Received: 12 January 2020 | Revised: 9 August 2020 | Accepted: 17 August 2020

IET Image Processing

DOI: 10.1049/ipr2.12048

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

Recurrent learning with clique structures for prostate sparse-view CT artifacts reduction

Tiancheng Shen[1,2] | **Yibo Yang**[1,2] | **Zhouchen Lin**[2] | **Mingbin Zhang**[3]

[1] Center for Data Science, Peking University, Beijing, China

[2] Key Laboratory of Machine Perception (MOE), School of Electronics Engineering and Computer Science, Peking University, Beijing, China

[3] Department of Urology, Affiliated Union Hospital of Fujian Medical University, Fuzhou, China

Correspondence
Mingbin Zhang, Department of Urology, Affiliated Union Hospital of Fujian Medical University, Fuzhou, China.
Email: 2012103020038@whu.edu.cn

Abstract

In recent years, convolutional neural networks have achieved great success in streak artifacts reduction. However, there is no special method designed for the artifacts reduction of the prostate. To solve the problem, the artifacts reduction CliqueNet (ARCliqueNet) to reconstruct dense-view computed tomography images form sparse-view computed tomography images is proposed. In detail, first, the proposed ARCliqueNet extracts a set of feature maps from the prostate sparse-view CT image by Clique Block. Second, the feature maps are sent to ASPP with memory to be refined. Thenanother Clique Block is applied to the output of ASPP with memory and reconstruct the dense-view CT images. Later on, reconstructed dense-view CT images are used as new input of the original network. This process is repeated recurrently with memory delivering information between these recurrent stages. The final reconstructed dense-view CT images are the output of the last recurrent stage. Our proposed ARCliqueNet outperforms the SOTA (state-of-the-art) general artifacts reduction methods on the prostate dataset in terms of PSNR (peak signal-to-noise ratio) and SSIM (structural similarity). Therefore, we can draw the conclusion that Clique structures, ASPP with memory and recurrent learning are useful for prostate sparse-view CT Artifacts here.

1 | INTRODUCTION

The prostate is a small, walnut-sized gland located deep inside the groin. Due to bad work, health and eating habits, prostate diseases are very common in men.

For example, prostate cancer is the second leading cause of cancer death in American men. Fortunately, because of a huge increase in imaging techniques, such as computed tomography (CT), prostate diseases are now easier to diagnose. Precise imaging of the prostate is very useful for treatment planning and many other diagnostic and therapeutic procedures for prostate diseases.

However, precise CT always brings a considerable amount of radiation dose, which is harmful to human health. One way to reduce radiation dose is sparse-view CT reconstruction, which is achieved by reducing the number of radiation angles, that is., views. This method is also called 'down-sampling streak artifacts' since reducing the number of radiation angles can be regarded as downsampling. Every coin has two sides. Fewer X-ray projection views result in more streak artifacts and worse image quality than CT images reconstructed from original dense projection views. Besides, because of large variations of prostate shape and indistinct prostate boundaries, streak artifacts make it harder for doctors to locate and diagnose. Balancing image quality and radiation dose level has become a well-known trade-off problem.

To resolve this issue, we propose a neural network architecture named artifacts reduction CliqueNet (ARCliqueNet) to reconstruct dense-view CT images from sparse-view CT images. The prostate has a wide variation in size and shape among different subjects due to pathological changes, so we introduce ASPP [1] to use multi-scale information to further enhance the efficiency of information acquisition, and Clique Block [2] to maximise information flow. What is more, due to the similar appearance of the prostate and its surrounding tissues (e.g. blood vessels, bladder, rectum, and seminal vessels), the lack of clear prostate boundaries motivates us to introduce the recurrent mechanism to refine reconstruction images stage

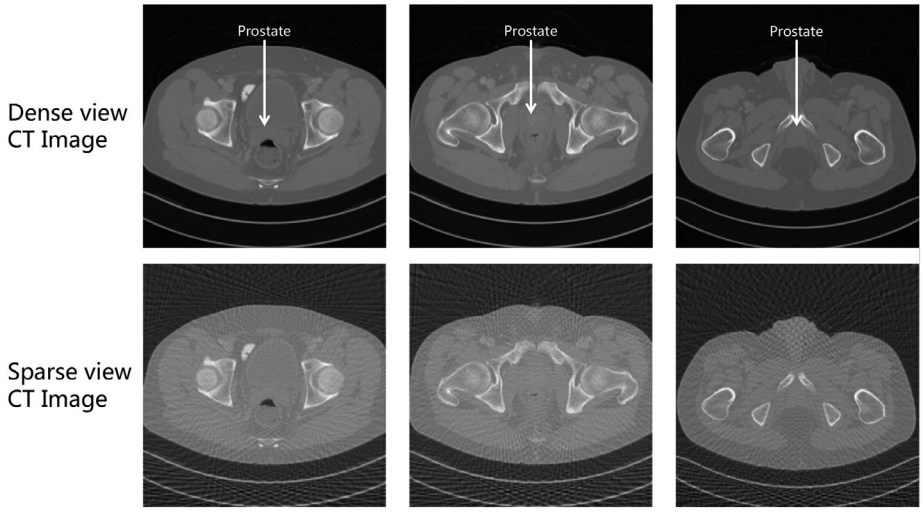

FIGURE 1 The prostate has a wide variation in size and shape among different subjects due to pathological changes, and there is lack of clear prostate boundaries due to similar appearance of prostate and its surrounding tissues (e.g. blood vessels, bladder, rectum and seminal vessels). And as shown in the above two rows, the streak artifacts, resulting from sparse view, make the prostate more confusing and hard to locate and analyse

by stage and add memory mechanism to ASPP. According to the two weaknesses mentioned above, which can be seen in Figure 1, ARCliqueNet involves Clique Block [2], ASPP [1] with memory and recurrent mechanism.

