CAAI Transactions on Intelligence Technology

Research Article

Guided filter-based multi-scale super-resolution reconstruction

ISSN 2468-2322
Received on 12th November 2019
Revised on 2nd March 2020
Accepted on 28th March 2020
doi: 10.1049/trit.2019.0065
www.ietdl.org

Xiaomei Feng[1,2], Jinjiang Li[1,2] ✉, Zhen Hua[1,2]

[1]*School of Electronic and Communications Engineering, Shandong Technology and Business University, Yantai 264005, People's Republic of China*
[2]*Co-innovation Center of Shandong Colleges and Universities: Future Intelligent Computing, Shandong Technology and Business University, Yantai 264005, People's Republic of China*
✉ *E-mail: lijinjiang@gmail.com*

Abstract: The learning-based super-resolution reconstruction method inputs a low-resolution image into a network, and learns a non-linear mapping relationship between low-resolution and high-resolution through the network. In this study, the multi-scale super-resolution reconstruction network is used to fuse the effective features of different scale images, and the non-linear mapping between low resolution and high resolution is studied from coarse to fine to realise the end-to-end super-resolution reconstruction task. The loss of some features of the low-resolution image will negatively affect the quality of the reconstructed image. To solve the problem of incomplete image features in low-resolution, this study adopts the multi-scale super-resolution reconstruction method based on guided image filtering. The high-resolution image reconstructed by the multi-scale super-resolution network and the real high-resolution image are merged by the guide image filter to generate a new image, and the newly generated image is used for secondary training of the multi-scale super-resolution reconstruction network. The newly generated image effectively compensates for the details and texture information lost in the low-resolution image, thereby improving the effect of the super-resolution reconstructed image.Compared with the existing super-resolution reconstruction scheme, the accuracy and speed of super-resolution reconstruction are improved.

1 Introduction

Single image super resolution (SISR) is a reconstruction technique that recovers high-resolution (HR) images from low-resolution (LR) images [1]. Recovery of HR from LR is an illness problem due to the loss of important information in LR [2]. However, since super-resolution reconstruction can be used as a monitoring facility [3], medical imaging [4] and other built-in modules [5, 6] for performing image restoration and recognition tasks. Therefore, super-resolution reconstruction has a wide range of applications and promotes the development of super-resolution reconstruction. The SISR method is mainly divided into three categories: based on learning [7]; based on interpolation [8]; and based on reconstruction. To improve the accuracy of super-resolution reconstruction, a reference-based super-resolution (RefSR) reconstruction method has been proposed in recent years. The RefSR method utilises HR detail in the reference image to achieve super-resolution reconstruction. The RefSR method is more competitive than the SISR method, but the key issue of the RefSR method is how to solve the problem of transmitting the high frequency detail of the reference image to the LR image.

With the great success of deep learning in computer vision [9, 10], the learning-based super-resolution reconstruction method has also become a research hotspot. For example, such as Random Forest [11] and Convolutional Neural Network (CNN) [12] were first applied to super-resolution reconstruction, which opened a new era of super-resolution reconstruction. Subsequent SR methods included SRCNN [13], which used a three-layer CNN to perform super-resolution reconstruction of LR images. VDSR [14] used the residual network for the first time to solve the SR problem. ESPCN [15] made the up-sampling data more accurate and reduced noise interference. The generative adversarial network SRGAN [16] introduced the perceptual loss and improved the visual effect after super-resolution reconstruction. The dense residual block EDSR [17] with skip connections was used to fully extract the features of different layers. The wide residual WDSR [18], which considered that the effective features were mainly concentrated in low-level features, and the wide residual network was used for effective learning. Later DBPN [19] used up and down iterative sampling to learn features more effectively. The learning-based super-resolution reconstruction method is to input the LR image into the network and learn the non-linear mapping relationship between LR and HR through convolution. Because the input LR image part features are lost, the mapping relationship learned by the convolutional network are also different, super-resolution reconstruction effects are uneven.

In recent years, with the emergence of perceptual loss [20] and adversarial loss [21], large-scale factor reconstruction has greatly improved visual effects [16, 22], but reconstructed images often exhibit artefacts. To further improve the accuracy of super-resolution reconstruction, the RefSR method introduced additional images to aid in super-resolution reconstruction tasks, but required careful selection of reference images. The reference image should have textures and content similar to the LR image, similar to the two initial images of image style migration [23, 24], which can be obtained from adjacent frames in the video [25], or from different visual images [26] or obtained from an external data set [27]. The reference images selected by these methods are all based on the reference of the LR image itself, and the LR image needs to be strictly aligned with the reference image [28], so the robustness is low.

In this paper, we propose a multi-scale super-resolution reconstruction method based on a guided image filter. Like the traditional SISR, we use a multi-scale method to build a network and fully exploit the non-linear mapping between LR and HR. The difference with the traditional SISR is that we draw on the idea of RefSR. Real HR image features are applied to the super-resolution reconstruction network by using guided image

filter. The algorithm is divided into two stages. In the first stage, the LR image is transmitted through the network to perform different scales of up-sampling for SR tasks. In the second stage, the HR image output from the first stage and the real HR image are subjected to a guide image filtering process to obtain a new image. The new image and real HR image are used to jointly train the network to achieve super-resolution reconstruction. In the first stage, in a multi-scale super-resolution reconstruction network, images of different scales at each level play different roles, and feature extraction is performed using a code–decoded structure with residual blocks [29], using the improved long short-term memory (LSTM) [30] processing inputs of different scales; Super-resolution reconstructed image distortion due to loss of LR features of the first stage input. In this paper, the method of guiding image filter is used to effectively supplement the details of LR image loss and texture information, and then train through the first-stage network to complete the super-resolution reconstruction task.

2 Related work

2.1 Super-resolution reconstruction

2.1.1 SISR reconstruction: With the development of deep learning, the deep learning-based reconstruction method has shown superior performance in both qualitative and quantitative analysis of images [31, 32]. The SRNN proposed by Timofte *et al.* [13] introduced CNN into the SR field for the first time. SRNN only used a three-layer network to extract features and used mean square error (MSE) as the loss function. The experimental results are good, which proves the effectiveness of deep learning. As the depth of the network increases, the results of the training will become more and more accurate, but the deep network also brings difficulties to the training. The problem of gradient disappearance or gradient explosion hinders the design of deeper networks. In 2016, Kim *et al.* [14] proposed a network ResNet that can be connected by jumping, and its network depth can reach 152 layers. Effectively solve the problem of gradient disappearance or explosion, and make the network develop to a deeper level. The combination of deep networks and residual blocks has emerged, such as EDSR [17], WDSR [18], DBPN [19]. EDSR super-resolution reconstruction is better, but the number of network layers is deeper and the number of parameters is larger, the time used for super-resolution reconstruction is longer. Compared with EDSR, WDSR adopts weight standardisation and removes many redundant convolutional layers, which has improved structure and performance. DBPN is different from the previous method. DBPN uses the projection unit to perform up-and-down iterative sampling. The extraction features are more comprehensive, and the super-resolution reconstruction results are better, but the network complexity is higher.

From the sampling method, SISR can be divided into four categories: predefined up-sampling, single up-sampling, progressive up-sampling, and iterative ups and downs. The predefined upsampling is to learn the non-linear mapping between LR and HR. Before inputting the network, the LR image is first interpolated to enlarge the image size to match the target image size, such as Bicubic. However, this method is easy to make noise and affects the quality of reconstruction. To solve this problem, a single up-sampling occurs, the predefined up-sampling interpolation operation is removed, and the LR deconvolution is performed on the last layer, such as FSRCNN [33] or ESPCN [15], but CNN has insufficient learning ability and poor reconstruction effect. Progressive upsampling uses the Laplacian pyramid network to progressively predict SR images, similar to a single upsampled stack. Enhanced ability to learn complex mappings, fewer parameters, shorter runtimes, and more effective for large-scale factors. The iterative up-and-down sampling DBPN [19] has two units of upper projection and lower projection, which implements iterative up-sampling and down-sampling. The network can deeply explore the direct interdependence between LR and HR. The network has higher complexity, but the super-resolution reconstruction is better.

From the loss function, the model based on deep learning usually trains the parameters by minimising the MSE between the real image and the network output image, but this does not represent the true visual experience of human beings. However, the perceptual loss can lead to better visual effects, Johnson *et al.* [21] demonstrate the effectiveness of perceptual loss on network training. For example, the perceptual loss is used in generative adversarial networks (GANs) [16], and the adversarial loss is introduced, and the perceptual correlation distance between the real value and the network output value is minimised, but the GAN-based perceptual loss method is based on distortion. The cost is to improve the perceived image quality, so the perceptual loss function still has some limitations for the super-resolution reconstruction task.

2.1.2 RefSR reconstruction: To increase the quality of super-resolution reconstruction, a method of assisting LR images for super-resolution reconstruction using additional reference images [34, 35] has emerged. These methods are called RefSR methods. The method Boominathan *et al.* [34] use DSLR to capture the HR features of the reference image as a reference, and apply a patch synthesis algorithm to reconstruct. Adding a patch registration in [35] to improve the algorithm before the nearest neighbour search, and then uses dictionary learning to reconstruct. Decompose the image in [36] into frequency subbands and apply patch matching for high-frequency subband reconstruction. Patch-based synthesis algorithms cannot handle the non-rigid deformation of irregularities, thus causing synthetic patches to have artefacts and blurring. Although the sliding window method proposed by Boominathan *et al.* [34] can effectively alleviate this problem, the calculation cost of the method is enormous. Compared with the existing RefSR method, we do not need to consider the corresponding problem between HR image and LR image. We use the guide graph filter to generate a new image as the training image, and use the L2 loss function to train, so that the reconstruction can be achieved real-time requirements.

