Combination Strategy of Deep Learning and Direct Back Projection for High-efficiency Computed Tomography Reconstruction

Yu Shi, Fanzhen Meng, Ye Mao, Chenfeng Li, Shouping Zhu*
Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University
266 Xinglong Section of Xifeng Road
Xi'an, Shaanxi 710126
spzhu@xidian.edu.cn

ABSTRACT
Back projection filtering (BPF) and filtered back projection (FBP) are classic CT reconstruction analytic methods. However, when dealing with the reconstruction of incomplete datasets, analytic methods may suffer from artifacts due to the global nature of filtering. The iterative algorithm is another CT reconstruction method. Although the iterative algorithm can get excellent results, it requires high computation cost. In this work, inspired by the idea of BPF, we propose a CT image reconstruction strategy based on a new convolutional neural network that can capture features with different levels. The proposed method can learn the mapping between the direct back projection and the ideal CT images and then realize accurate image reconstruction. It is suitable for not only complete projection data but also transverse truncated projection data with different regions of interest (ROIs) sizes. The experiment results show that the proposed method outperforms the conventional FBP and other state-of-the-art deep learning methods in the truncated reconstruction. At the same time, the proposed method requires a lower computation cost which is helpful for fast reconstruction in the clinical research.

CCS Concepts
• **Applied computing →Imaging →Computing methodologies → Supervised learning**

Keywords
Deep Learning; Back Projection Filtering; Interior Reconstruction.

1. INTRODUCTION
Computed Tomography (CT) is one of the major tools widely used in medical diagnosis and image guided radiation therapy (IGRT). The keys to accurate CT image reconstruction are not only advanced hardware devices but also efficient reconstruction algorithms. The most classic analytic reconstruction algorithm is the filtered back projection (FBP) algorithm which can get an accurate analytical solution when the projection data is complete. If the order of filtering and back projection is switched, we can get another analytic reconstruction algorithm, back projection filtering (BPF) algorithm [1]. However, when the data is incomplete, the analytic algorithms may suffer from artifacts in reconstruction images due to the global nature of filtering. When dealing with the reconstruction of incomplete data, iterative algorithms [2; 3] can achieve excellent results. The main disadvantage of the iterative algorithm is the high computation cost limiting its application in the clinic.

In the past few years, artificial intelligence has been developing fast and gains great achievements in many biomedical fields including medical image segmentation [4], artifact reduction [5], image denoising [6] and so on. Recently considering the excellent performance of deep learning in solving the inverse problem, it has been used in solving a special kind of ill-posed inverse problems, medical image reconstruction. Würlf et al. [7] mapped the FBP algorithm onto an artificial neural network (ANN) architecture consisting of a convolution layer, a fully connected layer and a rectified linear unit. In their approach, the limited angle CT data can be reconstructed much more accurately than that with FBP while the computational complexity remained same. As addressed in the article, most artifact compensation methods in CT can be mapped to convolutional neural networks (CNN).

As a kind of CNN, the U-Net [4] architecture has become one of the most popular neural networks, which is widely used in the field of medical image reconstruction. Jin et al. [8] have proposed a biomedical imaging strategy, called FBPConvNet, which is based on the modified U-Net and residual learning. The FBPConvNet uses FBP images of sparse-view CT as the input. Experiment results show it has better imaging performance compared to the iterative reconstruction methods based on total variation constraint. One of its shortcomings is that it is hard to transfer between different datasets limiting its application in clinical research. The same problem existed in Han's work [9], in which the modified U-Net architecture is used to solve the interior problem. Although the approach performs near-perfect reconstruction for a particular ROI size, it is not suitable for transverse truncated data with different ROI sizes. Besides, there are other deep learning architectures such as improved GoogLeNet [10], perform better than U-Net when dealing with particular medical reconstruction. It means U-Net may not be the perfect network or it has the potential to be better in medical image reconstruction.

