A Semantic Segmentation Algorithm for Water Plants Distribution Detection

Wangyi Wang[1,a], Jiaoyan Ai[1,b,*], Jianwu Zheng[1,c] and Dongyi Yao[1,d]

[1]*School of electrical engineering, Guangxi university, Nanning, China*
a.1454811452@qq.com, b. aijy@gxu.edu.cn
c. 1149569930@qq.com, d. 1179150880@qq.com
**corresponding author: Jiaoyan Ai*

Keywords: Dataset, density detection, semantic segmentation, deep learning.

Abstract: The planting density of water plants in landscape lake is very important to the water quality restoration. In order to accurately divide the area without water plants, this paper proposes a novel deep learning algorithm to explain the distribution of water plants, and the method based on the improved FC - DenseNet network structure, its structure is mainly consists of five encoding networks and the five decoding networks. The encoding network and the decoding network are connected together by concat layer, the last layer of the network is a pixel-level classifier. The encoding network structure is improved on the basis of DenseNet network, and the decoding network restores the image to the original resolution through corresponding up-sampling, so as to obtain a more accurate segmentation effect. Compared with FCN and other seven semantic segmentation network models, the results show that this research method has better segmentation performance and can accurately segment the sediment area without water plants.

1. Introduction

The eutrophication of water in landscape lake is becoming more and more serious, which seriously affects the city appearance and the living quality of residents.Aquatic animal and plant restoration technology refers to the cultivation of aquatic plants such as water plants in the landscape lakes, or the breeding of fish, shrimp and other animals.Water plants can use their own absorption, adsorption and metabolic functions, so as to decompose and absorb nitrogen and phosphorus and other nutrient-rich substances in the water.Too sparse planting density of water plants will increase inter-species competition and foraging, which will affect the growth and recovery of water plants. Therefore, reasonable planting density has a significant impact on the restoration effect of landscape lakes.In this paper, a semantic segmentation algorithm is proposed to solve the problem of density recognition.

Long et al. proposed the FCN[1] model to replace the fully connected layer of VGG16[2] with the fully convolution layer and classify the image at the pixel level, thus solving the problem of image segmentation at the semantic level.FCN can accept any size of the input image, use the deconvolution operation to upsample the last layer of the convolutional layer, make it back to the input image of the same size, which can have a prediction on each pixel, FCN achieves the state of - the - art segmentation of PASCAL VOC (20% relative improvement scheme to 62.2% IU on

2012).Vijay et al. proposed SegNet[3] network on the basis of FCN, and realized end-to-end pixel-level semantic segmentation by using a symmetric structure of encoder-decoder. FCN carried out up-sampling through deconconvolution operation, while SegNet achieved better results than FCN through feature fusion in the previous layer.The PSPNet[4] proposed by zhao et al. extends pixel-level features to pyramid pooling, and uses local and global information fusion operations to achieve an 80.2% IOU value on Cityscapes dataset.The DeepLabV3+[5] network proposed by Chen et al. Used depthwise seperable convolution and atrous spatial pyramid pooling module to achieve 82.1% MIOU value on Cityscapes dataset.The MobileNetV2[6] proposed by Mark et al. Uses a lightweight convolutional neural network to greatly reduce the parameters of the model, and reduces the computation amount of convolution by using deep separable convolution, which can be used in mobile devices.Over YOLO2[7] on the COCO dataset.All the above models have achieved good results on large datasets such as COCO or ImageNet, but so far, no model has been found that applies semantic segmentation algorithm to density recognition of water plants.The model in this paper introduces the basic structure of DenseNet[8] network and improves the FC-DenseNet[9] model on the self-made dataset.

In order to analyze the density distribution of water plants, a small data set on the density distribution of water plants was produced. The data set mainly contains sediment and water plants, so this is a two-category classification task. When labeling, only the area that does not contain water plants is bounded by a bounding box, and then the data set is used to transfer learning through 8 semantic segmentation models. And from the perspective of precision, recall, F1 score, IOU (Intersection over union) values, the performance of these 8 models on the data set is analyzed.