2.2 Guide image filter

The guide image filter is a filter with adaptive weights that can maintain the boundary while smoothing the image. Unlike Gaussian filtering and bilateral filtering, guided filtering is directional, selective for regions and edges, and is an anisotropic filter. Guided filtering uses the input image as a guide map to effectively distinguish the edges and regions of the image, and has edge sensing capability to better protect the edges and details of the image.

Guide image filtering is simply defined as follows:

$$q_i = \sum_j W_{ij}(I) p_j \tag{1}$$

where I is the guide image, P is the input image to be filtered, q is the filtered image, W is the weight value determined according to the guide map I, and W can be calculated by the following equation:

$$W_{ij}(I) = \frac{1}{|\omega|^2} \sum_{k(i,j)\in\omega_k} \left(1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\sigma^2 + \varepsilon}\right) \tag{2}$$

where μ_k is the mean of the pixels in the window, I_i, I_j are the values of two adjacent pixels, σ_k is the variance of the pixels within the window, and ε is the penalty value.

For (2) analysis, when I_i, I_j are located on both sides of the boundary, If the positive and negative signs of the $I_i - \mu_k$ and $I_j - \mu_k$ values are different, the weight value decreases. Instead, the weight value increases. Therefore, the weight value of the alien sign is much smaller than the weight value of the same number, and the pixels in the flat region have a larger weight, and the

smoothing effect is better. The pixel weights on both sides of the boundary are less, the smoothing effect is weaker, and the boundary is better.

The above theory proves the feasibility of guided filtering, but the running time is slow, and it is difficult to achieve real-time requirements. The fast-guided image filter proposed by He and Jian [37] is consistent with the original filtering effect. However, it is about 80% higher than the original running speed and reduces the time complexity from $O(N)$ to $O(N/s^2)$.

The fast guided image filtering is first driven by the local linear model as shown in (3), where I, p, and q are the guidance images, the filtered input image, and the filtered output image, respectively

$$q_i = a_k I_i + b_k, \quad \forall_i \in \omega_k \tag{3}$$

where i is the pixel index, and k is the index of the local square window ω of radius r, where a_k is (4) and b_k is (5)

$$a_k = \frac{(1/|\omega|)\sum_{i\in\omega_k} I_i p_i - \mu_k \overline{p}_k}{\sigma^2 + \varepsilon} \tag{4}$$

$$b_k = \overline{p_k} - a_k \mu_k \tag{5}$$

where μ_k and σ_k are the mean and variance of I in window k, respectively, and ε is the regularisation parameter that controls the smoothness.

The calculation method of the output image after filtering is as shown in the following equation:

$$q_i = \overline{a}_i I_i + \overline{b}_i \tag{6}$$

where $\overline{a}_i$ and $\overline{b}_i$ are the average values of a and b of window ω_i centred on i, respectively.

2.3 U-Net network

The U-Net network [38] is a variant of the fully convulted neural network. The network consists of two parts: the contracted path and the extended path. The image is first input into the network, convolved through two 3×3 convolution kernels (no padding operation), and adjusted by the non-linear activation function ReLU, downsampled through the pooling layer of 2×2. The shrinking path is subjected to five such operations, and each time the pooling layer is subjected to a downsampling, the final downsampling result is obtained. Then, through the extended path, the 2×2 convolution kernels is used for upsampling, the number of feature channels is halved, the number of channels corresponding to the contraction path portion is fused, and the features of the corresponding phase of the contraction path are spliced to the upsampling stage. The process also has five operations. At the last layer of the network, the convolution kernel is used for dimensionality reduction, and the result is mapped to the expected number. The network structure is shown in Fig. 1.

The shrink path is used to extract features and the extended path is used for accurate positioning. The U-Net network uses elastic deformation to enhance the data, thereby reducing the number of training samples, thus achieving end-to-end training for a very small number of samples, and the training results are superior to sliding window convolution [38]. Due to the small number of samples, the training speed of the network is also very fast. At first, the U-Net network was applied to medical images with a small number of samples, and later improved, and the processing of the SR problem [29] also achieved good results.

3 Proposed method

This paper proposes a multi-scale super-resolution reconstruction algorithm based on guide pattern filtering. In the first stage, referring to Fig. 2, the LR image input network is up-sampled

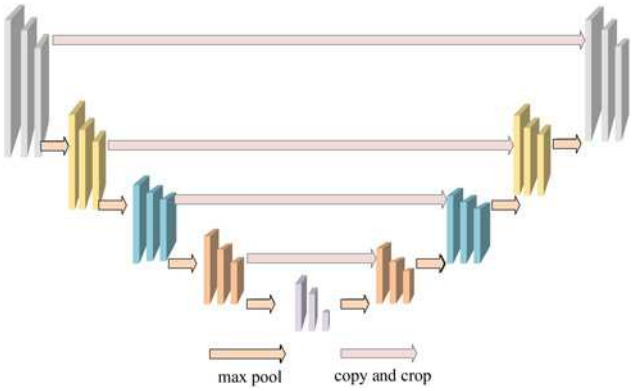

Fig. 1 *U-Net network structure*

according to different proportions to obtain a series of images LR_1, LR_2, LR_3 as sub-network input. First, the LR_1 image is trained using an encoding–decoding network with a residual block to obtain training result HR_1, up-sampling HR_1 and passing it to the next scale LR_2, trains again through the code–decoding network, obtains HR_2, up-samples HR_2 and passing it to the next scale LR_3, training again through the encoding–decoding network, and finally outputting the super-resolution image HR_3; In the second stage, the super-resolution image HR_3 obtained in the first stage and the real super-resolution image are subjected to the guide image filtering process, and the obtained image is used as the training image of the first stage to retrain the network for the super-resolution reconstruction task. The method flow chart is shown in Fig. 2.

3.1 Multi-scale super-resolution reconstruction network

Encoding–decoding networks have a wide range of applications in the field of computer vision. We improved the U-Net network, introduced residual blocks to improve the encoding–decoding network [39], and added the number of convolution layers to better extract LR features on different scales. In addition, for the problem of slow network convergence caused by too many convolutional layers, a hidden state loop module [40] is used internally to make the network quickly converge. There are re-used modules inside the network. The strategy in [29] is adopted, and ConvLSTM [41] is included in the last layer of the contraction path, which is used to hide the state and can achieve different scale connections.

In a multi-scale super-resolution reconstruction network, the LR image is upsampled into three scales, where the ith scale is half of the $(i+1)$th scale. For a code–decoder network, the input image is first convolved and features are extracted, after passing two ResBlocks, followed by a ConvLSTM, and then through two ResBlocks. The network structure is shown in Fig. 3. First, the first convolutional layer maps the image to $H \times W \times 32$. After the first residual block processing, the changed image becomes $H/2 \times W/2 \times 64$. After the second residual block, the image changes to $H/4 \times W/4 \times 128$. Image size and channel number do not change with ConvLSTM. The network moves further forward, and after two transposed layers, the image size changes to $H \times W \times 32$. The convolution kernel has a size of 5×5, the convolution step of the residual block and the transposed layer is 2, and the other convolution step is 1. ReLU is used as an activation function between convolutional networks. The network structure is shown in Fig. 3.

3.2 Extracting features

When the LR image passes through the contraction path of the encoding–decoding network, the decoder adjusts the image size through the corresponding transposed convolution layer [42] to

