Overview of Semantic Segmentation Algorithms Based on Threshold Method and Deep Learning

Jucheng Bao[1], Lingxiao Sun[2], Hongling Ye[3], Chenjia Zhang[4,5,*]

[1]Chongqing No.1 Secondary School, Chongqing, China
[2]Wuhan Britain-China School, Wuhan, China
[3]Chongqing No.1 Secondary School, Chongqing, China
[4]No.2 High School of East China Normal University, Shanghai, China

[5]williamzhangshanghai@outlook.com
*corresponding author

Abstract. With the development of hardware, the technology of artificial intelligence is more and more developed, and computer vision is once again a craze. Semantic segmentation is an important issue in computer recognition. Semantic segmentation makes the computer recognize the objects in a given image and classify and present them, which can save huge human time and cost, and the mature algorithms are more accurate than manual. The paper will focus on four sections: Firstly, different semantic segmentation algorithms were introduced. Then, the performance, advantages, and disadvantages of representative algorithms in each method were compared and analyzed, and the applications of semantic segmentation were listed. Finally, the optimization of semantic segmentation algorithms was introduced. The aim is to clarify the idea of the algorithms, the design intention, and the correlation between them, to understand the limitations of different algorithms, to enable the researcher to choose the most appropriate method, and to gain further new inspiration.

Keywords: Computer vision, Thresholding method, deep learning, semantic segmentation.

1. Introduction

Semantic segmentation is a key technique in computer vision that involves taking raw data (flat image) as input and converting it into a mask with highlighted areas of interest. Generally speaking, semantic segmentation is to classify each pixel in an image and assign a corresponding category label according to the target or region to which it belongs. It has long been a basic, pixel-level perception and understanding task in computer vision. It also relies on data sets to make judgments, and now mainly relies on pascalVOC. Semantic segmentation plays a vital role in many practical applications, such as medical image processing, robot vision, remote sensing image augmented reality, image compression and transmission, autonomous driving vision and intelligent video analysis and other fields. It can help us quickly identify such as traffic conditions, human body structure and so on in human daily life, so that people can quickly identify the situation and give answers.

Before the advent of deep learning, semantic segmentation was based on traditional image segmentation, including a variety of region - and boundary-based algorithms, such as K-means clustering [1], graph cut method [2], etc. After deep learning was proposed, semantic segmentation

ushered in an era of rapid development, and its performance was significantly improved. Full convolutional networks are also used for semantic segmentation for the first time, and Fully Convolutional Networks (FCN) algorithm makes convolutional neural networks gradually become a new model of image segmentation. FCN improves the accuracy and performance of semantic segmentation significantly through the three technologies of convolution, upsampling and skip structure. However, semantic segmentation still cannot recover the lost information in sampling. Therefore, many models also add rich and diversified methods to capture context information. Based on this demand, researchers have invented feature extraction networks such as Visual Geometry Group (VGG) [3] and GoogLeNet [4].

Semantic segmentation will be mainly applied to geographic information systems, medical image analysis and other neighboring areas before 2020, to help people make rapid judgments when encountering complex situations, simplify complex scenarios, and continuously strengthen its functions through deep learning. Of course, semantic segmentation still has many defects, such as low accuracy, and the following problems: Therefore, Wang et al. proposed a meta-learning based small sample semantic segmentation algorithm to address the problem of low segmentation accuracy for unknown new classes in existing small sample semantic segmentation models, and constructed an adaptive meta learning classifier using a vision transformer [5]. Qin et al. proposed a power line segmentation network to solve the problem of low prediction speed and accuracy in semantic segmentation algorithms. To improve the prediction speed of the network, a lightweight backbone feature extraction network was designed using Swin Transformer V2 in the encoder part, and a dynamic snake like spatial pyramid pooling module was proposed to improve segmentation accuracy. The improved algorithm achieved good results in accuracy [6].

To sum up, semantic segmentation plays an important role in real life, and plays different roles in each neighborhood, providing a lot of technologies for image neighborhood. Although there are still many defects in image segmentation nowadays, with the progress of technology and the innovation of algorithms, the significance of semantic segmentation will continue to be reflected.

2. Types of semantic segmentation
For computer, the pixels on an image are just a collection of 0s and 1s. The ultimate goal of the segmentation is to divide these elements into "background", "object A", "object B", etc.

2.1. P-tile method
As one a bit traditional method, P-tile applies a relatively simple idea: After knowing the approximate ratio of the image size, the computer needs to try parameters continuously, until the parameter can obtain values which is close to the expected ratio. Then it is used to define the threshold for segmenting the image. In addition, the background is defined as brighter than the objects. Though this method seems too simple and rash, while requires an approximate range ratio. It has strong resistance to the noise, and the calculation is relatively light. There are fewer requirements to the hardware[7].

