Residual Attention Fusion Network for Single Image Super-Resolution

Hao Zhang¹, Chuwen Lan², Zehua Gao*

¹School of Information and Communication Engineering, Beijing Laboratory of Advanced Information Networks, Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications, Beijing, 100876, China
²gaozehua@bupt.edu.cn

Abstract. Recently, a very deep convolutional neural network (CNN) demonstrated influential performance in the field of single image super-resolution (SISR). However, most of the CNN-based methods focus on designing deeper and wider network structures alone and do not use the hierarchical and global features in the input image. Therefore, we proposed a residual attention fusion network (RAFN), which is an improved residual fusion (RF) framework, to effectively extract hierarchical features for use in single-image super-resolution. The proposed framework comprises two residual fusion structures composed of several residual and fusion modules, and a continuous memory mechanism is realized by adding a long and short jump connection. The network focuses on learning more effective features. Furthermore, to maximize the power of the RF framework, we introduced global context attention (GCA) module that can model the global context and capture long-distance dependencies. The final RAFN was constructed by applying the proposed RF framework to the GCA blocks. Extensive experiments showed that the proposed network achieved improved performance in the SISR method with fewer parameters, as compared to the methods proposed in previous studies.

1. Introduction
Recently, single image super-resolution (SISR) has received wide attention owing to its ability to capture a low-resolution (LR) image and produce a visually high-resolution (HR) image. However, Image SR develops multiple HR outputs to any LR input, which is an ill-posed concern. Many SR methods, such as early interpolation-based methods, reconstruction-based methods, and recent learning-based methods, have been proposed to overcome this problem [1–10].

In recent years, deep convolutional neural networks (CNNs) have achieved groundbreaking results in multiple fields, including the field of SISR, and have made great progress compared to traditional methods. Dong et al [6], proposed an SRCNN with a three-layer network, which combined a CNN with image super-resolution. Furthermore, Kim et al. increased the depth of the VDSR [8] to 20 layers by introducing residual structures. After this pioneering work, many other CNN-based methods have been proposed, due to which further achievements have been made in the field of SISR [11–18].

Although there has been considerable progress in the field of SISR, existing CNN-based SR models possess certain limitations. (1) Most CNN-based SR methods do not entirely utilize the hierarchical information in the original LR image, which results in relatively low performance. (2) While existing CNN-based SR models focus on designing deeper or wider networks to learn advanced features with
more discriminative power, they rarely use the global context features of the image, which hinders the representational ability of CNNs.

Figure 1. Commonly use of residual block in super resolution network.

To overcome these limitations, we propose a residual fusion (RF) framework that fuses local residuals to achieve an improved, and powerful feature representation. Figure 1 shows a traditional network design with multiple residual modules stacked together to build a deep network to propagate the residual features of the previous module to the subsequent blocks through a long path. After a series of convolution operations, these features were fused with other features to form complex features that were difficult for the network, thereby limiting the expression ability of the network. As seen in Figure 2, the proposed RF framework comprises reorganized stacked residual modules that fuse the output of each layer. The fusion module is realized through a $1 \times 1$ convolution, which does not introduce excessive computation. The residual blocks are aggregated once in advance to further reduce network overhead. The fusion features and module input are then used as the input to the subsequent RF modules in the form of residuals to fully utilize the residual features and hierarchical features of the entire network.

While the residual features of the different residual blocks reflect different aspects of the spatial content, the relationship between these remaining features does not receive sufficient attention. Therefore, to further improve the performance of the RF framework, a global context awareness mechanism was developed to enhance the capture of long-range dependency using residual features. However, the existing global attention mechanisms in image SR are either weak or computationally strong. For example, the plain spatial attention demonstrated in [19] lacked a large receptive field essential for image SR, whereas the nonlocal mechanism demonstrated in [10,20] consumed a large number of computational resources. To overcome this problem, we introduced a lightweight and efficient global contextual attention network (GCNET) [21], which uses global attention pooling for context modeling and a bottleneck transform for feature fusion to capture the inter-channel dependencies and model the global context while saving computation, such as the SE Block [26].

2. Related Work

Various methods, such as interpolation-based, model-based, and deep learning-based methods, have been proposed for image SR-based methods that have rapidly developed and showed improved results as compared to traditional methods. In this section, we briefly review the deep neural networks based single-image super-resolution methods, which is similar to what has been proposed in this study.

