Cascaded Fully Convolutional DenseNet for Automatic Kidney Segmentation in Ultrasound Images

Zhiwei Wu[1], Jinjin Hai[1], Lijie Zhang[2], Jian Chen[1], Genyang Cheng[2], Bin Yan[1,*]
[1]PLA Strategy Support Force Information Engineering University
[2]The First Affiliated Hospital of Zhengzhou University
Zhengzhou, China
e-mail: ybspace@hotmail.com

***Abstract*—Although the extraction of traditional features has achieved good performance in kidney segmentation, the automatic kidney segmentation in whole ultrasound images is still challenging. Due to the characteristics of the kidney ultrasound images such as obvious noise, severe redundancy information and low image contrast, it is difficult to achieve accurate segmentation of the kidney by directly using the fully convolutional dense network (FC-DenseNet). Therefore, we propose a cascaded FC-DenseNet, that is, a model of coarse-to-fine segmentation. By training the coarse segmentation model, prediction probability that the pixel becomes the correct class label is calculated, and the proposal region of the kidney in the ultrasound image is obtained. By training the fine segmentation model, the proposal region of kidney is finely segmented and restored back to the original size. Compared with the original FC-DenseNet, cascaded FC-DenseNet has better performance in kidney segmentation. The mean intersection over union is increased by 1.43%, and the samples that are difficult to segment could be better segmented.**

Keywords-ultrasound image; kidney segmentation; coarse-to-fine segmentation; FC-DenseNet

I. INTRODUCTION

Ultrasound is widely used in medical imaging because of its non-invasive, real-time and low-cost. It plays an important role in medical diagnosis, treatment and prognosis. Ultrasound image of the kidney is an important reference for the diagnosis of nephropathy. It is often used for preoperative examination of renal puncture biopsy and real-time monitoring of nephropathy. However, due to the rigor of medical imaging diagnosis, the kidney segmentation results of ultrasound images are less practical. The segmentation accuracy is affected by many factors, such as the complex structure of the kidney, the uneven gray scale of image, the relatively high noise of the ultrasound imaging system, the low contrast of the image, and the inconspicuous kidney boundary.

As a key technology in the field of computer vision, image segmentation has always been the focus. The methods of traditional image segmentation divide the image into foreground and background. They are mostly unsupervised segmentation method. Kidney segmentation in ultrasound image could be regarded as the segmentation of foreground and background. Many methods of traditional image segmentation are applied to the kidney segmentation task and have achieved certain segmentation effects. The method based on texture and shape priors for kidney segmentation in ultrasound images was proposed [1]. By constructing a texture model, the texture similarity between the inner region and outer region of the kidney was measured and different regions according to different texture features were divided. Region splitting and merging algorithm was used for the real time kidney ultrasound image segmentation method [2]. The average calculation time per image was 8.5 seconds by this way. To obtain the outline of the kidney, the gradient vector flow from the gray level and binary boundary maps in the image was calculated [3]. By comparing five common image enhancement techniques, the literature [4] pointed out that image enhancement techniques of median filtering or histogram equalization should be considered before segmenting the kidney edges. To improve the detection sensitivity and specificity for the weak boundary of the kidney image, the snake balloon and the canny operator were combined in a fusion algorithm [5]. The kidney segmentation in the ultrasound image was achieved by the dynamic graph-cuts method that the intensity information of the original image and the texture feature map extracted by the Gabor filter was integrated [6]. The above methods could be categorized as the semi-automated kidney segmentation method in local ultrasound images by extracting traditional image features.

In recent years, Convolutional Neural Network (CNN) has achieved remarkable achievements in various fields such as image classification [7]-[9], target detection [10]-[12] and image segmentation [13]-[15] due to its strong feature representation and information extraction capabilities. Although many researchers pay attention to CNN, there are relatively few applications that apply it on kidney segmentation in ultrasound images. Literature [16] proposed an end-to-end learning method to achieve automatic kidney segmentation in ultrasound images by constructing a transfer learning network, a boundary distance regression network and a pixel classification network. Although automatic kidney segmentation in the whole ultrasound image was achieved, the internal contour segmentation of kidney was not involved.

