,b,*], *Yunfeng Yuan* [a,b], *Ruru Gou* [a,b], *Xiaonan Li* [c]

[a] *School of Cyber Security and Computer, Hebei University, Hebei 071002, China*
[b] *Hebei Machine Vision Engineering Research Center, Baoding, Hebei 071002, China*
[c] *College of Education, Hebei University, Hebei 071002, China*

A R T I C L E I N F O

Article history:
Received 23 November 2022
Received in revised form
5 February 2023
Accepted 7 February 2023
Available online xxxx

Keywords:
Semantic segmentation
Agricultural images
Deep learning
Convolution neural networks

A B S T R A C T

As an important research topic in recent years, semantic segmentation has been widely applied to image understanding problems in various fields. With the successful application of deep learning methods in machine vision, the superior performance has been transferred to agricultural image processing by combining them with traditional methods. Semantic segmentation methods have revolutionized the development of agricultural automation and are commonly used for crop cover and type analysis, pest and disease identification, etc. We first give a review of the recent advances in traditional and deep learning methods for semantic segmentation of agricultural images according to different segmentation principles. Then we introduce the traditional methods that can effectively utilize the original image information and the powerful performance of deep learning-based methods. Finally, we outline their applications in agricultural image segmentation. In our literature, we identify the challenges in agricultural image segmentation and summarize the innovative developments that address these challenges. The robustness of the existing segmentation methods for processing complex images still needs to be improved urgently, and their generalization abilities are also insufficient. In particular, the limited number of labeled samples is a roadblock to new developed deep learning methods for their training and evaluation. To this, segmentation methods that augment the dataset or incorporate multimodal information enable deep learning methods to further improve the segmentation capabilities. This review provides a reference for the application of image semantic segmentation in the field of agricultural informatization.

Contents

1. Introduction 00
2. Traditional semantic segmentation methods 00

* Corresponding author at: School of Cyber Security and Computer, Hebei University, Hebei 071002, China.
E-mail address: wenzhuyang@hbu.edu.cn (W. Yang).

Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2023.02.001

2.1. Semantic segmentation based on thresholding . . . 00
2.2. Semantic segmentation methods based on clustering . . . 00
3. Deep learning-based methods . . . 00
3.1. U-Net . . . 00
3.2. SegNet . . . 00
3.3. DeepLab . . . 00
3.4. Others . . . 00
4. Challenges and strategies . . . 00
4.1. Data augmentation . . . 00
4.2. Pixel-level accuracy realized by multiscale strategy . . . 00
4.3. Information complementarity realized by multimodal fusion . . . 00
5. Conclusion . . . 00
Declaration of Competing Interest . . . 00
References . . . 00

1. Introduction

Agricultural automation has grown rapidly in the past few years. Machine vision-based approaches have driven the progress of the agricultural industry. Vision-based technologies are used in scenarios such as pest identification [1], behavioral traits of livestock [2], etc., which can help reduce labor costs and time consumption. Semantic segmentation is one of the basic tasks in machine vision to achieve pixel-level classification. It is an important component of computer vision-based applications. Unlike making predictions about an image, semantic segmentation generates pixel-level descriptions of objects embedded in their spatial information. With the advancement of semantic segmentation methods, they have been used to address diversity and data-rich agricultural problems, such as crop cover and type analysis [3-5], forest tree species labeling [6], weed segmentation [7], predictive agriculture [8], pest and disease identification [9], etc.

Semantic segmentation techniques have replaced the traditional manual observation and measurement of phenotypic data, have played an extremely important role in modern agriculture. For example, in greenhouse farming [10], semantic segmentation techniques are used to monitor the growth status of crops, predict leaf canopy area and vegetation height, extract pixel points from fruit surfaces to determine fruit ripeness. Machine vision technology monitors the effect sowing process and improves the quality of crop production by harvesting the crop at the right time to achieve the optimum maturity level. In pest segmentation [11], it's feasible to achieve pest identification by extracting the texture, shape and size of insects in the image. This helps with pesticide control. In addition, there are many difficulties in the acquisition of some macroscopic information and the regionalization of parameters due to the problem of non-uniformity of the ground and near-ground environment with the increased spatial scale. Farm segmentation is essential for automatic navigation [12].

However, feature extraction and image segmentation are difficult jobs due to the complex environment, season and illumination changes, etc. Before deep learning began to be studied, the robustness of traditional segmentation methods to complex environments is greatly enhanced by color space conversion and combination of color channels. For extreme lighting conditions as well as sharp shadow edges, the vegetation index-based methods utilize separate color channels for pixel color and brightness, respectively [13]. Other methods use machine-learning based classification techniques such as decision trees [14-16]. In contrast, deep learning methods extract features autonomously according to different classification tasks, which allows them to handle complex and diverse application scenarios. The great success in other sectors has motivated the combination of traditional segmentation methods with deep learning methods to solve problems in agricultural sector. New developed deep learning models, including VGG, FCN, U-Net, SegNet, DeepLab, etc., are often used for implementing pixel-level segmentation nowadays.

To solve the segmentation challenges due to various growth stages and overlapping plant objects, an Encoder-Decoder deep network, which fed with 14 various vegetation indices as input for weed/crop/background segmentation, gets the best MIoU (Mean Intersection of Union) value of 80.8% [17]. For segmentation challenges with severe occlusions, the U-Net network is able to splice low-level features and high-level features preserving edge information while having a small computational burden. in the high-resolution image segmentation of fruit tree branches, the semantic segmentation model using U-Net [18] with a modified cross-entropy loss function [19] performs well. For complex image segmentation, the DeepLabV3+ model utilizes dilated convolution with increased receptive field. As counted by pigs, the MIoU value of the improved DeepLabV3+ is 74.62%, which is higher than other models [20]. In addition to CNN architectures, Generative Adversarial Networks (GANs) also perform well in semantic segmentation tasks. The discriminators in GANs can help learn relationships between pixels (often ignored in CNN architectures), which can improve the performance and accuracy of the networks [21]. The backbone networks that fit the characteristics of images can cope with type-complex agricultural datasets. We classify some agricultural image datasets by different image scales. Among them, the near-ground image dataset contains: pests and diseases [22,23], species understanding [24], plants phenotyping [25-27], etc. UAV (Unmanned Aerial Vehicle) datasets are often used for agricultural yield estimation [28], and remote sensing image datasets can be used for forest monitoring [29], etc.

Recently, some surveys on semantic image segmentation are available [30,31]. But up to now, there has not been a systematic summary of semantic segmentation of agricultural images. We review the recent advances in traditional semantic segmentation methods and deep learning methods used for agricultural images. In our literature, we also identify the challenges in agricultural image segmentation and the directions that can be investigated in the future.

2. Traditional semantic segmentation methods

Before the rise of neural network models, there are many traditional methods designed for solving the problems of agricultural image classification and semantic segmentation. Representative traditional algorithms include threshold-based methods, clustering-based methods, wavelet transforms, and random forest [32-35]. In this section, we introduce the main ideas and applications of some typical traditional semantic segmentation methods for agricultural images, such as thresholding-based segmentation, region-based segmentation, etc. The traditional semantic segmentation process that has been widely used in agricultural automation in recent years is shown in Fig. 1.

2.1. *Semantic segmentation based on thresholding*

Thresholding is one of the classical methods of image segmentation, which uses the difference in grayscale between the segmentation target and the background to set different gray thresholds to classify the image gray histogram, and considers pixels with gray values in the same gray range to be of the same class and have certain similarity by judging the characteristic attributes of each pixel point in the image.

