Follicular Ultrasound Image Segmentation based on Improved Deeplabv3

Tianlong Zeng[1, 2,a, *], Jun Liu[1,2]

[1]School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, P. R. China

[2]Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan 430081, P. R. China

[a] 942636327@qq.com

*Corresponding author

Keywords: DeepLabv3; ASPP; The ultrasound image segmentation technique of yellow cattle follicles; follicle monitoring.

Abstract: The ultrasound image segmentation technique of yellow cattle follicles plays an important role in the monitoring of the dynamic changes of follicles in yellow cattle. With the development of deep learning, the image segmentation technology based on deep learning neural network model has made great breakthroughs, such as DeepLabv3, an end-to-end semantic segmentation network model. However, the above algorithm can't meet the requirements of follicle segmentation in cattle follicle monitoring. Because of its deep network depth, multiple operational parameters, many iterations, and huge amount of computation, it has high requirements for the operating environment of the system (memory, CPU, GPU, etc.), and cannot be easily applied to the actual production practices. Therefore, an improved DeepLabv3 model was proposed in this paper. By removing the ASPP layer and moving the atrous convolution layer forward, the improved model is more suitable for the actual production of follicle segmentation, which is more in line with the economic cost of follicle monitoring. Finally, the experimental results show that the computational resources and the running time of the improved model decrease obviously when the average segmentation precision of follicular ultrasound image does not decrease significantly.

1. Introduction

Image segmentation is an important part of image processing and image analysis. Image segmentation is essentially a clustering process of pixels in an image. Traditional image segmentation algorithms, such as threshold segmentation, edge segmentation and image segmentation algorithm based on mathematical morphology, are not satisfactory in the face of complex images due to the single use of image information[1]. In order to further improve the accuracy and accuracy of image segmentation, the image segmentation algorithm based on machine learning is solved in high dimensional feature space. Some machine learning clustering algorithms are also gradually applied to image segmentation, such as K-means algorithm, Markov random field, support vector machine (SVM) and so on[2]. However, in the machine learning algorithm, the extraction method of image data features relies too much on human experience, and the characteristics of some high-dimensional spaces are inevitably neglected, resulting in the extracted features not being able to fully characterize the data, and the results obtained are also unsatisfactory.

In recent years, with the development of deep learning, the image segmentation task has made significant progress. The early application of deep learning to image segmentation is based on pixel classification at the pixel level of the CNN network. FCN is a network proposed by Jonathan Long et al. for image semantic segmentation[3]. The network replaces the fully connected layer in the traditional image segmentation network with the convolution layer, breaking through the pixel-level stereotype mode in the traditional CNN. Subsequently, in 2015, based on the coding-decoding architecture SegNet and U-Net network achieved better results in the semantic segmentation of

images [4]. In the same year, Fisher Yu et al. proposed that the Dilated Convolutions layer use a "context module" to aggregate multi-scale information. In 2015 and 2017, Liang-Chieh Chen and the Google team proposed DeepLabv1 and DeepLabv2 for image semantic segmentation[5]. DeepLabv1 combines DCNNs with fully connected CRF and applies Hole algorithm to DCNNs model innovatively to segment images. DeepLabv2 proposes atrous spatial pyramid pooling (ASPP), which extracts features at different scales by taking advantage of atrous convolution[7]. They improved the atrous spatial pyramid pooling layer, designed a series or parallel atrous convolution module to capture multi-scale context information, and achieved good results in segmentation accuracy and accuracy[8].

In view of the segmentation task of the cattle follicle ultrasound image in this study, the economic cost pressure it faces in the actual production practice is harsh, and the hardware resources of the system operation are also limited, so the above method is not suitable for the follicle monitoring system. Therefore, an improved DeepLabv3 neural network model was proposed in this paper which reduces the accuracy of follicular ultrasound image segmentation to an acceptable range, while the computational parameters are reduced in volume, the system training time is reduced, and the dependence on computing resources is reduced, so that it has better robustness for actual production.

2. Background

2.1 DeepLabv3

In 2015, Long et al. proposed Full Convolutional Network (FCN), which extended the original CNN structure and made intensive prediction without full connection layer, extending the segmentation algorithm from pixel level classification to image level. Liang-chieh Chen and Google team put forward DeepLab series model of image semantic segmentation in 2015 and 2017 respectively, and obtained the best segmentation effect of image semantic segmentation.