Dong et al. [6] introduced a novel shallow three-layer convolutional network (SRCNN) for image SR and achieved improved performance results as compared to previous studies. Inspired by this, Kim et al. stacked more convolution by introducing residual learning and designed a deeper VDSR [8] and DRCN [21] with over 16 layers. Furthermore, in ESPCN [22], Shi et al. proposed a subpixel convolution network to extend the features from the LR to HR space to build a deeper network. Lim et al. proposed a very deep network EDSR [12] with reduced memory consumption and computing time by using a subpixel convolution layer and modifying the structure of residual blocks by deleting the BN layer. Recent works such as Memnet [23], SRRResNet [24], and RDN [13] have introduced improve, denser connections.
Recently, attention mechanism has been widely used in computer vision tasks, by which a network can possess higher weight in the input containing the largest amount of information, and is usually combined with existing networks to further improve the network performance. Wang et al. [25] proposed a nonlocal neural network. Hu et al. [26] proposed a squeeze excitation network (SENET) that uses the relationship between channels as an advantage, and has made remarkable progress in image classification.

3. Residual Attention Fusion Network (RAFN)

![Figure 3. The architecture of our proposed residual attention fusion network (RAFN).](image)

3.1. Basic Network Architecture for Image SR

As seen in Figure 3, the proposed RAFN comprises four parts: shallow feature extraction, deep feature extraction based on the residual fusion framework, up-sampling module, and reconstruction section. $I_{LR}$ and $I_{SR}$ are the inputs and outputs of the RAFN, respectively. Based on the study in [12,25], we used a single convolution layer to extract the shallow feature $F_0$ from the LR input, which is given as:

$$F_0 = H_{sf}(I_{LR})$$

(1)

where $H_{sf}$ represents the shallow feature extraction module of the head. The extracted shallow feature $F_0$ is then sent to the body of the network for deep feature extraction, which comprises $K$ basic modules and RFB, and is given as:

$$F_k = R_k(F_{k-1}) = R_k(R_{k-1} - 1(\ldots(R_1(F_0)\ldots + F_0))$$

(2)

where $R_k$ represents the K-RF deep feature extraction module comprising d GCA. Based on the results, we can conclude that the proposed RFB not only obtains a deeper depth but also possesses a larger receptive field to ensure diversity in extracted features. $F_{k-1}$ is the input of the kth module, whereas $F_k$ is the corresponding output. The features are then amplified by the up-sampling module, given as:

$$F_{up} = H_{up}(F_k + F_0)$$

(3)

where $H_{up}$ and $F_{up}$ are the up-sampling module and amplification features, respectively. Global residual learning was used to alleviate the training difficulty. While several other options such as transpose convolution [28] and ESPCN [23] can be used as the upsampling module, previous studies suggest that it is better to adopt the latest CNN-based upsampling module to ensure a balance between the computational burden and network performance [28,13,29]. The amplified features were then mapped onto the SR images through a convolution network, which is given as:

$$I_{SR} = H_{R}(F_{up}) = H_{RAFN}(I_{LR})$$

(4)

where $I_{SR}$ is the super-resolution image, $H_R$ is the reconstruction layer, and RAFN is the network function. Loss functions such as RAFN, L2, L1, and perceived loss have been widely used. In this study, we used the L1 function to verify the effectiveness of RAFN, similar to the previous work. For a training set $\{I_{LR}, I_{HR}\}_{i=1}^{N}$ comprising N LR inputs and their corresponding HRs, the objective of RAFN training is to minimize the L1 loss function, which is given as:
where Θ is the parameter set of the network. The loss function was optimized using stochastic gradient descent. More details on the training are provided in section 4.1.

3.2. Residual Fusion (RF)
Residual learning and intensive learning are proven essential in image classification problems. Recently, EDSR [12] and ESRGAN [30] introduced residual learning into image SR to build a deeper network. Furthermore, RDN [13] introduced intensive learning into the network to extract multi-layer features, which further improved the performance of image SR. However, dense blocks increase the network size, making it difficult to train the number of network parameters.

