CT Image Super-Resolution Reconstruction via Pixel-Attention Feedback Network

Jianrun Shang, Guisheng Zhang, Wenhao Song, Mingliang Gao

School of Electrical and Electronic Engineering,
Shandong University of Technology,
Zibo, China
E-mail: jianrun_shang@163.com
E-mail:sdut_guisheng@163.com
E-mail: sdut_songwenhao@163.com
E-mail: mlgao@sdut.edu.cn

Qilei Li

School of Electronic Engineering and Computer Science,
Queen Mary University of London,
London, United Kingdom
E-mail: qilei.li@outlook.com

Jinfeng Pan*

School of Electrical and Electronic Engineering,
Shandong University of Technology,
Zibo, China
E-mail: pjfbysj@163.com
*Corresponding author

Abstract: Computed tomography (CT) imaging has been widely used in clinical medicine, and high-resolution CT images play a crucial role in the determination of lesions. To fully excavate the contributive information of initial features and improve the feature representation ability of the model, we propose a Pixel-Attention Feedback Network (PAFNet) for CT image super-resolution reconstruction. Specifically, the PAFNet adopts multi-feedback network as backbone to make full use of initial features. Subsequently, a gated feedback (GF) block is introduced to refine the underlying features using the feedback features. To enrich the output characteristics and pay attention to essential details, a pixel attention mechanism is adopted to the self-calibration convolution. The subjective and objective evaluation demonstrate the superiority of the proposed method over the state-of-the-art approaches.

Keywords: Super-Resolution; CT Image; Pixel Attention; Feedback Network.

Biographical notes:

Jianrun Shang is studying for the M.S. degree at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His

research interests include image super-resolution and deep learning.

Guisheng Zhang is studying for the M.S. degree at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include fake face and deep learning.

Wenhao Song is studying for the M.S. degree at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include computer vision and deep learning.

Mingliang Gao received his Ph.D. in communication and information systems from Sichuan University, Chengdu, China, in 2013. He is currently an associate professor at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His main research interests include computer vision and deep learning.

Qilei Li is currently a Ph.D. student with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom. He received an M.S. degree in signal and information processing from Sichuan University. His research interests are computer vision and deep learning.

Jinfeng Pan received the Ph.D. degree in singal and information processing at University of Chinese Academy of Sciences, Xian, China, in 2016. From 2005, she works in Dept. of electrical and electronic information engineering, Shandong University of Technology, Zibo, China. Her main research interests include deep learning and image processing.

1 Introduction

Computed tomography (CT) imaging has been broadly applied in medical diagnosis. A high-resolution image of organs and tissues is crucial for accurate disease analysis. However, because of the restriction of hardware facilities, the generated medical images often have low resolution and even lose some essential structural information.

Recently, many effective methods have been proposed to enhance the image resolution Li et al. (2020). The early methods Chang et al. (2004); Anbarjafari & Demirel (2010); Liu et al. (2013); Kumar & Singh (2013); Ruangsang & Aramvith (2017); Shen et al. (2009); Zhang et al. (2010, 2012) are based on interpolation or reconstruction. Although these methods are simple, they are often unable to retain fine detail information due to their limited ability to extract features. Subsequently, deep learning-based super-resolution methods have been the mainstream due to the powerful ability of feature representation. The image super-resolution using deep convolutional networks (SRCNN) Dong et al. (2014) is the cornerstone of deep learning application in the domain of image super resolution. Kim *et al.* Kim et al. (2016) improved the learning rate of SRCNN and eliminated the obstacles of gradient propagation by adopting residual learning. In recent years, the introduction of attention mechanism has greatly improved the performance of the network due to the recalibration of the feature responses towards the most informative and important components of the inputs Zhai et al. (2022*a*,*c*); Guo et al. (2022); Zhai et al. (2022*b*). Zhang *et al.* Zhang et al. (2018*a*) proposed

the residual channel attention network (RCAN) and applied the attention mechanism in the super-resolution domain.

To fully excavate the contributive information of intermediate features and improve the feature representation ability, we propose the PAFNet for CT image super-resolution. Compared with the single-feedback network Dong et al. (2014); Zhang et al. (2018*a*) which only propagates the highest-level features to the shallow layer, ignoring other high-level information captured in the receptive field of different sizes, the PAFNet employs a multi-feedback network as the backbone. The multi-feedback network can transfer several advanced features to the shallow layer, so that the reconstructed image contains valuable context information. Meanwhile, the PAFNet contains two key components, namely gated feedback (GF) block and pixel attention based on self-calibrated convolution (PASC). The feedback connection of two adjacent subnetworks can be established through multiple GF blocks and skip connections. The GF block can refine low-level features using high-level features, which are transported via the feedback connections. The PASC block is obtained by replacing the pooling layer and upsampling in self-calibration convolution with pixel convolution unit. To improve the feature representation ability of the model, the PASC block is built to extract the detail features and enhance the diversity of output features.