2. Related Work

Since the existing data sets are focused on the semantic segmentation of common objects, including road detection, medical image detection and other fields, they cannot be directly used for the training of water plants density semantic segmentation data sets. In order to realize the function of guiding the machine to automatically plant water plants, it is necessary to make a semantic segmentation data set dedicated to annotating the distribution of water plants. This section introduces the process and statistical data for making semantic segmentation water plants datasets.

The experimental images were collected from Jinghu and Biyun Lake of Guangxi University. Using the high-definition camera mounted on the unmanned ship, several images of water plants of different growth conditions are randomly collected in a natural environment, and some of the blurred images, images with relatively large repetitiveness, and images with large effects of light and shadow are removed, and the remaining 480 images The size is set to 960 × 720 pixels, unified in png format, and then the labeling tool labelme is used for labeling. The bottom mud area of each image is marked red, and the water plants area except the bottom mud is the background by default and black.

Image enhancement technology operates on the image by translating, flipping, rotating, scaling, and changing the brightness of the image data to generate new pictures to participate in training. Overfitting phenomenon. For the binary classification task, 480 images are enough, so the algorithm in this paper performs random image enhancement on the input image, that is, an image enhancement operation is performed before each input image is trained. This operation depends on a random value. If the random value is 1, the image enhancement operation is performed, otherwise the operation is not performed. The enhancement methods mainly include horizontal random flip, vertical random flip, clockwise rotation of 0-30 degrees, counterclockwise rotation of 0-30 degrees, brightness change range of 0.9-1.0. Some experimental images are shown in Figure 1.

Figure 1: water plant data set for density segmentation.

The 480 pre-processed images and the corresponding 480 labeled images are stored in three folders, which are 270 training sets, 90 verification sets and 180 test sets. The training set is sent to the model for training. The initialization parameters of the low-level feature extraction network are pre-trained parameters. Each input image undergoes random image enhancement to further train the model parameters; afterwards, the validation set image is used to train The parameters after the training set training are tuned; finally, the parameters after the training are loaded into the test module to evaluate the model with the test set images. The smaller the image resolution, the faster the training speed. Therefore, the appropriate adjustment of the picture size is helpful to improve the model segmentation efficiency. Therefore, all pictures are uniformly cropped to 512 × 512 pixels before training.

The target categories of this dataset are divided into two categories, namely background (water grass) and sediment . Use labelme as a labeling tool, a total of 480 pictures. Table 1 shows the number of pictures and the number of marked frames containing sediment. Figure 2 is an example of the annotation of the data set.

Table 1: Dataset label Quantity Statistic.

	Image numbers	Bounding box
Train set	270	2693
Test set	90	901
Validation set	120	1137

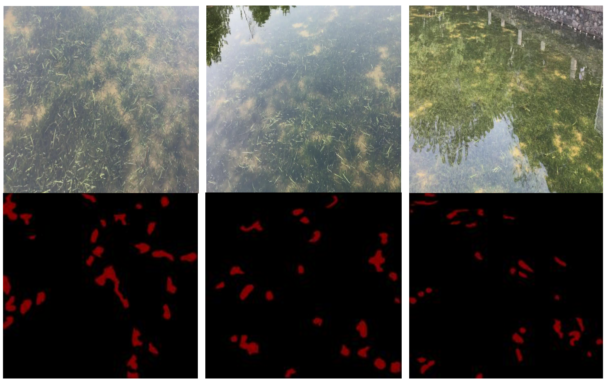


Figure 2: Annotated image in semantic segmentation data set of grass density.

3. Improved Fc-densenet Semantic Segmentaion Algorithm Analysis

In this paper, a fully convolutional neural network is used as the basic network for water plants density segmentation, and low-level image feature extraction is performed through a pre-trained feature extraction network. Then, after training with multiple water plants density images, the features of the sediment area are learned to achieve end-to-end image semantic segmentation.