To overcome these shortcomings, we propose an RF module that gives the network a greater depth and the ability to take full advantage of layered features while having fewer parameters. Figure 4 shows the details of the RF module comprising residuals learning, local feature fusion, and local long and short jump connections, and forms a continuous memory mechanism, which is given as:

\[
F_d = F_{d-1} + H_c(H_{c-1}(...(H_0(F_{d-1})...) + F_{d-1}))
\]  \hspace{1cm} (6)

Where \(F_d\) and \(F_{d-1}\) are the inputs and outputs of d-rb, respectively, and \(H_c\) is the c-th RF module. The first three residuals are aggregated in the middle of the module and locally short-linked as the later input. The RF module output of the previous layer is connected to all subsequent layers, so that each layer can extract local features and propagate the previous feature information to the end of the RF module without loss, to obtain features of more layers.

3.3. Global Context Attention (GCA)
We used the RF framework in conjunction with the spatial attention mechanism to maximize the power, considering we had to focus on the remaining features on the critical spatial content. The attention module must first be sufficiently light because it is plugged into every RF module in the network. Further, the attention module should utilize the global context of the image entirely to effectively complete the image SR task. Therefore, we introduced the global context attention (GCA) module, as shown in Figure 5, which is an improved version of NLNET [31], to obtain the non-local modeling capability for the global context and the computation-saving feature of SENET [26].

GCBlock given as:
\[ z_i = x_i + W_{v2}ReLU \left( LN \left( W_{v1} \sum_{j=1}^{N_p} e^{W_{k}x_j} x_j \right) \right) \]  \hspace{1cm} (7)

where \( \sum_{m=1}^{N_p} e^{W_{k}x_m} \) is the weight of global attention pooling and \( W_{v2}ReLU \left( LN \left( W_{v1} \cdot \right) \right) \) is a bottleneck transform.

3.4. Implementation Details
The RF framework was used in conjunction with the GCA module to build the final SR network (RAFN comprising 16 RF modules, each of which contained eight residuals and GCA modules. A 3 × 3 convolution kernel was used to extract shallow features. In each RF module, all 3 × 3 convolution kernels comprised 64 filters, except for the 1 × 1 convolution kernel of the feature fusion part. The final conversion layer had three output channels, allowing the network to output colored SR images.

4. Experiments

4.1. Settings

4.1.1. Datasets
Based on previous studies, we chose DIV2K [32] as the training set, which is a high-resolution dataset used in image recovery applications and comprises 800 training images, 100 validation images, and 100 test images. For testing, we used five standard benchmark datasets: SET5 [33], SET14 [34], BSD100 [35], URBAN100 [36], and MANGA109 [37]. A bicubic (BI) degradation model [9] was used to evaluate the results of all SRs in the Y channel of the transformed YCbCr space based on the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

4.1.2. Training Settings
During the training process, we performed data enhancement on 800 training images by randomly rotating the images to 90°, 180°, and 270°, and flipping them horizontally. From the training batch, 16 48 × 48 LR RGB patches were randomly extracted as inputs. We selected the ADAM optimizer training for our model, where \( \beta_1 = 0.9 \), \( \beta_2 = 0.99 \), and \( \epsilon = 10^{-8} \). The initial learning rate was 10^{-4} and was reduced to half every 200 backpropagation iterations. We implemented our model with NVIDIA 2080TI using the PyTorch framework.

4.2. Effects of RF Framework
As described in Section 3, our RAFN comprises two main components: the residual fusion module RF and the global context attention module GCA. To prove the effectiveness of the different modules, we compared the RF with the ordinary residual structure trained on the Set5 dataset using the same training set. The specific performance is listed in Table 1. Furthermore, to prove the SR performance of the residual fusion module proposed in this study, we deleted the global attention GCA module from the very deep network and retained the RF module and the long and short jump connections. As shown in Table 1, our module was advantageous over the PSNR and SSIM values on the SET5 dataset compared to the baseline (without RF and jump connection), with the RF module alone, thereby indicating that the proposed RF module is effective for image SR.

| Set5  | PSNR/SSIM (×2) | PSNR/SSIM (×3) | PSNR/SSIM (×4) |
|-------|----------------|----------------|----------------|
| Baseline | 38.09/0.9601 | 34.62/0.9279 | 32.45/0.8967 |
| RF     | 38.19/0.9613 | 34.73/0.9298 | 32.55/0.8991 |
4.3. Combination with GCA

Based on the above results, we combined the GCA module into the network to further demonstrate it on image SR. As compared to the network without the GCA module, the network comprising the GCA module showed improved performance. Based on the previous results, although it is difficult to further improve by deepening the network because of the very deep depth of the network, we were able to further improve the network performance through the GCA module, thereby demonstrating the effectiveness of the GCA module.

