TRANSFORM DOMAIN BASED MEDICAL IMAGE SUPER-RESOLUTION VIA DEEP MULTI-SCALE NETWORK

Chunpeng Wang⋆ *Simiao Wang*† *Bin Ma*⋆∗ *Jian Li*⋆∗ *Xiangjun Dong*⋆ *Zhiqiu Xia*‡∗

⋆ School of Information, Qilu University of Technology (Shandong Academy of Sciences)
†School of Electronic and Information Engineering, Liaoning Technical University
‡Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology

ABSTRACT

This paper proposes a new medical image super-resolution (SR) network, namely deep multi-scale network (DMSN), in the uniform discrete curvelet transform (UDCT) domain. DMSN is made up of a set of cascaded multi-scale fushion (MSF) blocks. In each MSF block, we use convolution kernels of different sizes to adaptively detect the local multi-scale feature, and then local residual learning (LRL) is used to learn effective feature from preceding MSF block and current multi-scale features. After obtaining multi-scale features of different MSF block, we use global feature fusion (GFF) to jointly and adaptively learn global hierarchical features in a holistic manner. Finally, compared with other prediction methods in spatial domain, we applied DMSN in UDCT domain, which enables a better representation of global topological structure and local texture detail of HR images. DMSN shows superior performance over other state-of-the-art medical image SR methods.

Index Terms— super-resolution, deep multi-scale network, uniform discrete curvelet transform, local residual learning, global feature fusion

1. INTRODUCTION

In clinical medicine, high-resolution (HR) medical images are visual and effective tools for physicians to make accurate diagnoses. However, acquisition of HR medical images is complicated by many factors. Low-resolution (LR) medical images will badly influence physicians' diagnoses; thus, super-resolution (SR) techniques for medical images [1] have gradually become extremely crucial.

Due to the powerful learning ability, CNN-based methods [2, 3, 4, 5, 6, 7, 8, 9, 10, 11] are widely used to address nature image SR tasks and have achieved impressive results. From the first SR network SRCNN [2] to the latest RCAN [12], the number of convolutional layer increases from 3 to 400, which proves that increasing the network depth can result in better SR results. In addition, the current deep SR network is a series of identical feature extraction blocks (FEB). The ability of each FEB to extract features plays a crucial role in the final SR performance. Based on this consideration, this paper proposes an efficient multi-scale fushion block to effectively exploit features.

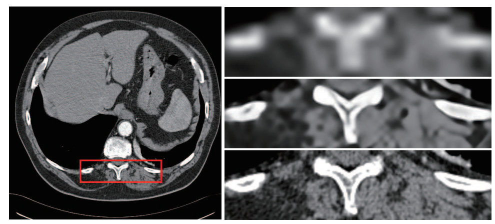


Fig. 1: The performance of our network: the left side is the original image. The right side is the red zone of the LR image (8×), the SR image, and the original image from top to bottom

Due to image transform domain can reserve context and texture information of image at different levels, image SR reconstruction in the transform domain has attracted some attention. As a classic image transformation, wavelet transform has been used for SR in natural and face images. However, the direct extension of a wavelet to 2D by the tensor product of two 1D wavelets is no longer optimal for representing a image that has features along smooth curves. To overcome this limitation, we use uniform discrete curvelet transform (UDCT) for SR in this paper, which is a real 2D image representation tool with multi-scale, multi-directional and anisotropic features. Fig. 1 shows the performance of our network, indicating the super-resolution medical images by the proposed DMSN have abundant details.

