Image Super-Resolution Based on Additional Self-Loop Supervision

Hong Chen\textsuperscript{1}, Li Xu\textsuperscript{2*}, Bing Luo\textsuperscript{1}, Shuai Du\textsuperscript{1}, Kunshu Xiao\textsuperscript{3} and Siwei Li\textsuperscript{3}

\textsuperscript{1}School of Computer and Software Engineering, University of Xihua, Chengdu, China
\textsuperscript{2}School of Science, University of Xihua, Chengdu, China
\textsuperscript{3}Xihua College of Xihua University, University of Xihua, Chengdu, China
Email: xuli19860715@163.com

Abstract. Relying on the prevalence of deep neural networks, only one image super-resolution (SR) research has taken a big step forward. Many network models use the difference between the real image and the generated super-resolution image as a loss function, ignoring the difference between the small-scale image and the super-resolution generated image. We adopt an end-to-end network model with two-way supervision, which can ensure that the image content is similar in both large and small scales. Only real natural images are used for supervision. As the scale of the image becomes larger, the difference becomes larger, and the effect of correcting image details becomes smaller. Adding the loss between the input image and the degraded image of the generated image can not only keep the content of the image similar on a small scale, but also ensure the image details. Compared with several state-of-the-arts, our method obtains the best results.

Keywords. Superpixel segmentation; convolutional neural network; image superresolution.

1. Introduction

Modern society is gradually infiltrated by the Internet, and digital images are focused on as a visual information carrier. Single image SR research is particularly important in digital imaging systems, that is, to recover a clear and natural high-resolution image from a low-resolution image, but it is a typical ill-conditioned problem. In recent years, due to the strong nonlinear representation capabilities of deep convolutional neural networks (CNN), it directly learns the end-to-end mapping from low-resolution images (LR) to high-resolution images (HR).

Image SR can be divided into supervised degraded image super-resolution and unsupervised degraded image super-resolution according to the degradation principle. Supervised degradation is to first assume an image degradation model, which simulates how high-resolution images in a natural environment degenerate into low-resolution images observed by the human eye, and then go through the reverse process to obtain SR results. There are two commonly used models. One is to first convolve a high-resolution image with a fuzzy kernel, and then use a bicubic down sampler with a scaling factor to act on the convolved high-resolution image to obtain a small size and low resolution. Rate image, and finally add an additive white Gaussian noise (AWGN). For instance, Yang et al proposed a dictionary-based SR method, and Dong et al. \cite{1} pioneered a three-layer convolutional network (SRCNN), which laid the foundation for a CNN-based method, FSRCNN is an improvement of the SRCNN network and a model based on strong supervised learning. The proposal of the residual network (ResNet) solves the problem that the previous network structure cannot be trained when the network structure is relatively deep. For example, DRCN \cite{2} introduces the existing recurrent neural network structure. Tai et al.
analyzed the network structure of ResNet, VDSR [3], and DRCN, and proposed a deeper network DRRN [4]. Li et al. proposed a feedback network SRFBN [5] that can feed back useful information from the next layer to the previous layer. Wang et al. [6] proposed a spatial feature modulation layer, using the segmentation mask of the image as the prior feature condition for SR. Zhang et al. used another simplified model to directly pass the high-resolution image through the bicubic down-sampler to obtain the degraded image.

Another type of unsupervised degradation is proposed to imitate the real super-resolution process from low resolution to high resolution. From the study of the nature of SR tasks, there are still differences between the model and the real super-resolution mechanism in real life. In response to this problem, the generative confrontation network proposed by Goodfellow et al. [7] played a role. As a result, Kupyn et al. [8] realized a method based on GAN blind motion deblurring, that is, without information about the blur kernel, given a blurred image, restore a clear image. Bulat et al. proposed that the super-resolution model of image degradation needs to be learned first, the “ZSSR” proposed by Shocher et al. does not rely on any other image samples and pre-training, and uses the internal self-similar information of the image to train a specific image network.