Compared with other semantic segmentation networks, such as FCN [13] and U-Net [17], FC-DenseNet [18] has better segmentation performance and less parameters. It is

978-1-7281-0831-5/19/$31.00 ©2019 IEEE

suitable for segmentation with less data. Therefore, FC-DenseNet is introduced to segment the inner and outer contours of kidney in ultrasound image. The size of the kidney area is relatively small compared to entire ultrasound image, and redundant information of ultrasound image is more. If only the original FC-DenseNet is used, the segmentation accuracy is relatively low. Therefore, cascaded FC-DenseNet for automatic kidney segmentation in ultrasound images is proposed. This algorithm uses two cascaded FC-DenseNet to focus on the inner and outer contours of the kidney. The parameters of the two networks are simultaneously optimized by means of joint training. The experimental results show that cascaded FC-DenseNet has a better segmentation result than the original FC-DenseNet.

II. METHEODOLOGY

A. FC-DenseNet

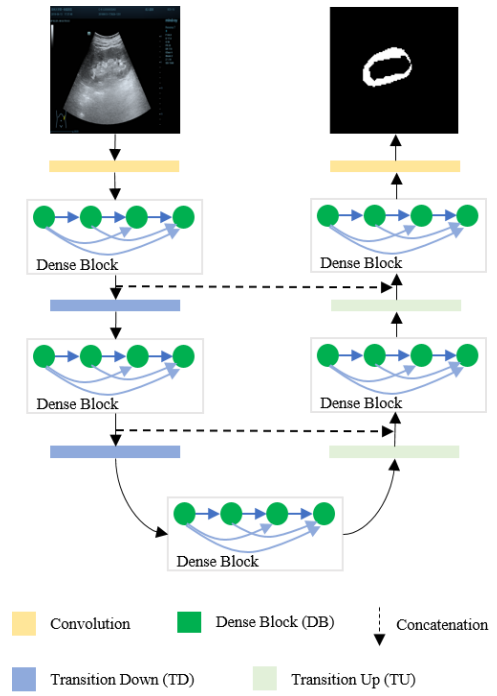


Figure 1. The architecture of FC-DenseNet.

A typical semantic segmentation network is generally composed of a downsampling path for extracting sparse semantic features, an upsampling path for restoring the original resolution, and a post-processing module for improving the segmentation result. In general, the segmentation network is training based on the pre-training classification network. The FC-DenseNet could be seen as a product of the DenseNet [19] mapping from the classification task to the segmentation task. It does not include the post-processing module nor the weight of the pre-trained classification network. The route is very similar to the full convolution network such as U-Net. As shown in Fig. 1, FC-DenseNet consists of four parts: convolution, dense block (DB), transition down (TD) and transition up (TU). DB is composed of batch normalization (BN) [20], rectified linear unit (ReLU) [21], a 3×3 convolution and Dropout [22]. A TD layer is composed of BN, ReLU, a 1×1 convolution followed by a 2×2 maximum pooling operation. TU contains a 3×3 transposed convolution with stride 2.

Compared with the DB in DenseNet, FC-DenseNet lacks 1×1 convolution. This situation will cause the number of feature maps in the upsampling path to increase linearly. The amount of calculation and the memory consumption will be greatly increased. In order to avoid these problems, the input and output of each module in FC-DenseNet are no longer directly connected. The information loss caused by the pooling is compensated by combining the feature map after the deconvolution operation in the upsampling path with the DB in the downsampling path.

B. Cascaded FC-DenseNet

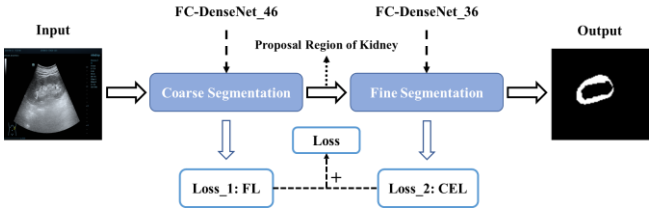


Figure 2. The architecture of cascaded FC-DenseNet.

As shown in Fig. 2, cascaded FC-DenseNet built on FC-DenseNet includes two modules: a coarse segmentation model and a fine segmentation model. The selection of the number of layers in the two basic networks is considered from data volume and segmentation performance. As shown in Tab. 1, the coarse segmentation model consists of a 46-layers FC-DenseNet. The training input of the model is the original image and the ground truth labeled by the professional doctor. The output is the probability value of each pixel belonging to a certain category. A proposal region of the kidney is formed according to the probability, and a coarse segmentation bounding box is obtained. Because the kidney region in the original image is relatively small, the category imbalance is serious. This problem will lead to the classifier being more biased towards the background class during training, and the segmentation performance is poor. Therefore, in the coarse segmentation model, we choose the focal loss [23] which could effectively solve the problem, as shown in equation (1).

$$FL = -\alpha_t (1-p_t)^{\gamma} \log(p_t) \quad (1)$$

where α_t is used to adjust the proportion of positive and negative examples. When α_t is used for the foreground class, $1-\alpha_t$ will represent the corresponding background class. p_t indicates the classification probability of different categories, which is calculated by sigmoid. Both γ and α_t are fixed values and do not participate in training.