Main contributions of this paper are listed as follows:

1. To the best of our knowledge, this is the first paper to introduce Clique Block [2] and ASPP [1] with memory into the streak artifacts reduction of CT. These designs are special for the prostate's properties.
2. We bring in recurrent mechanism to refine prostate CT images getting rid of streak artifacts stage by stage. This mechanism encourages the addition of memory for ASPP and is also helpful to reduce parameters needed.
3. Experiments show that ARCliqueNet outperforms general SOTA artifacts reduction methods on prostate CT images dataset from the American Association of Physicists in Medicine (AAPM) [3].

2 | RELATED WORK

2.1 | Basic module evolution

A number of deep networks with large model capacity have been proposed. These designs can be briefly divided into two categories: widen the network and deepen the network.

As for widening the network, the Inception module [4] fuses the features in different map sizes to construct a multi-scale representation. Wide residual networks [5] increase the width and decrease the depth to improve the performance. However, simply widening the network is easy to consume more runtime and memory.

As for deepening the networks, skip connections or shortcut paths are widely adopted to ease the training stage of the network, such as ResNet [6]. To further increase information flow, DenseNet [7] replaces the identity mapping in the residual block by concatenating operation so that new feature learning can be reinforced while keeping old feature reused. In each Clique Block [2], both forward and feedback are densely connected. The information flow is maximised and feature maps are repeatedly refined by attention mechanism.

2.2 | Streak artifacts reduction in computer vision

With the demand for radiation dose reduction is becoming more and more intense, great efforts have been devoted to improving sparse-view CT reconstruction's quality. Existing approaches to address the streak artifacts can be mainly divided into two categories.

On the one hand, classical methods are developed from mathematics problems, such as compressed sensing theory, total variation, dictionary learning, and so on. These methods include ASD-POCS [8], AwTV [9], ASDL [10] and so on.

On the other hand, with the rapid growth in deep learning, deep learning methods, such as Tight Frame U-Net(TF U-Net) [11], cascade of U-Nets [12], outperform traditional methods in terms of PSNR and SSIM.

Our method is a deep learning method incorporating Clique Block [2], ASPP [1] with memory and recurrent mechanism. These designs are intended to solving the lack of clear prostate boundaries due to similar appearance between the prostate and the surrounding tissues, and the wide variation in size and shape among different subjects.

3 | PROPOSED METHODS

In this section, we first overview the proposed ARCliqueNet architecture, then we introduce Clique Block [2] as encoder

