Improved semantic segmentation method using edge features for winter wheat spatial distribution extraction from Gaofen-2 images

Hao Yin,[a] Chengming Zhang,[a,*] Yingjuan Han,[b] Yonglan Qian,[c] Tao Xu,[d] Ziyun Zhang,[a] and Ailing Kong[a]

[a]Shandong Agricultural University, College of Information Science and Engineering, Taian, Shandong, China

[b]Key Laboratory for Meteorological Disaster Monitoring and Early Warning and Risk Management of Characteristic Agriculture in Arid Regions, CMA, Yinchuan, Ningxia, China

[c]National Meteorological Center/CMA, Beijing, China

[d]University of Jinan, Jinan, Shandong, China

Abstract. In the final feature map obtained using a convolutional neural network for remote sensing image segmentation, there are great differences between the feature values of the pixels near the edge of the block and those inside the block; ensuring consistency between these feature values is the key to improving the accuracy of segmentation results. The proposed model uses an edge feature branch and a semantic feature branch called the edge assistant feature network (EFNet). The EFNET model consists of one semantic branch, one edge branch, one shared decoder, and one classifier. The semantic branch extracts semantic features from remote sensing images, whereas the edge branch extracts edge features from remote sensing images and edge images. In addition, the two branches extract five-level features through five sets of feature extraction units. The shared decoder sets up five levels of shared decoding units, which are used to further integrate edge features and deep semantic features. This strategy can reduce the feature differences between the edge pixels and the inner pixels of the object, obtaining a per-pixel feature vector with high inter-class differentiation and intra-class consistency. Softmax is used as the classifier to generate the final segmentation result. We selected a representative winter wheat region in China (Feicheng City) as the study area and established a dataset for experiments. The comparison experiment included three original models and two models modified by adding edge features: SegNet, UNet, and ERFNet, and edge-UNet and edge-ERFNet, respectively. EFNet's recall (91.01%), intersection over union (81.39%), and F1-Score (91.68%) were superior to those of the other methods. The results clearly show that EFNET improves the accuracy of winter wheat extraction from remote sensing images. This is an important basis not only for crop monitoring, yield estimation, and disaster assessment but also for calculating land carrying capacity and analyzing the comprehensive production capacity of agricultural resources. [DOI: 10.1117/1.JRS.15.028501]

Keywords: convolutional neural network; remote sensing; semantic segmentation; winter wheat; edge feature; Gaofen-2.

Paper 200908 received Dec. 31, 2020; accepted for publication Apr. 26, 2021; published online May 25, 2021.

1 Introduction

Image semantic segmentation assigns a label value to each pixel in an image.[1] Researchers use the shallow features such as image color, grayness, and texture along with deep semantic features obtained via deep learning to assign each pixel in the image to its category.[2] Using semantic segmentation to extract crops in remote sensing images accurately and in real time can maximize

*Address all correspondence to Chengming Zhang, chming@sdau.edu.cn

agricultural productivity. Detailed crop coverage maps are essential for agricultural monitoring,[3,4] food security,[5,6] and responding to global environmental changes.[7,8]

Early image semantic segmentation algorithms extracted features primarily based on manual design; they include scale-invariant feature conversion,[9] directional gradient histogram [histogram of oriented gradients(HOG)],[10] and histogram back projection.[11] More recently, machine learning methods based on probabilistic map models have been proposed, such as the Markov random field,[12] Bayesian network,[13] and conditional random field (CRF).[14] However, these methods rely excessively on manually labeled feature libraries.[15,16] Hence, a large amount of manpower is required, which greatly restricts the practical application of these methods.

Deep learning methods have contributed to the development of image segmentation through mining pixel information, neural networks,[17] support vector machines,[18] decision trees,[19] and random forests.[20] Further, using a neural network can ensure that feature extraction and classification are completed concurrently. The deep belief network[21] achieved satisfactory performance in image segmentation.[22,23] When used for image segmentation, the network completely mines the characteristics of each pixel, and the spatial features between pixels can be better utilized. Therefore, the convolutional neural network (CNN) sets reasonable convolution kernels and batch normalization (BN) layers in the image characteristics and segmentation requirements to build a network structure.[24] This approach has achieved highly accurate results.[25–27] A fully convolutional network (FCN)[28] realizes end-to-end training, which is more conducive to the automatic extraction of refined winter wheat.[29,30] However, the accuracy of the extraction result is limited owing to the local receiving field and short-term context information.[31,32] Some encoder–decoder networks are used to gradually restore edge details,[33–36] such as UNet,[37] SegNet,[38] and DeepLabv3.[39] Despite their success, some details are lost after the encoder downsampling, which means that predictions are often less accurate near the boundaries. RefineNet[40] uses features from various levels to make semantic segmentation near the boundaries more precise. Romera et al.[41] proposed a lightweight real-time semantic segmentation model, ERFNet, which uses the decomposition convolution of the one-dimensional convolution kernel to replace the traditional convolution. The rise of the attention mechanism has seen its introduction into semantic segmentation, such as non-local neural networks,[42] CCNet,[43] and DANet.[44] This mechanism can more effectively performs accurate segmentation within a long-range context. However, the results obtained by upsampling with these models are still rough and some details in the images should not be ignored. These methods may be optimized by introducing prior knowledge.

Pixels will have strong grayscale jumps at the junction of different surface types. Edge detection algorithms are typically used to distinguish boundary pixels from internal pixels. Traditional edge detection methods use derivative operators to highlight grayscale changes and set thresholds through derivative values.[45] The Roberts operator uses the local difference operator to find the edge.[46] The Sobel and Prewitt operators use the maximum value obtained by convolution in the horizontal or vertical directions. The Canny operator judges the edge operator according to the signal-to-noise ratio, positioning accuracy, and unilateral response standards.[47,48] In most cases, the Canny algorithm performs better than most commonly used edge detection algorithm.[49]

With the advancement of imaging technology and hardware equipment, more feature information can be mined from high-resolution remote sensing images. Mixing pixel issues have also been mitigated. The edge pixels themselves may be mixed pixels, but the edge features obtained by capturing regional changes through edge detection can improve the edge fineness of the segmentation results. In image segmentation, the integration of edge features and semantic features improves the effectiveness of features in distinguishing different crops.[50–53] Chen et al.[54] proposed the use of a domain transform as filtering to generate segmentation specific edges in end-to-end trainable systems, thus optimizing semantic segmentation quality. Liu et al.[52] merged a semantic segmentation network and an edge network and used regularization methods to optimize the resultant composite network. Marmanis et al.[48] proposed a new cascade model combining edge and semantic segmentation networks. He et al.[55] introduced edge information as a priori knowledge and used the feature information extracted from the overall nested edge branch to correct the FCN segmentation result. However, these methods use end-to-end networks to combine semantic segmentation with edge detection, which requires a large number of parameters to be trained simultaneously and consumes more time and computing power.

The main purpose of this research is to develop a strategy to improve the semantic segmentation of CNNs by combining edge features, and to propose a remote sensing image segmentation model for obtaining the spatial distribution of winter wheat with high precision.

The main contributions of this article are as follows:

A new image segmentation dataset structure is proposed, i.e., a high-resolution remote sensing dataset is established based on Feicheng City, China. The dataset consists of image blocks of remote sensing images, corresponding artificial label maps, and corresponding generated edge maps. The edge map generated based on the edge detection algorithm is used as part of the model to participate in the training and testing of the model. A semantic segmentation dataset for winter wheat in Shandong Province, Gaofen-2, is established.

A new CNN framework with semantic features and edge features is constructed. The semantic branch and edge branch extract shallow semantic features and edge features, respectively. The shared decoder further extracts the shallow semantic features and performs upsampling to restore the feature map size. Finally, the feature fusion module is used to fuse multi-scale edge features and deep semantic features to obtain more distinguishable features with richer edge details. Furthermore, the framework is incorporated into other models to improve their segmentation results. Based on the proposed edge feature framework, a new CNN model fine feature network [edge assistant feature network (EFNet)] is established.