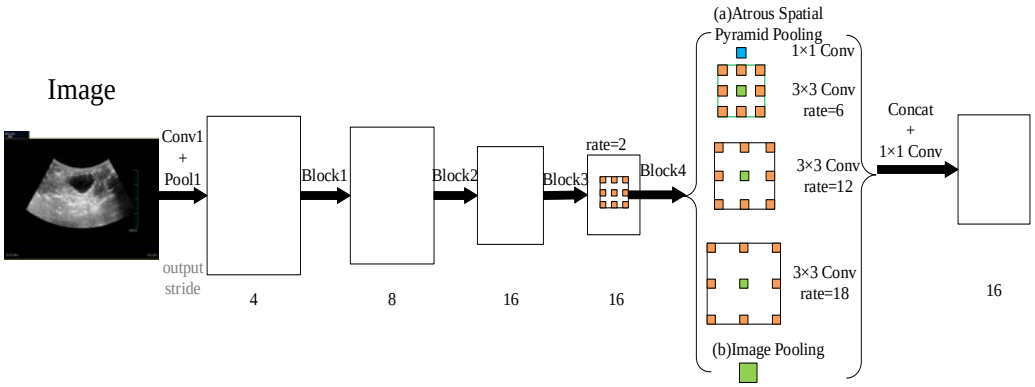

Figure 1. DeepLabv3 model structure.

Unlike most encoder-decoder network architectures, DeepLabv3 proposes a semantic segmentation architecture for controlling signal extraction and learning multi-scale context features. DeepLabv3 is an optimization based on DeepLabv2. It modifies the multi-scale convolution feature by modifying the ASPP module proposed by DeepLabv2, and encodes the global background based on the image hierarchy to obtain features.

Deeplabv3 takes ResNet pre-trained on ImagNet as its main feature extraction network. It adds a new residual block for multi-scale feature learning. The last ResNet block uses an atrous convolution instead of a regular convolution. In addition, each convolution within this residual block uses a different dilation rate to capture multi-scale context information. In addition, the top of this residual block uses the Atrous Spatial Pyramid Pooling (ASPP). ASPP uses convolution of different expansion rates to classify regions of any size.

The DeepLabv3 model structure is shown in figure 1.

2.2 Atrous Convolutions

Atrous convolutions also known as dilated convolutions, the dilation rate parameters are introduced to the convolution layer defines the convolution kernels spacing of each value when dealing with data.

Consider two-dimensional signals, for each location i on the output y and a filter w, atrous convolution is applied over the input feature map x:

$$y[i] = \sum_{k=1}^{K} x[i + r \bullet k]\omega[k] \tag{1}$$

where the atrous rate r corresponds to the stride with which we sample the input signal, which is equivalent to convolving the input x with upsampled filters produced by inserting r − 1 zeros between two consecutive filter values along each spatial dimension. Standard convolution is a special case for rate r = 1, and atrous convolution allows us to adaptively modify filter's field-of view by changing the rate value.

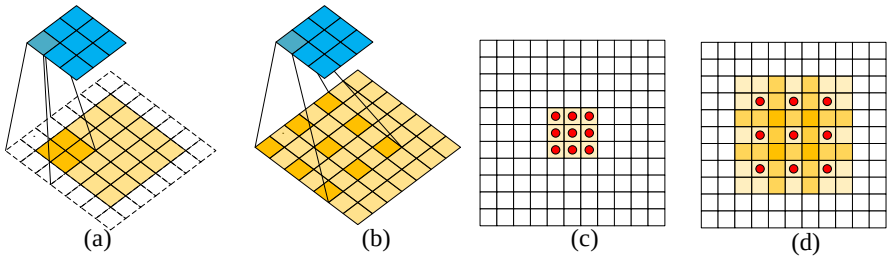

Figure 2. Atrous convolution and standard convolution

As shown in the figure 2, (a) and (c) show normal convolution, and (b) and (d) show 3*3 2-dilalated atrous convolution with a hole of 1. It can be seen that compared with the ordinary convolution, the receptive field of the atrous convolution is expanded from 5*5 to 7*7, and the receptive field is enlarged, so that each convolution layer is added without the loss of information in the pooling operation. The output contains a large range of information to make better use of context information.

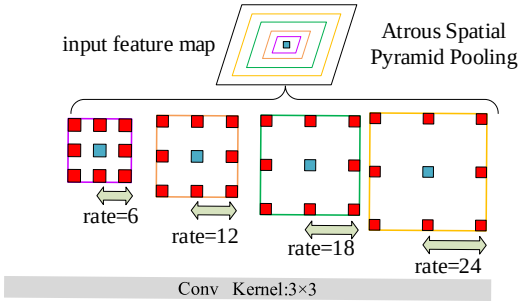

Figure 3. Atrous spatial pyramid pooling.

