Magnetic Resonance Image Super-Resolution via Multi-Resolution Learning

Wang Wentao[1,*], Zhang Kaixiang[1]

[1]Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, China
*Corresponding author

Abstract: High-resolution magnetic resonance images are of great significance for medical diagnosis. A convolutional neural network with multi-resolution learning is proposed for magnetic resonance image (MR) superresolution. The network is an improved deep residual network, which involves residual units for feature extraction, a deconvolution layer for multi-resolution up-sampling, and a multi-resolution learning layer. The proposed network performs the super-resolution task in the low-resolution space, which can accelerate the network. Multiresolution upsampling is put forward to integrate multiple residual unit information and to accelerate the network. Multi-resolution learning can adaptively determine the contributions of these upsampled high-dimensional feature maps to high-resolution MR image reconstruction. Experiment results indicate that the proposed method can achieve a good super-resolution reconstruction performance for magnetic resonance images, which is superior to the state-of-the-art deep learning methods.

1. Introduction

Magnetic Resonance (MR) imaging is a powerful, flexible and non-invasive imaging technique that provides high quality cross-sectional images of organs and structures in the body and is widely used in medical diagnosis.

In particular, High-Resolution (HR) provides doctors with clearer images so that they can more accurately analyse the body's tissue structure and focal areas and give precise diagnostic advice [1]. Therefore, the study of Super Resolution (SR) in MR images has started to receive a lot of attention from scholars [2].

Due to the powerful self-learning performance of Convolutional NeuralNetwork (CNN), it has been widely used in the field of machine vision [3-8], and more and more researchers have introduced CNN into MR image super-resolution reconstruction research, mainly using two strategies of pre-amplification and single upsampling for image super-resolution reconstruction.

Super-resolution reconstruction in magnetic resonance (MR) imaging is an active area of research. Various methods have been proposed to enhance the resolution of low-resolution MR images. In the context of MR image super-resolution reconstruction, the pre-amplification strategy is commonly employed to improve the performance of convolutional neural networks (CNNs) in generating high-resolution images.One approach mentioned in reference [9] is the use of bi-trivial interpolation or

inverse convolution as pre-amplification techniques. These methods upscale the input low-resolution images before feeding them into the CNN for learning. The purpose of pre-amplification is to provide the network with more detailed information from the low-resolution images, which can assist in generating higher-resolution reconstructions.

Different studies have applied CNNs to MR image super-resolution reconstruction. Pham et al. [10] utilized a super-resolution CNN for brain MR image reconstruction, while Oktav et al. [13-14] proposed a residual CNN-based method for cardiac MR image super-resolution. In these approaches, the input images were pre-amplified using deconvolution or residual connections before being processed by the CNN. It is worth noting that the network architecture used in the latter study bears similarities to VDSR [15]. Shi et al. [16] introduced a progressive wide residual network with fixed jump connections specifically for reconstructing high-resolution MR images of the brain. Similarly, the input images in their method were pre-amplified using bi-trivial interpolation prior to CNN learning. This strategy enables the network to operate in the high-resolution image space rather than directly learning from low-resolution images. However, it should be acknowledged that this pre-amplification approach may not necessarily yield optimal results and can increase computational complexity.

Overall, these pre-amplification strategies in MR image super-resolution reconstruction aim to enhance the network's ability to generate high-resolution images. While they may improve performance, it is important to consider the potential trade-offs in terms of computational complexity and optimality of the reconstruction results.The super-resolution reconstruction of MR images based on the single upsampling strategy does not require pre-amplification, and the image amplification operation is implemented at the end of the network structure by deconvolution or Pixel Shuffle, and the super-resolution reconstruction is done in the low-resolution image space, which will reduce the computational complexity significantly. Tanno et al [19] applied the Efficient Sub-Pixel Convolutional Neural Network (ESPCN) [20] to diffusion-weighted magnetic resonance imaging (dMRI) super-resolution reconstruction by implementing pixel shuffling in the effective sub-pixel convolutional layer at the end of the network. Although these methods can speed up the network, the layer-by-layer cascading of the network used will result in the loss of information from the current layer not being transmitted to the next layer.