This paper makes the following contributions: (1) The difference between natural images and medical imaging gives rise to significant differences in textural detail and edge structure; in light of this, we constructed a database applicable to medical image SR (SRMIdataset) to improve the learning effects of the CNN-based SR method; (2) We proposed a new

This work was supported by the National Natural Science Foundation of China (Nos: 61802212, 61872203, and 61502241). Data used in this publication were generated by the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC).

feature extraction module, multi-scale fushion (MSF) block, to construct our network in cascaded manner. In each MSF block, the features of multiple receptive fields are efficiently exploited through convolution kernels of different sizes. By using of local residual learning (LRL) and global feature fusion (GFF), our DMSN can jointly and adaptively learn hierarchical features in a holistic manner. (3) Existing CNN-based SR methods mostly concentrate on the spatial domain, leading to over-smooth reconstruction results; therefore, UDCT is applied to effectively restore global topology and local edge detail information of HR images; (4) Results showed that DMSN is significantly better than other excellent methods in terms of peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and augmentation of texture detail and edge structure of medical images.

2. RELATED WORK

In recent years, deep learning has aroused widespread interest as a method for overcoming the defects of conventional shallow learning methods. Dong *et al.* [2] pioneered the application of a CNN to image SR and unveiled a super-resolution convolutional neural network (SRCNN), which is significantly better than the output attained using conventional methods. Based on this, many CNN-based super-resolution algorithms have been proposed [4, 1, 3]. The above models share something in common—their network structures have fewer than 10 layers. However, other network models applied to computer vision indicate that depth of network does count in deep learning. As a result, researchers have started to apply deep network models to SR [13, 5, 6, 7, 8]. Recently, many CNN-based SR methods construct the entire SR network by concatenating a series of identical feature extraction blocks [14, 15, 12], indicating the ability of each block plays a key role in the SR performance of the deep network.

The above methods complete image SR in the spatial domain of the image but often generate overly smooth output that loses textural details. By contrast, image SR in the transform domain can preserve the image's context and texture information in different layers to produce better SR results. With that in mind, Guo *et al.* [9] designed a deep wavelet super-resolution (DWSR) network to acquire HR image by predicting "missing details" of wavelet coefficients of the LR image. Later, the same team [10] integrated discrete cosine transformation (DCT) into CNN and put forward an orthogonally regularized deep network (ORDSR). In addition, Huang *et al.* [11] applied wavelet transform to CNN-based face SR to validate that this method can accurately capture global topology information and local textural details of faces.

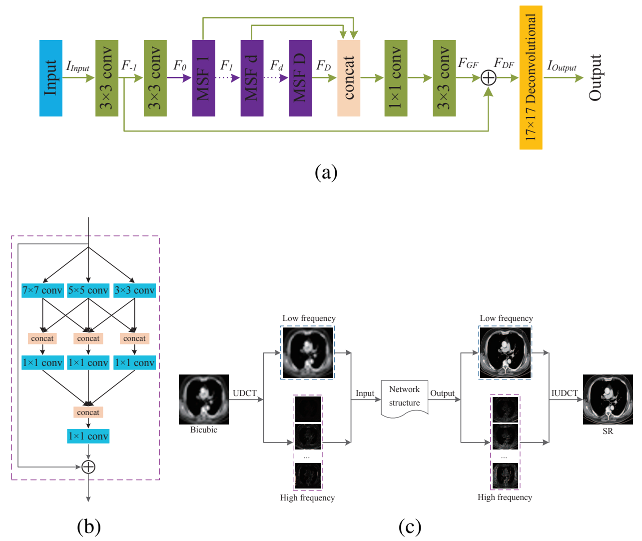


Fig. 2: Network structure: (a) DMSN, (b) MSF block, and (c) UDCT prediction

3. METHOD

3.1. Network Structure

Our network structure is shown in Fig. 2(a). It consists of three parts, the shallow feature extraction module, the multi-scale feature extraction module, and the up-sample module. We solve the following problem:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}^{SR}(F_{\theta}(I_i^{LR}), I_i^{HR}), \tag{1}$$

where $\theta = \{w_1, w_2, w_3, ..., w_m, b_1, b_2, b_3, ..., b_m\}$ stands for the weights and bias of the convolutional layer, N is the number of training samples. $\mathcal{L}^{SR}$ is the loss function for minimizing the difference between I_i^{LR} and I_i^{HR}.

