Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Progressive U-Net residual network for computed tomography images super-resolution in the screening of COVID-19

Defu Qiu, Yuhu Cheng, and Xuesong Wang

*Engineering Research Center of Intelligent Control for Underground Space, Ministry of Education, China University of Mining and Technology, Xuzhou, China; †School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China

ABSTRACT

Thin-slice computed tomography (CT) examination plays an important role in the screening of suspected and confirmed coronavirus disease 2019 (COVID-19) outbreak patients. Therefore, improving the image resolution of COVID-19 CT has important clinical value for the diagnosis and condition assessment of COVID-19. However, the existing single-image super-resolution (SISR) methods mainly increase the receptive field of convolution kernels by deepening and widening the network structure, and adopt the equal processing methods in the airspace and channel domains with different importance, and a large number of computing resources will be wasted on the unimportant features. We propose a progressive U-Net residual network (PURN) for COVID-19 CT images super-resolution (SR) to solve the practicality of existing models, to better extract features, and reduce the number of parameters. First, we design a dual U-Net module (DUM), which can efficiently extract low-resolution (LR) COVID-19 CT images feature. Second, the DUM module first performs up-block three times, and then down-blocks three times in order to learn the interdependence between high-resolution (HR) and LR images more efficiently. Finally, the local skip connection structure is introduced in the DUM module, and the global long skip connection structure is introduced in the reconstruction layer to further enrich the flow of reconstructed HR image information. Experimental results show that our algorithm effectively improves the SR reconstruction effect of COVID-19 CT images, restores its detailed features more sharply, and greatly improves the practicability of the algorithm.

1. Introduction

From January to March 2020, a novel coronavirus pneumonia epidemic occurred in Wuhan, Hubei province, and other regions, seriously endangering human health and forming a large-scale transmission. As a result, there were over 80,000 confirmed cases, over 4,000 deaths, and over 1,000 critically ill patients in China. For imaging examination of coronavirus disease 2019 (COVID-19), multi-slice spiral computed tomography (CT) shows unique advantages and brings good news to COVID-19 patients. However, for atypical patients, patients with small lesions, and patients with a large weight base, it is often necessary to increase the tube voltage and tube current to ensure the image quality, which greatly increases the patient’s radiation dose (Ai et al., 2020; Algaissi et al., 2020; Alshammar et al., 2020; Ayinde et al., 2020). With the continuous advancement and optimization of technology, imaging technology has achieved leaps and bounds (Barry et al., 2020), and increasing researches have applied deep convolution neural network to the SR reconstruction of medical CT images.

In recent years, scholars have made many achievements in SR research on natural images, and developed a series of reconstruction algorithms. According to the basic idea of the algorithm, the current SR reconstruction technology can be divided into interpolation based (Keys, 1981) and reconstruction based (Ge et al., 2016) and the method based on learning (Qiu et al., 2020, 2021; Tai et al., 2017; Wang et al., 2015). The method based on interpolation generally estimates the connected pixels between adjacent pixels through basis functions to fill the unknown points in the HR image. Although this method is efficient and intuitive, it cannot guarantee the accuracy of estimated pixels, so the reconstructed image often has problems such as artifacts and blur. Furthermore, the reconstruction method is based on signal processing theory, taking LR images as constraints, and using the prior information on the images to estimate the HR images. However, due to the lack of prior information, the HR images reconstructed by this type of method often have problems such as missing details. In order to break the limitations of the first two methods, related scholars introduced the idea of learning into the SR field. The SR reconstruction algorithms based on...
instance learning (Lu et al., 2018), based on sparse coding (Yang et al., 2010), and fixed neighborhood regression (Chang et al., 2004) have been successively used, and it is proposed that the SR reconstruction quality of the image is improved by another level.

Dong, Loy, He et al. (2016) took the lead in the application research of convolutional neural network in the field of image SR, and proposed an image SR method based on convolutional neural network (SRCNN). The SRCNN uses three convolutional layers to perform features extraction, non-linear mapping and image reconstruction operations on the image, and obtains a reconstruction effect that is significantly better than traditional algorithms. In order to speed up the reconstruction speed, Dong, Loy, Tang et al. (2016) combined deconvolution with the original algorithm and proposed fast image super-resolution reconstruction based on deconvolution (FSRCNN). With the deepening of research, scholars have found that deep networks can achieve better SR effects, but the deepening of the number of layers often leads to problems such as network gradient dispersion or gradient explosion, making it impossible to converge. Kim et al. (2016a) borrowed the idea of residual network (ResNet) (He et al., 2016), used deep residual network to accelerate network learning, and proposed a super-resolution reconstruction algorithm (VDSR) based on deep convolutional neural network.