2 Materials and Methods

2.1 *Dataset*

2.1.1 *Study area*

As shown in Fig. 1, Feicheng City is located in the central area of Shandong Province, China, at the western foot of Mount Tai. Its geographic coordinates are 35°53′–36°19′ N and 116°28′–116° 59′ E. The city is 48-km long in the north–south direction and 37.5-km wide in the east–west

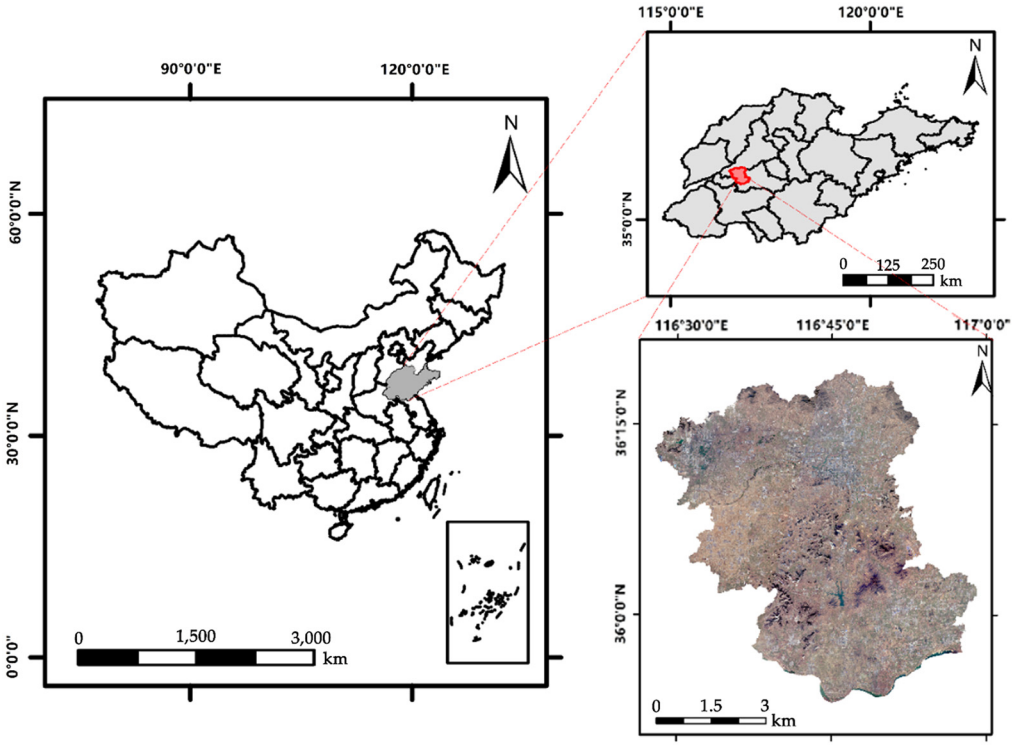

Fig. 1 Geographical location of Feicheng City, Shandong Province, China.

direction, with a total area of 1277 km^2. This area is an important crop production region in Northern China. The area is within a temperate continental semi-humid monsoon climate zone, with four distinct seasons and sufficient light and heat for crop growth. The annual average temperature is 13°C, annual average sunshine time is 2536.21 h, and annual average rainfall is 688.3 mm.

2.1.2 *Dataset Creation*

We collected 40 Gaofen-2 (GF-2) remote sensing images covering the entire study area from November 2017 to April 2018 and an additional 37 images from November 2018 to April 2019. Each GF-2 image is composed of a multispectral image and a panchromatic image. After preprocessing, each remote sensing image contained four channels: red, blue, green, and near-infrared, with a spatial resolution of 1 m. The primary land use types in the images collected included winter wheat, buildings, roads, woodland, mountains, and lakes. As winter wheat was the primary crop in the pre-processed images, we selected winter wheat as the extraction target to experimentally demonstrate the effectiveness of the method.

We divided the obtained Feicheng City GF-2 image into 7126 image blocks of equal size (512×512 pixels). We then selected 20% of all images (1425) for manual annotation, of which 900 were used for model training, 125 for model verification, and 400 for model quantitative analysis. When selecting image blocks, we chose representative and complex image blocks containing winter wheat under various terrains and adjacent to other ground objects. Convolutional neural networks rely on a large number of label maps to train the model and extract effective features. Data sets of sufficient quality can improve the accuracy of the results. This experiment prioritizes the accuracy of the data set as much as other similar models. During the production of the semantic segmentation data set, some pixels may be misinterpreted. In this experiment, through the guidance of professionals combined with ground surveys, we produced as accurate a map as possible. As shown in Fig. 2, four areas were selected for field investigation to verify the accuracy of the manual annotation graphs and correct the annotation maps in the dataset.

Unlike previous approaches, this study innovatively used the edge detection map as an important part of the dataset for model training and testing. The edge detection algorithm was used to create an edge detection map whose edge pixel value was equal to the original pixel value for each image block. As shown in Fig. 3, we established a dataset comprising the original, labeled, and edge images.

2.1.3 *Edge detection map*

Figure 4 shows the results of five different edge detection algorithms. Because of the sowing time and growing tendency of winter wheat, the spectrum, texture, and shape characteristics of the farmland varied. Compared with its counterparts, the edge map obtained by the Canny edge detection algorithm more accurately coincided with the edges of the actual winter wheat plot. Therefore, the Canny edge detection operator was selected to extract the edges of winter wheat production plots.

2.2 *EFNet Model*

In this experiment, a new CNN framework with improved edge features was developed (Fig. 5). First, semantic branches were used to extract the rich semantic information. Next, the edge information was generated by combining the edge branches with multi-scale semantic information. Finally, through the shared decoder module, multi-scale edge features and deep semantic features were combined to generate the final image segmentation results.

Based on the proposed model framework, the EFNet model is proposed for extracting features from remote sensing images. As shown in Fig. 6, the EFNet model network uses a semantic branch and an edge branch for feature extraction. The feature fusion unit in the shared decoder is used to fuse edge features of different scales with deep semantic features.

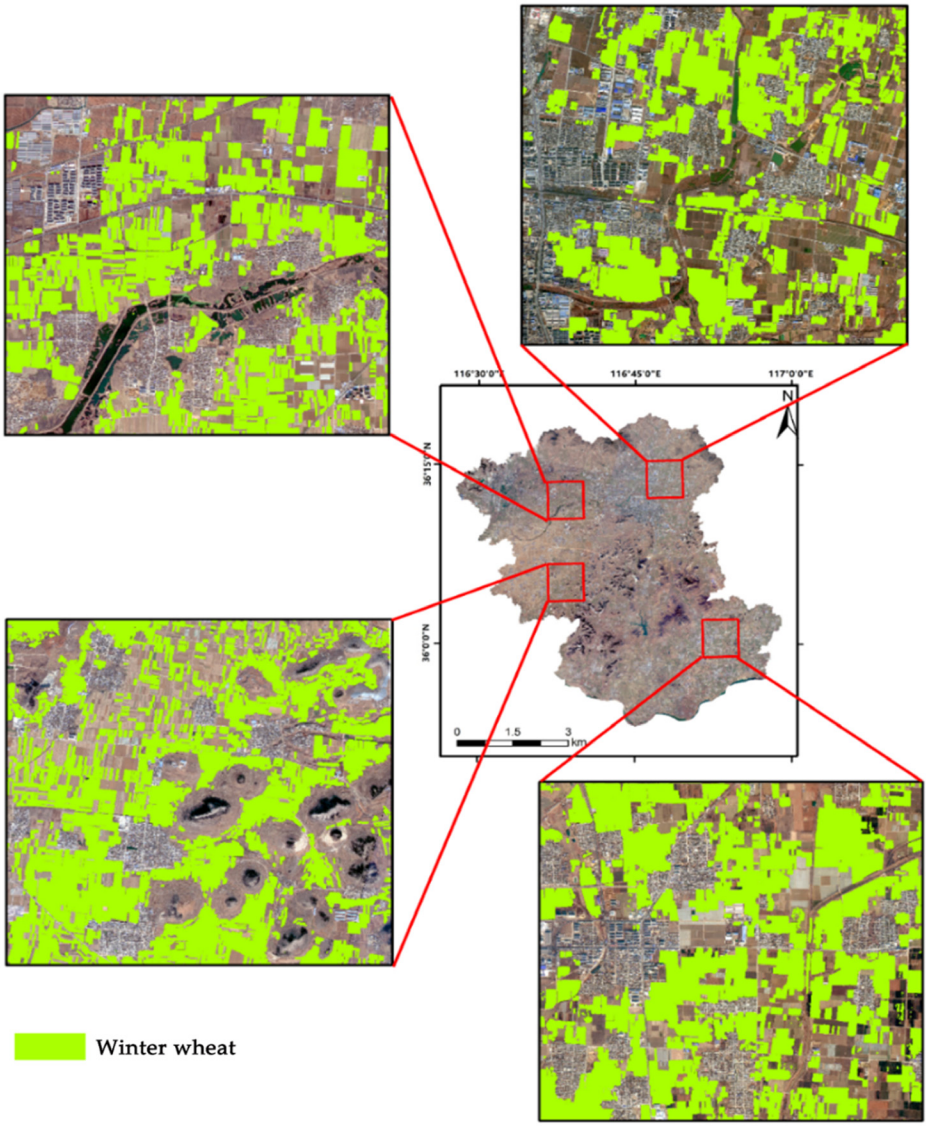

Fig. 2 Field investigation area.