The most widely-used image objective optimization function is the MSE function. However, Lim *et al.* [16] have demonstrated that training with MSE loss is not a good choice. As a better alternative, the MAE loss function can be defined as

$$\mathcal{L}^{SR} = \frac{1}{N} \sum_{i=1}^{N} \left\| I_i^{LR} - I_i^{HR} \right\|_1. \tag{2}$$

Zheng *et al.* [14] empirically found that their model with MSE loss can improve performance of a trained network with MAE loss. In order to avoid introducing unnecessary training tricks and reduce computations, we use the L1 function.

After shallow feature module, we obtain the F_0 and input it into the feature extraction module, which contains a set of cascaded MSF blocks. Here, we use global feature G by fusing features from all the MSF blocks.

$$F_{GF} = H_{GFF}([F_1, \cdots, F_D]), \tag{3}$$

where $[F_1, \cdots, F_D]$ denotes the concatenation of feature-maps produced by MSF blocks $1, \cdots, D$. H_{GFF} is a composite function of 1×1 and 3×3 convolution.

Global residual learning is then utilized to obtain the feature-maps before conducting up-scaling by

$$F_{DF} = F_{-1} + F_{GF}, \tag{4}$$

where F_{-1} represents the shallow feature-maps. All the other layers before GFF are fully utilized with our proposed MSF blocks. Finally, we input F_{DF} to 17×17 deconvolution layer to obtain the output of HR. Except for the deconvolutional layer, the other layers are followed by ReLu.

3.2. MSF block

The proposed MSF block is shown in Fig. 2(b). In each MSF block, we construct a three-bypass network and different bypass use different convolutional kernel. In this way, the information between those bypass can be shared with each other so that able to detect the image features at different scales. The operation can be defined as:

$$C_3 = \sigma(w^1_{3\times3} * F_{d-1} + b^1), \tag{5}$$

$$C_5 = \sigma(w^1_{5\times5} * F_{d-1} + b^2), \tag{6}$$

$$C_7 = \sigma(w^1_{7\times7} * F_{d-1} + b^3), \tag{7}$$

$$H_1 = \sigma(w^2_{1\times1} * [C_3, C_5] + b^4), \tag{8}$$

$$H_2 = \sigma(w^2_{1\times1} * [C_5, C_7] + b^5), \tag{9}$$

$$H_3 = \sigma(w^2_{1\times1} * [C_7, C_3] + b^6), \tag{10}$$

$$F_d = w^3_{1\times1} * [H_1, H_2, H_3] + b^7 + F_{d-1}, \tag{11}$$

where w and b stand for the weights and bias, respectively. F_{d-1} and F_d are the input and output of the d-th MSF block, respectively. $\sigma(x)$ denotes the ReLU function.

3.3. Uniform Discrete Curvelet Transform

Wavelet analysis can not "optimally" represent image functions with straight lines and curves. Curvelet transform is a very effective image representation method, which improves the processing ability of complex lines. Several discrete curvelet and curvelet like transforms have been proposed in the past years, which can be divided into discrete transforms based on the fast Fourier transform (FFT), or based on filter bank (FB) implementations. UDCT [17] is a new discrete curvelet transform that uses the ideas of both FFT-based discrete curvelet transform and filter-bank based contourlet transform, which has excellent frequency response and extremely low redundancy.

As shown in Fig. 2(c), a low-frequency subband and six high-frequency subbands of one-level UDCT are entered in the network structure as "Input". The seven subbands of the SR image are as "Output". Low-frequency subband is applied to effectively global topology while high-frequency subbands capture important structural information. It is worth mentioning that UDCT can be used in different SR networks, which is a simple and effective way to improve the performance. S-peaking of the role of UDCT, it is to take further experiment in Section 4.4. The detailed process of UDCT implementation can be found in [17].

4. EXPERIMENTS

In the experiments, the performance of the proposed DMSN is evaluated on both qualitative and quantitative aspects. PSNR and SSIM are used for quantitative evaluation. The contrasting methods selected in this part—very deep convolutional network (VDSR) [13], deep recursive residual network (DR-RN) [7], deep persistent memory network (MemNet) [8], and information distillation network (IDN) [14]—are all state-of-the-art deep learning SR methods.

