Research on Image Segmentation based on Full Convolutional Neural Network

Yanan Guo, Zengshun Zhao∗ and Zuohong Wub
Shandong University of Science and Technology, China

*Corresponding author e-mail: 15898984079@163.com, ∗1421079118@qq.com,
b2784903942@qq.com

Abstract. Image segmentation technology is an important branch in the field of computer vision. Image segmentation is the basis of image analysis processing, and image segmentation effect directly affects image subsequent processing. Aiming at the problem that the segmentation method based on traditional machine learning is not accurate, the edge information is lost and the robustness needs to be improved, this paper proposes an image segmentation algorithm based on improved full convolution neural network. The algorithm utilizes the better feature extraction ability of the deep learning model and the sensitivity of the cluster segmentation to the edge information, and further assists the segmentation with the Ncut algorithm. The experimental results show that compared with the traditional convolutional neural network image segmentation algorithm, the algorithm finally achieves higher segmentation accuracy. From the overall analysis, the segmentation method proposed in this paper is better than other methods.

1. Introduction
In recent years, the rapid development of image information technology has caused more and more information to be transmitted as images. This requires us to quickly and accurately distinguish the information in the image, but it is inevitable that only one image of the human eye will be retrieved. Omissions, and the efficiency is very low. Therefore, many scholars have developed intelligent algorithms to perform image segmentation, and then extract the parts of the image that are of interest to humans. Image segmentation is based on the grayscale, color, texture, shape and other features of the image. The image is divided into several non-overlapping regions, and these features are similar in the same region, and there are significant differences between different regions. 1]. Areas with unique properties in the segmented image can then be extracted for different studies. In recent years, many methods of image segmentation have been proposed by Chinese and foreign scholars, but the traditional image segmentation technique has lower accuracy. However, the segmentation method based on deep learning extracts the high-level semantic features of the image, which lacks the specific features of the image. The segmentation method combined with deep learning enables the segmentation results to include specific low-level features and high-level semantic features, thereby improving the accuracy of image segmentation.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
2. Related Work

For image segmentation techniques, they have been broadly divided into two major types over the years: segmentation based on traditional methods and segmentation based on deep learning. The traditional segmentation methods are roughly divided into threshold-based segmentation [2, 3, 4], graph theory segmentation [5], cluster segmentation [6, 7], and deep learning segmentation [8, 9, 10, 11]. However, the simple threshold segmentation algorithm is only realized by the gray distribution in the image, and the spatial relationship is not considered. However, the simple graph-based segmentation algorithm has high pixel false positive, and the segmentation effect is high. It is not ideal. It only relies on the clustering algorithm to be insensitive to images with large changes in features. Although the segmentation based on depth learning has higher segmentation accuracy, there is edge information loss. This paper proposes an image segmentation method that combines graph cut, clustering and deep learning, so that the three methods complement each other and improve the segmentation precision.

The algorithm block diagram of this paper is shown in Figure 1:

![Algorithm Block Diagram](image)

Figure 1. Block diagram of the algorithm

2.1. Segmentation based on Graph Theory

The graph-based segmentation method considers each pixel of an image as a node. The relationship between a pixel and a pixel is regarded as an edge, and the weight of the edge represents the similarity of adjacent pixels in terms of gray value, color, or texture. The segmentation process of the image is
the process of clipping the image. Each segment that is segmented corresponds to the subgraph in the image. To achieve the best result of the segmentation, the segmented subgraph must maintain the same similarity internally. The similarity between the subgraph and the subgraph remains minimal. Since each pixel has a weight, the graph theory-based method is not sensitive to the shape of the segmentation target, and plays a better role in image segmentation.

The Normalized cut (Ncut) algorithm used in this paper is a graph theory based segmentation method. It is based on the minimum segmentation algorithm and solves the problem that the minimum segmentation may separate a single vertex far from the large force from other vertices. Mistaken into two types of problems.

Using asso (A, V) to represent the weight of all points in A connected to all points in the image:

\[ N_{cut}(A, B) = \frac{cut(A, B)}{asso(A, V)} + \frac{cut(A, B)}{asso(B, V)} \]  

among them, \( asso(A, V) = \sum_{u \in A, v \in V} \omega(u, v) \); \( cut(A, B) = \sum_{u \in A, v \in B} \omega(u, v) \);

\[ N_{asso}(A, B) = \frac{asso(A, A)}{asso(A, V)} + \frac{asso(B, B)}{asso(B, V)} \]  

Formula (1) (2) is available in conjunction:

\[ N_{cut}(A, B) = 2 - N_{asso}(A, B) \]  

Ncut cuts the graph into two halves by a value smaller than the one cut out. Therefore, the Ncut algorithm can achieve better image segmentation tasks than the minimum segmentation.

