



Accurate image super-resolution using dense connections and dimension reduction network

Shiyu Hu¹ · Guodong Wang¹ · Yanjie Wang¹ · Chenglizhao Chen¹ · Zhenkuan Pan¹

Received: 30 July 2018 / Revised: 7 August 2019 / Accepted: 13 September 2019 /

Published online: 1 November 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Recently, the convolution neural network (CNN) has achieved significant performance improvements toward the image Super-Resolution (SR) problems. Yet, the existing benchmark arts exist multiple limitations, including make use of the feature information deficiently, accompany with the gradient disappearance phenomenon and have serious time consumption. The paper utilizes a newly designed fully convolutional neural network named Accurate Image Super-resolution Using Dense Connections and Dimension Reduction Network (**DCDRN**) to fully exploit the image features. Contextual information of image regions utilizes efficiently and accurately through uniting dense connections and cascading small filters multiple times. And such implementation can be regarded as feature extractors to fuse local and global image features. We newly introduce 1×1 CNNs parallelization structure in the image reconstruction section to reduce data dimensions of the previous layers, which alleviates the computational burden effectively while avoiding the context info losing. The calculation becomes more complex and the convergence becomes slower during training because of the pre-processed images. The proposed **DCDRN** invents a simple and effective method which processes the original image directly and the optimization of layers and filters of CNNs shorten the cost of training significantly. Experiments on benchmark datasets with different methods show that **DCDRN** achieves gratifying performance against state-of-the-art methods. Code is available at <https://github.com/doctorwgd/DCDRN>.

Keywords Super-resolution · Dense connections · Parallelization · Contextual information

1 Introduction

Single image Super-Resolution (SISR) method has been widely applied in multiple computer vision related tasks including image compression [14, 25, 40], visual tracking [24, 30, 37],

✉ Guodong Wang
doctorwgd@gmail.com

¹ College of Computer Science and Technology, Qingdao University, Qingdao 266071, People's Republic of China

intelligent surveillance [23, 26, 41]. On account of the variety and uncertainty of the mapping make the recovering of high-resolution (HR) image from low resolution image diversely. Large scale-factor creates redundant details and has a challenging to recover local features. The obstinate information redundancy also causes the absences of high frequency details. Interpolation method [22] is implementable but the ideal high-resolution output of linear model usually unavailable. Sparse-based techniques [13, 39] have been applied to enhance image priors assume that any natural image patch could be sparsely represented by multiple atomic dictionaries. And the dictionaries are frequently generated from the given image patch pool database, not to mention its heavy computation costs which brought by the conventional OMP (orthogonal matching pursuit) reconstruction procedure. Similar to the sparse representation based methods, other techniques used to simulate the mapping from LR to HR similarly, includes neighbor embedding [7], random forest [32].

In this paper, our approach develops dense connections inspired by DenseNet [13]. Dense connections have following advantages: solve the phenomenon of gradient disappearance effectively, promote the feature propagation and reduce the number of parameters. Dense connections of each layer obtain more contextual information and model the complex functions of the nonlinear layer while enhances the efficiency of gradient transmission in the network and the utilization of contextual information. Further, employing a reasonable number of dense connections in a very deep convolution neural network (CNN) [35] can potentially increase reconstruction performance. These will be discussed later in Section 3.2. Large dimension formatted since all the feature outputs are concatenated to the input layer of the reconstructed network. Therefore, 1×1 CNNs has employed to reduce dimension before generating HR pixels. It also should be noted that our experiment results conduct an opposite conclusion towards the common agreement, i.e., there exists a positive relationship between the network depth and the performance improvements.

1.1 Contribution

Different from previous works, the newly introduced dense connections integrate low-level features into high-level features to provide rich information. Using the proposed structure, large image region contextual information can be utilized accurately.

In the image reconstruction network, the 1×1 CNNs parallelization structure reduces the dimension of the output from previous layers to expedite the operation. 1×1 CNNs organize information across channels while increase the ability of expressing the model. The direct processing of the original image reduces the computational complexity. Optimized the number of layers and filters of CNN to reduce computational costs is essential. The method has been evaluated on three publicly available benchmark datasets and outperformed the most advanced methods currently available. In Fig. 1, we compared the PSNR and running times of various methods. As shown in the figure, the **DCDRN** has optimal performance than state-of-the-art methods.

2 Related work

Prior to the advent of convolutional neural networks, mainstream methods for reconstructing high-resolution images were dominated by matrix operations that traversed dictionary atoms for fast super-resolution. Representative work is A+ [36] and SelfEx [11]. The RFL [33] then

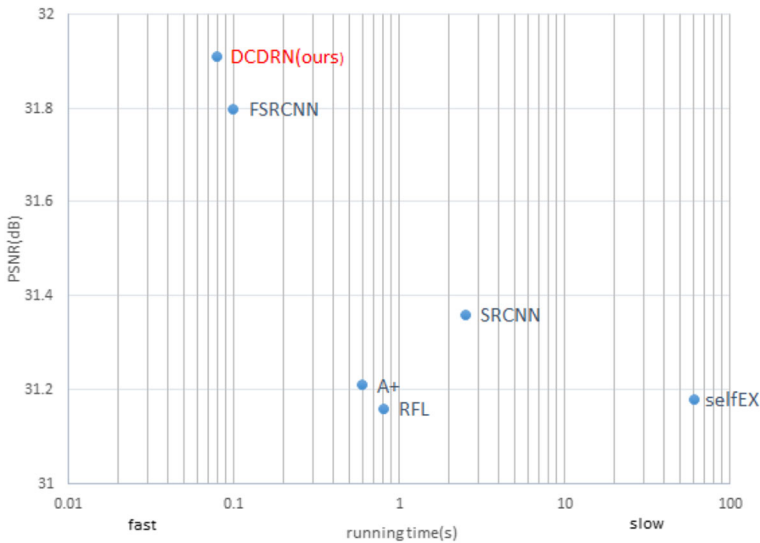


Fig. 1 Compared to the state-of-the-art methods [4, 6, 11, 33, 36], our DCDRN can significantly improves both the PSNR and computation costs over the scale factor $\times 2$ on the dataset *bsd100* [29]