The remainder of the paper is described as follows. Section 2 presents the previous work. Section 3 introduces the architecture and details of the PAFNet. Section 4 verifies the effectiveness by comparative experiments. Section 5 is the conclusion.

2 Related work

Image super-resolution technology has attracted the attention of researchers considering its broad application prospects. Therefore, the researchers have invested a lot of effort and proposed many effective methods, which can be roughly categorized into two branches, namely conventional algorithms and deep learning-based algorithms Dai et al. (2019).

The early conventional methods are basically based on bicubic interpolation. Benefiting by the low complexity and effectiveness, they became the mainstream at the time. However, these methods inevitably lead to image edge artefacts. To address this problem, Kumar *et al.* Kumar & Singh (2013) proposed an algorithm using discrete wavelet transform and interpolation successively.

The image super-resolution using deep convolutional networks Dong et al. (2014) is the first work of deep learning application in the field of super resolution. In order to achieve lightweight design and save computing resources, Dong *et al.* Dong et al. (2016) proposed the accelerating super-resolution convolutional neural network (FSRCNN) by adding a deconvolution layer and using a smaller filter on the basis of SRCNN. The residual dense network (RDN) Zhang et al. (2018*b*) achieves dense feature fusion and depth monitoring via utilizing local and global features adequately. Li *et al.* Li et al. (2019*b*) proposed a super-resolution feedback network (SRFBN) to refine the representation of low-level features through the feedback of high-level features. The attention mechanism can boost the feature representation ability in deep learning-based methods. The second-order attention network (SAN) Dai et al. (2019) was proposed to adjust channel features to obtain discriminative features. Improving convolutional networks with self-calibrated convolutions (SCNet) Liu et al. (2020) can adaptively establish remote spatial and channel dependencies around each spatial location through self-correcting operations.

3 The proposed approach

3.1 Network framework

The framework of PAFNet is depicted in Fig. 1. It includes a series of subnetworks, each of which containing five blocks, namely low resolution feature extraction (LRFE) block, gated feedback (GF) block, residual density (RD) block, pixel attention based on self-calibrated convolution (PASC) block and pixel attention based reconstruction (PAR) block.

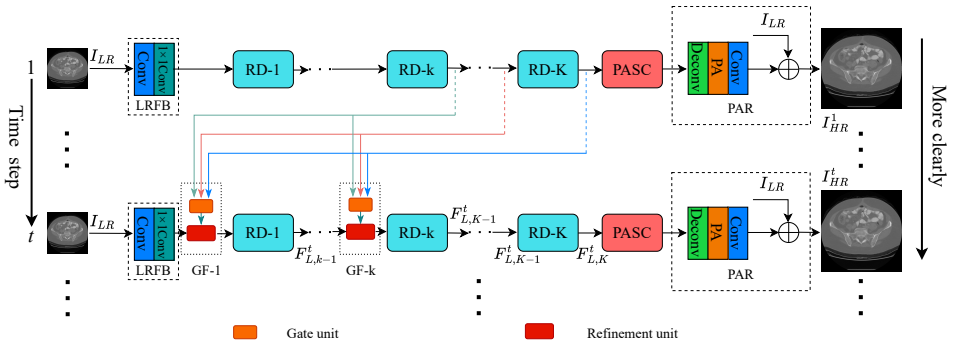


Figure 1: The framework of the proposed PAFNet for CT image super-resolution reconstruction.

The low-resolution image I_{LR} is fed to the LRFE block to extract shallow feature $F_{L,0}^{t}$. The operation of LRFE block is formulated as

$$F_{L,0}^{t} = H_{\text{LRFE}}\left(I_{\text{LR}}\right), \tag{1}$$

where $H_{\text{LRFE}}\left(\cdot\right)$ represents the function of the LRFE block. Then, the feature $F_{L,0}^{t}$ is transferred to multiple RD blocks to generate features at different levels.

The feature representation ability of a subnetwork is closely related to cascaded RD blocks in the subnetwork. The feedback connection between two adjacent subnetworks is realized by the RD blocks and the GF blocks. The high-level feature $F_{L,K}^{t}$ is defined as

$$F_{L,K}^{t} = H_{\text{GF-RD}}\left(F_{L,0}^{t}\right), \tag{2}$$

where $H_{\text{GF-RD}}\left(\cdot\right)$ is the combining function of RD blocks and GF blocks. $F_{L,K}^{t}$ denotes the output of the K-th RD block in the t-th subnetwork. L symbolizes the low-resolution space.