Inspired by these findings, we modified the conventional U-Net architecture and combined deep learning and direct back projection together for CT interior reconstruction in this paper. The reasons why we use direct back projection as input are 1) the direct back projection is a local operation in the direction of detec-

ISICDM 2019, August 24–26, 2019, Xi'an, China

ACM ISBN 978-1-4503-7262-6/19/08…$15.00

DOI: https://doi.org/10.1145/3364836.3364896

tors which will not be affected by the incomplete transverse truncated data, so the strategy will be suitable for both complete and transverse truncated data with different ROI sizes, 2) the calculation process of back projection can be accelerated by GPU which helps reduce reconstruction time combined with deep learning. Also, a new type of CNN which can capture different levels of features was constructed to learn a mapping between the back projection data and the ideal CT images.

In this study, the conventional FBP reconstruction algorithm and two deep learning reconstruction methods based on FBPConvNet [8] and BP+U-Net respectively were compared with the proposed approach and the results showed that the proposed approach had more accurate and stable imaging performance in terms of quantitative measures. In addition, the proposed approach had a lower computational cost. Considering the low computational cost and the universality for truncated data with different ROI sizes, the proposed reconstruction strategy is suitable for high-efficiency clinical research.

2. METHODOLOGY

2.1 Theory

In this section, we derive the expressions of projection and back projection for parallel-beam imaging which can be expanded to fan-beam imaging. It is assumed that *f(x, y)* is a density function of a 2D object and *p(s, θ)* is the projection [1]:

$$p(s,\boldsymbol{\theta}) = \int_{-\infty}^{\infty} f\left(s\boldsymbol{\theta} + t\boldsymbol{\theta}^{\perp}\right) dt \tag{1}$$

where $\boldsymbol{\theta} = (\cos\theta, \sin\theta)$, $\boldsymbol{\theta}^{\perp} = (-\sin\theta, \cos\theta)$, s is the distance from the detector rotation center to the beam and θ is the rotated angle. For the parallel-beam CT, the rotated angle is always equal to 180 degrees due to the symmetry. So the integral interval of back projection calculation is [0, π] and the back projection *b(x, y)* is expressed as:

$$b(x,y) = \int_0^{\pi} p(s,\boldsymbol{\theta})\big|_{s=x\cdot\boldsymbol{\theta}}\, d\theta \tag{2}$$

Back projection converts the projection called sinogram into an image which has the blurred outline of the object *f(x, y)*. In the process of back projection, the low-frequency components are over-weighted resulting in the blurred vision. Following the idea of *backprojection then filtering* (BPF) [1], we can do the filtering after back projection by multiplying the ramp filter $|\omega| = \sqrt{u^2 + v^2}$ in the 2D Fourier space. Then the expression of the object *f(x, y)* is written as the following form:

$$\begin{aligned} f(x,y) &= F^{-1}\left(F\left(b(x,y)\right)\cdot|\omega|\right) \\ &= b(x,y) * F^{-1}\left(|\omega|\right) \end{aligned} \tag{3}$$

where F denotes 2D Fourier transform, F^{-1} denotes 2D inverse Fourier and "$*$" denotes the convolution operation. However, it is difficult to calculate the 2D inverse Fourier of $\sqrt{u^2 + v^2}$ and perform convolution in the image domain. Besides, due to the global nature of the ramp filter, BPF is just suitable for complete projection data in most cases which limits its application.

As shown in equation (2), the calculation of *b(x, y)* is only related to the rotated angle and it works well for transverse truncated data which is incomplete in the direction of detector. Therefore, using the direct back projection as input, we propose a new reconstruction strategy where the mapping between back projection image *b(x, y)* and object *f(x, y)* is learned by the proposed CNN. The proposed method is suitable for both complete data reconstruction and interior reconstruction when maintaining computational cost the same as BPF. Equation (3) is rewritten as:

$$f(x,y) = NN\left(b(x,y)\right) \tag{4}$$

where NN denotes the proposed neural network system.

2.2 Network Architecture

The proposed network architecture is based on the U-Net++ [11] architecture which is initially designed for medical image segmentation and not suitable for CT reconstruction directly. In this paper, the architecture is modified to suit this special problem.