After that, Kim et al. (2016b) considered the problem of parameter scale and proposed the deep recursive convolutional network (DRCN), which uses a recursive network structure to share parameters between network structures, which effectively reduces the difficulty of training; in addition, the author also uses skip connection and integration strategies to further improve performance. Subsequently, Shi et al. (2016) proposed an efficient sub-pixel convolutional neural network (ESPCN), which uses LR images as input, and uses sub-pixel convolutional layers at the back end of the network structure to implicitly map LR images to HR images, effectively reducing computational complexity and improving reconstruction efficiency. Lai et al. (2017) proposed the Laplacian pyramid network (LapSRN), where the idea of Laplace pyramid is introduced into deep learning, and the experimental results prove the superiority of the step-by-step sampling operation. In addition, the residual results predicted at each level are monitored during the training process, which further improves the performance. Lim et al. (2017) proposed the enhanced deep residual network for single image super-resolution (EDSR) by removing the redundant modules from the literature (Dong, Loy, Tang et al., 2016) and using the L1 norm as the loss function. Y. Zhang et al., (2018) proposed the residual channel attention network (RCAN), by using the channel attention mechanism, a feature channel with rich information can be selected. The above network structures are mostly feed-forward structures, ignoring the interdependence of HR images and LR images and the error when upsampling LR images.

Although the above learning-based algorithms have achieved good SR effects on natural images, there are still some problems when applied to CT images. The single-scale network structure of VDSR cannot fully extract the main features of CT images, and there are only three layers of networks. The definition of CT images recovered by the SRCNN is insufficient, and the algorithm complexity is relatively high, which cannot meet the requirements of medical image processing speed. On the other hand, although the FSRCNN has been improved, the depth of its network is still not enough to extract the detailed information on medical images. At the same time, a network can only be applied to one magnification, which limits the application of the above two algorithms in the field of medical imaging.

In response to the above problems, we started from the characteristics of CT images and combined the idea of residual learning, and proposed the progressive U-Net residual network (PURN). By building a dual U-Net network module, we can more fully extract various features of LR CT images, introduce residual learning, and adopt the Adam optimization algorithm to speed up the convergence of the model. At the same time, we process the training set under multiple enlargements in the early stage of training so that the model can support multiple resolution enhancement at the same time.

The main contributions of this paper are as follows:

1. We propose a progressive U-Net residual network to realize image SR reconstruction, and through comparative experiments on the benchmark dataset, it is verified that the proposed algorithm has a significant improvement over the most advanced SISR methods.
2. We design a dual U-Net model, which can strengthen feature supervision from a shallow level, promote network convergence, and effectively improve the quality of image reconstruction.
3. We propose a local and global residual network structure, which effectively alleviates the problem of gradient disappearance, and the local residual block can transfer the original rich details of the image to the subsequent feature layer to extract richer details feature.

2. Related work

2.1. Residual learning

When training a very deep network structure, since the initialization parameters are very close to zero, it is easy to cause gradient dispersion when the network
reversely broadcasts the update parameters. This makes deepening the network structure not only unable to improve network performance, but even worse. In response to this problem, He et al. (2016) proposed the residual net (ResNet), using the idea of residual learning to alleviate the problem of gradient dispersion. The main idea is to add a direct connection channel to the network, allowing a certain percentage of the previous network output to be retained. However, there are certain difficulties in learning identity mapping, to avoid learning the parameters of identity mapping, the ResNet uses the network structure describe in Figure 1, namely \( H(x) = F(x) + x \). It can be converted to \( F(x) = H(x) - x \), where \( F(x) \) is the residual term. When the residual term is \( F(x) = 0 \), the identity mapping \( H(x) = x \) can be easily constructed. Compared to learning the identity mapping \( H(x) = x \), learning \( F(x) = 0 \) is easier.

Figure 1. Structural framework of residual learning.

Figure 2. Structural framework of residual learning.
2.2. Up-block and down-block module

The predefined upsampling network mainly uses the original LR image to be upsampled to the target’s HR space using the bicubic interpolation algorithm, and then it is input into the convolutional neural network for deep feature extraction. Representative networks include the SRCNN and VDSR, as shown in Figure 2(a). Although this type of method is relatively simple in operation, the image feature extraction part takes place in the HR space, which will greatly increase the training time of the overall network. At the same time, due to the bicubic algorithm being used to sample the image, the noise in the LR image can lead to inaccurate feature information, which has a great impact on the reconstruction results of the whole network. In order to solve the problems of the predefined upsampling network, the single upsampling network removes the upsampling module in the predefined structure of the network. All its mapping transformations are performed in the LR space. The pixel convolution layer or deconvolution layer reconstructs the image into an HR space, its representative network is an ESPCN, which greatly reduces the time complexity and space complexity of the overall network, but there are also some such networks. The problem is that the mapping relationship between their HR and LR images is only determined by the last layer so that a simple convolutional layer reflects the mapping relationship, which cannot fully explain the feature mapping information between the HR and LR images, as shown in Figure 2(b). In order to solve the problems of the predefined upsampling and single upsampling networks, we designed a feature extraction network based on U-Net ideas and residual ideas. This is manifest from Figure 2(c), where the network structure is not fixed in LR space or HR space. Instead, it is continuously transformed into HR and LR space, and it is sampled in low-resolution space by convolution layer, and then projected to high-resolution space by deconvolution layer to learn the nonlinear relationship between LR and HR.

