Single Image Super-Resolution Using a Deep Encoder-Decoder Symmetrical Network with Iterative Back Projection

Heng Liu[1], Jungong Han[2, *], Shudong Hou[1], Ling Shao[3], Yue Ruan[1]

[1] School of Computer Science and Technology, Anhui University of Technology, 243032, China

[2] School of Computing & Communications, Lancaster University, LA1 4YW, U. K.

[3] School of Computing Sciences, University of East Anglia, NR7 Tj4, U. K.

[*] Corresponding author: jungonghan77@gmail.com

Abstract: Image super-resolution (SR) usually refers to reconstructing a high resolution (HR) image from a low resolution (LR) image without losing high frequency details or reducing the image quality. Recently, image SR based on convolutional neural network (SRCNN) was proposed and has received much attention due to its end-to-end mapping simplicity and superior performance. This method, however, only using three convolution layers to learn the mapping from LR to HR, usually converges slowly and leads to the size of output image reducing significantly. To address these issues, in this work, we propose a novel deep encoder-decoder symmetrical neural network (DEDSN) for single image SR. This deep network is fully composed of symmetrical multiple layers of convolution and deconvolution and there is no pooling (down-sampling and up-sampling) operations in the whole network so that image details degradation occurred in traditional convolutional frameworks is prevented. Additionally, in view of the success of the iterative back projection (IBP) algorithm in image SR, we further combine DEDSN with IBP network realization in this work. The new DEDSN-IBP model introduces the down sampling version of the ground truth image and calculates the simulation error as the prior guidance. Experimental results on benchmark data sets demonstrate that the proposed DEDSN model can achieve better performance than SRCNN and the improved DEDSN-IBP outperforms the reported state-of-the-art methods.

Keywords: Single image super-resolution, Deep encoder-decoder, Symmetrical network, Iterative back projection

1. Introduction

With the rapid development of high definition (HD) display device, HR image or video becomes urgently in need. Traditional optical ways to get HR images is to improve image acquisition sensors or optical systems, namely to reduce the size of the imaging unit by adopting high precision imaging chips and other related devices [1]. However, due to the great cost and the hardware limitation, in practice people usually reconstruct HR image from LR one base on image processing way. Therefore, image SR technology naturally appears and holds a huge demand.

Depending on the number of input images, image SR methods can be classified as single image based SR and multiple images based SR. It is obvious that image SR can be realized in the frequency domain or the spatial domain based on classical signal processing techniques. Generally speaking, image SR can be implemented by three means, namely, interpolation based methods, reconstruction based methods and learning based ones. A simple and fast SR method is image interpolation, which is based on the smoothness assumption, including linear, bilinear and bi-cubic interpolation ways. However, the simple smoothness assumption unavoidably results in jaggy and ringing effects due to the discontinuities across image regions. For more effective image SR, more sophisticated prior information can be learned from image correspondences, such as sparsity constraints, self-similarity prior and exemplar prior.

Actually, interpolation and reconstruction based methods are merely signal processing while the learning based ones pay more attention to the understanding of the content and the structure of the image to be reconstructed. Learning based reconstruction methods, which are quite popular nowadays, make full use of the prior related to the image itself to provide strong constraints for better SR performance. Fig. 1 summarizes the taxonomy of image SR methods [1-3].
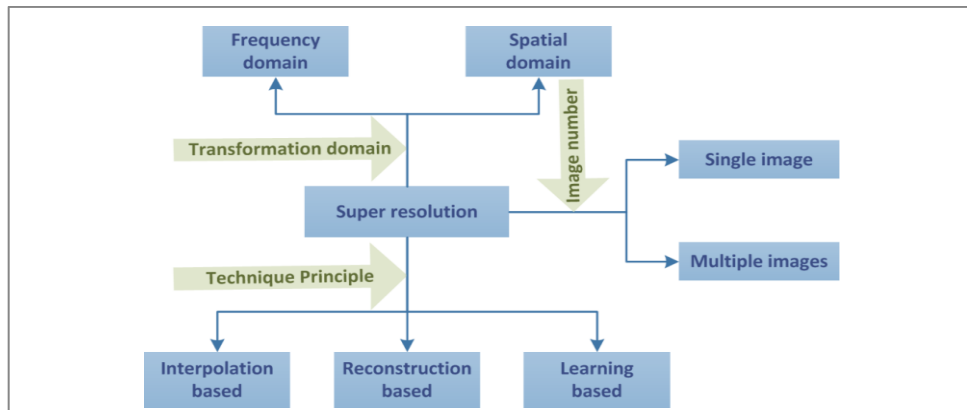

Fig. 1. The taxonomy of image SR methods.

Aiming at obtaining an HR image based on one LR input image, single image SR becomes the most attractive way among all SR solutions. The LR image is usually generated based on a degenerated model, which consists of diverse degradation operations such as geometric deformation, blur, down sampling and additive noise. Super-resolving an LR image is an inverse process of the degradation that manages to recover the missing high-frequency details in the original HR image, such as edges and textures.