2.2. Two-peak method
Among the data analysis diagrams, the grey level histogram shows the value and the distribution of each pixel clearly. In the processing of some images which has a large difference between the grey value of the objects and the background, the two-peak method can easily obtain the defined value in the valley between the two peaks in the grey level diagram to segment this image. This method is efficient due to it is easy to understand and only requires a small amount of calculation. On the other hand, it needs high-resolution images, and it is weak to the noise.

2.3. The Otsu methode
First, every pixel of the image is labeled and defined as different gray levels, from 0 to L-1. Then the computer processes these values. If the definition is T, the number of grey values which is less than T is recorded as N2, and their ratio is recorded as ω1. While the greater one is N1, and their ratio is ω2.

The average gay value of these two groups of pixels is recorded as μ1 and μ2. The total average grey level of the image is μ.

It can be seen from this: the ratio $\omega1 = N1 / (N1 + N2)$; $\omega2 = N2 / (N1 + N2)$

Image gray mean $\mu = \mu1 * \omega1 + \mu2 * \omega2$

The between-class variance of the two classes is:

$\sigma2 = \omega1 \times (\mu - \mu1)2 + \omega2 \times (\mu - \mu2)2$

$\sigma2 = \omega1 \times (\mu\text{-}\mu1)2 + \omega2 \times (\mu\text{-}\mu2)2$

It can be simplified as: $\sigma2 = \omega1 \times \omega2 \times (\mu1 - \mu2)2$

The final threshold selection formula is T = Argt ∈ { 0, 1, 2... L } Max (taking the maximum value of between-class variance σ2).

Though the calculation of this algorithm formula is faster and more stable than the above two. However, there will still appear problems of abnormal division when the noises are too many or the ratio of the target to the background is very different.

The methods introduced previously is a bit out-date, while newer algorithms based on deep learning will be mentioned.

2.4. FCN method

The principle of FCN is to cut the image into several pieces (X pixels* Y pixels), this progress is called 'cut'. Then, assign to them to obtain a smaller feature map, it is 'convolution'. The type of the image is determined by the learned fundamental function after knowing the importance of each piece. The definition line is drawn after the pieces are restored to the original image size.

Only with complex training and various artificially added regular functions can the machine learn a high-precision definition function. The edge is that this algorithm does not require specific image size, resolution or the ratio of the object to the background in the image. As soon as the model is trained properly, its accuracy remains high.

2.5. VGG algorithm

VGG algorithm is an early improvement of FCN algorithm. The algorithm uses a fixed 3 * 3(pixel) convolution core, and increases the depth of the network to improve the accuracy of recognition. In fact, VGG algorithm is a class of algorithms, there are VGG16 VGG19, and so on, the difference between the two is the number of convolution layer has differences. In contrast, the VGG algorithm training is more rigorous, using a fixed size of 224 * 224 RGB, over-averaging the image. After improvement, the structure of the algorithm is realized because of the fixed value. However, due to the increase of the full connection layer, the parameter increases rapidly and requires a lot of computing and resources, which requires higher hardware requirements.

2.6. The U-net algorithm

As an algorithm model composed of an encoder and a decoder, the encoder uses maximum pooling to sample down the current features, and takes the maximum, passing it step by step. In this way, the features can be extracted from a given image layer by layer, and the decoder can sample the processed features, restore the precision and image quality, segment the object, and produce the final output. Because the encoder and the decoder are symmetric, and each level of encoding and decoding has a path connection it has the characteristics of skipping the connection between the encoder and the decoder block. This structure allows information to be passed directly between encoder and decoder, which helps to preserve more spatial information in semantic segmentation tasks. The advantage of this algorithm is that its structure is simple and its segmentation is accurate in the case of small training set, but there is still some image loss in the decoding process [8].

3. Segmentation algorithm experimental results analysis

3.1. different data sets

The table lists several mainstream algorithms, including one based on encoding and decoding, one based on hole convolution, and one based on attention mechanism, and the Miou (Mean Intersection over Union) index on two commonly used datasets, CityScapes and Pascal VOC2012. Miou is a commonly used index to evaluate the accuracy of image semantic segmentation. The higher the value, the better the segmentation effect. [9]

3.2. Experimental form

The following details of the analysis of several mainstream semantic segmentation algorithms in common data sets on the MIOU experimental results comparison, as shown in Table 1:

Table 1. Miou experiments on commonly used data sets in mainstream semantic segmentation algorithms

Mainstream algorithms	Data set	mIoU/%
DeconvNet	PASCAL VOC2012	72.5
DeepLab V1	PASCAL VOC2012	71.6
DeepLab V2	PASCAL VOC2012	79.7
DeepLab V3	PASCAL VOC2012	86.9
DeepLab V3+	PASCAL VOC2012	89.0
PAN	CityScapes	78.6
DANer	CityScapes	81.5
OCNer	CityScapes	81.4
ENet	CityScapes	58.3
LinkNet	CityScapes	76.4
BiSeNet	CityScapes	68.4
Fast-SCNN	CityScapes	68.0
DFANet	CityScapes	71.3

3.3. Results analysis

Several conclusions can be drawn from these data:

Data set impact: different algorithms perform differently on different data sets. For example, the DeepLabV-V3 + algorithm performed well on the Pascal VOC2012 dataset and also on the CityScapes dataset, but overall performance on the CityScapes dataset was generally slightly lower than on the Pascal VOC2012 dataset.

