EPJ Web of Conferences **251**, 03058 (2021) https://doi.org/10.1051/epjconf/202125103058
*CHEP 2021*

# Deep learning based low-dose synchrotron radiation CT reconstruction

*Ling* Li [1, 2], *Yu* Hu [1, *]

[1]Institute of High Energy Physics, CAS, 100049 Beijing, China
[2]University of Chinese Academy of Sciences, 100049 Beijing, China

**Abstract.** Synchrotron radiation sources are widely used in various fields, among which computed tomography (CT) is one of the most important. The amount of effort expended by the operator varies depending on the subject. If the number of angles needed to be used can be greatly reduced under the condition of similar imaging effects, the working time and workload of the experimentalists will be greatly reduced. However, decreasing the sampling angle can produce serious artifacts and blur the details. We try to use a deep learning model which can build high quality reconstruction sparse data sampling from the angle of the image and ResAttUnet are put forward. ResAttUnet is roughly a symmetrical U-shaped network that incorporates similar mechanisms to ResNet and attention. In addition, the mixed precision is adopted to reduce the demand for video memory of the model and training time.

## 1 Introduction

Synchrotron radiation is electromagnetic radiation emitted by charged particles moving at varying speeds in a magnetic field near the speed of light. Synchrotron radiation source is widely used in atomic physics, molecular environment science, radiation therapy, nanoscience, nuclear physics, medicine, medical diagnosis and other fields because of its wide spectrum distribution, high brightness and high collimation. The main experimental methods of synchrotron radiation are also varied, including X-ray lithography, high spatial resolution X-ray imaging, X-ray diffraction and scattering, etc. In practical applications, some experiments are very time-consuming, and some experiments require high operating accuracy such as computed tomography (CT). If the needed number of angle measurements can be greatly reduced under the condition of similar imaging effects, the working time and workload of the experimentalists will be greatly reduced.

Analytical algorithms and iterative algorithms are the two main types of traditional CT reconstruction methods. The theoretical cornerstone of the analytic algorithms is Radon Transform[1]. One of the most widely used analytical algorithms is the filtering back projection reconstruction algorithm (FBP)[1]. If the number of sampling angles is reduced, the angle between the two pauses will also increase, and the reliability of the data in the interpolation region will be reduced, so the error caused by interpolation will be larger. The

* Corresponding author: huyu@ihep.ac.cn

iterative algorithms are based on the assumption of discretization. Setting the discrete points on the plane as unknowns, the equations are set up by using the rays passing through the object. To reconstruct the image is essentially to solve the equations. The core of iterative algorithms is to use gradient descent method to reduce the error until the solution of the equations is iterated. Because of the large amount of data, overdetermined or underdetermined, ill-conditioned problems, the iterative method is often used instead of the traditional method to solve the equations.

The high similarity between iterative algorithms and neural networks makes it possible for neural networks to be applicable to the realization of CT reconstruction conjecture. Many researchers have attempted to use deep learning to post-process CT images and achieved excellent results[3, 4]. The landmark network architecture in this field is Unet[2]. Unet is often used in CT image segmentation. In this paper, the author tries to add a structure similar to ResNet[5] on the basis of Unet network structure to reduce the effect of vanishing gradients of neural networks, so that the network can improve its performance with the deepening of depth. At the same time, the attention mechanism[6,7] is also added to the network, which can make the network reconstruction pay more attention to the subtle structure of the image through the weighted algorithm. These structures are often the focus of the experimenter's attention, such as fine lines or gaps in an object's interior. Because of the intensity and high brightness of synchrotron radiation source, its CT data is often much larger than the general medical CT data. However, the neural network training process has a large demand for video memory, so the mixed precision training technique is adopted in this paper. The default data type of neural networks is FP32. Using FP16 not only reduces video memory usage, but also speeds up training. However, directly replacing FP32 with FP16 will cause gradients overflow/underflow and rounding error. In mixed precision training, FP16 is used for storage and multiplication, FP32 is used for accumulation, and the idea of Loss Scaling is used to minimize the error caused by using FP16. In the experimental results section of the fourth chapter, it is shown that the use of mixed precision training reduces video memory occupation and training time.

## 2 Synchrotron Radiation CT Reconstruction Method Based on Neural Network

The network ResAttUnet proposed in this paper is relatively symmetric on the whole, and it cannot complete the transformation from sinusoidal domain to plane domain, so it cannot directly reconstruct the original data into images. We first need to use FBP to convert the original data from the sinusoidal domain to the plane domain, and then use the neural network to remove artifacts, refine the image and other work.