Then, $F_{L,K}^{t}$ is fed to PASC block to generate more powerful features. The final refined feature F_{PASC}^{t} is formulated as

$$F_{\text{PASC}}^{t} = H_{\text{PASC}}\left(F_{L,K}^{t}\right), \tag{3}$$

where $H_{\text{PASC}}\left(\cdot\right)$ represent the functions of PASC block.

Subsequently, we utilize PAR block to upsample the refined features F^t_{PASC} to generate the final results I_{SR}, which is formulated as

$$I_{SR} = H_{\text{PAR}}\left(F^t_{L,K}, I_{\text{LR}}\right), \tag{4}$$

where $H_{\text{PAR}}(\cdot)$ denotes the function of the PAR block, and I^t_{SR} is the super-resolution image generated by the t-th subnetwork. The PAR block consists of a deconvolution layer, a pixel attention layer, and a 3×3 convolutional layer.

The L_1 loss function is adopted to train the proposed PAFNet model. It is denoted as

$$L_1(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left\|I^t_{\text{HR}} - I^t_{\text{SR}}\right\|_1, \tag{5}$$

where T denotes the number of subnetworks. I^t_{HR} indicates the high-resolution image in the t-th subnetwork, and θ symbolizes the parameter to be trained.

3.2 PASC block

To generate a weighted feature matrix ($C \times H \times W$) in three dimensions, we employ the pixel attention Zhao et al. (2020). Pixel attention generates the attention coefficients of all the pixels in the feature map, which helps preserve the subtle detail features and the diversity of the output features. As depicted in Fig. 2, the pixel attention contains a 1×1 convolution layer and a sigmoid function. Then, the input features will multiply with available attention maps to obtain the final weighted feature maps.

Suppose F_{k-1} and F_k represent the input and output of pixel attention, respectively. The mathematical formulation of the pixel attention can be expressed as

$$F_k = H_{PA}\left(F_{k-1}\right) \cdot F_{k-1}, \tag{6}$$

where $H_{PA}(\cdot)$ represents the operation of the convolution layer and sigmoid function.

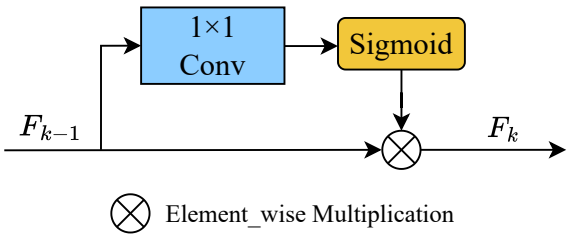


Figure 2: Structure of the pixel attention block

To enhance the feature representation, we build the pixel attention based on self-calibrated convolution (PASC) block, which is depicted in Fig. 3. The PASC block has a similar network architecture with SCNet Liu et al. (2020). It includes two branches, and each branch has a 1×1 convolution layer at the beginning, which is called $f_{split}(\cdot)$. Given the input high-level feature $F^t_{L,K}$, the two branches are denoted as

$$F^{t,1}_{L,K} = f^1_{split}\left(F^t_{L,K}\right), \tag{7}$$

$$F_{L,K}^{t,2} = f_{split}^{2}\left(F_{L,K}^{t}\right). \tag{8}$$

Where $F_{L,K}^{t,1}$ and $F_{L,K}^{t,2}$ denote the two intermediate features, and their channel number are half of $F_{L,K}^{t}$.

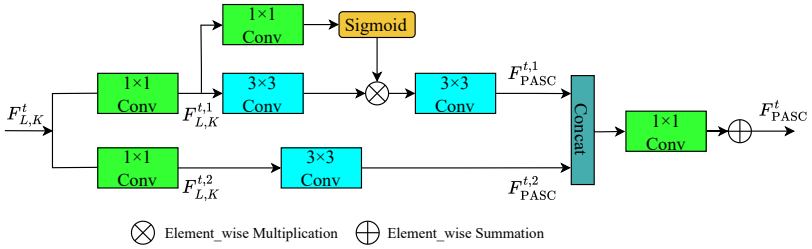


Figure 3: Structure of the PASC block.

The upper branch contains two 3×3 convolutional layers, and only the first layer is equipped with pixel attention. The lower branch sustains the initial features by concatenating 3×3 and 1×1 convolution layers. The upper branch converts the high-level feature $F_{L,K}^{t}$ to the split feature $F_{\text{PASC}}^{t,1}$, and the lower branch converts the high-level feature $F_{L,K}^{t}$ to the split feature $F_{\text{PASC}}^{t,2}$. At last, $F_{\text{PASC}}^{t,1}$ and $F_{\text{PASC}}^{t,2}$ are concatenated and transferred to a 1×1 convolution layer to obtain the refined feature F_{PASC}^{t}.