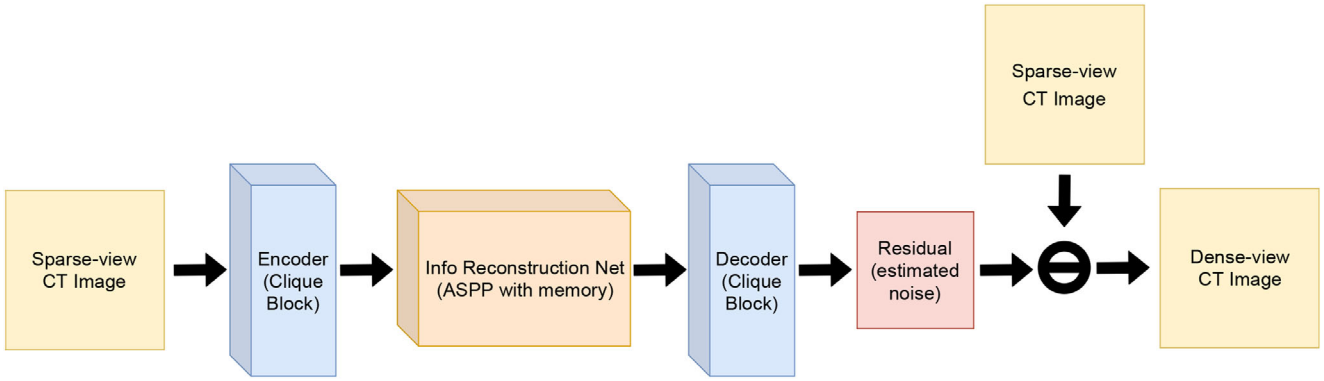

FIGURE 2 The architecture of ARCliqueNet in one recurrent stage

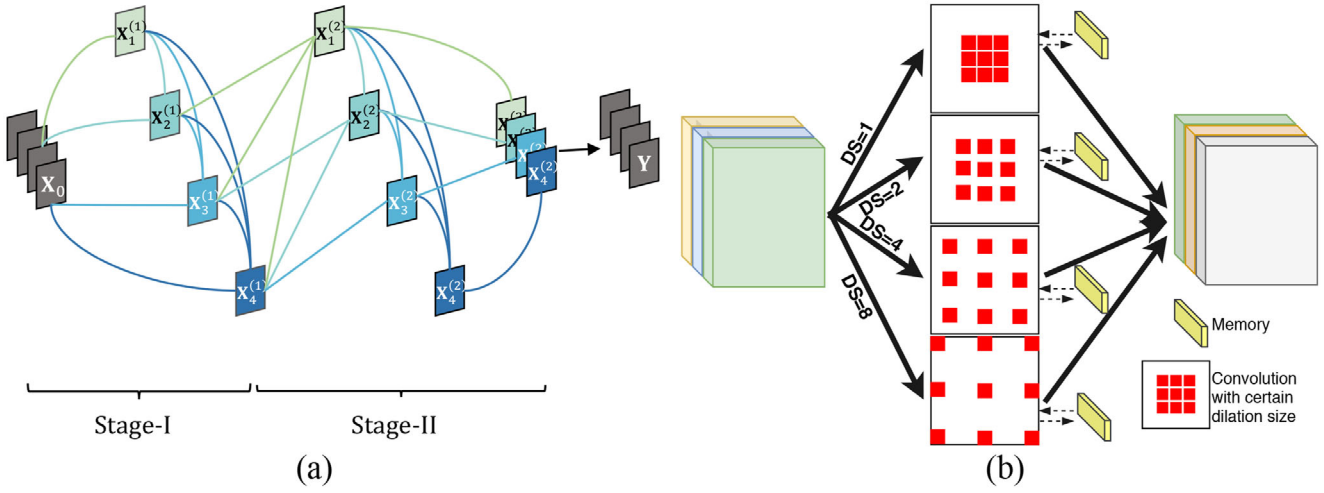

FIGURE 3 (a) Clique Block. (b) ASPP with memory

and decoder, recurrent mechanism, and ASPP with memory as information reconstruction net, which are the key parts of ARCliqueNet.

3.1 | Overview

As shown in Figure 2, our ARCliqueNet mainly consists of two modules: Clique Block [2] and recurrent mechanism including ASPP [1] with memory. Clique Block [2] is responsible for the conversion between images and a set of feature maps. ASPP [1] is originally designed to combine information from different receptive sizes. Due to the appearance of the recurrent mechanism, the memory mechanism is added to ASPP in order to adapt to it and exploit its potential.

3.2 | Clique Block as encoder and decoder

Clique Block [2] is showed in Figure 3(a). The reasons why Clique Block [2] is chosen as the main module are illustrated as follows. First, the forward propagation of Clique Block [2] contains two stages. The first stage does the same things as Dense Block [7]. Besides, the second stage distills the feature further. Second, compared with the Dense Block [7], the Clique Block [2] contains more skip connections, so the information flow among layers can be more efficient.

Suppose a Clique Block [2] has l layers, the channel of input is c_1, and the channel of each layer in a Clique Block [2] is c_2. The input and the output of the Clique Block [2] are denoted by $\mathbf{X}_0 \in \mathbb{R}^{c_1 \times h \times w}$ and $\mathbf{Y} \in \mathbb{R}^{(lc_2) \times h \times w}$, respectively. The weight between layer i and layer j is represented by $\mathbf{W}_{ij}$. The feed-forward pass of the Clique Block [2] can be mathematically described as the following equations. For stage one,

$$\mathbf{X}_i^{(1)} = \sigma\left(\sum_{k=1}^{i-1} \mathbf{W}_{ki} * \mathbf{X}_k^{(1)} + \mathbf{W}_{0i} * \mathbf{X}_0\right), \tag{1}$$

where $*$ is the convolution operation, σ is the activation function. For stage two,

$$\mathbf{X}_i^{(2)} = \sigma\left(\sum_{k=1}^{i-1} \mathbf{W}_{ki} * \mathbf{X}_k^{(2)} + \sum_{k=i+1}^{l} \mathbf{W}_{ki} * \mathbf{X}_k^{(1)}\right), \tag{2}$$

For the output of Clique Block,

$$\mathbf{Y} = \left[\mathbf{X}_1^{(2)}, \mathbf{X}_2^{(2)}, \mathbf{X}_3^{(2)}, \ldots, \mathbf{X}_l^{(2)}\right], \tag{3}$$

where $[\cdot]$ represents the concatenation operation.

A important difference between original Clique Block [2] and the Clique Block [2] in ARCliqueNet is that Clique Block [2] in ARCliqueNet has no batch normalisation (BN) [13] layers. Because BN [13] can reduce the internal covariate shift of feature maps, BN [13] is widely used in deep neural network's training stage. Through applying BN [13], each normalised scalar feature has zero mean and unit variance. So these features have the same distribution and are independent of each other. However, in the artifacts reduction problem, streak artifacts in different layers have different distributions in intensity, color, shapes and etc. Therefore, BN [13] contradicts the characteristics of our proposed model, and we choose to remove BN [13] from the original Clique Block [2].

Besides, since BN [13] has to keep a normalised copy of the feature map in GPU, removing BN [13] can also reduce the demand on GPU memory. This benefit can help us to enlarge model capacity or increase batch size.

3.3 | ASPP with memory

ASPP actually is an atrous version of spatial pyramid pooling (SPP), in which the concept has been used in SPPNet [14]. In ASPP, parallel atrous convolutions with different dilation sizes (DS) are applied in the input feature map and fuse the processed feature maps together. Some of the papers also call atrous convolution as dilated convolution. As the prostate has different scales and shapes in different objects' CT images, ASPP helps to account for different scales and shapes which can improve the precision.