The Unet network was proposed in 2015 and soon became a reference in the filed of CT medical images. The network as a whole presents a symmetrical U-shape, which gives the input and output data the same dimension. On the left side of the U-shaped structure, with the increasing depth of the network, the area perceived by the filter is constantly enlarged, and so is the field of vision. The output image can also integrate features of different scales. The right side of the U shape uses transposed convolutions to increase the image size. Images of the same depth on the left and right sides of U-type are connected by skip connection. This structure can make the data on the right side of U-type supplement the relatively original information on the left side. Since the current detector pixel has been improved, generally reaching 2K, and more than 10K detectors will be used in the future, which will increase the memory, so the parameters of the Unet model we used as the control will be half of the original.

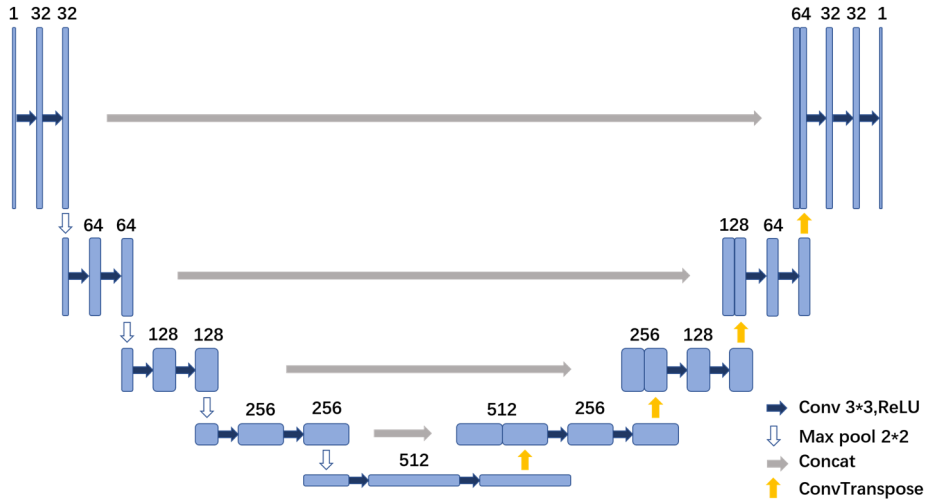


**Fig. 1.** The original Unet model [2]

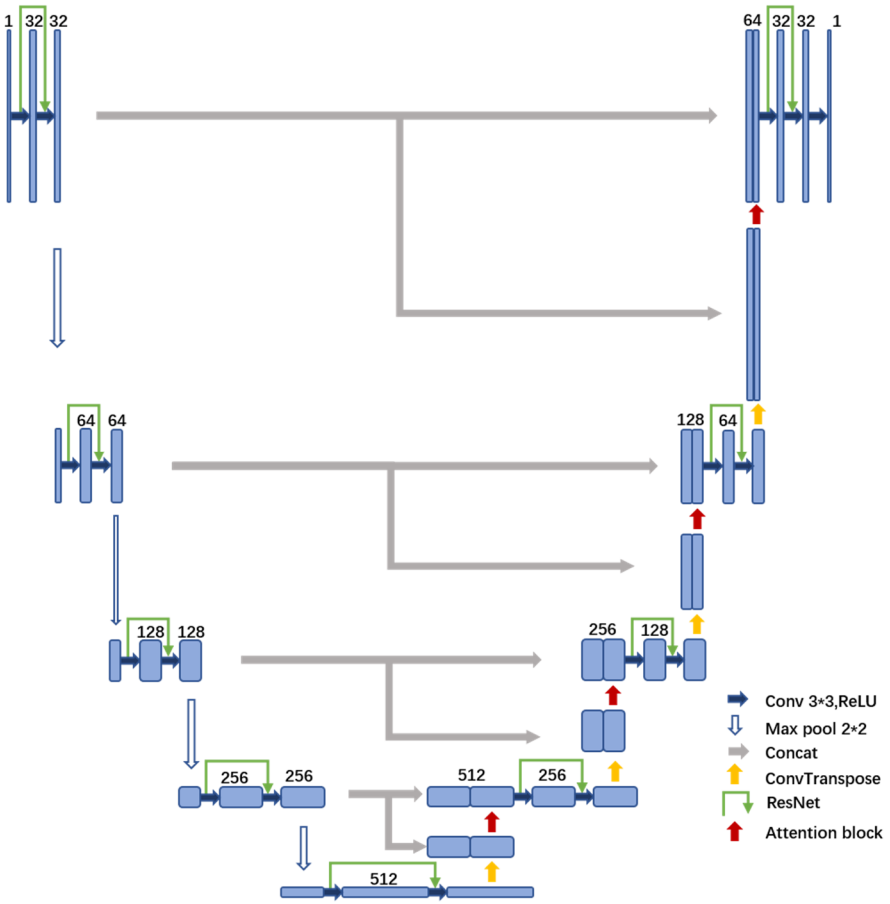


**Fig. 2.** The architecture of ResAttUnet

The ResAttUnet model proposed in this paper is shown in **Fig. 2**. Its overall framework is similar to the Unet, but on the basis of the Unet it joins the two structures, ResNet-like structure and attention mechanism. Both the left and right sides of the U-shape have a structure similar to ResNet. ResNet is proposed to alleviate the gradient vanishing problem of deep neural networks. In principle, deep neural networks have higher representational power than shallow networks: however, experimental evidence suggests that the vanishing gradients problem can hinder their performance. ResNet connects the neural network layer near the input to the layer near the output. If the output of a layer is already a good fit for the desired result, adding another layer will not make the model worse.

The ResNet-like structural data used in this article skips only one convolution, and its output is actually the output of the second convolution plus the input. Their specific differences are shown in **Fig. 3**.

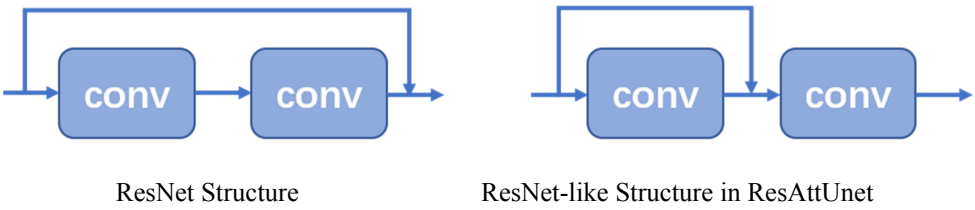