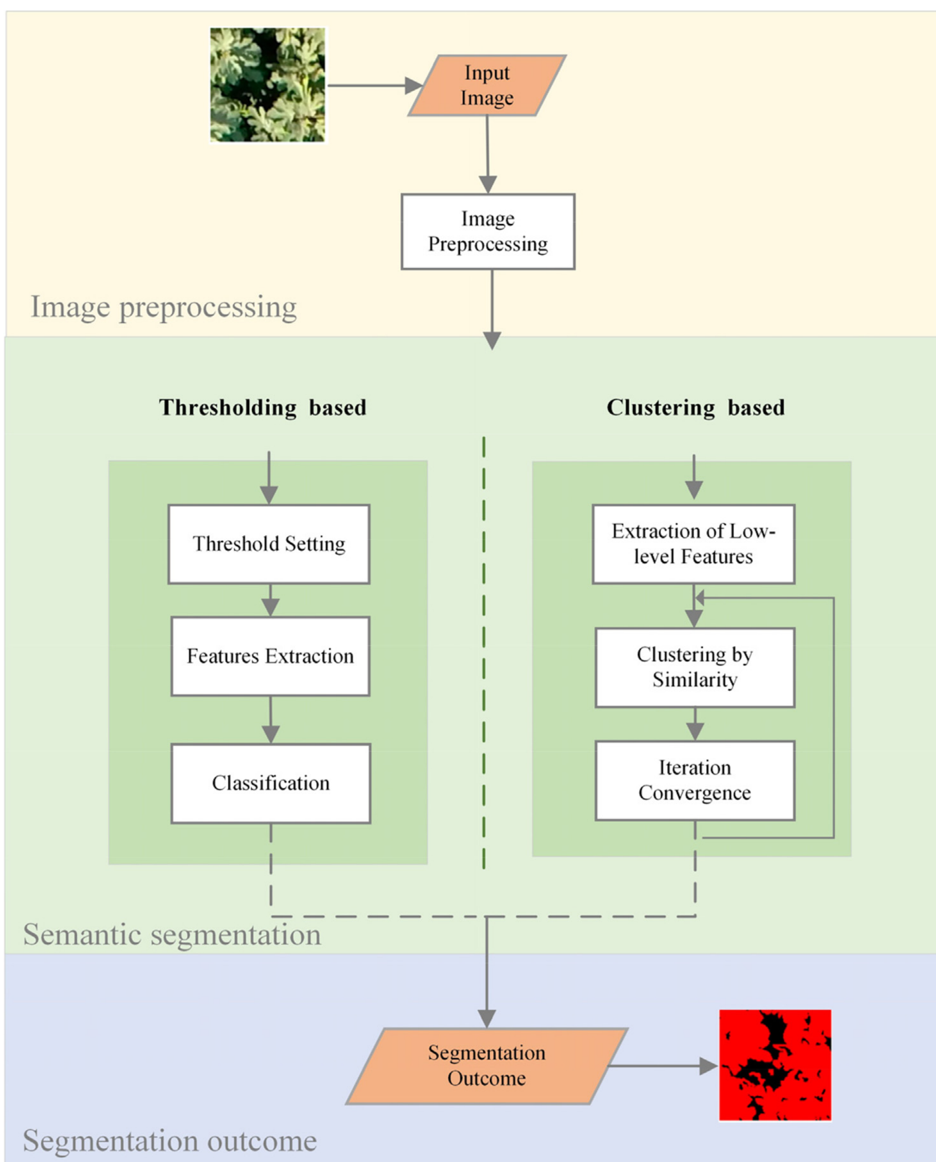

Fig. 1 – Semantic segmentation process based on traditional methods.

The key of threshold segmentation method is to find the appropriate threshold to segment the image accurately. In order to obtain the best threshold, a global threshold method Otsu is proposed [36]. Otsu method is a master work of image segmentation using threshold value. The standard used in Otsu algorithm to measure the difference is the most common inter-class variance. When the threshold segmentation makes the inter-class variance maximum, it means that the misclassification probability is minimum. However, when the illumination is not uniform, it is difficult to find an appropriate single threshold value to segment the image accurately. The proposed adaptive threshold method solves this problem to a certain degree [37]. It calculates the mean value, median value, and Gaussian-weighted average value of a neighborhood to determine the local threshold according to the brightness distribution in different regions of the image.

In vegetation image segmentation, various threshold-based methods follow the setting of equality between vegetation and other objects in the image which is not entirely valid at different stages of plant growth [38]. On the other hand, color-indexed techniques that convert RGB values to grayscale do not always produce good discriminating grayscale images. This is caused by the influence of lighting conditions on the values of RGB images. Thus, although the color index and threshold index-based methods provide simplicity and advanced technology, they have certain limitations due to the effect of light variation conditions on the achieved vegetation segmentation results, especially in sunny and cloudy conditions. Over the past decade, many works have been done in agriculture yield estimation using machine learning techniques. For segmentation of citrus fruit images, the watershed algorithm was used after converting the original RGB images to HSV ones and obtained a $R2$ value of 0.93 [39]. In order to detect the appropriate threshold value, the HSV color space is used and then the Hue channel is extracted to generate the tonal histogram. Appropriate thresholds are detected using filtered hue histograms and fitted Gaussian curves of hue histograms [13]. The method can generate accurate and stable vegetation segmentation performance with an average accuracy of 87.29% and a standard deviation of 12.5% [38]. We summarized the vegetation indices that had outstanding effects in previous studies are shown in Table 1.

Table 1 – Commonly used vegetation indices [40].

Index	Definition
ExG	2×G–R–B
ExR	1.3×R–G
ExGR	ExG-ExR
MExG	1.262×G–0.884×R–0.311×B
GB	G-B
NDI	128×(((G–R)/(G+R))+1)
VEG	$G/(R^{a}\times B^{(1-a)})$, a=0.667
CIVE	0.441×R–0.811×G+0.385×B+18.78745
COM	ExG+CIVE+ExGR+VEG
NCIVE	(CIVE+188.0176)/417.4351
MNGRDI	((G–R)/(G+R)+1)/2

When segmenting foreground and background, the index-based method is used for approximate pre-segmentation to calculate the segmentation threshold. When it combined with Otsu, the semantic segmentation accuracy of plant image achieves to 97.4%. When the NCIVE image index is computed, its robustness to segmentation in a real-world environment has been significantly improved [40].

In addition, the advantage of the index-based methods is that they can achieve comparable segmentation performance but lower computational cost to some complex algorithms. So, they are widely used for vegetation segmentation tasks.

2.2. *Semantic segmentation methods based on clustering*

The semantic segmentation method based on clustering is to gather the pixel points with similar features into the same area, and then the segmentation result is obtained after repeated iteration and clustering until convergence. From the perspective of computer vision, superpixels belong to over segmentation, which are pixel blocks composed of neighboring pixels with similar features (color, texture, etc.). The similarity between pixel features is used to group pixels, and a small number of superpixels are used to represent image features instead of a large number of pixels so that edge information is preserved. This can greatly reduce the complexity of image post-processing and increase the computing speed. Superpixel is often used as preprocessing steps for image segmentation algorithms.

This technology has been widely used in computer vision application including posture estimation, target recognition and image segmentation, etc.

Simple linear iterative clustering (SLIC) algorithm is an efficient superpixel generation algorithm derived from the extension of K-means clustering algorithm. In recent years, more superpixel algorithms based on SLIC have been investigated and these algorithms improve the boundary adherence capability [41,42].

Moreover, an adaptive SLIC (ASLIC) [41] method uses the maximum distance of each cluster to dynamically normalize the color and spatial proximity, which performs well in terms of speed, but performs poorly on weak boundaries.

In order to better adapt to the weak boundary, the SLIC is processed by multi-level hierarchical processing, and the original image is segmented by simple linear iterative clustering method based on local information (LI-SLIC) [43]. Neighboring superpixels belonging to the same object are merged according to the similarity of the probability distribution, and the number of superpixels is reduced. LI-SLIC gets improved robustness but slower speed compared with some other clustering methods.

When the SLIC method is used in practical agricultural application scenarios, shadows and uneven illumination of images obtained in natural environments often pose significant segmentation difficulties. For example, when segmenting the plant disease images, the soil and other objects in the image will affect the segmentation process [44]. The color of the soil is almost the same as that of the plant infected area with early/late blight while the texture is different. To solve this problem, color equalization and superpixel operation are performed on the original images to eliminate the effect of uneven lighting. It achieved 97.2% accuracy on the plant village dataset [45].

Aiming to eliminate the impact of various illumination, SLIC combined with a graph-based segmentation algorithm, was used to generate region proposals, and colour model with ensemble neural networks specific to each intensity was proposed. It achieved an average accuracy of 74.3% [46]. Superpixel technology can not only be used as an optimization method in image preprocessing stage, but also be combined with deep learning methods for image segmentation.

In recent years, deep learning methods incorporating with superpixels have been extensively studied, such as Segmentation Aware Loss (SEAL) [47], Superpixel Sampling Networks (SSN) [48], and Superpixel Segmentation with Fully Convolutional Networks (FCN) [49]. When segmenting carcasses using the SLIC algorithm in combination with deep learning techniques, the superpixels are formed first for the carcass images using the SLIC algorithm. Then these superpixels are manually labeled for different classes of objects (carcasses, background), as shown in Fig. 2.

This Superpixel + CNN architecture obtains encouraging segmentation results in carcass image dataset. However, the choice of the number of superpixels k has a great impact on the segmentation results. For instance, when $k = 100$ and 1000, the pixel accuracy reached 84.3% and 96.1%, and the IoU were 77.3% and 92.2%, respectively [50].

3. Deep learning-based methods

Due to lack of generalization ability, the traditional segmentation methods are difficult to fit the model in practical applications. Deep learning methods are widely used in the field of agricultural vision due to their superior performance. Their applications include classification, detection [51,52] and semantic segmentation [53]. We summarize the semantic segmentation process based on deep learning into three stages: feature extraction, semantic segmentation, and post-processing, as shown in Fig. 3.

There have been many deep learning-based architectures designed for semantic segmentation. In this section, we give a brief introduction of the classical deep learning-based methods designed for semantic segmentation of agriculture images by presenting their models, principles, and applications. These methods described in this section not only can be applicable to solve the semantic segmentation problems for agricultural images, but also can be applied to other application scenarios.

