RU-Net for Heart Segmentation from CXR

Yu Lyu*, Wei-Liang Huo and Xiao-Lin Tian
Macau University of Science and Technology, Avenida WaiLong, Taipa, Macau, China

*Corresponding author's email: 1909853gii20002@student.must.edu.mo

Abstract. Cardiovascular disease is one of the top causes of death in the world. In order to release heavy workload for doctor, automated segmentation methods using deep learning are proposed by researchers. Due to limitation of medical images, we proposed a novel model RU-Net based on the combination of U-Net and Residual Network for heart segmentation. We replaced Res path from direct skip connection from encoder to decoder. We use Jaccard similarity coefficient to compare the result of our method and U-Net with public dataset called Japanese Society of Radiological Technology (JSRT). The experiment result demonstrates the accuracy of our method.

Keywords. Chest X-ray radiography, heart segmentation, U-Net, Res path, convolutional neural network

1. Introduction

According to the World Health Organization, heart disease is the lead killers worldwide. Heart segmentation represents an essential step in cardiovascular therapy. A research [1] shows that asymptomatic patients with atrial fibrillation have stoke more than twice as many as symptomatic patients with atrial fibrillation in 2017. There are a lot of way to diagnose for heart disease, such as Positron Emission Tomography (PET), Chest X-ray radiography (CXR), Magnetic Resonance Images (MRI), Computer Tomography (CT), and etc.

Chest X-ray radiography is commonly modalities of medical images in routine physical examination due to cheap, fast and low dose of radiation. CXR includes ribs cage, lungs filed, heart and great vessels. Manual labeling in the massive medical images is time consuming. Nowadays, low resolution and 2-D projection causes challenges of chest radiograph diagnose. Furthermore, high accuracy and reliable segmentation of medical image is needed to support clinical practitioners. The segmentation of CXR provides some information about shape and size of heart to evaluate clinical conditions including Pleural effusion, Cardiac hypertrophy, Pneumothorax and Emphysema. This enablesdoctor to detect and treat some disease of heart with organ segmentation. Therefore, image segmentation plays a profound role of disease diagnosis.

Usually, CXR uses to diagnose for lung disease because of clear structure of lung. In recent years, doctor use experience and observation to diagnose common heart disease, such as Cardiac hypertrophy (enlargement of heart), pulmonary hypertension and etc. However, doctor need to deal with a massive number of CXR at work. To deal with this situation,
This work was funded by The Science and Technology Development Fund, Macau SAR (File no. 0009/2018/A).

We use some image processing for heart segmentation for heart diagnose for low prize than other diagnose method for heart. Medical image segmentation and classification is becoming the focus of research in the recent years because people pay attention about their health. There is a large amount of paper for medical image segmentation for heart segmentation and lung segmentation by CXR. The challenging task that heart segmentation by using CXR is limitation of medical image scale. Dataset for medical image with accurate annotation is generally small due to doctor’s heavy workload. CXR is difficult for segmentation due to overlap lung fields, ribs and heart mask with blurred border.

For overcoming these difficulties, we propose an accurate and efficient heart segmentation model in this paper. Our architecture combined with U-Net and Residual Network, while we made a couple of vital improvements over U-Net model by adding a res path and changing the processing before max pooling and after up-sampling. In our paper, our model can reach higher accuracy than U-Net efficiently. We use Jaccard similarity coefficient to compare the result of our method and U-Net with public dataset called Japanese Society of Radiological Technology (JSRT). The experiment result shows the comparison of accuracy of our method and U-Net.

The composition of our paper is as follows. Section II reviews the most common methods for medical images segmentation and classification. In Section III, we demonstrates our Res path and the architecture of RU-Net. In Section IV, we compare the experimental results of our method and U-Net. Section V shows the conclusion of our paper.