In addition to with regard to increasing network depth and parameters, other networks, such as NLRN [9] and RCAN [10], also improve performance in view of feature correlation in in dimensional or channel size. Different network structures will design loss functions for training optimization. The existing methods mainly use real images (GT) and generated images (SR) to compare pixel-level loss, which can only be used for targets with strong supervision. Johnson et al. proposed perceptual loss [11] to compare high-level perception and semantic differences between images.

The modern CNN-based SR models still turn towards some limitations. Most CNN-based SR methods do not utilize of the original LR image information full, resulting in relatively low performance. In order to solve this problem, we propose a new network structure to achieve more powerful function expression and performance learning. Specifically, we propose a low-resolution image loss to better constrain the end-to-end model. This mechanism enables our network to make better use of LR information. By designing a new network module (DB) to degenerate the SR image we generated into a new small image LSR, the end-to-end learning network adopts a two-way supervision mode to optimize the network to ensure the restoration of image details. In addition, the recursive structure reduces network parameters and deepens the network depth. As shown in figure 1, compared with other excellent SR methods, our method can obtain better visual quality and restore more image details.

![Figure 1](image-url)

**Figure 1.** Visually compare the SR (x4) results of the image Urban100:img004.

### 2. Related Work

The study of the SR of a single image is the inverse process of learning the artificially designed image degradation process. In the super-resolution data pre-processing stage of a single image, the image pair
of LR-HR needs to be formed first. Usually, the high-resolution image is reduced to a low-resolution image by the bicubic downsampling method, and Gaussian noise is added to simulate the real low-resolution image. Resolution image, this method is the main data preprocessing method in the current image super-resolution research.

2.1. Super-Resolution Methods Based on Residual Network
The emergence of residual networks has brought a better development to image super-resolution, because the target image and the input image have a lot of similar information, so only the residual part between the HR image and the LR image needs to be learned, so the model complexity and the difficulty of learning can be greatly reduced. Kim proposed a deeper network VDSR based on the residual structure. DRCN uses the idea of residual learning, namely Skip-Connection, to deepen the network structure and increase the network experience field. Lai et al. proposed LapSRN [12], through step-by-step upsampling, one-level prediction residuals, when doing high-power upsampling, the output of intermediate low-power upsampling results can also be obtained. SRDenseNet [13] is proposed by Tong et al. It combines low-level features and high-level features through skip links to improve the performance of super-resolution reconstruction. Although the use of faster and deeper convolutional neural networks has made breakthroughs in the accuracy and speed of single-image super-resolution, the generated images often cannot satisfy our visual perception and lack detailed information. Ledig et al. proposed a neural network the network’s perceptual loss optimization algorithm SRGAN to improve the authenticity of the image. By stacking the modified residual blocks, Lim et al. proposed a network EDSR with average depth and breadth [6], a multi-scale deep super-resolution system (MDSR) sampled from different multiples to increase network parameters. Its performance has also been improved. The significant performance improvement shows that the depth of the representation network plays a key role in image SR. Other state-of-the-art methods such as MemNet [14] and RDN [15] form deep networks based on dense blocks and focus on utilizing all hierarchical features from all convolutional layers. Hui et al. proposed the structure of the information distillation block [16].

2.2. Super-resolution Methods Based on Recursive Structure
Although the increase in network depth and parameters can improve performance, it will take up huge storage resources and There is a problem of exceeding the storage capacity, for which recursive structure is often used. The recurring neural network (RNN) is an extension of the time dimension, representing the transmission and accumulation of information from the time dimension to the rear dimension. Based on the preliminary probability of information, the neural network is represented as a hidden layer of neural network, and a hidden layer of neural network is output before entry. A single image super-resolution task can be composed of two parts, one is the encoder and the other is the decoder. The required image features are obtained by encoding the image. The higher-level semantics depends on the underlying semantics and their combination. DRCN applies the existing RNN to super-resolution processing. Since the convolution weights are shared, the amount of parameters will be greatly reduced. Similar to DRCN, DRRN uses recursive learning, that is, multiple copies of a basic jump connection block to achieve multi-path network area, and memory overhead and computational complexity have been significantly reduced. Tai et al. proposed a new image super-resolution persistent memory network (MemNet for short). Li et al. used a backhaul mechanism to improve the super-resolution effect, which can deepen the network without introducing new parameters.