In the original U-Net architecture, the combination of shallow and deep features is implemented by a simple skip connection. The original U-Net with the wide receptive field may work well for the image with a fixed ROI size. However, it may fail to generate a high-quality image for transverse truncated data with different ROI sizes. The reasons are that the weights of shallow and deep features are not immutable for reconstructions of truncated data with different ROI sizes and the linear concatenation in original architecture may not solve the complex problem. In the U-Net, the number of down-sampling layers determines the size of the receptive field and different sizes of receptive fields can capture features with different depths. Therefore, it is expected to succeed in the reconstruction of transverse truncated data with different ROI sizes if we deeply aggregate shallow and deep features instead of a simple skip connecting operator such as concatenation. The deep aggregate is implemented by U-Net++.

Considering the particularity of the transverse truncated data, the modified architecture of U-Net++ was used to CT reconstruction. The basic idea of U-Net++ is the deep combination of deep features and shallow features which is suitable for CT truncated reconstruction with different ROI sizes.

The diagram of our modified aggregation architecture is shown in Figure 1. The network uses the direct back projection of complete data and truncated data with a particular ROI size as input. In the diagram, the feature extraction operator is a black arrow, consisting of a convolution layer whose convolutional kernel size is 3×3 followed by a rectified linear unit (ReLU) and a batch normalization module. Two consecutive feature extraction operators constitute the basic module. The max-pooling layer which is red down arrow in Figure 1 is used between adjacent feature extraction layers. After going through the basic module, the features at high levels of the feature pyramid are up-sampled by deconvolution operation (yellow up arrow) both in the encoder and decoder. In the same layer of the feature pyramid, the features are grouped together by concatenation (blue arrows). Then the fused images are encoded by the basic module to generate the new features. The new features are grouped again and the above steps are repeated. The proposed architecture inherits the dense connections composed of the concatenation and the basic module in U-Net++ but the drop out layer in the basic module is replaced by batch normalization.

2.3 Experiments

Our experimental setups were as follows. Firstly, we compared our proposed reconstruction strategy with the conventional FBP algorithm by reconstructing transverse truncated projections with different ROI sizes. Secondly, a deep learning method based on FBP called FBPConvNet was compared with the proposed

method. Although FBPConvNet mentioned above was designed for sparse view CT, a similar network architecture proposed by Han [9] outperformed iterative algorithms for transverse truncated data with a particular ROI size so we did not compare to iterative methods. Finally, in order to prove the utility of the combination of shallow and deep features, we also compared the proposed method to another reconstruction strategy called BP+U-Net that the network architecture was the U-Net where the combination of features was implemented by a simple skip connection. The input of BP+U-Net was still the back projection.

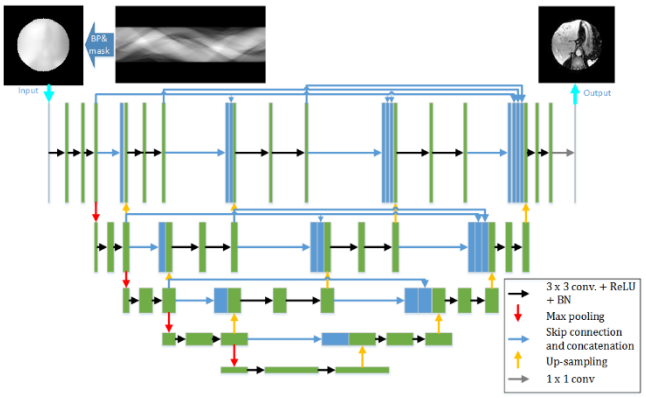

Figure 1. The architecture of the proposed deep network for back projection filtering

2.3.1 Dataset

The data set contained 6,985 128×128 artifacts-free ideal CT images. The sinograms were obtained by forward projection operation called simulated CT. In the simulated CT, the number of detector elements was set to 370 whose pixel size was 0.3 mm. The distance between source and object (SOD) was 250 mm and the distance between source and detector (SDD) was 500 mm. The image size was 38.4 mm×38.4 mm corresponding to the object. The training dataset contained 5,000 of 6,985 images and the rest of the images made up the test set. Techniques of data augmentation such as horizontal and vertical flips were used to avoid overfitting. In the training set, 70 elements in the middle of 370 detector elements were used to simulate the transverse truncated sinogram and full detector elements were applied to generate complete data. In the test set, 70, 130, 170, 210, 250, 290 and 370 detector elements in the middle were used to generate the truncated sinograms with different ROI sizes and complete data. Then back projection operator was used to generating back projection images with a mask making back projection images zero outside the ROI.