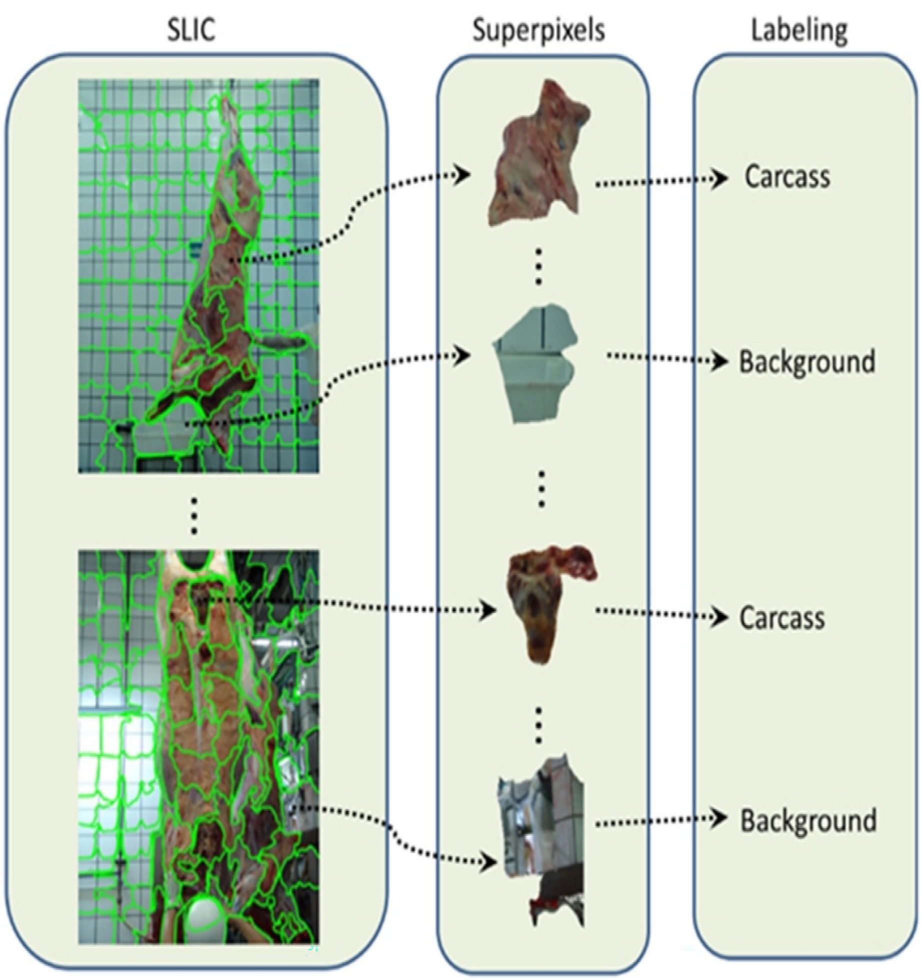

Fig. 2 – Creation and labeling of superpixel [50].

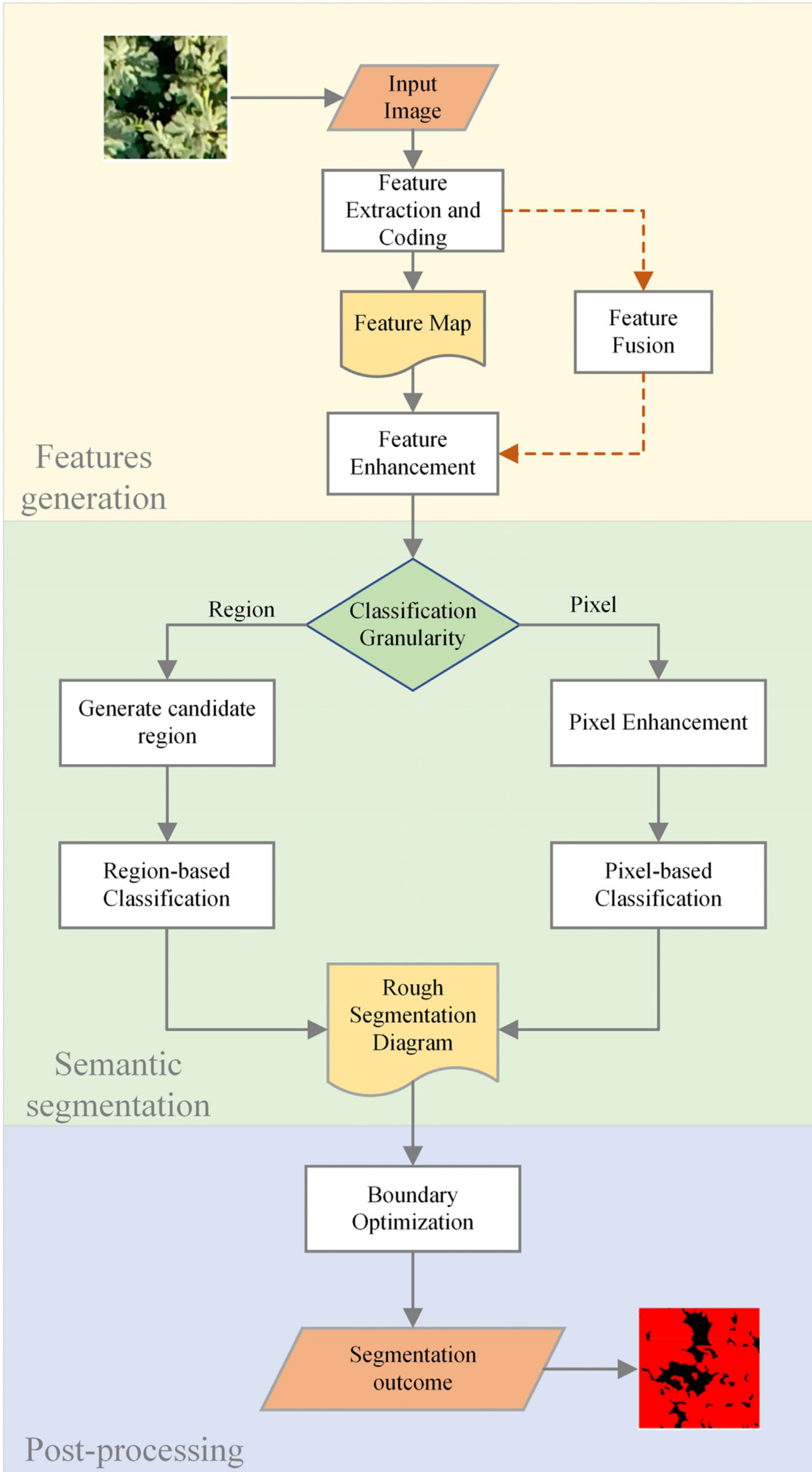

Fig. 3 – Semantic segmentation process based on deep learning.

As one of the first high impact CNN-based segmentation, FCN (Fully Convolution Network) [54] provides key ideas for the development of Encoder-Decoder networks. The important idea of FCN is to input an image of arbitrary size and get the output of the same size by efficient learning and inference of the model. FCN replaces the fully-connected layer with a convolutional layer. To keep the same size with the original image for the segmentation process, deconvolution layer is used to bilinearly up-sample the feature map. A prediction can be made for each pixel and the spatial information in the original input image can be retained. In this section, we review the new models derived from Encoder-Decoder networks and their applications in agricultural image segmentation.

3.1. U-Net

The U-Net architecture [18] structure based on FCN consists of three components: (1) down-sampling stage that abstracts features level by level; (2) up-sampling stage that reconstructs the features; (3) a final convolutional layer to achieve classification. The architecture is illustrated in Fig. 4.

Because of its streamlined structure, U-Net can be well applied to agricultural images with few samples. In addition, U-Net has strong local information extraction ability. It can well combine the features of global and local details of images, and fuse low-resolution information with regularity and high-resolution information with blurred boundaries and complex gradients, which can achieve accurate segmentation of some agricultural images. Table 2 summarizes the segmentation scenarios using U-Net as a backbone network, along with their improvements and experimental results.

But usually, the sizes of pixel blocks are much smaller than the whole image. The algorithm can only extract some local feature information, which leads to the limitation of the classification performance. Secondly, the computer needs to do a lot of repetitive calculations due to the repetition of adjacent pixel blocks. In recent years, U-Net has been successfully applied in tasks such as plant phenotype observation and pest and disease detection.

3.2. SegNet

SegNet [55] is a deep network for image semantic segmentation proposed by Cambridge to solve the problem of automatic driving or intelligent robot. Based on Caffe, SegNet is one of the classic models of Encoder-Decoder network. It follows the FCN architecture and the semantic segmentation network is obtained by VGG16. The architecture of SegNet is shown in Fig. 5.