The purpose of the atrous convolution is to provide a larger receptive field without using the pooling layer (the pooling layer will result in information loss) and with the same amount of calculation, and the spatial resolution will not be reduced. This is very useful in the image segmentation task, because the large receptive field can detect and segment of large targets, high resolution can accurately locate the target. In addition, by setting the different atrous convolution parameters dilation rates, different receptive fields can be obtained, that is, multi-scale context information can be obtained.

2.3 Atrous Spatial Pyramid Pooling

Atrous spatial pyramid pooling (ASPP) achieves more robust segmentation results with multi-scale information. ASPP adopts multiple atrous convolutional layers with different dilation

rates to detect feature maps and capture objects and image context in multiple proportions to adapt to feature extraction and accurate positioning of objects with different resolutions in images. ASPP The DeepLabv3 model structure is shown in figure 3.

3. An Improved Deeplabv3 Model

The dynamic monitoring system of follicles in yellow cattle is mainly based on the ultrasound image of follicles to judge the development of cattle follicles. The study object of this paper is an ultrasound slice image of several yellow cattle follicles. It can be seen from the figure 5 that the segmentation task of this study belongs to the segmentation of single-category objects, and most of the images contain only a single segmentation target. Therefore, the latest deep learning semantic segmentation network architecture DeepLabv3 is very suitable for the segmentation task of this study. However, DeepLabv3 network architecture, like other deep learning segmentation models, has a large parameter volume and heavy computing task. The target of real-time monitoring of the cattle follicles is particularly dependent on computing resources, especially demanding requirements for the CPU and GPU of the system. Therefore, excessive economic costs make it impossible to apply it to actual production practices.

In the follicular monitoring system, the development of the follicle is judged mainly according to the size of the follicle volume, that is, the target area of the ultrasound image of the follicle, and thus the accuracy of segmentation of the follicle edge is not high. Based on this, this paper proposes an improved DeepLabv3 model to complete the segmentation task of this study.

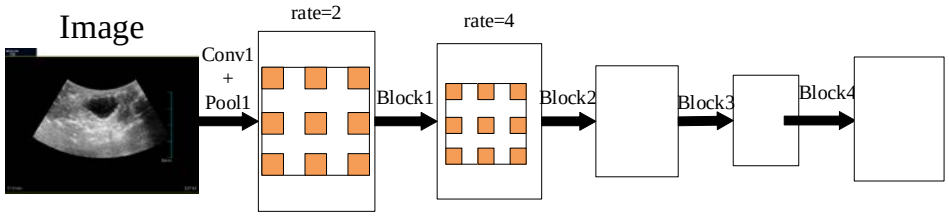

Figure 4. improved DeepLabv3 model structure.

As can be seen from figure 5, most of the follicular ultrasound images have only one follicle target, and the shape, size and spatial position of the follicle are relatively stable, so it is unnecessary to use the idea of multi-scale image feature extraction to cope with the situation of different resolutions. Therefore, the improved DeepLabv3 model proposed in this paper first removes the original ASPP structure to simplify the network and reduce computational parameters. Secondly, the follicle data set in this study uses ultrasonic imaging technology, whose image quality is not high, containing a lot of speckle noise, the target edge is fuzzy, and the contrast is not obvious. After directly removing the ASPP structure, the segmentation accuracy will be greatly reduced. Moreover, the atrous convolutional layer in the network is in the post-convolution layer, so it is very sensitive to the noise in the image. Therefore, the improved DeepLabv3 in this paper adjusts the hole convolutional layer to the convolution layer of the first two layers. The advantage of this is that it can improve the receptive field of large-resolution features, make the large-resolution feature map of the shallow network more convolution in the large receptive field, improve the weight of the high-frequency information in the network, and make better use of the coarse Scale information to compensate for the problem of reduced segmentation accuracy due to the removal of the ASPP structure.

In addition, because the segmentation task of this study is simple and the target features are relatively thin, in order to further reduce the computational volume and improve the segmentation speed, the improved DeepLabv3 model abandons the original ResNet101 network and adopts a shallower ResNet50 network architecture.

The improved DeepLabv3 model structure is shown in the following figure 4.

4. Experimental Results and Analysis

The ultrasound cattle follicle image set of this experiment is derived from the National Natural Science Foundation of China, "Research on Key Techniques for Monitoring Dynamic Changes of Follicles in Nanyang Yellow Cattle Based on 3D Ultrasound Imaging", including 6000 three-dimensional ultrasound slices from 115 yellow cattle follicles.

The software environment of the experimental code running in this paper is TensorFlow1.6, python3.5, and the hardware environment is 8GB RAM, Intel i5-8400 CPU, and NVDIA GTX1070.