**Fig. 3.** The difference between ResNet and ResNet-like Structure in ResAttUnet

The attention mechanism can make the model ignore irrelevant information and pay attention to important information and improve the level of details in the reconstructed images. As shown in Fig. 4, the data on the left and right sides of the U-shaped structure are input into two 1*1 convolutions respectively, and then concatenated into an additional 1*1 convolution. The output filter of this convolution is 1. Finally, the data on the right side of the U-shaped structure is multiplied by the convolution weight to get the output.

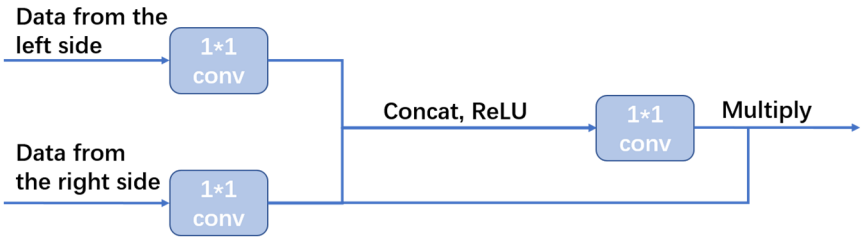


**Fig. 4.** The structure of the attention block

## 3 The experimental results

### 3.1 The training and test data sets

First, the simulated tomographic data is used to train the proposed network in the training process. We used the randomly generated foam phantoms 3D objects. To generate a foam phantom, non-overlapping randomly-placed spheres with randomly intense and varying sizes were removed from a cylinder of a single material. Tomographic projections of the objects

were simulated using TomoPhantom[8]. In addition, we have a set of real data for testing. Our actual data source is Beijing Synchrotron Radiation Facility (BSRF) [9].

This experiment uses sparse reconstructions of real data and simulated data, with a subsampling ration of 20, that is, one out of every 20 angles of the original data is selected as the data for reconstruction. There are obvious artifacts in sparse reconstruction images. The sparse reconstructed images of the simulated data are used as the training set and validation set, and the real value of the simulated data are used as the training set label and validation set label.

### 3.2 Conducting experiments

After the data set is read in, the data is uniformly processed so that each pixel is represented by 8-bit binary. The scale of simulated data and real data as well as their image after sparse reconstruction are unified to prevent model training failure due to inconsistency of the data scale.

The next step consists in building the model. The final number of filters in each layer on the left side of the U type is 32, 64, 128, 256 and 512 respectively. The number of filters in each layer on the right side of the U corresponds to that on the left side of the U. The parameters of the Attention block are determined by the number of filters in the same layer on the left of the U and the number of filters in the convolution block below it.