2.3. U-Net

The U-Net network structure is the network structure proposed by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in the 2015 IEEE International Symposium on Biomedical Imaging (ISBI) competition (Ronneberger et al., 2015). The network structure is composed of a shrinking sub-network and an expanding sub-network, forming a U-shaped structure, so it is named U-Net. The Unet network is shown in Figure 1, the U-Net first extracts feature information by downsampling through convolution and pooling, and then up-samples through transposed convolution and crops the previous low-level feature maps for fusion for precise positioning. We then repeat this process until the output feature map is obtained, and finally the segmentation map is obtained through the activation function. Although U-Net has the advantages of fewer parameters and simple network structure, it is insufficient in depth compared to networks such as ResNet and VGG. Therefore, in this experiment, the traditional U-Net network model is improved, the residual network is combined with the U-Net network model, and we propose the progressive U-Net residual network.

3. Methodology

The reconstruction algorithm in this paper applies a deep convolution network, so that in the process of feature extraction, it can be directly operated on the LR image frequently. In addition, the increase in depth increases the accuracy of reconstruction. The PURN method is shown in Figure 3, which consists of six groups of up sampling modules and six groups of down sampling modules to form a dual U-Net network structure, and skip connection.

![Figure 3](image-url). Structural framework of PURN for COVID-19 CT images processing.
3.1. Shallow feature extraction block

The shallow feature extraction block consists of a convolutional layer and an activation layer, which is used to extract features from the input LR image. The $Y$ and $F_0$ are used to represent the input and output of the network, the feature extraction can be represented by Equation (1).

$$F_0 = \sigma(W_0 \ast Y + b_0),$$

where $\ast$ denotes the convolution operation, $W$ denotes the convolution kernel, $b$ denotes the bias, $F$ denotes the feature extracted from the LR $Y$. Before performing the convolution operation, in order to prevent the feature map after the convolution from getting smaller and smaller, the zero-padding operation is performed before each convolution, to ensure that the image after each convolution is still the original size.

3.2. Dual U-Net module

Compared with the U-Net network, we design a dual U-Net module (DUM) that can deepen the number of network layers and improve the quality of image reconstruction; and more skip connections can be added in the middle of the network, which can better combine the feature information of LR image and extract more detailed features, as shown in Figure 4. In addition, the ResNet has the advantage of fast convergence and can reduce the amount of model data, and the ResNet makes the model easier to train, which can prevent model degradation, but also prevent the gradient from disappearing, and the loss does not converge.

3.3. Up- and down-block module

In the up-block and down-block module network structure, we use an iterative up-and-down sampling network. In the feature extraction process, instead of using ordinary convolutional layers to extract image features like the above mainstream algorithms, we are constantly extracting features during upsampling and downsampling, as shown in Figure 5(a,b). In this process, we first use upsampling to map the image to the HR space, and add the projection error from LR to HR to the HR feature map; then use downsampling to map the image from the HR space to the LR space. In the resolution space, the projection error from HR to LR is added to the LR feature map. Through this alternate

![Figure 4. Structural framework of the dual U-Net module.](image1)

![Figure 5. Structural framework of the (a) up-block and (b) down-block.](image2)
use of upsampling and downsampling, the mapping relationship between image LR and HR is continuously adjusted, and their non-linear relationship is optimized, to achieve the purpose of improving the final LR to HR reconstruction effect.

In the entire feature extraction module, our PURN algorithm uses multiple convolutional layers and deconvolutional layers alternately to transform image information into HR and LR spaces. Furthermore, every time an image is projected to an HR space through up projection or to an LR space through down projection, it will bring some errors with the original image, and by continuously optimizing these error relations, the mapping relationship between HR and LR images can be more accurate. Based on this idea, our PURN algorithm will focus on learning the direct mapping relationship between HR and LR space. For more convenient description, the up-block method is called upsampling, and the down-block method is called downsampling.