3. The Proposed Method
The study of the SR of a one image is the inverse process of learning the artificially designed image degradation process. In the super-resolution data pre-processing stage of a single image, the image pair of LR-HR needs to be formed first. Usually, the high-resolution image is reduced to a low-resolution image by the bicubic downsampling method, and Gaussian noise is added to simulate the real low-resolution image. this method is the main data preprocessing method in the current image super-resolution research.
As shown in figure 2, our proposed network structure is improved on the basis of SRFBN, adding a degenerate module DB, which is composed of three convolutional layers. Degenerate the generated SR image through the DB module to obtain a new small-scale LSR image, calculate the loss of LR and LSR, and calculate the total loss jointly with the real image and SR image to optimize the model.

Next, the network is divided into three parts to explain, feature extraction module, feedback module (FB), image reconstruction module. The first is the feature extraction module, which extracts the feature $f_{LR}$ from the LR image, expressed by the following formula:

$$f_{LR} = B_{FE}(LR)$$  \hspace{1cm} (1)

In the above formula, $B_{FE}$ represents the feature extraction function, and then the feature is reused through a recursive structure, and the result of the previous iteration is added to the next iteration process to provide advanced information. This process is expressed in mathematical formula as:

$$f_{t+1} = B_{FE}(f_{LR}, f_t)$$  \hspace{1cm} (2)

where $B_{FB}$ represents the feedback module FB, in which dense jump connections are used for effective cross-level information fusion to restore better SR images.

Finally, the image is reconstructed through the image reconstruction block. The module includes a deconvolution layer and a convolution layer. After the reconstruction block, the feature map is enlarged to obtain a high-frequency image residual image, which can be expressed as:

$$Res = B_{RB}(f_T)$$  \hspace{1cm} (3)

Here $f_T$ is the result of the last iteration, $T$ is the number of iterations, and $B_{RB}$ is the image reconstruction module.

$$SR = Res + F_B(LR)$$  \hspace{1cm} (4)

$Res$ represents the residual image we learned from the model, which is the image obtained by direct interpolation of the $LR$ image, and finally we can get our final super-resolution image $SR$. Where $F_B$ is the interpolation function.

4. Loss function

In the field of super-resolution, the loss function is used to measure the recovery error and optimize the guidance model. In the early days, researchers usually used pixel-level L2 loss, but later discovered that it could not measure the reconstruction quality very accurately. Therefore, a combination of multiple loss functions (for example, Perceptual loss, Adversarial loss [17], Texture Loss [18]) is used to better measure the reconstruction error and produce more realistic and higher-quality results. Through 3x3 convolution to reduce the image, the realistic low-resolution input image is generated, and the network model is more in line with the real image super-resolution process.
Figure 3. DB module in network structure.

In this paper, the L1 loss optimization model is used to measure between GT and the generated SR image to obtain $\text{Loss}_1$. In addition, the loss of LR and LSR is combined. Through the degradation module DB we designed, as shown in figure 3 above, the SR image is reduced to obtain LSR, and the difference between LR and LSR is measured to obtain $\text{Loss}_2$. Finally, the total loss is used to perform the network model Optimize learning. The formula is as follows:

$$\text{Loss}_{\text{total}} = \text{Loss}_1 + \text{Loss}_2$$

(5)

Strengthen the loss function and promote the network to learn more realistic super-resolution tasks. As the similarity between the input image and the output image becomes higher and higher, the performance of our model is superior.

5. Experiments

We use DIV2K and Flickr2K datasets to train the network model. The image data preprocessing process uses the down-sampling operation of the natural image GT with a scale factor of 4 to obtain a small-scale image, sends the paired data set to the network for training, and uses Set5 to verify during training.

5.1. Datasets

We conducted tests on public benchmark data sets and compared them with other superior performance algorithms, including Set5, Set14, BSD100, Urban100, DIV2K, Flickr2K and Manga109. Among them, DIV2K and Flickr2K are used to train network models. The other data sets are test sets.