CAAI Trans. Intell. Technol., 2020, Vol. 5, Iss. 2, pp. 128–140

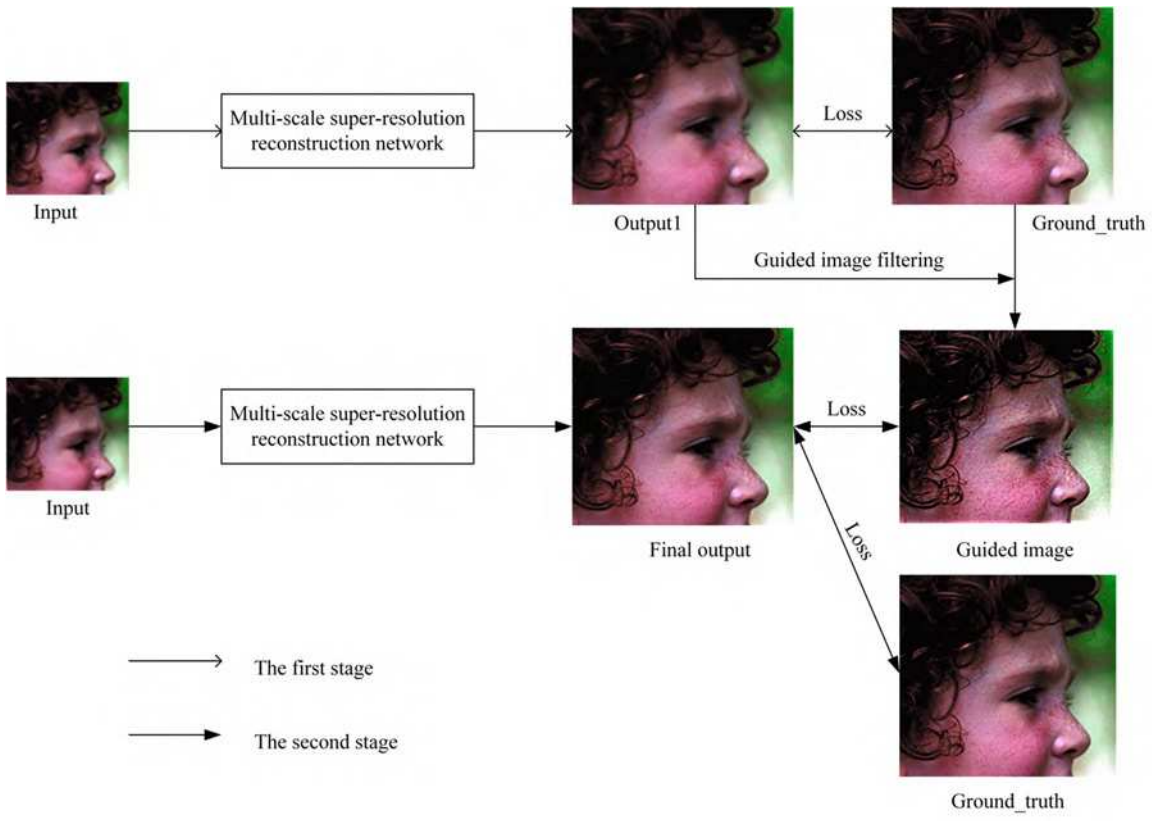

Fig. 2 *Algorithm flowchart*

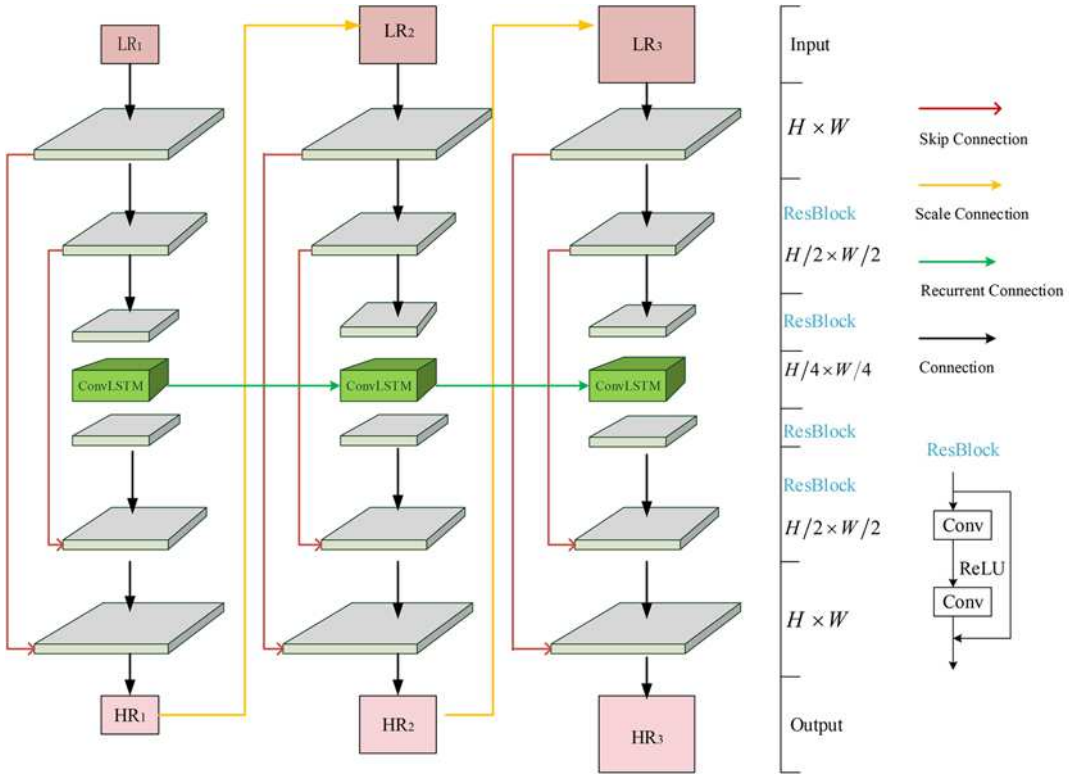

Fig. 3 *Multi-scale super-resolution reconstruction network structure*

generate a large amount of data, and the network needs to capture the effective information in the data. LSTM can establish stable long-term dependence in various studies [39, 42]. However, LSTM only controls the time and does not process the space, considering the spatial correlation of the data. Xingjian *et al*. [41] improved FC-LSTM [43] and used convolution instead of full

connectivity. ConvLSTM estimates the value of a pixel by the input of a pixel's adjacent area and the previous state.

After the input LR image is downsampled, ConvLSTM is inserted into the intermediate stage as a natural selection of sequential input. The network structure includes the following:

$$f_i = \mathrm{Net_E}(J_i^{\mathrm{H}};\ \theta_{\mathrm{E}}) \tag{7}$$

$$g_{\mathrm{i}},\ h_i = \mathrm{ConvLSTM}(f_i,\ h_{i-1};\ \theta_{\mathrm{LSTM}}) \tag{8}$$

$$I_0^{(i)} = \mathrm{Net_D}\left(g_{\mathrm{i}},\ H_i^{\mathrm{E}};\ \theta_{\mathrm{D}}\right) + I_0^{L\uparrow} \tag{9}$$

where $\mathrm{Net_E}$ is the CNN of the encoder, θ_{E} is the parameter, and f_i is the encoder network output. $\mathrm{Net_D}$ is the convolutional network of the decoder, θ_{D} is the parameter, and g_{i} is the decoder network input. h_i is the hidden state of the ith step ConvLSTM. H_i^{E} is the sum of the intermediate feature maps of the encoder network and is used for jump connections. $I_0^{(i)}$ is the output image at the ith time. $I_0^{L\uparrow}$ is the result of Bicubic sampling of I_0^L.

3.3 Multi-scale super-resolution reconstruction based on guided filtering

The image processing in computer vision and computer graphics mostly involves the concept of image filtering to suppress noise or extract useful image structures. Our commonly used linear translation invariant (LTI) filters, including Gaussian filters and Sobel filters, are widely used in image blurring, edge detection, and feature extraction. The filter kernel of the LTI filter is fixed, so the filtering does not consider the content of the image. In addition, although the LTI filter is effective, it takes a long time to calculate and cannot meet the real-time requirements. Bilateral filtering determines the weight by guiding the intensity of the image or colour similarity, and weights the average of the pixels, so the filter can filter based on the image content considerations. The bilateral filter can smooth the edges, but gradient inversion artefacts may appear near the edges and the calculation complexity is high.

Unlike the above two types of filtering, the guide image filter is based on image content filtering. The guide image may be the input image itself or other different images. Guide image filters not only have the characteristics of bilateral filters to keep the edges smooth, but also have no gradient inversion artefacts. The guided image filtering uses a local linear structure, and the calculation is relatively small. For greyscale and colour images, the guided filter has an O (N) time (in pixels N) exact algorithm Fig. 4.

In the first stage multi-scale super-resolution reconstruction network, we use real super-resolution images to train the network, and the loss function is L2. Such as formula (10). In the first stage, the large-scale super-resolution reconstructed image has blurred edges and artefacts, and the reconstruction effect of the details is not satisfactory. This shows that in the process of convolution, the mapping relationship between LR images and HR images has not been fully and effectively learned. To make the network learn the mapping relationship as fully as possible, we can increase the number of convolutions, or expand the receptive field. In this paper, to make the network learn a more effective mapping relationship, we use a guided image filter. The guided image filter can extract the structure information from the guide image and merge the structure information into the input image to achieve the integration of one information. First, a guided filtering operation is performed on the HR image output in the first stage, and the guided image is a real HR image. We use guided filtering to supplement the details that fail during the super-resolution reconstruction process and improve the performance of the super-resolution reconstruction image

$$L = \sum\nolimits_{i=1}^{n} \frac{K_i}{N_i} \left\| I^i - I_*^i \right\|_2^2 \tag{10}$$

where K_i is the proportional weight, usually $K_i = 1.0$. N_i is the number of I^i elements. I_*^i is a true HR image, and I^i is a HR image of the network output.

The image optimised by guided filtering cannot be used as the final super-resolution image. We combine the images after guided

Fig. 4 *Some examples showing the difference between guide image filter and ground truth images*

a As ground truth images,
b As guide image filter images,
c Is the result of (a) subtracts (b)

CAAI Trans. Intell. Technol., 2020, Vol. 5, Iss. 2, pp. 128–140

Table 1 Comparison of network structure

Method	Input	Reconstruction	GRL	LRL	Pyramid	Residual	Concatenation
SRCNN	LR + bicubic	direct	—	—	—	—	no concatenation
VDSR	LR + bicubic	direct	✓	—	—	✓	no concatenation
lapSRN	LR	progressive	✓	—	✓	✓	deep concatenation
EDSR	LR	direct	—	✓	—	✓	deep concatenation
DBPN	LR + bicubic	direct	—	✓	—	✓	deep concatenation
ours	LR	progressive	✓	—	✓	✓	ConvLSTM concatenation

filtering with real HR images to jointly train the network. Such as formula (11). When the convolutional network learns the mapping relationship between LR and real HR images, the mapping relationship that may be learned is not comprehensive. However, we also have guided filtered images, using the same CNN, to learn the mapping relationship between LR and guided filtered images again, making the learned mapping relationship more comprehensive. Real HR images help super-resolution images to enhance detail, while guided filtered images ensure that the edges of super-resolution reconstruction are clear, achieve noise reduction, and reduce artefacts, making the network have higher performance

$$\arg\min_{\Omega} \sum_{(I_*^i, I^*, I^i)\in\varphi} \left\{ \lambda_1 (I^* - I^i)^2 + (1-\lambda_1)(I_*^i - I^i)^2 \right\} \tag{11}$$

where I^* represents the filtered image, I_*^i is the real HR image, I^i is the high- resolution image output by the network, and parameter λ_1 is used to control the weight of the real image and the guided filtered image.