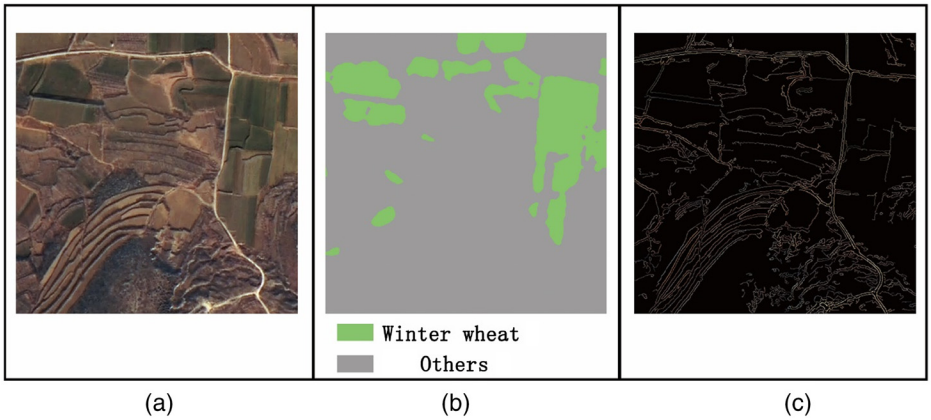

Fig. 3 An example of the winter wheat dataset: (a) image block; (b) manually labeled image corresponding to (a); (c) edge image corresponding to (a).

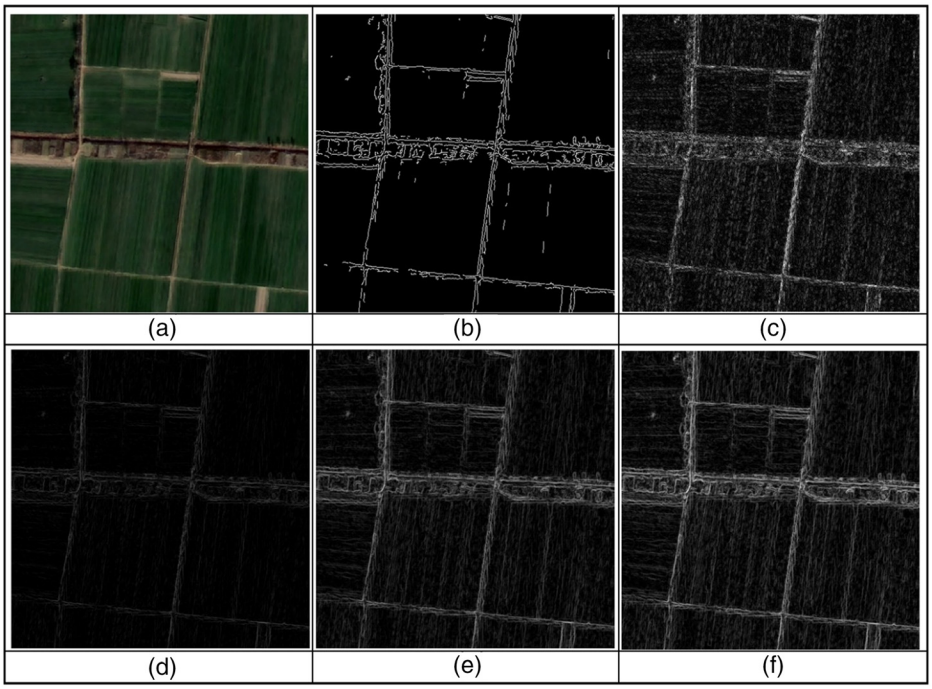

Fig. 4 Comparison of the segmentation results of five edge detection algorithms: (a) image block; (b) Canny edge detection map; (c) Laplacian edge detection map; (d) Roberts edge detection map; (e) Prewitt edge detection map; (f) Sobel edge detection map.

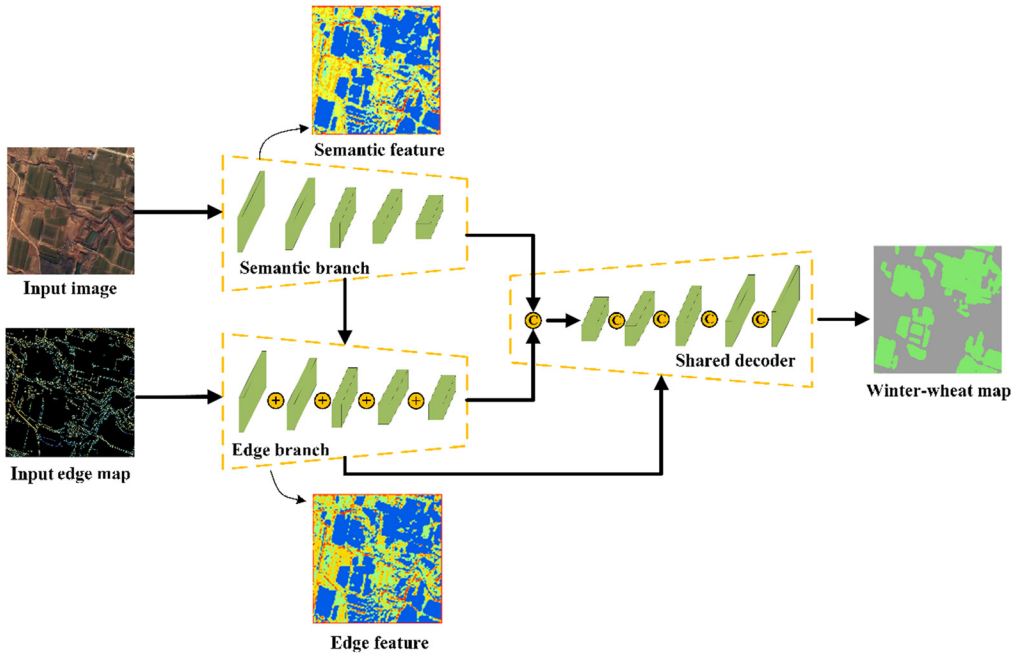

Fig. 5 Schematic diagram of the CNN framework with improved edge features.

2.2.1 *Semantic branch*

The semantic feature extraction unit is composed of five feature extraction units connected in series, which can extract five levels of semantic feature information for each pixel. Each feature extraction unit contains two or three convolution layers, one BN layer, one activation layer, and one 2×2 maximum pooling layer. Table 1 lists the number and size of the convolution kernels used by each convolution layer.

