



FAN: Frequency Aggregation Network for Real Image Super-resolution

Yingxue Pang*, Xin Li*, Xin Jin,
Yaojun Wu, Jianzhao Liu, Sen Liu, and Zhibo Chen[†]

CAS Key Laboratory of Technology in Geo-spatial Information Processing and
Application System,
University of Science and Technology of China, Hefei 230027, China
{pangyx,lixin666,jinxustc,yaojunwu,jianzhao}@mail.ustc.edu.cn
elsen@iat.ustc.edu.cn chenzhibo@ustc.edu.cn

Abstract. Single image super-resolution (SISR) aims to recover the high-resolution (HR) image from its low-resolution (LR) input image. With the development of deep learning, SISR has achieved great progress. However, It is still a challenge to restore the real-world LR image with complicated authentic degradations. Therefore, we propose FAN, a frequency aggregation network, to address the real-world image super-resolution problem. Specifically, we extract different frequencies of the LR image and pass them to a channel attention-grouped residual dense network (CA-GRDB) individually to output corresponding feature maps. And then aggregating these residual dense feature maps adaptively to recover the HR image with enhanced details and textures. We conduct extensive experiments quantitatively and qualitatively to verify that our FAN performs well on the real image super-resolution task of AIM 2020 challenge. According to the released final results, our team SR-IM achieves the fourth place on the X4 track with PSNR of 31.1735 and SSIM of 0.8728.

Keywords: Frequency aggregation network (FAN), Real image super-resolution (RealSR), AIM 2020 challenge.

1 Introduction

Single image super-resolution (SISR) task aims to recover the high-resolution (HR) image from its low-resolution (LR) input image, where the LR image is acquired by applying some downsampling settings to the HR image. Whether the early traditional methods like [8,2] or recent methods [6,15,19,27], SISR has drawn more and more attention because of its wide range of application, such as medical image[11], surveillance [36] and security [10]. With the development of convolutional neural networks (CNN) and several high quality super-resolution datasets, Recent models [6,15,19,27,20,16] have achieved remarkable success in

*The first two authors contributed equally to this work.

[†]Corresponding author.

SISR task. A well-designed CNN can effectively capture the non-linear mapping and automatically learn the mapping function based on various high-quality super-resolution dataset. However, these models can perform well on the “clean” standard benchmarks, and they often fail to be applied in real-world scenarios. They cannot handle the real-world image super-resolution problem with complicated degradation processes, as shown in Fig. 1.



Fig. 1. Examples of SISR cannot be generalized to RealSR. (a) Real low-resolution image, which is captured by low quality DSLR camera. (b) Processed by general SR model EDSR. (c) Processed by our FAN.

Therefore, the real-world image super-resolution (RealSR) has attracted more and more attention. Different from the general SISR, the purpose of RealSR is to learn a general model to restore LR image captured under practical scenarios to high-resolution. This model focuses on solving the degradation problem that is more complex than bicubic downsampling. The paired LR and HR images are captured by various DSLR cameras. The diverse degradation differences among devices can be expressed as Eq.1

$$\begin{aligned} I_{HR} &= (I * k_{HR}) + n_{HR}, \\ I_{LR} &= (I * k_{LR}) \downarrow_s + n_{LR}, \end{aligned} \quad (1)$$

where k_{LR} and k_{HR} represent degradation kernel of high quality and low quality DSLR cameras respectively. \downarrow_s denotes the multiple of downsampling. n_{HR} and n_{LR} are additive noise. I represents the real world image. I_{HR} and I_{LR} are the image pairs collected with high-quality and low-quality DSLR cameras respectively. RealSR recovers the I_{HR} from I_{LR} , which more satisfies the needs of industrial applications. According to Eq. 1, we can get the I_{LR} from I_{HR} as Eq.2.

$$I_{LR} = (I_{HR} * k_{HR}^{-1} * k_{LR}) \downarrow_s + n_{LR} - (n_{HR} * k_{HR}^{-1} * k_{LR}) \downarrow_s. \quad (2)$$

From the above equation, we can find that the RealSR is more complicated than general SISR. When the degradation of high-quality DSLR camera is small, the Eq. 2 will be degraded to general SISR problem.