3.3 Feedback connection

The feedback connection of two adjacent subnetworks is achieved through multiple GF blocks and skip connections. As depicted in Fig. 1, the GF block contains two key subassemblies, namely gate unit and refinement unit. The GF blocks is adopted to take advantage of multiple high-level features to refine the low-level features Li et al. (2019*a*). The high-level features are transported via feedback connection, and the low-level features are extracted from shallow RD blocks. GF blocks are typically placed in front of shallow RD blocks. The outputs of these RD blocks are low-level features which need to be refined. The outputs of the deep RD blocks are identified as high-level features, which are fed to the GF blocks at the next time step through the long skip connection. The gate unit is to select the conducive information from high-level features, and the refinement unit will use the information to achieve the refinement of low-level features.

Given that $S_M = \{1, 2, ..., M-1, M\}$ is the set of selected indexes of the shallowest RD blocks and the input of S_M is low-level features. The $D_N = \{N, N+1, ..., K-1, K\}$ is regarded as the set of selected indexes of the deepest $(K - N + 1)$ RD blocks. The output of D_N is utilized to refine the low-level features. At the t-th time, if $k \in S_M$ and $t > 1$, the output of the k-th RD blocks $F_{L,k}^{t}$ can be obtained via

$$F_{L,k}^{t} = H_{\text{RD,k}}\left(H_{\text{RU,k}}\left(\left[F_{H,k}^{t}, F_{L,k-1}^{t}\right]\right)\right). \tag{9}$$

In other cases, the output of the k-th RD block can be derived by

$$F_{L,k}^{t} = H_{\text{RD,k}}\left(F_{L,k-1}^{t}\right), \tag{10}$$

where $H_{\mathrm{RD,k}}(\cdot)$ denotes the operation of the k-th RD block, and $H_{\mathrm{RU,k}}(\cdot)$ symbolizes the function of the refinement unit. $\left[F_{H,k}^{t}, F_{L,k-1}^{t}\right]$ represents the combination of $F_{H,k}^{t}$ and $F_{L,k-1}^{t}$. $F_{H,k}^{t}$ represents the high-level information which is selected and enhanced from high-level features. It is transmitted to the k-th GF block. The high-level features are collected by the deepest RD blocks and then delivered via a series of feedback connections. Hence, the high-level information $F_{H,k}^{t}$ is calculated as

$$F_{H,k}^{t}=\begin{cases} H_{\mathrm{GU,k}}\left(\left[F_{L,N}^{t-1},...,F_{L,K}^{t-1}\right]\right), & \text{if } k<N, \\ H_{\mathrm{GU,k}}\left(\left[F_{L,k}^{t-1},...,F_{L,K}^{t-1}\right]\right), & \text{otherwise}, \end{cases} \tag{11}$$

where $H_{\mathrm{GU,k}}(\cdot)$ is the operation of the gate unit in the k-th GF block. To reduce the computational burden, each of the gate unit and refinement unit consists of a 1×1 convolution layer.

4 Experiments and discussion

4.1 Datasets

We perform the evaluation on two large-scale medical image datasets, *i.e.,* Low Dose CT dataset [a] and QIN LUNG CT dataset [b]. The Low Dose CT (LDCT) McCollough et al. (2020) is collected from 299 clinically performed CT examinations on patients. We group the samples into two splits, *i.e.* LDCT Part_A and LDCT Part_B, according to the scanned position. The former split includes 2,272 chest images (1,822 images for training and 450 images for test). The later split contains 1,132 abdominal scan images (892 images for training and 240 images for test). The QIN LUNG CT dataset Kalpathy-Cramer et al. (2015) contains 3954 images, which are published in the TCIA Cancer Imaging Archive. We select 328 images for training and 150 images for test from the QIN LUNG CT dataset.

