Multi-window back-projection residual networks for reconstructing COVID-19 CT super-resolution images

Defu Qiu [a,b], Yuhu Cheng [a,b], Xuesong Wang [a,b,*], Xiaoqiang Zhang [a,b]

[a] Engineering Research Center of Intelligent Control for Underground Space, Ministry of Education, China University of Mining and Technology, Xuzhou, 221116, China
[b] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

ARTICLE INFO

ABSTRACT

Background and objective: With the increasing problem of coronavirus disease 2019 (COVID-19) in the world, improving the image resolution of COVID-19 computed tomography (CT) becomes a very important task. At present, single-image super-resolution (SISR) models based on convolutional neural networks (CNN) generally have problems such as the loss of high-frequency information and the large size of the model due to the deep network structure.

Methods: In this work, we propose an optimization model based on multi-window back-projection residual network (MWSR), which outperforms most of the state-of-the-art methods. Firstly, we use multi-window to refine the same feature map at the same time to obtain richer high/low frequency information, and fuse and filter out the features needed by the deep network. Then, we develop a back-projection network based on the dilated convolution, using up-projection and down-projection modules to extract image features. Finally, we merge several repeated and continuous residual modules with global features, merge the information flow through the network, and input them to the reconstruction module.

Results: The proposed method shows the superiority over the state-of-the-art methods on the benchmark dataset, and generates clear COVID-19 CT super-resolution images.

Conclusion: Both subjective visual effects and objective evaluation indicators are improved, and the model specifications are optimized. Therefore, the MWSR method can improve the clarity of CT images of COVID-19 and effectively assist the diagnosis and quantitative assessment of COVID-19.

© 2021 Published by Elsevier B.V.

1. Introduction

After the coronavirus invaded the lungs, it would diffuse along alveolar pores, which would lead to alveolar swelling, exudation of alveolar septum fluids, and thickening of alveolar septum, etc. All these will increase the CT number of the lungs, namely the lungs will become white. There is no exudation of granulocytes in viral infections, the alveoli are clean, and the air is still inside, so there is often a ground glass shadow without a substantial white mass change. Therefore, super-resolution (SR) reconstruction technology is urgently needed to improve the resolution of COVID-CT as an important basis for the diagnosis of COVID-19 [1].

Single-image super-resolution (SISR) reconstruction is an important image processing technology in the field of computer vision, which is widely used in medicine, video, security, and remote sensing. In the actual scene, due to the limitations of the existing medical hardware conditions, low-quality low-resolution (LR) medical images are often obtained. For example, the medical images of the current hospital CT detector as the detection and diagnosis of diseases often lack the key detection details or parts of the scene. Therefore, it is necessary to overcome the resolution limitations of the existing hardware system and use SISR reconstruction technology for enhancement the spatial resolution of the image. The core idea of this technology is to reconstruct the super-resolution image with high pixel density by analyzing the key semantic information or signal information of LR image, inferring the lack of real details.

At present, the research of SISR reconstruction is mainly divided into three stages. The earliest and most intuitive method is interpolation method based on sampling theory [2,3]. The advantage of this method is that it runs fast and is suitable for parallel computing, but it can't introduce additional useful high-frequency information, so it is difficult to get sharp high-definition image.

* Corresponding author.
E-mail addresses: defu.qiu@cumt.edu.cn (D. Qiu), wangxuesongcumt@163.com (X. Wang).
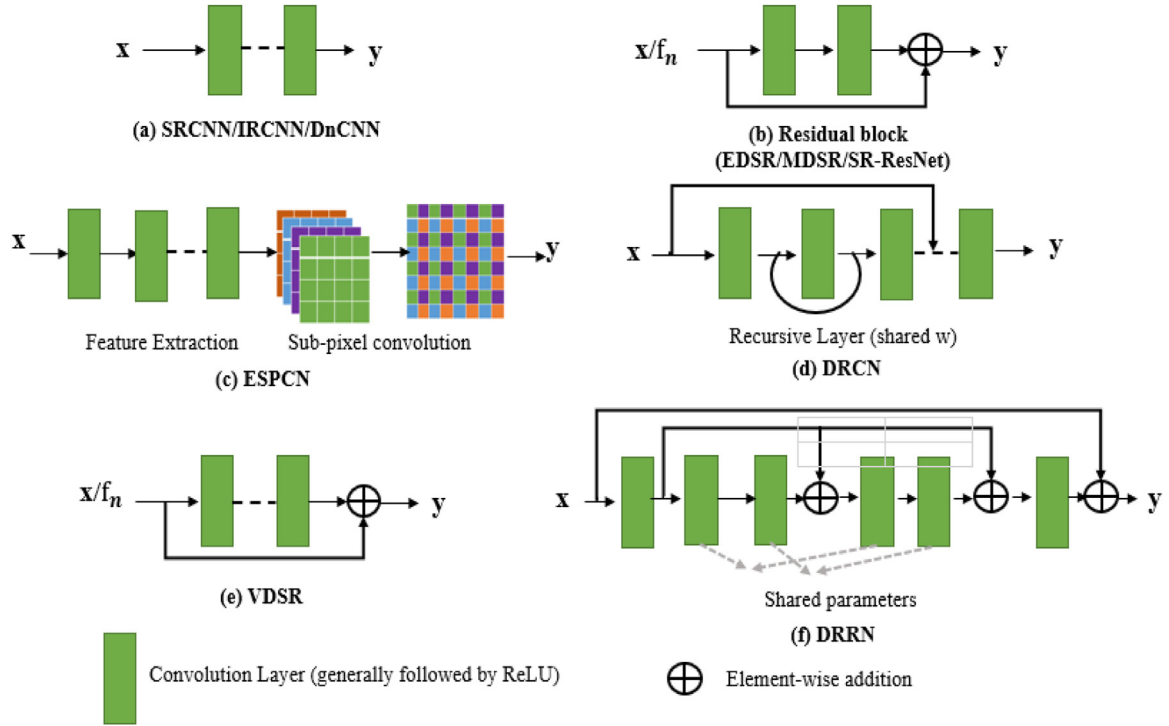
Fig. 1. Single image super-resolution network architecture of deep network.