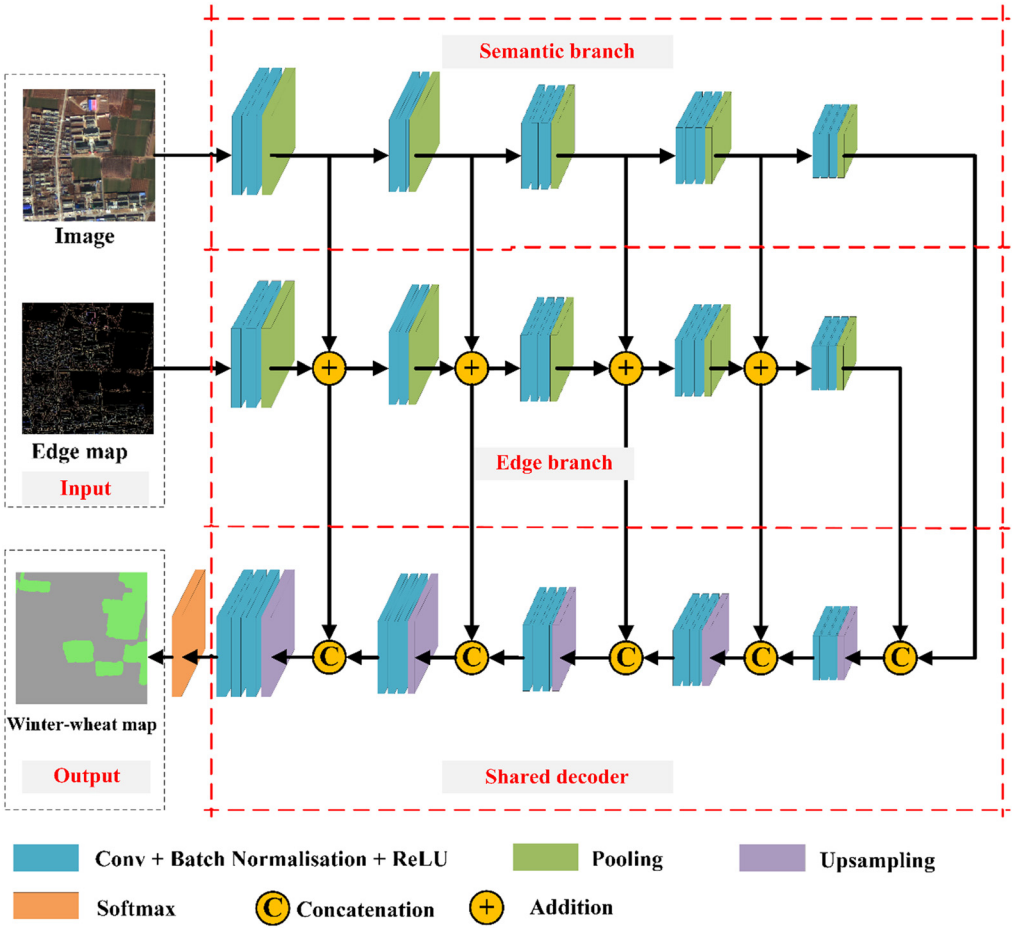

Fig. 6 EFNet model network structure.

Table 1 Number of convolution kernels in each convolution layer.

Convolution layer	Number of convolution kernels	Size of convolution kernels
1, 2	32	3
3, 4	64	3
5, 6	128	3
7	128	1
8, 9	256	3
10	256	1
11, 12	512	3
13	512	1

The continuous use of multiple 3×3 convolutional layers can achieve the effect of a larger convolution kernel and reduce the computational complexity. The activation layer is widely applied to the convolution of neural networks and uses the rectified linear activation function as the activation function. The maximum pooling layer discards some of the eigenvalues but retains the eigenvalues of the key points in the region. The computational complexity is reduced, but high classification accuracy is still guaranteed. To ensure that sufficient semantic features can

be extracted, this network structure avoids the problem that the use of an excessively deep convolution structure may cause significant noise in feature values. It is helpful for pixels of the same kind to generate pixel-by-pixel feature vectors with good stability. Because the feature value of the key point of the suspected edge pixel in the edge feature map is the weighted sum of the edge feature and the semantic feature of the same scale, the pixel point retained by the largest pool is the most representative edge feature key point in the receptive field.

2.2.2 *Edge branch*

The edge feature extraction unit is the same as the semantic feature extraction unit (Table 1), but with the feature compensation module added between each of the five feature extraction units. The working principle of feature compensation module is as follows: the semantic feature map after the same number of feature extraction units is multiplied by the trainable coefficient γ and added to the edge feature map [Eq. (1)] to develop a new edge feature map. γ can be automatically adjusted during model training, and the value of γ at each level is determined separately when training the model.

$$f_{\text{edge}}(n) = f_{\text{edge}}(n) + f_{\text{semantic}}(\text{n}) \times \gamma. \tag{1}$$

2.2.3 *Shared decoder*

The shared decoder is composed of five feature fusion modules, which fuse shallow edge features and deep semantic features to obtain fusion features (Fig. 7). The EFNet model adopts a gradual recovery strategy. Each upsampling layer adjustment doubles the number of rows and columns. Three convolutional layers are used to fuse two feature maps and adjust the feature values.

2.2.4 *Classifier and loss function*

The EFNet model uses the Softmax function [Eq. (2)] to calculate the probability of pixels belonging to each category in the feature map generated by the decoder. The EFNet model uses

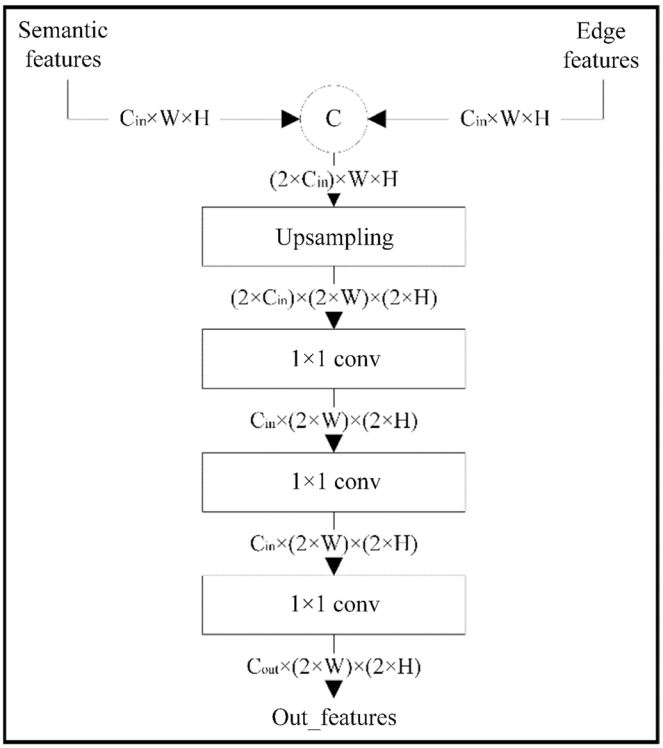

Fig. 7 Feature fusion module.

the category number corresponding to the maximum probability value as the pixel category. S_i is the Softmax value of the i'th node, V_i is the output value of the i'th node, and j is the number of output nodes, that is, the number of classification categories.

$$S_i = \frac{e^{V_i}}{\sum_j e^{V_j}}. \tag{2}$$

The EFNet model inputs the output (x) of Softmax and the label map (label) into the cross-entropy loss function [Eq. (3)] to calculate the loss value.

$$\text{loss}(x, \text{label}) = -\log \frac{e^{x[\text{label}]}}{\sum_{j=1}^{N} e^{x_j}} = -x[\text{label}] + \log\left(\sum_{j=1}^{N} e^{x_j}\right). \tag{3}$$

2.3 *EFNet Model Training*

2.3.1 *Optimizer*

The EFNet model uses adaptive moment estimation (Adam) as the optimizer. The Adam algorithm uses the gradient's first moment estimation (mean of the gradient) and second moment estimation (uncentered variance of the gradient) to adjust the learning rate of each parameter dynamically and calculate the update stride.

The Adam algorithm combines the advantages of the adaptive gradient and root-mean-square propagation optimization algorithms. The parameter update is not affected by the scaling transformation of the gradient, and it can automatically adjust the learning rate within the initial learning rate range. In addition, the hyperparameters usually require only minimal fine-tuning or do not need to be adjusted.

2.3.2 *Training method*

The EFNet model is used in an end-to-end manner. The specific training steps are as follows:

1. Determine the hyperparameters in the training process and initialize the parameters of the EFNet model.
2. Input the image blocks, label maps, and corresponding edge detection maps in the training set as training data for the EFNet model.
3. Use EFNet to perform a forward calculation on the current training data to obtain the output value.
4. Calculate the loss value of the true and predicted values according to Eq. 3. Use the Adam optimizer to update the parameters of the EFNet model according to the loss value.
5. Obtain a new experience file and complete the training.
6. Repeat steps 3, 4, and 5 for the specified number of times.

2.4 *Experimental Setup*

Three classic CNN models and three models combined with an edge feature improvement framework (Fig. 5) were selected for comparative experiments (Table 2). The three classic models that follow the encoder–decoder structure are widely used in semantic segmentation and have achieved notable results.

SegNet realizes end-to-end pixel-level image segmentation, and the information lost in the pooling process can be obtained in the decoder through upsampling. SegNet is used to explore the effect of adding only edge features or not adding edge features to the encoder. UNet allows learning of the missing features in the encoder pool at each stage of the decoder and connects the different scale feature maps of the encoder to the upsampled feature maps of the decoder at each stage. It is also used to explore the effect of combining edge features after the encoder combines

Table 2 Models used for comparison.

Number	Name	Description
1	UNet	Original UNet
2	UNet-edge	Improved UNet combined with edge features
3	ERFNet	Original ERFNet
4	ERFNet-edge	Improved ERFNet combined with edge features
5	SegNet	Original SegNet
6	EFNet	Ours

low-level semantic features. ERFNet uses non-bottleneck-1D to achieve a balance between accuracy and parameter volume and is also used to explore the effect of combining edge features on networks with fewer parameters.