2. Related work

On account of importance of image segmentation and complexity of manual label, a massive of automated segmentation methods [2] are widely explored. In recent years, medical image segmentation and classification are becoming more and more popular topics [3]. Many researches in deep learning for medical image segmentation and classification have explored for brain [4], liver [5], lung [6], pelvis [7] and heart [8].

With the recent advancement in deep learning, a vast number of approaches use convolutional neural networks (CNN) [9] for biological image segmentation [10]. CNN (2012) used a sliding window method to predict the semantic tag of each pixel and obtained the best performance. However, the drawback of CNN is obvious, which cost much time for calculation and finding a suitable and fixed patch size. Jonathan Long et al. proposed an architecture called Full Convolution Networks (FCN) [11] with end-to-end architecture to reduce image loss from convolution and pooling operation. They use full convolutional layers instead of full connected layers to improve efficiency for reduce calculation with arbitrary input. After convolutional operation, they obtained an image with low resolution and high dimensional feature called heatmap. The result of segmentation and classification is same size as input image by deconvolutional operation. In the same year, [12] proposed U-Net to segment with multiple organs based on CNN and gained a terrific performance. They found this delicate and small model perform an excellent result for medical image segmentation. They use data augmentation for shift and rotation to enlarge the dataset which contains a scarce amount of data for only 30 images. The strong use of data augmentation avoids overfitting for deep learning and train a good model for medical biological image segmentation. U-Net is similar to FCN which begins deconvolutional operation to up sampling at fifth step for high dimensional feature and low dimensional feature fusion.

Figure 1 shows U-Net architecture, which is an encoder and decoder network with symmetric structure. The contracting path is regard as encoder for capturing feature. The expanding path is regard as decoder for precise location.
Figure 1. U-Net architecture. It is encoder-decoder architecture with skip connection for corresponding layers. In the part of encoder, every step consists of two 3x3 convolution followed ReLU and 2x2 max pooling. In the part of decoder, every step consists of 2x2 deconvolution and two 3x3 convolution followed ReLU. The skip connection is plain short cut for corresponding layer before and after the max pooling operation and deconvolutional operation.

The reduced sampling path must first go through two 3x3 convolutions, and then perform rectified linear unit (ReLU) [13] and 2x2 maximum merge operation. For every step for expanding path of up sampling, it composes a 2x2 deconvolution and repeated two 3x3 convolution followed a ReLU. They used a 1x1 convolutional operation at the final layer to map the feature vector to the required classes. Skip connection connects the corresponding layers before and after the max pooling operation and deconvolutional operation for fusion of high dimensional feature and low dimensional feature. However, U-Net has three drawbacks. Firstly, it is too much information with low resolution kept in the skip connection. The extracted image is blurred due to repeated information of low resolution. Secondly, it is blurred edge due to low resolution. Thirdly, it is difficult to optimize extracted feature of high level for pooling operation. In a U-Net based segmentation architecture, the delineation of the heart would be uncertainty due to low resolution.

Badrinarayanan V et al. proposed a novel architecture based on FCN called SegNet [14]. SegNet contains encoder network to extract feature and decoder network to predict classification of pixels. Chen L C et al. proposed DeepLab [15] and shown great improvement from SegNet by using atrous convolution instead of down sampling.

3. Methodology
In this section, we demonstrate the architecture details of the proposed methods for heart and lung segmentation of CXR image.

3.1. Res path
In the U-Net, skip connection between the corresponding layers of encoder and decoder is plain direct connection. Skip connection connects the corresponding layer between encoder and decoder. Although it makes low dimensional feature and high dimensional feature fusion, it could not address different scale of medical image segmentation and classification. Inspired by inception blocks [16], we use 1x1 convolution to handle multi-scale medical image processing.