As we remove the streak artifacts in multiple stages based on recurrent mechanism, useful information for reduction in previous stages can guide the learning in later stages. So we incorporate the RNN architecture with memory mechanism to make full use of the useful information in previous stages. In detail, the atrous convolutions in ASPP are replaced with convolutional GRU with dilation to fully exploit recurrent mechanism and ASPP's potential. ASPP with memory is used as an information reconstruction net and shown in Figure 3(b).

3.4 | Recurrent mechanism

As there are various streak artifacts of different characteristics in a sparse-view CT image, we think it is better to refine reconstructed dense-view CT images stage by stage. Meanwhile, the recurrent mechanism is suitable for decomposing the denoise task into multiple stages.

In each recurrent stage, our proposed model predicts the whole residual, that is, streak artifacts. Our scheme can be formulated as:

$$I_s^1 = I_{\text{ori}},\ H^0 = \text{None}, \tag{4}$$

$$\widehat{R^n} = F_n(I_s^n, H^{n-1}),\ 1 \le n \le N, \tag{5}$$

$$I_s^{n+1} = \widehat{I_d^n} = I_{\text{ori}} - \widehat{R^n}, \tag{6}$$

where I_{ori} indicates the original sparse-view CT image, I_s^n represents the input of the nth recurrent stage, N is the number of recurrent stages, F_n indicates the computing process of the nth recurrent stage, H^{n-1} represents the hidden states, that is, the feature maps of the previous stage in ASPP with memory, $\widehat{R^n}$ indicates the output of the nth recurrent stage, and $\widehat{I_d^n}$ is the predicted dense-view CT image as well as intermediate artifacts-free image after the nth recurrent stage.

The recurrent mechanism with N stages is shown in Figure 4. In order to illustrate the effect of recurrent mechanism, we also present the metric of different stages in Figure 4, and the difference between the previous stage and the later stage in Figure 5.

3.5 | Loss function

The loss function is defined as the sum of all recurrent stages' MSE loss, which is formulated as:

$$L(\Theta) = \sum_{n=1}^{N} \left\| \widehat{R^n} - R \right\|_F^2, \tag{7}$$

where Θ represents the network's parameters, R is the residual between the original sparse-view CT image and dense-view CT image, and $\widehat{R^n}$ indicates the output of the nth recurrent stage.

4 | EXPERIMENTS

4.1 | Dataset

We evaluate ARCliqueNet on a dataset from AAPM [3], which consists of 2,378 CT images from 10 patients and is the most commonly used benchmark dataset in artifact reduction field. For training, we use five male patients' data containing CT images from chest to hip, whose range includes prostate. For testing, we use the other patients' CT images only containing the prostate. For the training set, we use the 2D FBP reconstruction images from 60, 120, 180, and 240 projection views as input. And we adopt residual learning, so the residual images, that is, streak artifacts, are used as the label. The residual images are the difference between the dense-view (720 views) reconstructions and the sparse-view reconstructions.