After the model of coarse segmentation, the problem is simplified to kidney segmentation on a local ultrasound image. Therefore, as shown in Tab. 1, the fine segmentation model consists of a 36-layers FC-DenseNet with a smaller number of network layers. The training input for this model

is the local original image corresponding to the output of the coarse segmentation model and the ground truth in the corresponding location. The cross-entropy loss function in the fine segmentation model is used, as shown in equation (2).

$$CEL = -\frac{1}{N}\sum_{i=1}^{N} y_i \log p_i \tag{2}$$

where N is the number of pixels of the image, y_i is the class label of pixel i, and p_i indicates prediction probability that the pixel i becomes the correct class label.

TABLE I. THE ARCHITECTURE OF SUBNETWORK IN CASCADED FC-DENSENET.

Architecture	
46-Layers FC-DenseNet	***36-Layers FC-DenseNet***
Input	Input
3 × 3 Convolution	3 × 3 Convolution
DB(4 layers) + TD	DB(4 layers) + TD
DB(4 layers) + TD	DB(4 layers) + TD
DB(4 layers) + TD	DB(4 layers) + TD
DB(4 layers) + TD	DB(4 layers)
DB(4 layers)	TU + DB(4 layers)
TU + DB(4 layers)	TU + DB(4 layers)
TU + DB(4 layers)	TU + DB(4 layers)
TU + DB(4 layers)	1 × 1 Convolution
TU + DB(4 layers)	Softmax
1 × 1 Convolution	-
Softmax	-

As shown in equation (3), the loss function of the cascaded FC-DenseNet could be expressed as the sum of two models. The two models are optimized by joint training.

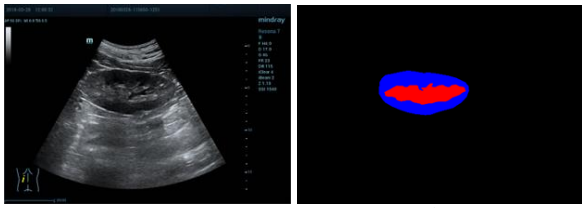
$$Loss = FL + CEL \tag{3}$$

III. EXPERIMENTS

A. Data

The data used in our experiments are digital kidney ultrasound images provided by the First Affiliated Hospital of Zhengzhou University in Henan Province. There are 461 images in the data set derived from 68 patients with nephropathy. The original data are gray-level digitized kidney ultrasound images with a resolution of 1260 (width) by 910 (height) pixels saved as standard DICOM format. As shown in Fig. 3, the original label of kidney is portrayed in red and blue. The boundary of blue area represents the outer outline of the kidney, and the boundary of red area represents the inner outline of the kidney. The data set is randomly divided into a training set and a test set, and there is no intersection between the two data sets. Among them, the training set contains 362 images, and the test set contains 99 images.

Due to the relatively large size of the kidney ultrasound images, FC-DenseNet needs to reduce the resolution of the feature map by multi-layer downsampling operation, which will increase the number of parameters to be learned and easily lead to over-fitting for the network. Therefore, we adjust the size of the input image to 512×512. In order to achieve the segmentation of the inner and outer contours, the label is only read in blue channel.



(a) (b)

Figure 3. The original kidney ultrasound image and corresponding label.

B. Metrics

In this paper, we chose the pixel accuracy (PA), mean pixel accuracy (MPA), mean intersection over union (MIoU), and frequency weighted intersection over union (FWIoU) to quantitatively evaluate the segmentation performance of the kidney segmentation algorithm. The calculation is shown as follows:

$$PA = \frac{\sum_i n_{ii}}{\sum_i t_i} \tag{4}$$

$$MPA = \frac{1}{n_c l}\sum_i \frac{n_{ii}}{t_i} \tag{5}$$

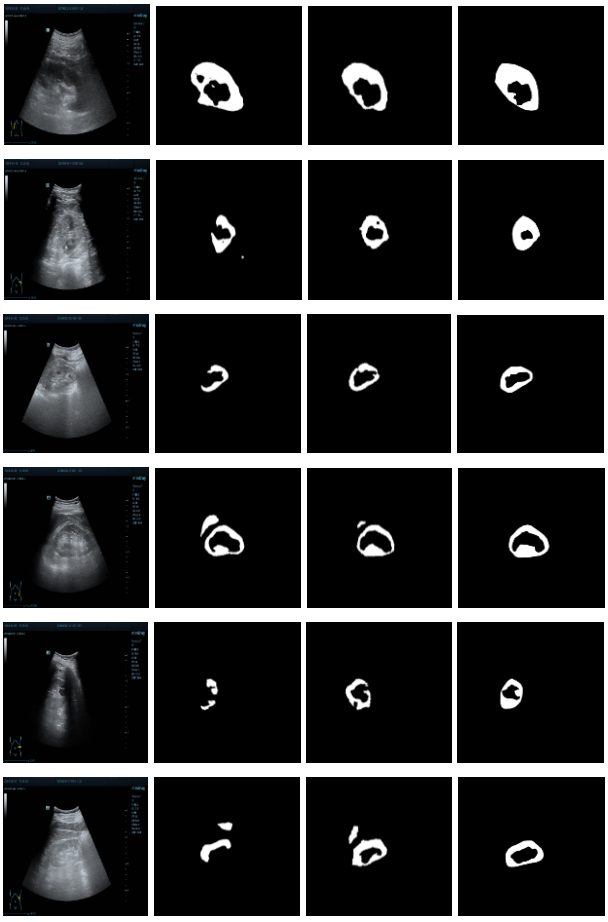
$$MIoU = \frac{1}{n_c l}\sum_i \frac{n_{ii}}{(t_i + \sum_j n_{ji} - n_{ii})} \tag{6}$$

$$FWIoU = (\sum_k t_k)^{-1}\sum_i \frac{t_i n_{ii}}{(t_i + \sum_j n_{ji} - n_{ii})} \tag{7}$$

where $n_c l$ is the number of classes in the ground truth, n_{ij} is the number of pixels in which the class i is predicted to be the class j, and t_i is the total number of pixels in the class i.

C. Results and Discussion

We evaluated the proposed model on the collected kidney dataset. In order to improve the autonomous learning ability of the sub-model, we update the gradient and network parameters of the coarse and fine segmentation models by respectively using Gradient Descent and Adam optimizer [24]. The initial learning rate of the network is set to 0.001. The exponential decay is used to update the learning rate, and the learning rate decay is set to 0.995. The training batch size of the network is set to 1. Before the kidney ultrasound images are sent to the network, the pixel values are normalized to 0-1 and subtract the mean pixel value.



(a) (b) (c) (d)

Figure 4. Segmentation results of Kidney. (a) Image. (b) Original FC-DenseNet. (c) Cascaded FC-DenseNet. (d) Ground Truth.

To confirm the performance of the algorithm proposed in this paper, we compared it to the original FC-DenseNet. Fig. 4 shows the qualitative segmentation results of the two algorithms. Compared with the original FC-DenseNet, the segmentation performance of cascaded FC-DenseNet is better, which is mainly reflected in the following three aspects: the edge preservation is better, the interference from non-kidney regions is less, and the recognition sensitivity is higher for the regions with less noticeable grayscale differences. The segmentation results of original FC-DenseNet are smoother. At the same time, it has good recognition accuracy in the case of easy segmentation.

TABLE II. THE QUANTITATIVE COMPARISONS OF THE PROPOSED AND ORIGINAL FC-DENSENET ALGORITHMS

Methods	PA	MPA	MIoU	FWIoU
FC-DenseNet	0.9881	0.8772	0.8111	0.9785
Cascaded FC-DenseNet	0.9889	0.8908	0.8254	0.9799

The quantitative segmentation results are showed in Tab. 2. The PA on the segmentation test set is 0.9889, the MPA is 0.8908, the MIoU is 0.8254, and the FWIoU is 0.9799. The increase of MIoU could reflect that the algorithm in this paper is more competitive.

In order to verify the role of focal loss in the coarse segmentation model, this paper compares it with cross-entropy loss function (CEL). The segmentation results under different combinations of loss function are calculated. As shown in Tab. 3, the combination of FL and CEL loss function has better performance.