After that, some scholars proposed that the information of the corresponding high-resolution (HR) part can be inferred from the LR image. Relying on the technique of neighborhood embedding [4,5], sparse coding [6–8], the algorithm of learning the mapping function between LR image and HR image is studied and proposed. However, when the image does not contain enough repetitive patterns, this method tends to produce non-detailed sharp edges.

The main contributions of this paper are described as follows:

(1) Expand the network structure horizontally to avoid the vertical depth of the network. The extended network uses the multi-window up-projection and down-projection residual module (MWUD) to extract the key information of the same feature map at the same time from the shallow network, to obtain more complete high/low frequency information in the original image as soon as possible.

(2) The residual network extracts features. The dilated convolution is used to expand the receptive field, and the image high/low frequency information is extracted layer by layer through 3 repetitive and continuous residual modules.

2. Deep neural network for super-resolution reconstruction

In recent years, methods based on deep learning have become the most active research direction in the field of SR. Since the SR-CNN [9] model proposed by Dong C et al. successfully used the convolutional neural network technology to reconstruct and generate higher-definition images, such methods have come to the fore. It uses many external HR images to construct a learning library, and generates a neural network model after training. In the process of LR image reconstruction, the prior knowledge obtained by the model is introduced to obtain the high-frequency detail information of the image, to achieve the excellent image reconstruction effect.

After that, FSRCNN [10], ESPCN [11] and other models have made some improvements on each part of the network structure based on SRCNN, and increased the number of network layers, focusing on learning the end-to-end mapping relationship from LR

images to HR images, as shown in Fig. 1. However, with the deepening of the network layer, the training cost increases gradually. At the same time, due to the increase of channel number, filter size, step size and other super parameters, it is very difficult to design a reasonable network structure. Then, He et al. [12] proposed the ResNet to solve the above problems. Although it is suitable for image classification, its residual idea and strategy of repeated stacking modules can be applied to all computer vision tasks. In addition, the ResNet also proved that shortcut connection and recursive convolution can effectively reduce the burden of neural network carrying a lot of key information.

Subsequently, residual network based super-resolution models, such as DRCN [13], DRNN [14], LapSRN [15], SRResNet [16], and EDSR [17] have been proposed, as shown in Fig. 1. These models are implemented by linear superposition of single size convolution modules to achieve the vertical deepening of the network, to pursue higher expression and abstract ability. However, for the super-resolution technology, it is very important to extract rich and complete feature information from the original image. If the network is deepened vertically, the high frequency information will be lost in the process of layer-by-layer convolution and filtering calculation, which will affect the authenticity of the final mapping generated super-resolution image. In addition, the number of model parameters will also increase exponentially. If the training dataset is limited, it is easy to produce over fitting; the model specifications will increase, and it is not easy to reconstruct and transplant; the amount of calculation will also increase, resulting in the training difficulty multiplied and difficult to apply.