The main function of upsampling is to learn the reconstruction error relationship when the image is projected from the LR space to the HR space (Mei et al., 2020). Then, a convolution operation is performed on the HR feature map to obtain the LR feature map. Subtracting the two LR feature maps can obtain the difference between the LR feature maps. In fact, these differences are important feature information lost in the process of feature extraction, and deconvolution of these differences feature images is added to the HR feature images obtained for the first time. In this way, an upsampling operation is completed, and the projection error from low-resolution space to high-resolution space is learned.

$$\hat{I}_0 = (I_0 \ast U) \uparrow s,$$

(2)

The convolution operation is performed on the upsampling $I_0$, and the downsampling feature map is obtained.

$$\hat{I}_1 = (I_0 \ast D) \downarrow s,$$

(3)

We calculated the high-frequency residuals of the LR feature maps $I_0$ and $\hat{I}_1$ twice, and perform deconvolution operations on the errors.

$$E_r = ((\hat{I}_1 - I_0) \ast U) \uparrow s,$$

(4)

We add $E_r$ to the HR feature map obtained after the first convolution to get the entire upsampling result.

$$\hat{I}_1 = I_0 + E_r,$$

(5)

where $\hat{I}_1$ represents the upsampling image obtained after deconvolution, $I_0$ represents the original LR image, $\ast$ represents the convolution operator, $U$ represents the upsampling convolution kernel, $\uparrow s$ represents the upsampling operator, and $\hat{I}_1$ represents the downsampling image obtained after the convolution operation, $D$ represents the downsampling convolution kernel, $\downarrow s$ represents the downsampling operator, and $E_r$ represents the feature map obtained after deconvolution of the high-frequency residual parts of the two LR images.

The downsampling process is opposite to the upsampling operation, and it is mainly used to learn the projection error relationship when the HR image is projected into the LR space. In this process, the image will be transformed into HR and LR space with the convolutional layer and deconvolution layer first, and the difference in the LR space will be obtained, and then these difference feature maps will be merged into the LR space. From the rate feature map, the corresponding downsampling feature mapping information is obtained.

We perform convolution operations on the image to obtain the downsampling feature map.

$$\hat{I}_1 = (I_1 \ast D') \downarrow s,$$

(6)

The deconvolution is performed on the upsampling image.

$$\hat{I}_2 = (I_1 \ast U') \uparrow s,$$

(7)

We calculate the high-frequency residuals of the LR feature maps $\hat{I}_1$ and $\hat{I}_2$ twice, and perform the convolution operation.

$$E_r' = ((\hat{I}_2 - \hat{I}_1) \ast D') \downarrow s,$$

(8)

We add $\hat{I}_1$ to the LR feature map obtained after the first convolution to get the result of downsampling.

$$\hat{I}_2 = \hat{I}_1 + E_r',$$

(9)

where $\hat{I}_1$ represents the LR image obtained after downsampling, $D$ represents the convolution kernel used for downsampling, $\hat{I}_2$ represents the image obtained after upsampling, $U$ represents the convolution kernel used for upsampling, and $E_r$ represents the feature map obtained after convolution of the high-frequency residual parts of the two LR images.

### 3.4. Sub-pixel convolutional upsampling layer

We reduce the dimensionality of the output obtained by the previous deep feature extraction block to the same dimension as the initial LR input, adaptively fuse and retain the previously learned features, perform global feature fusion and output, and reconstruct an HR image.

$$H_{\text{high}} = G_u(F_d(Y, R(x))),$$

(10)

where $Y$ denotes the initial input of the image, $R$ denotes is the high-frequency feature information learned by the nonlinear mapping, $F_d$ denotes the calculation operation of the local feature fusion layer, $G_u$ denotes the upsampling operation, and $H_{\text{high}}$ denotes the reconstructed HR image.
4. Experimental analysis

4.1. Dataset and implementation details

In the experiment, we use the public dataset DIV2K (Gao & Grauman, 2017) as the training dataset, which mainly includes 800 training data, 100 verification images, 100 test images, and the dataset images include rich scenes and rich edge and texture details. While training the model, we use 800 training images from the dataset. In order to avoid underfitting during the training process, the 800 images are augmented by random rotation of 90°, 180°, 270°, and horizontal rotation to expand the training data to 8 times the original sizes so that the model has enough training data, and can adapt to the problem of image reconstruction with different tilt angles (Cheng et al., 2020). For testing, we use three standard benchmark datasets: Set5 (Bevilacqua et al., 2012), Set14 (Zeyde et al., 2010), and Urban100 (Huang et al., 2015). At the same time, we use the values of peak signal-to-noise ratio (PSNR) (Wang et al., 2004) and structural similarity (SSIM) (K. Zhang et al., 2018) as the objective criteria for evaluating prediction results.