2.3.2 Network training

The proposed method was trained by pairs of back projection images and original ideal images. The number of training image pairs was 30,000 containing complete data and transverse truncated back projection with 70 detector elements left after data augmentation. The classical mean squared error (MSE) was used as the loss function.

The networks were implemented and trained by Pytorch 1.0 [12] on a PC with Intel Xeon Sliver 4110 CPU @ 2.10 GHz and an NVIDIA GTX 1080Ti (11G) GPU. The hyper-parameters were set as follows: the number of epochs was 100 and the initial learning rate was 10^{-4} dropped to 5×10^{-5} finally. Stochastic gradient descent (SGD) optimizer was used to train the network. The weight decay and batch size were 10^{-4} and 32 respectively. It took about 8 hours to train the proposed model.

3. EVALUATION

3.1 Quantitative Metric

The structural similarity (SSIM) [13] was used as the quantitative metric to compare the similarity between networks' outputs and ideal images, SSIM is given by

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{5}$$

where μ_x, μ_y were the average of x, y respectively, σ_x, σ_y were the variances of x, y respectively and σ_{xy} was a covariance of x and y. In addition, the peak signal to noise ratio (PSNR) was used as another metric defined as

$$PSNR(x,y)=10\log_{10}\left(\frac{\left(\max\left(x\otimes x\right)\right)^2}{\|x-y\|_2^2}\right) \tag{6}$$

where $\otimes$ denotes element-wise multiplication. Besides, the reconstruction time over the test dataset is recorded to quantify the computational cost.

3.2 Experimental Results

We compared our proposed method with one analytic reconstruction method and two other deep learning reconstruction methods using qualitative and quantitative metrics. As shown in Figure 2, both the proposed strategy and BP+U-Net had good performance in different ROI reconstructions, but analytical results and FBPConvNet's results had a bright outline in some cases. It indicated that back projection was suitable for transverse truncated data reconstruction with different ROI sizes due to its local nature.

We calculated the mean of PSNR and SSIM in ROIs for quantitative evaluation in Table 1. The proposed strategy had the highest PSNR value and SSIM value in all situations except for complete data (370 elements) and the nearest complete data (290 elements). FBP got the highest scores in metrics for complete back projection reconstruction because the untruncated reconstruction was a well-posed problem and the analytical method such as FBP can get a precise and unique solution. FBPConvNet applied the FBP reconstructed results as the input which was very close to the ground truth for nearest complete data, therefore it worked better than the other methods in this situation. However, it did not suit CT reconstruction with different ROI sizes considering its poor performance in the cases that the detector elements were few. Similar to the qualitative assessment, the backprojection-based reconstruction strategy scored the highest in most truncated reconstructions as shown in the quantitative results. However, the proposed method benefitting from a dense combination of shallow and deep features had more stable SSIM and higher PSNR than BP+U-Net. It means that the proposed method works well in all transverse truncated reconstructions.

Besides, the computation time for the proposed method containing back projection and reconstruction was 1706.2 seconds, in which reconstruction costed 69.1s, over the whole test set including 1985 reconstruction tasks. The time cost for the FBPConvNet which outperformed iterative reconstruction was 1780.2s over the test set, where reconstruction costed 51.1s. In addition, the computation

time for the proposed method could be less by applying developed acceleration technology.

4. CONCLUSIONS

In this paper, we proposed a combination strategy of deep learning and back projection for high-efficiency CT reconstruction. Following the idea of BPF, we use back projection as input. The modified U-Net++ was used as the backbone which can combine shallow and deep features deeply. Numerical experiment results show that the proposed strategy performs better in quantitative metrics for transverse truncated data reconstruction with different ROI sizes and it has a lower computational cost. The most important is that the network was only trained by two cases but it seemed suitable for all transverse truncated reconstruction with different ROI sizes in inference. Therefore, the proposed strategy is suitable for an efficient reconstruction in clinical practice where flexible collimation would be preferred.