To solve the above problems, an MR super-resolution reconstruction network based on a depth residual network [21] is designed and applied to MR image super-resolution reconstruction in this paper. The network has multiple jump-connected residual units, which can extract image feature maps of different resolutions layer by layer[11]; the feature maps of different resolutions output from multiple residual units are fused with Low Resolution (LR) MR images by a deconvolution layer for multi-resolution upsampling respectively, in order to avoid information loss during network transmission to the maximum extent; finally, the multi-resolution Finally, multi-resolution learning is performed on the multi-resolution fused feature maps, and adaptive learning determines the contribution of each resolution upsampled fused feature map to the super-resolution reconstruction of MR images, and finally realizes the super-resolution reconstruction of MR images.

2. Multi-resolution learning convolutional neural networks

2.1 Network Architecture

The multi-resolution learning convolutional neural network designed in this paper is shown in and consists of three stages: feature extraction, multi-resolution upsampling and multi-resolution learning. The network input is a low-resolution MR image and the output is a high-resolution MR image, and the whole super-resolution reconstruction process is an end-to-end learning process.
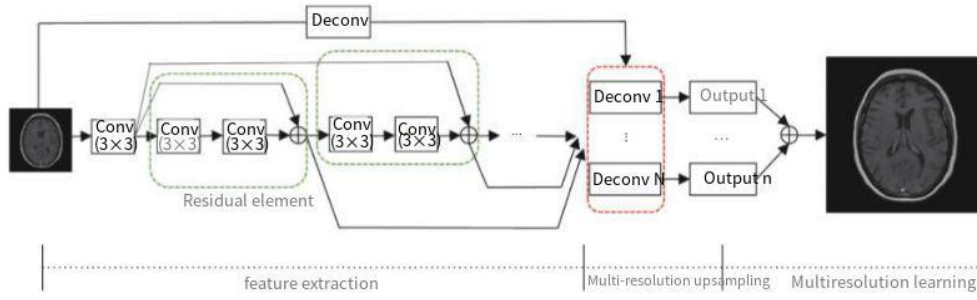
Figure 1: The architecture of the proposed network

In the feature extraction stage, the low-resolution MR image is first passed through the first convolutional layer for obtaining coarse feature maps, after which several residual units are cascaded layer by layer to achieve coarse to fine feature extraction. To avoid information loss, the coarse feature map extracted from the first convolutional layer is superimposed on the output of each residual unit, which is used as the output of the residual unit and fed to the next residual unit, and also passed to the multi-resolution upsampling deconvolution layer. Each residual unit has two convolutional layers, all of which, including the first layer, have 128 convolutional kernels each, with a convolutional kernel size of 3×3[12].

In the multi-resolution upsampling deconvolution layer, the superimposed feature map output from each residual unit is deconvoluted to obtain the upsampled feature map of each residual unit, and then fused with the upsampled feature map of the low-resolution MR image respectively to achieve multi-resolution upsampling. The fused upsampled feature maps are input to the multi-resolution learning layer, and the contribution of the fused upsampled feature maps of each resolution to the reconstructed high-resolution MR image is determined by adaptive learning, and the final weighted fusion achieves super-resolution reconstruction of MR images.

2.2 Residual units

For deep networks, cascading multiple convolutional layers will improve feature extraction, but will also lead to increasingly difficult network convergence. To improve the speed of network optimisation, residual units have been widely used in deep networks [22]. Therefore, the paper also uses the residual strategy to construct residual units and cascade multiple residual units to improve feature extraction and obtain feature maps with different resolutions. The structure of each residual unit is shown in Figure 2, with two convolutional layers, each with 128 convolutional kernels, and a convolutional kernel size of 3×3.
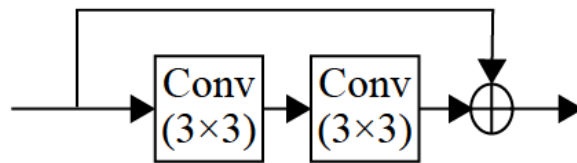