There are two versions of SegNet, the SegNet and the Bayesian SegNet [57]. The authors of the SegNet also provide a shallow BASIC version according to the depth of the network. From Fig. 5 we can see that the structure of SegNet is quite similar to that of U-Net. The difference is that SegNet uses Pooling indices to restore the original positions of the feature

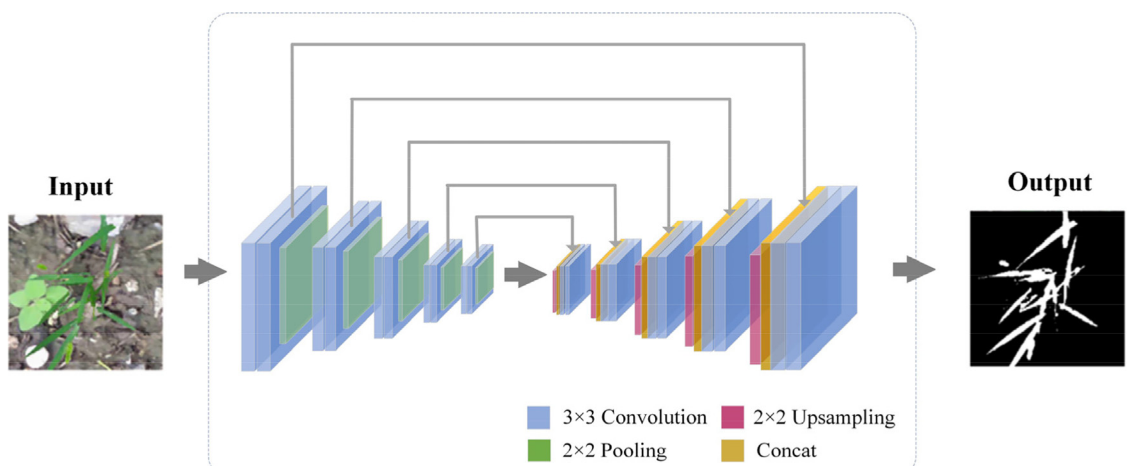

Fig. 4 – U-Net and its application in weed image segmentation [7].

Table 2 – Segmentation scenarios with their improvements and experimental results using U-Net as backbone network.

Application scenario	Methodology	Result
Weed segmentation [7]	Reduce Encoder-Decoder layers	89.45% (IoU)
Forest cover segmentation [29]	Compared with state-of-the-art CNN architectures such as DeepLabV3+, PSPNet, etc. in terms of standard performance metrics	95.13% (Dice)
Forest cover segmentation [84]	Add Airborne Laser Scanning (ALS) technology	62.00% (IoU)
Wheat farm segmentation [12]	A multi-scale global attention module (MGA) is used to aggregate multi-scale global environmental information and form enhanced features using improved attention mechanisms	93.88% (Dice)

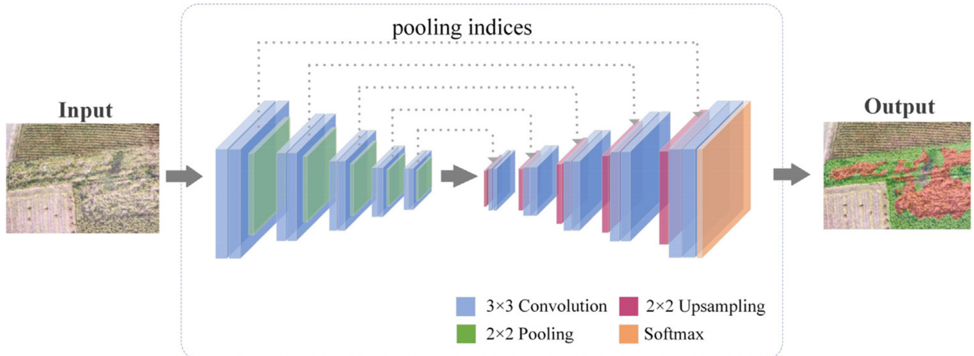

Fig. 5 – SegNet and its application in sunflower UAV image segmentation [56].

Table 3 – Segmentation scenarios with their improvements and experimental results using SegNet as backbone network.

Application scenario	Methodology	Result
Sunflower planting areas segmentation [56]	Conditional Random Field (CRF) and image fusion were used to improve the model performance	89.8% (IoU)
Landsat-8 satellite image segmentation [58]	Fine-tuning the forest images to improve their generalization ability by an effective transfer learning procedure on an initialized CNN pre-training model	82.1% (IoU)

maps. Table 3 summarizes the segmentation scenarios using SegNet as a backbone network, along with their improvements and experimental results.

Although SegNet is fast in convergence, it does not fully consider the pixel-to-pixel relationship. In the field of agriculture, SegNet is more advantageous in large volume target extraction for high resolution remote sensing images, such as cultivated land images. SegNet is also successfully applied in the field of orchard 3D visualization and fruit picking vision system, point cloud.

3.3. *DeepLab*

DeepLab [59-62] improves FCN by employing atrous convolution. The atrous convolution increases the receptive field by adding holes, so that the convolutional output contains a larger range of information and preserves the spatial features of the image.

DeepLabV1 [59] and DeepLabV2 [60] are end-to-end frameworks with excellent performance. They adopt VGG (DeepLabV1) and ResNet (DeepLabV2) as encoders, while using atrous convolution and fully connected Conditional Random Fields (CRF) to improve the ability of the model to capture details.

The original DeepLabV3 [61] still takes ResNet [63] as its backbone. However, the last block of ResNet is modified with the addition of atrous convolution. It uses ASPP (Atrous Spatial Pyramid Pooling) instead of CRF to capture contextual information at different scales. ASPP enhances the ability of extracting dense features. However, the boundary information of the segmented target is seriously lost due to the existence of pooling and convolution with stride. Comparatively, DeepLabV3+ applies depth separable convolution extracted from the Xception model to convolve the space of each channel separately and achieves higher accuracy. The architecture of DeepLabV3+ [62] is shown in Fig. 6.

Table 4 summarizes the segmentation scenarios using DeepLabV3+ as a backbone network, along with their improvements and experimental results.

DeepLabV3 + network is excellent at extracting more dense features. At the same time, DeepLabV3+ largely expands the receptive field and obtains a multi-scale global background, but is still limited by the local area. Secondly, the huge number of parameters bring tremendous computational burden. Now DeepLabV3+ network is mostly used to achieve high-precision segmentation of high-altitude images [66–68]. Combining machine learning technology and remote sensing data, especially high spatial and temporal resolution, hyperspectral remote sensing images, DeepLabV3+ network can identify land cover and crop types at different scales, and then combine with meteorological data to produce agro-climatic zoning and agro-meteorological disaster risk zoning.

The aforementioned CNN network architectures perform well in the semantic segmentation tasks of agricultural images. Because of their excellent performance, CNNs have been studied extensively, such as Densely Connected Convolutional Network (DenseNet) [69] and ShuffleNet [70], which have also been used for agricultural image segmentation.

3.4. *Others*

VGG network, AlexNet, GoogLeNet, and some other deep learning models [63,71-75] are often used as the backbones of Encoder-Decoder networks for agricultural image semantic segmentation due to their excellent performance.

VGG network and Inception are often modularized and applied to do the down-sampling task in Encoder-Decoder

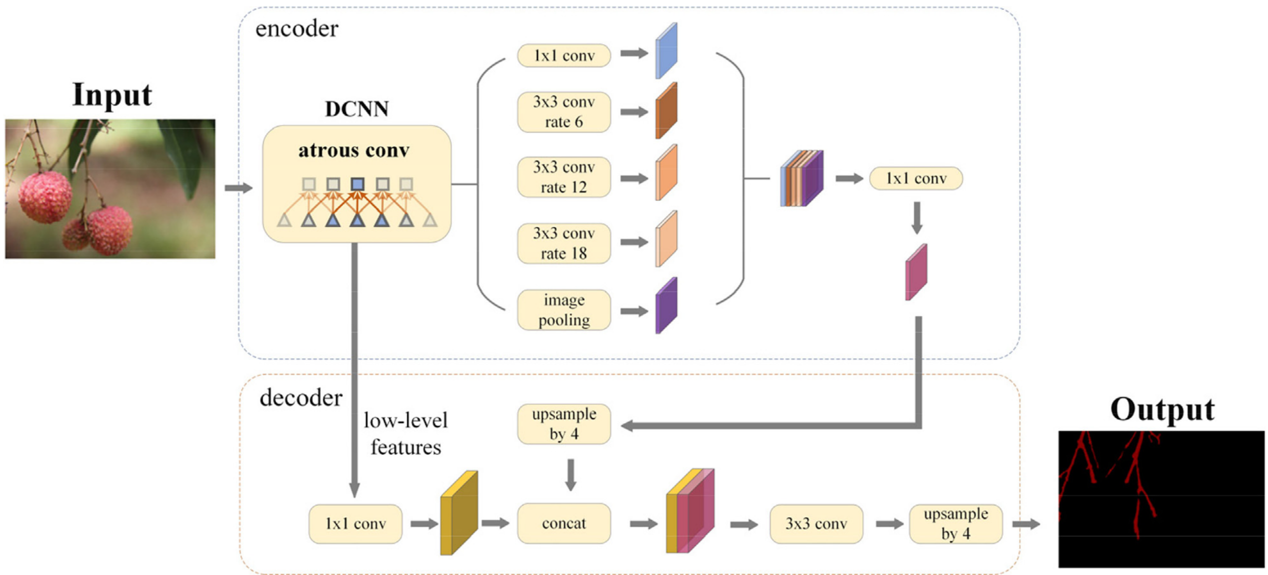

Fig. 6 – DeepLabV3+ and its application in litchi branch image segmentation [64].