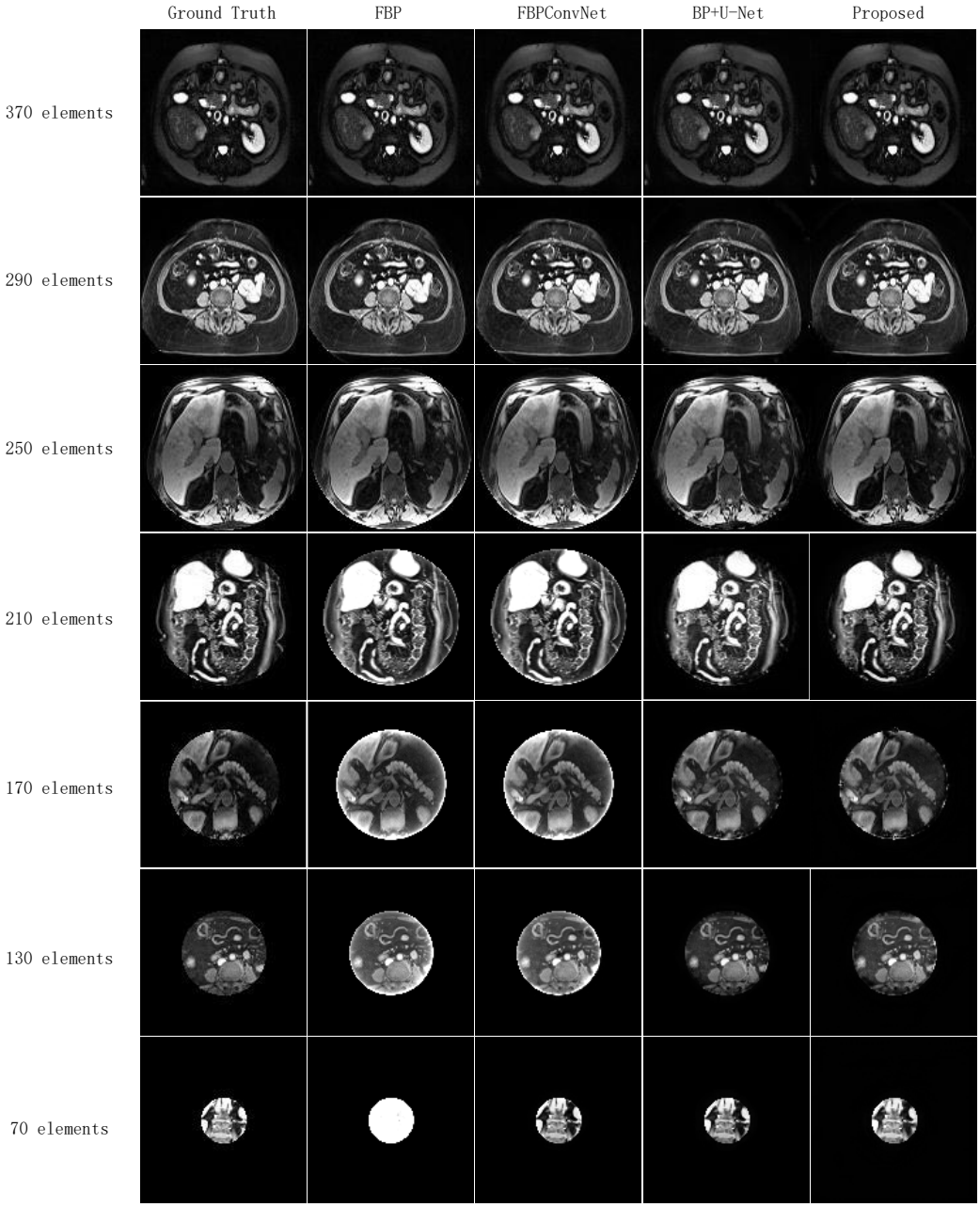

Figure 2. Reconstruction results using four methods for different ROI sizes

Table 1. Means of PSNR and SSIM of four methods over test dataset for different ROI sizes.