Compressive sensing theory was firstly introduced as a sparse representation method for image super-resolution in [4-6]. This approach assumes that HR patches hold the same representation coefficients with LR patches and usually requires a lot of training data to jointly learn a couple of dictionaries, which are utilized for the reconstruction of HR and LR images, respectively. In addition, the image local structures tend to recur within and across different image scales. Therefore, image SR can be realized based on self-similar examples instead of external data sets, some recent works of which can be found in [12-14]. Similar to the sparse representation based methods, exemplar images based SR ones also require large numbers of training data to learn the mapping relationship. It should be noted that for each scale down-sampling, all these image SR methods require to retrain their learning models [11, 17] at corresponding scale.

Recently, deep learning, especially convolutional neural network (CNN), achieves great progress in many low-level computer vision tasks, such as image dehazing [31], image denoising [22, 30] and image SR [8, 19, 29], etc. Benefitting from

powerful end-to-end mapping learning, image SR based on convolutional neural network (SRCNN [7, 8]) obtains significant improvement. Moreover, many other deep learning based image SR methods have also demonstrated their good performance in [9, 18].

In this work, our main contributions are summarized as follows:

- We propose a deep encoder-decoder symmetrical network (DEDSN) for image SR. The network only consists of convolution and deconvolution layers. The convolution and deconvolution operations are purely symmetrical, which can ensure that the reconstructed image does not discard useful image details and can avoid the size reduction of the output image with respect to the input image.

- We combine DEDSN with the iterative back projection (IBP) algorithm to form a new model, named DEDSN-IBP. Each step of the IBP algorithm is converted to a specific layer in the new network, which is trained as a part of the whole network. The simulation error between the reconstructed LR image and the down-sampling of the ground truth image is iteratively utilized to update the final reconstructed HR image. The merit of this new model is the continuous prior introduction of the ground truth image in the SR procedure.

The remainder of this work is organized as follows. Section 2 provides an overview of the related work. Section 3 proposes the deep encoder-decoder symmetric network (DEDSN) in detail. Section 4 describes the improved DEDSN-IBP compositional model. Section 5 presents extensive experimental results with comparative discussions. Finally, Section 6 concludes the paper.

2. Related Work

In [4-6], Yang et al. propose a sparse coding based super-resolution (ScSR) approach. Given the LR/HR training patch pairs, ScSR can learn a couple of HR and LR dictionaries with the assumption that LR and HR patch pairs share the same sparse coding coefficients. Firstly, the input LR image is divided into overlapped patches and then the coefficients can be obtained through the learned LR dictionary. Then, the required HR patches can be estimated by the HR dictionary with acquired coding coefficients. Due to the division of overlapped patches, the sparse representation based SR method ignores the consistency of pixels in overlapped patches, which is a strong constraint for image reconstruction. Taking note of this problem, Gu et al. [18] raise a convolutional sparse coding (CSC) based method for image super-resolution. CSC first decomposes the input image into sparse LR feature maps by the learned LR filters to avoid dividing overlapped patches. Then the LR feature maps are mapped onto HR feature maps through the mapping function. Finally, the HR image can be estimated based on the convolution of HR feature maps with learned HR filters.

In addition, by combining sparse representation with non-local similarity, Li et al. [35] present a self-learning image SR method, which aims to preserve the image details while keeping the lower computation load. For multi-frame SR, a double sparsity structure [33] has been recently proposed to overcome the separation problem of block matching and sparse coding in traditional procedure.

The pioneer work of deep learning based image SR was introduced by Dong et al in [7, 8]. In their work, SRCNN is designed to directly learn the non-linear mapping between LR and HR images, which consists of three convolution layers. SRCNN implements patch extraction and representation, non-linear mapping and image reconstruction, to simulate the processing of a sparse coding procedure to generate HR images. However, SRCNN ignores the image prior, which is a significant component for image recovery. Liang et al. [34] introduce Sobel edge prior so as to capture gradient information to accelerate the training convergence. In fact, the method does reduce the training time but the resultant reconstruction enhancement is limited. Wang et al. [9] directly extended the conventional sparse coding with the merits of deep learning. Based on the learned iterative shrinkage and thresholding algorithm, a feed-forward neural network is implemented. To exploit both external and self-similarities, Wang et al. [19] also produced a deep joint super-resolution model which contains complex fine-tuning operations and is not end-to-end.

Recently, encoder-decoder deep network consisting of convolution and deconvolution operators has been applied to semantic segmentation [20, 21] and image denoising [22]. However, whether and how such an encoder-decoder network can be applied to image SR needs to be investigated.

To the best of our knowledge, iterative back projection (IBP) algorithm, proposed by Irani et al. in [16], is usually utilized as a post processing step in many reconstruction based image SR methods. First, an initial estimate of an HR image is regarded as an intermediate result. Then the result is mapped onto the actual LR image to acquire an LR simulation image using the degradation operation. Then the difference between the LR simulation image and the observed LR image is calculated as the simulation error, which can be used for HR image updating. Such processing is executed iteratively to get the final result. IBP processing can be described as the following equations:

$$y^0 = W_k x^0 + n_k, \tag{1}$$

$$x^1 = x^0 + H^{BP}(y - y^0), \tag{2}$$

where x^0 and x^1 denote the initial estimation and the updated estimation of HR images respectively; y^0 represents the LR simulation acquired by the degradation operation; y is the observed LR image; and H^{BP} represents the back projection operator.