The original image is input into the convolutional neural network model for convolution operation to obtain the initial feature map. The first layer of convolutional layer uses three feature extraction networks to perform comparative experiments for low-level feature extraction, which are ResNet50, ResNet101, and MobileNetV2. Comparative experiments with different models equipped with different feature extraction networks have found that ResNet50 achieves better results than the other two. This kind of network is better, so the final data comparison is to use ResNet50 as a low-level feature extraction network. The first convolutional layer contains 48 convolution kernels of size 3 × 3. The convolutional layer contains three basic operations, namely Batch Normalization, ReLuctant and Convolution operations. The batch normaliazation layer is after each convolution re-adjust the data to a standard Gaussian distribution, which can speed up the training and effectively suppress the gradient disappearance and gradient explosion problems. In addition, the parameter adjustment process also becomes simple, the requirements for initialization become lower, and can be used Larger learning rate. The ReLu layer is a nonlinear activation function. Compared to sigmoid and tanh, its derivative is easier to solve, and it can increase the nonlinear capability of the network to prevent the gradient disappearance problem when the neural network is trained to the deep layer. The expression is f (x) = max (0, x). The convolution kernel used in this paper except the last layer, the size of all other convolution kernels is 3 × 3, and the output tensor of the first layer is (1,512,512,48). Expression (1) is the calculation formula of the convolutional layer:

$$X_i^k = f(\sum_{i\in\delta_j}^{M} W_i^k \mathrm{X}_i^{k-1} + b_i^k) \tag{1}$$

In the above formula, f represents the ReLu activation function, the output of the k-1 hidden layer is, the input image is the mapping weight matrix of the k hidden layer, and the offset matrix of the k hidden layer .

The 48 low-level feature maps obtained after the first layer of convolution enter the Dense Block module. The algorithm in this paper contains a total of 11 Dense Blocks, of which 5 Blocks are connected to the down sampling layer (Transition Down) and 5 Blocks are connected to the up sampling. The layer (Transition Up) is restored to the original resolution. Each block contains 4 convolutional layers, and the growth rate of each convolutional layer is set to 12, which means that the size of the convolution kernel is 12. The same operation included in each convolutional layer is BN + ReLu + Conv. Figure 3 shows the internal connection mechanism of Block.

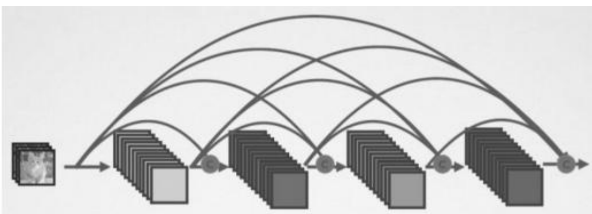


Figure 3: Dense connection mechanism inside a Dense Block.

Each layer inside the Dense Block is concatenated with the previous layer in the Channel dimension (Concat) and serves as an input for the next layer. For an L layer network, Each Dense Block contains $\frac{L(L+1)}{2}$ connections.The Transition Down and Transition Up operations are connected between Dense blocks.The operation of Transition Down adopts Pooling to reduce the feature graph, the Pooling layer adopts the maximum Pooling to conduct the Down sampling operation, and the Pooling operation is conducted every time， The original $M \times \mathrm{N}$ image is now $\frac{1}{2}M \times \frac{1}{2}N$. The Transition Up layer uses deconvolution to restore the image to its original resolution. The last convolution layer adopts the convolution kernel of 1×1. The sampling of 1×1 convolution can compress the number of channels and enhance the nonlinear expression ability of the network. The whole flow chart is shown in Figure 4:

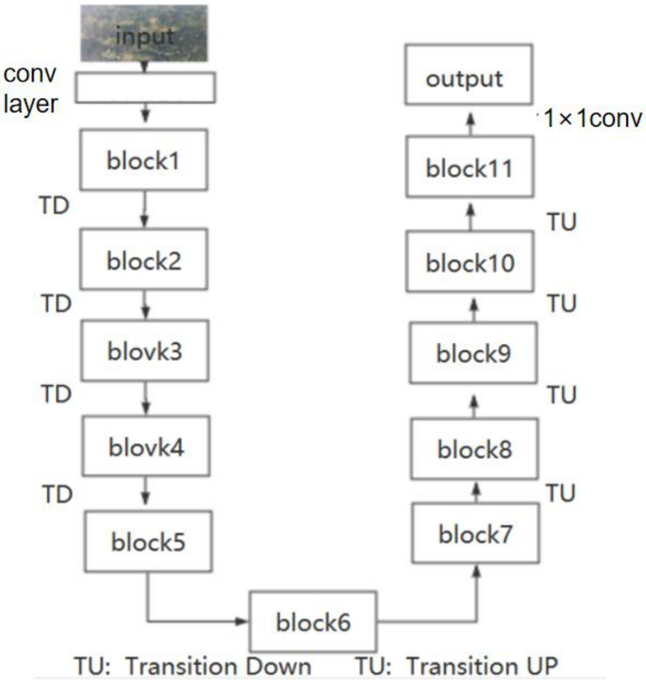


Figure 4: Based on the improved fc-densenet network structure.

After extracting the features of the image through the deep convolutional neural network, it is sent to the classification layer for classification. In this paper, Softmax is used to classify each pixel. The calculation formula is as follows:

$$P(S_i) = \frac{e^{g_i}}{\sum_k^n e^{g^k}} \tag{2}$$

Set a total of n numerical classification for S_k , k ϵ (0, n), where n said the number of classification, a classification of k, I said g_i values of the classification.After the classification of pixels, the loss function is used to evaluate the training model. The final output vector of softmax [Y1,Y2] and the real value of the sample are used to calculate the cross entropy loss. The formula is as follows:

$$H_{y'}(y) = -\sum_i y_i' \log(y_i) \tag{3}$$

Where y_i' represents the value of the ith in the actual label of y, and y_i is the output vector of softmax [y1,y2...].Obviously, the more accurate the prediction is, the smaller the result value will be. Finally, the final loss value will be calculated by using the summation average.RMSProp algorithm is selected to optimize the loss function, which controls the amount of historical information acquisition by adding a attenuation coefficient. The global learning rate is set by dividing the global learning rate by the square root of the square sum of the historical gradient squares controlled by the attenuation coefficient, so the learning rate of each parameter is different. In the more gentle parameter interval gradient update faster, and suppress the steep parameter interval, make it more gentle, speed up the neural network training speed. Figure 5 shows the relationship between the loss function and the number of iterations.

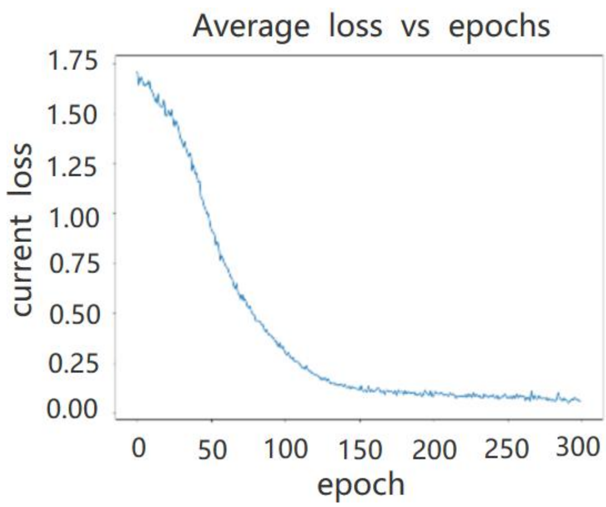


Figure 5: Relation between the loss function and the number of iterations.

After about 150 iterations, the loss function is almost minimized.