The algorithm in this paper is shown in Algorithm 1.

Algorithm 1: Multi-scale super-resolution reconstruction based on guided image filter

Require: LR image, Real high-resolution I_*^i

i **Input:** Low-resolution Image
ii **Output:** High-resolution Image
iii Initialise Downsample the Low-resolution image to get LR_1, LR_2, LR_3
iv **while** LR_i **do**
v Train the input low-resolution image. According to the (10): $L = \sum_{i=1}^{n} \frac{K_i}{N_i} \| I^i - I_*^i \|_2^2$
vi Network output to be improved high-resolution image
vii **end while**
viii Filter image I^* obtained by *(6)*: $q_i = \sum_j W_{ij}(I) p_j$
ix **while** LR_i **do**
x Joint training with filtered images I^* and real HR I_*^i images by *(11)* :

$$\arg\min_{\Omega} \sum_{(I_*^i, I^*, I^i)\in\varphi} \left\{ \lambda_1 (I^* - I^i)^2 + (1-\lambda_1)(I_*^i - I^i)^2 \right\}$$

xi Output high quality high-resolution image
xii **end while**

4 Experimental analysis

The computer used in the experiment was configured as Intel Core i7-6700K CPU@3.40 GHz, NVIDIA TITAN X GPU, 16 GB RAM, Win10 operating system. Use TensorFlow to build a platform framework [44].

The data set uses DIV2K [45]. Since the image size in the DIV2K is large, the image is first reduced. The DIV2K consists of 800 training images and 100 test images and 100 proof images. Since the 800 training images did not publish real images, our test was selected in 100 verified images. The DIV2K validation set has less data. It also uses four benchmark datasets Set5 [46], Set14 [47], B100 [48], and Urban100 [49] for performance ratios. In addition, the Manga109 dataset was used to perform super-resolution reconstruction of characters in the comics.

The LR image is upsampled using Bicubic to obtain input images of different scales. The variables are initialised using the Xavier method [50]. 2×10^4 iteration of (8). According to experience [29], $\lambda = 0.01$ in (8).

4.1 Model comparison

To verify the effectiveness of the proposed method, it is compared with other traditional methods in both qualitative and quantitative aspects. The comparison methods are: bicubic interpolation, anchor domain method (A+) [51], and deep learning-based SRCNN [33], VDSR [14], LapSRN [52], EDSR [17], DBPN [19], compare these methods with peak signal to noise ratio (PSNR), structural similarity (SSIM) values and visual effects.

As can be seen from Table 1, most CNNs use a recurrent neural network (RNN), which can implement a deep feedforward network in which all layers share the same weight. However, theory and experience show that this kind of learning cannot be preserved for a long time [53]. This article uses LSTM networks of special implicit units to achieve long-term preservation of input. The LSTM network is a special unit called memory cells, similar to accumulators and gated neurons. The LSTM can be used in the encoding and decoding network, which will have a weight connected to itself at the next time step, copying the true value of its state and accumulating external signals, and the LSTM network is then proven to be more efficient than traditional RNNs. Global residual learning (GRL) represents the difference between a network learning real HR image and an upsampled (using bicubic interpolation or learning filter) LR image. Local residual learning (LRL) represents a local hop between intermediate convolutional layers. Connected. Pyramid indicates whether there is a pyramid structure, and Concatenation indicates whether to use cascading.

4.1.1 Model parameters: As can be seen from Fig. 5, this paper proposes a more compromised algorithm. Although the number of parameters is not the least, the super-resolution reconstruction

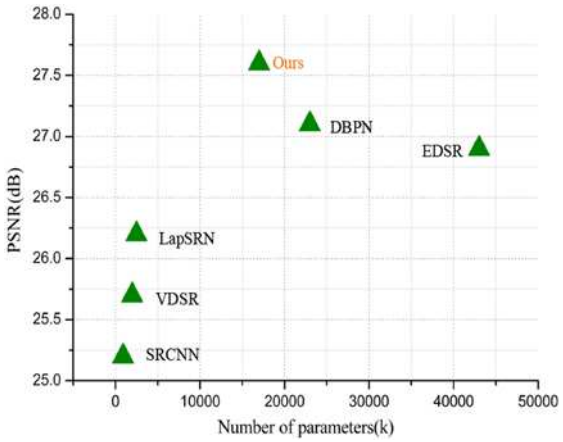

Fig. 5 *Comparison of the number of parameters*

performance is optimal. Compared with the DBPN algorithm, the number of parameters is reduced by 23%, and the number of parameters is reduced by 58% compared with the EDSR algorithm.

4.2 Running time

Comparing the running time of the algorithm with other algorithms, it can be seen from Fig. 6 that the LapSRN algorithm takes the shortest time to run, and the similarity between this paper and the LapSRN algorithm is that this paper also uses the pyramid structure to realise the coarse to fine super-resolution reconstruction task. However, since the algorithm needs to reconstruct the network twice through super-resolution, the running time is long, but the performance of the algorithm is optimal.

4.3 Quantitative analysis

The quantitative comparison was performed using PSNR and SSIM [54] to evaluate the reconstructed image quality. PSNR is based on error-sensitive image quality evaluation, the error between the calculated pixel points, as in (12). When the error between the original HR image and the reconstructed super-resolution image is less, the denominator of (12) is less, and the larger the PSNR value, the better the reconstruction effect

$$\mathrm{PSNR} = 10 \cdot \log_{10} \frac{PQ}{\|y - \hat{y}\|} \tag{12}$$

where P is the HR image size and Q is the LR image size. y is the original HR image and $\hat{y}$ is the SR image.

The SSIM is to evaluate the quality of the super-resolution reconstructed image by measuring the similarity between the original HR image and the reconstructed HR image. SSIM is evaluated in terms of brightness, contrast, and structure. The mean value is used for luminance estimation, the standard deviation is used for contrast estimation, and the covariance is used for measurement of SSIM. As in (13). The larger the SSIM value, the less image distortion. That is, the closer the HR of the original image is to the reconstructed super-resolution image, the better the reconstruction effect

$$\mathrm{SSIM} = \frac{(2\mu_y\mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)} \tag{13}$$

where μ_y is the average grey value of the original HR image and σ_y is the variance of the original HR image. $\mu_{\hat{y}}$ is the average grey value of the reconstructed HR image. $\sigma_{\hat{y}}$ is the variance of the HR image after reconstruction. $\sigma_{y\hat{y}}$ is the covariance of the original image and the reconstructed image. C_1, C_2 is a constant.

Table 2 shows the quantitative test results of different methods on different data sets. From the overall analysis, it can be seen from the data in the table that the PSNR and SSIM indices are gradually increasing on different data sets, indicating that the super-resolution reconstruction method is getting better and better. The PSNR and SSIM of the Set5 data set are relatively higher than other data sets. This is because the Set5 data set is mainly composed of natural scenes, and the super-resolution reconstruction effect is better. The PSNR and SSIM of Urban100 data set are relatively low compared with other data sets. This paper also implements the super-resolution reconstruction of the Manga109 data set. Although there is no comic character image in the training set data, the super-resolution reconstruction effect of the comic image is better, as can be seen from Table 2.