IBP may be combined with most SR approaches for reconstruction enhancement. For example, works in [4, 18] apply IBP updating result as the HR image of initial estimation x^0 in Eq. (1). In this work, we integrate the IBP algorithm into the proposed

DEDSN model to form a new network so as to improve single image SR performance. The flow chart of the proposed compositional model can be viewed in Fig. 3.

3. Deep Encoder-decoder Symmetric Network (DEDSN)

Unlike some other deep learning models – early stacked auto-encoder (SAD) [10] or recent coupled auto-encoder (CAD) [29] which were designed for unsupervised feature learning or intrinsic representation learning, here the encoder means the convolution operation and the decoder means deconvolution operation. Each encoder in the encoder network performs convolution with certain kernels (a filter bank) to produce a set of feature maps whereas the role of the decoder network is to map the low resolution feature maps to high resolution feature maps. Actually, the encoders acquire the feature maps of the input image and acts as feature extractor, which encode the primary image content with noise decreased. Then based on these encoded feature maps the following decoder layers are utilized to restore the image content and recover the image details gradually. Naturally, since the symmetrical configuration is applied to the encoder-decoder network, the sensitivity limitation of the context in image blocks in classical CNN models (including SRCNN) will be overcome. Therefore, motivated by encoding-decoding principles, we consider to utilize the convolution and deconvolution symmetrical network for image SR. The overall architecture of the proposed DEDSN model is shown in Fig. 2.
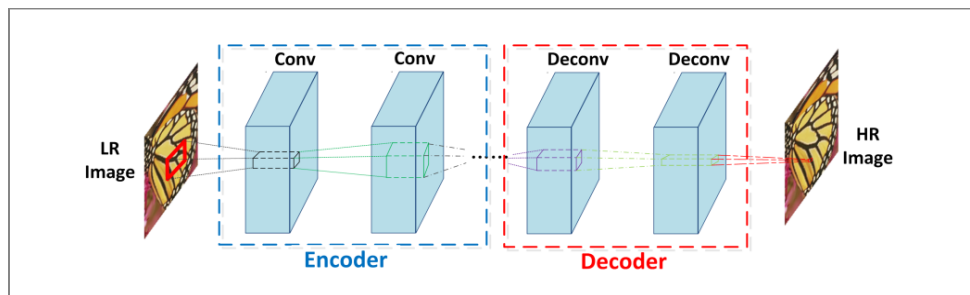

Fig. 2. Overall architecture of the proposed DEDSN model. The network consists of symmetrical convolution and deconvolution layers (12 layers in total).

3.1. DEDSN Architecture

As displayed in Fig. 2, the proposed DEDSN model is made up of multiple convolution layers (encoder) and symmetrical deconvolution layers (decoder), in which convolution operation serves as feature extractor and deconvolution layers are used to recover the image content details. Behind each convolution and deconvolution layer, instead of ReLU, parametric rectified linear

unit (PReLU) is added to enrich nonlinearity of the model. Compared with convolution layers, in which multiple inputs within a filter window are integrated into a single output, deconvolution layers unite a single input with multiple outputs.

Recent literature shows that the convolution and deconvolution structure has been successfully exploited in image segmentation [23], semantic segmentation [20, 21] and object detection [24]. In these works, convolution and deconvolution were served as encoder and decoder separately. For image restoration, only using the convolution operator will be difficult to remember all contents of an image due to much pixel correlation information or details could be lost in the feed-forward process of feature extraction or encoding. In our work, the cascading deconvolution layers will learn to recover the image details based on the feature maps of convolution layers. We use neither pooling nor unpooling in the network as no matter what pooling operation is applied, it will discard useful image details. The network takes an LR image as input and predicts image details by going across convolution and deconvolution layers, and then outputs an HR image finally.

Unlike existing not too deep networks, deeper neural networks can possess more layers which result in acquiring more image context related clues that will be useful for fine reconstruction. This behaviour is equivalent to the use of more and gradually increasing receptive fields to process images. This may be why a deeper network give superior performance. Therefore, we use a deeper model (12 vs. 3 [8]) in our DEDSN architecture for image SR. The detailed configuration of our network can be found in Section 5.

3.2. Highlights

Encode-decoder model has been applied to semantic segmentation [20, 21] and object detection [24], however, there are some obvious differences between these models and the proposed DEDSN architecture. For example, in the semantic segmentation network, a dense pixel-wise class probability map is obtained by successive pooling, deconvolution and non-linear activation, whereas our DEDSN model does not introduce any form pooling operations. Our model does not adopt pooling layer because pooling operation contributes to acquiring image abstractions but the downside is that the image details are discarded. The second highlight of the proposed model is that the full convolutional and deconvolutional symmetrical structure can avoid the size reduction of the output image, which is usually occurred in asymmetric convolutional networks. Actually, this highlight indicates that the proposed DEDSN model can work for multiple scales down-samples image SR.

