A Comparison of U-Net Series for CT Pancreas Segmentation

Yufei Wang 23020221154122 Informatics Class, Linya Zheng 23020221154182 Informatics Class, Fan Zhang 23020221154180 Informatics Class, Yayao Hong 23020221154161 Informatics Class

Abstract

The incidence of pancreatic diseases has been increasing in recent years. Especially, pancreatic cancer is the most malignant tumour in the digestive system with an extremely low survival rate and no significant improvement in the overall survival of patients in recent years. Therefore, the early detection and diagnosis of pancreatic cancer is of great clinical significance. With the advances in medical imaging equipment and image processing techniques, deep learning convolutional neural networks have achieved a great success in the field of medical image segmentation. These networks are able to achieve an accurate organ segmentation by deeply learned high-level features. Using convolutional neural networks to segment the pancreas from medical images and extract meaningful regions of interest can help characterise and identify tissue and improve the diagnosis and treatment of pancreatic diseases. This thesis introduces and modifies U-Net series for pancreas segmentation and aims to solve the problems of large variation in shape and location, and complex background of many neighbouring tissues in the pancreas segmentation.

Introduction

The pancreas is a glandular organ found in the abdomen and consists of endocrine and exocrine tissues that play a vital role in glucose metabolism and digestion respectively (Pandey, Tekchandani, and Verma 2020).

The incidence of pancreatic cancer is increasing, yet pancreatic cancer remains difficult to treat, and most pancreatic cancers are diagnosed late, missing the best time for treatment. In recent years, the number of patients with pancreatic cancer has increased each year and the overall 5-year survival rate is low. The overall prognosis for pancreatic cancer has not improved significantly compared to other malignancies (e.g. liver cancer, lung cancer, etc.) and the treatment outcomes are not ideal. However, clinical breakthroughs in the early detection and diagnosis of pancreatic cancer have been relatively lacking. Often, when abdominal pain, weight loss or yellowing of the skin (jaundice) occurs, the cancer is already at an advanced stage and the best time for treatment is missed. Therefore, the early diagnosis and treatment of pancreatic cancer is of great research importance and value.

Pancreatic cancer is usually diagnosed by medical imaging in the form of ultrasound or computed tomography (CT) scans. With the rapid development of computer image processing and pattern recognition and imaging histology technology, early diagnosis of pancreatic lesions based on big data analysis of medical images has become an important way and method to improve the early diagnosis rate of pancreatic cancer today. Based on the results of these processes and analyses, the early diagnosis of pancreatic disease can be significantly improved and more efficient and accurate treatment can be implemented.

With the rapid development of deep neural networks in recent years, we have witnessed many achievements in the field of medical image analysis and computer-aided diagnosis (CAD). For example, we can automatically segment organs from CT scans. Pancreatic segmentation can help in clinical diagnosis and treatment, but the difficulties arise mainly from the high anatomical variability or the small size of the target organ (Zhou et al. 2017). Among the different abdominal organs, pancreatic segmentation is particularly difficult and presents a significant technical challenge. The main reason for this is the complex anatomy of the pancreas, which extends from behind the stomach to the left upper abdomen near the spleen. Factors such as the shape, size and location of the pancreas are influenced by gender, age and obesity. The pancreas is also surrounded by a large number of other tissues such as the stomach, spleen, blood vessels and duodenum. These tissues are closely connected to the pancreas in the CT image, the grey scale contrast is not obvious and the pancreas is only a small part of the total CT volume. In this case, the pancreas segmentation is likely to be interfered with by the background, and the segmentation results become inaccurate, especially when the border areas of the pancreas are difficult to identify.

Convolutional neural networks are highly accurate and robust, and have made great achievements in the field of medical image segmentation. U-Net (Ronneberger, Fischer, and Brox 2015) convolutional neural networks are commonly used for image segmentation and are based on fully convolutional networks, with a modified and extended architecture that can handle fewer training images and produce more accurate segmentations. In this paper, various convolutional neural networks with U-Net architecture are used to segment pancreatic images and compare the effectiveness of different methods for pancreatic segmentation.