obtains a high-resolution image by training the random forest which has the leaf nodes with regression functions.

SRCNN [4], the earliest method takes 2 to 4 convolutional layers for super-resolution [5] to acquire more significant performance and asserts CNN can use the end-to-end manner to learn the mapping from LR to HR. Although SRCNN achieved good results by combining deep learning with super-resolution, it still has some limitations:

- Over-reliance on information in small image domains;
- Has poor ability to convergence during training;
- The network structure only trains a single scale.

Hence, FSRCNN [6] replaced a large kernel into a small one and fine-tuning the final deconvolution layer result in a great faster than SRCNN. The authors of VDSR [16] concluded increase the depth of layer is more sensible than employ larger convolution filters and accelerated the training process by introducing residual learning and gradient clipping [2] into the network. VDSR proves residual network is particular suitable for solving super-resolution problems.

After extensive application of residual learning, the network structure tends to be deeply reconstructed. Deepened the network and reduced the number of parameters by sharing weights called DRCN [15] increased the computational complexity obviously. In Residual Encoder-Decoder Networks (RED) [27], the symmetrical structure that each convolutional layer has a deconvolution layer as a response has strong capacity to train very deep networks and achieve prominent performance. The algorithm enables inverse propagation signals transmit to the bottom layer immediately and settles the problem of gradient disappearance. Yet, the calculation gets more complexity and the time of training is longer than others because of the structure. Certainly, ResNet [10] and Highway Network [34] train more than 100 layers of network by bypassing the path between convolutional layers. The stochastic depth [12]

randomly discards certain layers and features to improve the performance of the residual network. While above studies reflect the trend of “deeper is better”, Romano recommended Rapid and Accurate Image Super Resolution (RAISR) [31] invented a shallow and faster learning-based method. Although performance is slightly lower than [15, 16, 27, 38].

Recently, growing numbers of methods have constructed a novel network by combining several convolution neural networks for super-resolution. For instance, Image super-resolution using a dilated convolutional neural network (DCNN) [19] and a two-channel convolutional neural network for image super-resolution (SDSR) [20] used differentiated convolution and multi-channel convolution also achieved remarkable performance respectively. FractalNet [17] combines the parallel networks of several different depths recently proposed and creates many short paths in the network. Symmetric dense connections have been introduced in an encoding-decoding network [28] for image restoration tasks. These all have something in common: dense connections have established between layers for efficient training.

3 Proposed method

3.1 Model overview

From Fig. 2, the network mainly divided into two elements: feature extraction network and image reconstruction network. We cascade a pair of layers (convolution, activation) in the feature extraction network. After capturing the local and global features adequately, the whole outputs are stacked in the concat layers and send into the reconstruction network as inputs. The deconvolution layers are used to reconstruct the details of the image and finally low resolution (LR) Images through CNN layers converted to high resolution (HR) images.

Figure 3 shows the loss feature information of model during the training process. Each pixel up-sampled undergoes a lot of calculations and misses more semantic information, especially with a growing scale factor. In the initial feature extraction network, we optimize the number

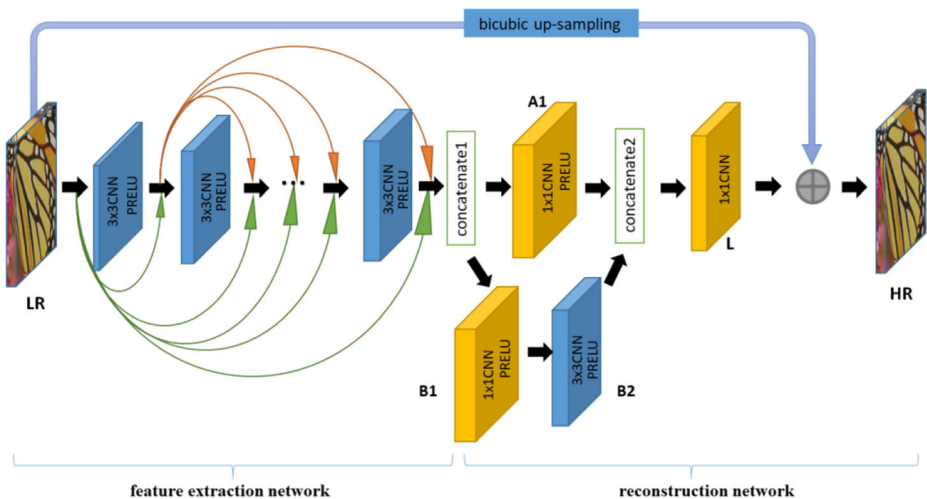


Fig. 2 Our network structure. Low resolution (LR) images through CNNs converted to high-resolution (HR) images. We employ 64 filters for every convolutional layer in feature extraction network. The blue area in the figure uses 3×3 CNNs, while the yellow part indicates the 1×1 CNNs