Figure 2: Residual unit

Assuming that the input low-resolution MR image is x, 128 coarse feature maps can be obtained by the first convolution layer of the network.

$$U_0 = f_{\text{conv}}(\boldsymbol{x}) \tag{1}$$

This is then passed on to each residual unit. The output of each residual cell is therefore

$$U_i = U_0 + f_{\text{conv}}\left(f_{\text{conv}}\left(U_{i-1}\right)\right) \tag{2}$$

where U_{i-1} and $U_i(i=1,2,\ldots,N)$ denote the input and output of the ith residual unit and N represents the number of residual units.

2.3 Multi-resolution upsampling

The residual units will inevitably encounter information loss in network cascade transmission, and the information loss caused by them will accumulate more and more as the network depth increases, eventually affecting the subsequent high-resolution reconstruction of detailed information. In order to avoid information loss in network cascade transmission and accelerate the network, this paper designs a multi-resolution upsampling scheme to achieve the fusion of the output feature map information of multiple residual units in order to maximize the retention of detail information.

In order to retain maximum detail information in the low-resolution image, the upsampled feature maps of the low-resolution MR image are fused with the output feature maps of each residual unit separately, so the input of the deconvolution layer is N+1 feature maps and the output is N fused upsampled feature maps $\hat{y}_i$

$$\hat{y}_i = f_{\text{dec}}(x) + f_{\text{dec}}(U_i), i=1,2,\cdots,N \tag{3}$$

where U_i is the output feature map with the i-th residual unit and $f_{\text{dec}}(.)$ is the deconvolution operation.

2.4 Multi-resolution learning

Obviously, the fused upsampled feature maps output from the deconvolution layer contain different resolution image information and inevitably contribute differently to the final high-resolution MR image. Therefore, a multi-resolution learning strategy is proposed to adaptively learn a fused upsampled feature map $\hat{\mathbf{y}}_i$ of each resolution weighted to super-resolution reconstruction of the MR high-resolution image $\hat{y}$, i.e.

$$\hat{y} = \sum_{i=1}^{N} w_i \hat{\mathbf{y}}_i \tag{4}$$

where w_i denotes the contribution of the fused upsampled feature map $\hat{\mathbf{y}}_i$ Clearly, the total loss L of the whole network is composed of the loss L_i of each residual unit and the loss L_{sum} of the multi-resolution learning layer, i.e.

$$L = \sum_{i=1}^{N} L_i + L_{\text{sum}} \tag{5}$$

$$L_i = \frac{1}{2} y - \hat{y}_i^2 \tag{6}$$

$$L_{\text{sum}} = \frac{1}{2} \| \mathbf{y} - \mathbf{y} \|^2 \tag{7}$$

where y represents the real high-resolution image. Stochastic gradient descent is used to optimise the network, and gradient cropping is employed to improve convergence performance and suppress

gradient explosion.

3. Experimental results and discussion

3.1 Dataset and pre-processing

The experimental data in this paper were selected from the Brain-Tumor-Progression subset of the Cancer Medicine Image Public Dataset (https://wiki.canerimagingarchive.net/display/Public/Brain-Tumor-Progression), and a randomly selected A total of 548 images from 24 MR image sequences were used to build the training set, and 646 images from 28 MR image sequences were randomly selected for testing. All MR images in the sub-dataset were used as the true HR images, and the LR images after double triple interpolation were used as the true LR images.

For the training set, the real LR image is cropped into multiple m $\times$ m sized LR image blocks. Due to the intrinsic properties of deconvolution [18], the corresponding real HR image is cropped into multiple [(m-1)$\times$n+1] $\times$ [(m-1)$\times$n+1] sized HR image blocks (n is the magnification). In this paper, super-resolution reconstructions with magnifications of 2, 3 and 4 are discussed. Thus, for $\times$2, $\times$3 and $\times$4, the LR/HR image block sizes in the training set are 142/272, 102/282,82/292 pixels respectively; the cropped LR images are in steps of 11, 7 and 6 pixels respectively; thus, approximately 175 000 training samples are ultimately available.