Metrics \ Methods		FBP	FBPConv Net	BP+U-Net	Proposed
PSNR	370 elements (full)	**43.06**	41.98	37.79	35.70
	290 elements	31.82	**32.72**	29.48	30.05
	250 elements	25.33	25.74	25.90	**26.53**
	210 elements	17.71	18.22	23.05	**25.09**
	170 elements	12.63	13.34	21.77	**25.42**
	130 elements	7.91	9.84	20.88	**25.49**
	70 elements	0.50	25.40	26.55	**26.96**
SSIM	370 elements (full)	**0.9903**	0.9890	0.9881	0.9738
	290 elements	0.9311	**0.9456**	0.9003	0.9137
	250 elements	0.8890	**0.8990**	0.8688	**0.8901**
	210 elements	0.8094	0.8218	0.8337	**0.8755**
	170 elements	0.7024	0.7176	0.8205	**0.8830**
	130 elements	0.5685	0.6074	0.8170	**0.8911**
	70 elements	0.2975	0.9116	0.9197	**0.9290**

5. ACKNOWLEDGMENTS

This work was partly supported by the National Key Research and Development Program of China under Grant No. 2016YFC0103800, the National Natural Science Foundation of China under Grant Nos. 81227901, 61471279, 61901339，the Fundamental Research Funds for the Central Universities Nos. JB191205, JB181201, and the China Postdoctoral Science Foundation No. 2018M643589, the National Natural Science Foundation of Shaanxi Province under Grant No.2019JQ-655.

6. REFERENCES

[1] Zeng, G.L. 2009. *Medical Image Reconstruction*. DOI= https://doi.org/10.1007/978-3-642-05368-9.

[2] Sidky, E.Y., Kao, C.M., and Pan, X. 2006. Effect of the data constraint on few-view, fan-beam CT image reconstruction by TV minimization. In *Nuclear Science Symposium Conference Record*. DOI= https://doi.org/10.1109/NSSMIC.2006.354372.

[3] Ludwig, R., Frank, B., Christof, F., and Marc, K. 2011. Improved total variation-based CT image reconstruction applied to clinical data. *Physics in Medicine Biology 56*, 6 (2011), 1545-1561.

[4] Ronneberger, O., Fischer, P., and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing & Computer-assisted Intervention* (Munich, Germany, October 2015). DOI= https://doi.org/10.1007/978-3-319-24574-4_28

[5] Zhang, C. and Xing, Y. 2018. CT artifact reduction via U-net CNN. In *Society of Photo-optical Instrumentation Engineers Medical Imaging* (Houston, Texas, United States, 2018). DOI= https://doi.org/10.1117/12.2293903

[6] Gondara, L. 2016. Medical Image Denoising Using Convolutional Denoising Autoencoders. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)* (Barcelona, Spain, 2016). DOI= https://doi.org/10.1109/ICDMW.2016.0041

[7] Würfl, T., Ghesu, F., Christlein, V., and Maier, A. 2016. *Deep Learning Computed Tomography*. In *International Conference on Medical Image Computing & Computer-assisted Intervention* (Athens, Greece, October 2016). DOI= http://dx.doi.org/10.1007/978-3-319-46726-9_50

[8] Jin, K.H., Mccann, M.T., Froustey, E., and Unser, M. 2017. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing 26*, 9, 4509-4522.

[9] Han, Y., Gu, J., and Ye, J.C. 2017. Deep Learning Interior Tomography for Region-of-Interest Reconstruction. https://arxiv.org/abs/1712.10248

[10] Xie, S., Zheng, X., Chen, Y., Xie, L., Liu, J., Zhang, Y., Yan, J., Zhu, H., and Hu, Y. 2018. Artifact Removal using Improved GoogLeNet for Sparse-view CT Reconstruction. *Scientific Reports 8*, 1, 6700.

[11] Zhou, Z., Rahman Siddiouee, 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 (DLMIA 2018)* (Granada, Spain, September 2018). Springer International Publishing, Cham, 3-11. DOI= https://doi.org/10.1007/978-3-030-00889-5_1

[12] Pytorch Homepage: https://pytorch.org.

[13] Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P., 2004. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans on Image Process 13*, 4, 600-612.