3. Multi-window back-projection residual networks

To solve the problems of incomplete extraction of feature information in the original input image and large model scale due to the deep longitudinal structure of the network, we propose a multi-window back-projection residual networks for reconstructing COVID-19 CT super-resolution images. The model mainly includes multi-window up-projection and down-projection residual
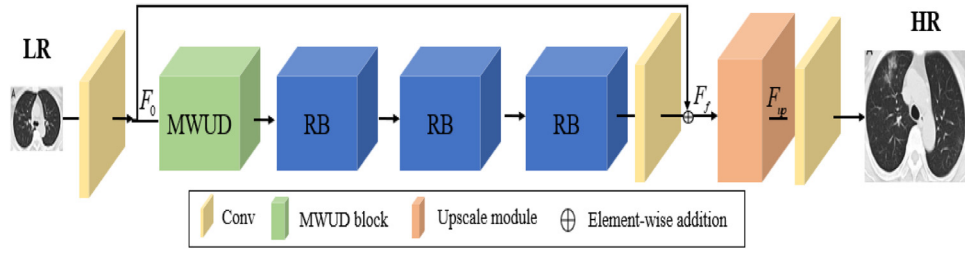
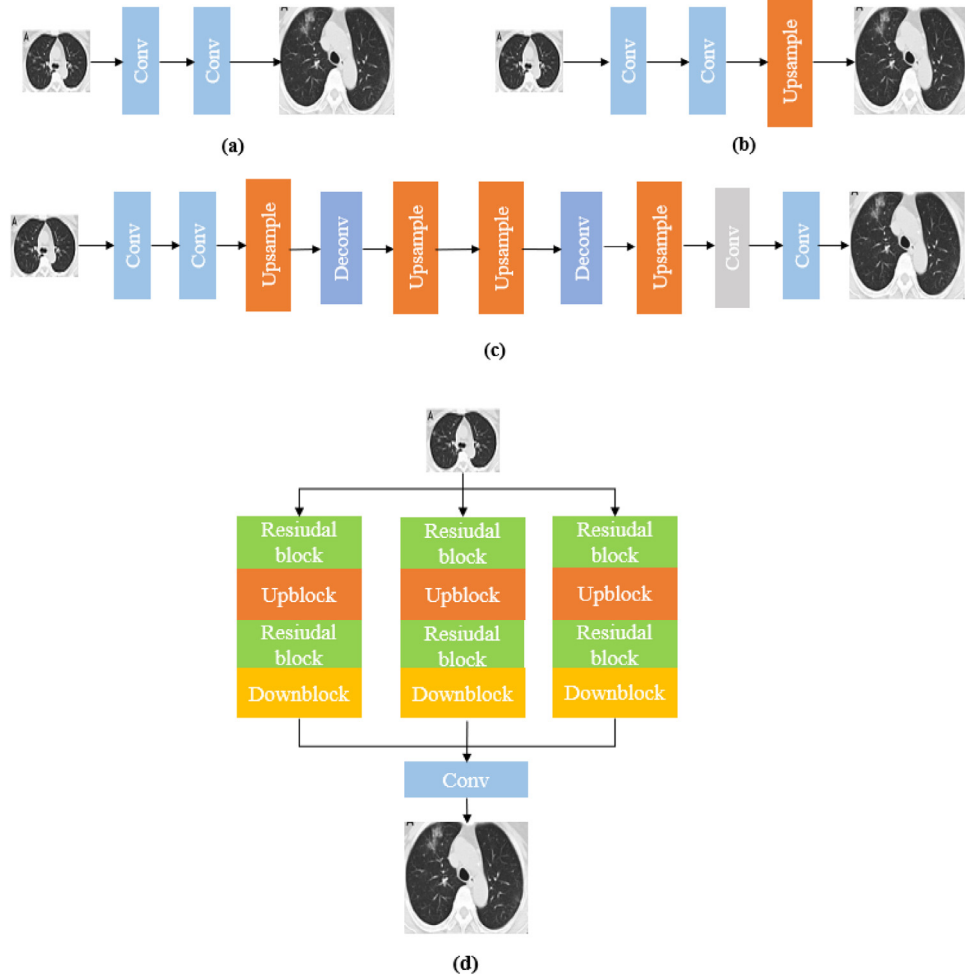**Fig. 2.** The architect of multi-window back-projection networks.

Fig. 3. Comparisons of deep network super-resolution. (a) Predefined upsampling (e.g., VDSR [18], DRRN [18], SRCNN [9]). (b) Single upsampling (e.g., ESPCN [11], FSRCNN [10]). (c) Progressive upsampling uses a Laplacian pyramid network. (d) multi-window up-projection and down-projection rsidual module is proposed by our MWSR which uses the interconnected up- and down sampling stages to obtain many HR features of different depths.

module (MWUD) and 3 Residual block modules (RB), MWUD is up-projection and down-projection residual module, RB is residual block module. as shown in Fig. 2.

3.1. Multi-window up-projection and down-projection residual module

In the deep network, it is divided into three types. Fig. 3 is (a) Predefined upsampling, (b) Single upsampling, (c) Progressive upsampling. In this article, we proposed the multi-window up-projection and down-projection residual module, as shown in Fig. 3(d).

The main purpose of multi-window up-projection and down-projection residual module (MWUD) is to explore the interdependence between LR and HR as an efficient iterative process. In addition, the MWUD module can provide more information for each

bottom-up or top-down mapping and increase the flow of information. When the sampling multiple is large, the corresponding size of convolution kernel or deconvolution kernel is large, which makes the convergence speed of network slow and easy to fall into suboptimal results [20]. In each MWUD, the use of skip connections to fuse intermediate-scale features and the use of 1×1 convolution kernel to reduce the dimensionality of the fused high-dimensional features improve the feature utilization rate reduces the complexity of the network, which can obviously optimize the network structure.

As shown in Fig. 3 (d), multi-window back-projection network consists of three up-projection and down-projection residual modules, each MWUD module consists of an up-projection (upblock), a down-projection (downblock), and two residual block modules. Consequently, in the deep network, in order to increase the re-

Table 1
The network architect setting of MWSR.

Network module			Kernel size	Stride	Padding	Dialted rate	Input size	Output size
Initial layer			3*3	1	1	1	H*W*1	H*W*64
Residual block	Conv 1		3*3	1	1	1	H*W*64	H*W*64
	Conv 2		1*1	1	0	1	H*W*64	H*W*4
	Conv 3		1*1	1	0	1	H*W*4	H*W*64
Upblock	1	DeConv	4*4	2	1	1	H*W*64	H*W*128
		Conv	4*4	2	1	1	H*W*128	H*W*64
		DeConv	4*4	2	1	1	H*W*64	H*W*128
	2	DeConv	4*4	2	4	3	H*W*64	H*W*128
		Conv	4*4	2	4	3	H*W*128	H*W*64
		DeConv	4*4	2	4	3	H*W*64	H*W*128
	3	DeConv	4*4	2	7	5	H*W*64	H*W*128
		Conv	4*4	2	7	5	H*W*128	H*W*64
		DeConv	4*4	2	7	5	H*W*64	H*W*128
Downblock	1	Conv	4*4	2	1	1	H*W*128	H*W*64
		DeConv	4*4	2	1	1	H*W*64	H*W*128
		Conv	4*4	2	1	1	H*W*128	H*W*64
	2	Conv	4*4	2	4	3	H*W*128	H*W*64
		DeConv	4*4	2	4	3	H*W*64	H*W*128
		Conv	4*4	2	4	3	H*W*128	H*W*64
	3	Conv	4*4	2	7	5	H*W*128	H*W*64
		DeConv	4*4	2	7	5	H*W*64	H*W*128
		Conv	4*4	2	7	5	H*W*128	H*W*64
RB module	Conv 7		3*3	1	1	1	H*W*64	H*W*64
	Conv 8		1*1	1	0	1	H*W*64	H*W*4
	Conv 9		1*1	1	0	0	H*W*4	H*W*64
Middle layer			3*3	1	1	1	H*W*64	H*W*64
Upscale			3*3	1	1	1	H*W*64	2H*2W*64
Reconstruction layer			3*3	1	1	1	2H*2W*64	2H*2W*1