4.2 Evaluation metrics

Two evaluation metrics, *i.e.,* peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) Wang et al. (2004); Li et al. (2022), are adopted for objective assessments. The PSNR is the ratio between the powers of peak value and noise Georgescu et al. (2022). It is denoted as:

$$PSNR\left(I_{SR},I_{HR}\right)=10\cdot\log 10\times\left(\frac{L^{2}}{\frac{1}{N}\sum_{i=1}^{N}\left(I_{SR}\left(i\right)-I_{HR}\left(i\right)\right)^{2}}\right), \tag{12}$$

where L represents the maximum pixel, and N denotes the number of all pixels in I_{SR} and I_{HR}. The SSIM is proposed under the assumption that the human perceptual system has a strong adaptability to the image structure. The SSIM takes the structural similarity

[a][Online]. Available: https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=52758026

[b][Online]. Available: https://wiki.cancerimagingarchive.net/display/Public/QIN+LUNG+CT

into account by combining the contrast, luminance and texture of the images. And it can be formulated as:

$$SSIM\,(x,y) = \frac{2u_x u_y + k_1}{u_x^2 + u_y^2 + k_1} \cdot \frac{\sigma_{xy} + k_2}{\sigma_x^2 + \sigma_y^2 + k_2}, \tag{13}$$

where x, y represent two images, σ_{xy} symbolizes the covariance between x and y, u and σ^2 represent the average value and variance, and k_1, k_2 denote constant relaxation terms. Higher scores of the PSNR and SSIM denote better reconstruction effect.

4.3 Implementation details

The deep-learning framework configurations are PyTorch 1.8.0, CUDAToolkit 10.2, cuDNN 8.1.1, Python 3.8, and two paralleled NVIDIA 3060 GPUs. The medical LR image is cropped into 16 image patches with size of 48×48 in every iteration. Adam Kingma & Ba (2014) is adopted to optimize the proposed PAFNet. The original learning rate is set to 2×10^{-4}, and it is halved every 2×10^5 iteration.

4.4 Comparative analysis

To evaluate the performance of the PAFNet, we conduct comparative experiments with some mainstream methods, *e.g.,* SRCNN Dong et al. (2014), FSRCNN Dong et al. (2016), RDN Zhang et al. (2018*b*), and SRFBN Li et al. (2019*b*). The objective evaluation and subjective results are evaluated on the LDCT Part_A, LDCT Part_B and QIN LUNG CT test sets with scale factors of $\times 2$, $\times 3$, and $\times 4$.

Table 1 Comparative results on the LDCT Part_A, LDCT Part_B McCollough et al. (2020) and QIN LUNG CT Kalpathy-Cramer et al. (2015) test sets with scale factor $\times 2$. The best rests are highlighted in **bold**.

Methods	LDCT Part_A		LDCT Part_B		QIN LUNG CT	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SRCNNDong et al. (2014)	43.79	0.9822	33.82	0.9488	34.20	0.9352
FSRCNNDong et al. (2016)	44.14	0.9837	34.39	0.9502	35.65	0.9362
RDNZhang et al. (2018*b*)	44.53	0.9841	35.05	0.9508	37.15	0.9401
SRFBNLi et al. (2019*b*)	47.32	0.9878	35.41	0.9523	38.49	0.9800
PAFNet (Ours)	**48.39**	**0.9885**	**35.75**	**0.9544**	**40.77**	**0.9830**

The results of PSNR and SSIM on LDCT and QIN LUNG CT test sets are illustrated in Table 1, 2 and 3. The results prove that PAFNet acquires the best performance on the LDCT Part_A, LDCT Part_B and QIN LUNG test sets. The PSNR and SSIM scores of the proposed PAFNet on three test sets with scale factors of $\times 2$ are (45.01, 0.9814), (31.91, 0.8864) and (36.19, 0.9569) respectively. Compared with SRCNN Dong et al. (2014) which is the cornerstone of deep learning in the field of super resolution, the proposed method increases the score of PSNR by 4.6 dB and SSIM by 0.0063 on the LDCT Part_A test set. On Part_B test set, the PAFNet improves the score of PSNR by 1.36 dB and SSIM by 0.0042 compared with the FSRCNN Dong et al. (2016). With scale factors of $\times 3$, the proposed PAFNet ranks the first place in both PSNR and SSIM compared with other SOTA

Table 2 Comparative results on the LDCT Part_A, LDCT Part_B McCollough et al. (2020) and QIN LUNG CT Kalpathy-Cramer et al. (2015) test sets with scale factor $\times 3$. The best rests are highlighted in **bold**.

Methods	LDCT Part_A		LDCT Part_B		QIN LUNG CT	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SRCNNDong et al. (2014)	39.12	0.9633	29.86	0.8072	31.85	0.8578
FSRCNNDong et al. (2016)	38.87	0.9623	30.19	0.8116	32.28	0.8612
RDNZhang et al. (2018*b*)	44.70	0.9668	31.79	0.8853	33.24	0.8911
SRFBN Li et al. (2019*b*)	44.16	0.9804	31.75	0.8843	34.55	0.9512
PAFNet (Ours)	**45.01**	**0.9814**	**31.91**	**0.8864**	**36.19**	**0.9569**