4. Training of Fully Convoluted Densenet Neural Network Model Using

4.1.Experimental Environment and Parameter Setting

Based on the improved fc-densenet model is carried out on the win10 operating system, using the Tensorflow framework as the back-end, Python as the programming language, graphics card for NVIDIA GTX1080 Ti 11 GB, equipped with Intel Core i7 processor.

In this paper, RMSProp algorithm is used to optimize the network model.Epoch was set to 300 and Batch Size was set to 1. Each incoming image was resized to 512×512 and random image enhancement was performed on it.

4.2.Evaluation Indicators Used to Analyze Different Models

Before the comparison of the model, it was found through the comparison experiment that when the ResNet50 network was used as the feature extraction network of the first convolution layer, the model had better effect.Next, the model is compared with DeepLabV3, DeepLabV3+, FCN, PSPNet, MobileNetV2, SegNet, GCN and other models. The test set contains 120 pictures. The mathematical formulas are as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$recall = \frac{TP}{TP + FN} \tag{5}$$

$$f1 = \left(\frac{1}{n} \sum \frac{2 \cdot precision \cdot recall}{precision + recall} \right)^2 \tag{6}$$

$$IOU = \frac{TP}{TP + FP + FN} \tag{7}$$

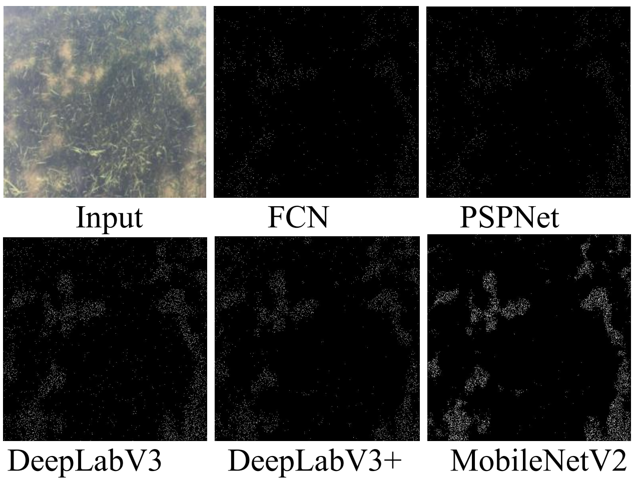
TP indicates that sediment classification is correct;TN means the classification of water plants is correct;FP means that water plants is classified as sediment.FN indicates that the sediment is classified as water plants.Precision represents Precision, and its mathematical expression is:

$$precision = \frac{TP}{TP + FP} \tag{8}$$

F1 Score combines the results of accuracy and recall rate. When F1 Score is high, it indicates that the model is better.

5. Effect Comparison of Different Segmentation Methods

In the final model, the modified 56-layer fully convoluted DenseNet is used as the network structure to segment the water plants and sediment.In order to evaluate the performance of the model in this paper, the model proposed in this paper is compared with the above seven models. There are 120 images in the test set, and the image enhancement method is the same as the training set.Accuracy, recall rate, f1 score and IOU value were used as the evaluation criteria of segmentation results.



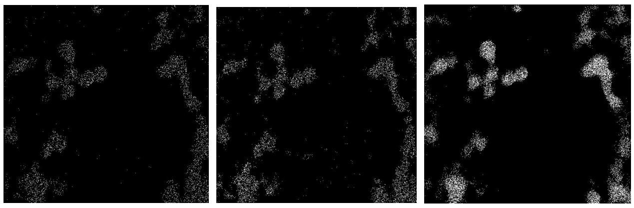


Figure 6: Comparison of experimental results of different models.