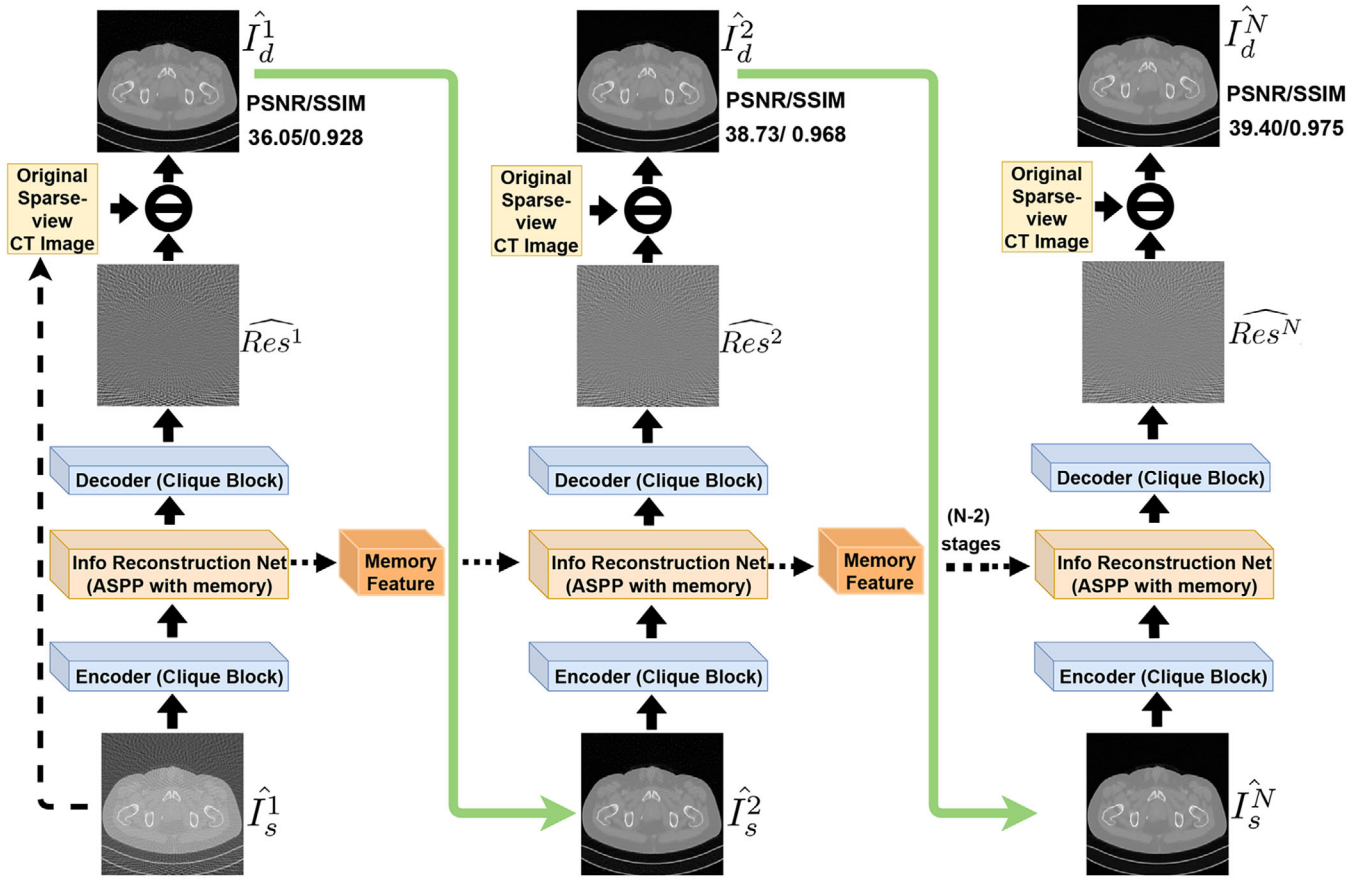

FIGURE 4 The architecture of ARCliqueNet of N recurrent stages

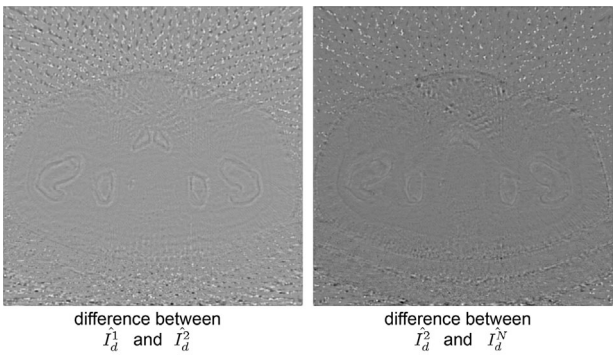

FIGURE 5 The difference between $\widehat{I_d^1}$ and $\widehat{I_d^2}$ as well as $\widehat{I_d^2}$ and $\widehat{I_d^N}$ in Figure 4

4.2 | Experimental setup

In our experiments, the size of each image in the system is 1*512*512 (1 is the number of channels). In order to avoid the influence of outliers, we normalise the dataset according to the upper and lower 25 points of all pixels' values. For a fair comparison, all architectures are trained by Adam algorithm. For evaluation metrics, we adopt SSIM and PSNR.

In our proposed ARCliqueNet, the layer number in a Clique Block [2] is 3, and the stage number is set as 2. And the dilation sizes in ASPP with memory is 1, 2, 4, 8, which ensure the fastest growth in receptive field size without missing any pixel in its receptive field. Unlike most neural network designed for computer vision, we avoid dropout [17], batch normalisation [13] and instance normalisation [18], because they reduce the flexibility of features and not suitable for the artifact reduction problem.

TABLE 1 Quantitative comparison between ARCliqueNet and other SOTA methods on AAPM's prostate CT dataset

SSIM/PSNR	60 views	120 views	180 views	240 views
FBP [15]	0.651/25.18	0.897/33.16	0.975/39.46	0.992/44.27
TF U-Net [11]	0.971/38.43	0.989/43.30	0.994/46.35	0.996/48.31
DD-Net [16]	0.967/37.56	0.986/41.72	0.991/44.33	0.993/45.56
ARCliqueNet	**0.974/39.07**	**0.991/44.56**	**0.995/47.75**	**0.997/49.55**

4.3 | Experimental results

4.3.1 | Comparison between ARCliqueNet and other SOTA models

In Table 1, we present the average PSNR and SSIM values of FBP, other SOTA methods, and ARCliqueNet. All methods are