Table 4 – Segmentation scenarios with their improvements and experimental results using DeepLabV3+ as backbone network.

Application scenario	Methodology	Result
Litchi branch [64]	Xception feature extraction model used by DeepLabV3+ was improved at different layers	76% (IoU)
Swine herds segmentation [29]	DeepLabV3+ deep learning method introduces lightweight row and column-based attention modules that can speed up the efficiency of feature computation	74.62% (IoU)
Grape leaf black rot spots segmentation [65]	A feature fusion branch based on a feature pyramid network is added to the DeepLabV3+ encoder	84.8% (IoU)

network, and used for replacing the original encoder or forming different encoding stages with the original encoder. In the up-sampling stage, the use of deconvolution allows for a clever combination of the Inception module and up-sampling, as seen in Fig. 7. The added Inception modules perform superposition and decomposition on the up-sampled convolutional layers to capture multiscale information.

4. Challenges and strategies

For the excellent performance of CNNs in vision tasks in other fields, researchers are committed to the successful application of deep learning methods to the semantic segmentation of agricultural images. However, there are some peculiarities in agricultural images: difficulties in image acquisition, hierarchical relationships between different classes, complex shooting environment for agricultural images, etc. There are still rooms for improvement when applying deep networks to semantic segmentation of agricultural images.

4.1. *Data augmentation*

One of the shortcomings of using supervised learning methods in agricultural image segmentation tasks is the lack of sufficient labeled datasets, which often affects the training process of the network [76]. The robustness and generalization of the deep learning models mentioned in Section 3 are affected by the diversity of the training data and the amount of data, so data augmentation is crucial when using deep learning methods for small sample problems. For example, in the segmentation of plant diseases and insect pests, the onset conditions of cases are different, and some lesion image samples are scarce [11]. It is difficult to have sufficient data to support training in actual projects. In the small sample problem, the traditional methods usually carry out geometric transformation or color transformation on the existing data, however, they do not substantially augment the data set. There are other ideas to solve this problem. One of these methods is utilizing multiple channels of an image [53]. The multiple channels (original RGB data, vegetation index, HSV color channel and edge detector as shown in Fig. 8) divided from the images are used as the input data of CNN, which enhanced model generalization with restricted training data.

Image synthesis can also be used to augment input data. For weed segmentation [25], weed images were used as the "foreground" and other images were used as the "background". These synthesized images with "foreground" and

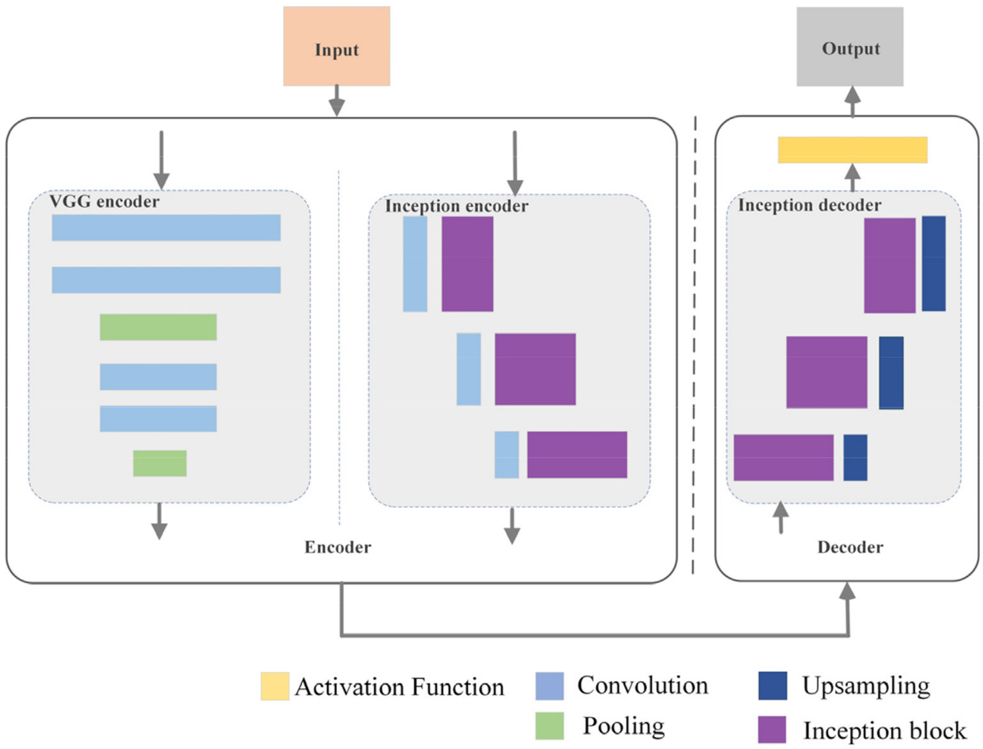

Fig. 7 – Semantic segmentation model using VGGNet, GoogLeNet as backbones.

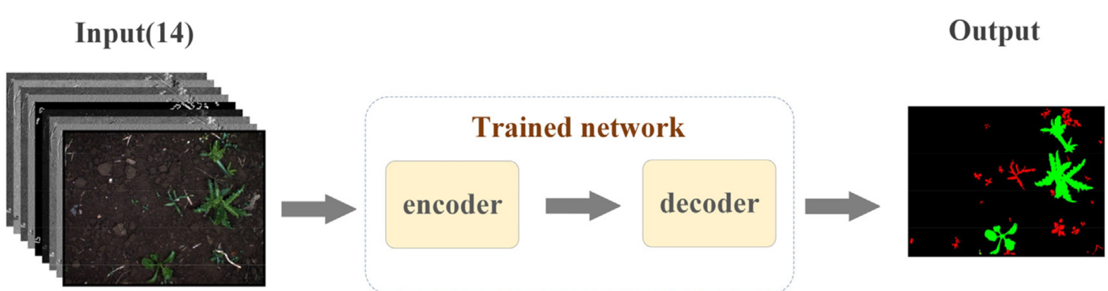

Fig. 8 – The segmentation result of Encoder-Decoder network using multiple channels of an image [53] as input data.

"background" are used for pre-training to augment dataset and reduce the number of manually labeled samples as shown in Fig. 9.

The problem of small samples can also be solved using the GAN-based image generation method [21,77]. Samples consistent with the real data distribution are generated through image generation to expand the existing data set, and the mixed data set can enhance the learning ability of a model. Compared with existing GANs, its variation models, such as, DCGAN (Deep Convolutional Generative Adversarial Networks) [78], C-GAN (Conditional Generative Adversarial Network) [79], AR-GAN (Activation Reconstruction Generative Adversarial Network) [80] have better performance. When C-GAN is used to enhance the leaf image data of tomatoes [79], overfitting can be effectively suppressed as shown in Fig. 10. Besides, to synthesize more realistic tomato leaf images, AR-GAN [80] retains the semantic information from the mask.

The above image augmentation strategies can also be used for other small sample application scenarios to achieve better segmentation results.

