

Image Segmentation Based on Improved Unet

Xiaojin Li, Wenhua Qian*, Dan Xu and Chunyu Liu

School of Information, Yunnan University, Kunming, 650504, China

*Corresponding author email: whqian@ynu.edu.cn

Abstract. In order to help doctors diagnose and treat liver lesions and accurately segment liver images, this paper proposes an improved Unet network, which adds compression extraction modules and full-scale connection blocks, extracts input image features, and achieves accurate segmentation of liver images. The compression extraction module distributes weights to convolutional layers of different sizes, which is conducive to the extraction of image spatial information and context information. Full-scale blocks are connected by skipping, combining the higher semantic information from the decoder and corresponding the lower semantic information from the encoder to strengthen the ability to extract tumor edge information. This article includes 25 cases from the Lits liver dataset. The dataset is classified as the training dataset and the test dataset, and the image blocks are extracted after gray-scale normalization and input to the network to acquire the final segmentation results. The segmentation result is evaluated by F1 score. Comparing multiple sets of experiments, compared with general network structures such as Unet and AttenUnet, it shows that the network architecture proposed in the Dissertation improves the accuracy and efficiency of liver image segmentations.

Keywords: Gray level normalization; Dice score; Segmentation accuracy; The liver.

1. Introduction

1.1. Background

Liver cancer is the deadliest cancers, its mortality rate ranks second in the world, and it is an important factor leading to human death [1]. Generally, early diagnosis of liver cancer can significantly reduce patient mortality. Medical image segmentation of liver is a hot research topic at present. In modern pathology, liver images rely on MRI and CT scans, and different equipment or technologies are used to collect images of the liver. Traditional automatic liver segmentation methods are based on thresholds, regions, edges, etc. [2-4]. These methods have a great influence on subjective factors, which make the traditional segmentation algorithm unable to accurately locate the liver position and cannot fully describe the liver image features, resulting in poor segmentation accuracy, deviations and errors. The segmentation method based on convolutional neural network [5] (CNN) can provide the latest accuracy for different computer vision problems, and has achieved success in different modes in the field of the imaging segmentation, including lung segmentation, diabetes recognition, retina Blood vessel segmentation, brain tumor detection, etc. [6-7]

1.2. Research Work

In 2016, Ben-Cohen et al.[8]used FCN to solve the problem of liver segmentation. Unet uses FCN as the backbone to fully excavate the multi-scale features of the image. Seo et al.added residual paths on the basis of Unet to improve the accuracy of liver segmentation. The above-mentioned neural network realizes the automatic segmentation of liver by independently classifying pixels in the segmentation



process, but there are certain defects, which are mainly manifested in the poor segmentation accuracy of target boundaries and small-volume targets.

1.3. Organizational Structure

The structure of this paper is as follows: Section two mainly introduces the Network structure, and Section three introduces the experiment, including Data set and Evaluation index. Section four introduces the conclusion of this article.

2. Network Structure

The network model proposed in the paper is shown in Figure 1. The model is mainly composed of compressed extraction block, scale connection block, and down-sampling block. The compression extraction module focuses on the relationship between channels, the model can automatically learn the importance of different channel features, and the full-scale connection block connects the high-level compression extraction block and the low-level full-scale connection block through a complete jump, and merges The small-scale feature map in the encoder, the same-scale feature map, and the largest scale feature map from the decoder capture the low-level and the higher level semantic information at the full scale, and restore the feature map. The model encoder path consists of 5 cascaded compressed extraction blocks x^1, x^2, x^3, x^4, x^5 and 4 down-sampling layers, each down-sampling layer is composed of a 3×3 convolutional layer with a step size of 2. The decoder consists of 4 full-scale connection blocks.

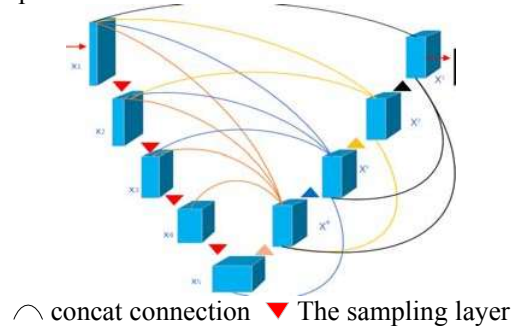


Figure 1. Architecture of the fully connected attention Unet network.

Because Unet relies on fixed weights in the process of image feature extraction, a compressed extraction block[9] is introduced to distribute the weights of feature extraction to all stages of the path. The compressed extraction block is mainly composed of the compressed part and the extracted part. The importance of the channel is predicted, and the importance of different channels is obtained, and then it is fused to the corresponding channel of the previous feature map. Figure 2 shows that the input image $U(h, w, c)$ is first made the global average pooling F_{sq} , and the output data is passed through two-level fully connected F_{ex} , and finally limited to the range of $[0, 1]$ with the sigmoid activation function. The value as F_{scale} is multiplied by the channel c of the input image as the input data of the next stage.