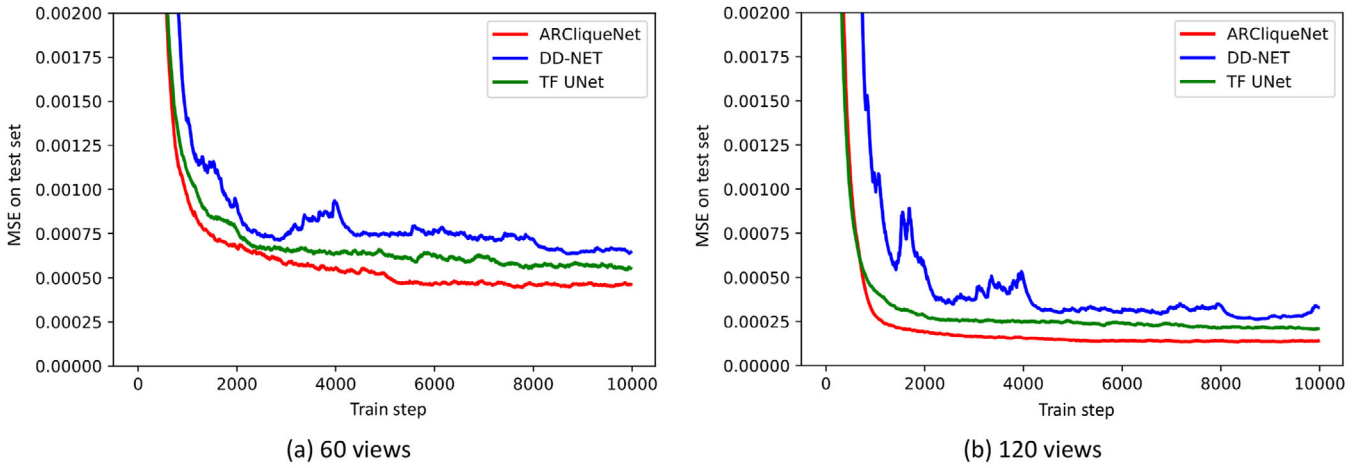

FIGURE 6 MSE curve of 60 and 120 views on test data during the training process

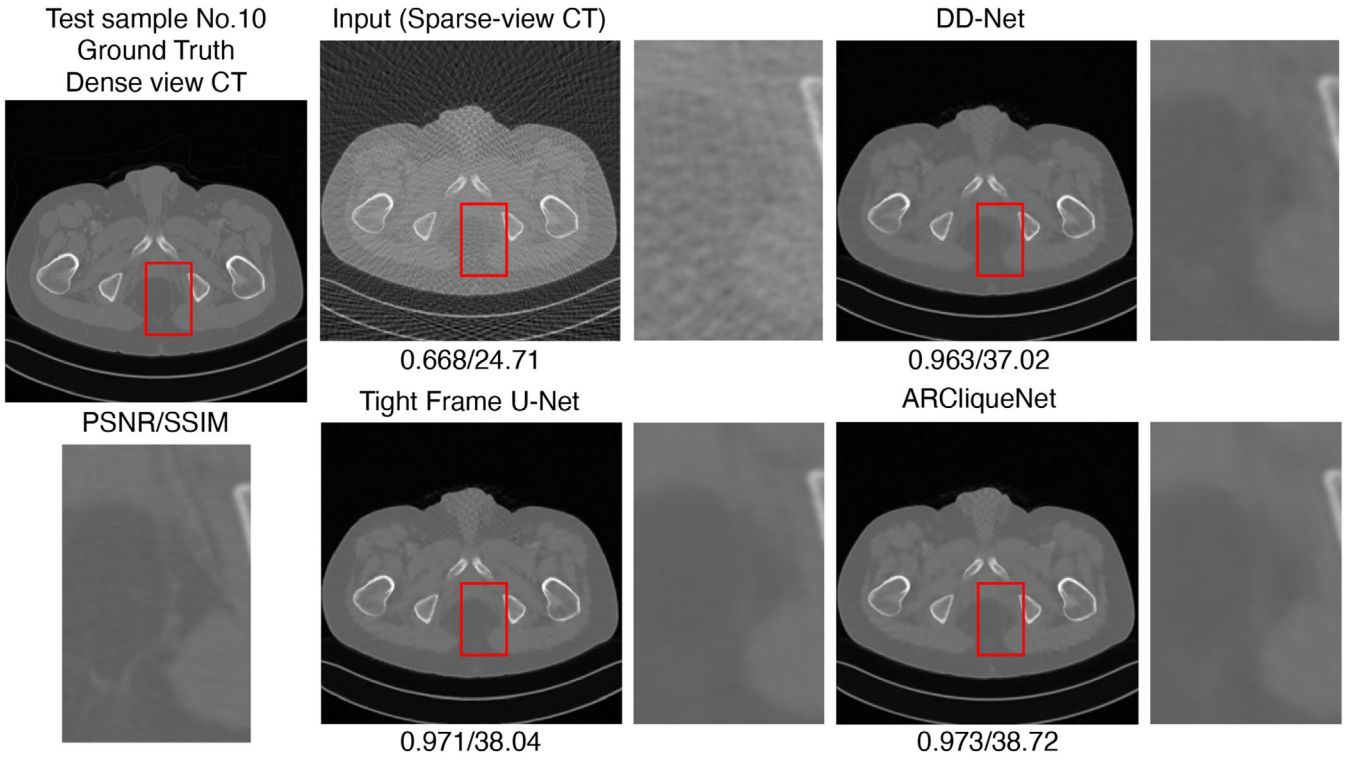

FIGURE 7 Qualitative comparison between ARCliqueNet and other SOTA methods on test sample No. 10 in AAPM's prostate CT dataset of 60 views. It is easy to find that the quality of ARCliqueNet's reconstructed dense-view CT images is better than Tight Frame U-Net's and DD-Net's

significantly better than FBP [15]. Our ARCliqueNet achieves much better results than the other two SOTA methods in all different views.

Besides, the MSE curve on test data during the training process is presented in Figure 6. We can observe that our ARCliqueNet converge at the lowest MSE value. We take the 60 and 120 views, for example, and the results of 180 and 240 views are similar to 60 and 120 views' MSE curves.

Qualitative results of 60 view are shown in Figures 7 and 8. In the figure, we can easily see the outputs of our ARCliqueNet have less noise and clearer boundaries, especially in the red box area, which is related to the prostate and its surroundings.