Related Work

Medical image segmentation Medical image segmentation is to segment the target areas such as organs, lesions and surgical tools in medical images. Compared with natural image segmentation, medical image segmentation has the advantages of clear target, fixed structure, and large overall image similarity. However, due to the low imaging quality of medical imaging equipment, medical images also have problems such as high noise, fuzzy boundary, and low resolution. Deep learning-based approaches are popular and state-of-the-art for medical imaging applications. Deep learning models have a hierarchical architecture, consisting of many layers. These layers are used for self-feature extraction. Full Convolution Network (FCN) is one of the most representative research works. It realizes end-to-end training of segmentation network when the size of input image is the same as that of output image. It can segment images at pixel level and significantly improves the effect of image segmentation. Inspired by FCN, in view of the rich spatial information of medical images, in order to avoid the loss of spatial information in the process of sampling under the network, researchers began to gradually apply the network based on the encoder decoder structure, such as U-Net (Oktay et al. 2018).

Pancreas Segmentation Deep convolution neural networks are widely used to segment the pancreas from CT images in the literature. Zhou et al (Zhou et al. 2017) shrinked the input pancreas region to predict the segmentation mask by a fixed-point model for pancreas segmentation. In earlier work, Zhou et al (Zhou et al. 2017) proposed a coarse-to-fine approach based on different neural network models. The coarse-scale model takes the whole body as its input and roughly localises the pancreas. In addition, the fine-scale model takes as input a small area covering the pancreas, which helps to eliminate background noise and thus provides better segmentation accuracy, especially in the border region. However, the algorithm deals with two separate stages, lacks optimisation of the global energy function and limits its ability to integrate multi-stage visual cues. Yu et al (Yu et al. 2018) presented a recurrent saliency transformation network with a saliency transformation module that uses a probability map to update weights. The saliency transformation module is their key innovation. While Bobo et al (Bobo et al. 2018) applied fully connected convolutional neural networks to segment the pancreas from magnetic resonance images, Ozan et al (Oktay et al. 2018) embedded an attention gate module into U-Net to suppress irrelevant regions in CT image. Hao Li et al (Li et al. 2019) proposed a pancreas segmentation model based on spatial context information (SCU-Net) and bidirectional convolutional long and short-term memory (BiCLSTM) networks. The BiCLSTM network was introduced to stimulate the interaction between the information streams from bidirectional sequences. A new loss function for inter-slice restriction and regularisation was also introduced. More recently, Li et. al. (Li et al. 2021) developed a probabilistic map guidance based bi-directional recurrent U-Net for pancreas segmentation.

Proposed Solution

This section describes U-Net based pancreas segmentation methods in details. We first segment the pancreas using U-Net, U-Net++ and U-Net3+ networks and compare their segmentation results. Then, we replace standard convolution by depthwise separable convolution so that it improves U-Net series with better segmentation accuracy and reduced parameters.

Data Augmentation

Since deep learning-driven medical image segmentation often requires a large amount of data for training to obtain a model with strong generalization capability, we employ three data augmentation methods of contrast adjustment, Gaussian filtering, and warping to increase the size of training sets.

Contrast adjustment Adjusting the contrast of a picture mainly solves the problem of low contrast caused by a small range of image gray levels, which serves to enlarge the range of image gray levels, thus making the image clearer.

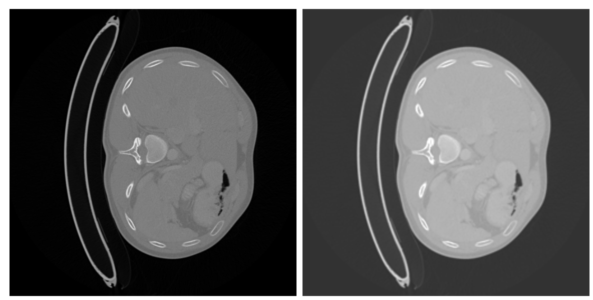


Figure 1: Contrast adjustment

Gaussian filtering Gaussian filtering is a linear smoothing filter, which is suitable for eliminating Gaussian noise and is widely used in the noise reduction process of image processing, in which the whole image will look blurrier compared with the original image, and some noise points on the original image will be eliminated after the weighted sum of surrounding pixels.