4.2. *Pixel-level accuracy realized by multiscale strategy*

UAV images and remote sensing images are widely used in agriculture. They have common characteristics: high resolution. The high performance of convolutional neural network largely depends on fine-grained spatial details and sufficient context information, both of which trigger high computational cost. This severely limits their real-time processing capabilities. In some scenes of agricultural remote sensing image segmentation, like cultivated land segmentation [81], aquatic products monitoring [82], etc., due to the characteristics of high-resolution images, such as abundant details of ground objects and great differences in categories, the existing convolution neural network image segmentation methods

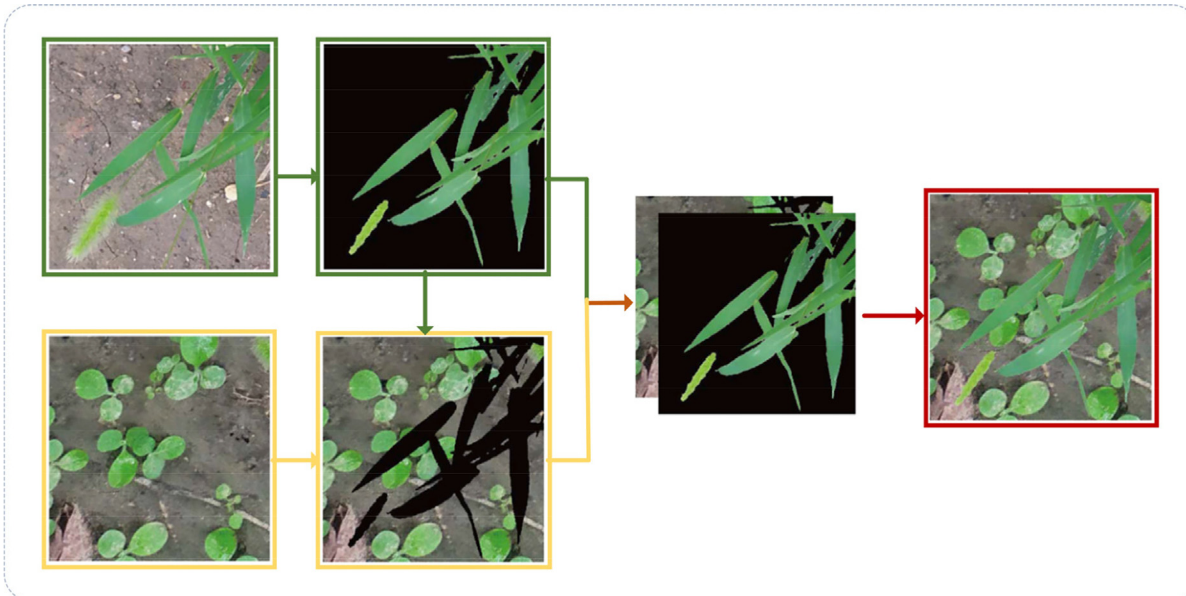

Fig. 9 – Synthetic process of weed dataset [25].

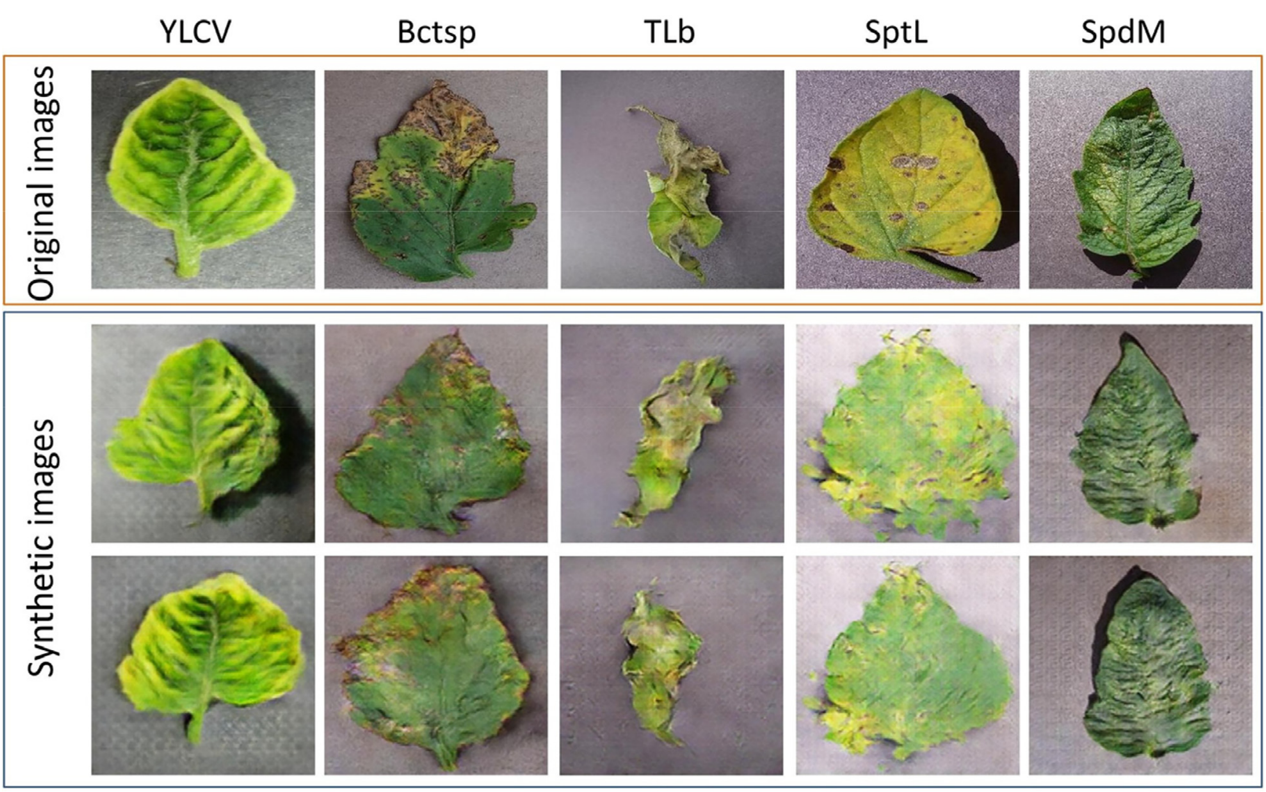

Fig. 10 – Original tomato leaf images (Yellow Leaf Curl Virus, Bacterial Spot, Late Blight, Septoria leaf spot and Two Spotted Spider Mite) and their synthetic images generated by C-GAN [79]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

generally have the problems of low segmentation accuracy and inaccurate boundary of ground objects.

Multiscale convolutional neural network has been used for solving these problems. In addition to the ASPP (Atrous Spatial Pyramid Pooling) module used in DeepLab network, DDCN (Dynamic Dilated Convolutional Network) [83] can be used for multi-scale strategy. The key idea of DDCN is still to maintain the resolution of the image. Specifically, the model does not down-sample the dynamically generated input data (original image and probability distribution of possible input sizes) during the training process. The abundant texture features brought by high spatial resolution trigger huge differences in image features of the same crop. The networks introduced in Section 3.1 have multi-scale advantages in themselves, and the insertion of modules such as ASPP, DDCN will further enhance the networks. The methods of DeepLabV3+ and DDCN incorporating multiscale strategies obtained 95.15%, 95.46% accuracy in semantic segmentation of UAV images of citrus orchards, which are higher than 94.96%, 94.96% of U-Net, SegNet [84]. Meanwhile, different expansion rates in the convolution layer can make full use of high-resolution images for multi-scale measurement and segmentation of wheat ears, and this will help to overcome the influence of dramatic changes in the shape of plant growth stage on the segmentation accuracy [85,86].

4.3. *Information complementarity realized by multimodal fusion*

Due to the complexity of the agricultural image shooting environment, a multi-modal fusion approach can be used for agricultural semantic segmentation. Multiple classes of information are provided to complement each other for the same segmentation scene. Multi-modal fusion methods take advantage of multiple information sources to generate more accurate predictions than a single source when used for semantic image segmentation. Features from different information sources are fused into different stages of the segmentation network according to their characteristics.

In the field of agriculture, the relevant modalities contain RGB-D imaging [88-90], near-infrared imaging [91-93], polarimetric imaging [94-96], etc. For instance, the Freiburg Forest dataset [87] contains color, depth, and near-infrared images annotated with six categories of classification. Because of the diversity of plant types in natural forest environments, this dataset provides enhanced normalized vegetation index images to improve the accuracy of segmentation. At the same time, textual information can be used as a complement to image information. The disease recognition model based on "image-text" multimodal collaborative representation and knowledge assistance (ITK-Net) constructed using semantic embedding method can achieve 99% accuracy [97]. And Fuse-Net is a clear example of incorporating the auxiliary depth information into an Encoder-Decoder segmentation framework [98]. The abstract features obtained from the depth-encoder are simultaneously fused to the RGB branch as the network goes deeper.

Finally, some promising future directions are discussed, in which we suggest to consider the following aspects. (1) Supervised methods always need a large number of labeled samples, which is a huge burden for agricultural image segmentation tasks. Agricultural images often have complex textures and complicated scenes. In future works, weakly supervised or unsupervised learning methods will have great potential, and transfer learning can also be used to reduce the number of training samples by transfer the learned knowledge to new constructed network. (2) 3D agricultural images often contain high-value information that is of great guidance to agricultural production. Efficient processing of 3D agricultural images is also a fantastic work in the future.

5. Conclusion

Semantic segmentation methods used for agricultural tasks have developed rapidly in recent years and have been widely used in cultivated land segmentation and Predictive Agriculture (PA). Semantic segmentation technology also has potential in other agriculture-related fields, including plant water stress index detection, water erosion assessment, pollutant identification, and food disease or defect identification. These models can also be used in environmental informatics to estimate the impact of various physical or artificial processes on the environment. The progress of image capturing equipment makes the collected images have higher resolution and more sensitive to color and texture features. CNNs and their variants have demonstrated great success in numerous computer vision applications. Traditional methods are often used as tools in the pre-processing stage or derived into functional modules in deep convolutional networks to enhance training results. We review the traditional segmentation methods for agricultural images and the segmentation methods based on the advanced CNN architecture, which enable the transformation of semantic segmentation methods to adapt to various data patterns and characteristics of agricultural images.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the Post-graduate's Innovation Fund Project of Hebei University (HBU2022ss037), and the High-Performance Computing Center of Hebei University.