3.3. Model Training

For image SR, we expect to obtain the mapping relationship between HR output and LR input from training data. Given a training dataset $\{x^{(i)}, y^{(i)}\}_{i=1}^{N}$, our goal can be depicted by the following equation.

$$y = F(x, \Phi), \tag{3}$$

where x, y denote the LR image and the estimated HR image separately, and Φ represents the learned parameter. Given an LR image x and the corresponding HR image x^H, we expect to minimize the difference between y and x^H. Formally, this operation can be expressed as follows:

$$\min_{\phi} \frac{1}{N} \sum_{i=1}^{N} \|F(x_i, \Phi) - x_i^H\|, \tag{4}$$

where N denotes the number of samples in the dataset. Then we train the network using the Caffe package [26] on a workstation with 2 Intel Xeon E5 CPUs and 1 Titan X GPU. During training, we use stochastic gradient descend (SGD) to update the network parameters.

4. DEDSN-IBP Compositional Model

4.1. Motivation

Despite rather good results have been achieved by using DEDSN, there is still further space for reconstruction improvement. One improvement motivation is to impose image prior in the image reconstruction process. In our opinion, iterative back projection (IBP) can continually introduce the ground truth image prior by producing the LR simulation image and calculating the simulation error, which can be used to improve the reconstruction effect by updating the estimated HR image iteratively. Thus, we implement IBP through neural network operations, and make them integrated into convolutional neural network to form a new network model, termed DEDSN-IBP. The following Fig. 3 gives a diagram of the proposed compositional model.
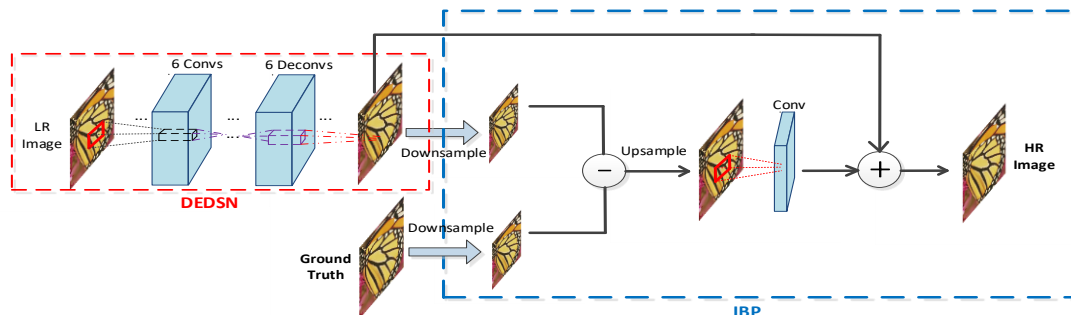
Fig. 3. The compositional network model of DEDSN with IBP.

4.2. Architecture of DEDSN-IBP

Based on DEDSN model and motivated by above IBP principle, the proposed DEDSN-IBP network consists of 16 layers including convolutional part, deconvolution part, two sub-sampling layers (one for down-sampling and the other for up-sampling), one difference layer, and one update (convolution) layer. The specific functions of these layers are described in detail as follows:

(1) Convolutional part. The first six layers of the model are convolutional part which is the same as those in the proposed DEDSN. These layers all take the form of convolution and then they can extract different scales image features by cascading multiple kernels filtering (convolutions) gradually.

(2) Deconvolutional part. The second six layers of the model are deconvolution part to recover the image details gradually with cascading different kernels backward filtering (deconvolutions) based on the extract feature maps of the image.

(3) Output layer of image SR. This layer performs a simple convolution operation to get the SR reconstruction result based on the deconvolution results which holds abundant image details.

(4) Down-sampling layer. This layer down-samples the image derived from the output layer of DEDSN. As a result, an LR simulation version of the super-resolved image is acquired.

(5) Difference layer. This layer calculates the difference between the down-sampled version of the original HR image and the corresponding acquired LR simulation version. The difference may be treated as the simulation error and will be introduced as a prior guidance for image SR.

(6) Up-sampling layer. This layer up-samples the simulation error to generate the simulation error of the HR version.

(7) Update (convolution) layer. This layer performs a convolution operation on the above simulation error. Then the final SR image is updated based on the synthesis of the output of DEDSN and the convolution result of the simulation error.

Actually, IBP can also be applied with SRCNN to get better performance. Additional experimental results demonstrate that the SRCNN-IBP model surpasses the performance of SRCNN because of the introduction of the ground truth image prior. However, compared to the proposed DEDSN-IBP model, there are still some obvious limitations for the SRCNN-IBP. First, the depth of the SRCNN-IBP network is not deep enough for extracting more useful image information. Although Dong et al. [8] argued that deeper models do not result in better performance, recent works [23, 24] reveal that network depth is of crucial importance and appropriate deeper models can usually lead to a remarkable performance improvement. This may be because that a deeper network can extract

more abstract and abundant image features. Second, the pooling layer in the SRCNN-IBP model may lead some image details to be lost. In image restoration problems, such as image SR, image details are sensitive and significant for accurate image reconstruction.

5. Experimental Results

In this section, we first provide implementation details of our experiments. Then, we compare the proposed method with classical and recent state-of-the-art methods, including bi-cubic interpolation, ScSR [4], and SRCNN [8]. For evaluation, we use 'Set5' [27] and 'Set14' [28] as test datasets. They consist of 5 and 14 natural images separately, which are often used for benchmarking in current image SR works.