Set5 and Set14 are commonly used data sets, and Set14 contains more categories than Set5. DIV2K consists of 800 images for training and 100 images for testing and verification. Flickr2K has 2650 pictures, including people, animals, landscapes, etc. BSD100 is a data set composed of 100 various images. Urban100 is a relatively new data set, all of which are urban scenes. The Manga109 dataset is a collection of 109 test images in a comic.

5.2. Comparisons

After a large number of experimental comparisons, the superiority of our experimental effect is confirmed. Firstly, test images are prepared, and LR images are generated by performing bicubic down-sampling and degradation operations on each test data set. After a lot of experiments, in order to show the performance of this method, it is compared with other super-resolution methods with superior performance, including: Li et al.’s SRFBN method, Tong et al.’s SRDenseNet method, Haris et al.’s DBPN [19] method, Zhang et al. RDN method, Ai et al.’s RNAN [20] and Lim et al.’s EDSR [21] method. We use PSNR and SSIM two evaluation indicators for objective comparison. Table 1 shows the results of PSNR and SSIM for each method of image x4.

The Urban:img011 image in figure 4 is enlarged and found that the “lines” of the method results of EDSR, DBPN, ERCA, and SRFBN are very blurred, while the method clearly shows the sense of lines. Compared with other methods, our effect is also the best. EDSR generates false tablecloths. The results of DBPN, ERCA, and SRFBN methods only have diagonal stripes on one side, and the other side completely disappears.
Table 1. PSNR and SSIM values of different data sets in different SR methods (x4), where the optimal value has been bolded.

| Method    | Set5  | Set14  | Urban100 | BSD100 | Manga109 |
|-----------|-------|--------|----------|--------|----------|
|           | PSNR/SSIM | PSNR/SSIM | PSNR/SSIM | PSNR/SSIM | PSNR/SSIM |
| SRDenseNet | 32.02/0.893 | 28.50/0.778 | 26.05/0.782 | 27.53/0.734 | 29.49/0.899 |
| DBPN       | 32.42/0.897 | 28.75/0.786 | 26.38/0.794 | 27.67/0.739 | 30.90/0.913 |
| RDN        | 32.47/0.899 | 28.81/0.787 | 26.61/0.821 | 27.72/0.742 | 31.00/0.915 |
| SRFBN      | 32.47/0.898 | 28.81/0.787 | 26.60/0.802 | 27.72/0.741 | 31.15/0.916 |
| EDSR       | 32.48/0.898 | 28.81/0.787 | 26.64/0.803 | 27.72/0.742 | 31.03/0.915 |
| RNAN       | 32.52/0.899 | 28.82/0.787 | 26.67/0.805 | 27.72/0.742 | –/–      |
| Ours       | **32.59/0.900** | **28.89/0.789** | **26.87/0.809** | **27.78/0.742** | **31.46/0.919** |

Figure 4. The top experimental results of different methods. The bottom part shows compare the SR (x4) results of the image Urban100:img011.

As shown in figures 5, on the img002 image in the dataset Set14, we can see from the restoration effect of the “tablecloth” that it is also our method to achieve better results, recovering some black and
white grids, and other methods can only get rough striped tablecloths. Tablecloth black and white checkered shape.

![Figure 5](image)

**Figure 5.** Visually compare the SR (x4) results of the image Set14:img002. On the right is the enlarged effect picture of the green box on the left.

### 6. Conclusion

We propose an end-to-end learning single-image SR method based on two-way supervision. The smaller the difference between the generated low-resolution image LSR and the input image, it means that some details of the current image restoration are more accurate. We designed a degraded network to restore the generated super-resolution image to a low-resolution image LSR. The loss obtained by comparing the input image with the LSR is combined with the loss of the natural image and the generated SR image to correct the network. This method is in the supervision of image details is enhanced at different scales to make the image details clearer. After verifying a large number of data sets, which proves the superiority of our model.

### Acknowledgments

This work was supported in part by National Natural Science Foundation of China (No. 61801398), The Young Scholars Reserve Talents program of Xihua University, The program for Vehicle Measurement, Control and Safety Key Laboratory of Sichuan Province (No. QCCK2019-005) and The Talent plan of Xihua College of Xihua University (No. 2021055).