In this paper, we propose a frequency aggregation network (FAN) for complex real-world image super-resolution problem. Specifically, we extracted the low-frequency, middle-frequency and high-frequency of LR image with multi-scale representation. Then we utilize different branches to restore the corresponding frequency components of LR and aggregate them adaptively with channel attention mechanism to generate the HR image. From different frequency components, our framework can be more robust to solving different distortions. In order to solve the complex degradation kernels, we take RDB [40] as our basic unit, and redesign GRDB [17] into CA-GRDB by introducing a channel mechanism to adaptively aggregate the representations of different receptive fields, which can further improve the representation ability of the network. Extensive experiments have validated the effectiveness of our FAN for real image super-resolution problem. According to the released final results, our team SR-IM achieves the fourth place on the X4 track with PSNR of 31.1735 and SSIM of 0.8728 by applying our FAN on the real image super-resolution task of AIM 2020 challenge [34,35].

Our contributions can be summarized as follows:

- We propose a frequency aggregation network (FAN) for real image-world super-resolution problem.
- We redesign the novel CA-GRDB module by introducing a channel attention mechanism to further improve the representation ability of network.
- According to the released final results, our team SR-IM achieves the forth place on the X4 track with PSNR of 31.1735 and SSIM of 0.8728 on the real-world image super-resolution task of AIM 2020 challenge [34].

2 Related Works

2.1 Single Image Super-Resolution

Dong et al. proposed the SRCNN [6,7] which can be roughly seen as the first SISR work based on CNN. They adopted an end-to-end supervised learning model to restore the HR images with its corresponding bicubic downsampling LR images. Compared to traditional methods [8,2], they reconstructed high-quality HR images with clearer details and higher metric scores. SRCNN proves the effectiveness of CNN in solving SISR problem and inspires plenty of works to be proposed to improve the qualitative and quantitative results. Kim et al. [18] proposed DRCN in which they designed a deeper recursive layer with the receptive field of 41 to demonstrate their performance in common benchmarks. But methods extended by SRCNN [18,30] usually utilize the pre-defined upsampling operator which increases unnecessary computational cost and lead to reconstruction artifacts in some cases. To tackle this problem, LapSRN [19] constructed a set of cascaded sub-networks to progressively predicts the sub-band high frequency residuals. It replaced the pre-defined upsampling with the learned transposed convolutional layers to remove the undesired artifacts and reduce the computational cost.

Considered the methods mentioned above, we still have the difficulty to recover the HR image with more high-frequency details when using larger upsampling factors. Since those methods focus to optimize the pixel loss and ignore the image quality, the results often lack high-frequency details or fail to maintain the perceptual fidelity compared to its high-resolution “ground truth”. Ledig et al. [21] proposed SRGAN, the first SISR work which utilized Generative Adversarial Network(GAN), to solve the problem. They combine a perceptual loss, an adversarial loss and a content loss to infer HR images for 4x upsampling factors and show hugely significant gains in perceptual quality. Wang et al. [32] proposed ESRGAN which is an enhanced version by reconsidering the entire SRGAN and modifying it in network architecture, adversarial loss and perceptual loss. And ESRGAN brings the subjective state-of-the-art algorithm to SR. Zhang et al. [37] proposed RankSRGAN to introduce a ranker to learn the behavior of perceptual metrics and address the indifferentiable perceptual metrics problem.