In addition, combined with CUDA 11.1 and PyTorch 1.8.1, we use Python code to implement PURN algorithm, and train and evaluate the algorithm through many experiments on NVIDIA GeForce RTX 3080 GPU and Ubuntu 18.04 operating systems, and the feature size of the PURN network for COVID-19 CT images SR is shown in Table 1.

Moreover, the initial learning rate is set to $10^{-4}$, the ADAM optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$, and use $L_1$ norm as the loss function.

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^{n} \| H_{\text{high}}(x_i) - F_{\text{high}}(x_i) \|^2,$$

where $i$ represents each pixel, $H_{\text{high}}$ represents the reconstructed HR image, and $F_{\text{high}}$ represents the original HR image.

CT imaging examination is one of the main methods of COVID-19 diagnosis, its value lies in the detection of lesions, the judgment of the nature of the lesions, and the assessment of the severity of the disease in order to facilitate clinical classification. At the same time, CT scan has the advantages of HR imaging and multi-plane reconstruction and observation, and the early CT manifestations of COVID-19 have certain characteristics. Therefore, chest CT examination plays a critical role in its early diagnosis and treatment, and is suitable for imaging screening of patients with high clinical suspicion. In order to build a COVID-19 diagnosis model based on deep learning, we tested it on the public COVID-CT dataset (COVID-19 CT images: https://github.com/UCSD-AI4H/COVID-CT). The COVID-CT (Qiu et al., 2019) dataset contains 349 CT slices of 216 COVID-19 patients and non-COVID-19 463 CT slices of the patient. It mainly collects medical images from websites and publications, that is, collected from 760 COVID-19-related pre-printed documents, which mainly come from medRxiv and bioRxiv.

4.2. Comparison with the state-of-the-arts

Our PURN method not only integrates residual ideas but also combines U-Net to improve the quality of predicted images to verify the effectiveness of the algorithm in this paper, we compare and analyze the prediction results of the algorithm in this paper and some mainstream algorithms on different datasets (Jiang et al., 2020). The experimental results are shown in Table 2, in the four public datasets Set5, Set14, BSD100, and Urban100, some mainstream algorithms for comparison include Bicubic (Lu et al., 2018), A+ (Timofte et al., 2015), SCNN (Timofte et al., 2015), SRCNN (Chang et al., 2004), FSRCNN (Kim et al., 2016a), VDSR (Dong, Loy, He et al., 2016), DRCN (Shi et al., 2016), LapSRN (Lai et al., 2017), DRRN (Kim et al., 2016b). Furthermore, we compare the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and subjective visual effects of the state-of-the-art algorithms when the scaling factors are 2, 3, and 4 on different datasets.

It is manifest from the Table 2 that our PURN algorithm has achieved excellent performance on the public test dataset, especially when the scaling factor is 2, our PURN algorithm achieves sub-optimal PSNR in the Urban100 dataset, but in the Set5 dataset, Set14 dataset achieves the best PSNR. In particular, compared with other state-of-the-art algorithms on the Set14 dataset, the improvement is the most obvious, with an increase of 0.46 dB relative to DRRN. In addition, our PURN method has reached the optimal value of the SSIM index.

In addition, in order to clearly observe the comparison results of detailed information, we zoomed in on some areas of the reconstructed results, and the experimental results are shown in Figure 6, it can be found that the reconstructed image of the Bicubic and SRCNN can hardly observe contours, and so on. For detailed information, although the reconstruction results of the FSRCNN, VDSR, and LapSRN algorithms have been slightly improved, some of the detailed information is still missing. The reconstruction results of the DRCN and DRRN algorithms obtain a better

| Network module | Input size | Output size |
|----------------|------------|-------------|
| Initial layer DUM up-block | H*W*1 | H*W*8 |
| up-block | H*W*4 | H*W*16 |
| up-block | H*W*16 | H*W*32 |
| up-block | H*W*32 | H*W*64 |
| down-block | H*W*64 | H*W*32 |
| down-block | H*W*32 | H*W*16 |
| down-block | H*W*16 | H*W*8 |
| Reconstruction layer | H*W*8 | H*W*1 |
Table 2. Average PSNR/SSIM for scale factors (2x, 3x, 4x) of various SISR methods, the best and second-best results are highlighted and underlined.