### References

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

[2] Kim J, Kwon Lee J, Mu Lee K 2016 Accurate image super-resolution using very deep convolutional networks *2016 IEEE Conference on Computer Vision and Pattern Recognition* pp 1646-1654.

[3] Kim J, Lee J K, Lee K M 2016 Deeply-recursive convolutional network for image superresolution *2016 IEEE Conference on Computer Vision and Pattern Recognition* pp 1637-1645.

[4] Tai Y, Yang J, Liu X 2017 Image super-resolution via deep recursive residual network *2017 IEEE Conference on Computer Vision and Pattern Recognition* pp 3147-3155.

[5] Li Z, Yang J, Liu Z, et al. 2019 Feedback network for image superresolution *2019 IEEE Conference on Computer Vision and Pattern Recognition* pp 3867-3876.

[6] Wang X, Yu K, Dong C, et al. 2018 Recovering realistic texture in image super-resolution by deep spatial feature transform *2018 IEEE Conference on Computer Vision and Pattern Recognition* pp 606-615.
[7] Zhang H, Goodfellow I, Metaxas D, et al. 2019 Self-attention generative adversarial networks 2019 International Conference on Machine Learning pp 7354-7363.

[8] Kupyn O, Budzan V, Mykhailych M, et al. 2018 Deblurgan: Blind motion deblurring using conditional adversarial networks 2018 IEEE Conference on Computer Vision and Pattern Recognition pp 8183-8192.

[9] Liu D, Wen B, Fan Y, et al. 2018 Non-local recurrent network for image restoration Advances in Neural Information Processing Systems 2018 1673-1682.

[10] Zhang Y, Li K, Li K, et al. 2018 Image super-resolution using very deep residual channel attention networks 2018 in Proceedings of the European Conference on Computer Vision pp 286-301.

[11] Johnson J, Alahi A and Li F 2016 Perceptual losses for real-time style transfer and super-resolution 2016 in European Conference on Computer Vision pp 694-711.

[12] Lai W S, Huang J B, Ahuja N, et al. 2017 Deep Laplacian pyramid networks for fast and accurate super-resolution 2017 IEEE Conference on Computer Vision and Pattern Recognition pp 624-632.

[13] Tong T, Li G, Liu X, et al. 2017 Image super-resolution using dense skip connections 2017 IEEE International Conference on Computer Vision pp 4799-4807.

[14] Tai Y, Yang J, Liu X, et al. 2017 MemNet: A persistent memory network for image restoration 2017 IEEE International Conference on Computer Vision pp 4539-4547.

[15] Zhang Y, Tian Y, Kong Y, et al. 2018 Residual dense network for image super-resolution 2018 IEEE Conference on Computer Vision and Pattern Recognition pp 2472-2481.

[16] Hui Z, Wang X and Gao X 2018 Fast and accurate single image super-resolution via information distillation network 2018 IEEE Conference on Computer Vision and Pattern Recognition pp 723-731.

[17] Wang X, Yu K, Wu S, et al. 2018 ESRGAN: Enhanced super-resolution generative adversarial networks 2018 in Proceedings of the European Conference on Computer Vision pp 1586-1590.

[18] Zhang Z, Wang Z, Lin Z, et al. 2019 Image super-resolution by neural texture transfer 2019 IEEE Conference on Computer Vision and Pattern Recognition pp 7982-7991.

[19] Haris M, Shakhnarovich G and Ukita N 2018 Deep back-projection networks for superresolution 2018 IEEE Conference on Computer Vision and Pattern Recognition pp 1664-1673.

[20] Ai W, Tu X, Cheng S, et al. 2020 Single image super-resolution via residual neuron attention networks 2020 IEEE International Conference on Image Processing pp 1586-1590.

[21] Lim B, Son S, Kim H, et al. 2017 Enhanced deep residual networks for single image super-resolution 2017 IEEE Conference on Computer Vision and Pattern Recognition pp 136-144.