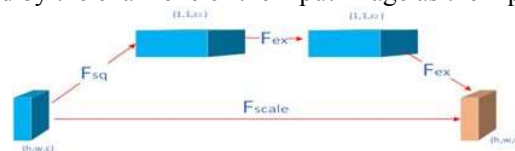
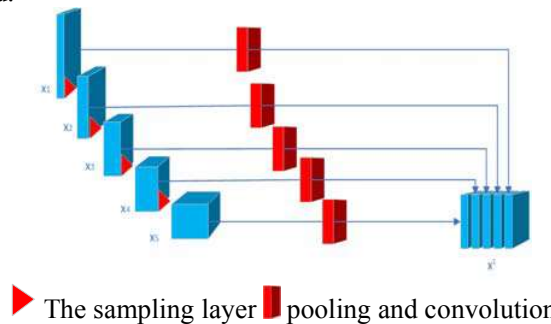


Figure 2. Compressed extraction module diagram.

Figure 3 depicts the $x1$ full-scale connection[10] process; among them, $x1, x2$, and $x3$ pool down the encoder layer through different maximum pooling operations, so as to transfer the low-level semantic information of the bottom layer and unify the size of the feature map. $x5$ is up-sampling by bilinear interpolation to enlarge the resolution of the feature map. In order to unify the size of the feature map, $x1$ must be reduced by 8 times, $x2$ must be reduced by 4 times, $x3$ must be reduced by 2 times, and $x5$ must be enlarged by 2 times. Then, the cascaded feature maps of the five scales of $x1, x2, x3, x4$, and $x5$ are convolved through $64 \ 3 \times 3$ convolution kernels in order to merge the shallow semantic informations and the deep information, and execute the feature Aggregation mechanism. Finally,

through the channel feature fusion of Unet, and through 3*3 convolution and Relu operations, a new feature map x_1 is obtained.



► The sampling layer pooling and convolution

Figure 3. Full scale connection block diagram.

The model in this paper uses binary cross entropy (BCE) as the loss function. The binary cross entropy formula is:

$$l_{bce} = - \sum_{(a,b)} [G T(a,b) \log(SE G(a,b)) + (1 - G T(a,b)) \log(1 - SE G(a,b))] \quad (1)$$

Among them, is the expert label of pixel (a, b) and is the predicted probability of the segmented object. The BCE loss function is pixel by pixel. It does not consider the labels of the neighborhood, and weights both the segmented pixels and the background pixels, which helps the convergence of the loss function.

3. Experiment

3.1. Data Set

This paper selects 25 case images from the LITS2017 [11] data set for experiments. Each patient image contains a liver image and a real segmentation label image. Among them, 20 cases are used as the training set for training, and 5 cases are used as the test set, and the image size is 512×512. The expert's annotation results are used as the gold standard for training and testing.

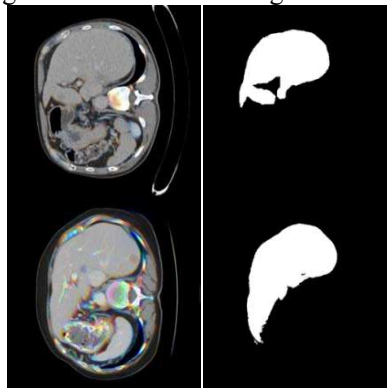


Figure 4. LITS2017 dataset.

The resolution of the image is uniformly cut to 256X256. Because the contrast of the image is not the same, the Z-Score method is used to standardize each image, and the data of different magnitudes are unified into the same magnitude, and the calculated Z -Score value measurement to ensure the comparability between data. Use formula (2) to calculate:

$$x = \frac{x - \mu}{2a} \quad (2)$$

Among them, μ is the average value of the data, σ is the deviation of the overall data, and x is the individual observation value.

3.2. Evaluation Index

For the sake of better experiment, verify the segmentation effect, enhance the data, rotate the image, interchange left and right, zoom in and out, and expand the image.

F1 score is an important indicator to measure the accuracy of two classification problems, and it is the harmonic average of recall and precision. When using F1 score to evaluate model accuracy, the classification result of each pixel will affect the score. The higher the F1 score, the better the accuracy of the resulting change map. Calculate using formula (3)(4)(5).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (5)$$

3.3. The Experimental Results

In order to prove that the network has better results for liver segmentation, Unet, AttenUnet[12], the network of this paper are used to simultaneously segment the Lits data set. Figures 5 show the comparison between the test results of different networks for liver segmentation and the manual annotation results of experts. From top to bottom, each row is the real segmentation label, the segmentation result of this article, the Unet segmentation result, and the Attention Unet segmentations result. It can be seen from Figures five that Unet leads to the lack of liver boundary details and obvious over-segmentation. The liver boundary of AttentionUnet segmentation is more obvious, but the segmented liver boundary edge area still has unsmooth links. The segmentation method in this paper strengthens the image boundary to obtain a segmentation result similar to the segmentation label, and the result at the liver image point is more substantial, With high accuracy, effectively refine the boundaries of the liver. Solve the over-segmentation and under-segmentation problems of other algorithms.