Image super-resolution is usually evaluated using Peak Signal-To-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [23]. The higher the PSNR and SSIM values, the better the image reconstruction effect.

3.2 Network parameters configuration and hardware implementation

The weights were initialised using the method from the literature [24], with weight decay set to 0.000 1, momentum to 0.9, minimum batch size to 128 and epoch to 100. The initial learning rate was set to 0.1 and decayed by a factor of 10 after every 20 epochs. All deep learning models were trained and tested using the PyTorch package. All network models were trained on a computer with Intel®Core™ i5-9600KCPU@3.40 GHz x 6 and an NvidiaGTX2080Ti GPU, unless otherwise stated. All test tasks were implemented on a computer with an Intel Xeon E5-2360 v3 CPU2.40 GHz$\times$16 and no GPU.

3.3 Discussion of multi-resolution upsampling

The effect of a multi-resolution upsampling strategy on the effect of network reconstruction is discussed here by fixing the number of residual units in the network to nine. In addition to the single upsampling strategy in the literature [18], an additional dual upsampling strategy is defined in this paper to illustrate that the fusion of low-resolution MR image information is also useful for super-resolution reconstruction. The so-called double upsampling is based on the literature [18], where the low-resolution MR images are upsampled and then superimposed on the single-sampling network output to achieve super-resolution reconstruction of the fused high-resolution MR images. Again, here the magnification is only considered for the case of 2.

The single-sampling strategy network model only captures high-dimensional features by cascading residual units to achieve coarse-to-fine feature extraction, so some detailed information in the low-resolution MR images may be lost layer by layer during the cascade transfer process, affecting the final high-resolution MR image reconstruction. As shown in Table 1, the single upsampling strategy network model did obtain the worst super-resolution reconstruction. Because the dual upsampling strategy network model incorporates the initial low-resolution MR image

information, it obtains higher PSNR and SSIM values than the single-sampling strategy network model. In contrast, the multi-resolution upsampling strategy network in this paper passes the feature maps output from all residual units to the multi-resolution upsampling layer, and separately fuses the initial low-resolution MR image information, thus obtaining the best super-resolution reconstruction results for high-resolution MR images.

Table 1: MR Image SR performance of the proposed network models with different upsampling strategies (Scale factor=2).

Evaluation indicators	Up-sampling strategy		
	Single upsampling	Double upsampling	Multi-resolution upsampling
PSNR	45.33	45.56	**45.67**
SSIM	0.9899	0.9943	**0.9945**

3.4 Comparison of reconstruction performance with other super-resolution methods

After the above discussion, the residual units in the proposed network model were fixed by nine and compared with some of the latest MR image super-resolution deep learning networks (SRCNN [10], de-CNN [13- 14], FSCWN [16] FSRCNN [17] and ESPCN [19]). As shown in Table 2, the deep learning methods all obtained far better image super-resolution results than the bicubic interpolation methods.

Table 2: MR Image SR performance of various image SR methods

Evaluation indicators	MR image super-resolution reconstruction methods						
	Scale	Bicubic	SRCNN	ESPCN	FSRCNN	de-CNN	FSCWN	Ours
PSNR	×2	40.55	44.96	44.39	43.46	44.26	44.03	46.56
	×3	35.12	36.73	36.27	37.52	36.34	-	38.75
	×4	32.31	33.92	33.53	34.25	32.60	-	35.39
SSIM	×2	0.987 1	0.977 5	0.965 9	0.990 8	0.9911 0.9911		0.9987
	×3	0.941 4	0.961 0	0.954 5	0.965 7	0.9641	-	0.9696
	×4	0.921 6	0.932 0	0.933 8	0.932 0	0.9344	-	0.9389