Unlike previous feed-forward approaches, Haris et al. [13] introduced the back-projection into the reconstruction process of SISR. They showed that combining the up and down sampling, along with error feedbacks, encourages to get better results. Then, it seems to be a new trend to exploit the attention mechanism which can adaptively process visual information and focus on salient areas. RCAN [38] was proposed to apply the channel attention to capture the dependencies among channels. RNAN [39] utilized the local and non-local attention modules to get feature representation and dependencies. Similarly, SAN [5] also incorporated the non-local attention mechanism to capture long-range spatial contextual information. In addition, multi-pass methods also prove its effectiveness in SISR in recent work [4,28,22,12]. Within it, different paths may perform different operations to extract corresponding feature maps. By fusing these feature maps coming from each path, the whole network provides better modelling capabilities and generalization. We hence get inspired to construct our three-path aggregation network to capture different frequency information of LR images. And we demonstrate that our method performs well on the AIM 2020 challenge, came the fourth place on the X4 track with PSNR of 31.1735 and SSIM of 0.8728.

2.2 Real-World Image Super-Resolution

The SISR methods mentioned above usually utilize the clean and ideal datasets which mostly adopt simple and uniform degradation (e.g. bicubic downsampling) to construct LR images from HR images without any distracting artifacts (e.g. sensor noise, image compression, non-ideal PSF, etc). However, since the degradation in real world images is more complicated, SISR models inevitably fail in real-world image super-resolution with limited application and generalization.

To overcome the problem, several real-world super-resolution challenges [26,25] have been hold to attract more participants to come up their solutions. Cai et al. [1] build a real-world single super-resolution dataset where paired LR-HR images on the same scene are captured by adjusting the focal length of a digital camera. Shocher et al. [29] proposed ZSSR , the first unsupervised CNN-based

SR method, to train a small image-specific CNN at test time by exploiting the internal recurrence information of the image. Fritsche et al. [9] proposed DS-GAN to generate LR-HR pairs with similar natural image characteristics. Ji et al. [14] proposed a new degradation framework with various blur kernels and noise injections to solve the realistic image super-resolution problem. However, in this work, we propose FAN to extract the different frequency components of the LR image and then aggregate them to recover the HR image with more high frequency details.

3 Method

In this section, we will illustrate the overall architecture of our FAN and explain each component in detail, including hierarchical feature extractor (HFE), the main body with three branches and fusion module.

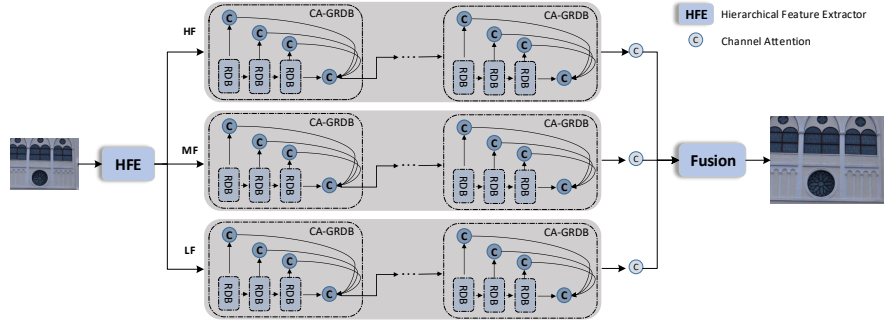


Fig. 2. The proposed network architecture: FAN

3.1 Overall Structure of FAN

Our end-to-end frequency aggregation network (FAN) is shown in Fig. 2. We first extract different frequency components of the LR image through the hierarchical feature extractor (HFE), and then pass them to each branch to obtain corresponding feature map. FAN contains three branches with same structure to process high-frequency component (HF), middle-frequency component (MF) and low-frequency component (LF) respectively. Followed three branches, FAN utilizes the fusion module to recover the HR image with enhanced details and textures from three residual dense feature maps.