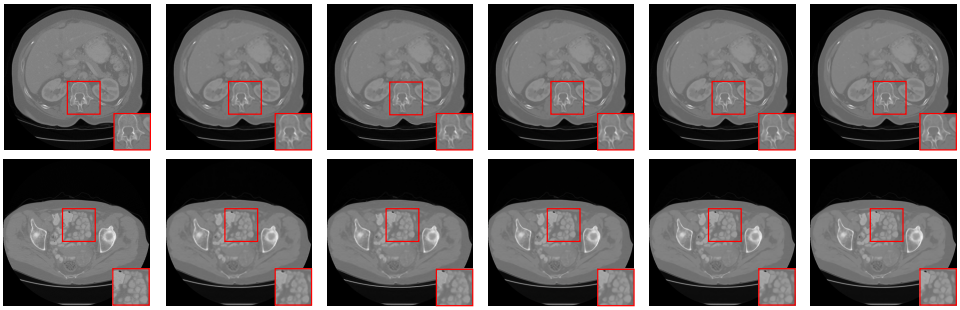
Table 3 Comparative results on the LDCT Part_A, LDCT Part_B McCollough et al. (2020) and QIN LUNG CT Kalpathy-Cramer et al. (2015) test sets with scale factor $\times 4$. The best rests are highlighted in **bold**.

Methods	LDCT Part_A		LDCT Part_B		QIN LUNG CT	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SRCNNDong et al. (2014)	36.63	0.9465	28.46	0.8337	27.48	0.8381
FSRCNNDong et al. (2016)	37.06	0.9363	28.49	0.8215	27.55	0.8668
RDNZhang et al. (2018*b*)	40.78	0.9546	29.83	0.8346	30.43	0.8462
SRFBN Li et al. (2019*b*)	41.05	0.9714	30.06	0.8398	31.78	0.9226
PAFNet (Ours)	**42.55**	**0.9748**	**30.22**	**0.8420**	**33.46**	**0.9339**

methods. The mean PSNR and SSIM values obtained by PAFNet are increased by 5.49 dB and 0.0385 compared with FSRCNN Dong et al. (2016) on the LDCT Part_A test set. On the LDCT Part_B test set, the values of PSNR and SSIM are 0.12 dB and 0.0011 higher than RDN Zhang et al. (2018*b*). Moreover, the proposed method also acquires the highest PSNR and SSIM scores on three test sets with scale factor of $\times 4$. Specifically, the proposed PAFNet improves the PSNR and SSIM by 2.95 dB and 0.0658 on the QIN LUNG CT test set, respectively. On the QIN LUNG CT test set, the proposed PAFNet improves PSNR and SSIM by 1.5 dB and 0.0034 compared with SRFBN Li et al. (2019*b*) which also employs the feedback connections.

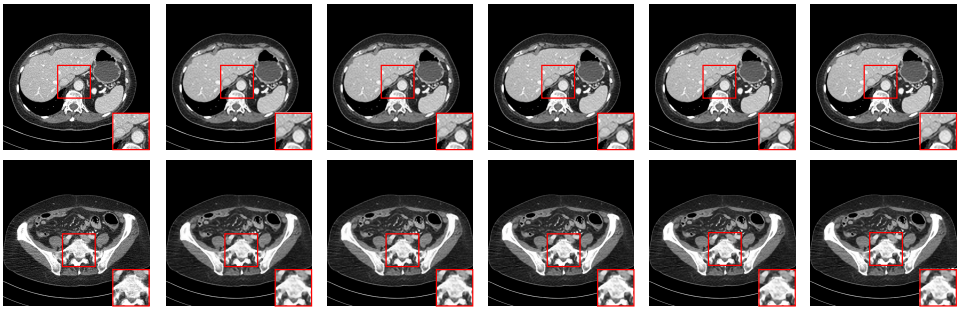
Fig. 4 illustrates the results of the subjective visual comparison on LDCT Part_A test set with scale factor of $\times 4$. The first row of Fig. 4 depicts that the abdominal scan images reconstructed by RDN Zhang et al. (2018*b*), SRFBN Li et al. (2019*b*) and the PAFNet are sharper than those reconstructed by SRCNN Dong et al. (2014) and FSRCNN Dong et al. (2016). Although there is visible difference in resolution ratio of the abdominal scan images reconstructed by RDN Zhang et al. (2018*b*), SRFBN Li et al. (2019*b*) and the PAFNet, the contrast of reconstructed image by PAFNet is enhanced. The second row of Fig. 4 shows that the intestinal tract scan image reconstructed by the proposed PAFNet has sharper edges compared with those images reconstructed by SRCNN Dong et al. (2014), FSRCNN Dong et al. (2016), RDN Zhang et al. (2018*b*) and SRFBN Li et al. (2019*b*). The reconstructed images of SRCNN Dong et al. (2014) and FSRCNN Dong et al. (2016) contain noise and have edge overlap. Fig. 5 depicts the subjective results on LDCT Part_B test set with scale factor of $\times 4$. The first row of Fig. 5 depicts that the image reconstructed by the proposed