The SRCNN uses a pre-amplification strategy before network learning that may introduce some artifacts and lose some prior information, and the network has only three convolutional layers, so the SRCNN has the worst super-resolution reconstruction performance among all the deep learning methods. reconstruction, which can learn the non-linear mapping relationship between LR images to HR images directly from LR images, and thus obtain better reconstruction results than SRCNN, although their network depths are not much different from SRCNN. Although de-CNN and FSCWN use a pre-amplification strategy, they use residual learning to cascade the network reasonably deeply and their network depth is much greater than the first three network models, thus these two factors lead to higher PSNR and SSIM values obtained by de-CNN and FSCWN than ESPCN and FSRCNN. It should be noted that the network can only directly achieve 2X super-resolution reconstruction due to the fact that FSCWN uses 2×2 pooling to reduce the computational effort in the pooling layer to reduce the image. In contrast, the network in this paper uses multiple residual units to capture the high-dimensional feature maps of the low-resolution MR images at different resolutions, and uses multi-resolution upsampling to achieve the fusion of the high-dimensional feature maps and the low-dimensional feature maps (obtained from the low-resolution MR images), avoiding the loss of information in the cascade transfer of image details in the residual units, while the multi-resolution

learning strategy adaptively balances the impact of each resolution feature map on the reconstruction of high-resolution MR images. As a result, the network in this paper obtains the optimal super-resolution reconstruction of high-resolution MR images. A comparison of 2 visual examples of super-resolution reconstructions (magnification 2) is given in Figure 3, and the results are consistent with the objective data comparison shown in Table 2.

     (a) GT   (b) Ours   (c) SRCNN   (d) ESPCN

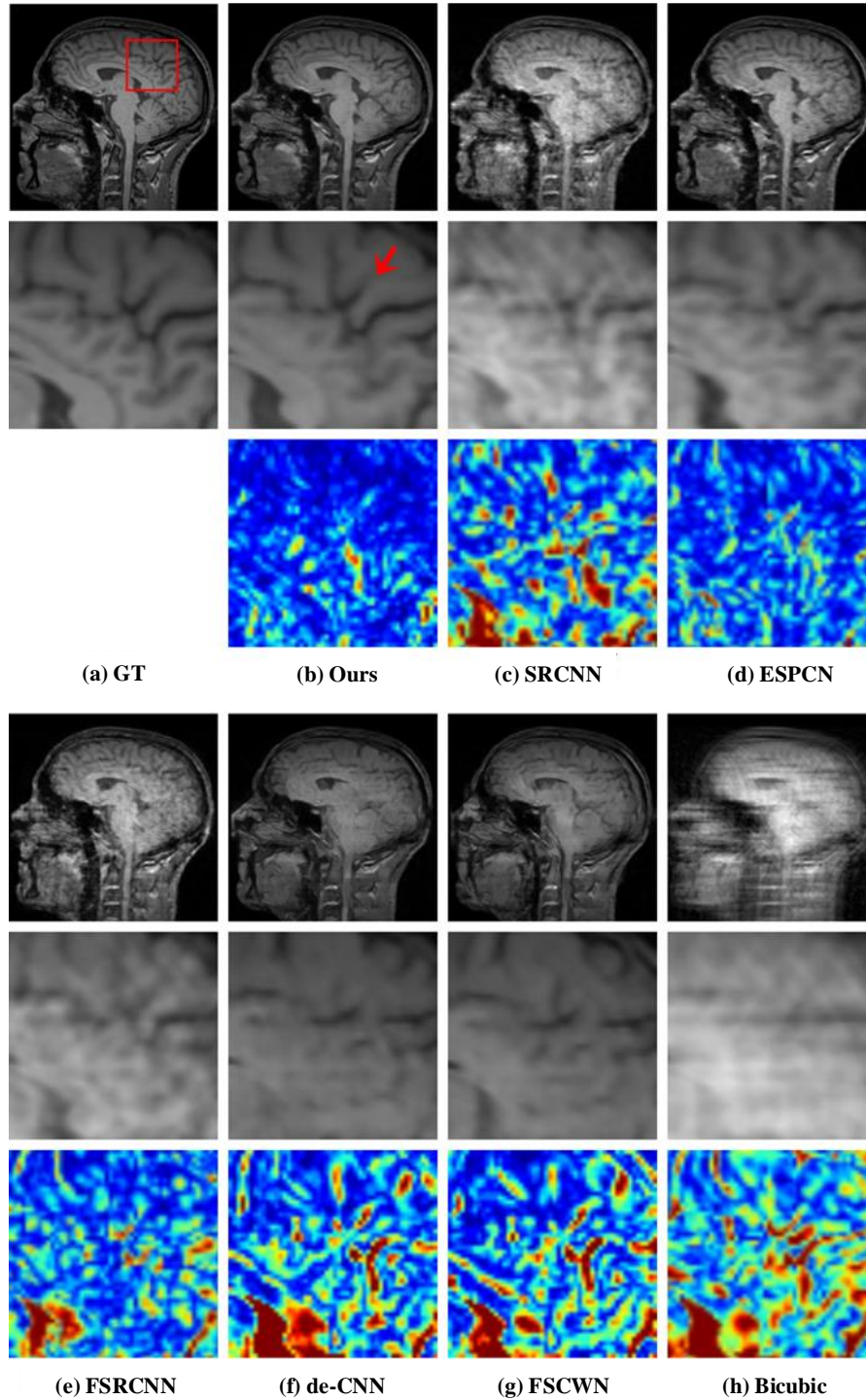   (e) FSRCNN   (f) de-CNN   (g) FSCWN   (h) Bicubic

