Deep Generate Residual Similar Feature Networks for Image Super-Resolution

Xian Li\textsuperscript{1, 2}, Jiahuan Zhang\textsuperscript{3}, Minjie Wang\textsuperscript{4} and Gang Xu\textsuperscript{1}
\textsuperscript{1}Institute of Materials Technology and Engineering CAS, Ningbo315201, China
\textsuperscript{2}University of Chinese Academy of Sciences
\textsuperscript{3}School of Information Engineering, Yangzhou University, China
\textsuperscript{4}School of Mechatronic Engineering and Automation, Shanghai University, China
Email: \{lixian,zhangjiahuan,xugang\}@nimte.ac.cn; wangminjie920705@163.com,

Abstract. In this paper, Driven by advanced convolutional neural networks, we present a deep generate residual similar feature networks to improve super-resolution performance. Many researches have found that many CNN network training of SR requires skill and computing equipment. The input low-resolution features and CNN intermediate features contain rich similar feature maps. The different upscaling factor is used for needing the different model, which increase computational complexity. For the problem raised above, we proposed the generate residual feature (GRF) module to generate more high-frequency information in the residual feature. Each generated residual structure contains the short skip connection and long skip connection. Furthermore, we extract the similar feature in the residual feature by considering the interrelation among residual feature. Qualitative and quantitative assessments on benchmark datasets shows that we use different methods to achieve the same effect as best results of SR, while we make the network light weighted. Meanwhile, Our experiments shows that the pedestrian detection in the monitoring scene has achieved good results.

1. Introduction
A single super-resolution is a common problem in computer vision question. The purpose is to get a single high-resolution image from a single low-resolution image. There are many deep learning methods stand by convolutional neural networks (CNN) show better performance on image super-resolution problems. Image super-resolution technology is widely used in security surveillance, medical imaging, and satellite image processing [1]. The image super-resolution (SR) problem is an inverse problem. There are different CNN methods [2][4][5] based on different degradation models. Among them, Dong et al. [2] present a SRCNN networks that can learn a mapping for image SR. As the performance of convolutional neural networks continues to increase, their methods perform better on super-resolution of images. There are two classic convolutional neural networks for image super-resolution VDSR [6] and SRGAN [7]. Lim et al. [8] present EDSR for image SR by using simplified residual blocks.

However, according more methods, although a simple adding residual block can make the network more deeper, it is difficult to improve performance. Moreover, improving the performance of the network is still an exploration process for the image SR by deepening the depth of the network. Most of the current methods are to extract high-frequency information in low-resolution images by deepening the network, ignoring the generation of features and the relationship between features in the network.

For solving the above problem, we present a deep generate residual similarity feature network (DGRSFN) to obtain the depth residual feature map, and robustly obtain similar features of the high
frequency information in the residual feature. At the same time, we improve the propagation and computational power of image information in the network through the short-skip connection between the residual short block and the long-skip connection between the residual blocks.

Overall, there are three points in our contribution: (1) We present the deep generate residual similar feature network (DGRSFN) for more precise super resolution. (2) We propose the generate residual structure in the network to get more generate residual feature. (3) We propose the similar feature extraction in the residual feature for more high-frequency information.

2. Presented Method

2.1. Model Structure

As shown in Figure 1, DGRSFN mainly composed of four parts: low-level feature extraction, deep generate residual feature, Similar Feature extraction, Reconstruction part. We present $I_{LR}$ and $I_{HR}$ as the low resolution and high resolution image training pairs. And the $I_{SR}$ is the output of our SR model.

We define the low-level extraction feature $F_L$ from the $I_{LR}$ input

$$F_L = H_{LF}(I_{LR}) \tag{1}$$

Where $H_{LF}()$ denotes the convolution operator of low-level feature extraction. $F_L$ is used for generate residual feature with GRF module. So, we can get the generate residual feature $F_{RF}$:

$$F_{RF} = H_{GRF}(I_L) \tag{2}$$

Where $H_{GRF}$ denotes the deep generate residual feature structure, which contains similar feature extraction in residual feature structure. For making better use of the similar features in the residuals feature maps. Therefore, we can get a deep network depth and get more useful information about similar feature. We can reduce computational complexity and get more useful information.

In the reconstruction part, we are inspired by ESPCN [9] and EDSR [8] we make a upscaling and get the result of SR, we denotes the scaling factor is $M$. By the reconstruction part, we can get the $I_{SR}$ result of our method with different upscaling factor.

$$I_{SR} = H_{Rec-M}(F_{RF}) = H_{DGRSFN}(I_{LR}) \tag{3}$$

Figure 1. The architecture of deep generate similar feature network (DGRSFN).
Where $H_{Re-M}(\cdot)$ denote the reconstruction part with different upscaling factor $M$ and $H_{DGRSFN}(\cdot)$ denote the function of our method.