PAFNet has a clearer representation of blood vessels in the human liver region compared with those images reconstructed by other methods. The second row of Fig. 5 proves that the lumbar disc scan image reconstructed by the proposed PAFNet can retain the bony details on the surface of the lumbar spine while improving the resolution. The images reconstructed by other methods can improve the resolution but lose the bony details of the lumbar surface, which is not conducive to the diagnosis of lumbar diseases. Fig. 6 is the visualization results on QIN LUNG CT test set with scale factor of $\times 4$. Human lung texture feature is difficult to capture. As can be seen from the first row of Fig. 6, the image reconstructed by the proposed method contains numerous textures and have a favourable resolution. In the second row of Fig. 6, the femoral head site scan images reconstructed by SRCNN Dong et al. (2014), FSRCNN Dong et al. (2016), RDN Zhang et al. (2018*b*) and SRFBN Li et al. (2019*b*) contain artefacts to some extends. The reconstructed image by PAFNet is clear for the representation of femoral head periphery.



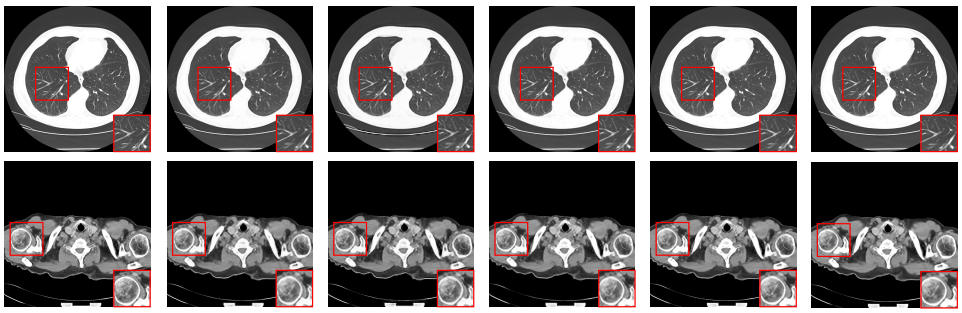
(a) HR (b) SRCNN (c) FSRCNN (d) RDN (e) SRFBN (f) Ours

Figure 4: Qualitative comparisons on LDCT Part_A dataset



(a) HR (b) SRCNN (c) FSRCNN (d) RDN (e) SRFBN (f) Ours

Figure 5: Qualitative comparisons on LDCT Part_B dataset



(a) HR (b) SRCNN (c) FSRCNN (d) RDN (e) SRFBN (f) Ours

Figure 6: Qualitative comparisons on QIN LUNG CT dataset

5 Conclusion

In this paper, we propose a Pixel-Attention Feedback Network (PAFNet) for CT image super-resolution reconstruction. The proposed PAFNet employs a multi-feedback network as the backbone to transmit multiple intermediate high-level features so as to refine the low-level features. It consists of two components, namely a gated feedback (GF) block and a pixel attention based on the self-calibrated convolution (PASC) block. The GF block can refine the low-level features using high-level features delivered by feedback connection. In addition, the PASC block enables the model to generate discriminative features, and enhance the diversity of output features and grasp delicate details. The comparative experiments on LDCT and QIN LUNG CT datasets illustrate that the proposed PAFNet outperforms the state-of-the-art methods objectively and subjectively.

References

Anbarjafari, G. & Demirel, H. (2010), 'Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image', *ETRI journal* **32**(3), 390–394.

Chang, H., Yeung, D.-Y. & Xiong, Y. (2004), Super-resolution through neighbor embedding, *in* 'Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.', Vol. 1, IEEE, pp. I–I.

Dai, T., Cai, J., Zhang, Y., Xia, S.-T. & Zhang, L. (2019), Second-order attention network for single image super-resolution, *in* 'Proceedings of the IEEE/CVF conference on computer vision and pattern recognition', pp. 11065–11074.

Dong, C., Loy, C. C. & Tang, X. (2016), Accelerating the super-resolution convolutional neural network, *in* 'European conference on computer vision', Springer, pp. 391–407.

Dong, C., Loy, C. C., He, K. & Tang, X. (2014), Learning a deep convolutional network for image super-resolution, *in* 'European conference on computer vision', Springer, pp. 184–199.