5.1. Implementation Details

We use 91 images [2] for model training. In our work, training is only performed on a single channel. For color images, we first convert the original images into YCbCr colour space and then extract features from the luminance component Y. Each image was cropped into small patches with size 50 x 50 (larger than 33 x 33 in SRCNN [8]). The cropped ground truth patches were regarded as the HR patches, namely the label in experiments, and the corresponding LR pairs can be obtained by bicubic interpolation. The goal of DEDSN is to learn the mappings from LR to HR by minimizing the Euclidian loss function between the reconstructed HR image and the ground truth. In [8], all the convolution layers will produce a smaller size output compared to the input. They evaluate the difference between the central pixels of network output and the original label. While we get an output image whose size is the same as the input image due to the deconvolution introduction which avoids the loss calculation only depending on the central parts of the super-resolved and the label images. Since the network is symmetrical, the proposed DEDSN and the DEDSN-IBP models can be applied on arbitrary sized images reconstruction though a fixed image size is always used in training.

The specific configuration of our model can be referred in Table 1. All the layers' parameters are initialized with the "Xavier" strategy. In addition, in our experiments the learning rate is set as 10^{-5} and a batch size of 128 is used in the training protocol. Momentum and weight decay parameters are set to 0.9 and 0.0001, respectively.

Table 1. Configuration of our two networks. "conv3" and "deconv3" stand for convolution and deconvolution kernels of size 3 x 3, similar for others. 1, 32, 64 and 128 are the number of feature maps after each convolution and deconvolution.

DEDSN-IBP Model			DEDSN Model
data			Data
(conv1-1) x 1	DEDSN		(conv3-32) x 1
			(conv3-64) x 2
			(conv3-128) x 3
	down-sampling	down-sampling	(deconv3-128) x 3
eltwise(difference)			(deconv3-64) x 2
up-sampling			(deconv3-1)
conv3-1(update)			
eltwise(sum)			

Some feature maps extracted from different convolution and deconvolution layers are shown in Fig. 4 separately. It is easy to see that most feature maps are prone to zero.
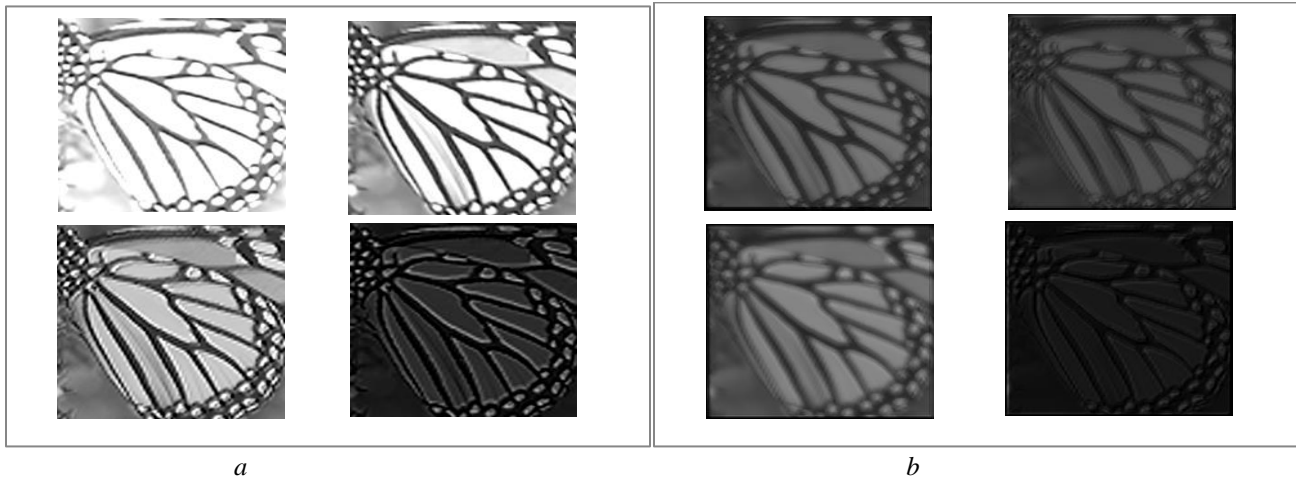

a   *b*

Fig. 4. Several feature maps examples of convolution and deconvolution layers**.** a. feature maps extracted from the third convolution layer; b. feature maps extracted from the fourth deconvolution layer.