ceptive field and reduce the amount of calculation, it is necessary to carry out subsampled (pooling or S2 / conv). Although the receptive field can be increased, the spatial resolution is reduced. In this paper, we introduce dilated convolution into the up projection and down-projection of MWUD module without losing the resolution and expanding the receptive field. Furthermore, 128 feature fusing layer and 64 Feature Fusing Layers (FFL) are introduced in up-projection model and down-projection model to reduce the MWSR network parameters, the parameters are shown in Table 1. In the MWUD module, different division rates lead to different receptive fields, which can capture multi-scale context information of COVID-19 CT images.

The up-projection model maps each other step by step between the LR feature map and the HR feature map, and the down-projection model maps each other step by step between the HR feature map and the LR feature map, as shown in Fig. 4. Accordingly, back-projection extracts image features by up-projection and down-projection. It can be understood that the network is a process of continuous self-correction. Its purpose is to avoid the single-step nonlinear mapping error when only up-sampling at the end of the network [20], and to improve the super-resolution performance.

Furthermore, the input of up-projection model is a LR feature map, which is mapped three times between the LR image and the HR image. The three mappings are: the first mapping maps the LR feature map to the HR feature map, the second mapping maps from the HR feature map to the LR feature map, and the third mapping maps the LR feature map to the HR feature map for output. Meanwhile, the down-projection model is very similar to the up-projection model, which is the inverse process of the up-projection model. Thus, the input of the down-projection model is a HR feature map, and after three mappings, the final output is a LR feature map.

The up-projection and down-projection residual module are defined as follows: up-projection scale up:

$$H_0^n = \left(L^{n-1} * p_n\right)\uparrow^s \tag{1}$$

up-projection scale down:

$$L_0^n = (H_0^n * g_n)\downarrow_s, \tag{2}$$

up-projection residual:

$$e_n^1 = L_o^n - L^{n-1}, \tag{3}$$

up-projection scale residual up:

$$H_1^n = \left(e_n^1 * q_n\right)\uparrow^s, \tag{4}$$

output feature map:

$$H^n = H_0^n + H_1^n, \tag{5}$$

down-projection scale down:

$$L_1^n = (H_2^n * g_n)\downarrow_s, \tag{6}$$

down-projection scale up:

$$H_3^n = (L_1^n * p_n)\uparrow^s, \tag{7}$$

down-projection residual:

$$e_h^n = H_3^n - (H_2^n * k_n), \tag{8}$$

down-projection scale up:

$$H_3^n = (L_1^n * p_n)\uparrow^s, \tag{9}$$

down-projection scale residual up:

$$L_2^n = \left(e_h^n * g_n\right)\downarrow_s, \tag{10}$$

output feature map:

$$L^n = L_0^n + L_1^n, \tag{11}$$

where $*$ is the convolution operation. $\uparrow^s$ and $\downarrow_s$ are the upsampling and subsampled operation with scale factor S, respectively. p_n is the upsampling deconvolutional layer of the $n - th$ UD. g_n is the subsampled convolutional layer of the $n - th$ UD. q_n is 128-dimensional feature fusion layer of the $n - th$ UD. k_n is 64-dimensional feature fusion layer of the $n - th$ UD [21].
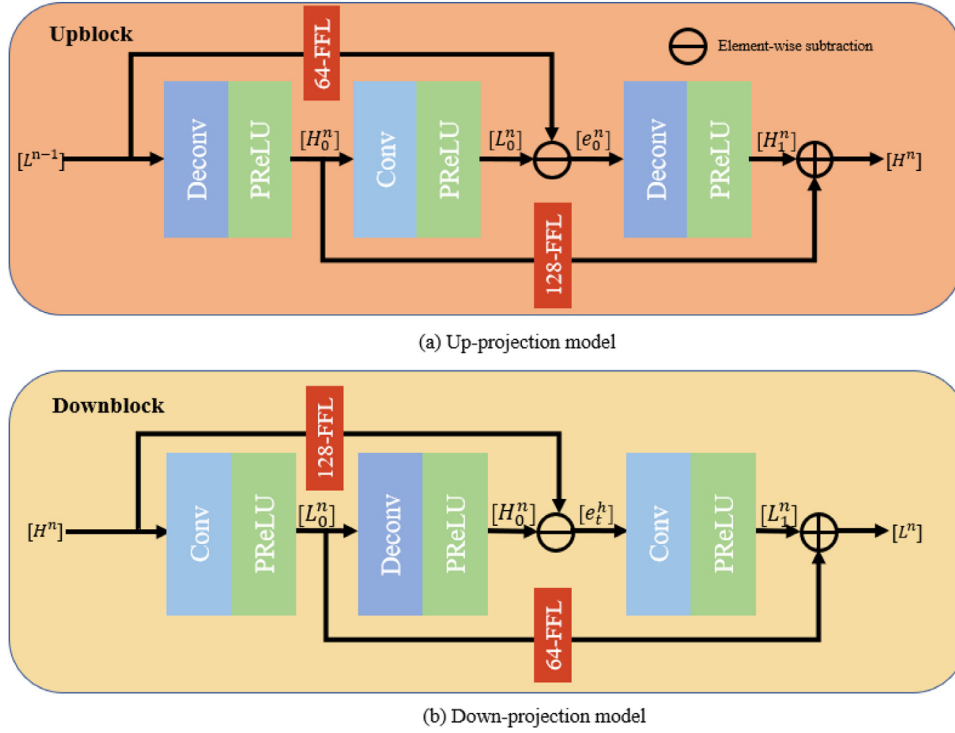
(a) Up-projection model