Figure 3: Results of the reconstructed No.1 MR images with different algorithms

While training time is an important issue for deep learning, testing time, especially in the CPU, is even more critical for practical applications. As shown in Table 3, the pre-amplification strategy

performs super-resolution reconstruction in HR image space, while the single upsampling strategy performs super-resolution reconstruction in LR image space, so the former networks (SRCNN, de-CNN and FSCWN) take more time than the latter networks (FSRCNN and ESPCN) when the difference in depth between the two networks is not particularly large. For example, the SRCNN, which has only 3 layers of 9-1-5 convolution, takes less time to reconstruct. The network in this paper is also a multi-resolution upsampling strategy, which also performs super-resolution reconstruction in LR image space, and therefore takes less time than the pre-amplification strategy networks de-CNN and FSCWN with about the same network depth. However, compared to the two single upsampling networks FSRCNN (7 convolutional layers and 1 deconvolutional layer with no more than 56 convolutional kernels) and ESPCN (3 convolutional layers and 1 pixel shuffle layer with no more than 64 convolutional kernels in each convolutional layer), the network in this paper has more layers (19 convolutional layers and 1 deconvolutional layer) and more convolutional kernels (128 convolutional kernels in each convolutional layer) Therefore, the network in this paper will take slightly more time to reconstruct.

Table 3: Running time for testing via different methods

Magnification	MR super-resolution reconstruction methods					
	SRCNN	FSRCNN	ESPCN	de-CNN	FSCWN	Ours
x2	0.23	0.13	0.05	1.16	1.29	1.02
x3	0.21	0.08	0.04	1.14	-	0.47
x4	0.24	0.07	0.03	1.13	-	0.32

4. Summary and outlook

MR image super-resolution is of great research and display significance for aiding medical diagnosis. In this paper, a multi-resolution learning convolutional neural network based MR image super-resolution method is proposed. The input of the network is a low-resolution MR image, and the information is passed through multiple residual units in LR image space, and the output of high-resolution MR image reconstruction is directly achieved by fully capturing image details at different resolutions through multi-resolution upsampling and multi-resolution learning. Experimental results show that for different magnifications ($\times 2$, $\times 3$ and $\times 4$), the network in this paper can improve the PSNR/SSIM values to 45.46 / 0.991 4, 37.65 / 0.968 9 and 34.09 / 0.939 7 respectively with less CPU test time consumption, and the super-resolution reconstruction is better than some of the latest deep learning methods for MR images. Since Markov random fields can find correlations between HR/LR images, the introduction of Markov random fields can be considered to reduce the number of convolutional kernels to further accelerate the network proposed in this paper.