5.2. Quality Evaluation

We evaluate and compare the performance of our proposed two models in terms of peak signal to noise ratio (PSNR) and structure similarity (SSIM) [32]. The models are tested on Set5 and Set14, which are commonly adopted in the evaluation of

existing SR algorithms. The visual qualities of the SR results generated by our method and other competing ones are depicted in Fig. 5.
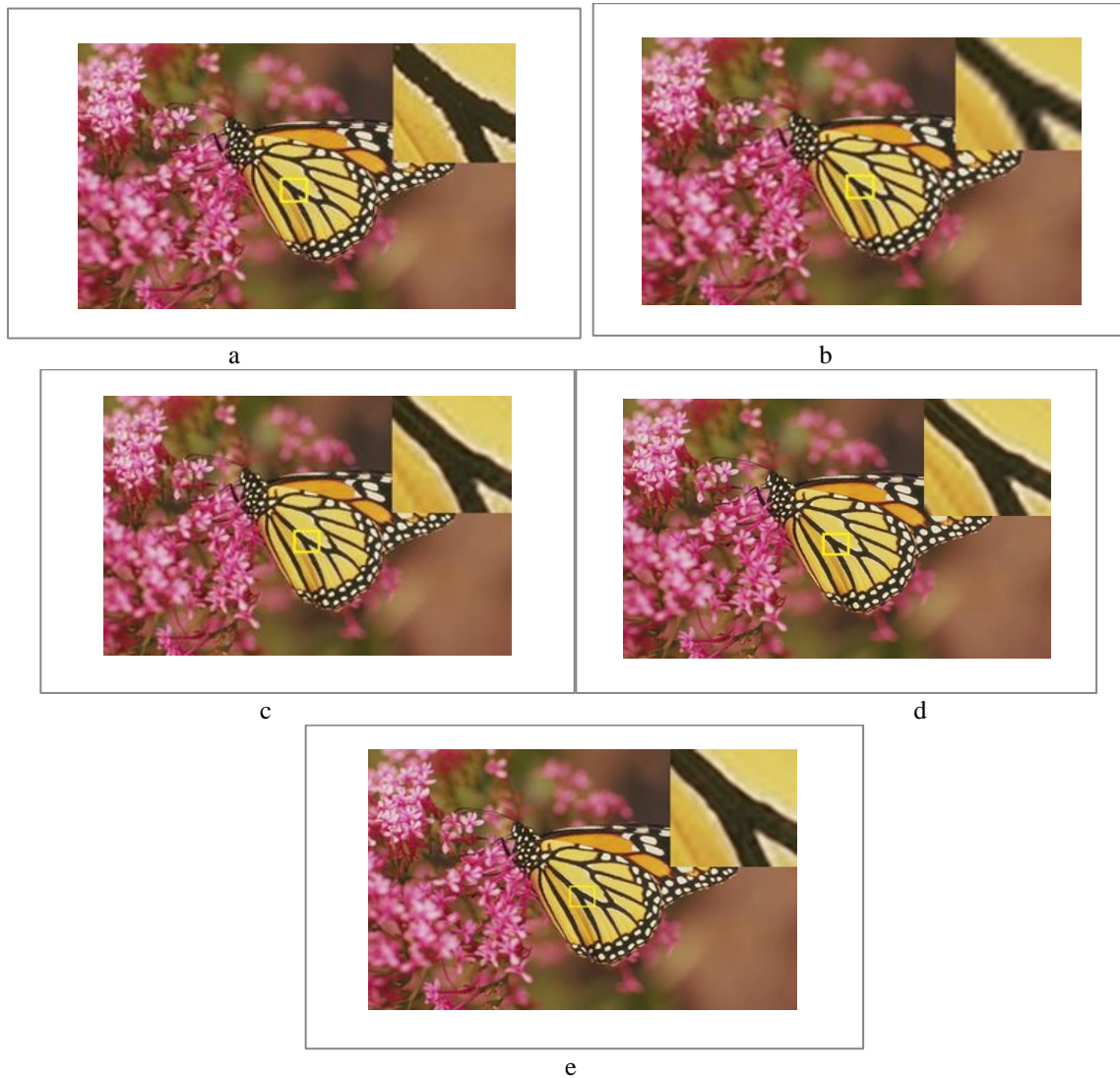

Fig. 5. Comparison of PSNR on different image SR methods with our models. a. Original Image; b. Bicubic: 29.49dB; c. SRCNN: 32.51dB; d. SRCNN-IBP: 32.52dB; e DEDSN: 32.85dB.

Table 2 and Table 3 show the comparison results (the super-resolution factor is 3) in terms of PSNR and SSIM between the proposed method and the competing methods. The best results are highlighted in bold. As one can see, our proposed method improves the results in terms of both PSNR and SSIM. We can draw some conclusions from the results.

Table 2. The result of PSNR (dB) on test images using different methods.

Images	Bicubic	ScSR	SRCNN	DEDSN	DEDSN-IBP
butterfly	24.05	25.69	27.51	27.89	**27.93**
baby	33.95	34.30	35.03	35.04	**35.10**
bird	32.49	34.11	34.74	**35.14**	**35.14**
head	32.91	33.14	33.55	33.70	**33.78**
comic	23.16	23.94	24.43	24.69	**24.75**
baboon	23.01	**23.47**	23.35	23.42	**23.44**
ppt3	23.90	25.15	26.24	26.52	**26.60**
Lena	31.73	32.66	33.40	33.65	**33.71**
zebra	26.73	28.24	28.93	**28.95**	28.90
average	27.99	28.96	29.69	29.89	**29.93**

Table 3. The result of SSIM on test images using different methods.