REFERENCES

[1] Yuan Y, Chen L, Wu H, Li L. Advanced agricultural disease image recognition technologies: A review. Inform Process Agric 2022;9(1):48–59.
[2] Qiao Y, Truman M, Sukkarieh S. Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming. Comput Electron Agric 2019;165 104958.
[3] Yang S, Chen Q, Yuan X, Liu X. Adaptive coherency matrix estimation for polarimetric SAR imagery based on local heterogeneity coefficients. IEEE Trans Geosci Remote Sens 2016;54:6732–45.
[4] Kussul N, Lavreniuk M, Skakun S, Shelestov A. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci Remote Sens Lett 2017;14:778–82.
[5] Jadhav J, Singh R. Automatic semantic segmentation and classification of remote sensing data for agriculture. Mathematical Models in Engineering 2018;4:112–37.
[6] Dechesne C, Mallet C, Le Bris A, Gouet-Brunet V. Semantic segmentation of forest stands of pure species as a global optimization problem. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences 2017;IV-1/W1:141-148.

[7] Zou K, Chen X, Wang Y, Zhang C, Zhang F. A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field. Comput Electron Agric 2021;187 106242.
[8] Anand T, Sinha S, Mandal M, Chamola V, Yu F. AgriSegNet: Deep aerial semantic segmentation framework for IoT-assisted precision agriculture. IEEE Sens J 2021;21:17581–90.
[9] Singh P, Verma A, Alex J. Disease and pest infection detection in coconut tree through deep learning techniques. Comput Electron Agric 2021;182 105986.
[10] Ma J, Zheng F, Zhang L, Sun Z. Disease recognition system for greenhouse cucumbers based on deep convolutional neural network. Trans Chinese Soc Agric Eng 2018;34 (12):186–92.
[11] Zhao S, Liu J, Bai Z, Hu C, Jin Y. Crop pest recognition in real agricultural environment using convolutional neural networks by a parallel attention mechanism. Front Plant Sci 2022;13 839572.
[12] Liu G, Bai L, Zhao M, Zang H, Zheng G. Segmentation of wheat farmland with improved U-Net on drone images. J Appl Remote Sens 2022;16(3) 034511.
[13] Riehle D, Reiser D, Griepentrog H. Robust index-based semantic plant/background segmentation for RGB-images. Comput Electron Agric 2020;169 105201.
[14] Yang W, Wang S, Zhao X, Zhang J, Feng J. Greenness identification based on HSV decision tree. Inform Process Agric 2015;2(3–4):149–60.
[15] Mortensen A, Dyrmann M, Karstoft H, Jørgensen R. Semantic segmentation of mixed crops using deep convolutional neural network. Organising Committee CIGR 2016:1–6.
[16] Zhuang S, Wang P, Jiang B. Segmentation of green vegetation in the field using deep neural networks. World Congress on Intelligent Control and Automation (WCICA). IEEE; 2018. p. 509-514.
[17] Wang A, Xu Y, Wei X, Cui B. Semantic segmentation of crop and weed using an Encoder-Decoder network and image enhancement method under uncontrolled outdoor illumination. IEEE Access 2020;8:81724–34.
[18] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention; 2015. p. 234-241.
[19] Purkait P, Zach C, Reid I. Seeing behind things: Extending semantic segmentation to occluded regions. International Conference on Intelligent Robots and Systems (IROS). IEEE; 2019. p. 1998-2005.
[20] Liu C, Su J, Wang L, Lu S, Li L. LA-DeepLab V3+: A novel counting network for pigs. Agriculture 2022;12(2):1–16.
[21] Zhang J, Rao Y, Man C, Jiang Z, Li S. Identification of cucumber leaf diseases using deep learning and small sample size for agricultural Internet of Things. Int J Distrib Sens Netw 2021;17(4):1–13.
[22] Tassis L, de Souza J, Krohling R. A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images. Comput Electron Agric 2021;186 106191.
[23] Gonçalves J, Pinto F, Queiroz D, Villa F, Barbedo J, Del P. Deep learning architectures for semantic segmentation and automatic estimation of severity of foliar symptoms caused by diseases or pests. Biosyst Eng 2021;210:129–42.
[24] Skovsen S, Dyrmann M, Mortensen A, Laursen M, Gislum R, Eriksen J, et al. The Grassclover image dataset for semantic and hierarchical species understanding in agriculture. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. IEEE/CVF; 2019. p. 2676-2684.
[25] Sa I, Che Z, Popović M, Khanna R, Liebisch F, Nieto J, et al. WeedNet: Dense semantic weed classification using multispectral images and MAV for smart farming. IEEE Rob Autom Lett 2017;3(1):588–95.
[26] Fuentes-Pacheco J, Torres-Olivares J, Roman-Rangel E, Cervantes S, Juarez-Lopez P, Hermosillo-Valadez J, et al. Fig plant segmentation from aerial images using a deep convolutional Encoder-Decoder network. Remote Sens (Basel) 2019;11(10):1157.
[27] Kang J, Liu L, Zhang F, Shen C, Wang N, Shao L. Semantic segmentation model of cotton roots in-situ image based on attention mechanism. Comput Electron Agric 2021;189 106370.
[28] Yang M, Tseng H, Hsu Y, Tsai H. Semantic segmentation using deep learning with vegetation indices for rice lodging identification in multi-date UAV visible images. Remote Sens (Basel) 2020;12(4):633.
[29] Bragagnolo L, da Silva R, Grzybowski J. Amazon forest cover change mapping based on semantic segmentation by U-Nets. Eco Inform 2021;62 101279.
[30] Maheswari P, Raja P, Apolo-Apolo O, Pérez-Ruiz M. Intelligent fruit yield estimation for orchards using deep learning based semantic segmentation techniques—A review. Front Plant Sci 2021;12 684328.
[31] Singh R, Rani R. Semantic segmentation using deep convolutional neural network: A review. In Proceedings of the International Conference on Innovative Computing & Communications (ICICC); 2020.
[32] Gilles J. Empirical wavelet transform. IEEE Trans Signal Process 2013;61(16):3999–4010.
[33] Ţălu Ş. Mathematical methods used in monofractal and multifractal analysis for the processing of biological and medical data and images. ABAH Bioflux 2021;4(1):1–4.
[34] Gao J, Wang B, Wang Z, Wang Y, Kong F. A wavelet transform-based image segmentation method. Optik 2020;208 164123.
[35] Guijarro M, Riomoros I, Pajares G, Zitinski P. Discrete wavelets transform for improving greenness image segmentation in agricultural images. Comput Electron Agric 2015;118:396–407.
[36] Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybernet 1979;9(1):62–6.
[37] Wu Z, Gao Y, Li L, Xue J, Li Y. Semantic segmentation of high-resolution remote sensing images using fully convolutional network with adaptive threshold. Connect Sci 2019;31 (2):169–84.
[38] Hassanein M, Lari Z, El-Sheimy N. A new vegetation segmentation approach for cropped fields based on threshold detection from hue histograms. Sensors 2018;18(4):1253.
[39] Dorj U, Lee M, Yun S. An yield estimation in citrus orchards via fruit detection and counting using image processing. Comput Electron Agric 2017;140:103–12.
[40] Castillo-Martínez M, Gallegos-Funes F, Carvajal-Gámez B, Urriolagoitia-Sosa G, Rosales-Silva A. Color index based thresholding method for background and foreground segmentation of plant images. Comput Electron Agric 2020;178 105783.
[41] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell 2012;34 (11):2274–82.
[42] Giraud R, Ta V, Papadakis N. Robust superpixels using color and contour features along linear path. Comput Vis Image Underst 2018;170:1–13.
[43] Di S, Liao M, Zhao Y, Li Y, Zeng Y. Image superpixel segmentation based on hierarchical multi-level LI-SLIC. Opt Laser Technol 2021;135 106703.
[44] Khan S, Narvekar M. Novel fusion of color balancing and superpixel based approach for detection of tomato plant diseases in natural complex environment. J King Saud University-Computer and Information Sciences 2020;30 (58):13495.