2.2. Clustering based Image Segmentation

Clustering-based image segmentation clusters around k points in space and categorizes the objects closest to them. The iterative method is used to update the values of each cluster center one by one until the best clustering result is obtained. Objects of the same class are represented by the same color, objects of different classes are represented by different colors, and finally the segmentation result is obtained.

In this paper, the K-means algorithm is used to cluster the color of the image. The image to be segmented is shown in Figure 2.

![Figure 2. Original image](image)

There are animals and grasses in the image, in which the animals are segmented foreground, and the grass is the background. We use the K-means algorithm to cluster the image colors in 2, as shown in Figure 3:
The final segmentation result is obtained by the K-means algorithm. The biggest advantage of this algorithm is its simplicity and speed. The key to the algorithm is the initial center selection and distance formula. In this paper, the initial center is randomly chosen, and the distance formula is chosen using the Euclidean distance.

2.3. Image segmentation based on deep learning
The image segmentation based on deep learning mainly uses the structure of the deep learning model to extract features. Finally, the pixel-level classification result, that is, the semantic segmentation result of the image is obtained. The deep learning models proposed for image segmentation in recent years are mostly realized by the downsampling-upsampling structure.

The FCN network is modified by the CNN network. It preserves the convolutional layer pooling layer in the CNN structure and discards the fully concatenated layer. Therefore, it can input images of any size and pass specific convolution kernels and steps in the convolution pooling layer. Long convolution continuously extracts features, and the resolution of the image is also reduced. In order to restore the resolution of the image, the extracted feature map is deconvolved, the image is restored to the original size, and the semantic segmentation of the image is completed.

The full convolutional neural network used in this paper is to improve the U-net network. The convolution method of the traditional U-net network is a simple convolution. In this paper, the BN layer will be added in the convolution process [12], and the activation function is composed of the Relu function. Become a Leaky Relu function. The network framework is shown in Figure 4.
The improved U-net network proposed in this paper consists of 9 convolutional blocks, 4 maximal pooling layers and 4 upsampling layers. Since there is no fully connected layer in the network, images of any size can be input.

1) At the input layer, this paper proposes to input a 500-320 3-channel image.
2) In the convolutional block, each convolution block includes a convolutional layer, a bulk regularization layer, and a Leaky Relu activation function. The convolution layer is extracted with a convolution kernel of size 3*3 and step size 1. The batch regularization layer can allow a large learning rate to pass and improve the convergence speed of the training model; and the Leaky Relu activation function is better than the Relu function. The convergence speed is further improved.
3) In the maximal pooling layer, the pooling with a step size of 2*2 is used to reduce the dimension of the convolutional layer output feature map and reduce the unnecessary calculation.
4) In the up sampling layer, the step size is 2*2, and the reduced image resolution after pooling is restored to restore the image size to the original image size.
5) Add a sigmoid activation function to the output layer and the last convolutional layer to output the same probability heat map as the input image.

2.4. Fusion of Segmentation Results of Different Algorithms
The same original image is obtained by Ncut algorithm, K-means algorithm and improved U-net network to obtain three segmentation result images. However, different algorithms obtain different image types. We replace three images into Uint8 type, and then proceed. Binary operation, finally the intersection of the processed three images, and then the conditional random field model optimization, to obtain the final segmentation results.
3. Experimental Process and Results Analysis

In order to verify the validity and accuracy of the image segmentation method proposed in this paper, the VOC2011 data set was selected for experiment. The experiment process was completed in 2.6GHz processor, 6G memory windows7 Ultimate, and the experimental program was run in matlab2014b and python3.6.

3.1. Experiment Procedure

In this paper, the segmentation method we use is a combination of traditional segmentation method and deep learning segmentation method. Firstly, the image to be segmented is normalized by noise, and it is extracted by Ncut algorithm. The -means algorithm uses clustering method to extract low-level features such as pixel color and texture. The FNC network is used to train the VOC dataset, and the same image to be tested is tested to extract the high-level semantic features of the image. Finally, the segmentation results of different algorithms are performed. Fusion, so that the final segmentation results include low-level image colors, textures and other specific features also contain advanced semantic features to make the segmentation results more accurate.

3.2. Experimental results and analysis

The same image is segmented by different algorithms. The segmentation result is shown in Figure 5:

![Segmentation results of different algorithm](image)

**Figure 5.** Segmentation results of different algorithm

In order to objectively evaluate the effect of each algorithm segmentation, this paper chooses Accuracy, Precision, Recall, F1-measure and average absolute error indicators. The calculation formula of each indicator is as follows:

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (5)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (6)
\]

\[
F1\text{-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (7)
\]
\[ MAE(X, h) = \frac{1}{m} \sum_{i=1}^{m} |h(X^{(i)}) - y^{(i)}| \]  

TP indicates that the foreground pixel class is correctly divided into the number of samples of the foreground pixel class, TN indicates that the background pixel class is correctly divided into the number of samples of the background pixel class, and FP indicates that the background pixel class is divided into the samples of the foreground pixel class. The number, FN, represents the number of samples in which the foreground pixel class is divided into background pixel classes. \( m \) is the number of pixels, \( i \) is the \( i \)-th pixel, \( h(X^{(i)}) \) is the segmentation result graph after processing by an algorithm, and \( y^{(i)} \) is the GT graph.