4.1. Medical Image Database

Image databases of four body parts, found in The Cancer Imaging Archive (TCIA) [18]—breast, brain, lung and kidney—are integrated to create a database applicable to medical image SR. This database comprises 400 medical images—100 images for each body part. A total of 280 medical images (70 images for each body part) compose a training set; the remaining 120 images compose a test set.

4.2. Implementation Details

Data augmentation is performed on the 280-image training dataset described in Section 4.1. Inspired by [7, 13], the flipped and rotated versions of training images are considered; specifically, we rotate the original images by 90°, 180°, and 270° and flip them horizontally. After that, for each original image, we have seven additional augmented versions. The training images are split into 41×41 patches, with the step of 31, by considering both the training time and storage complexities. Our network contains 18 MSF blocks. The number of feature maps used in all the convolutional layers is 64. The learning rate in initialized to 10^{-4} for all layers and decreases half for every 50 epochs. Training our model takes roughly one day with Tesla P40 GPUs.

4.3. Evaluation of Results

In this section, we evaluate the performance of our method on four databases (*i.e.*, breast, brain, lung, and kidney). PSNR and SSIM [19] are used to measure the image quality. For fair comparison, we use the released codes of the above models and train all models with the same training set. The PSNR and SSIM values for comparison (scale: $4\times$ and $8\times$) are shown in Table 1; values in bold font indicate optimal values. The table shows that when evaluated on four databases, our proposed

Table 1: Comparison of PSNR/SSIM for different methods.

Dataset	scale	Bicubic	VDSR	DRRN	MemNet	IDN	Ours
Breast	4	30.514/0.879	32.053/0.898	32.411/0.905	32.551/0.908	32.482/0.906	**32.743/0.911**
	8	26.736/0.801	28.134/0.821	28.311/0.827	28.456/0.836	28.431/0.833	**28.743/0.844**
Brain	4	32.766/0.907	34.362/0.922	34.795/0.931	34.952/0.935	35.041/0.937	**35.246/0.944**
	8	28.249/0.822	29.221/0.840	29.469/0.849	29.528/0.849	29.549/0.851	**29.913/0.857**
Lung	4	25.053/0.825	29.775/0.868	30.139/0.878	30.192/0.885	30.156/0.881	**30.454/0.899**
	8	22.432/0.737	24.208/0.784	24.508/0.792	24.546/0.801	24.511/0.797	**24.825/0.804**
Kidney	4	28.369/0.848	31.754/0.899	32.146/0.906	32.231/0.914	32.210/0.911	**32.518/0.921**
	8	24.949/0.751	26.257/0.777	26.455/0.796	26.513/0.805	26.412/0.799	**26.891/0.811**

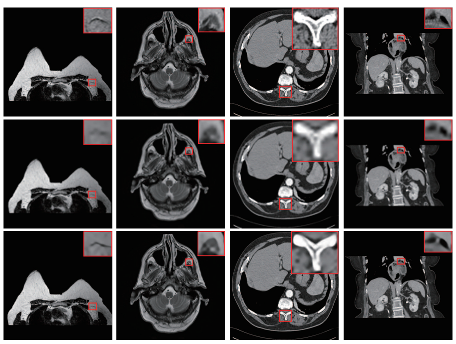


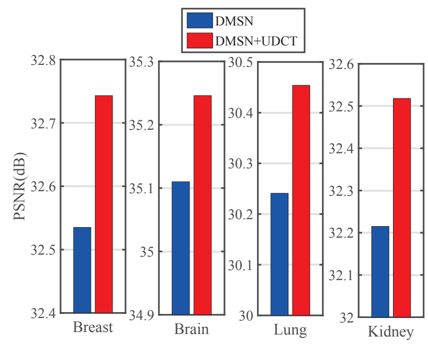
Fig. 3: Qualitative results of our network. The first line is the original images, the second line is the bicubic interpolation images, and the third line is the SR images

DMSN obtains higher PSNR and SSIM on average than other methods.