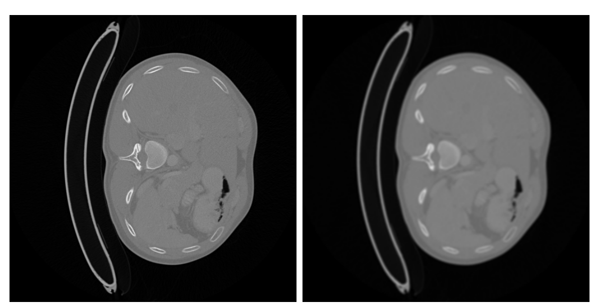


Figure 2: Gaussian filtering

Warping A coarse displacement grid with random displacement is generated for each grid point, then this grid is interpolated to calculate the displacement of each pixel in the input image, and finally the input image is deformed using the displacement vector and the spline difference.

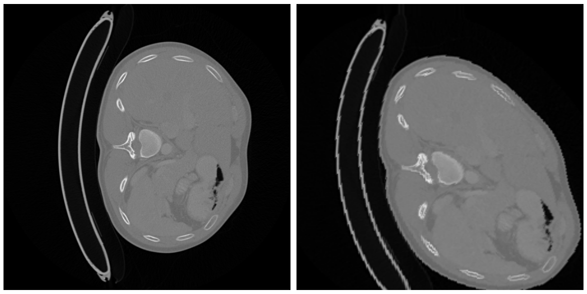


Figure 3: Warping

U-Net Architectures

Figure 4 shows the architectures of U-Net series.

U-Net consists of an encoder and an decoder and has lowresolution features and limited receptive field. U-Net++ contains one encoder path and several decoder paths.

Particularly, U-Net++ reshape skip connections to accumulate features at multiple scales on the different decoder paths. While U-Net++ implements embedded and dense skip connections, it obtains multiscale but not fullscale features. U-Net3+ employs fullscale skip connections to combine low-level details with high-level semantics from input feature maps as well as it implements deep supervision to achieve hierarchical feature representations from fullscale combined feature maps, which can simultaneously improve accuracy and reduce training parameters.

More importantly, we embed depthwise separable convolution into U-Net, U-Net++, and U-Net3+ instead of standard convolution to improve segmentation accuracy and reduce training parameters. Figure 5 shows an example of the conversion of a standard convolution into a depthwise separable convolution. All three architectures of U-Net+DWC, U-Net++DWC, and U-Net3+DWC will be demonstrated to significantly reduce the amount of training parameters and to improve the average dice coefficient in pancreas segmentation. Additionally, we use the connected components operator to remove noise or small false-positive segmentation.

Model Building

Figure 6 illustrates the comparison pipeline of pancreas segmentation. We first perform data augmentation on pancreatic CT images and their corresponding labels, and then we use these images and labels as input to our U-Net series of models for training. After training these models, we input other pancreatic CT images into the trained model to segment the pancreas in the images.

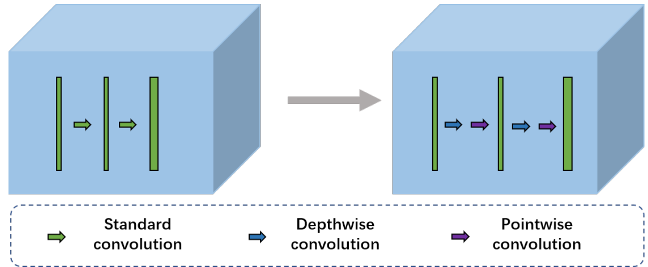


Figure 5: Depthwise separable convolution consists of depthwise and pointwise convolutions.