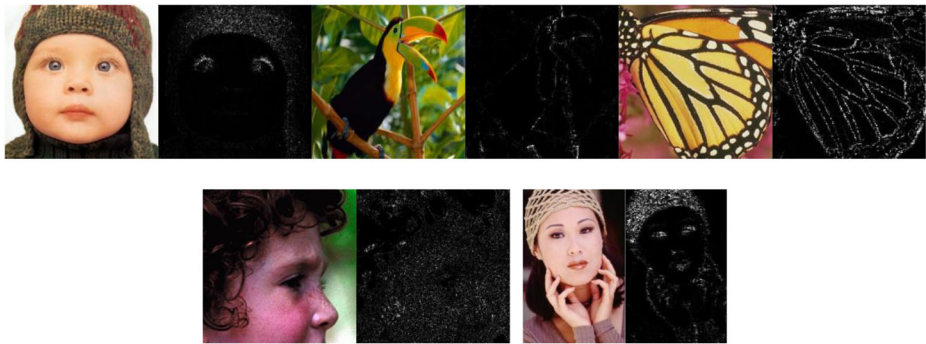


Fig. 3 The loss feature information of Set5

of filters at each layer and transfer the extracted features directly to the image reconstruction part through dense connections. Each output was used as the input to the rest of the layers. In this paper, the activation layer is added after each convolutional layer to enhance the expression ability of the network while increasing the characteristic nonlinear conduction. The PReLU activation function used to address the possible “dying ReLU” problem [8]. Getting more local features while prevent learning a large negative deviation term for better performance. The Table 1 shows 7 group convolution layers make the best performance. These will future discussed later in Section 4.2.

The deconvolution layer proposed by Matthew D. Zeiler [42] is common to learn up-sampled filters in the reconstruction network. The input dimension is quite large since all the features connected before the reconstruction network. Superposition training of the deconvolution layer is essential in order to obtain better reconstruction performance. The proposed structure is similar to the Network in Network [21] contains one or more 1×1 CNNs. In Fig. 4b shows the feature map extracted in the image reconstruction section. The same as the rest of the deep learning network, the proposed model focuses on learning the remaining output, even in the case of shallow (less than or equal to 7) module.

3.2 Benefit of feature combination

As shown in Fig. 4, the output of feature maps is demonstrated accurately. The formula is: $x_l = H_l(|x_0, x_1, \dots, x_{l-1}|)$, x_l represents the input of l-th layer, H_l includes the weights of PReLU and 3×3 CNNs. $|x_0, x_1, \dots, x_{l-1}|$ indicates the output from layer 0 to layer $l - 1$. Each layer can directly obtain a gradient from the loss function to obtain the input signal to reduce the loss of the feature.

Table 1 The comparison table with different parameters(Scale Factor = 3, test dataset = Set 5). The italic part is the best performance of PSNR

	Layer 5	Layer 6	Layer 7	Layer7 (filters = 32)	Layer 7 (filters = 128)
PSNR(dB)	33.21	33.30	<i>33.56</i>	33.47	33.32
Epochs	53	48	59	54	51
Steps	65,540	59,357	72,960	67,705	63,067

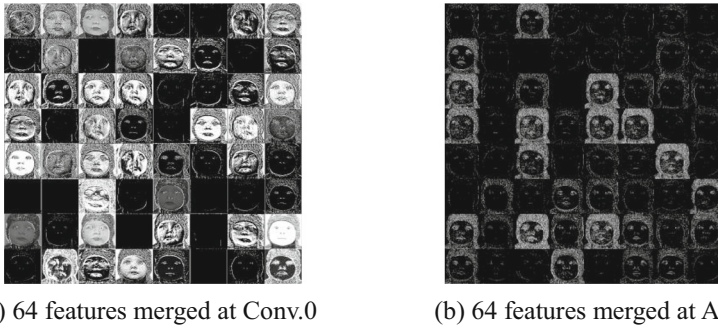


Fig. 4 Features obtained from different layers

Dense connections are committed to improving network performance from the perspective of feature reuse and splice the features extracted from each layer and connected low-level and high-level feature-linked feature maps and then use as input to the image reconstruction network. The output of each layer in front is equivalent to the input of each subsequent layer. In addition, the combination of feature maps generated from feature extraction network used for SR reconstruction. In the experimental part, the SR results compared with different network parameters in Table 1.

3.3 The structure of dimension reduction

Explained previous section, the transposed layer of convolution (also known as the deconvolution layer) proposed by [42] has widely application in up-sampled. In order to obtain better reconstruction performance, it is necessary to increase the depth of the deconvolution layer.

We propose a parallel structure consists of three 1×1 CNNs to avoid the expensive calculation. In Fig. 2, A1 and B1 after the concat layer is mainly to reduce the input dimension and the network parameters before generating the HR image. The last occurrence of L performs a final dimensionality reduction operation on the advanced features to facilitate the addition of the interpolated image to reconstruct the final HR image. Advantages of the structure are:

- Reduce the dimension of the input to the previous layer to perform less calculation with transferring more accurate information and less information loss, make more nonlinear combinations to enhance the potential representation of the network.
- Reduce the number of CNNs or transpose CNNs filters significantly.
- The calculation of 1×1 CNNs is at least 9 times less than 3×3 CNNs.