Max pooling operation would loss some information of image and keep high level feature of medical image. There is some diversity of feature between encoder and decoder needed to be alleviated. Hence, we put some convolutional operations in the skip connection. Therefore, instead of direct connection between encoder and decoder, we use a chain of 3x3 convolutional operation with a residual connection for 1x1 convolutional operation and concatenate for four times with decoder information. Figure 2 shows our improved skip connection and call it Res path.
3.2. The architecture of RU-Net

The architecture of the RU-Net is shown in Figure 3. The proposed model called U-Net combined with residual network (RU-Net) encompasses encoder and decoder. Different from U-Net, we made some important improvement by residual path [17] and concatenation operation. The U-Net contracting path contains a sequence of two 3x3 convolution operations with a max pooling operation as follow for four times. Despite that this modification gave great performance than slide window, it is still quite demanding for clear border for organ segmentation. In order to keep relation between our model and U-Net architecture, we calculate the value of parameter W as following:

$$W = \theta \times U$$

Parameter U represents the number of filters in the corresponding layers of U-Net and $\theta$ is scaler coefficient. We compare RU-Net and U-Net easily for decomposing W to 0 by controlling the number of parameters. We choose $\alpha = 0.28$ for little lower of parameters than U-Net. Increasing filters in the corresponding convolutional layers performs better result than equal number of filters. Therefore, we assign [W], [2W], [6W] filters for successive convolutional layers respectively for our best result.

3.2. The architecture of RU-Net

For every step of down sampling, we take three output from three 3x3 convolution layers and concatenate them together. For the sake of improving the efficacy in medical image segmentation to obtain spatial feature, we take non-linear addition with 1x1 convolutional layers. After that, we use 2x2 max pooling operation to extract high dimensional feature. On the contrary for up sampling, we use 2x2 up sampling operation for enlarge feature image. Then we take same method as every step of down sampling before max pooling for image processing. There are four steps for up sampling operation and down sampling operation respectively. In order to map the feature, we use 1x1 convolutional layer at the
last layers. Res path is used to replace the skip connection with res path to alleviate the diversity of feature between decoder and encoder.

In our RU-Net model, we use concatenation of three 3x3 convolutional layer with residual connection instead of using 3x3 convolutional layer for three times to reduce calculation and bottleneck. All the aviation operation except last layer is ReLU aviation function and batch normalization [18]. The last layer of aviation function is sigmoid aviation function.

4. Experiments
The RU-Net has been implemented using Keras with the well-known tensorflow framework. An Intel Xeon E5-2698 v4 2.2GHz(20-Core) CPU and one Tesla V100(32 GB memory for each GPU) is used to implement the experiment. Since RU-Net is improved on the famous U-net network, RU-Net and U-net are both experimented on the dataset mentioned above.

4.1. CRX images pre-processing and dataset
CXR images are with low contrast and blurred border for organ segmentation. In this paper, the adaptive histogram equalization (CLAHE) is used for image enhancement.

Japanese Society of Radiological Technology (JSRT) [19] is public dataset of CXRs with accurate annotation for lung fields and heart masks. JSRT encompasses 154 nodule and 93 non-nodule 12-bits grayscale images with high resolution 2048x2048. As we known, machine learning needs massive data to optimize the model and avoid overfitting. Due to the limited dataset of JSRT has only 247 images, image augmentation is needed.

Data augmentation is a common method in image processing of medical image segmentation and classification, because the public dataset with accurate annotation is commonly small to cause overfitting. The common data augmentation divides into basic image manipulations [20] and deep learning approaches [21]. In this paper, we use some basic image manipulations method for data augmentation [22]. We use geometric transformations and rotation for data augmentation to increase the data from JSRT.

4.2. Training methodology
In fact, our method is a kind of semantic segmentation and pixels belonging to the same category are all classified into one category. RU-Net uses the ADAM to train the model and adaptively calculate different learning rates for different parameters. Since in semantic segmentation pixels of the same type only take up a small part of the entire image, it is far from enough to judge the segmentation results by indicators such as recall rate. In this article the Jaccard coefficient is used to compare the similarity and difference between CXR datasets. The standard evaluations metrics can be expressed mathematically as follow:

\[
J(A, B) = \frac{|A \cap B|}{|A \cup B|}
\]  (1)

We proposed a dataset A for our heart segmentation result and a dataset B for training with accurate annotation. Jaccard is used to compare the similarity and diversity of our heart segmentation result and training with accurate annotation. The value of Jaccard index is equal to the value of similarity between two datasets. The high value of Jaccard similarity coefficient means good performance for our model with high similarity of our result and accurate segmentation.