3.2 Hierarchical Feature Extractor

Method [3] has proved that features extracted by different scale factors can capture different frequency information of the image. Inspired by it, we design the

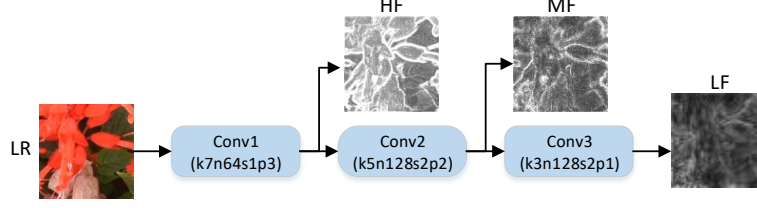


Fig. 3. Hierarchical feature extractor (HFE) with corresponding kernel size (k), number of feature maps (n), stride (s) and padding size (p) for each convolutional layer.

hierarchical feature extractor (FHE), which is shown in Fig. 3, to progressively extract different frequency components. Specifically, we gradually input the LR image into three convolutional layers and output the hierarchical feature maps in each layer. As [3] claimed, the larger feature map size is, the higher frequency component it contains. Therefore, we design the first convolutional layer (Conv1) with form of $k7n64s1p3$ to output the same size feature map (HF) as the LR image. The remaining two (Conv2 and Conv3) with forms of $k5n128s2p2$ and $k3n128s2p1$ produce the feature maps (MF and LF) which have half and quarter size of the input LR image. We indicate the convolutional layer by corresponding kernel size (k), number of feature maps (n), stride (s) and padding size (p). Among three frequency components, LF contains the low-level frequency of LR image with roughly structure and HF represents the high-level frequency with fine-grained details such as edges and textures. MF indicates the intermediate value between them. The entire frequency component extraction process can also be defined as Eq. 3.

$$HF = Conv1(I_{LR}), MF = Conv2(HF), LF = Conv3(MF) \quad (3)$$

3.3 Main Body: Three Branches

Since the degradation of DSLR camera is complex and contains multiple distortions such as blur and noise, we need to improve the network representation ability. After HFE, we hence propose the main body with three same branches to handle each frequency feature map separately. Inspired by [38] and [17], we introduce channel attention mechanism into the grouped residual dense block (GRDB) termed as CA-GRDB to adaptively aggregate those residual dense blocks (RDBs) [40] with different attention weights. We cascade four CA-GRDBs in each branch where each CA-GRDB consists of three RDBs, resulting 12 RDBs of each branch shown in Fig. 2.

Then the specific structure of CA-GRDB and RDB is shown in Fig. 4. In CA-GRDB, we cascade three RDBs and apply channel attention after each RDB.

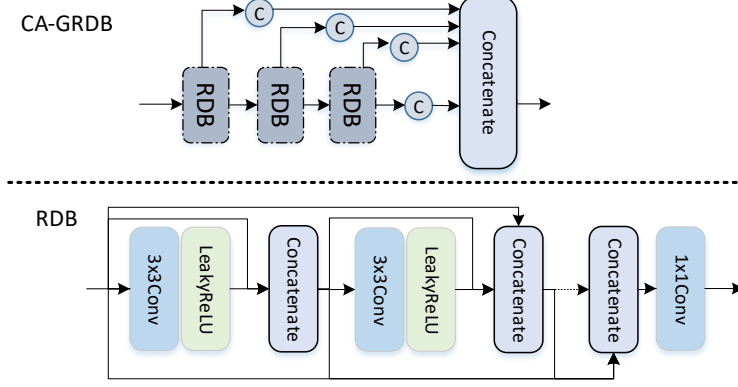


Fig. 4. Components of CA-GRDB and RDB, where C denotes channel attention module from [38] and the convolutional layers in RDB are from [17].

Then the feature maps from channel attention are concatenated together as fusing operation. As for the RDB, we replace ReLU with LeakyReLU to avoid dying ReLU problem [24]. And each RDB is constructed by 8 layers with the form 3x3 Conv-LeakyReLU-Concat via dense connection and residual connection.

3.4 Fusion Module

The fusion module is shown in Fig. 5. Like [31] claimed, Non-Local module (NL) tends to capture the long-term dependencies of features to compensate the lost information due to the local neighborhood filtering in convolutional layers. To fuse the features of different branches, we utilize non-local attention and convolution layer to further capture the correlation between different frequency features. Then we exploit pixel-shuffle layer used in [23] and convolutional layer twice to upsample the frequency feature to the original image size, which is the generated HR image. Notice that we employ 2x upsampling operation using pixel-shuffle layer and we detail each convolutional layer with corresponding kernel size (k), number of feature maps (n), stride (s) and padding size (p) in Fig. 5.