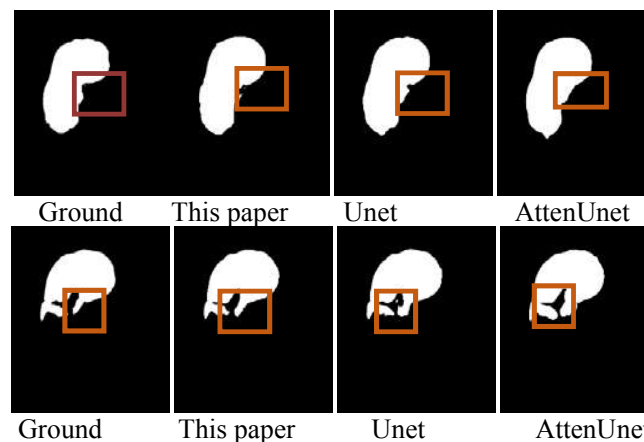


Figure 5. Segmented image.

Table 1 compares the Unet, Attention Unet and the network structure of this article for liver segmentation experiments. It can be known that under the evaluation of relevant indicators, the method proposed in the paper performs better than other network models in liver image segmentation. The lits data set achieves an F1 score of 94.4306%. This article will improve The attention block is embedded in the framework of Unet network, and the high-order consistency between pixel categories is learned

through training, which effectively improves the segmentation accuracy of small scale objective and target boundaries.

Table 1. The evaluation of different network structures on the Lits test data set.

Method	F1 score
Unet	0.943854
AttenUnett ^[16]	0.944306
This paper patarticle	0.945211

4. Conclusion

To solve the problems of missing information and blurring edge in liver segmentation, this paper proposes an improved Unet network architecture for attention mechanism. The feature maps of different levels are fused to guide the feature maps to learn feature expression. The classification task and the segmentation task are combined through the mixed loss function to retain high-level and low-level semantic information and improve the segmentation accuracy.

Acknowledgement

This research was funded by the grants (No. 61662087, 61761046) from the Research Natural Science Foundation of China, the Research Foundation of Yunnan Province (No. 2019FA044), Provincial Foundation for Leaders of Disciplines in Science and Technology (2019HB121), the Postdoctoral fund of the Ministry of education of China (No. 2017M621591), Jiangsu Planned Projects for Postdoctoral Research Funds in 2017.

References

- [1] Bosch F X, Ribes J, Mireia Díaz, et al. Primary liver cancer: worldwide incidence and trends.[J].Gastroenterology, 2004, 127(5-suppl-S1):S5-S16.
- [2] Zheng Zhou, Zhang Xuechang, Zheng Siming, et al. CT liver image segmentation based on region growth and unified level set[J]. Journal of Zhejiang University (Engineering Science Edition), 2018, 52(12):145-159.
- [3] Liu J , Wang S , Linguraru M G , et al. Tumor sensitive matching flow: A variational method to detecting and segmenting perihepatic and perisplenic ovarian cancer metastases on contrast-enhanced abdominal CT.[J]. Medical Image Analysis, 2014, 18(5):725-739.
- [4] Linguraru M G , Richbourg W J , Liu J , et al. Tumor Burden Analysis on Computed Tomography by Automated Liver and Tumor Segmentation[J].IEEE Transactions on Medical Imaging, 2012, 31(10):1965-1976.
- [5] Lecun Y, Boser B, Denker J, et al. Backpropagation Applied to Handwritten Zip Code Recognition[J].Neural Computation, 2014, 1(4):541-551.
- [6] Pang S , Du A , Orgun M A , et al. CTumorGAN: a unified framework for automatic computed tomography tumor segmentation[J]. European Journal of Nuclear Medicine and Molecular Imaging, 2020, 47(10):2248-2268.
- [7] Kumar Singh V , Rashwan H A , Romani S , et al. Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network[J]. Expert Systems with Application, 2020, 139(Jan.):112855.1-112855.14.
- [8] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4):640-651.
- [9] Jie, Hu, Li, et al. Squeeze-and-Excitation Networks.[J]. IEEE transactions on pattern analysis and machine intelligence, 2019.

- [10] Huang H , Lin L , Tong R , et al. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation[C]// ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.
- [11] Bilic, Patrick , et al. "The Liver Tumor Segmentation Benchmark (LiTS)." (2017).
- [12] Oktay O , Schlemper J , Folgoc L L , et al. Attention U-Net: Learning Where to Look for the Pancreas[J]. 2018.