Considering the BI degradation model is widely used to simulate the LR images in image SR, we compared our RAFN with the 10 latest image SR methods: SRCNN, FSRCNN, VDSR, LAPSRN, MEMNET, EDSR, SRMD, NLRN, DBPN, and RDN, to verify its effectiveness. Table 2 shows the quantitative comparison results of BI models X2, X3, and X4SR.

4.4. Results with Bicubic (BI) Degradation Model

Compared with previous methods, the proposed RAFN achieved approximate or improved quantitative results. For scale x2, RAFN achieved the best results on B100, the best SSIM on Urban100, and the highest PSNR on MANGA109, whereas for scale x3, the RAFN outperformed the previous methods on all data sets. Our RAFN also exhibited excellent performance in proportional X4, achieving the best results achieved on both PSNRs. Furthermore, compared to other methods, RAFN performed particularly well on Urban100 and MANGA109 datasets considering both datasets contained rich structured content, which allowed RAFN to gradually aggregate hierarchical information into more representative features. This can be further verified from the SSIM score, which focuses on the visible structures in the image, achieved in this study. For example, on the Urban100 (×2) dataset, both PSNR and SSIM performed well, indicating that our RAFN is capable of recovering better visible structures. A similar phenomenon was observed on the SET14 (×4) dataset. The visual comparisons are shown in Figures 6 and 7, which prove that our RAFN can reconstruct improved structural details.

Table 2. Quantitative results with BI degradation model. Best results are highlighted.

| Method   | Scal | Set5     | Set14     | B100     | Urban100 | Manga109 |
|----------|------|----------|-----------|----------|----------|----------|
|          |      | PSN      | SSIM      | PSN      | SSIM     | PSN      | SSIM     |
| Bicubic  | ×2   | 33.66    | 0.929     | 30.24    | 0.868    | 29.56    | 0.843    | 26.88    | 0.840 | 30.80 | 0.939 |
| SRCNN    | ×2   | 36.66    | 0.954     | 32.45    | 0.906    | 31.36    | 0.887    | 29.50    | 0.894 | 35.60 | 0.966 |
| FSRCNN   | ×2   | 37.05    | 0.956     | 32.66    | 0.909    | 31.53    | 0.892    | 29.88    | 0.902 | 36.67 | 0.971 |
| VDSR     | ×2   | 37.53    | 0.950     | 33.05    | 0.913    | 31.90    | 0.896    | 30.77    | 0.914 | 37.22 | 0.975 |
| LapSRN   | ×2   | 37.52    | 0.959     | 33.08    | 0.913    | 31.08    | 0.895    | 30.41    | 0.910 | 37.27 | 0.974 |
| MemNet   | ×2   | 37.78    | 0.959     | 33.28    | 0.914    | 32.08    | 0.897    | 31.31    | 0.919 | 37.73 | 0.974 |
| EDSR     | ×2   | 38.11    | 0.960     | 33.92    | 0.919    | 32.32    | 0.901    | 32.93    | 0.935 | 39.10 | 0.977 |
| SRMD     | ×2   | 37.79    | 0.960     | 33.32    | 0.915    | 32.05    | 0.898    | 31.33    | 0.920 | 38.07 | 0.976 |
| DBPN     | ×2   | 38.09    | 0.960     | 33.85    | 0.919    | 32.19    | 0.899    | 32.55    | 0.932 | 38.89 | 0.977 |
| RDN      | ×2   | 38.24    | 0.964     | 34.01    | 0.921    | 32.34    | 0.901    | 32.89    | 0.935 | 39.18 | 0.978 |
| RAFN(Ours) | ×2 | **38.25** | **0.961** | **33.95** | **0.920** | **32.36** | **0.901** | **33.12** | **0.937** | **39.40** | **0.978** |