3.4 Model for multiple and large scales

Table 3 shows the results among different methods with large-scale factors. **DCDRN** yields more details than other existed methods indicate the algorithm has superiority when the scale factor becomes larger. Most methods with small scale factor work well and gradually deteriorate when the scale factor increases, because when the scale factor increases, the sub-

pixel edge recovery is more difficult and the information loss is more serious. Almost all methods based on deep learning associate with bicubic interpolation, the proposed algorithm combines it with dense connections to make the image recovery more apparent. As a result, we have observed that multi-scale training can improve large-scale performance.

4 Experimental results

This section shows the concrete performance of our method on several datasets. First, we introduce several datasets for testing and training, and then give detail parameters of the algorithm. Finally compare our methods with some state-of-the-art methods.

4.1 Datasets for training and testing

4.1.1 Training dataset

Different methods of deep learning have various training dataset. Dataset we used is 91 images from Yang et al. [39] and 200 images from Berkeley Segmentation Dataset [1]. We take 291 images as training data and flipped the images horizontally and vertically shown in Fig. 5. During the training phase, employed dataset Set5 [3] to evaluate performance and supervise if the model is over-fitting or convergent. The proposed method takes a color (RGB) image convert to an YCbCr image and only processing the Y channel.

4.1.2 Testing dataset

Three benchmark datasets, Set5, Set14 [43] and bsd100, are used for various networks commonly. Timofte et al. used the Berkeley segmentation dataset of natural images in [29]. All the super-resolution methods in Table 3 test on these three datasets to make the comparison results more convincing.

4.2 Training parameters

The learning rate has been set as 0.001, reduced twice while loss stop to decrease after 5 training epochs, training will terminative if the learning rate is lower than 0.00002. Different from the rest of deep learning super-resolution models of other technologies, we have fixed 7



Fig. 5 Augment result of “15,088” (bsd100) flipped horizontally, vertically

group of 3×3 CNNs in the feature extraction network and 64 outputs of each layer in order to reduce the parameters. Figure 4a shows the 64 feature maps extracted by the first layer during training. More detailed training process and specific parameters are shown in Table 2. Each training image is divided into 32 patches, using 64 patches as a mini-batch. A total of 118,384 batch data were generated of Set5. We optimized the regression target based on gradient descent of back-propagation [18]. L2 norm produce fuzzy predictions inevitably and the recovered high-resolution images tend to be more smoothly when training the network. Hence, the proposed DCDRN employ the L1 norm (the penalty factor is multiplied by 0.0001) to optimize the network model and set the momentum parameter to 0.9. The L1 norm refers to the difference between the absolute values of each element in the vector and called “Sparse Rule Operator”. It is used to improve the generalization ability by preventing over-fitting. L1-regularization reflects the mapping from LR to HR and obtains better hierarchical features.

Let x represents the interpolated low-resolution image and y be the high-resolution image. Given a training set, the aim is to learn a model f with an estimated value of $\hat{y} = f(x)$. \hat{y} is the target high-resolution image we estimated. Most of the deep-learning-based super-resolution methods use Rectified linear Units (ReLU) as an activation function:

$$\max(0, x_i) = \begin{cases} 0, & \text{if } x_i \leq 0 \\ x_i, & \text{if } x_i > 0 \end{cases}$$

In the less than zero part, ReLU all abandoned that causes many features lost, increasing the uncertainty of the image recovery. Therefore, a new activation function PReLU introduced by He et al. [9] used in our model. The expression is:

$$y_{ji} = \begin{cases} x_{ji}, & \text{if } x_{ji} \geq 0 \\ a_{ji}x_{ji}, & \text{if } x_{ji} < 0 \end{cases}$$

where

$$a_{ji} \sim U(l, u), l < u \& l, u \in [0, 1)$$

Table 2 The training parameters of each layer (While the scale factor = 2)

Network	Layer	Kernel size	Input	Total output
Feature extraction network	Conv_0	$3 \times 3 \times 64$	1	64
	Conv_1		64	128
	Conv_2		128	192
	Conv_3		192	256
	Conv_4		256	320
	Conv_5		320	384
	Conv_6		384	448
Image reconstruction network	A1	$1 \times 1 \times 64$	448	64
	B1	$1 \times 1 \times 32$	448	32
	B2	$3 \times 3 \times 32$	32	32
	L	$1 \times 1 \times 4$	96	4

Comparing the two equations, PReLU introduced a new parameter a_{ji} its value often set as 0.01 and make the value of the negative distribution retained. This method retains the original advantages of ReLU and obtains as many characteristics as possible. Above, employ PReLU as the activation function of the model. Each CNN layer is initialized with what He.et.al said. Also initialize the bias and PReLU to zero. Our training time took nearly 3 h on GTX 1080Ti.

4.3 Comparisons with state-of-the-art methods

Peak signal-to-noise ratio (PSNR) is used to compare the accuracy of the proposed method and other SR algorithms. Figure 6 shows the test results for different scale factors and the relationship between the number of iteration (k) and PSNR (dB) when using different scale factors. Numbers of experiments illustrate in Table 1 while looking for the best model. Changed the number of convolution layers, the number of filters and so on. Table 3 exhibits the comparison of PSNR for $\times 2$, $\times 3$, $\times 4$, respectively. The italic indicates the best performance and the bold shows the second best. It is obviously that the proposed algorithm has the best performance for these datasets. Dense connections make the feature maps learning increased lead to some slightly bigger time consumption than FSRCNN. While the image quality is improved, also increases the test time. Comprehensive consideration of both accuracy and time, our model has the best performance. The implementation of each algorithm uses different platforms and libraries, it is unfair to simply calculate the training and testing time. Comprehensive overall performance is needed to judge the quality of a model. We keep the other settings the same as the benchmark model.