References

[1] Shi F, Cheng J, Wang L, Etal. LRTV: MR image super- resolution with low-rank and total variation regularizations [J]. IEEE Transactions on Medical Imaging, 2015, 34(12): 2459-2466.
[2] Plenge E, Poot D H J, Bernsen M, et al. Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? [J]. Magnetic Resonance in Medicine, 2012, 68(6):1983-1993.
[3] Wu J, Cai N, Li F, etal. Automatic detonator code recognition via deep neural network [J]. Expert Systems with Applications, 2020, 145: 113121.
[4] Fu X, Cai N, Huang K, et al. M-net: a novel u-net with multstream feature fusion and multscale dilated convolutions for bile ducts and hepatolith segmentation [J]. IEEE Access, 2019, 7: 148645-148657.
[5] Wu J, Cai N, Chen W, etal. Automatic detection of hardhats worn by construction personnel: a deep learning approach and benchmark dataset [J]. Automation in Construction, 2019, 106: 102894.

[6] Weng S W, Peng Y H, Wei B, et al. A two-stage-video forgery based on inception-v3 network [J]. Journal of Guangdong University of Technology, 2019, 36(6): 16-23.

[7] Gao J Y, Liu W Y, Yang Z G. Object tracking combined with attention and feature fusion [J]. Journal of Guangdong University of Technology, 2019, 36(4): 18-23.

[8] Ye W J, Gao H J, Weng S W, et al. A two-stage effect rendering method for art font based on CGAN network [J].Journal of Guangdong University of Technology, 2019, 36(3): 47-55.

[9] Haris M, Shakhnarovich G, Ukita N. Deep backprojection networks for superresolution [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1664-1673.

[10] Pham C H, Ducournau A, Fablet R, et al. Brain MRI super-resolution using deep 3D convolutional networks [C] //2017 IEEE 14th International Symposium on Biomedical Imaging. Melbourne: IEEE, 2017: 197-200.

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

[12] Dong C, Loy Cc, He K, et al. Image super-resolution us- ing deep convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2): 295-307.

[13] Oktay O, Bai W, Lee M, et al. Multi-input cardiac image super-resolution using convolutional neural networks[C]// International Conference on Medical Image Computing and Computer-assisted Intervention. Istanbul: Springer, Cham, 2016: 246-254.

[14] Giannakidis A, Oktay A O, Keegan J. Super-resolution reconstruction of late gadolinium cardiovascular magnetic resonance images using a residual convolutional neual network 25th Scientific Meeting of the Int[C]//International Society for Magnetic Resonance in Medicine. Hawaii: [s.n.], 2017.

[15] Kim J, Kwon Lee J, Mu Lee K. Accurate image super- resolution using very deep convolutional networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1646-1654.

[16] Shi J, Li Z, Ying S, et al. MR image super-resolution via wide residual networks with fixed skip connection [J]. IEEE Journal of Biomedical and Health Informatics, 2018, 23(3): 1129-1140.

[17] Mcdonagh S, Hou B, Alansary A, et al. Context- sensitive super-resolution for fast fetal magnetic resonance imaging[C]//Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment. Canada: Springer, Cham, 2017: 116-126.

[18] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network[C]//European Conference on Computer Vision. Amsterdam: Springer, Cham, 2016: 391-407.

[19] Tanno R, Worrall D E, Ghosh A, etal. Bayesian image quality transfer with CNNs: exploring uncertainty in dMRI super-resolution [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Istanbul: Springer, Cham, 2017: 611-619.

[20] Shi W, Caballero J, Huszár F, etal. Real-time single image and video super-resolution using an efficient subpixel convolutional neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1874-1883.

[21] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017: 3147-3155.

[22] He K, Zhang X, Ren S, etal. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.

[23] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[24] He K, Zhang X, Ren S, et al. Delving deep into recti- fiers: Surpassing human-level performance on image net classification[C]//Proceedings of the IEEE International Conference on Computer Vision. Boston: IEEE, 2015: 1026-1034.