[45] Mohanty S, Hughes D, Salathé M. Using deep learning for image-based plant disease detection. Front Plant Sci 2016;7:1419.
[46] Lu H, Cao Z, Xiao Y, Li Y, Zhu Y. Region-based colour modelling for joint crop and maize tassel segmentation. Biosyst Eng 2016;147:139–50.
[47] Tu W, Liu M, Jampani V, Sun D, Chien S, Yang M, et al. Learning superpixels with segmentation-aware affinity loss. CVPR 2018:568–76.
[48] Jampani V, Sun D, Liu M, Yang M, Kautz J. Superpixel sampling networks. ECCV 2018:363–80.
[49] Yang F, Sun Q, Jin H, Zhou Z. Superpixel segmentation with fully convolutional networks. CVPR 2020:13961–70.
[50] Gonçalves D, Moares W, Pistori J, Costa G, Araujo R, Pereira A, et al. Carcass image segmentation using CNN-based methods. Information Processing in Agriculture 2020;8 (4):560–72.
[51] Zheng Y, Kong J, Jin X, Wang X, Zhang N. CropDeep: The crop vision dataset for deep-learning-based classification and detection in precision agriculture. Sensors 2019;19(5):1058.
[52] Ale L, Sheta A, Li L, Wang Y, Zhang N. Deep learning based plant disease detection for smart agriculture. Globecom Workshops (GC Wkshps) IEEE 2019:1–6.
[53] Milioto A, Lottes P, Stachniss C. Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. international conference on robotics and automation. IEEE; 2018.p. 2229-2235.
[54] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE; 2015. p. 3431-3440.
[55] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional Encoder-Decoder architecture for image segmentation. Transactions on Pattern Analysis and Machine Intelligence. IEEE; 2017. p. 2481-2495.
[56] Song Z, Zhang Z, Yang S, Ding D, Ning J. Identifying sunflower lodging based on image fusion and deep semantic segmentation with UAV remote sensing imaging. Comput Electron Agric 2020;179 105812.
[57] Mukhopadhyay P, Mallick S. Bayesian deep learning for seismic facies classification and its uncertainty estimation. Technical Program Expanded Abstracts SEG 2019:2488–92.
[58] Wang J, Liu W, Gou A. Numerical characteristics and spatial distribution of panoramic street green view index based on SegNet semantic segmentation in Savannah. Urban For Urban Green 2022;69 127488.
[59] Chen L, Papandreou G, Kokkinos I, Murphy K, Yuille A. Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv preprint arXiv:1412.7062, 2014.
[60] Chen L, Papandreou G, Kokkinos I, Murphy K, Yuille A. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 2017;40(4):834–48.
[61] Chen L, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv: 1706.05587, 2017.
[62] Chen L, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European conference on computer vision. ECCV; 2018. p .801-818.
[63] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. European conference on computer vision 2016:630–45.
[64] Peng H, Xue C, Shao Y, Chen K, Xiong J, Xie Z, et al. Semantic segmentation of litchi branches using DeepLabV3+ model. IEEE Access 2020;8:164546–55.
[65] Yuan H, Zhu J, Wang Q, Cheng M, Cai Z. An improved DeepLabV3+deep learning network applied to the segmentation of grape leaf black rot spots. Front Plant Sci 2022;13:795410.
[66] Alzu'bi A, Alsmadi L. Monitoring deforestation in Jordan using deep semantic segmentation with satellite imagery. Eco Inform 2022;70:101745.
[67] Sun K, Wang X, Liu S, Liu C. Apple, peach, and pear flower detection using semantic segmentation network and shape constraint level set. Comput Electron Agric 2021;185 106150.
[68] Jo Y, Lee S, Lee Y, Kahng H, Park S, Bae S, et al. Semantic segmentation of cabbage in the South Korea Highlands with images by unmanned aerial vehicles. Appl Sci 2021;11 (10):4493.
[69] Huang G, Liu Z, Maaten L, Weinberger K. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE; 2017. p. 4700-4708.
[70] Zhang X, Zhou X, Lin M, ShuffleNet SJ. An extremely efficient convolutional neural network for mobile devices. In: The IEEE conference on computer vision and pattern recognition. CVPR; 2018. p. 6848–56.
[71] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[72] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Proces Syst 2012;25:1097–105.
[73] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D et al. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE; 2015. p. 1-9.
[74] Chollet F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE; 2017. p. 1251-1258.
[75] Shehab L, Fahmy O, Gasser S, El-Mahallawy M. An efficient brain tumor image segmentation based on deep residual networks (ResNets). Journal of King Saud University-Engineering Sciences 2020;33(6):404–12.
[76] Kamilaris A, Prenafeta-Boldú F. A review of the use of convolutional neural networks in agriculture. J Agric Sci 2018;156(3):312–22.
[77] Espejo-Garcia B, Mylonas N, Athanasakos L, Vali E, Fountas S. Combining generative adversarial networks and agricultural transfer learning for weeds identification. Biosyst Eng 2021;204:79–89.
[78] Wu Q, Chen Y, Meng J. DCGAN-based data augmentation for tomato leaf disease identification. IEEE Access 2020;8:98716–28.
[79] Abbas A, Jain S, Gour M, Vankudothu S. Tomato plant disease detection using transfer learning with C-GAN synthetic images. Comput Electron Agric 2021;187 106279.
[80] Nazki H, Yoon S, Fuentes A, Park DS. Unsupervised image translation using adversarial networks for improved plant disease recognition. Comput Electron Agric 2020;168 105117.
[81] Potlapally A, Chowdary P, Shekhar S, Mishra N, Madhuri C, Prasad A. Instance segmentation in remote sensing imagery using deep convolutional neural networks. In: 2019 International Conference on contemporary Computing and Informatics. IEEE; 2019. p. 117–20.
[82] Pahlevan N, Chittimalli S, Balasubramanian S, Vellucci V. Sentinel-2/Landsat-8 product consistency and implications for monitoring aquatic systems. Remote Sens Environ 2019;220:19–29.
[83] Nogueira K, Dalla Mura M, Chanussot J, Schwartz W, Dos Santos J. Dynamic multicontext segmentation of remote

sensing images based on convolutional networks. IEEE Trans Geosci Remote Sens 2019;57(10):7503–20.

[84] Osco L, Nogueira K, Ramos A, Pinheiro M, Furuya D, Gonçalves W, et al. Semantic segmentation of citrus-orchard using deep neural networks and multispectral UAV-based imagery. Precis Agric 2021:1–18.

[85] Ma J, Lin Y, Liu H, Du K, Zheng F, Wu Y, et al. Improving segmentation accuracy for ears of winter wheat at flowering stage by semantic segmentation. Comput Electron Agric 2020;176 105662.

[86] Banasiak P, Berezowski P, Zapłata R, Mielcarek M, Duraj K, Stereńczak K. Semantic segmentation (U-Net) of archaeological features in airborne laser scanning—example of the Białowieża forest. Remote Sens (Basel) 2022;14(4):995.

[87] Valada A, Oliveira G, Brox T, Burgard W. Deep multispectral semantic scene understanding of forested environments using multimodal fusion. International Symposium on Experimental Robotics 2017:465–77.

[88] Song Y, Xu F, Yao Q, Liu J, Yang S. Navigation algorithm based on semantic segmentation in wheat fields using an RGB-D camera. Information Processing in Agriculture 2022. https://doi.org/10.1016/j.inpa.2022.05.002.

[89] Giang T, Khai T, Im D, Ryoo Y. Fast detection of tomato sucker using semantic segmentation neural networks based on RGB-D images. Sensors 2022;22(14):5140.

[90] Masuda T. Leaf area estimation by semantic segmentation of point cloud of tomato plants. Proceedings of the IEEE/CVF International Conference on Computer Vision 2021:1381–9.

[91] Mishra P, Sadeh R, Bino E, Polder G, Boer M, Rutledge D, et al. Complementary chemometrics and deep learning for semantic segmentation of tall and wide visible and near-infrared spectral images of plants. Comput Electron Agric 2021;186 106226.

[92] Anand T, Sinha S, Mandal M, Chamola V, Yu F. AgriSegNet: Deep aerial semantic segmentation framework for IoT-assisted precision agriculture. IEEE Sens J 2021;21 (16):17581–90.

[93] Bhagat S, Kokare M, Haswani V, Hambarde P, Kamble R. MS-Net: A CNN architecture for agriculture pattern segmentation in aerial images. International Conference on Computer Vision and Image Processing 2022. p. 489-500.

[94] Zhang R, Chen J, Feng L, Li S, Yang W, Guo D. A refined pyramid scene parsing network for polarimetric SAR image semantic segmentation in agricultural areas. IEEE Geosci Remote Sens Lett 2021;19:1–5.

[95] Verma U, Chauhan A, MM M, Pai R. DeepRivWidth: Deep learning based semantic segmentation approach for river identification and width measurement in SAR images of Coastal Karnataka. Computers & Geosciences 2021;154:104805.

[96] Mehra A, Jain N, Srivastava H. A novel approach to use semantic segmentation based deep learning networks to classify multi-temporal SAR data. Geocarto Int 2022;37 (1):163–78.

[97] Zhou J, Li J, Wang C, Wu H, Zhao C, Teng G. Crop disease identification and interpretation method based on multimodal deep learning. Comput Electron Agric 2021;189 106408.

[98] Hazirbas C, Ma L, Domokos C, Cremers D. FuseNet: Incorporating depth into semantic segmentation via fusion-based CNN architecture, Lecture Notes in Computer Science; 2017. p. 213-228.