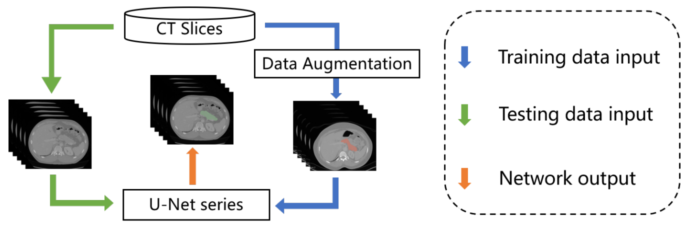


Figure 6: Pipeline to compare various U-Net architectures for pancreas segmentation

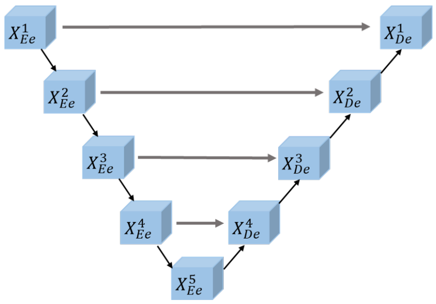
Experiments

Dataset

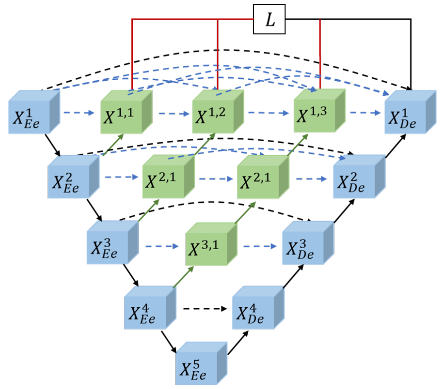
Our experiments use datasets from the National Institutes of Health Clinical Center These datasets contains 82 contrast CT images and their corresponding annotated labels. These CT scanning parameters were of 512×512 pixels and the slice thickness ranges from 1.5-2.5 mm. We separate these data into 62 and 20 CT volumes for training and testing, respectively. We also performed data enhancement on the training datasets.

Post-process

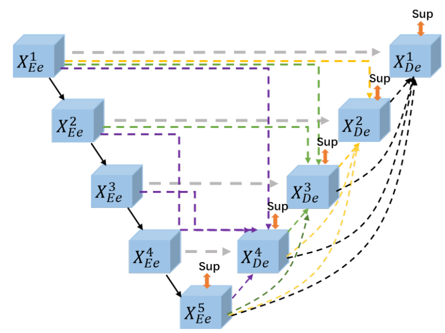
At the beginning of the experiment, we found that when the trained model made predictions, there was unnecessary noise in the output results. Noise will not only lead to loss of accuracy when evaluating model segmentation results, but also may cause unnecessary misdiagnosis if the results with noise are directly applied to the clinic. In the results of the whole pancreas segmentation, the noise appears as some connected components that are much smaller than the segmented pancreas, so we added an operation to remove too small connected components on the basis of the predicted results. In actual operation, the final result after noise reduction is 1% - 2% higher than the original result in average DSC. As shown in Figure 7, the DSC of Case_82 before noise reduction is 69.51%. After noise reduction processing, unnecessary noise is removed, and the DSC increases to 75.39%.



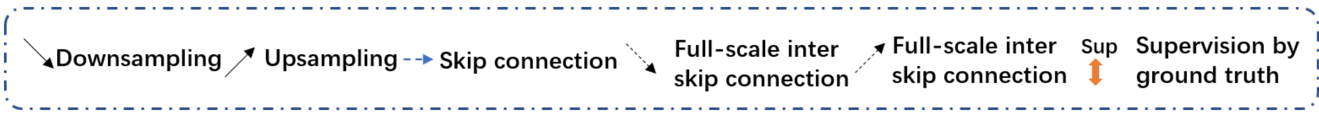
(a) U-Net (Ronneberger, Fischer, and Brox 2015)



(b) U-Net++ (Zhou et al. 2018)



(c) U-Net3+ (Huang et al. 2020)



(d) Operations used in U-Net series

Figure 4: Architectures of U-Net, U-Net++ and U-Net3+ with various operations

Table 1: Comparison of the performance of the six pancreas segmentation methods

Approaches	U-Net	U-Net++	U-Net3+	U-Net+DWC	U-Net++DWC	U-Net3+DWC
Dice (%)	74.46	73.90	75.55	**76.13**	75.41	75.77
Precision (%)	74.43	72.16	75.13	72.76	74.51	77.03
Recall (%)	77.11	78.58	77.67	**81.13**	78.58	76.58
Parameters	39.4M	47.2M	27.0M	14.3M	18.4M	**3.15M**

Before After

Figure 7: The compared segmentation results of before and after using postprocessing.