The restoration of buildings is a very representative work in super resolution. In Figs. 7, 8, and 9, DCDRN restored a more pronounced edge profile with the buildings. Because of the dense connections, its features learn repetitively and guarantee quality of high resolution image. And feature maps are concatenating twice to ensure that the residual learning is fully performed, it obtains more contextual information and more conducive to the restoration of images in DCDRN. The performance can be observed by connecting the features of all layers indicate there are complementary information between SR different levels of feature mapping and have strong similarities. Finally, our method is the clearest for the outline recovery of bird and green peppers showed in Figs. 10 and 11. Through these images, the edges of other methods are blurred proved our method recovers better high resolution images.

As expected, our network achieved better results than other networks after employing dense connections and dimension reduction which shows the combination of shallow features and deep features improve SR reconstruction performance significantly.



Fig. 6 Steps (k) vs Performance (PSNR (dB))

Table 3 PSNR for scale factor with $\times 2$, $\times 3$ and $\times 4$ on datasets Set5, Set14, bsd100. Italic indicates the best performance and bold indicates the second

Dataset	Scale	Bicubic PSNR	A+ PSNR/time	RFL PSNR/time	selfEX PSNR/time	SRCNN PSNR/time	FSRCNN PSNR/time	Ours PSNR/time
Set5	$\times 2$	33.66	36.54/0.58	36.54/0.63	36.49/45.78	36.66/2.19	37.00 /0.068	<i>37.65</i> /0.097
	$\times 3$	30.39	32.58/0.32	32.43/0.49	32.58/33.44	32.75/2.23	33.16 /0.027	<i>33.56</i> /0.143
	$\times 4$	28.42	30.28/0.24	30.14/0.38	30.31/29.18	30.48/2.19	30.71 /0.015	<i>31.11</i> /0.145
Set14	$\times 2$	30.24	32.26/0.86	32.26/1.13	32.22/105.00	32.42/4.32	32.63 /0.160	<i>33.07</i> /0.156
	$\times 3$	27.55	29.13/0.56	29.05/0.85	29.16/74.69	29.28/4.40	29.43 /0.061	<i>29.83</i> /0.136
	$\times 4$	26.00	27.32/0.38	27.24/0.65	27.40/65.08	27.49/4.39	27.59 /0.029	<i>27.91</i> /0.122
bsd100	$\times 2$	29.56	31.21/0.59	31.16/0.80	31.18/60.09	31.36/2.51	31.80 /0.098	<i>31.95</i> /0.078
	$\times 3$	27.21	28.29/0.33	28.22/0.62	28.29/40.01	28.41/2.58	28.60 /0.035	<i>28.83</i> /0.060
	$\times 4$	25.96	26.82/0.26	26.75/0.48	26.84/33.57	26.90/2.51	26.98 /0.019	<i>27.31</i> /0.052

5 Conclusion and future works

The accurate super-resolution method based on convolution neural network proposed in this paper makes use of the dense connections and 1×1 CNNs structure. The network for feature extraction is optimized structurally and both local and global features are merged through dense connections and transmitted to the reconstructed part of the network.

In the reconstructed network, the architecture of the 1×1 CNNs enables the model to obtain better reconstruction capabilities, and the amount of calculation is smaller than others. In addition, the needless preprocessed images and handled the original size of the image

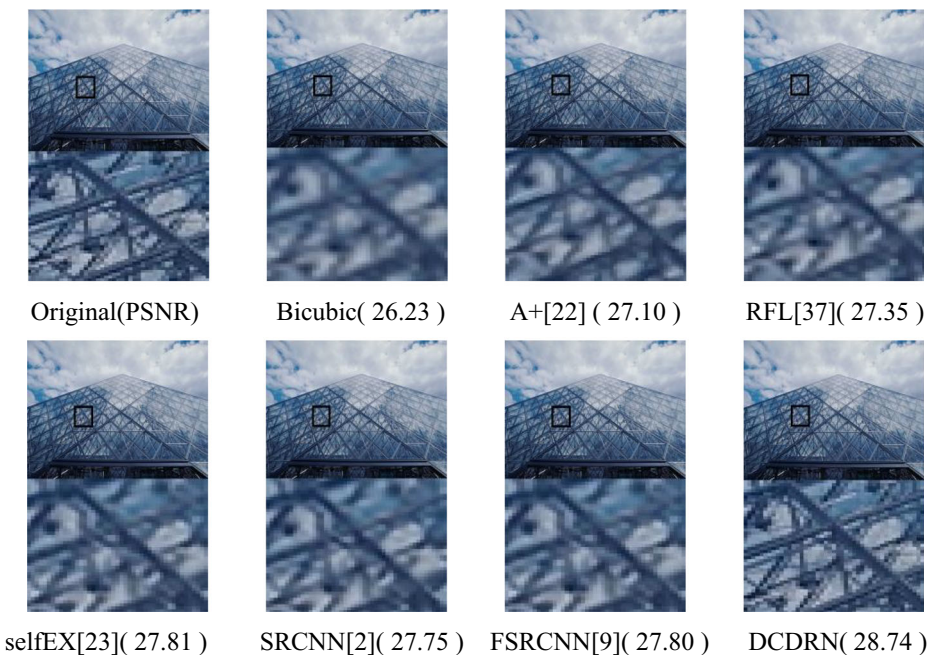


Fig. 7 Super-resolution results of “img_045” (bsd100) with scale factor $\times 2$. For the outline of the building, our method is more accurate and we can see the structure between the glasses more clearly