(b) Down-projection model

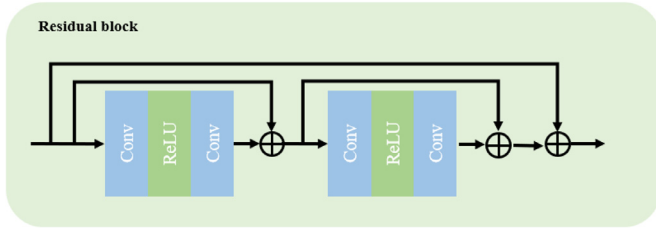**Fig. 4.** Structure of up-projection model and down-projection model.

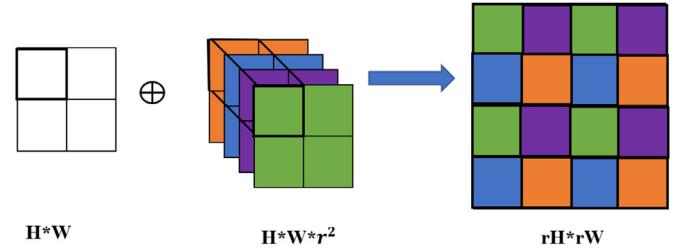**Fig. 5.** Structure of residual block model.

Fig. 6. Upsampling process on the subpixel convolution layer.

3.2. Residual block module

The MWSR model extracts the high and low frequency information of the image layer by layer through three repeated and continuous residual blocks, and fuses the initial feature map and the output of the above three residual blocks to merge the information flow through the network into the reconstruction module [22,23]. Thus, we construct three RB modules to extract deep features, as shown in Fig. 5. The operation of RB can be described as:

$$T = \left[Q_3^1, M_1, M_2, M_3 \right], \tag{12}$$

where [ ] is the connection operation between features, Q_3^1 is the initial feature graph, M_3 is the output of the third residual module, and T is the output of global feature fusion.

3.3. Sub-pixel convolutional upsampling layer

In the actual camera imaging process, due to hardware limitations, each pixel in the generated image represents a whole block of color nearby. In fact, at the microscopic level, there are many pixels, namely sub-pixels, between actual physical pixels. In the field of SR image reconstruction, the sub-pixels that cannot be detected by the sensor can be approximated by algorithms [11], which is equivalent to inferring high-frequency information such as missing texture details in the image.

The sub-pixel convolution is used to complete the mapping of LR images to HR images in the high-multiple reconstruction part of the MWSR model, as shown in Fig. 6. Assuming that the target multiple is γ and the input LR feature map size is $H*W$, we convolve it with the r^2 sub-pixel convolution kernel with channel number $H*W$ to obtain $H*W*r^2$ pixel values, and then rearrange them to form a target image with size $rH*rW$.

3.4. Network architect

The proposed the MWSR method is illustrated in Fig. 2. It can be divided into three parts: multi-window back projection, depth feature extraction and sub-pixel convolution layer. Further, set I_{LR} and I_{LR} denote the input image and the reconstructed image of MWSR, and the input image I_{LR} is extracted by an initial layer to obtain the initial feature map $L_{initial}$, $C_{initial}$ is the initial layer, and the operation of the initial layer can be defined as:

$$L_{initial} = C_{initial} * I_{LR}, \tag{13}$$

In the multi-window up-projection and down-projection residual modules, MWSR performs the back-projection for the initial feature map $L_{initial}$, and the operation of back-projection is described as:
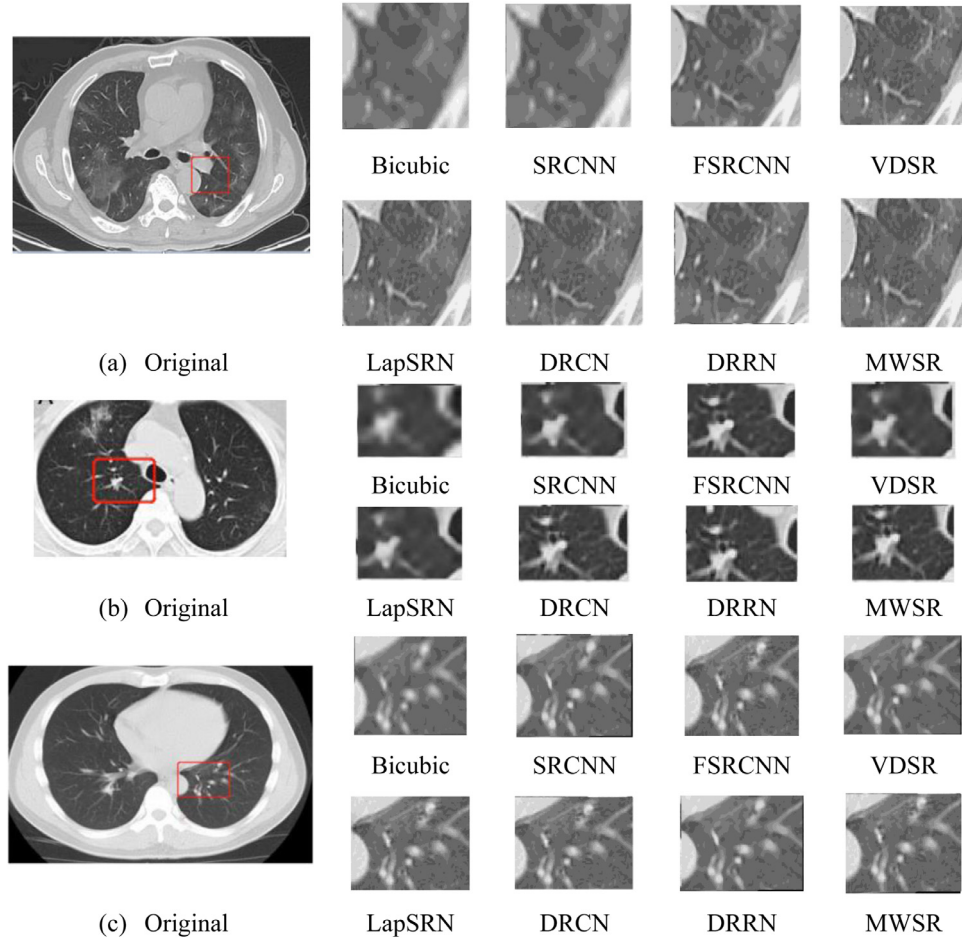
$$L_{bp} = f_{bp}(L_{initial}), \tag{14}$$

Fig. 7. Visual quality of the proposed MWSR and other state-of-art methods for 4× scale factor on COVID-19 CT images.

where $f_{bp}(\bullet)$ denotes the operation of back-projection, I_{bp} denotes the extracted shallow feature maps by back-projection.

Then, MWSR uses three RB modules to extract deep feature maps from the extracted shallow feature maps, and the operation of deep feature extraction is

$$L_{deep} = f_{deep}(L_{bp}), \tag{15}$$

where $f_{deep}(\bullet)$ denotes the deep feature extraction operation, L_{deep} denotes the extracted feature maps by deep feature extraction.

Finally, MWSR uses the sub-pixel convolutional layer to upsample, and the operation of upscaling can be formulated as:

$$L_{up} = f_{up}(L_{initial} + C_{middle} * L_{deep}), \tag{16}$$

where $f_{up}(\bullet)$ denotes the upscaling operation, C_{middle} denotes the middle layer which is a convolutional layer, L_{up} denotes the upscaling feature maps.

4. Experiments and analysis

4.1. Dataset and protocol

The public datasets BSD500 and T91 [24] were used in the experiment, and the two training sets have a total of 591 images. Due to the depth model usually benefits from large amounts of data, 591 images are not sufficient to push the model to its best performance [29]. Thus, to make full use of the dataset, we use the MATLAB to expand the data of the BSD500 and T91 training set images by two methods, namely scaling and rotation. Each image was scaled by the ratios of 0.7, 0.8, and 0.9. Additionally, each im-

age was respectively rotated by 90°, 180°, and 270°, and 9456 images were finally obtained. In addition, a lot of tests and comparisons have been made on the public benchmark datasets Set5 [25], Set14 [26], and Urban100 [27].

4.2. Training details and implementation

In this paper, a 48×48 RGB image cut from I_{LR} is used as input, and the quality of the generated SR image is evaluated by I_{HR} of the target magnification. In the proposed MWSR, the parameters are initialized to a gaussian distribution with a mean of 0 and a standard deviation of 0.001, and the initialization of the bias is set to 0. The network trained a total of 200 epochs. To speed up the training process, an adjustable learning rate strategy was used to train the network. The initial learning rate was set to 0.1, and the learning rate decreased to 0.1 times the original learning rate every 10 epochs [30]. When the learning rate decreased to 0.0001, it is then kept at 0.0001 and the batch size is set to 128. In addition, use Adam optimizer [25] to set $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$ respectively. In this paper, L1 norm is selected as the loss function training model [28]. Compared with L2, it has sparsity, which can realize automatic feature selection, and has fewer parameters, which can better explain the model. L1 loss function to optimize MWSR as follow:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} ||f_{GBPN}(I_{LR}^i) - I_{HR}^i||_1, \tag{17}$$

where θ denotes the parameters of MWSR.

Table 2

Quantitative evaluation results of state-of-the-art SR methods: average PSNR and SSIM for scale factors ($\times 2$, $\times 3$, $\times 4$). Red numbers indicate the best and green numbers indicate the second-best performance.