Results and Analysis

Figure 8 compares the segmented pancreas regions of using the different methods. Figure 9 illustrates the segmentation result of using our proposed pancreas segmentation approach. We also use the dice similarity coefficient, precision, and recall to quantitatively evaluate the segmentation results. Table 1 summarizes the quantitative results of the compared six pancreas segmentation approaches. It can be seen that the U-Net3+ model has the best average segmentation effect among the three models and is the most stable, followed by U-Net, and U-Net++ is slightly inferior in pancreas segmentation. From the perspective of network parameters, U-Net3+ effectively reduces the number of parameters of the U-Net network, while the U-Net++ network has the largest number of parameters, but since both U-Net++ and U-Net3+ networks use deep supervision methods, so The training is still more time-consuming than the training of the U-Net network. After using the depth separable convolution, the segmentation result of U-Net+DWC surpassed the segmentation result of U-Net3+ network and its improved U-Net3+DWC network, and the average DSC reached 76.13%. In terms of parameters, the parameters of the three improved networks have been greatly reduced compared with the original network. The network parameters of U-Net3+DWC are the lowest among all networks, and U-Net++DWC is The largest number of parameters among the three improved networks. The same as the original network, in terms of time, since U-Net+DWC does not use a deep supervision module, U-Net+DWC has shorter training time than U-Net++DWC and U-Net3+DWC.

More specifically, the training time of U-Net is about 7 hours, while U-Net++ takes about 13 hours, and U-Net3+ takes about 25 hours. So in terms of time cost and effect, the U-Net network still has great advantages. Although U-Net3+ has higher segmentation results than U-Net, the training time is as high as 3 times that of U-Net. However, U-Net++ is not outstanding enough in pancreas segmentation in terms of

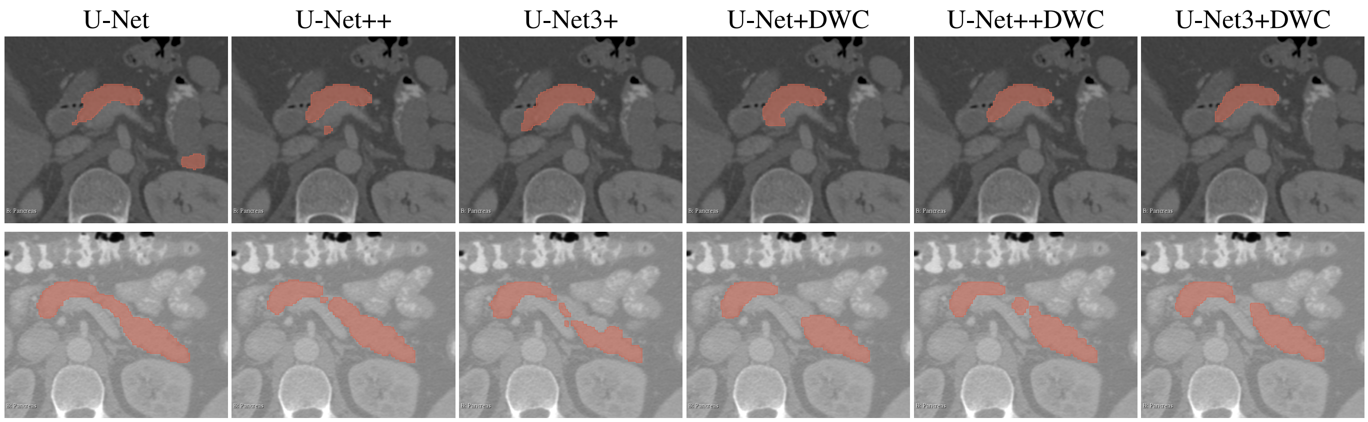
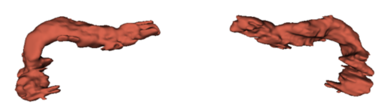