As shown in Table 2, two configuration levels were used for each experiment: the original model and the CNN with improved edge features.

By setting different numbers of edge feature extraction units, four sets of ablation experiments were designed to explore the performance of the edge feature compensation unit and fully understand the network.

We used the horizontal flipping, vertical flipping, random radiation transformation, 90 deg rotation, and hue, saturation, value color space contrast transformation steps for image enhancement on the training dataset. All image enhancement techniques were used only for model training data sets.

2.5 *Evaluation Metrics*

Four metrics were used to evaluate the experimental results in our experiment: intersection over union (IOU), precision, recall, and F-score (F score). The results of winter wheat extraction can be divided into true positive (pixels correctly identified as winter wheat), true negative (pixels correctly identified as other types), false positive (pixels that are misclassified as winter wheat), and false negative (pixels that are misclassified as others).

The pixel-wise IOU score evaluates segmentation accuracy.

$$\mathrm{IoU} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP} + \mathrm{FN}). \tag{4}$$

Precision is used to indicate the proportion of winter wheat pixels that are accurately classified compared to all pixels identified as winter wheat.

$$\mathrm{Precision} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}). \tag{5}$$

Recall indicates the proportion of accurately classified winter wheat pixels compared to all actual winter wheat pixels.

$$\mathrm{Recall} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}), \tag{6}$$

The F-score is an index used to evaluate the accuracy of classification models in statistics. The F-score can be regarded as a weighted average of model accuracy and recall.

$$F_{\mathrm{score}} = 2 \times (\mathrm{Precision} \times \mathrm{Recall})/(\mathrm{Precision} + \mathrm{Recall}). \tag{7}$$

3 Results

The three models SegNet, ERFNet, and UNet were selected for the experiments in the early stage. Then, the models with improved edge features were selected. As shown in Fig. 8, we

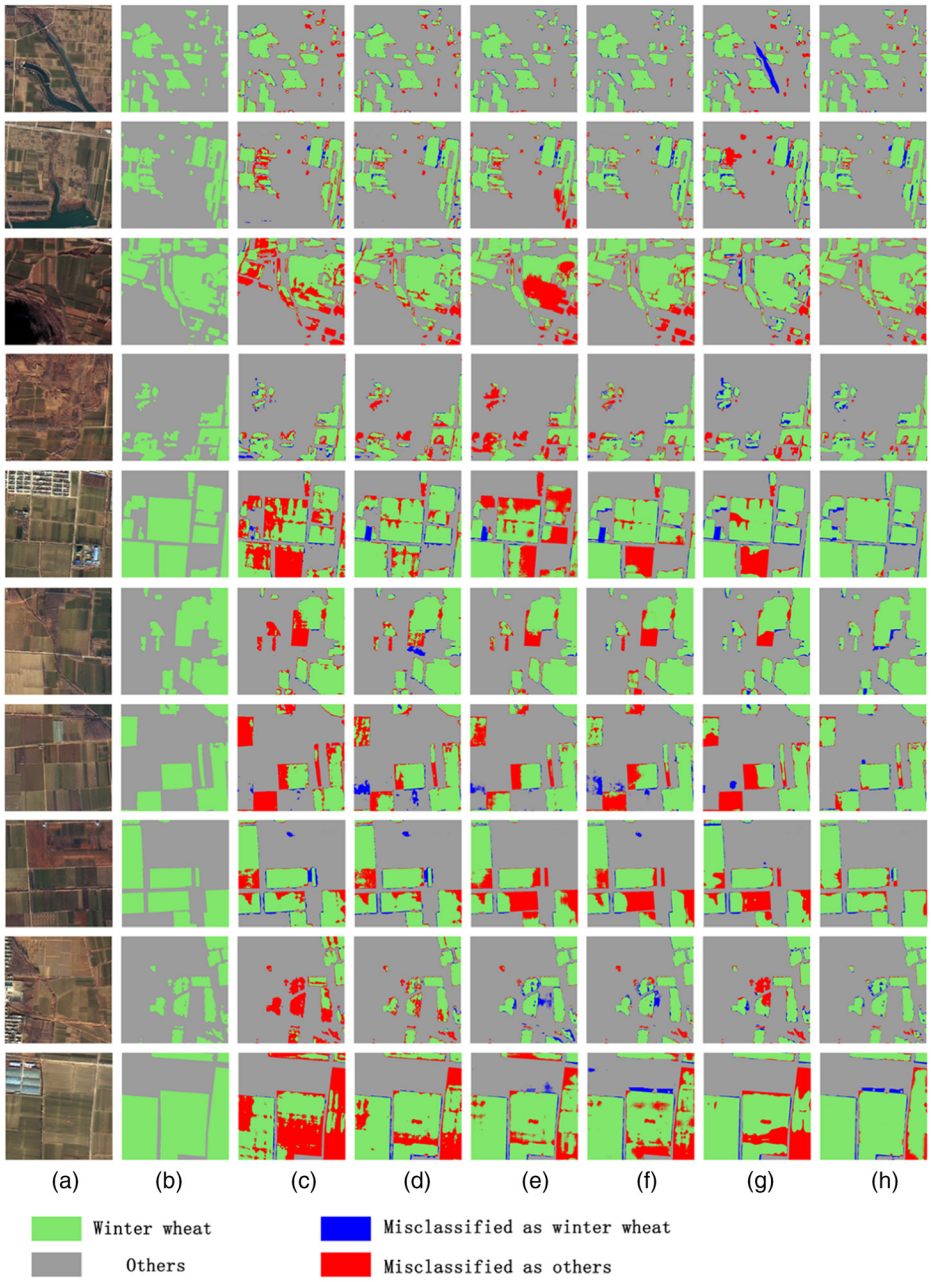

Fig. 8 Comparison of the segmentation results for 10 randomly selected image blocks: (a) original images; (b) manually labeled images corresponding to (a); (c) UNet; (d) UNet-edge; (e) ERFNet; (f) ERFNet-edge; (g) SegNet; (h) EFNet.

compared and analyzed the results of the six experiments. In the first row, the spectrum and texture of the water area and winter wheat are similar, and the SegNet model mistakenly identified the river as winter wheat. However, this problem was solved after introducing edge features. Affected by topography and landform, winter wheat planting plots are small in area and irregular in shape (second, third, and fourth rows of Fig. 8). The simple combination of semantic features cannot accurately segment winter wheat.

After introducing edge features, small planting areas were also effectively segmented. Affected by the growth of winter wheat, the spectral gaps of winter wheat in the same plot are relatively large, resulting in them not being fully recognized (fifth and sixth rows of Fig. 8). In addition, the color of some winter wheat with poor growth conditions was lighter. The model that combined edge features better resolved this problem than the original model. Specifically, EFNet identified almost all winter wheat plots [Figs. 5(h) and 8] in the image block [Figs. 5(a) and 8].

Because of the different planting times of winter wheat, the spectra between various plots could also have large differences (ninth and tenth rows of Fig. 8). The introduction of edge features increased the difference between the edge pixels and the interior; hence, the feature values of the same category were more similar. In other challenging conditions, such as the roads and plots in rural areas being easy to confuse (seventh and eighth rows of Fig. 8), EFNet demonstrated better performance and obtained more refined edges.

The fusion of edge features and semantic features is used to solve the problems of inconsistency within classes and indistinguishability between classes. High-quality features improve the accuracy of winter wheat results under different terrains and different maturity levels; further, they improve the edge fineness segmentation results, making the results more accurate and realistic. Among all models, EFNet obtained the best segmentation results. The results show that the edge feature branch can improve the boundary accuracy and reduce the semantic ambiguity between different regions of the same category.

Through statistical analysis of the experimental results, we can intuitively observe that the four evaluation indicators of the three groups of improved models were superior to those of the original model (Table 3). The recall rate and IOU improved, F-scores improved to a certain extent, and the accuracy rate improved slightly. Compared with the other methods, UNet, ERFNet, SegNet, ERFNet-edge, and UNet-edge, EFNet achieved the best performance in terms of recall, IoU, and F-score. In addition, EFNet's IOU and F-score were significantly higher than those of SegNet, by 4.07% and 2.75%, respectively. The recall of the three models when combined with edge features increased by 1.83%, 2.81%, and 5.58%, respectively. This means that the edge features are effective in network improvement.

As shown in Fig. 9, the spatial distribution information of winter wheat planting in the whole of Feicheng City was obtained using the trained EFNet model. The extraction results of the whole city reached a very high precision.