The running results of each algorithm are shown in Table 1:

| Algorithm name | Accuracy | Precision | Recall | F1-measure | CalMAE |
|---------------|----------|-----------|--------|------------|--------|
| Threshold     | 0.54     | 0.49      | 0.61   | 0.51       | 0.39   |
| Morphological | 0.62     | 0.64      | 0.62   | 0.63       | 0.28   |
| Ncut          | 0.84     | 0.83      | 0.84   | 0.83       | 0.05   |
| K-means       | 0.73     | 0.78      | 0.89   | 0.81       | 0.13   |
| Improved U-net| 0.88     | 0.82      | 0.91   | 0.85       | 0.04   |
| Our algorithm | 0.94     | 0.95      | 0.90   | 0.91       | 0.03   |

It can be concluded from Table 1 that the proposed algorithm has strong numerical support in the Accuracy, Precision, F1-measure and CalMAE indicators, and the effect is slightly worse in the Recall index, but the Recall coefficient indicates the recall rate. The low detection rate in the algorithm is due to the high segmentation precision and accurate foreground background segmentation. Although other algorithms such as Ncut have higher Recall values, there are more false positive pixels. The F1-measure coefficient can be seen from the formula. It is a fusion of Recall coefficient and Precision coefficient, which can reflect the segmentation result of images in each algorithm. Therefore, from the overall analysis, the proposed algorithm is superior to other algorithms.

4. Conclusion
The image segmentation method combined with full convolutional neural network, Ncut graph theory segmentation and K-means cluster segmentation proposed in this paper is superior to simple traditional methods or pure deep learning methods in that it uses both deep learning models. The advantage of quickly extracting the high-level semantic features of the image is to take advantage of the K-means clustering method to segment the edge information of the image. At the same time, the Ncut graph method is used to reduce the segmentation error. Finally, the full-join conditional random field optimization is used to further improve the accuracy of the algorithm. Sex. The experimental results show that the proposed algorithm and other algorithms segment the same image, and the segmentation effect is better than other algorithms. The results obtained by the algorithm proposed by this paper will be more conducive to further accurate analysis of the image.

Acknowledgments
This work was financially supported by the fund of the National Natural Science Foundation of China (Grant No. 61403281), it was also funded by the Special Fund for China Postdoctoral Science Foundation (2015T80717).

References
[1] He jun, Ge hong, Wang yufeng. A Survey of Image Segmentation Algorithms [J]. Computer Engineering and Science, 2009, 31 (12):58-61.
[2] Wang qian, Peng zhong, Liu li. An Image Segmentation Algorithm Based on Adaptive Threshold[J]. Journal of Beijing Institute of Technology, 2003 (04):521-524.

[3] Dong zhongyan, Jiang lixing, Wang junya, Xiao gai. One-dimensional Otsu improved algorithm based on image complexity [J]. computer science, 2015 (S1).

[4] Zhang yun, Jiao bin. Two-dimensional maximum entropy threshold segmentation algorithm based on Canny operator [J]. Journal of Shanghai Dianji University, 2018, 21 (03):1-5.

[5] Zhang congcong. Color Image Segmentation Based on Improved Watershed Algorithm and NCut Algorithm [D]. Guangzhou, South China University of Technology, 2016.

[6] Xu xiaoli. Research on Image Segmentation Algorithm Based on Cluster Analysis [D]. Harbin, Harbin Engineering University, 2012.

[7] Li guang, Wang zhaoying, Hou zhijiang. Color image segmentation algorithm based on K-means clustering and region merging [J]. Computer application, 2010, 30 (02): 354-358.

[8] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3431–3440, 2015.

[9] Wang C, Yan X, Smith M, et al. A unified framework for automatic wound segmentation and analysis with deep convolutional neural network [J], Conf Proc IEEE Eng Med Biol Soc, 2015, 2015 (8): 2415-2418.

[10] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [M]// Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Springer International Publishing, 2015: 234-241.

[11] Tan zhiguo, Ou jianping, Zhang jun, Shen xiangeng. Depth image segmentation algorithm based on multi-feature combination [J]. Computer Engineering and Science, 2018, 40 (08): 1429-1434.

[12] Kaiming He, Xiangyun Zhang, Shaoqing Ren, Jian Sun, et al. Deep Residual Learning for Image Recognition [J]. Conference on Computer Vision and Pattern Recognition. In IEEE, 2016.