Georgescu, M.-I., Ionescu, R. T., Miron, A.-I., Savencu, O., Ristea, N.-C., Verga, N. & Khan, F. S. (2022), 'Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution', *arXiv preprint arXiv:2204.04218*.

Guo, X., Gao, M., Zhai, W., Li, Q., Pan, J. & Zou, G. (2022), 'Multiscale aggregation network via smooth inverse map for crowd counting', *Multimedia Tools and Applications*.

Kalpathy-Cramer, J., Napel, S., Goldgof, D. & Zhao, B. (2015), 'Qin multi-site collection of lung ct data with nodule segmentations', *Cancer Imaging Arch* **10**, K9.

Kim, J., Lee, J. K. & Lee, K. M. (2016), Accurate image super-resolution using very deep convolutional networks, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 1646–1654.

Kingma, D. P. & Ba, J. (2014), 'Adam: A method for stochastic optimization', *arXiv preprint arXiv:1412.6980*.

Kumar, G. & Singh, K. (2013), 'Image super resolution on the basis of dwt and bicubic interpolation', *International Journal of Computer Applications*.

Li, K., Yang, S., Dong, R., Wang, X. & Huang, J. (2020), 'Survey of single image super-resolution reconstruction', *IET Image Processing* **14**(11), 2273–2290.

Li, Q., Li, Z., Lu, L., Jeon, G., Liu, K. & Yang, X. (2019*a*), 'Gated multiple feedback network for image super-resolution', *arXiv preprint arXiv:1907.04253*.

Li, X., Gao, M., Shang, J., Pan, J. & Li, Q. (2022), 'A complexity reduction based retinex model for low luminance retinal fundus image enhancement', *Network Modeling Analysis in Health Informatics and Bioinformatics* **11**(1), 30.

Li, Z., Yang, J., Liu, Z., Yang, X., Jeon, G. & Wu, W. (2019*b*), Feedback network for image super-resolution, *in* 'Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition', pp. 3867–3876.

Liu, J., Gan, Z. & Zhu, X. (2013), Directional bicubic interpolation—a new method of image super-resolution, *in* '3rd International Conference on Multimedia Technology (ICMT-13)', Atlantis Press, pp. 463–470.

Liu, J.-J., Hou, Q., Cheng, M.-M., Wang, C. & Feng, J. (2020), Improving convolutional networks with self-calibrated convolutions, *in* 'Proceedings of the IEEE/CVF conference on computer vision and pattern recognition', pp. 10096–10105.

McCollough, C., Chen, B., Holmes, D., Duan, X., Yu, Z., Xu, L., Leng, S. & Fletcher, J. (2020), 'Low dose ct image and projection data [data set]', *The Cancer Imaging Archive*.

Ruangsang, W. & Aramvith, S. (2017), Efficient super-resolution algorithm using overlapping bicubic interpolation, *in* '2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)', IEEE, pp. 1–2.

Shen, H., Ng, M. K., Li, P. & Zhang, L. (2009), 'Super-resolution reconstruction algorithm to modis remote sensing images', *The Computer Journal* **52**(1), 90–100.

Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. (2004), 'Image quality assessment: from error visibility to structural similarity', *IEEE transactions on image processing* **13**(4), 600–612.

Zhai, W., Gao, M., Anisetti, M., Li, Q., Jeon, S. & Pan, J. (2022*a*), 'Group-split attention network for crowd counting', *Journal of Electronic Imaging*.

Zhai, W., Gao, M., Souri, A., Li, Q., Guo, X., Shang, J. & Zou, G. (2022*b*), 'An attentive hierarchy convnet for crowd counting in smart city', *Cluster Computing*.

Zhai, W., Li, Q., Zhou, Y., Li, X., Pan, J., Zou, G. & Gao, M. (2022*c*), 'Da2net: a dual attention-aware network for robust crowd counting', *Multimedia Systems*.

Zhang, H., Zhang, L. & Shen, H. (2012), 'A super-resolution reconstruction algorithm for hyperspectral images', *Signal Processing* **92**(9), 2082–2096.

Zhang, L., Zhang, H., Shen, H. & Li, P. (2010), 'A super-resolution reconstruction algorithm for surveillance images', *Signal Processing* **90**(3), 848–859.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B. & Fu, Y. (2018*a*), Image super-resolution using very deep residual channel attention networks, *in* 'Proceedings of the European conference on computer vision (ECCV)', pp. 286–301.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B. & Fu, Y. (2018*b*), Residual dense network for image super-resolution, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 2472–2481.

Zhao, H., Kong, X., He, J., Qiao, Y. & Dong, C. (2020), Efficient image super-resolution using pixel attention, *in* 'European Conference on Computer Vision', Springer, pp. 56–72.