4.3.2 | Ablation study about Clique Block and recurrent mechanism with memorised ASPP

In this part, we conduct experiments to compare the effect of Clique Block [2] and recurrent mechanism with memorised ASPP [1]. Since ASPP with memory working with recurrent

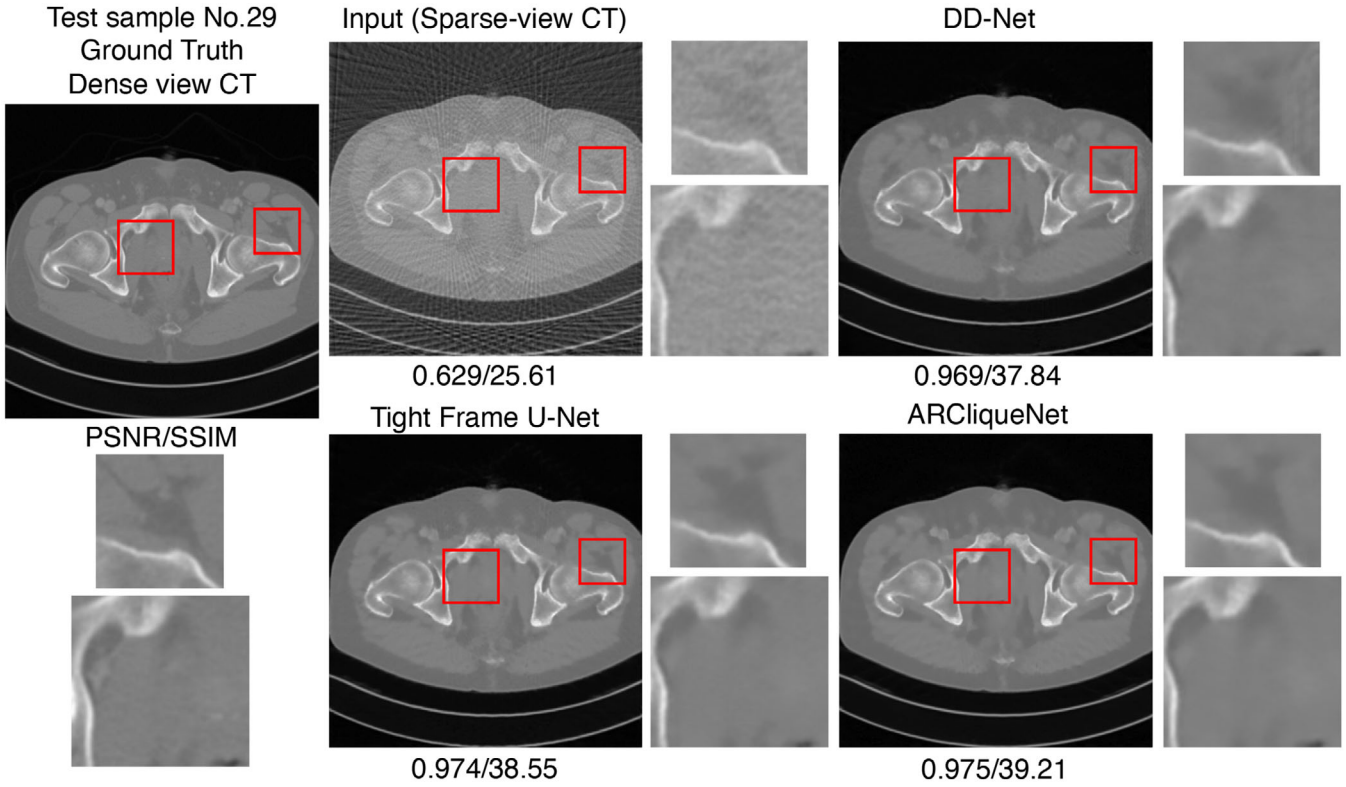

FIGURE 8 Qualitative comparison between ARCliqueNet and other SOTA methods on test sample No. 29 in AAPM's prostate CT dataset of 60 views. It is easy to find that the quality of ARCliqueNet's reconstructed dense-view CT images is better than Tight Frame U-Net's and DD-Net's

TABLE 2 Ablation study about the existence of Clique Block and recurrent mechanism with memorised ASPP on 120 views prostate CT dataset

Clique Block	PSNR/SSIM	Recurrent and ASPP with memory	PSNR/SSIM
✓	0.991/44.56	✓	0.991/44.56
×	0.989/44.07	×	0.985/42.86

TABLE 3 Quantitative comparison of the performance contribution of different recurrent stages

SSIM/PSNR	60 views	120 views	180 views	240 views
Stage 0 (input)	0.651/25.18	0.897/33.16	0.975/39.46	0.992/44.27
Stage 1	0.932/36.08	0.985/42.85	0.994/46.83	0.996/48.87
Stage 2	0.968/38.55	0.990/44.32	0.995/47.68	0.997/49.50
Stage 3	**0.974/39.07**	**0.991/44.56**	**0.995/47.75**	**0.997/49.55**

mechanism can play an effective role, we bind them together in the study. In Table 2, we report the ablation study's results. It is obvious that both Clique Block [2] and recurrent mechanism with memorised ASPP [1] are beneficial to the performance. In addition, the recurrent mechanism with memorised ASPP [1] contributes more than Clique Block [2].

4.3.3 | The performance contribution of different recurrent stages

In the previous subsection, we find that the recurrent mechanism with memorised ASPP [1] contributes more than Clique Block [2]. So, we conduct experiments to compare the performance contribution of different recurrent stages.