| Algorithm         | Scale | Set 5 | Set 14 | Urban 100 |
|-------------------|-------|-------|--------|-----------|
|                   |       | PSNR  | SSIM   | PSNR      | SSIM     | PSNR | SSIM |
| Bicubic (Lu et al., 2018) | 2x    | 33.69 | 0.931  | 30.250 | 0.702688 | 0.841 |
| A+ (K. Zhang et al., 2018) |       | 36.60 | 0.955  | 32.320 | 0.906295 | 0.895 |
| SRCNN (Timofte et al., 2015) |       | 36.58 | 0.954  | 32.350 | 0.905295 | 0.897 |
| SRCNN (Chang et al., 2004) |       | 36.72 | 0.957  | 32.510 | 0.908295 | 0.896 |
| FSRCNN (Kim et al., 2016a) |       | 37.05 | 0.956  | 32.660 | 0.909298 | 0.902 |
| VDSR (Dong, Loy, He et al., 2016) |       | 37.53 | 0.959  | 33.050 | 0.9130077 | 0.914 |
| DRCN (Shi et al., 2016) |       | 37.63 | 0.959  | 33.060 | 0.9123076 | 0.914 |
| LapSRN (Lai et al., 2017) |       | 37.52 | 0.959  | 33.080 | 0.9133041 | 0.910 |
| DRRN (Kim et al., 2016b) |       | 37.74 | 0.959  | 33.230 | 0.9143123 | 0.919 |
| PURN (Ours) |       | 37.90 | 0.961  | 33.690 | 0.9203122 | 0.920 |
| Bicubic (Lu et al., 2018) | 3x    | 30.41 | 0.869  | 27.790 | 0.77524460735 | 27.790 |
| A+ (K. Zhang et al., 2018) |       | 32.62 | 0.909  | 29.150 | 0.82026050799 | 29.150 |
| SRCNN (Timofte et al., 2015) |       | 32.62 | 0.908  | 29.160 | 0.81826210801 | 29.160 |
| SRCNN (Chang et al., 2004) |       | 32.78 | 0.909  | 29.320 | 0.82326250801 | 29.320 |
| FSRCNN (Kim et al., 2016a) |       | 33.18 | 0.914  | 29.370 | 0.82426430808 | 29.370 |
| VDSR (Dong, Loy, He et al., 2016) |       | 33.67 | 0.921  | 29.780 | 0.83227140829 | 29.780 |
| DRDN (Shi et al., 2016) |       | 33.83 | 0.922  | 29.770 | 0.83227150828 | 29.770 |
| LapSRN (Lai et al., 2017) |       | 33.82 | 0.922  | 29.870 | 0.83227070828 | 29.870 |
| DRRN (Kim et al., 2016b) |       | 34.03 | 0.924  | 29.960 | 0.83527530764 | 29.960 |
| PURN (Ours) |       | 34.12 | 0.924  | 30.270 | 0.84227042082 | 30.270 |
| Bicubic (Lu et al., 2018) | 4x    | 28.43 | 0.811  | 26.220 | 0.71525314665 | 26.220 |
| A+ (K. Zhang et al., 2018) |       | 30.32 | 0.860  | 27.340 | 0.75243407215 | 27.340 |
| SRCNN (Timofte et al., 2015) |       | 30.41 | 0.863  | 27.390 | 0.75124520726 | 27.390 |
| SRCNN (Chang et al., 2004) |       | 30.50 | 0.863  | 27.520 | 0.75324530725 | 27.520 |
| FSRCNN (Kim et al., 2016a) |       | 30.72 | 0.866  | 27.610 | 0.75324620728 | 27.610 |
| VDSR (Dong, Loy, He et al., 2016) |       | 31.35 | 0.883  | 28.020 | 0.76825180754 | 28.020 |
| DRDN (Shi et al., 2016) |       | 31.54 | 0.884  | 28.030 | 0.76825140752 | 28.030 |
| LapSRN (Lai et al., 2017) |       | 31.54 | 0.885  | 28.190 | 0.77225210756 | 28.190 |
| DRRN (Kim et al., 2016b) |       | 31.68 | 0.888  | 28.210 | 0.77225440764 | 28.210 |
| PURN (Ours) |       | 32.18 | 0.895  | 28.590 | 0.78225090786 | 28.590 |
visual experience, but compared to the algorithm in this paper, there is still a lack of sharp edge information. Therefore, by comprehensively comparing the above eight algorithms, our model can extract the original features of LR COVID-19 CT images more accurately and comprehensively, reconstruct its detailed

**Figure 6.** Comparison of rendering of COVID-19 CT images with SR magnification of 4 by our PURN and other state-of-the-art algorithms.

**Figure 7.** Comparison with the performance and running time on COVID-19 CT dataset for 4x enlargement.
texture more clearly and sharply, and at the same time, the edge information of the reconstructed image has good continuity, and no artifacts, the reconstructed image is the closest to the real image.