Since the purpose of this experiment is to determine the development of follicles through the size of follicle, so as to achieve the goal of follicle monitoring, and the size of follicle volume is reflected in the size of the target area in the ultrasonic image, this paper adopts pixel average segmentation accuracy as the segmentation effect index. The formula for pixel segmentation accuracy is defined as follows:

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \tag{2}$$

Where k represents the number of categories (from C_0 to C_k, where C_0 represents the background category), and p_{ij} represents the number of pixels that belong to class I but are predicted to be class j. For the specific task of this study, $k = 1$: only one category of follicles was included.

In this experiment, the following ablation experiments were performed: the follicle ultrasound image dataset was trained using the Deeplabv3 network model, the Deeplabv3 without ASPP network model, and the improved Deeplabv3 model. To be fair, all three models use the same configuration training network in the same environment that is independent of each other.

The training results of these three models are shown in the following table 1.

Table 1. training results of these three models.

method	parameters	accuracy (%)	training speed(img/sec)
DeepLabv3	93.2M	95.1	1.35
DeepLabv3 without ASPP	54.7M	89.3	3.08
Improved DeepLabv3	56.4M	92.7	2.96

As can be seen from the table 1, the original DeepLabv3 model has a very good segmentation effect on the follicular ultrasound data set, with an average pixel segmentation accuracy of 95.1%. But it is also obvious that it has a relatively large number of parameters, with 93.2MB of data. When training, the speed is also relatively slow, only 1.44 images per second. When we removed the ASPP structure from the network, the training parameters were reduced by almost half and the speed was much faster, reaching 3.05 images per second. However, its segmentation accuracy decreased significantly, by 5.8%. The improved DeepLabv3 network model proposed in this paper achieved a segmentation accuracy of 92.7%, which was only 2.4% lower than the original. However, the training time was significantly reduced, and the training speed was more than twice as fast as the original. At the cost of sacrificing less accuracy, it brings considerable time cost benefits, and makes the improved model more suitable for practical production and more practical value.

The results of follicle segmentation in this experiment are shown in the figure 5:

In figure 5, column a is the original image, column b is the DeepLabv3 segmentation result, column c is the improved DeepLabv3 segmentation result, and column d is the ground truth of the image.

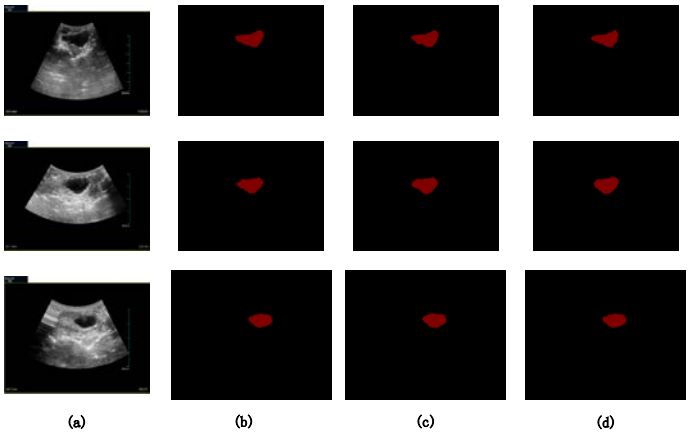

Figure 5. The results of follicle segmentation

5. Conclusion

According to the characteristics of the follicle ultrasound image, this paper improves the Deeplabv3 network structure, removes the original ASPP structure, and moves the atrous convolution layer forward, so that the new network reduces the segmentation accuracy within the acceptable range, and the calculation amount is reduced. Therefore, the speed of image segmentation is improved, the requirements of the system hardware environment are lowered, and the economic cost of follicle monitoring is more suitable, which can be better applied in actual production practice.

The improved DeepLabv3 network model reduces the amount of computation to a certain extent, speeds up the training time of the network, and broadens the hardware and environment cost of the system, but at the expense of certain segmentation accuracy. Designing a semantic segmentation network which emphasizes both training speed and segmentation accuracy is the direction for further research.

References

[1] Liu J, Li P F. A Mask R-CNN Model with Improved Region Proposal Network for Medical Ultrasound Image[C]// International Conference on Intelligent Computing. Springer, Cham, 2018.

[2] Wang Z, Cai W, Smith C D, et al. Residual Pyramid FCN for Robust Follicle Segmentation[J]. 2019.

[3] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4):640-651.

[4] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37 (9):1904-16.

[5] Chen L C, Papandreou G, Kokkinos I, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[J]. Computer Science, 2014(4):357-361.

[6] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 40(4):834-848.

[7] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions[C]// ICLR. 2016.

[8] Chen L C, Papandreou G, Schroff F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation[J]. 2017.