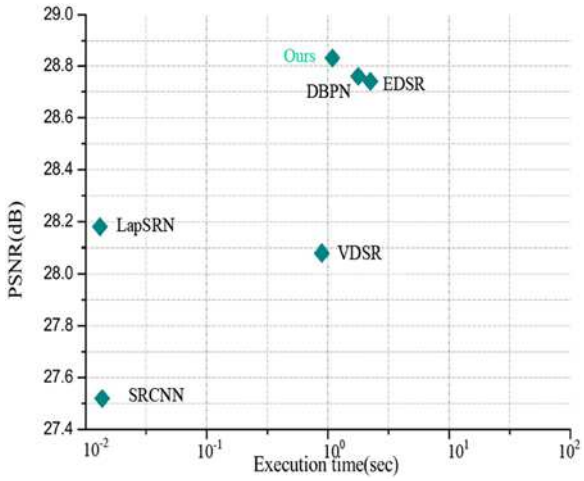

Fig. 6 *Comparison of running time*

Table 2 Compare with the state-of-the-art SR algorithms: average PSNR/SSIM for scale factors ×2, ×4 and ×8

		Set5		Set14		B100		Urban100		Manga109	
Algorithm	Scale	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	2	33.64	0.929	30.05	0.871	29.54	0.845	26.87	0.84	30.84	0.932
A+	2	36.55	0.954	32.42	0.908	31.22	0.887	29.22	0.859	35.32	0.966
SRCNN	2	36.65	0.955	32.97	0.904	31.36	0.888	29.53	0.896	36.61	0.972
VDSR	2	37.53	0.957	32.98	0.914	31.9	0.896	30.76	0.915	37.16	0.975
LapSRN	2	37.52	0.958	33.08	0.915	31.81	0.896	30.41	0.919	37.26	0.973
EDSR	2	***38.11***	***0.962***	**33.93**	***0.919***	***32.32***	***0.902***	**33.55**	**0.934**	39.11	0.977
DBPN	2	38.09	0.961	33.85	0.918	32.26	0.901	33.02	0.931	***39.32***	***0.978***
ours	2	**38.12**	**0.964**	***33.89***	**0.921**	**32.34**	**0.905**	***33.51***	***0.932***	**39.35**	**0.979**
Bicubic	4	28.42	0.81	26.01	0.702	25.94	0.843	23.14	0.656	24.93	0.788
A+	4	30.28	0.86	27.32	0.749	26.82	0.709	24.32	0.718	27.03	0.852
SRCNN	4	30.48	0.862	27.49	0.75	26.92	0.71	24.52	0.722	27.67	0.859
VDSR	4	31.35	0.883	28.01	0.767	27.29	0.725	25.21	0.756	28.83	0.887
LapSRN	4	31.54	0.885	28.19	0.772	27.32	0.728	25.21	0.756	29.49	0.891
EDSR	4	32.45	0.895	28.81	0.786	27.71	0.741	***27.29***	***0.802***	31.42	0.915
DBPN	4	***32.46***	***0.898***	***28.82***	***0.788***	***27.72***	***0.743***	27.08	0.795	***31.51***	***0.916***
ours	4	**32.48**	**0.899**	**28.83**	**0.789**	**27.74**	**0.745**	**27.31**	**0.805**	**31.43**	**0.918**
Bicubic	8	24.38	0.656	23.17	0.567	23.68	0.545	20.73	0.514	21.46	0.645
A+	8	25.51	0.691	23.97	0.596	24.18	0.567	21.38	0.546	22.36	0.679
SRCNN	8	25.32	0.687	23.86	0.594	24.11	0.564	21.31	0.545	22.37	0.683
VDSR	8	25.71	0.71	24.22	0.611	24.36	0.575	21.55	0.561	22.38	0.672
LapSRN	8	26.15	0.739	24.43	0.626	24.54	0.587	21.8	0.581	23.38	0.735
EDSR	8	26.97	0.775	24.95	0.641	24.79	0.595	23.11	0.629	24.56	0.778
DBPN	8	***27.21***	***0.782***	***25.13***	***0.649***	***24.89***	***0.603***	**23.26**	**0.623**	**25.53**	**0.801**
ours	8	**27.23**	**0.783**	**25.16**	**0.651**	**24.91**	**0.605**	***23.24***	***0.621***	***25.49***	***0.799***

CAAI Trans. Intell. Technol., 2020, Vol. 5, Iss. 2, pp. 128–140

Analyse the data in Table 2. Bold indicates the best performance and bold italic indicates the second. It can be seen from the table that this paper is better for small-scale super-resolution reconstruction, better than DBPN, EDSR and other methods. For large-scale super-resolution reconstruction, the proposed network outperforms some data sets. The most advanced DBPN method, but did not make a significant breakthrough in the reconstruction of the Urban100 dataset. For the Set5 and Set14 datasets with simple structure, the performance of the proposed algorithm is better than t state-of-the-art SR algorithms.

The PSNR and SSIM values are analysed. Comparing the A+ method with Bicubic, the PSNR index is improved by 1.5–3.0 dB, and it can be seen that the performance has been significantly improved. The SRCNN used the method of CNN to perform super-resolution reconstruction. However, due to the simple network structure and fewer features extracted, the super-resolution reconstruction effect is slightly improved, and the increase is not large. The PSNR index has increased and the range is between 0.1 and 0.57 dB. The VDSR method uses the residual network method to introduce the residual network into the field of super-resolution reconstruction. The PSNR of VDSR is higher than that of SRCNN, and the lifting range is 0.31–0.88 dB, which improves the performance of super-resolution reconstruction.

Therefore, the network model of super-resolution reconstruction is moving in a deeper direction. Unlike previous networks, LapSRN uses pyramidal progressive up-sampling and cascades the features extracted at each level to achieve super-resolution reconstruction. The EDSR method uses dense blocks and uses a residual network for jumping connection, which increases the depth of the network, and the network depth reaches 32 layers, and more features are extracted, so the super-resolution reconstruction effect is better. DBPN adopts the method of up-and-down iterative sampling, using the upper and lower projection units, the network structure is more complicated, but the reconstruction effect is better. When the method proposed in this paper is used for large-scale super-resolution reconstruction, the reconstruction effect on some data sets is not as good as DBPN. However, compared with DBPN, the proposed network structure is simple, the number of parameters is small, and the running speed is fast. Compared with complex network structures such as DBPN and EDSR, this paper has certain advantages.

The PSNR, SSIM, and IFC [55] after super-resolution reconstruction are compared in the DIV2K data set. The experimental results are shown in Table 3. It can be seen from the table that the algorithm is superior to most of the comparison algorithms and obtains better experimental results.

4.4 Qualitative analysis

Table 2 shows the quantitative comparison of different data sets under different methods. The values of PSNR and SSIM are analysed to evaluate the results of super-resolution reconstruction. However, since the numerical values do not fully explain the problem, the following figure further studies the super-resolution reconstruction results through qualitative comparison.

In Figs. 7–9, the test image is extracted on the data set, wherein the red box represents the selected position, and the selected positions are amplified by different scale factors by different methods, Each method corresponds to the magnification of the selected location and the overall super-resolution reconstruction. Compare the experimental results of different methods. Fig. 7*a* is an extraction of the 'img-002' image in the Set5 data set, and the magnification factor is 2. Super-resolution reconstruction results were evaluated by the visual effects of the parrot's eye. It can be seen from the figure that compared with the original HR, the Bicubic method has the worst visual effect, and the reconstructed image artefacts and edge blurring are more serious, and the texture and detail information of the image cannot be clearly displayed; Compared with the Bicubic algorithm, the A+ algorithm improves the edge blur, but there are still more serious artefacts. The reconstruction ability of the image details is poor, and the visual effect is not good. The SRCNN after the A+ algorithm introduces a CNN into the field of super-resolution reconstruction. It can be seen from the experimental results that the SRCNN algorithm has a great improvement on the processing of artefacts and edge blur conditions compared with the previous methods. However, still cannot meet the human visual requirements, because the reconstructed image also has artefacts. This is because the SRCNN network structure is relatively simple, feature extraction is not sufficient, and the available features are less, resulting in unsatisfactory reconstruction results. However, a problem has been explained by the SRCNN network, that is, the learning-based super-resolution reconstruction method is effective for super-resolution reconstruction.

VDSR uses global residuals, which not only deepens the depth of the network, but also improves the convergence speed of the network. Super-resolution reconstruction is superior to SRCNN. It can be seen from the reconstructed image that VDSR is better than SRCNN for detail reconstruction, but there is still a gap between the result and the actual HR image. This is because both SRCNN and VDSR use the dual method to predefined up-sampling the image. This method is easy to introduce noise and affect the quality of image reconstruction. Therefore, the reconstructed image is prone to artefacts. LapSRN uses progressive up-sampling to avoid noise from predefined up-sampling. The LapSRN network adopts two branches, one branch extracts features and the other branch performs image reconstruction. By extracting features of different scale images, and the reconstruction ability of high-frequency features is improved. The EDSR method uses a large number of dense blocks to extract different depth features. The super-resolution reconstruction is better, but the number of network parameters is larger and the running time is longer. The DBPN algorithm uses dense projection units, and the super-resolution reconstruction is better. The algorithm is compared with the DBPN algorithm. The reconstructed image achieves similar visual effects, but the quantitative analysis of the experimental results is better than the DBPN algorithm.