| Method   | Scal | Set5     | Set14     | B100     | Urban100 | Manga109 |
|----------|------|----------|-----------|----------|----------|----------|
|          |      | PSN      | SSIM      | PSN      | SSIM     | PSN      | SSIM     |
| BI       | ×3   | 30.39    | 0.868     | 27.55    | 0.774    | 27.21    | 0.738    | 24.46    | 0.734 | 26.95 | 0.855 |
| SRCNN    | ×3   | 32.75    | 0.909     | 29.30    | 0.821    | 28.41    | 0.786    | 26.24    | 0.798 | 30.48 | 0.911 |
| FSRCNN   | ×3   | 33.18    | 0.914     | 29.37    | 0.824    | 28.53    | 0.791    | 26.43    | 0.808 | 31.10 | 0.921 |
| VDSR     | ×3   | 33.67    | 0.921     | 29.78    | 0.832    | 28.83    | 0.799    | 27.14    | 0.829 | 32.02 | 0.934 |
| LapSRN   | ×3   | 33.82    | 0.922     | 29.87    | 0.832    | 28.82    | 0.798    | 27.07    | 0.828 | 32.21 | 0.935 |
| MemNet   | ×3   | 34.09    | 0.924     | 30.01    | 0.835    | 28.96    | 0.800    | 27.56    | 0.837 | 32.51 | 0.936 |
| EDSR     | ×3   | 34.65    | 0.928     | 30.52    | 0.846    | 29.25    | 0.809    | 28.80    | 0.865 | 34.17 | 0.947 |
| SRMD     | ×3   | 34.12    | 0.925     | 30.04    | 0.838    | 28.97    | 0.802    | 27.57    | 0.839 | 33.00 | 0.940 |
4.5. Model Complexity Analysis

Assuming we can obtain improved image super partition results by limiting the computing resources, we compared the performance of the models and found that the proposed model achieved similar performance as RDN (22M) and EDSR (22M) with fewer parameters (11M). This indicated that RAFN can achieve a good balance between model performance and complexity.
5. Conclusion

In this paper, we developed a very deep residual attention fusion network (RAFN) to overcome the limitations of image SR. The network effectively combines the residual fusion (RF) structure as a basic module allowing each module to fully utilize the characteristics of the local layer through the polymerization characteristics, and the residual memory block length and local connection to bypass low frequency information and focus on learning high-frequency information. Furthermore, we introduced the global context attention (GCA) module to maximize the function of the RF module, which allows the network to focus on the highly informative input. We conducted a large number of SR experiments using the BI degradation model to verify the effectiveness of the proposed RAFN. Experiments have demonstrated that RAFN can significantly reduce the computing cost and achieve better visual effects. Future work will focus on maintaining super-resolution performance while continuing to reduce the computing cost.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities NO. 2020RC02; National Natural Science Foundation of China (61905021, 61672108, 61976025); Shenzhen Technical Plan Project (JCYJ20180305164708625).

References

[1] Zhang, L., Wu X.L. (2006) An edge-guided image interpolation algorithm via directional filtering and data fusion. IEEE Trans. Image Processing, 15(8):2226–2238.
[2] Zhang, K.B, Gao, X.B, Tao, D.C, Li, X.L. (2012) Single image super-resolution with non-local means and steering kernel regression. IEEE Trans. Image Processing, 21(11):4544–4556.
[3] Timofte, R., Smet, V.D., Gool, L.V. (2013) Anchored neighborhood regression for fast example-based super-resolution. In: ICCV.920–1927.
[4] Timofte, R., Smet, V.D., Gool, L.V. (2014) A+: adjusted anchored neighborhood regression for fast super-resolution. In: ACCV. Springer.111–126.
[5] Peleg, T., Elad, M. (2014) A statistical prediction model based on sparse representations for single image super-resolution. IEEE Trans. Image Processing, 23(6):2569–2582.
[6] Dong, C., Loy, C.C., He, K., Tang, X. Learning a deep convolutional network for image super-resolution. In: ECCV. Springer.184–199.
[7] Jia, B.Hu., Singh, A., Ahuja, A.N. (2015) Single image super-resolution from transformed self-exemplars. In: CVPR. 5197–5206.

[8] Kim, J., Lee, J.K., Lee, K.M. (2016) Accurate image super-resolution using very deep convolutional networks. In: CVPR. 1646–1654.

[9] Zhang, K., Zuo, W.M., Zhang, L. (2018) Learning a single convolutional super-resolution network for multiple degradations. In: CVPR. 3262–3271.

[10] Dai, T., Cai, J.R., Zhang, Y.B., Xia, S.T., Zhang, L. (2019) Second-order attention network for single image super-resolution. In: CVPR. 11065–11074.

[11] Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H. (2017) Deep laplacian pyramid networks for fast and accurate super-resolution. In: CVPR. 5835–5843.