2.2. Generate Residual Feature (GRF)

![Figure 2. The results of generate residual feature.](image1)

![Figure 3. The similar feature in the generate residual feature.](image2)

In this part, we show the design of the presented generate residual feature structure. Inspired by the generation adversarial network (GAN) [10], there have been many researches about generating adversarial networks in recent years. We propose generate residual features in the residual network [11], which contain low frequency and high frequency information. By define the residual block in our network, we get the generate residual feature ($F_{SR}$). As shown in Figure 2, we can find the 64 generate residual feature, which contain low-frequency and high-frequency information. However, we want to get more detail high-frequency information to build the SR image from the more generate residual feature.

2.3. Similar Feature Extraction (SFE)

As mentioned in section 2.1, Depth-generated residual features are rich in high-frequency information, so it is necessary to centrally filter this information in the network during the reconstruction process. Inspired by [4][11] we introduce the Similar feature extraction module. We exploit the interrelation among residual feature by as show the Figure 3 the similar feature extract module and some similar example sample. Here, we option the sigmoid function to activate and get similar features $F_{SR}$.

$$F_{SF} = f\left(\text{Cov}(g(\text{Conv}(F_{RF})))\right)$$

Where $f(\cdot)$ and $g(\cdot)$ denote activation function, respectively. In the similar feature extraction module, Firstly, we use the Average pooling to Lightweight residual feature. Secondly, the basic convolution to reduce the feature map and expand the feature map. Moreover, the similar feature extraction module deepens the depth of the network and makes the network more robust.

2.4. Loss Function

Our network parameters are optimized by loss function. We investigated most of the existing SR algorithms loss functions $L_1$ and $L_2$, then the perceptual loss and adversarial loss [8] [7]. To validate the effects of our network model (DGRSFN), we chose to optimize with the $L_1$ Loss function as most
previous methods. We give a training set \( \{ I_{LR}^i, I_{HR}^i \}_N \) that contains low-resolution and corresponding high-resolution image pairs. Therefore, our loss function is

\[
\min_{\Theta} L(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \| H_{DGRSFN}(I_{LR}^i) - I_{HR}^i \|_1
\]

We use stochastic gradient descent (SGD) to optimize this cost function. More detailed experiments are presented in the next section.

3. Experiments

3.1. Experiments Showing

![Figure 4. Small-size people with different direction in the monitoring.](image)

Our training images from DIV2K dataset [13]. We use 800 training images. More experiments of testing implement in five datasets: Set5 [14], Set14 [15], BSD100 [16], Urban [17], Manga109 [18] and Small-size people datasets (as the figure 4) in monitoring [9]. The degradation model we used in bicubic interpolation (BI) downsampling. The SR value results are evaluate with average PSNR and SSIM on Y channel of the YCbCr image color space by Matlab.

In preliminary feature extraction, we use the convolution to extract the previous feature. We use Pytorch to experiment with a 1080Ti GPU. We use Adam optimization, and the learning rate is half by per \(2 \times 10^5\) iteration. In the reconstruction part, \(M\) is set \(2, 3, 4\). Then, we can get different scaling factor model for different SR questions.

3.2. Qualitative Result

As shown in the Table 1. We have compared the results of some super-resolution algorithms at this stage. On the numerical results, we conclude that this method is different from the previous method. And we get the same good results.

3.3. Visual Result

![Figure 5. Comparison of super-resolution visual effects of a butterfly image SR in scale 2.](image)
Table 1. For bilinearly interpolated images; numerical comparisons are made under different methods: We use APSNR and ASSIM to represent PSNR/SSIM by average about scale factor $2\times, 3\times, 4\times$.