4 Discussion

4.1 *Importance of Edge Features for CNN Models*

Traditional CNN models generally use high-level semantic features extracted by multiple feature extraction units for semantic segmentation; however, such features cause a loss of accuracy owing to upsampling. Because the CNN uses image blocks as the basic unit when extracting features, as the receptive field continues to expand, the pixel inconsistency in the image blocks increases, and the inconsistency of the extracted semantic features further increases. Therefore, other information needs to be introduced to improve the consistency of semantic features. In the high-level semantic feature map, different types of pixels in the pixel sensing field at the edges of the image blocks are mixed. After multiple convolutions, because of being affected by the feature values of different types of surrounding pixels, the feature values of the edge pixels are considerably different from those of the pixels of the same category located inside the image block. Therefore, the difficulty of the model segmentation of the edge pixel is high and the

Table 3 Comparison of our technique with network extraction techniques. The best result for each metric is highlighted in boldface.

Model name	Recall (%)	Precision (%)	IOU (%)	F-score (%)
ERFNet-edge	88.27	89.61	79.40	86.96
ERFNet	86.44	89.45	78.49	86.24
UNet-edge	84.04	**90.37**	77.30	85.79
UNet	81.23	88.92	73.51	82.55
SegNet	85.43	88.81	77.32	85.84
EFNet	**91.01**	88.44	**81.39**	**88.59**

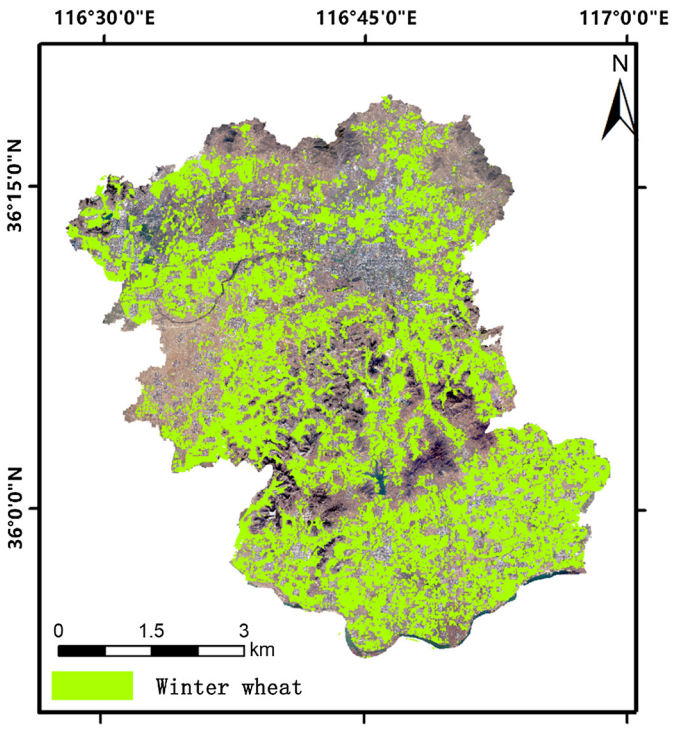

Fig. 9 Extraction results of winter wheat from Feicheng City.

segmentation effect is poor. The edge branch adopts a hierarchical extraction strategy and uses a convolution operation to extract the feature information of the object edge. Finally, the model uses a hierarchical fusion strategy to fuse multi-level edge features into the final feature map, which improves the robustness of features and the consistency of features of similar objects.

In the early experiments, the features extracted from the edge map were directly applied to the encoder part. The multi-scale edge features extracted from the edge graph were integrated with semantic features by adding, multiplying, or concatenating operations. New features replaced low-level semantic features in subsequent feature extraction. After many experiments and a thorough analysis of the results, the results of these methods are not significantly better than those of the original models. Since the value of non-edge pixels in the edge map is 0, the feature map obtained after multiple convolutions loses a great deal of feature information.

The CNN framework with improved edge features (Fig. 5) used low-level semantic feature maps and feature maps extracted from edge maps to obtain an edge feature map. The edge features obtained using this method retained the differences between the pixels of the original edge feature map and supplemented the pixel feature values at non-edge points. Enlarging the feature difference between edge pixels and internal pixels caused the feature values of pixels of the same category to be closer so that the model more accurately distinguished pixels of different categories. In the shared decoder, the multi-scale edge feature maps are connected with different levels of deep semantic features. The connected feature map is upsampled and input to the continuous three-layer convolutional layer, and the feature value is adjusted to fully integrate the two features. Completely fusing the two features using this method compensated for the loss of accuracy of pooling, enlarged the differences between the feature values of different category pixels at the edge, and reduced the differences in feature values between pixels within the same category. EFNet introduced edge information as a priori information and further extracted edge features to guide semantic segmentation. Therefore, introducing edge features into the CNN model can improve the network structure and the model segmentation accuracy.

4.2 *Comparison of Edge Features and Low-level Semantic Features*

The pixel values of the edge pixels in the edge image were the same as the corresponding pixel values in the original image; however, the pixel values of the non-edge pixels in the edge image

were 0. This resulted in a large difference between the feature values extracted from the edge pixels and internal pixels. In the encoder part, the same feature extraction unit was used for feature extraction of two types of inputs. The same convolution kernel led to similarities between the two types of feature maps; however, the semantic feature map displayed all the local information of the pixels in the original image, and the edge feature focused on the edge pixels in the edge image. The semantic feature information was richer and more suitable for further extracting deep features. The edge pixels were susceptible to several different types of nearby pixels; hence, the feature value exhibited a large difference within the same type. Compared with the semantic features of the same level, the edge feature in the edge branch widened the gap between the feature value of the edge pixel and those of the surrounding pixels. With a deepening of the network level and the addition of semantic features, the internal pixel features of the same category differ from those of different categories. When the receptive field of pixels in the edge features is large, the eigenvalues are affected by the edges of other categories. Hence, edge features are not suitable for networks with several layers.

As shown in Fig. 10, compared with semantic features of the same level, the feature value difference within the winter wheat plots in the edge feature map is small and is considerably different from other surrounding features. Roads, bare land, and other crops can also be effectively distinguished, and edge features in hilly areas are also more distinguishable than low-level semantic features. The edge features have strong positioning details and distinguishing edges.

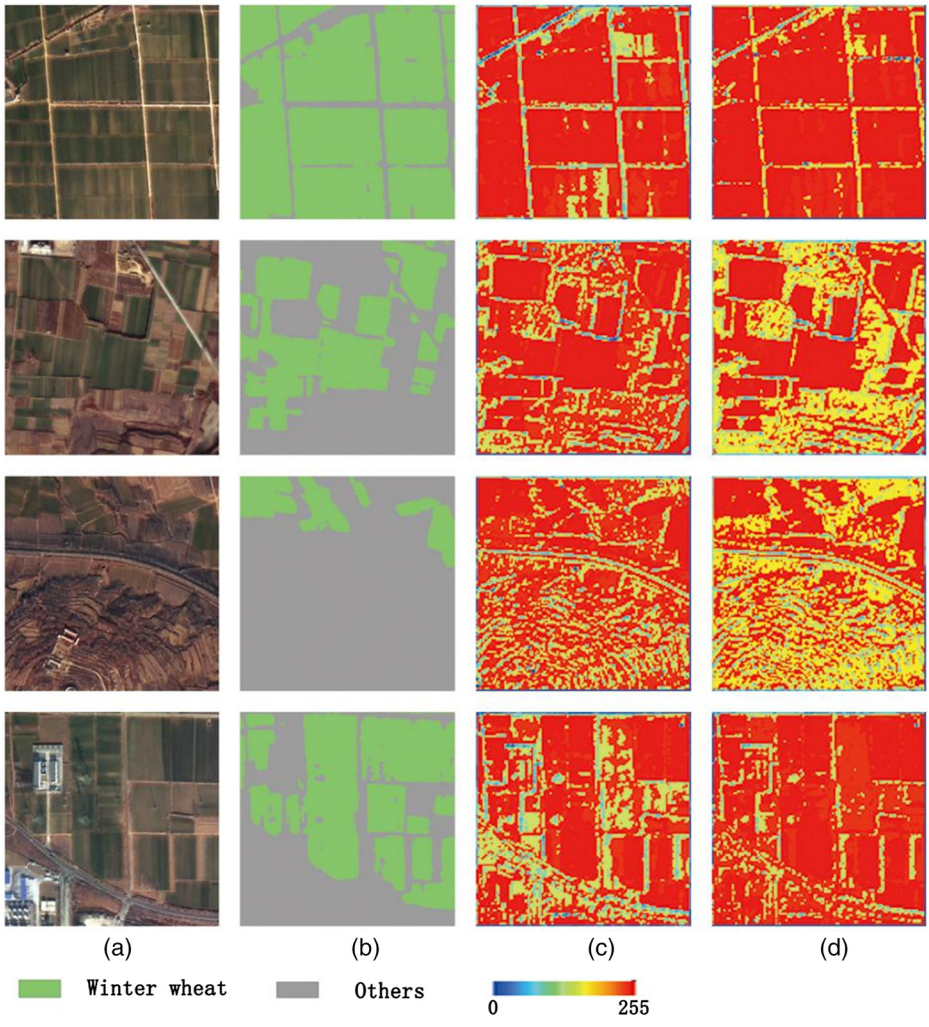

Fig. 10 Comparison of the feature maps for four randomly selected image blocks: (a) image block; (b) manually labeled image corresponding to (a); (c) semantic feature map of fine feature network (EFNet) after two feature extraction units; (d) edge feature map of EFNet after two feature extraction units.