TABLE III. THE QUANTITATIVE COMPARISONS OF THE PROPOSED UNDER DIFFERENT LOSS FUNCTION COMBINATIONS

Combinations	PA	MPA	MIoU	FWIoU
CEL+CEL	0.9874	0.8801	0.8078	0.9777
FL+CEL	0.9889	0.8908	0.8254	0.9799

IV. CONCLUSIONS

In order to realize the automatic segmentation for the inner and outer contours of kidney in whole ultrasound image, a cascaded FC-DenseNet based on joint training is proposed. Cascaded FC-DenseNet is a coarse-to-fine segmentation network. Through the joint training method, two models of coarse segmentation and fine segmentation are obtained. The coarse segmentation model could be seen as a pre-processing operation that could perform functions of denoising and removing redundant information. The fine segmentation model could be seen as a high-performance classifier that extracts the regions belonging to the kidney by extracting high-level semantic features. The proposed method could effectively improve the problem of low segmentation accuracy when FC-DenseNet is trained alone. However, the kidney segmentation in the whole ultrasound image is still challenging. It is a comprehensive problem. If only a single or a few aspects are considered, the segmentation result is not reliable enough. Therefore, the next step is to improve this problem.

ACKNOWLEDGMENT

This work was funded by the National Key R&D Program of China under grant 2018YFC0114500 and National Natural Science Foundation of China (No. 61701089 and No.61601518).

REFERENCES

[1] J. Xie, Y. Jiang, and H. T. Tsui, "Segmentation of kidney from ultrasound images based on texture and shape priors," IEEE Transactions on Medical Imaging, vol. 24, 2005, pp. 45-57.

[2] S. Dahdouh, et al., "SPIE Proceedings [SPIE SPIE Medical Imaging - Lake Buena Vista, FL (Saturday 7 February 2009)] Medical Imaging 2009: Ultrasonic Imaging and Signal Processing - Real-time kidney ultrasound image segmentation: a prospective study," 2009, 7265: 72650E.

[3] M. Kop. Arpana and Hegadi Ravindra, "Kidney segmentation from ultrasound images using gradient vector force," IJCA, Special Issue on RTIPPR, 2010, pp.104-109, Published By Foundation of Computer Science.

[4] W. M. Hafizah, "Comparative evaluation of ultrasound kidney image enhancement techniques," International Journal of Computer Applications, vol. 21, 2011, pp. 15-19.

[5] Xiao-Pei P, “Segmentation of ultrasound kidney images based on fusion algorithm,” Modern Computer, vol. 3, 2016, pp. 27-32.

[6] Q. Zheng, et al., “A dynamic graph-cuts method with integrated multiple feature maps for segmenting kidneys in ultrasound images,” Academic Radiology, 2018.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in Neural Information Processing Systems, 2012, pp. 1097-1105.

[8] C. Szegedy, et al., “Going deeper with convolutions,” in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015, pp. 1-9.

[9] K. He, et al., “Deep residual learning forimage recognition,” in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 770-778.

[10] J. Redmon, et al., “You only look once: unified, real-time object detection,” in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 779-788.

[11] R. Girshick, “Fast R-CNN,” in Proceedings of IEEE International Conference on Computer Vision, Santiago, Chile, Dec. 2015, pp. 1440-1448.

[12] S. Ren, et al., “Faster R-CNN: towards realtime object detection with region proposal networks,” Advances in Neural Information Processing Systems, 2015, pp. 91-99.

[13] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431-3440.

[14] L. C. Chen, et al., “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” IEEE transactions on pattern analysis and machine intelligence, vol. 40, 2018, pp. 834-848.

[15] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE transactions on pattern analysis and machine intelligence, vol. 39, 2017, pp. 2481-2495.

[16] S. Yin, et al., “Subsequent boundary distance regression and pixelwise classification networks for automatic kidney segmentation in ultrasound images,” arXiv preprint arXiv:1811.04815, 2018.

[17] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” International Conference on Medical image computing and computer-assisted intervention, Springer, Cham, 2015, pp. 234-241.

[18] S. Jégou, et al., “The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation,” Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on. IEEE, 2017, pp. 1175-1183.

[19] G. Huang, et al., “Densely connected convolutional networks,” Proceedings of the IEEE conference on computer vision and pattern recognition, vol. 1, 2017, pp. 3.

[20] S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” 2015, http://arxiv.org/abs/1502.03167.

[21] V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” in Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, June 2010, pp. 807-814.

[22] G. E. Hinton, et al., “Improving neural networks by preventing co-adaptation of feature detectors,” Computer Science, 2012.

[23] T. Y. Lin, et al., “Focal loss for dense object detection,” IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 99, 2017, pp. 2999-3007.

[24] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” 2014, http://arxiv.org/abs/1412.6980.