| Methods  | Scale | Set5 APSNR | Set5 ASSIM | BSD100 APSNR | BSD100 ASSIM | Urban100 APSNR | Urban100 ASSIM | Manga109 APSNR | Manga109 ASSIM |
|----------|-------|------------|------------|--------------|--------------|----------------|----------------|----------------|----------------|
| Bicubic  | 2     | 33.97 0.932| 30.24 0.870| 29.56 0.843  | 26.88 0.840  | 30.80 0.934  |                 |                 |                 |
| SRCNN[2] | 2     | 36.65 0.954| 32.45 0.907| 31.36 0.888  | 29.50 0.895  | 35.60 0.966  |                 |                 |                 |
| VDSR[6]  | 2     | 37.53 0.958| 32.97 0.913| 31.90 0.896  | 30.77 0.914  | 37.16 0.974  |                 |                 |                 |
| EDSR[8]  | 2     | 38.11 0.960| 33.92 0.919| 32.32 0.901  | 32.93 0.935  | 39.10 0.977  |                 |                 |                 |
| DGRSFN   | 2     | 38.27 0.961| 34.10 0.921| 32.41 0.902  | 33.34 0.938  | 39.43 0.978  |                 |                 |                 |
| Bicubic  | 3     | 30.39 0.868| 27.55 0.774| 27.21 0.738  | 24.46 0.375  | 26.95 0.856  |                 |                 |                 |
| SRCNN[2] | 3     | 32.75 0.909| 29.30 0.821| 28.41 0.786  | 26.24 0.799  | 30.48 0.912  |                 |                 |                 |
| VDSR[6]  | 3     | 33.67 0.921| 29.78 0.832| 28.83 0.799  | 27.14 0.829  | 32.01 0.934  |                 |                 |                 |
| EDSR[8]  | 3     | 34.65 0.928| 30.52 0.846| 29.25 0.809  | 28.80 0.865  | 34.17 0.948  |                 |                 |                 |
| DGRSFN   | 3     | 34.74 0.929| 30.35 0.843| 29.32 0.8111 | 29.08 0.870  | 34.43 0.949  |                 |                 |                 |
| Bicubic  | 4     | 28.42 0.810| 26.10 0.704| 25.96 0.669  | 23.15 0.659  | 24.92 0.789  |                 |                 |                 |
| SRCNN[2] | 4     | 30.48 0.862| 27.50 0.751| 29.90 0.710  | 24.52 0.722  | 27.58 0.855  |                 |                 |                 |
| VDSR[6]  | 4     | 31.35 0.883| 28.02 0.768| 27.27 0.726  | 25.18 0.754  | 28.83 0.887  |                 |                 |                 |
| EDSR[8]  | 4     | 32.46 0.896| 28.80 0.787| 27.71 0.742  | 26.64 0.803  | 31.02 0.914  |                 |                 |                 |
| DGRSFN   | 4     | 32.63 0.900| 28.87 0.788| 27.77 0.743  | 26.82 0.808  | 31.21 0.917  |                 |                 |                 |

3.4. Pedestrian Detection Performance

Figure 6. In the small-scale pedestrian detection number comparison of the surveillance scene.

4. Conclusion

We propose the deep generate residual similar feature networks(DGRSFN) for single image SR. Specifically, Our residual feature extraction block passes the low-frequency detail information and makes the residual network deeper through residual learning and cross-connection. Furthermore, we extract similar features in the Generate Residuals module. Our networks make the network more flexible and extracts richer high-frequency detail features for the features. Extensive quantitative and qualitative evaluation of the base datasets shows that we use different methods to achieve the same results as SR’s most advanced technology while using fewer parameters. At the same time, we test the detection of small people in the monitoring scene and achieved good results.
5. Acknowledgements
This research is supported by Supported by the Key Program of the Science and Technology Service Network Program (STS) of the Chinese Academy of Sciences, project number KFJ-STS-QYZD-118.

6. References
[1] Thornton, M.W., Atkinson, P.M., Holland, D.A.: Sub-pixel mapping of rural land over objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping. International Journal of Remote Sensing 27(3), 473–491 (2006)
[2] Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. vol. 38, pp. 295–307 (2016)
[3] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: CVPR (2018)
[4] Sajjadi, M.S.M., Schölkopf, B., Hirsch, M.: Enhancenet: Single image super-resolution through automated texture synthesis. 2017 IEEE International Conference on Computer Vision (ICCV) pp. 4501–4510 (2017)
[5] Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1646–1654 (2016)
[6] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A.P., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 105–114 (2017)
[7] Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (July 2017)
[8] Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
[9] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Bing, X., Bengio, Y.: Generative adversarial networks. Advances in Neural Information Processing Systems 3, 2672–2680 (2014)
[10] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 770–778 (2016)
[11] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: CVPR (2018)
[12] Agustsson, E., Timofte, R.: NTIRE 2018 challenge on single image super-resolution: Dataset and study. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (2018)
[13] Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M. L.: Low-complexity single image super-resolution based on nonnegative neighbor embedding. BMVC (09 2012)
[14] Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse representations. In: International Conference on Curves Surfaces (2012)
[15] Martin, D., Fowlkes, C., Malik, J., Tal, D.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Computer Vision, IEEE International Conference on (ICCV). vol. 02, p. 416 (2001)
[16] Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: CVPR (2015)
[17] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., Aizawa, K.: Sketch-based manga retrieval using manga109 dataset. Multimedia Tools and Applications 7 6(20), 21811–21838 (2017)