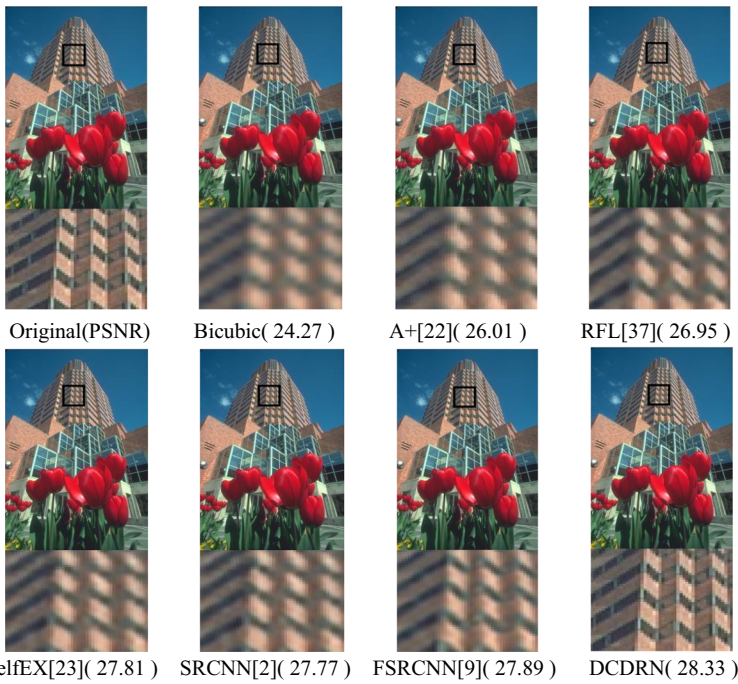


Fig. 8 Super-resolution results of “img_095” (bsd100) with scale factor $\times 3$. Our method restored better details. Better reconstruct the outline of the windows, you can clearly see the structure between the wall and the windows

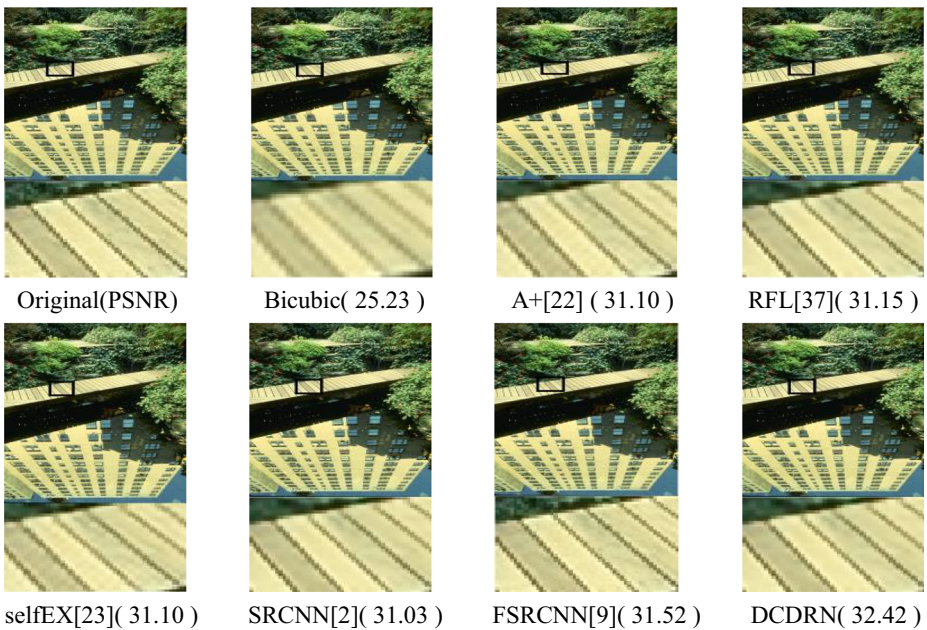


Fig. 9 Super-resolution results of “img_021” (bsd100) with scale factor $\times 2$. Our method recovers sharp lines

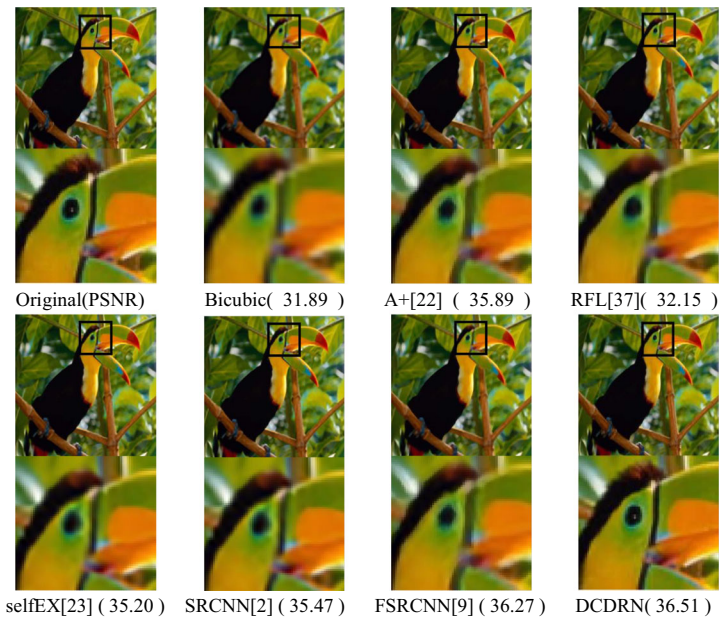


Fig. 10 Super-resolution results of “img_002” (Set5) with scale factor $\times 3$. The veins in the image are sharpest in our results