The performance and running time of our PURN method and other state-of-the-art methods to reconstruct a CT image are compared. Figure 7 shows that the average running time of our PURN on the COVID-19 CT dataset for 4x enlargement can achieve 0.39 s, as well as VDSR, but the PSNR is better at 0.19 dB than VDSR. Therefore, our PURN method can obtain the highest values of PSNR under the condition of better running time, which can meet the real-time application in actual scenarios.

5. Conclusion

In this paper, we propose the Progressive U-Net Residual Network (PURN) for COVID-19 CT image super-resolution. In PURN, in view of the loss of details and blur artifacts in the SR reconstruction of COVID-19 CT images, we design a dual U-Net module (DUM) to extract LR image features, which can alleviate the problem of underutilization of feature information and loss of high-frequency information in the learning process. At the same time, we make full use of the difference between feature images, so that we can find more useful high-frequency information in the reconstruction of the predicted image. The experimental results prove the superiority of our PURN method in both the PSNR and SSIM indicators, and the detailed information on the COVID-19 CT image reconstruction is more abundant. Furthermore, our method is also of great significance for experts to find lesions more accurately and improve the accuracy of clinical diagnosis, and provides new ideas for the theoretical research of medical image SR reconstruction methods. For future implementation, we propose an algorithm that performs SR at any scale, which is closer to actual application scenarios, and can integrate medical image registration and multi-frame image SR algorithms to further improve the accuracy of SR

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China [Grant No. 61772532, 61976215].

ORCID

Defu Qiu http://orcid.org/0000-0003-4921-9145
Xuesong Wang http://orcid.org/0000-0002-0371-0142

References

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., & Xia, L. (2020). Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. Radiology, 296 (2), 0033–8419, 206042. https://doi.org/10.1148/radiol.2020200642

Algaissi, A. A., Alharbi, N. K., Hassanain, M., & Hashem, A. M. (2020). Preparedness and response to COVID-19 in Saudi Arabia: Building on MERS experience. Journal of Infection and Public Health, 13 (6), 834–838. 1876–0341. https://doi.org/10.1016/j.jiph.2020.04.016

Alshammary, T. M., Altebainawi, A. F., & Alzenzi, K. A. (2020). Importance of early precautionary actions in avoiding the spread of COVID-19: Saudi Arabia as an example. Saudi Pharmaceutical Journal, 28 (7), 898–902. 1319-0164. https://doi.org/10.1016/jjsp.202005005

Ayinde, O. O., Usman, A. B., Posi, A., & Gbolahan, A. (2020). A cross-sectional study on Oyo state health care worker’s knowledge, attitude and practice regarding corona virus disease 2019 (COVID-19). Advances in Infectious Diseases, 10 (3), 6. https://doi.org/10.4236/aid.2020.103002

Barry, M., Ghonem, L., Alsharidi, A., Awadh, A., Al-Shahrani, F., & Naif, H. (2020). Coronavirus disease 2019 (COVID-19) pandemic in the kingdom of Saudi Arabia: Mitigation measures and hospitals preparedness. Journal of Natural Science and Medicine, 3 (3), 1–11. https://doi.org/10.1016/j.jnsm.2020.04.016

Bevlacqua, M., Roumy, A., Guillemot, C., & Alberi-Morel, M. L. (2012). Low-complexity single-image super-resolution based on nonnegative neighbor embedding. British machine vision conference (pp. 1–10). BMVA Press. https://doi.org/10.5244/C.26.135

Chang, H., Yeung, D. Y., & Xiong, Y. (2004). Super-resolution through neighbor embedding. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR) (pp. 275–282).

Cheng, X., Fu, Z., & Yang, J. (2020). Zero-shot image super-resolution with depth guided internal degradation learning. ECCV. Springer.

Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38 (2), 295–307. https://doi.org/10.1109/TPAMI.2015.2439281

Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the super-resolution convolutional neural network. European conference on computer vision (pp. 391–407). Springer. https://doi.org/10.1007/978-3-319-46475-6_25

Gao, R., & Grauman, K. (2017). On-demand learning for deep image restoration. Proceedings of the IEEE international conference on computer vision (pp. 1086–1095).