4.3. k-fold cross validation
Cross validation is to build models and verify model parameters. The acquired sample data is separate into different training datasets and testing datasets. The training dataset is used to train the model, and the testing dataset is used to evaluate the prediction of the model. On the multiple datasets of different training datasets and testing datasets can be acquired. A sample in a training dataset possibly becomes a sample in the testing dataset for the next time. S-fold cross-validation divides the sample data into A randomly. For each time, A1 is randomly selected as the training set, and the other is used as the testing
dataset. When this round is completed, randomly select A1 copies as the training dataset. After many rounds, we select the loss function to evaluate the different model and parameter.

4.4. Results

The experiment uses five-fold cross-validation method and following table shows the results of 5-fold cross-validation for RU-Net and U-net. From TABLE I the results demonstrate that using RU-Net, it is highly possible to acquire superior results in less training epochs comparing with U-Net architecture.

| Model   | Parameters | Evaluation scheme | Jaccard (%) |
|---------|------------|-------------------|-------------|
| U-net   | 31,084,008 | 5-fold CV         | 87.7828±0.4857 |
| RU-Net  | 30,482,465 | 5-fold CV         | 85.5774±0.0028 |

In order to improve readability, the Jaccard Index values have been converted into percentage ratios (%). Figure 4 shows that RU-Net is superior to the basic U-Net when segmenting the medical images of dataset.

![Figure 4. Comparison of different segmentation method result curve](image)

![Figure 5. Initial CXR image (a), CXR image segmentation results (b), RU-Net segmentation results (c) and U-net segmentation results (d). Both the results seem to the results close to the CXR image segmentation results (b).](image)
Since the model has some gradient fluctuations, the final convergence accuracy is obviously better than U-net. Figure 5 shows part of the segmentation result. From left to right is the Initial CXR image, CXR image segmentation results, well-known U-net result and proposed result.

The most important thing for image segmentation is its boundary part. As the state-of-the-art model for medical biological image segmentation, the proposed model has shown favorable results in the experiments. However, there are still many medical images shown bad result for both model, and our results are slightly better than U-net they are still unconvincing. Figure 6 shows the problem mentioned above and the results for RU-Net segmentation and U-Net segmentation. It will be the direction of our future research and optimization.

5. Conclusion
In this paper, the model RU-Net based on U-Net with res path and complicated concatenation with different size of convolution. The JSRT public dataset is used to evaluate the proposed method by Jaccard similarity coefficient. The segmentation results demonstrate that our model can achieve better performance segmentation results than U-Net.