directly. With these conditions, the proposed model achieves better performance with less calculation. After numerous experiments and failures, the model has finally put forward. Another noteworthy aspect of this study is uses an integrated learning network. We combine the two hierarchical structure networks. Deep learning itself has a good ability to solve complex problems. Classic ensemble learning achieves better results with less computation

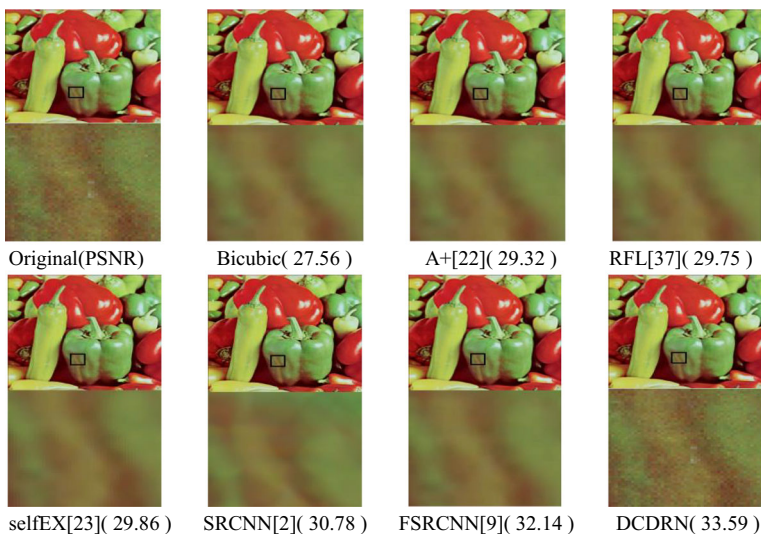


Fig. 11 Super-resolution results of “img_0012” (Set14) with scale factor $\times 4$. For the edge of green pepper, our method is best

even if the problem is non-negligible. Besides, the integrated model makes parallelization easier and faster. Gradient clipping is used to ensure the training stability. Our method greatly outperforms existing methods on the baseline image. We believe that the proposed DCDRN is easily applied to other image recovery problems such as denoising and defogging.

Acknowledgements This work was supported by the National “Twelfth Five-Year” development plan of science and technology (No.2014BAG03B05) and the National Natural Science Foundation of China (No. 61772294).

References

1. Arbelaez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(5):898–916
2. Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. In: *IEEE Transactions on Neural Networks*, pp. 157–166
3. Bevilacqua M, Roumy A, Guillemot C, Alberi M (2012) Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: *British Machine Vision Conference*, pp. 1–10
4. Dong C, Loy C, He K, Tang X (2014) Learning a deep convolutional network for image super-resolution. In: *European Conference on Computer Vision*, 184–199
5. Dong C, Loy C, He K, Tang X (2015) Image super-resolution using deep convolutional networks. In: *TPAMI*, pp. 295–307
6. Dong C, Loy C, He K, Tang X (2016) Accelerating the Super-Resolution Convolutional Neural Network. In: *European Conference on Computer Vision*, pp. 391–407
7. Gao X, Zhang K, Tao D, Li X (2012) Image super-resolution with sparse neighbor embedding. In: *IEEE Transactions on Image Processing*, pp. 3194–3205
8. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: *IEEE International Conference on Computer Vision*, pp.1026–1034
9. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance. In: *ImageNet Classification*, pp. 1020–1034
10. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778
11. Huang J, Singh A, Ahuja N (2015) Single image super-resolution using transformed self-exemplars. In: *CVPR*, pp 5197–5206
12. Huang G, Sun Y, Liu Z, Sedra D, Weinberger KQ (2016) Deep networks with stochastic depth. In: *European Conference on Computer Vision*, pp. 646–661
13. Huang G, Liu Z, Laurensvander M, Kilian Q (2017) Densely connected convolutional networks. In: *Computer Vision and Pattern Recognition*, arXiv:1608.06993
14. Irani M, Peleg S (1991) Improving resolution by image registration. In: *Graphical Models and Image Processing*, pp. 231–239
15. Kim J, Lee K, Lee M (2016a) Deeply-recursive convolutional network for image super-resolution. In: *Computer Vision and Pattern Recognition*, pp 1637–1645
16. Kim J, Lee K, Lee M (2016b) Accurate image super-resolution using very deep convolutional networks. In: *Computer Vision and Pattern Recognition*, pp 1646–1654
17. Larsson G, Maire M, Shakhnarovich G (2017) Ultra-deep neural networks without residuals. In: *International Conference on Learning Representations*, arXiv:1605.07648
18. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. In: *IEEE*, pp. 2278–2324
19. Li G, Wu Q, Qiu L, Huang X (2018) Image super-resolution using a dilated convolutional neural network. *Neurocomputing*, 1219–1230
20. Li S, Ren F, Lei G, Yue G, Hou C (2018) A two-channel convolutional neural network for image super-resolution. *Neurocomputing* 275:267–277
21. Lin M, Chen Q, Yan S (2014) Network in network. In: *International Conference on Learning Representations*, arXiv:1312.4400
22. Liu S, Wu X (2006) An edge-guided image interpolation algorithm via directional filtering and data fusion. In: *TIP*