Algorithmic: DeepLab V 3 + is one of the best performing algorithms on both datasets, demonstrating that the combination of hole convolution and attention mechanism is effective.

Application of the attention mechanism: the introduction of the attention mechanism generally improves the segmentation performance, indicating that the attention mechanism can help the model better focus on the important feature areas.

In conclusion, the experimental data show that there are some differences between different algorithms in semantic segmentation, and the choice of data set is also very important to the algorithm.

4. Application of semantic segmentation

4.1. Medical identification

Medical imaging methods continue to enrich, such as computed tomography (CT), magnetic resonance imaging (MRI), radiography, and so on. Manual inspection is not only time-consuming, but also prone

to errors in situations where time is tight or exceptionally subtle. Semantic segmentation can recognize and segment medical images, detect and quantify the patient's condition, and help patients to intervene treatment as soon as possible. And highlighting areas of interest to significantly reduce the workload of medical personnel. Among the methods described above, the threshold method is the most simple and effective when the contrast between the lesion area and the background is obvious. Such as bimodal analysis of tumors and other abnormal masses, can be very clear that the region of variation. Full convolutional network segmentation is effective and accurate, and its applications include brain tumor segmentation [10], abdominal CT or MRI segmentation [11], eye region segmentation, etc[12]. However, medical image segmentation is faced with many challenges, such as high noise, motion artifacts, and difficulty in obtaining a large number of annotated training sets.

4.2. Traffic analysis
Semantic segmentation in traffic analysis can be applied to automatic driving, intelligent road monitoring, infrastructure management. This is mainly about autopilot. The development of automatic driving can reduce traffic pressure, regulate urban traffic and ensure driving safety. Semantic segmentation identifies obstacles and driving areas, and provides information such as traffic signs, pedestrians and roads for self-driving vehicles to realize their environmental perception. Traffic scenes in two categories are mainly identified by shallow learning methods, such as Otsu and other threshold methods for fast vehicle detection. The depth learning method can improve the accuracy and solve the multi-class complex traffic environment. The representative neural networks such as VGGNet are good at single label classification, and based on this Long's semantic segmentation method based on FCN [13] its semantic segmentation task can use the learned features. ResNet, VGG16 and other classic work through deepening the network to gradually extract high-level semantic information to solve the target segmentation error accuracy problem. In automatic driving, segmentation is faced with the challenges of real-time requirement, abnormal situation (traffic accident) processing and so on.

4.3. Logistics sorting
With the development of sensing, grasping and robotic engineering technology, under the help of semantic segmentation for object detection and classification, logistics sorting is gradually explored by human manual completion and automated completion in semi-structure scene, improve efficiency, but also reduce labor costs. Semantic segmentation minimizes the potential hazards or high costs of misplacing products in a logistics warehouse by adding semantic tags to each pixel and separating different entities. Traditional detection methods can not meet the real-time needs, and the semantic segmentation model based on depth learning can process images in real time and keep goods flowing. Some examples are express cartons with small chromatic aberration. VGG algorithm can better extract express cartons to predict and sort.

5. Optimization of algorithms and their development
Traditional image segmentation methods such as thresholding method use mathematical knowledge to segment images, low computational cost, fast segmentation speed, but cannot guarantee accuracy in details and are not suitable for processing complex images. Thresholding technique Otsu method [14], can be improved by using local uniform information. However, the thresholding method selects a single threshold to divide the image into foreground and background, and it is difficult to divide the details with a single threshold in the face of complex images.

Deep learning based semantic segmentation, such as full convolutional networks, significantly improves segmentation accuracy and efficiency compared to traditional image segmentation. To further optimize the model three aspects of accuracy, expertise, and response time can be considered.

Boosting accuracy can be considered:

(1) Using stochastic gradient descent

(2) Reinforcing the learning mechanism (adjusting the learning rate to improve the model learning ability)

(3) Adjusting the ratio of test set and dataset (to prevent overfitting)

(4) Add antagonistic samples. (to improve perturbation resistance)

Improvement of specialization can be considered:

(1) Excluding non-optional ranges

(2) Using more specialized test sets & datasets.

Reducing response time can be done by:

(1) Reducing the number of unnecessary convolutional layers.