[12] Lim, B., Son, S.Y., Kim, H., Nah, S.J., Lee, K.M. (2017) Enhanced deep residual networks for single image super-resolution. In: CVPR. 1132–1140.

[13] Zhang, Y.L., Tian, Y.P., Kong, Y., Zhong, B.N., Fu, Y. (2018) Residual dense network for image super-resolution. In: CVPR. 2472–2481.

[14] Zhang, Y.L., Li, K.P., Li, K., Wang, L.C., Zhong, B.N., Fu, Y. (2018) Image super-resolution using very deep residual channel attention networks. In: ECCV. Springer. 294–310.

[15] Li, Z., Yang, J.L., Liu, Z., Yang, X.M., Jeon, G., Wu, W. (2019) Feedback network for image super-resolution. In: CVPR. 3867–3876.

[16] He, J.W., Dong, C., Qiao, Y. (2019) Modulating image restoration with continual levels via adaptive feature modification layers. In: CVPR. 11056–11064.

[17] Hu, X.C., Mu, H.Y., Zhang, X.Y., Wang, X.L., Tan, T., Sun, J. (2019) Metasr: A magnification-arbitrary network for super-resolution. In: CVPR. 1575–1584.

[18] Zhang, Y.L., Li, K.P., Li, K., Zhong, B.N., Fu, Y. (2019) Residual non-local attention networks for image restoration. In: ICLR (Poster).

[19] Hu, Y.T., Li, J., Huang, Y.F., Gao, X.B. (2018) Channel-wise and spatial feature modulation network for single image super-resolution. CoRR. abs/1809.11130.

[20] Liu, D., Wen, B.H., Fan, Y.C., Loy, C.C., Huang, T. (2018) Non-local recurrent network for image restoration. In: NeurIPS. 1680–1689.

[21] Yue, C. (2019) Genet: Non-local networks meet squeeze-excitation networks and beyond. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops.

[22] Kim, J., Lee, J.K., Lee, K.M. (2016) Deeply recursive convolutional network for image super-resolution. In: CVPR. 1637–1645.

[23] Shi, W.Z., Caballero, J., Huszar, F., Totz, J., Aitken, A., Bishop, B., Rueckert, D., Wang, Z.H. (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: CVPR. 1874–1883.

[24] Tai, Y., Yang, J., Liu, X.M. (2017) Image super-resolution via deep recursive residual network. In: CVPR. 2790–2798.

[25] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W. (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR. 105–114.

[26] Wang, F., Jiang, M.Q., Qian, C., Yang, S., Li, C., Zhang, H.G., Wang, X.G., Tang, X.O. (2017) Residual attention network for image classification. In: CVPR. 6450–6458.

[27] Hu, J., Shen, L., Sun, G. (2018) Squeeze-and-excitation networks. In: CVPR. 7132–7141.

[28] Dong, C., Loy, C.C., Tang, X.O. (2016) Accelerating the super-resolution convolutional neural network. In: ECCV. Springer.

[29] Haris, M., Shakharovich, G., Ukita, N. (2018) Deep backprojection networks for super-resolution. In: CVPR.

[30] Wang, X.T., Yu, K., Wu, S.X., Gu, J.J., Liu, Y.H., Dong, C., Qiao, Y., Loy, C.C. (2018) ESRGAN: enhanced super-resolution generative adversarial networks. In: ECCV. Springer. 63–79.

[31] Wang, X., Girshick, R., Gupta, A., He, K. (2018) Non-local neural networks. In: CVPR.
[32] Timofte, R., Agustsson, E., Gool, L.V., Yang, M.H., Zhang, L., Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M. (2017) NTIRE 2017 challenge on single image super-resolution: Methods and results. In: CVPR.1110–1121.

[33] Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M. (2012) Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In BMVC.1–10.

[34] Zeyde, R., Elad, M., Protter. M. (2010) On single image scale-up using sparse-representations. In Curves and Surfaces.Springer.711–730.

[35] Martin, D., Fowlkes, C., Tal, D., Malik, J. (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In ICCV.416–425.

[36] Huang, J.B., Singh, A., Single, A. (2015) image super-resolution from transformed self-exemplars. In: CVPR.5197–5206.

[37] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., Aizawa, K. (2017) Sketch-based manga retrieval using manga109 dataset. Multimedia Tools Appl.76(20):21811–21838.