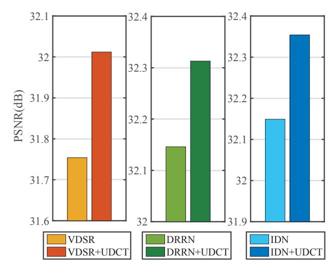
Fig. 3 shows the visual effects on scale 4×. The first line in the figures is the original images, the second line is the bicubic interpolation images, and the third line is the SR images obtained using our network. We see that the images reconstructed by the proposed method have abundant detail, which are very close to the original images. We can thus conclude that the method proposed in this paper can be effectively applied to medical image SR.

4.4. Effectiveness of UDCT

Given that in this paper, we introduce to predict UDCT coefficients in the field of medical image SR, we evaluate the effect of the contribution. We use five methods (VDSR, DRRN, Memnet, IDN, and DMSN) and integrate them with UDCT prediction. Fig. 4(a) shows the comparison results of DMSN across different databases. Fig. 4(b) shows the average comparison result of VDSR, DRRN, and IDN. From Fig. 4, we can see both methods improve significantly when integrated with UDCT. Experimental results demonstrate that UDCT prediction is superior to spatial domain; the improvements are consistent across various networks and benchmarks.



(a) DMSN in different databases



(b) VDSR, DRRN, Memnet, and IDN

Fig. 4: Effectiveness of UDCT prediction

5. CONCLUSION

In this paper, a CNN-based medical image SR method is proposed. Our end-to-end network DMSN contains a set of cascaded MSF blocks, which effectively exploit multi-scale feature to improve the SR performance. In addition, UDCT is applied to the network structure to effectively restore missing structural and edge information in the LR image to further improve the SR performance. Quantitative results show that the proposed method is much better than other state-of-the-art methods, remarkably boosting restoration ability of textural structure and edge details of medical images.

6. REFERENCES

[1] Kensuke Umehara, Junko Ota, and Takayuki Ishida, "Application of super-resolution convolutional neural network for enhancing image resolution in chest ct," *J. Digit. Imaging*, vol. 31, no. 4, pp. 441–450, 2018.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision (ECCV)*, 2014, pp. 184–199.

[3] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1874–1883.

[4] Chao Dong, Chen Change Loy, and Xiaoou Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference on Computer Vision (ECCV)*, 2016, pp. 391–407.

[5] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1637–1645.

[6] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS)*, 2016, pp. 2810–2818.

[7] Ying Tai, Jian Yang, and Xiaoming Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3147–3155.

[8] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu, "Memnet: A persistent memory network for image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4539–4547.

[9] Tiantong Guo, Hojjat Seyed Mousavi, Tiep Huu Vu, and Vishal Monga, "Deep wavelet prediction for image super-resolution," in *The IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 104–113.

[10] Tiantong Guo, Hojjat S Mousavi, and Vishal Monga, "Orthogonally regularized deep networks for image super-resolution," *arXiv preprint arXiv:1802.02018*, 2018.

[11] Huaibo Huang, Ran He, Zhenan Sun, Tieniu Tan, et al., "Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1689–1697.

[12] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, "Image super-resolution using very deep residual channel attention networks," in *European Conference on Computer Vision (ECCV)*. Springer, 2018, pp. 294–310.

[13] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646–1654.

[14] Zheng Hui, Xiumei Wang, and Xinbo Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 723–731.

[15] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu, "Residual dense network for image super-resolution," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[16] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *The IEEE conference on computer vision and pattern recognition (CVPR) workshops*, 2017, pp. 136–144.

[17] Truong T Nguyen and Hervé Chauris, "Uniform discrete curvelet transform," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3618–3634, 2010.

[18] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al., "The cancer imaging archive (tcia): maintaining and operating a public information repository," *J. Digit. Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[19] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.