Algorithm	Scale	Set 5		Set 14		Urban 100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic [3]	2×	33.69	0.931	30.25	0.870	26.88	0.841
A+ [7]	2×	36.60	0.955	32.32	0.906	29.25	0.895
SCN [28]	2×	36.58	0.954	32.35	0.905	29.52	0.897
SRCNN [9]	2×	36.72	0.955	32.51	0.908	29.53	0.896
FSRCNN [10]	2×	37.05	0.956	32.66	0.909	29.88	0.902
VDSR [18]	2×	37.53	0.959	33.05	0.913	30.77	0.914
DRCN [13]	2×	37.63	0.959	33.06	0.912	30.76	0.914
LapSRN [15]	2×	37.52	0.959	33.08	0.913	30.41	0.910
DRRN [19]	2×	37.74	0.959	33.23	0.914	31.23	0.919
MWSR (ours)	2×	37.86	0.960	33.53	0.919	31.25	0.919
Bicubic [3]	3×	30.41	0.869	27.79	0.775	24.46	0.735
A+ [7]	3×	32.62	0.909	29.15	0.820	26.05	0.799
SCN [28]	3×	32.62	0.908	29.16	0.818	26.21	0.801
SRCNN [9]	3×	32.78	0.909	29.32	0.823	26.25	0.801
FSRCNN [10]	3×	33.18	0.914	29.37	0.824	26.43	0.808
VDSR [18]	3×	33.67	0.921	29.78	0.832	27.14	0.829
DRCN [13]	3×	33.83	0.922	29.77	0.932	27.15	0.828
LapSRN [15]	3×	33.82	0.922	29.87	0.832	27.07	0.828
DRRN [19]	3×	34.03	0.924	29.96	0.835	27.53	0.764
MWSR (ours)	3×	33.96	0.924	30.25	0.843	27.24	0.832
Bicubic [3]	4×	28.43	0.811	26.22	0.715	23.14	0.658
A+ [7]	4×	30.32	0.860	27.34	0.751	24.34	0.721
SCN [28]	4×	30.41	0.863	27.39	0.751	24.52	0.726
SRCNN [9]	4×	30.50	0.863	27.52	0.753	24.53	0.725
FSRCNN [10]	4×	30.72	0.866	27.61	0.755	24.62	0.728
VDSR [18]	4×	31.35	0.883	28.02	0.768	25.18	0.754
DRCN [13]	4×	31.54	0.884	28.03	0.768	25.14	0.752
LapSRN [15]	4×	31.54	0.885	28.19	0.772	25.21	0.756
DRRN [19]	4×	31.68	0.888	28.21	0.772	25.44	0.764
MWSR (ours)	4×	31.64	0.887	28.46	0.784	25.23	0.757

In addition, combined with CUDA 10.0 and PyTorch 1.20, we use Python code to implement MWSR algorithm, and train and evaluate the algorithm through many experiments on NVIDIA GeForce RTX 1080ti GPU and Ubuntu16.04 operating systems.

4.3. Comparison with state-of-the-art methods

We compare our model with the 9 state-of-the-art SR methods, including Bicubic [3], A+ [7], SCN [28], SRCNN [9], FSRCNN [10], VDSR [18], DRCN [13], LapSRN [15], and DRRN [19]. The specific implementation of these models has been officially published on the Internet; thus, these algorithms can be executed on the same test dataset for fair comparison. In addition, the quality of the generated super-resolution image is evaluated by two common objective evaluation indexes: peak signal to noise ratio (PSNR) and structural similarity (SSIM) [31].

We show the quantitative results in the Table 2, it shows the evaluation results of 10 kinds of super-resolution algorithms, which are magnified by 2×, 3× and 4× respectively on three public test datasets. As can see that the performance of the MWSR method outperforms other state-of-the art methods in different multiples and different test datasets, further, the MWSR method successfully reconstruct the detailed textures and improves the image perception quality, and realizes the model lightweight and oper-

ation efficiency optimization [32]. At the best performance of 2× enlargement on Set14 dataset, the MWSR method achieve 33.53 dB which better 3.28 dB, 1.21 dB, 1.18 dB, 1.02 dB, 0.87 dB, 0.48 dB, 0.47 dB, 0.45 dB, 0.30 dB than Bicubic, A+, SCN, SRCNN, FS-RCNN, VDSR, DRCN, LapSRN, and DRRN respectively. Then, to further verify the performance of the MWSR, COVID-19 CT images are used to evaluate the visual quality of MWSR and other state-of-art methods (COVID-CT: https:, github.com/UCSD-AI4H/COVID-CT). The COVID-CT-Dataset has 349 CT images containing clinical findings of COVID-19 from 216 patients, and the images are collected from COVID19-related papers from medRxiv, bioRxiv, NEJM, JAMA, Lancet, etc. CTs containing COVID-19 abnormalities are selected by reading the figure captions in the papers. All copyrights of the data belong to the authors and publishers of these papers.

The results of 4× enlargement on the COVID-19 CT images dataset is visually shown in Fig. 7. By enlarging the reconstructed results in some regions, it can be found that the COVID-19 CT images reconstructed by Bicubic and SRCNN algorithms can hardly observe the contour and other details. While the reconstruction results of FSRCNN, VDSR and LaPSRN algorithms lack the detail information of COVID-19 CT; the reconstruction results of DRCN and DRRN algorithm obtain rich detail information, but lack clear edge information; on the contrary, the COVID-19 CT images reconstructed by the algorithm proposed in this paper can complete the high-frequency information more accurately and completely. Whether it is the detail information or the edge, it can predict the more real new pixel value after magnification according to the overall semantic of the image.

5. Conclusion

In order to solve the problems of incomplete extraction of feature information and large scale of COVID-19 CT original input image due to the deep vertical structure of the network, we propose a super-resolution model based on multi-window back-projection residual networks (MWSR). The model combines three windows to extract the key information of the same feature map at the same time, which can effectively use the feature map of each layer from the shallow network, and improve the probability of detecting high-frequency information. More importantly, compared with the vertical deepening network structure, this horizontal expansion network structure can obtain the complete COVID-19 CT images target features earlier. The experimental results show that MWSR can reconstruct COVID-19 texture features more effectively than other popular models. For future implementation, we will focus on optimizing the up-sampling operation of the high-resolution reconstruction part, and calculate the more realistic and effective mapping relationship between the low-resolution feature space and the high-resolution feature space.