Ge, D., Yao, X., & Lian, Z. (2016). Binocular vision calibration and 3D reconstruction with an orthogonal learning neural network. Multimedia Tools and Applications, 75 (23), 15635–15650. https://doi.org/10.1007/s11042-015-2845-5

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. IEEE conference on computer vision and pattern recognition (pp. 770–778). IEEE Computer Society. https://doi.org/10.1109/CVPR.2016.90

Huang, J. B., Singh, A., & Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. IEEE conference on computer vision and pattern recognition (pp. 5197–5206). IEEE. https://doi.org/10.1109/CVPR.2015.7299156
Jiang, K., Wang, Z., Yi, P., Wang, G., Gu, K., & Jiang, J. (2020). ATMEN: Adaptive-threshold-based multimodal fusion network for compressed face hallucination. IEEE Transactions on Multimedia, 22 (10), pp. 2734-2747. https://doi.org/10.1109/TMM.2019.2960586

Keys, R. (1981). Cubic convolution interpolation for digital image processing. IEEE Transactions on Acoustics, Speech, and Signal Processing, 29(6), 1153–1160. https://doi.org/10.1109/TASSP.1981.1163711

Kim, J., Lee, J. K., & Lee, K. M. (2016a). Accurate image super resolution using very deep convolutional networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1646–1654). arXiv:1511.04587

Kim, J., Lee, J. K., & Lee, K. M. (2016b). Deeply-recursive convolutional network for image super-resolution. Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1637–1645). https://doi.org/10.1109/CVPR.2016.1

Lai, W.-S., Huang, J.-B., Ahuja, N., & Yang, M. H. (2017). Deep laplacian pyramid networks for fast and accurate super resolution. IEEE conference on computer vision and pattern recognition (pp. 5835–5843). IEEE Computer Society. https://doi.org/10.1109/CVPR.2017.618

Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 136–144). https://doi.org/10.1109/CVPRW.2017.151

Lu, Y., Zhou, Y., Jiang, Z., Guo, X., Yang, Z. (2018). Channel attention and multi-level features fusion for single image super-resolution. 2018 IEEE Visual Communications and Image Processing (VCIP) (pp. 1–4). IEEE.

Mei, Y., Fan, Y., Zhou, Y., Huang, L., Huang, T. S., & Shi, H. (2020) Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5689–5698, https://doi.org/10.1109/CVPR42600.2020.00573.

Qiu, D., Cheng, Y., Wang, X., & Zhang, X. (2020). Multi-window back-projection residual networks for reconstructing COVID-19 CT super-resolution images, 200, 105934. Computer Methods and Programs in Biomedicine. https://doi.org/10.1016/j.cmpb.2021.105934

Qiu, D., Zhang, S., Liu, Y., Zhu, J., & Zheng, L. (2019). Super-resolution reconstruction of knee magnetic resonance imaging based on deep learning. Computer Methods and Programs in Biomedicine, 187(12), 105059. https://doi.org/10.1016/j.cmpb.2019.105059

Qiu, D., Zheng, L., Zhu, J., & Huang, D. (2021). Multiple improved residual networks for medical image super-resolution. Future Generation Computer Systems, 116, 200–208. https://doi.org/10.1016/j.future.2020.11.001

Ronneberger, O., Fischer, P., & Brox, T. (2015) U-Net: Convolutional networks for biomedical image segmentation. International conference on medical image computing and computer-assisted intervention (pp. 234–241). Springer.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. IEEE conference on computer vision and pattern recognition (pp. 1874–1883). IEEE Computer Society. https://doi.org/10.1109/CVPR.2016.207

Tai, Y., Yang, J., & Liu, X. (2017). Image super-resolution via deep recursive residual network. 2017 IEEE conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2790–2798). https://doi.org/10.1109/CVPR.2017.298

Timofte, R., De Smet, V., & Van Gool, L. (2015). A+ : Adjusted anchored neighborhood regression for fast super-resolution. In D. Cremers, I. Reid, H. Saito, & M.-H. Yang (Eds.), Computer vision-ACCV 2014 (pp. 111–126). Springer Int.

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. https://doi.org/10.1109/TIP. 2003.819861

Wang, Z., Liu, D., Yang, J., Han, W., & Huang, T. (2015). Deep networks for image super-resolution with sparse prior. 2015 IEEE International Conference on Computer Vision (ICCV) (pp. 1–15). IEEE. ArXiv preprint arXiv:1507.08905

Yang, J., Wright, J., Huang, T. S., & Ma, Y. I. (2010). Image super-resolution via sparse representation. IEEE Transactions on Image Processing, 19(11), 2861–2873. https://doi.org/10.1109/TIP.2010.2050625

Zeyde, R., Elad, M., & Protter, M. (2010). On single image scale-up using sparse-representations. International conference on curves & surfaces (pp. 711–730). Springer. https://doi.org/10.1007/978-3-642-27413-8_47

Zhang, K., Zuo, W., & Zhang, L. (2018). Learning a single convolutional super-resolution network for multiple degradations. 2018 IEEE/CVF conference on computer vision and pattern recognition (pp. 3262–3271). IEEE. ArXiv pre-print arXiv:1712.06116

Zhang, Y., Li, K., Kai, L., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. Proceedings of the European Conference on Computer Vision (ECCV), 11211(7), 294–310. https://doi.org/10.1007/978-3-030-01234-2_18