We use Adam optimizer and the Mean Squared Error (MSE) loss. Since mixed accuracy training is adopted, "eps=$10^{-7}$" need to be declared when setting the Adam optimizer. The initial learning rate is $3 \times 10^{-4}$, and as the epoch increased, the learning rate decreased dynamically. For every 10 epochs, the learning rate becomes 0.95 of the original size. Mixed precision training can be set to the level, O0 for pure FP32 training, O3 for pure FP16 training. Going from O1 to O3, the amount of operation calculated in FP16 increases. We set the parameter OPT_LEVEL for mixed precision training as "O2", that is, except for the BATCH norm, almost all calculations are performed using FP16. Set the parameter OPT_LEVEL ="O2" for mixed precision training. The training runs for 65 epochs.

We compare our results to the standard Unet model and the ResAttUnet model trained using FP32 and consistent hyper-parameters.

## 4 Experimental results and analysis

### 4.1 Evaluation criteria

Our evaluation criteria mainly include two indicators: PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity). These two indexes are often used for image evaluation. MSE (Mean Squared Error) is the mean energy difference between the real image and the noisy image, and the difference is the noise. PSNR is the ratio of the peak signal energy and MSE. PSNR is defined as follows:

$$PSNR = 10log_{10}\frac{MaxValue^2}{MSE} = 10log_{10}\frac{2^{bits}-1}{MSE} \quad (1)$$

The SSIM formula is based on three comparative measures between samples x and y: luminance, contrast, and structure. Its final definition is as follows.

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

$\mu_x$ and $\mu_y$ are the mean values of x and y of the sample image respectively. $\sigma_x^2$ and $\sigma_y^2$ are the variances of x and y. $\sigma_{xy}$ is the covariance of x and y. $c_1$, $c_2$are two constants.

### 4.2 Experimental results and analysis

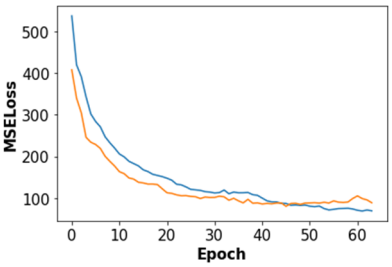


**Fig. 5.** The training loss (blue) and the validation loss (orange)

**Fig. 5** shows the training and validation losses evolution over the 65 training epochs. The total training time is 821 seconds and the average training and validation losses for $1200 \times 1200$ images are 64.4 and 82.2 respectively. The output images of training set, validation set and test set are shown in **Fig. 6**. The test set and validation set output results are shown in the figure below.