Compared with performing $\times 2$ magnification, the difficulty of performing $\times 4$ magnification has increased. It can be seen from the enlarged result in Fig. 8 that the image after super-resolution reconstruction is prone to artefacts and the details are blurred. The improved method mainly extracts more effective features by convolution, thus updating the LR to HR non-linear mapping relationship. As the algorithm continues to improve and extract

Table 3 Compare with the state-of-the-art SR algorithms on the DIV2K data set: average PSNR/SSIM/IFC for scale factors $\times 2$, $\times 4$ and $\times 8$

	$\times 2$			$\times 4$			$\times 8$		
Algorithom	PSNR	SSIM	IFC	PSNR	SSIM	IFC	PSNR	SSIM	IFC
Bicubic	32.42	0.903	6.345	28.12	0.756	3.643	25.17	0.246	0.837
A+	34.55	0.934	8.312	29.27	0.807	3.005	25.93	0.687	1.024
SRCNN	34.58	0.931	7.141	29.32	0.809	2.637	26.06	0.692	0.975
VDSR	35.44	0.942	8.391	29.81	0.823	3.005	26.22	0.698	1.062
LapSRN	35.32	0.941	8.586	29.89	0.826	3.131	26.12	0.699	1.115
EDSR	35.73	0.946	8.921	30.12	0.853	3.356	26.53	0.726	1.265
DBPN	**36.01**	**0.953**	**9.436**	30.43	0.887	3.632	26.72	0.784	1.308
ours	35.98	0.951	9.241	**30.61**	**0.905**	**3.814**	**26.93**	**0.816**	**1.399**

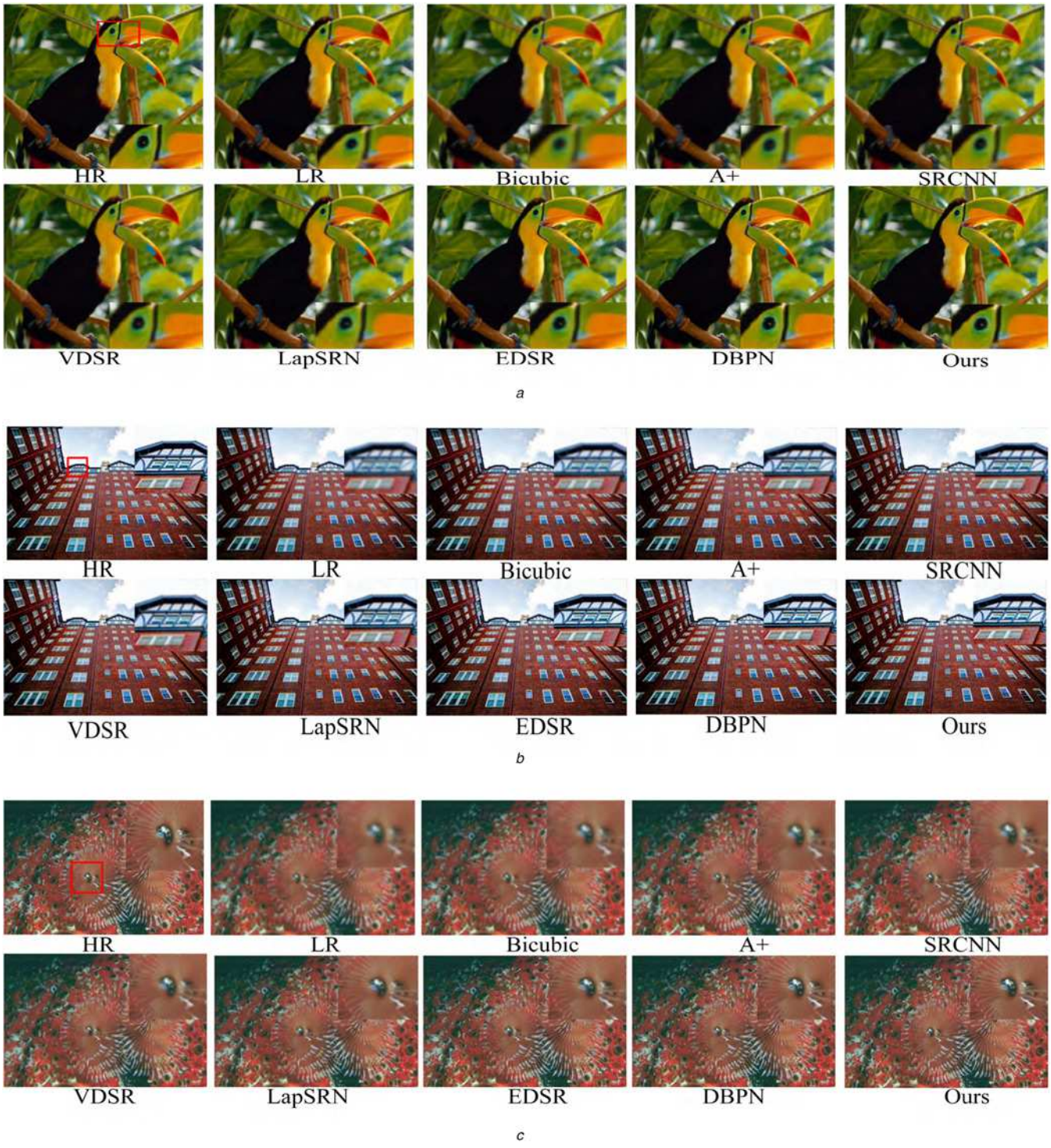

Fig. 7 *Super-resolution reconstruction with a scale factor of 2*

a Set5 img-002,
b Urban100 img-034,
c B100 12084

more effective features, the effect of smaller-scale super-resolution reconstruction is getting better and better.

For larger scale super-resolution reconstruction tasks, the performance of some methods may decrease. The PSNR and SSIM indices in Table 2 can be verified. This method may not be as good as the DBPN algorithm for large-scale reconstruction tasks on some data sets, but the reconstruction effect on small scale is better than DBPN algorithm. The DBPN network adopts up and down iterative sampling, and the projection unit can extract HR error features in LR and extract LR error features in HR to realise self-correction mechanism. However, since the network uses multiple projection units, the network structure is relatively complicated, the number of parameters is large, and the requirements for hardware devices are high. Comparing the algorithm of this paper with the DBPN network, this paper does not use multiple projection units, but uses a multi-scale method. DBPN uses multiple projection units to extract features from different depths of the image. Multiple convolutional layers and deconvolution layers are required, but multiple successive convolution and deconvolution operations are prone to internal

CAAI Trans. Intell. Technol., 2020, Vol. 5, Iss. 2, pp. 128–140

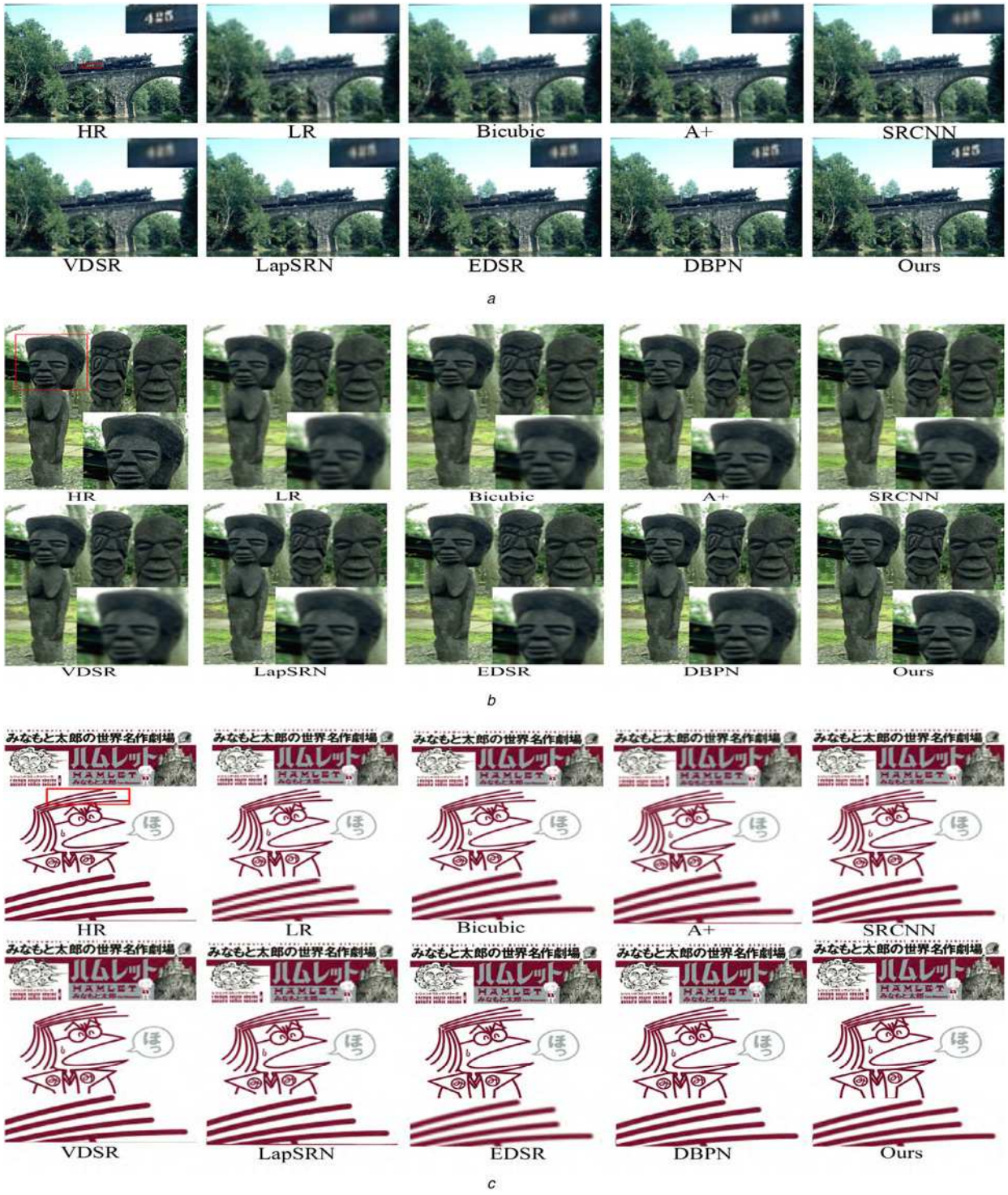

Fig. 8 *Super-resolution reconstruction with a scale factor of 4*

a B100 img-069,
b B100 img-001,
c Manga109 Hamlet

covariate migration [53], thereby affecting the results of feature extraction. It can be seen from Fig. 8*b* that the reconstructed image of the DBPN algorithm still has artefacts, and there is a certain degree of ambiguity at the edges, and the reconstruction ability of the details is still lacking. Compared with the algorithm in this paper, the super-resolution image of this paper has fewer artefacts, clear details and texture.