Figure 8: The compared segmentation results of using the six pancreas segmentation methods

(a) U-Net

(b) U-Net++



(c) U-Net3+

(d) U-Net+DWC

(e) U-Net++DWC

(f) U-Net3+DWC

Figure 9: Visual comparison of the 3D results of using the six pancreas segmentation methods

training time and training effect. For the network improved by using depth separable convolution, the segmentation results of the U-Net series network have been improved to a certain extent. Although the training time of U-Net+DWC network and U-Net++DWC network is slightly increased compared with the original network, the effect can be increased by 1.67% and 1.51%, which have great advantages.

Without DWC, U-Net3+ outperforms the other two networks. DWC generally improves the performance of U-Net series, particularly, U-Net+DWC works better than the other five segmentation methods. This implies that depthwise separable convolution in U-Net works much more effectively than redesigned dense skip connections in U-Net++ and fullscale skip connections and deep supervision in U-Net3+. Additionally, the computational efficiency was significantly improved by DWC.

Conclusion

This work thoroughly compares the performance of 2D U-Net, 2D U-Net++ and 2D U-Net3+ for CT pancreas segmentation. The U-Net network is one of the classic networks in the field of semantic segmentation. It has achieved good results in natural image segmentation. This work attempts to apply the U-Net to the specific experiment of pancreas segmentation, and explore the effects of U-Net network, U-Net++ and U-Net3+ in the field of medical image processing, especially the effect of pancreas segmentation. While these 2D network architectures work well, U-Net3+ generally outperforms the other two architectures. More interestingly, we introduce depthwise separable convolution instead of standard convolution into the three network models. Such a modification can significantly reduce the network parameters and improve the average dice similarity coefficients. More importantly, this work also finds that depthwise separable convolution can more effectively improve U-Net compared with redesigned dense skip connections in U-Net++ and fullscale skip connections and deep supervision in U-Net3+. Our future work is to explore a deep learning coarse-to-fine 3D networks in accordance with this comparative study to further improve pancreas segmentation.

References

Bobo, M. F.; Bao, S.; Huo, Y.; Yao, Y.; Virostko, J.; Plassard, A. J.; Lyu, I.; Assad, A.; Abramson, R. G.; Hilmes, M. A.; et al. 2018. Fully convolutional neural networks improve abdominal organ segmentation. In *Medical Imaging 2018: Image Processing*, volume 10574, 750–757. SPIE.

Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.-W.; and Wu, J. 2020. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1055–1059. IEEE.

Li, H.; Li, J.; Lin, X.; and Qian, X. 2019. Pancreas segmentation via spatial context based u-net and bidirectional lstm. *arXiv preprint arXiv: 1903.00832*.

Li, J.; Lin, X.; Che, H.; Li, H.; and Qian, X. 2021. Pancreas segmentation with probabilistic map guided bi-directional recurrent UNet. *Physics in Medicine & Biology*, 66(11): 115010.

Oktay, O.; Schlemper, J.; Folgoc, L. L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N. Y.; Kainz, B.; et al. 2018. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Pandey, S.; Tekchandani, H.; and Verma, S. 2020. A literature review on application of machine learning techniques in pancreas segmentation. In *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*, 401–405. IEEE.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Yu, Q.; Xie, L.; Wang, Y.; Zhou, Y.; Fishman, E. K.; and Yuille, A. L. 2018. Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8280–8289.

Zhou, Y.; Xie, L.; Shen, W.; Wang, Y.; Fishman, E. K.; and Yuille, A. L. 2017. A Fixed-Point Model for Pancreas Segmentation in Abdominal CT Scans. In Descoteaux, M.; Maier-Hein, L.; Franz, A.; Jannin, P.; Collins, D. L.; and Duchesne, S., eds., *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, 693–701. Cham: Springer International Publishing. ISBN 978-3-319-66182-7.

Zhou, Z.; Rahman Siddiquee, M. M.; Tajbakhsh, N.; and Liang, J. 2018. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, 3–11. Springer.