As shown in Fig. 8 and Table 3, EFNet (which introduces edge features in the decoder) is superior to edge-UNet (which introduces both edge features and low-level semantic features), UNet (which introduces low-level semantic features), and SegNet (which does not use other features to compensate for deep semantic features).

In the extraction process, edge features that combine low-level semantic features have the advantages of two features. The number of feature maps is only half of the sum of the two types of feature maps, which reduces the computational complexity and avoids feature redundancy. The CNN of the encoder–decoder structure can effectively improve the accuracy of semantic segmentation by optimizing the network with improved edge features.

4.3 *Ablation Study*

To verify the influence of edge features on the model completely, four sets of ablation experiments were performed with 1–4 sets of edge features and compared with the SegNet model without edge features and EFNet with five sets of edge features. As shown in Table 4, we calculated four evaluation indicators of six modeling results. When an edge feature unit was added, the recall increased from 85.43% to 88.74%. When two edge feature units were added, the recall increased from 85.43% to 90%, the IOU increased from 77.32% to 79.60%, and the F-score increased from 85.84% to 87.38%. With an increase of three edge feature units, the precision continually rose and gradually increased beyond the value of the original model. When the fifth edge feature unit was added, the precision dropped by 2.04%.

As shown in Fig. 11, with the introduction of edge features, various roads between fields can be effectively distinguished and the edges of the segmentation results become smoother. Some unrecognized winter wheat plots were successfully identified, and some roads between plots with complex terrain were distinguished. The experiment also confirmed that adding two or more edge feature units can greatly improve the accuracy of semantic segmentation.

5 Conclusions

The edge feature branch proposed in this research uses an edge detection map combined with semantic features to extract improved edge features. The EFNet model was proposed based on this methodology. EFNet, which combines the advantages of semantic features and edge features to improve the model structure, is an effective method to improve model accuracy.

The main contributions of this research are as follows:

A semantic segmentation dataset for winter wheat in Feicheng City, Shandong Province, China was established. The original images were used to assign values to the edge pixels in the edge detection image and used as part of the semantic segmentation dataset for experiments.

A CNN framework with improved edge features was proposed. Combining low-level semantic features and edge features to compensate for high-level semantic features improves the accuracy of the model. The framework was validated using multiple classic models.

Table 4 Comparison of the results of six groups of ablation experiments. The best result of each metric is highlighted in boldface.

Model name	Recall (%)	Precision (%)	IOU (%)	F-score (%)
SegNet	85.43	88.81	77.32	85.84
SegNet-1edge	88.74	85.25	77.26	85.81
SegNet-2edge	90.00	87.17	79.60	87.38
SegNet-3edge	89.30	89.73	80.98	88.25
SegNet-4edge	89.63	**90.48**	81.04	88.38
EFNet	**91.01**	88.44	**81.39**	**88.59**

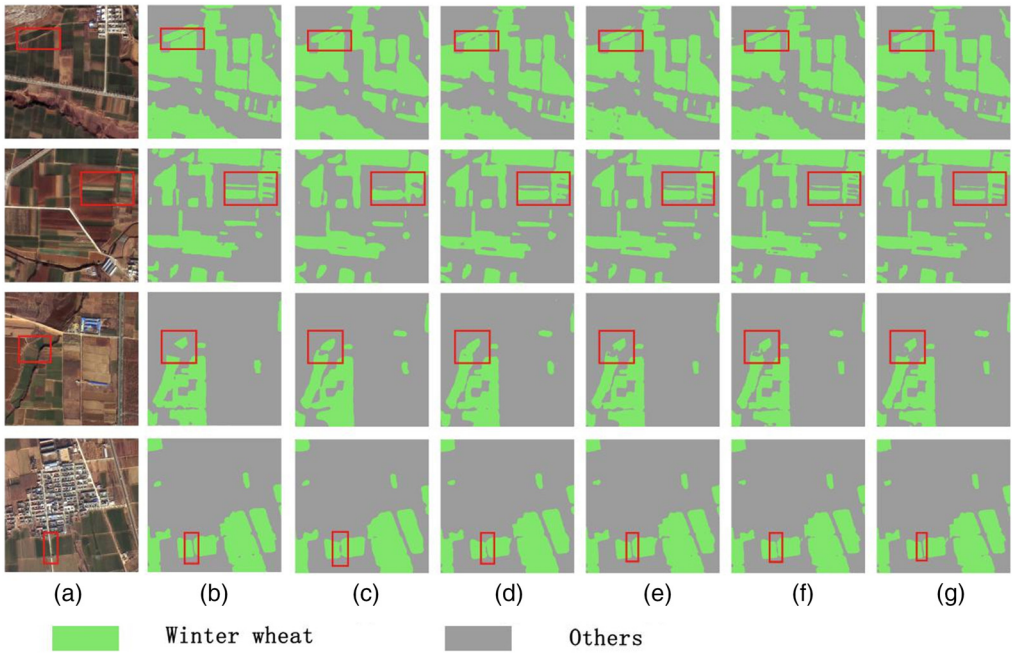

Fig. 11 Segmentation result of ablation experiments: (a) image block; (b) manually labeled image corresponding to (a); (c) SegNet-1edge; (d) SegNet-2edge; (e) SegNet-3edge; (f) SegNet-4edge; (g) EFNet.

The proposed EFNet model improved the semantic segmentation using edge features for extracting winter wheat from remote sensing images.

EFNet is more suitable for high-resolution images with more information. This experiment was intended to improve the features' inter-class distinction and intra-class consistency, and is also suitable for low-resolution remote sensing image segmentation. Furthermore, it is possible to adjust and build a new network using the edge CNN framework based on the experimental images and the planting characteristics of the extracted crops. The proposed EFNET requires pixel-by-pixel label images to construct a training dataset and thus the task of visual interpretation creates a large workload. The purpose of semantic segmentation of supervised CNN model based on labeled graph is to free human beings from tedious and repetitive tasks. A small number of error flags will not affect the results. In the final result, the segmentation precision of the edge is as close as possible to the precision of the mark graph. In subsequent research, we intend to introduce a semi-supervised method in the training process to reduce EFNet's dependence on label maps.

Acknowledgments

This research was funded by the Key Research and Development Program of Ningxia (Grant No. 2019BEH03008); the Applied Foundation of Qinghai (Grant No. 2021-ZJ-739); the Science Foundation of Shandong (Grant No. ZR2020MF130); the arid meteorological science research fund project by the Key Open Laboratory of Arid Climate Change and Disaster Reduction of CMA (Grant No. IAM201801). We thank the Supercomputing Center in Shandong Agricultural University for technical support.