23. Liu S, Lu MY, Liu GC, Pan Z (2017) A novel distance metric: generalized relative entropy. *Entropy* 19(6): 269
24. Liu S, Pan Z, Song HB (2017) Digital image watermarking method based on DCT and fractal encoding. In: *IET Image Processing*, pp.815–821
25. Liu S, Bai WL, Liu GC, Li WH, Srivastava M (2018) Parallel fractal compression method for big video data. *Complexity* 2016976
26. Liu G, Liu S, Muhammad K, Kumar A (2018) Object tracking in vary lighting conditions for fog based intelligent surveillance of public spaces. In: *IEEE Access*, pp. 29283–29296
27. Mao X, Shen J, Yang Y (2016a) Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: *Neural Information Processing Systems*, pp. 2802–2810
28. Mao X, Shen C, Yang Y (2016b) Image restoration using convolutional auto-encoders with symmetric skip connections. In: *Proceedings of the Neural Information Processing Systems*, arXiv:1606.08921
29. Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *International Conference on Computer Vision*, volume 2, pp. 416–423
30. Pan Z, Liu S, Fu W (2016) A review of visual moving target tracking. *Multimedia Tools and Applications*, pp. 16989–17018
31. Romano Y, Isidoro J, Milanfar P (2017) Rapid and accurate image super resolution. *IEEE Transactions on Computational Imaging* 3:110–125
32. Salvador J, Pérez-Pellitero E (2015) Naive Bayes super-resolution forest. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 325–333
33. Schuler S, Leistner C, Bischof H (2015) Fast and accurate image upscaling with super-resolution forests. In: *CVPR*, pp. 3791–3799
34. Srivastava RK, Greff K, Schmidhuber J (2015) Training very deep networks. In: *Advances in Neural Information Processing Systems*, pp. 2377–2385
35. Thom M, Schweiger R, Palm G, Roethermel A (2013) Learning convolutional neural networks from few samples. In: *International Symposium on Neural Networks*, pp 1884–1890
36. Timofte R, DeSmet V, VanGool L (2014) Adjusted anchored neighborhood regression for fast super-resolution. In: *ACCV*, pp. 111–126
37. Tu WQ, Pop F, Jia W, Wu J, Iacono M (2018) Visual attention feature (VAF): a novel strategy for visual tracking based on cloud platform in intelligent surveillance systems. *Journal of Parallel and Distributed Computing*, pp. 182–194
38. Yamanaka J, Kuwashima S, Kurita T (2017) Fast and accurate image super resolution by deep CNN with skip connection and network in network. In: *Computer Vision and Pattern Recognition*, arXiv:1707.05425
39. Yang J, Wright J, Huang T, Ma Y (2010) Image super resolution via sparse representation. *IEEE Trans Image Process* 19(11):2861–2873
40. Yang J, Wang Z, Lin Z, Cohen S, Huang T (2012) Coupled dictionary training for image super-resolution. In: *IEEE Transactions on Image Processing*, pp. 3467–3478
41. Yang C, Ma C, Yang M (2014) Single-image super-resolution: a benchmark. In: *European Conference on Computer Vision*, pp. 372–386
42. Zeiler MD, Krishnan D, Taylor GW, Fergus R (2010) Deconvolutional networks. In: *Computer Vision and Pattern Recognition*, pp. 2528–2535
43. Zeyde R, Elad M, Protter M (2012) On single image scale-up using sparse-representations. In: *International Conference on Curves and Surfaces*, pp. 711–730

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Shiyu Hu obtained her B.Eng. degree from Qingdao Agricultural University in 2017. Now she is a master degree candidate in Qingdao University. His research interests include machine learning and image processing.



Guodong Wang Male, Born in Weifang City, Shandong Province, China, in February, 1980. Now he is an associate professor in College of Computer Science and Technology, Qing University. He received bachelor degree in 2001 and master degree in 2004 in control theory and control engineer, Qingdao University of Science and Technology, and received Ph.D degree in pattern recognition and intelligent system in Huazhong University in 2008. His research Interest include: Variational Image Science, Face recognition, Intelligent video surveillance, 3D reconstruction and Medical image processing and Analysis.



Yanjie Wang obtained his B.Sc. degree from Qingdao University in 2017. Now he is a master degree candidate in Qingdao University. His research interests include machine learning and image processing.



Chenglizhao Chen received the PhD. degree in Technology of Computer Application from Beihang University, 2017. He is currently an assistant professor at Qingdao University. His research interests include pattern recognition, computer vision, and machine learning.



Zhenkuan Pan was born in 1966. He received his Ph.D. Degree in engineering mechanics from Shanghai Jiao Tong University, China in 1992. Currently he is a professor of College of Computer Science and Technology at Qingdao University. His research interests include dynamics and control of multibody systems, computer simulation and variational image processing. He has published numerous papers and conference papers in the area of image processing and object recognition.