Since the images in the Urban100 dataset are mostly composed of urban photos, there are many self-similar structures in the image, and the shooting distance is far, so the super-resolution reconstruction is difficult. The PSNR and SSIM values of the 'img-092' image reconstruction result in the Urban100 data set are analysed. It can be seen from Fig. 10 that the reconstruction performance of the large-scale factor is not significantly improved compared with the DBPN method, but it is better than the EDSR algorithm. For different data sets, the network performance is different. Through experimental analysis, the proposed algorithm is superior to most of the comparison methods and has certain superiority.

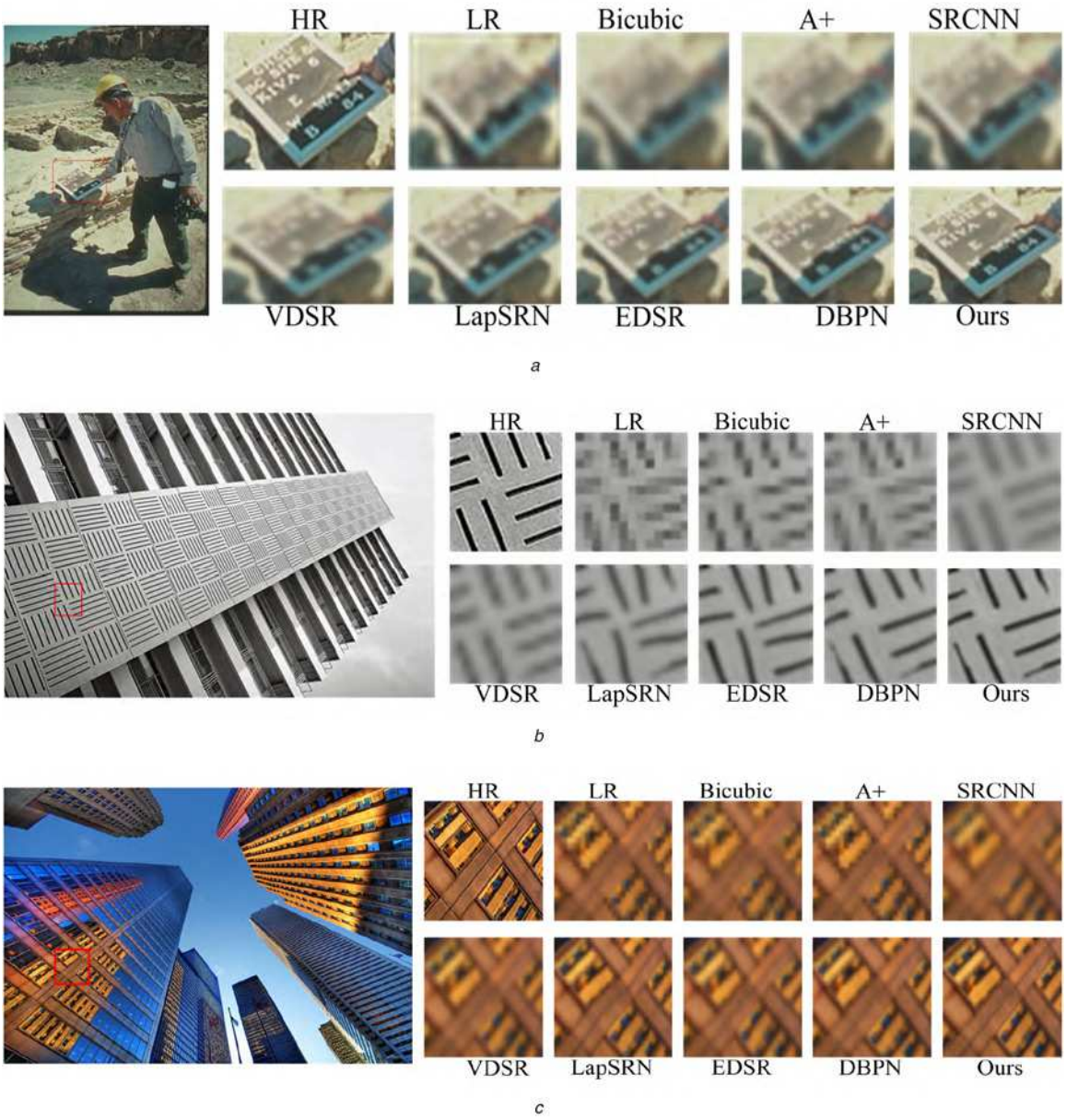

Fig. 9 *Super-resolution reconstruction with a scale factor of 8*

a B100 img-099,
b B100 img-092,
c Urban img-012

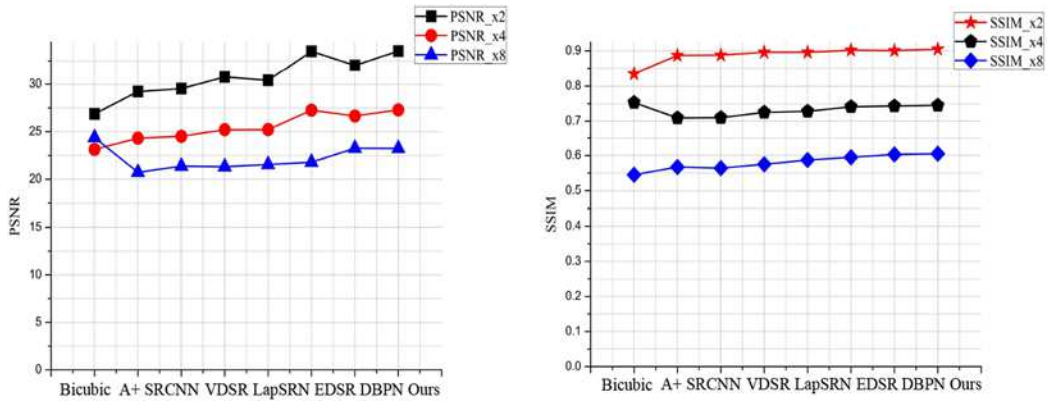

Fig. 10 *'img-092' scale up the different scale factors and compare the PSNR index with the SSIM index. PSNR index on the left and SSIM index on the right*

CAAI Trans. Intell. Technol., 2020, Vol. 5, Iss. 2, pp. 128–140

5 Conclusion

The traditional super-resolution reconstruction network is to input a single LR image, and the network is subjected to feature extraction through different levels of convolution kernels. In order to extract more effective features, the number of convolution layers is larger and the network depth is deeper. In order to prevent the gradient disappearance or gradient explosion caused by the increase of network depth, a jump connection occurs. At the same time, in order to extract the features of different layers as much as possible, the structure of the dense residual block is used. Generally, the network with the jump connection and the dense residual block has a deep network depth, the network structure is complex, and the parameters are many. Different from the traditional CNN, the network uses multiple sub-networks to take images of different scales of LR images as input of sub-networks, and then perform super-resolution reconstruction tasks at this scale separately in each sub-network, using sequential connections. The LR to HR features learned at different scales are concatenated by ConvLSTM, and finally the super-resolution reconstruction results of the original LR images are output. The network structure is relatively simple, with fewer parameters and faster convergence.

The multi-scale super-resolution method based on guide image filter proposed in this paper. The coding–decoding network is used to learn the features of different scales, and the network is trained with the real HR image features. The experimental results are improved in PSNR index and SSIM. In the future work, try new convolutional networks and replace the ConvLSTM network in multi-scale networks with new convolutional networks to make the results of super-resolution reconstruction more accurate.

6 Acknowledgments

The authors acknowledge the National Natural Science Foundation of China (grant nos. 61772319, 61976125, 61873177 and 61773244), and the Shandong Natural Science Foundation of China (grant no. ZR2017MF049).