Images	Bicubic	ScSR	SRCNN	DEDSN	DEDSN-IBP
butterfly	0.787	0.836	0.886	0.894	**0.910**
baby	0.888	0.881	0.908	0.908	**0.913**
bird	0.914	0.913	0.941	0.947	**0.951**
head	0.806	0.725	0.826	0.831	**0.838**
comic	0.689	0.736	0.768	0.778	**0.788**
baboon	0.619	0.520	0.663	0.665	**0.669**
ppt3	0.861	0.881	0.912	**0.913**	0.911
Lena	0.846	0.789	0.872	0.875	**0.877**
zebra	0.782	0.814	0.837	**0.839**	0.837
average	0.799	0.788	0.846	0.850	**0.861**

Frome these two tables, it is obvious that the symmetrical encoder-decoder with IBP model achieves best reconstruction results than the competing methods. These experimental results and comparisons validate that symmetrical convolution and deconvolution operations do work well for image SR and iterative back projection (IBP) can actually improve the performance of DEDSN. The results also indicate that the deeper network may contribute to the finer image details recovery and more hierarchal features extraction. Through the proposed architecture, we consider that the deeper the network the better SR performance it will achieve.

6. Conclusions

In this work, in order to overcome the pooling limitation and the size reduction of the output image in traditional CNN, we firstly proposed a deep encoder-decoder symmetrical network (DEDSN) for image SR. With symmetrically combining convolution and deconvolution operations, this model achieves image SR with better performance than SRCNN and other state-of-the art methods by extracting hierarchal image features and recovering image details gradually. In addition, considering the merit of iterative back projection (IBP) algorithm, we further propose a compositional model which combines DEDSN with the network realization of IBP. Based on the introduced image prior, the DEDSN-IBP model achieves the best performance compared with other competing methods.

In future work, more image priors, such as edges or image phase information, will be introduced into our SR model to enrich and maintain the structure features and the details recovering in super-resolved images. In addition, based on some domain adaptation and domain consistency machine learning techniques [36, 37, 38, 39], the simultaneous image super-resolution and classification with different specific image domains will be our another research perspective. Moreover, applying the proposed networks to different applications, such as visual saliency detection [40,41], image hashing [42,43] and action recognition [44,45] is also worth studying.

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] N. Kamal, B. Thomas, Super-resolution: a comprehensive survey, Machine Vision and Applications, 25 (6) (2014) , pp. 1423-1468.

[2] H. Su, J. Zhou, Z. Zhang, Survey of Super-resolution Image Reconstruction Methods, Acta Automatica Sinica, 39 (8) (2013), pp. 1202-1213.

[3] J. Yang, T. Huang, Image super-resolution: Historical overview and future challenges, in: Super-resolution imaging, 2010, pp. 1-34.

[4] J. Yang, J. Wright, Z. Lin, and S. Cohen, Coupled dictionary training for image super-resolution, IEEE Trans. Image Processing, 21 (8) (2012), pp. 3467-3478.

[5] J. Yang, J. Wright, T. Huang, and Y. Ma, Image super-resolution as sparse representation of raw image patches, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008. pp. 1-8.

[6] J. Yang, J. Wright, T. Huang, and Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Processing, 19 (11) (2010), pp. 2861-2873.

[7] C. Dong, C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: Proceedings of Europe Conference on Computer Vision, 2014, pp.184-199.

[8] C. Dong, C. Loy, K. He, and X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Analysis and Machine Intelligence, 38 (2) (2016), pp. 295-307.

[9] Z. Wang, D. Liu, J. Yang, and T. Huang, Deep Networks for Image Super-Resolution with Sparse Prior, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp.370-378.

[10] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzago, Extracting and composing robust features with denoising autoencoders, in: Proceedings of International Conference on Machine Learning, 2008, pp. 1096-1103.

[11] R. Timofte, V. De, L. Gool, A+: Adjusted anchored neighborhood regression for fast super-resolution, in: Proceedings of Asia Conference on Computer Vision, 2014, pp. 111-126.

[12] C. Yang, J. Huang, M. Yang, Exploiting self-similarities for single frame super-resolution, in: Proceedings of Asia Conference on Computer Vision, 2010, pp. 497-510.

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

[14] Y. Zhu, Y. Zhang, Y. Yuille, Single image super-resolution using deformable patches, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2917-2924.

[15] A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet classification with deep convolutional neural networks, in: Proceedings of Confernce on Neural Information Processing Systems, 2012, pp. 1097-1105.

[16] M. Irani, S. Peleg, Improving resolution by image registration, CVGIP: Graphical Models and Image Processing, 53 (3) (1991), pp.231-239.

[17] K. Zhang, B. Wang, W. Zuo, H. Zhang, L. Zhang, Joint Learning of Multiple Regressors for Single Image Super-Resolution, IEEE Signal Processing Letters, 23 (1) (2016), pp.102-106.

[18] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, Convolutional Sparse Coding for Image Super-Resolution, in: Proceedings of International Conference on Computer Vision, 2015, pp. 1823-1831.

[19] Z. Wang, Y. Yang, Z. Wang, S. Chang, et al., Self-Tuned Deep Super Resolution, in: Proceedings of Computer vision and Pattern Recognition Workshop on Deep Vision, 2015, pp. 1-8.