The segmentation effect of different networks is shown in Figure 6: the segmentation effect of FCN is not obvious, and a large part of the sediment area is not segmented.PSPNet is slightly clearer than FCN, but too much water plants divides into sediment;The segmentation effect of DeepLabV3 and DeepLabV3_plus is not very different, and the latter has less misclassification than the former.In general, the segmentation effect of MobileNetV2 is very good, with only a few misclassification, but the classification of sediment is not very detailed.The same SegNet did not divide the sediment area in detail, and a small part of the water weed area was misclassified.The GCN(Graph Convolutional Network) is a Graph Convolutional Network. Different from the traditional CNN, the GCN processing data is a Graph Structure, that is, a Non Euclidean Structure, a non-euclidean Structure, and a topological Structure.Such as social network connection, information network and so on.The GCN model used in this paper has a detailed segmentation of sediment area, but some misclassification exists.Finally, the improved fc-densenet algorithm adopted the 56-layer fully convoluted DenseNet network structure, and the parameters such as the number of Dense blocks and the Growth Rate were optimized on the original network model. The algorithm in this paper not only has a more detailed segmentation effect on the sediment region than the above algorithm, but also has few misclassification in the water plants region.The pairs of evaluation indicators are shown in Table 2.

Four evaluation indexes are used to evaluate 8 different models.As can be seen from the above table, the algorithm in this paper achieves the best results in terms of accuracy, recall rate, F1 score and IOU value.

The accuracy on the test set is as high as 84%, and the algorithm in this paper is short in running time, the processing time of an image is 10 seconds.The end - to - end training can be realized. In terms of test effect, the algorithm in this paper has more advantages than other algorithms.

Table 2: Comparison of evaluation indexes of different segmentation methods.

Model	Pre	Recall	F1	IOU
FCN	0.53	0.48	0.49	0.50
PSPNet	0.59	0.54	0.57	0.52
DeepLabV3	0.71	0.69	0.70	0.63
DeepLabV3+	0.71	0.70	0.73	0.64
MobileNetV2	0.75	0.70	0.74	0.68
SegNet	0.79	0.74	0.77	0.73
GCN	0.80	0.75	0.77	0.75
Our Model	0.84	0.79	0.81	0.80

6. Conclusions

In this paper, an improved fc-densenet algorithm is proposed to segment the density of water plants in landscape.Compared with other cutting-edge semantic segmentation algorithms, the proposed algorithm achieves better results.The accuracy on the test set was 84%.Due to the difficulty in labeling the sediment and sediment, and the boundary line between sediment and sediment is not very clear, the accuracy of the model is greatly affected.

The improved 56-layer fully convoluted DenseNet network can segment the area of water plants containing sediment more accurately. Compared with the existing semantic segmentation algorithm model, the number of model parameters used in this paper is smaller and the operation speed is faster.

For more complex situations, the algorithm in this paper can also more accurately divide the sediment area, which greatly exceeds the efficiency of manual segmentation and opens up a new shortcut for the subsequent automatic planting of water plants.

Up to now, the data set of water plants is still an early stage, which not only needs to expand the amount of data, but also needs to increase the diversity of data, such as adding more land lawns, so that the training model has better generalization performance and can be used for a variety of purposes.In the area of algorithm design, the research in this paper shows that different models have different emphasis on the segmentation of water plants, which provides more choices for different USES of the algorithm.

Acknowledgments

This work was financially supported by National Nature Science Foundation of China(61563002) and Guangxi Innovation-driven Development(AA17202032-2).

References

[1] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[2] Simonyan, Karen, Zisserman, Andrew. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer ence, 2014.

[3] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017:1-1.

[4] Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network[J]. 2016.

[5] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.

[6] Mark Sandler, Andrew G. Howard, Menglong Zhu, AndreyZhmoginov, and Liang-Chieh Chen. Mobilenetv2: Invertedresiduals and linear bottlenecks. mobile networks for classifi-cation, detection and segmentation.CoRR, abs/1801.04381,2018.

[7] J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger.InComputer Vision and Pattern Recognition (CVPR), 2017IEEE Conference on, pages 6517–6525. IEEE, 2017.

[8] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. CoRR, abs/1608.06993, 2016.

[9] Jégou, Simon, Drozdzal M, Vazquez D, et al. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation[J]. 2016.