A limitation of the FCN architecture is the low resolution of the final output segmented image after multiple convolutional and pooling layers. As well as the localization of FCN based methods limits the capture of long-distance dependencies in the feature map. For this Dzmitry Bahdanau [15] and others investigated attention mechanisms to merge or replace these models. This also laid the foundation for the later Transformer model.

6. Conclusion

In its current state, the limitations of the thresholding method are exposed. Only by providing multiple thresholds is it possible to improve accuracy in the face of more complex images, however, this involves underlying algorithmic logic and is difficult to implement. Deep learning is a new path for semantic segmentation. Although the hardware requirements are much higher, the accuracy has risen dramatically. For example, full convolutional networks significantly improve segmentation accuracy and efficiency compared to traditional image segmentation. Therefore, without sticking to the traditional threshold segmentation methods, improving or creating new convolutional networks is a better and more value-retaining option.

The paper roughly introduces several mainstream semantic segmentation algorithms, such as the older P-tile method, bimodal method, maximum inter-class variance method; and the newer FCN algorithm, VGG algorithm and U-net algorithm, and gives a list of evaluation indexes about the results of the U-net algorithm; the paper discusses some of the above methods in the actual scenarios, such as hospitals, intersections, and sorting stations, and apply algorithms that are more in line with the essence of the problem, and give targeted advice; finally, the paper points out that the general-purpose convolutional network is a better and more efficient option. It also discusses how some of the above methods can be applied to real-world scenarios such as hospitals, intersections and sorting stations, where the algorithms are more appropriate to the problem. The paper hopes to help new developers to understand the past algorithms to get new ideas for building algorithms and to specialize different algorithms.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1]   Wang, H., Yang, X., Liu, T., et al. (2024). A novel deep-learning model for detecting small-scale anomaly temperature zones in RDTS based on attention mechanism and K-Means clustering. Optical Fiber Technology, 88, 103969-103969.

[2]   Xiang, L., Brett, C., Alistair, Y. (2005). Model-based Graph Cut Method for Segmentation of the Left Ventricle. Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 20053059-62.

[3]   Chen, H., Haoyu, C. (2019). Face Recognition Algorithm Based on VGG Network Model and SVM//Asia Pacific Institute of Science and Engineering. Proceedings of 2019 3rd International Conference on Machine Vision and Information Technology (CMVIT 2019). College of computer and information science, Southwest University; College of Mathematics and Statistics, Southwest University, 2019:8.

[4]   Zhong, Z., Jin, L., Xie, Z. (2015). High Performance Offline Handwritten Chinese Character Recognition Using GoogLeNet and Directional Feature Maps. CoRR, abs/1505.04925.

[5]   Wang, L. Z., Mou, C. S. (2024). Small sample semantic segmentation algorithm based on meta learning. Journal of Jiangsu University (Natural Science Edition), 45(05): 574-580+620.

[6]   Qin, L. M., Wang, C. J., Bian, H. Q. et al. (2024). Power Line Semantic Segmentation Network Integrating Transformer and DeepLabv3+. Modern Electronic Technology, 47(17): 109-116. DOI:10.16652/j.issn.1004-373x.2024.17.018.

[7]   Han, S. Q., Wang, L. (2002). Overview of Threshold Methods for Image Segmentation Systems. Engineering and Electronic Technology.

[8]   Zhang, Y. Z., Yu, Q., Su, J. S. et al. (2023). From U-Net to Transformer: A Review of the Application of Deep Models in Medical Image Segmentation. computer application.

[9]   Liao, W. S., Li, X. W., Xu, C. (2021). A review of research on semantic segmentation of driving scenes based on deep learning. Beijing Union University Beijing Key Laboratory of Information Service Engineering.

[10]   Kamnitsas, K., Bai, W. Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M. Lee, M., Kainz, B., Rueckert, D. and Glocker, B. (2018). Ensembles of multiple models and architectures for robust brain tumor segmentation. in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer, Cham, pp. 450–462.

[11]   Christ, P. F., Ettlinger, F., Grün, F., Elshaera, M. E. A., Lipkova, J., Schlecht, S., Ahmaddy, F., Tatavarty, S., Bickel, M., Bilic, P., et al. (2017). Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv.2017, arXiv:1702.05970.

[12]   Edupuganti, V. G., Chawla, A., Amit, K. (2018). Automatic optic disk and cup segmentation of fundus images using deep learning. In Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10, pp. 2227–2231.

[13]   Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. IEEE Computer Vision and Pattern Recognition. 3431-3440.

[14]   Yang, X., Xu, G. and Zhou, T. (2021). An effective approach for ct lung segmentation using region growing, in Journal of Physics: Conference Series, vol. 2082, p. 012001.

[15]   Bahdanau, D., Kyunghyun, C. and Yoshua, B. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. arXiv preprint airXiv:1409.0473.