[20] S. Hong, H. Noh, B. Han, Decoupled Deep Neural Network for Semi-supervised Semantic Segmentation, in: Proceedings of Conference on Neural information Processing System, 2015, pp. 1495-1503.

[21] H. Noh, S. Hong, B. Han, Learning Deconvolution Network for Semantic Segmentation, in: Proceedings of International Conference on Computer Vision, 2015, pp. 1520-1528.

[22] X. Mao, C. Shen, Y. Yang, Image Denoising Using Very Deep Fully Convolutional Encoder-Decoder Networks with Symmetric Skip Connections, arXiv preprint, arXiv: 1603.09056, 2016

[23] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", arXiv preprint, arXiv: 1602.07360, 2016.

[24] J. Yang, B. Price, S. Cohen, H. Lee, and M.-Y., Yang, Object Contour Detection with a Fully Convolutional Encoder-Decoder Network, arXiv preprint, arXiv: 1603.04530, 2016.

[25] J. Long, E. Shelhamer, T. Darrell, Fully Convolutional Networks for Semantic Segmentation, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431-3440.

[26] Y. Jia, E. Shelhamer, J. Donahue, etc., Caffe: Convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM international conference on Multimedia, 2014, pp. 675-678.

[27] M. Bevilacqua, A.Roumy, C. Guillemot, M. A. Morel, Low-complexity single-image super-resolution based on nonnegative neighbour embedding, in: British Machine Vision Conference, 2012.

[28] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, in: Proceedings of International Conference on Curves and Surfaces, 2012, pp. 711-730.

[29] K. Zeng, J. Yu, R. Wang, C. Li, D. Tao. Coupled Deep Autoencoder for Single Image Super-Resolution, IEEE Transactions on Cybernetics, 47 (1) (2017): 27-37.

[30] Y. Wu, H. Zhao, L. Zhang, Image Denoising with Rectified Linear Units, in: Proceedings of International Conference on Neural Information Processing, 2014, pp.142-149.

[31] B. Cai, X. Xu, K. Jia and D. Tao, DehazeNet: An End-to-End System for Single Image Haze Removal, arXiv preprint, arXiv: 1601.071602, 2016.

[32] Wang Z, Bovik A, Sheikh H, Simoncelli E. Image Quality Assessment: From Error Visibility to Structural Similarity, IEEE Transactions on Image Processing, 13(4) (2004), pp. 600-612.

[33] Kato Toshiyuki, Hideitsu Hino, and Noboru Murata, Double sparsity for multi-frame super resolution, Neurocomputing, 240 (2017), pp. 115-126.

[34] Y. Liang, J. Wang, S. Zhou, Y. Gong, and N. Zheng, Incorporating image priors with deep convolutional neural networks for image super-resolution, Neurocomputing, 194 (2016). pp. 340–347.

[35] J. Li, J. Wu, H. Deng, J. Liu, A self-learning image super-resolution method via sparse representation and non-local similarity, Neurocomputing, 184 (2016), pp.196-206.

[36] L. Zhang, D. Zhang, Robust visual knowledge transfer via extreme learning machine-based domain adaptation, IEEE Transactions on Image Processing, 25(10) (2016), pp. 4959-4973.

[37] L. Zhang, J. Yang J, D Zhang, Domain class consistency based transfer learning for image classification across domains, Information Sciences, 418(2017), pp. 242-257.

[38] L. Zhang, D. Zhang, Evolutionary Cost-sensitive Extreme Learning Machine, IEEE Transactions on Neural Networks and Learning Systems, 99 (2016), pp.1-16.

[39] L. Zhang, D. Zhang, Visual Understanding via Multi-Feature Shared Learning With Global Consistency, IEEE Transactions on Multimedia, 18 (2) (2016), pp. 247-259.

[40] D. Zhang, D. Meng, J. Han, Co-saliency detection via a self-paced multiple-instance learning framework, IEEE Trans. on Pattern Analysis and Machine Intelligence, 39(5) (2017), pp. 865-878.

[41] J. Han, G. Cheng, Z. Li, D. Zhang, A unified metric learning-based framework for co-saliency detection, IEEE Trans. on Circuits and Systems on Video Technology, DOI: 10.1109/TCSVT.2017.2706264.

[42] Z. Lin, G. Ding, J. Han, J. Wang, Cross-view retrieval via Probability-based semantics-preserving hashing, IEEE Trans. on Cybernetics, 47(12) (2017), pp. 4342-4355.

[43] Y. Guo, G. Ding, L. Liu, J. Han, L. Shao, Learning to hash with optimized anchor embedding for scalable retrieval, IEEE Trans. on Image Processing 26(3) (2017), pp. 1344-1354.

[44] B. Zhang, Y. Yang, C. Chen, L. Yang, J. Han, L. Shao, Action recognition using 3D histograms of texture and a multi-class boosting classifier, IEEE Trans on Image Processing 26 (11) (2017), pp. 4648-4660.

[45] J. Han, L. Shao, D. Xu, J. Shotton, Enhanced computer vision with Microsoft Kinect sensor: A review. IEEE Trans on Cybernetics 43(5) (2013), pp. 1318-1334.