6. References
[1] Christow S P, et al. I1669Increased rate of previous stroke in asymptomatic/minimally symptomatic versus symptomatic patients with newly detected atrial fibrillation in western Europe - results from the GLORIA-AF registry EP Europace iii355-iii356
[2] Shin H C, Roth H R, Gao M, et al 2016 Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning IEEE transactions on medical imaging 35(5) 1285-1298 https://doi.org/10.1109/TMI.2016.2528162
[3] Schindelin J, Rueden C T, Hiner M C, et al. 2015 The ImageJ ecosystem: An open platform for biomedical image analysis Molecular reproduction and development 82(7-8) pp 518-529 https://doi.org/10.1002/mrd.22489
[4] Acharya U R, Oh S L, Hagiwara Y, et al. 2018 Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals Computers in biology and medicine 100 270-278 https://doi.org/10.1016/j.compbiomed.2017.09.017
[5] Li X, Chen H, Qi X, et al. 2018 H-DenseUNet: hybrid densely connected U-Net for liver and tumor segmentation from CT volumes IEEE transactions on medical imaging 37(12) pp 2663-2674 https://doi.org/10.1109/TMI.2018.2845918
[6] Shen W, Zhou M, Yang F, et al. 2017 Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recognition 61 pp 663-673 https://doi.org/10.1016/j.patcog.2016.05.029
[7] Lei Y, Wang T, Tian S, et al. 2020 Male pelvic multi-organ segmentation aided by CBCT-based synthetic MRI Physics in Medicine & Biology 65(3) 035013 https://doi.org/10.1088/1361-6560/ab63bb
[8] Dou Q, Yu L, Chen H, et al. 2017 3D deeply supervised network for automated segmentation of volumetric medical images Medical image analysis 41 pp 40-54 https://doi.org/10.1016/j.media.2017.05.001
[9] H. Greenspan, B. van Ginneken and R. M. Summers 2016 Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique IEEE Transactions on Medical Imaging 35(5) pp 1153-1159 https://doi.org/10.1109/TMI.2016.2553401

[10] K. H. Jin, M. T. McCann, E. Froustey and M. Unser 2017 Deep Convolutional Neural Network for Inverse Problems in Imaging IEEE Transactions on Image Processing 26(9) pp 4509-4522 https://doi.org/10.1109/TIP.2017.2713099

[11] Long J, Shelhamer E, Darrell T 2015 Fully convolutional networks for semantic segmentation Proceedings of the IEEE conference on computer vision and pattern recognition pp 3431-3440 https://doi.org/10.1109/TPAMI.2015.7298965

[12] Ronneberger O, Fischer P, Brox T 2015 U-net: Convolutional networks for biomedical image segmentation International Conference on Medical image computing and computer-assisted intervention Springer, Cham pp 234-241 https://doi.org/10.1007/978-3-319-24574-4_28

[13] LeCun, Y., Bengio, Y. & Hinton, G 2015 Deep learning Nature 521 pp 436–444 https://doi.org/10.1038/nature14539

[14] Badrinarayanan V, Kendall A, Cipolla R 2017 SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation IEEE Transactions on Pattern Analysis & Machine Intelligence 1-1 https://doi.org/10.1109/TPAMI.2016.2644615

[15] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille 2018 DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs IEEE Transactions on Pattern Analysis and Machine Intelligence 40(4) pp 834-848 https://doi.org/10.1109/TPAMI.2017.2699184

[16] C. Szegedy et al. 2015 Going deeper with convolutions IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA 1-9 https://doi.org/10.1109/CVPR.2015.7298594.

[17] K. He, X. Zhang, S. Ren and J. Sun 2016 Deep Residual Learning for Image Recognition 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV pp 770-778 https://doi.org/10.1109/CVPR.2016.90.

[18] Sergey Ioffe and Christian Szegedy 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift Proceedings of the 32nd International Conference on International Conference on Machine Learning 37 (ICML’15) JMLR.org pp 448-456 ArXiv abs/1502.03167

[19] Shiraishi J, Katsuragawa S, Ikezoe J, et al. 2000 Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists’ detection of pulmonary nodules American Journal of Roentgenology 174(1) pp 71-74 https://doi.org/10.2214/ajr.174.1.1740071

[20] Bagherinezhad H, Horton M, Rastegari M, et al. 2018 Label Refinery: Improving ImageNet Classification through Label Progression

[21] Yi X, Walia E, Babyn P 2019 Generative Adversarial Network in Medical Imaging: A Review Medical Image Analysis 58 101552 https://doi.org/10.13140/ RG.2.2.17702.16967

[22] Shorten C, Khoshgoftaar T M 2019 A survey on Image Data Augmentation for Deep Learning Journal of Big Data 6(1) pp 1-48 https://doi.org/10.1186/s40537-019-0197-0