References

1. R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS J. Photogramm. Remote Sens.* **145**, 60–77 (2018).

2. L. Liu et al., "Deep learning for generic object detection: a survey," *Int. J. Comput. Vision* **128**(2), 261–318 (2020).
3. F. Waldner, G. S. Canto, and P. Defourny, "Automated annual cropland mapping using knowledge-based temporal features," *ISPRS J. Photogramm. Remote Sens.* **110**, 1–13 (2015).
4. G. Azzari, M. Jain, and D. B. Lobell, "Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries," *Remote Sens. Environ.* **202**, 129–141 (2017).
5. L. See et al., "Improved global cropland data as an essential ingredient for food security," *Global Food Secur.* **4**, 37–45 (2015).
6. A. R. Phalke et al., "Mapping croplands of Europe, Middle East, Russia, and Central Asia using Landsat, Random Forest, and Google Earth Engine," *ISPRS J. Photogramm. Remote Sens.* **167**, 104–122 (2020).
7. S. Liang et al., "Effects of winter snow cover on spring soil moisture based on remote sensing data product over farmland in northeast China," *Remote Sens.* **12**(17), 2716 (2020).
8. W. G. Alemu and G. M. Henebry, "Land surface phenology and seasonality using cool earthlight in croplands of eastern Africa and the linkages to crop production," *Remote Sens.* **9**(9), 914 (2017).
9. H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," *Comput. Vision Image Understanding* **113**(3), 345–352 (2009).
10. R. Kadota et al., "Hardware architecture for HOG feature extraction," in *Proc. Int. Conf. Intell. Inf. Hiding and Multimedia Signal Process.*, Kyoto, Japan, pp. 1330–1333 (2009).
11. M.-C. Chuang et al., "Tracking live fish from low-contrast and low-frame-rate stereo videos," *IEEE Trans. Circuits Syst. Video Technol.* **25**(1), 167–179 (2015).
12. Y. Li et al., "Fast and accuracy extraction of infrared target based on Markov random field," *Signal Process.* **91**(5), 1216–1223 (2011).
13. L. Zhang and Q. Ji, "A Bayesian network model for automatic and interactive image segmentation," *IEEE Trans. Image Process.* **20**(9), 2582–2593 (2011).
14. J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE Trans. Biomed. Eng.* **64**(1), 16–27 (2017).
15. S. T. Yekeen and A. L. Balogun, "Advances in remote sensing technology, machine learning and deep learning for marine oil spill detection, prediction and vulnerability assessment," *Remote Sens.* **12**(20), 3416 (2020).
16. D. Zhang et al., "A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution," *Remote Sens. Environ.* **247**, 111912 (2020).
17. X. Han et al., "Pre-trained AlexNet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification," *Remote Sens.* **9**(8), 848 (2017).
18. F. Löw et al., "Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines," *ISPRS J. Photogramm. Remote Sens.* **85**, 102–119 (2013).
19. F. Löw, C. Conrad, and U. Michel, "Decision fusion and non-parametric classifiers for land use mapping using multi-temporal RapidEye data," *ISPRS J. Photogramm. Remote Sens.* **108**, 191–204 (2015).
20. J. Xia, N. Yokoya, and A. Iwasaki, "Hyperspectral image classification with canonical correlation forests," *IEEE Trans. Geosci. Remote Sens.* **55**(1), 421–431 (2017).
21. H. Lee et al., "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. 26th Ann. Int. Conf. Mach. Learn.*, pp. 609–616 (2009).
22. Y. Deng, R. Chen, and C. Wu, "Examining the deep belief network for subpixel unmixing with medium spatial resolution multispectral imagery in urban environments," *Remote Sens.* **11**(13), 1566 (2019).
23. Z. Zhao et al., "The generalized gamma-DBN for high-resolution SAR image classification," *Remote Sens.* **10**(6), 878 (2018).

24. A. J. X. Guo and F. Zhu, "A CNN-based spatial feature fusion algorithm for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.* **57**(9), 7170–7181 (2019).
25. S. Akodad et al., "Ensemble learning approaches based on covariance pooling of CNN features for high resolution remote sensing scene classification," *Remote Sens.* **12**(20), 3292 (2020).
26. Z. Li et al., "Deep multiple instance convolutional neural networks for learning robust scene representations," *IEEE Trans. Geosci. Remote Sens.* **58**(5), 3685–3702 (2020).
27. M. Aspri, G. Tsagkatakis, and P. Tsakalides, "Distributed training and inference of deep learning models for multi-modal land cover classification," *Remote Sens.* **12**(17), 2670 (2020).
28. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Boston, MA, pp. 3431–3440 (2015).
29. X. Wang et al., "Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches," *Remote Sens.* **12**(11), 1744 (2020).
30. F. Li et al., "Improved winter wheat spatial distribution extraction from high-resolution remote sensing imagery using semantic features and statistical analysis," *Remote Sens.* **12**(3), 538 (2020).
31. L. E. Falqueto et al., "Oil rig recognition using convolutional neural network on Sentinel-1 SAR images," IEEE Geosci. *Remote Sens. Lett.* **16**(8), 1329–1333 (2019).
32. L. Ding, J. Zhang, and L. Bruzzone, "Semantic segmentation of large-size VHR remote sensing images using a two-stage multiscale training architecture," *IEEE Trans. Geosci. Remote Sens.* **58**, 5367–5376 (2020).
33. M. Freudenberg et al., "Large scale palm tree detection in high resolution satellite images using U-Net," *Remote Sens.* **11**(3), 312 (2019).
34. S. Wei et al., "Multi-temporal SAR data large-scale crop mapping based on U-Net model," *Remote Sens.* **11**(1), 68–85 (2019).
35. Q. Yang et al., "Mapping plastic mulched farmland for high resolution images of unmanned aerial vehicle using deep semantic segmentation," *Remote Sens.* **11**, 2008 (2019).
36. Z. Du et al., "Smallholder crop area mapped with a semantic segmentation deep learning method," *Remote Sens.* **11**(7), 888 (2019).
37. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
38. V. Badrinarayanan et al., "SegNet: a deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).
39. L.-C. Chen et al., "Encoder–decoder with atrous separable convolution for semantic image segmentation," *Lect. Notes Comput. Sci.* **11211**, 801–818 (2018).
40. G. Lin et al., "RefineNet: multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Honolulu, HI, pp. 1925–1934 (2017).
41. E. Romera et al., "ERFNet: Efficient residual factorized convnet for real-time semantic segmentation," *IEEE Trans. Intell. Transp. Syst.* **19**(1), 263–272 (2018).
42. X. Wang et al., "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, pp. 7794–7803 (2018).
43. Z. Huang et al., "CCNET: Criss-cross attention for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 603–612 (2019).
44. J. Fu et al., "Dual attention network for scene segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, pp. 3146–3154 (2019).
45. M. Mittal et al., "An efficient edge detection approach to provide better edge connectivity for image analysis," *IEEE Access* **7**, 33240–33255 (2019).
46. P. Melin et al., "Edge-detection method for image processing based on generalized type-2 fuzzy logic," *IEEE Trans. Fuzzy Syst.* **22**(6), 1515–1525 (2014).
47. W. Rong et al., "An improved CANNY edge detection algorithm," in *IEEE Int. Conf. Mech. and Autom.*, pp. 577–582 (2014).

48. D. Marmanis et al., "Classification with an edge: improving semantic image segmentation with boundary detection," *ISPRS J. Photogramm. Remote Sens.* **135**, 158–172 (2018).
49. F. A. Pellegrino, W. Vanzella, and V. Torre, "Edge detection revisited," *IEEE Trans. Syst. Man Cybern.* **34**(3), 1500–1518 (2004).
50. C. Yang, "Semantic boundary refinement by joint inference from edges and regions," in *Proc. IEEE Int. Conf. Image Process.*, pp. 3105–3109 (2017).
51. X. Han et al., "The edge-preservation multi-classifier relearning framework for the classification of high-resolution remotely sensed imagery," *ISPRS J. Photogramm. Remote Sens.* **138**, 57–73 (2018).
52. S. Liu et al., "ERN: edge loss reinforced semantic segmentation network for remote sensing images," *Remote Sens.* **10**(9), 1339–1361 (2018).
53. Y. Liu et al., "Multiscale U-shaped CNN building instance extraction framework with edge constraint for high-spatial-resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.* 1–15 (2020).
54. L.-C. Chen et al., "Semantic image segmentation with task-specific edge detection using CNNs and a discriminatively trained domain transform," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Las Vegas, NV, pp. 4545–4554 (2016).
55. C. He et al., "Remote sensing image semantic segmentation based on edge information guidance," *Remote Sens.* **12**(9), 1501 (2020).

Hao Yin received his master's degree in information science and engineering from Shandong Agricultural University. His main research interests include land use monitoring and image segmentation. Currently, he is mainly engaged in the research of remote sensing technology for agriculture and environment.

Chengming Zhang is currently working as a professor at the College of Information Science and Engineering, Shandong Agricultural University. His main research areas include remote sensing and geographic information system in land use monitoring and evaluation. He presided over a number of agricultural remote sensing projects by Ministry of Science and Technology and Shandong Province. Currently, he is mainly engaged in the research of remote sensing technology in agriculture and environment.

Biographies of the other authors are not available.