7 References

[1] Yang, C.Y., Ma, C., Yang, M.H.: 'Single-image super-resolution: A benchmark'. The IEEE Conf. on European Conf. on Computer Vision, Cham, 2014, pp. 372–386
[2] Baker, S, Kanade, T.: 'Limits on super-resolution and how to break them', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, **24**, (9), pp. 1167–1183
[3] Zhang, L, Zhang, H, Shen, H, *et al.*: 'A super-resolution reconstruction algorithm for surveillance images', *Signal Process.*, 2010, **90**, (3), pp. 848–859
[4] Thornton, M.W., Atkinson, P.M., Holland, D.A.: 'Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping', *Int. J. Remote Sens.*, 2006, **27**, (3), pp. 473–491
[5] Fan, Y., Yu, J., Huang, T.S.: 'Wide-activated deep residual networks based restoration for BPG-compressed images'. The IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 2621–2624
[6] Liu, D., Wang, Z., Fan, Y., *et al.*: 'Robust video super-resolution with learned temporal dynamics'. Proc. of the IEEE Int. Conf. on Computer Vision, Venice, Italy, 2017, pp. 2526–2534
[7] Yang, J., Wright, J., Huang, T.S., *et al.*: 'Image super-resolution via sparse representation', *IEEE TIP*, 2010, **19**, (11), pp. 2861–2873
[8] Duchon, C.E.: 'Lanczos filtering in one and two dimensions', *J. Appl. Meteorol.*, 1979, **18**, (8), pp. 1016–1022
[9] Denton, E., Chintala, S, Szlam, A, *et al.*: 'Deep generative image models using a Laplacian pyramid of adversarial networks'. Advances in Neural Information Processing Systems, Montreal, QC, Canada, 2015, pp. 1486–1494
[10] Huang, G., Liu, Z., Weinberger, K.Q.: 'Densely connected convolutional networks'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Hawaii, HI, USA, 2017, pp. 4700–4708
[11] Dong, C., Loy, C.C., He, K., *et al.*: 'Image super-resolution using deep convolutional networks', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2016, **38**, (2), pp. 295–307
[12] Dong, C., Loy, C.C., He, K., *et al.*: 'Accelerating the super-resolution convolutional neural network'. European Conf. on Computer Vision, Amsterdam, Holland, 2016, pp. 391–407
[13] Timofte, R., Rothe, R., Gool, L.V.: 'Seven ways to improve example-based single image super resolution'. IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 1865–1873
[14] Kim, J., Lee, J.K, Lee, K.M.: 'Accurate image super-resolution using very deep convolutional networks'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 1646–1654
[15] Bengio, Y., Simard, P., Frasconi, P.: 'Learning long-term dependencies with gradient descent is difficult', *IEEE Trans. Neural Netw.*, 2002, **5**, (2), pp. 157–166
[16] Ledig, C., Theis, L., Huszar, F., *et al.*: 'Photo-realistic single image super-resolution using a generative adversarial network'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Hawaii, HI, USA, 2017, pp. 4681–4690
[17] Lim, B., Son, S., Kim, H., *et al.*: 'Enhanced deep residual networks for single image super-resolution'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops, Hawaii, HI, USA, 2017, pp. 136–144
[18] Yu, J., Fan, Y., Yang, J., *et al.*: 'Wide activation for efficient and accurate image super-resolution', 2018, arXiv:1808–08718
[19] Haris, M., Shakhnarovich, G., Ukita, N.: 'Deep back-projection networks for super-resolution'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 1664–1673
[20] Goodfellow, I., Pouget-Abadie, J., Mirza, M., *et al.*: 'Generative adversarial nets'. Advances in Neural Information Processing Systems, Montreal, QC, Canada, 2014, pp. 2672–2680
[21] Johnson, J., Alahi, A., Fei-Fei, L.: 'Perceptual losses for real-time style transfer and super-resolution'. European Conf. on Computer Vision, Switzerland, Zurich, 2016, pp. 694–711
[22] Sajjadi, M.S.M., Bernhard, S., Hirsch, M.: 'Enhancenet:single image super-resolution through automated texture synthesis'. IEEE Conf. on Computer Vision and Pattern Recognition, Hawaii, HI, USA, 2017, pp. 4491–4500
[23] Chen, T.Q., Schmidt, M.: 'Fast patch-based style transfer of arbitrary style', 2016, arXiv: 1612–04337
[24] Gatys, L.A., Ecker, A.S., Bethge, M.: 'Image style transfer using convolutional neural networks'. IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 2414–2423
[25] Liu, C, Sun, D.: 'A Bayesian approach to adaptive video super resolution'. IEEE Conf. on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 2011, pp. 209–216
[26] Zheng, H., Ji, M., Wang, H., *et al.*: 'Crossnet: an end-to-end reference-based super resolution network using cross-scale warping'. European Conf. on Computer Vision, Munich, Germany, 2018, pp. 88–104
[27] Zhu, Y., Zhang, Y., Yuille, A. L.: 'Single image super-resolution using deformable patches'. IEEE Conf. on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 2917–2924
[28] Yue, H., Sun, X., Yang, J., *et al.*: 'Landmark image super-resolution by retrieving web images', *IEEE Trans. Image Process.*, 2013, **22**, (12), pp. 4865–4878
[29] Tao, X., Gao, H., Liao, R., *et al.*: 'Detail-revealing deep video super-resolution'. Proc. of the IEEE Int. Conf. on Computer Vision, Venice, Italy, 2017, pp. 4472–4480
[30] Hochreiter, S, Schmidhuber, J.: 'Long short-term memory', *Neural Comput.*, 1997, **9**, (8), pp. 1735–1780
[31] Dong, C., Loy, C.C., He, K., *et al.*: 'Learning a deep convolutional network for image superresolution'. European Conf. on Computer Vision, Switzerland, Zurich, 2014, pp. 184–199
[32] Wang, Z., Liu, D., Yang, J., *et al.*: 'Deep networks for image super-resolution with sparse prior'. IEEE Int. Conf. on Computer Vision, Santiago, Chile, 2015, pp. 370–378
[33] Shi, W., Caballero, J., Huszár, F., *et al.*: 'Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network'. IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 1874–1883
[34] Boominathan, V, Mitra, K, Veeraraghavan, A.: 'Improving resolution and depth-of-field of light field cameras using a hybrid imaging system'. Int. Conf. on Computational Photography, Santa Clara, CA, USA, 2014, pp. 1–10
[35] Wu, J., Wang, H., Wang, X., *et al.*: 'A novel light field super-resolution framework based on hybrid imaging system'. Vis. Commun. Image Process., Singapore, 2015, pp. 1–4
[36] Zheng, H., Guo, M., Wang, H., *et al.*: 'Combining exemplar-based approach and learning-based approach for light field super-resolution using a hybrid imaging system'. Computer Vision and Pattern Recognition, Hawaii, HI, USA, 2017, pp. 2481–2486
[37] He, K., Jian, S.: 'Fast guided filter', 2015, arXiv preprint arXiv:1505.00996
[38] Ronneberger, O., Fischer, P., Brox, T.: 'U-net: convolutional networks for biomedical image segmentation', *MICCAI*, 2015, pp. 234–241
[39] Sutskever, I., Vinyals, O., Le, Q.V.: 'Sequence to sequence learning with neural networks'. Advances in Neural Information Processing Systems, Montreal, QC, Canada, 2014, pp. 3104–3112
[40] Graves, A.: 'Generating sequences with recurrent neural networks', 2013, arXiv:1308.0850
[41] Xingjian, S., Chen, Z., Wang, H., *et al.*: 'Convolutional LSTM network: a machine learning approach for precipitation nowcasting'. Advances in Neural Information Processing Systems, Montreal, QC, Canada, 2015, pp. 802–810
[42] Pascanu, R., Mikolov, T., Bengio, Y.: 'On the diffificulty of training recurrent neural networks'. Int. Conf. on Machine Learning, Paris, France, 2013, pp. 1310–1318
[43] Tao, X., Gao, H., Shen, X., *et al.*: 'Scale-recurrent network for deep image deblurring'. IEEE CVPR, Salt Lake City, UT, USA, 2018, pp. 8174–8182
[44] Abadi, M, Agarwal, A, Barham, P, *et al.*: 'Tensorflow: large-scale machine learning on heterogeneous distributed systems', Software available from tensorflow. 2016, arXiv preprint arXiv:1603.04467

[45] Timofte, R., Agustsson, E., Van Gool, L., *et al*.: 'Ntire 2017 challenge on single image super-resolution: methods and results'. Computer Vision and Pattern Recognition Workshops, Hawaii, HI, USA, 2017, pp. 114–125
[46] Bevilacqua, M., Roumy, A., Guillemot, C., *et al*.: 'Low-complexity single-image super-resolution based on nonnegative neighbor embedding'. Proc. of the British Machine Vision Conf., Guildford, UK, 2012, pp. 135.1–135.10
[47] Zeyde, R., Michael, E., Matan, P.: 'On single image scale-up using sparse-representations'. Int. Conf. on Curves and Surfaces, Berlin, Heidelberg, 2010, pp. 711–730
[48] Martin, D.R., Fowlkes, C., Tal, D.: 'A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics'. Proc. of the IEEE Int. Conf. on Computer Vision, Vancouver, BC, Canada, 2002, pp. 1110–1121
[49] Huang, J.B., Singh, A., Ahuja, N.: 'Single image super-resolution from transformed self-exemplars'. IEEE CVPR, Boston, MA, USA, 2015, pp. 5197–5206
[50] Sergey, I., Christian, S.: 'Batch normalization: 'accelerating deep network training by reducing internal covariate shift'. Int. Conf. on Machine Learning, Lile, France, 2015, pp. 448–456
[51] Timofte, R., De Smet, V., Van Gool, L.: 'A+: adjusted anchored neighborhood regression for fast super-resolution'. Asian Conf. on Computer Vision, Singapore, 2014, pp. 111–126
[52] Lai, W.S., Huang, J.B., Ahuja, N., *et al*.: 'Deep laplacian pyramid networks for fast and accurate super-resolution'. IEEE CVPR, Hawaii, HI, USA, 2017, pp. 624–632
[53] Lecun, Y., Bengio, Y., Hinton, G.: 'Deep learning', *Nature*, 2015, **521**, (7553), p. 436
[54] Wang, Z., Bovik, A.C., Sheikh, H.R., *et al*.: 'Image quality assessment: from error visibility to struc-tural similarity', *Image Process. IEEE Trans.*, 2004, **13**, (4), pp. 600–612
[55] Sheikh, H.R., Bovik, A.C., De Veciana, G.: 'An information fidelity criterion for image quality assessment using natural scene statistics', *IEEE Trans. Image Process.*, 2005, **14**, (12), pp. 2117–2128