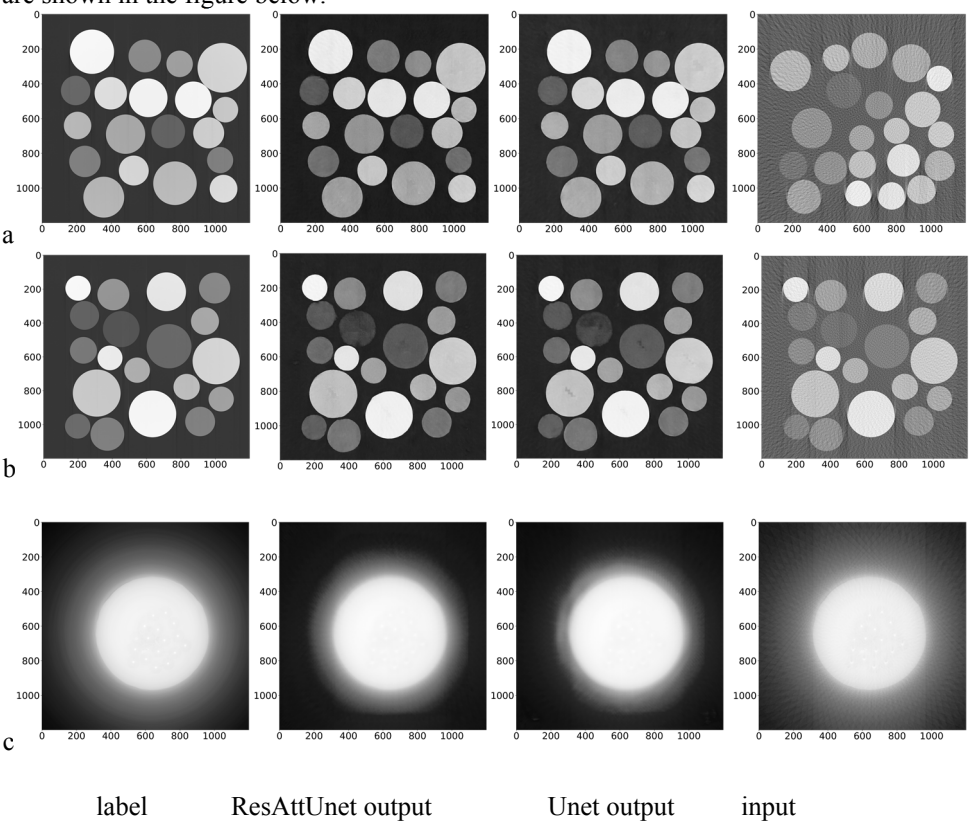


**Fig. 6.** The output of ResAttUnet and Unet

**Fig. 6** shows, from left to right, the label, ResAttUnet output, Unet output and the input data. The a, b, c lines represent the networks output corresponding to the training, validation and test set repectively. According to the image results, there is not much difference between ResAttUnet and Unet on the training set. In the validation set, the output image of the ResAttUnet model has some incomplete peripheral circles, while the output image of Unet has obvious holes in the inner circles. In the real data test set, both ResAttUnet and Unet have roughly removed artifacts, but the small round holes in the Unet output image are larger than the label. We averaged the PSNR and SSIM results on the entire training, validation and test sets (table 1). O0 and O2 denote the opt level. O2 is the opt level used in our experiment, and O0 as the control group indicates that mixed precision training is not used. As can be seen from the data in Table 1, ResAttUnet performance is very close to Unet, independently of the use of mixed precision. On the other hand, Table 2, clearly shows the reduced memory usage and training time achieved by using mixed precision.

**Table 1.** PSNR and SSIM of models

| Model | PSNR | | | SSIM | | |
|---|---|---|---|---|---|---|
| | Training set | Validation set | Test set | Training set | Validation set | Test set |
| **ResAttUnet(O0)** | 30.1283 | 29.0519 | 26.7286 | 0.9913 | 0.9891 | 0.9881 |
| **ResAttUnet(O2)** | 30.6647 | 29.0860 | 25.3447 | 0.9923 | 0.9894 | 0.9842 |
| **Unet(O2)** | 30.1168 | 29.0257 | 24.6563 | 0.9913 | 0.9890 | 0.9825 |

**Table 2.** Effects of mixed precision training on video memory occupancy and training time

| Model | O0 | | O2 | | | |
|---|---|---|---|---|---|---|
| | video memory (MiB) | training time(s) | video memory (MiB) | training time(s) | video memory saving | training time saving |
| **ResAttUnet** | 7021 | 1401.1953 | 4773 | 809.8071 | 32.01% | 42.20% |
| **Unet** | 5575 | 953.6296 | 4451 | 611.7596 | 20.16% | 35.84% |

The experimental results suggest that the proposed model is affective, greatly reducing image reconstruction time and thus simplifying the CT image reconstruction at synchrotrons. The proposed ResAttUnet architecture relies on convolutional kernels with fixed size, thus it optimally reconstructs local features. Transformer architectures [10] or fully connected layers migh improve the attention to global information.

## 5 Conclusion

In this paper, we propose a deep learning-based network model ResAttUnet for low-dose synchrotron radiation CT reconstruction. ResAttUnet adds ResNet blocks and the attention mechanism to Unet. The proposed method is tested on simulated and real datasets and compared with the Unet model. Experimental results show that the ResAttUnet proposed in this paper is effective in low-dose synchrotron radiation CT reconstruction. The use of mixed precision can reduce video memory requirements and training time.

## Reference

1. G. Zeng, *Medical image reconstruction: a conceptual tutorial,* New York: Springer (2010)

2. O. Ronneberger, P. Fischer, T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, International Conference on Medical image computing and computer-assisted intervention. Springer, Cham 234 (2015)
3. K. H. Jin, M. T. McCann, E. Froustey, M. Unser, IEEE Transactions on Image Processing **26**, 9, 4509-4522 (2017)
4. Z. Liu, T. Bicer, R. Kettimuthu, D. Gürsoy, F. De Carlo, I. Foster, J. Opt. Soc. Am. A **37**, 422-434 (2020)
5. K. He, X. Zhang, S. Ren, J. Sun, *Deep residual learning for image recognition*, 2016 IEEE conference on computer vision and pattern recognition (2016)
6. S. Chaudhari, G. Polatkan, R. Ramanath, V. Mithal, arXiv:1904.02874 (2019)
7. A. Vaswani, et al., *Attention is all you need*, Advances in neural information processing systems (2017)
8. D. Gürsoy, F. De Carlo, X. Xiao and C. Jacobsen, J. Synchrotron Rad. **21**, 1188-1193 (2014)
9. E. Tang and D. Xian, Rev. Sci. Instrum. **63**, 1575 (1992)
10. A. Dosovitskiy, et al., arXiv:2010.11929 (2020)