We report the SSIM and PSNR of different stages' output of different views in Table 3, where it can be concluded that every recurrent contributes to the final performance and later stages contribute less than earlier ones. This is because memory feature helps the previous recurrent stages guide the later stages' learning and the noise in later stages is minor than earlier stages. Therefore, we can draw the conclusion that recurrent mechanism with memorised ASPP benefits the refining of artifacts reduction stage by stage.

4.3.4 | Parameters cost and computational complexity

This kind of medical research makes sense only when it is put into application, so the parameters cost is fairly important. After calculating the number of parameters used in Tight Frame U-Net [11], DD-Net [16] and ARCliqueNet, we find out that ARCliqueNet costs only 43.97% parameters of DD-Net and 0.455% parameters of TF U-Net. The numerical results are illustrated in Table 4. Hence our ARCliqueNet needing less

TABLE 4 Parameters cost and computational complexity comparison among Tight Frame U-Net, DD-Net and our proposed ARCliqueNet

	Number of Params(K)	FLOPs(G)
TF U-Net [11]	40,837	290.68
DD-Net [16]	423	4.58
ARCliqueNet	186	49.57

TABLE 5 Parameters cost and computational complexity comparison of ARCliqueNet's different parts

	Params (K)	FLOPs (G)
Encoder	4.15	1.69
Info reconstruction net	177.32	46.15
Decoder	4.35	1.74
Sum	185.81	49.57

parameter to realise high performance is more suitable for real application than other SOTA methods.

As for computational complexity, Tight Frame U-Net [11] is 290.6 GFLOPs, DD-Net [16] is 4.58 GFLOPs and ARCliqueNet is 49.58 GFLOPs. The comparison is illustrated in Table 4. Due to the recurrent mechanism, our ARCliqueNet needs forward propagation three times in our experiments. Besides, our ARCliqueNet does not downsample the feature map. These two points result in that the ARCliqueNet needs more FLOPs compared with DD-Net [16], but our proposed model still needs fewer FLOPS than Tight Frame U-Net [11].

In addition, the parameters cost and computational complexity comparison of ARCliqueNet's different parts are presented in Table 5. We can observe that the information reconstruction net takes up most of the parameters cost and computational complexity.

5 | CONCLUSION

In this work, we investigate the utilisation of Clique Block and recurrent mechanism with memorised ASPP for sparse-view CT slice reconstruction. Since the prostate has a wide variation in size and shape among different subjects due to pathological changes, we introduce Clique Block to further extract useful information. Besides, for the lack of clear prostate boundaries, we bring in recurrent mechanism and ASPP with memory to get multi-scale information and refine the reconstructed dense-view CT images stage by stage. The experimental results show that our proposed ARCliqueNet achieves very good performance on the streak artifacts reduction task of prostate CT images. We are sure that the Clique Block and ASPP with memory can be easily applied to many more medical tasks and achieve high performance.

REFERENCES

1. Chen, L.-C., et al.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. IEEE Trans. Pattern Anal. Mach. Intell. 40(4), 834–848 (2017)
2. Yang, Y., et al.: Convolutional neural networks with alternately updated clique. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Utah, pp. 2413–2422 (2018)
3. AAPM Low Dose CT Grand Challenge. https://www.aapm.org/grandchallenge/lowdosect/ (2019). Accessed 3 July 2019
4. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, pp. 1–9 (2015)
5. Zagoruyko, S., Komodakis, N.: Wide residual networks. In: Proceedings of the British Machine Vision Conference, York, UK, pp. 87.1–87.12 (2016)
6. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, pp. 770–778 (2016)
7. Huang, G., et al.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, pp. 4700–4708 (2017)
8. Sidky, E.Y., Pan, X.: Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Phys. Med. Biol. 53(17), 4777 (2008)
9. Liu, Y., et al.: Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. Phys. Med. Biol. 57(23), 7923 (2012)
10. Chen, Y., et al.: Artifact suppressed dictionary learning for low-dose CT image processing. IEEE Trans. Med. Imaging 33(12), 2271–2292 (2014)
11. Han, Y., Ye, J.C.: Framing U-Net via deep convolutional framelets: Application to sparse-view CT. IEEE Trans. Med. Imaging 37(6), 1418–1429 (2018)
12. Kofler, A., et al.: A U-Nets cascade for sparse view computed tomography. In: International Workshop on Machine Learning for Medical Image Reconstruction, Springer, Cham, pp. 91–99 September (2018)
13. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceedings of the International Conference on International Conference on Machine Learning, Lille, France, pp. 448–456 (2015)
14. He, K., et al.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Machine Intell. 37(9), 1904–1916 (2015)
15. Kak, A.C., et al.: Principles of computerized tomographic imaging. Med. Phys. 29(1), 107–107 (2002)
16. Zhang, Z., et al.: A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. IEEE Trans. Med. Imaging 37(6), 1407–1417 (2018)
17. Srivastava, N., et al.: Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15(1), 1929–1958 (2014)
18. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, pp. 1501–1510 (2017)

How to cite this article: Shen T, Yang Y, Lin Z, Zhang M. Recurrent learning with clique structures for prostate sparse-view CT artifacts reduction. *IET Image Processing*. 2021;1–8.
https://doi.org/10.1049/ipr2.12048