Ethical approval

No ethics approval was required.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

Acknowledgment

References

[1] PJ Kahrilas, KW Altman, AB Chang, et al., The use of bronchoscopy during the COVID-19 pandemic: CHEST/AABIP guideline and expert panel report, Chest (2020) S001236921658893X.

[2] L Zhang, X. Wu, An edge-guided image interpolation algorithm via directional filtering and data fusion, IEEE Trans. Image Process. 15 (8) (2006) 2226–2238.

[3] D Qiu, L Zheng, J Zhu, G Han, Y Peng, et al., Multiple improved residual networks for medical image super-resolution, Future Gen. Comput. Syst. (2021) 200–208.

[4] T Roweis S, K Saul L, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[5] X Gao, K Zhang, D Tao, et al., Image super-resolution with sparse neighbor embedding, IEEE Trans. Image Process. 21 (7) (2012) 3194–3205.

[6] J Yang, Z Wang, Z Lin, et al., Coupled dictionary training for image super-resolution, IEEE Trans. Image Process. 21 (8) (2012) 3467–3478.

[7] R Timofte, Vincent De Smet, Luc Van Gool, A+: adjusted anchored neighborhood regression for fast super-resolution, in: Asian Conference on Computer Vision, Springer, Cham, 2014, pp. 111–126.

[8] E Michael, D. Dmitry, Example-based regularization deployed to super-resolution reconstruction of a single image, Comput. J. (1) (2018) 15–30.

[9] C Dong, C C Loy, K He, et al., Learning a deep convolutional network for image super-resolution, in: European Conference on Computer Vision, Springer, Cham, 2014, pp. 184–199.

[10] C Dong, C C Loy, X Tang, et al., Accelerating the super-resolution convolutional neural network, in: European Conference on Computer Vision, Springer, Amsterdam, 2016, pp. 391–407.

[11] W Shi, J Caballero, F Huszár, et al., Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Las Vegas, 2016, pp. 1874–1883.

[12] K He, X Zhang, S Ren, et al., Deep residual learning for image recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Las Vegas, 2016, pp. 770–778.

[13] J Kim, J Kwon Lee, K Mu Lee, Deeply-recursive convolutional network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1637–1645.

[14] Y Tai, J Yang, X Liu, et al., Image super-resolution via deep recursive residual network, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Las Vegas, 2017, pp. 2790–2798.

[15] W.-S Lai, J.-B Huang, N Ahuja, et al., Deep laplacian pyramid networks for fast and accurate superresolution, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Las Vegas, 2017, pp. 5835–5843.

[16] C Ledig, Z Wang, W Shi, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Hawaii, 2017, pp. 105–114.

[17] B Lim, S Son, H Kim, et al., Enhanced deep residual networks for single image super-resolution, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society, Hawaii, 2017, pp. 136–144.

[18] J Kim, J Kwon Lee, K Mu Lee, Accurate image super resolution using very deep convolutional networks, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1646–1654.

[19] Y Tai, J Yang, X. Liu, Image super-resolution via deep recursive residual network, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2790–2798.

[20] R Timofte, E Agustsson, L V Gool, et al., NTIRE 2017 Challenge on single image super-resolution: methods and results, in: IEEE Conference on Computer Vision & Pattern Recognition Workshops, IEEE, 2017, pp. 1110–1121.

[21] M Haris, G Shakhnarovich, N. Ukita, Deep back-projection networks for super-resolution, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 1664–1673.

[22] Y Zhang, K Li, Li Kai, L Wang, B Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: ECCV (7), volume11211 of Lecture Notes in Computer Science, 2018, pp. 294–310.

[23] J Liu, W Zhang, Y Tang, et al., Residual feature aggregation network for image super-resolution, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020, pp. 2356–2365.

[24] J Kim, K Lee J, M Lee K, Accurate image super-resolution using very deep convolutional networks, in: Computer Vision and Pattern Recognition, IEEE, 2016, pp. 1646–1654.

[25] M Bevilacqua, A Roumy, C Guillemot, et al., Low-complexity single-image super-resolution based on nonnegative neighbor embedding, in: British Machine Vision Conference, BMVA Press, Guildford, 2012, pp. 1–10.

[26] R Zeyde, M Elad, M. Protter, On single image scale-up using sparse-representations, in: International Conference on Curves & Surfaces, Springer, Avignon, 2010, pp. 711–730.

[27] J B Huang, A Singh, N. Ahuja, Single image super-resolution from transformed self-exemplars, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Boston, 2015, pp. 5197–5206.

[28] Z Wang, D Liu, J Yang, et al., Deep networks for image super-resolution with sparse prior, in: IEEE International Conference on Computer Vision, 2015, pp. 1–15.

[29] K K L Wong, G Fortino, D. Abbott, Deep learning-based cardiovascular image diagnosis: a promising challenge, Future Gen. Comput. Syst. (2020) 802–811.

[30] D Qiu, S Zhang, Y Liu, et al., Super-resolution reconstruction of knee magnetic resonance imaging based on deep learning, Comput. Methods Programs Biomed. (2019) 105059.

[31] Z Wang, A.C Bovik, H.R Sheikh, et al., Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[32] K Zhang, W Zuo, L. Zhang, Learning a single convolutional super-resolution network for multiple degradations, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018, pp. 3262–3271.