3.5 Loss Functions

We first employ pixel-wise $L1$ loss to measure the reconstruction error and then exploit MSE loss to fine tune our model. And the overall loss function can be represented as follows,

$$\mathcal{L}_{ALL} = \mathcal{L}_{L1} + \mathcal{L}_{MSE} \quad (4)$$

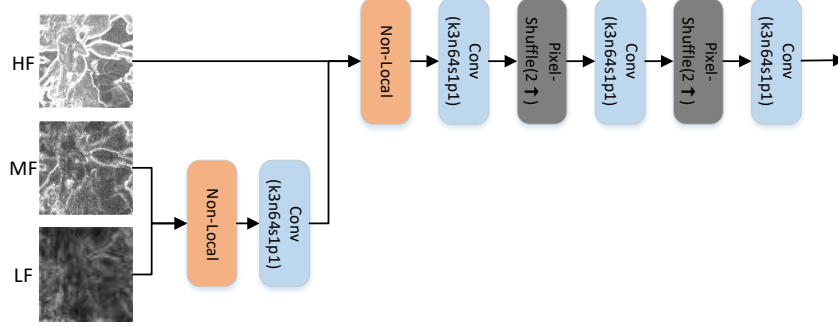


Fig. 5. Fusion module, which integrate non-local attention [31], pixel-shuffle layer [23] and convolution layer to capture the correlation between different frequency features. Each convolutional layer is denoted with corresponding kernel size(k), number of feature maps(n), stride(s) and padding size (p). (2 ↑) means 2x upsampling operation using pixel-shuffle layer.

4 Experiments

In this section, we will illustrate the RealSR dataset and evaluation metrics used in our experiments. And we will describe the implementation details of our training and testing. We compare our FAN with seven state-of-the-art methods qualitatively and quantitatively. To further verify the effectiveness of our each FAN module, We also conduct several ablation studies.

4.1 Dataset

We train our FAN model based on 19,000 LR-HR pairs of training images provided by AIM 2020 Real Image Super-Resolution Challenge-Track3, where the paired LR-HR images are captured by various camera in practical scenarios. In addition, the validation dataset contains 20 pairs of images. Although the test dataset provides 60 pairs of LR-HR images, the HR images are not available in the test stage. Therefore, we test our FAN model on validation dataset and compare with other state-of-the-art methods in this work. According to the released final results, we achieved the fourth place on the X4 track with PSNR of 31.1735 and SSIM of 0.8728 on test dataset.

4.2 Evaluation Metrics

(1) **Peak Signal-to-Noise Ratio (PSNR)**: PSNR is one of the most widely used full-reference quality metrics. It measures the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects

the quality of its representation. In other words, PSNR metric reflects the intensity differences between the real HR image and the recovered HR image. A higher PSNR score means that the intensity of two images is more close. Mathematically, it is high related to MSE loss mentioned above:

$$PSNR = 10 \log_{10} \left(\frac{R^2}{MSE} \right) \quad (5)$$

where R^2 denotes the maximum fluctuation in the input image data type.

(2) **The Structural SIMilarity Index (SSIM)** [33]: SSIM evaluate the image quality based on luminance, contrast and structure from the perspective of image formation. We apply it to compute the perceptual distance between the real HR image and the recovered image. A higher SSIM score means that the luminance, contrast and structure of two images are more similar. Mathematically,

$$\begin{aligned} \mathcal{L}_{SSIM} &= [l(\hat{I}, I)]^\alpha \cdot [c(\hat{I}, I)]^\beta \cdot [s(\hat{I}, I)]^\gamma \\ l(\hat{I}, I)^\alpha &= \frac{2\mu_{\hat{I}}\mu_I + C_1}{\mu_{\hat{I}}^2 + \mu_I^2 + C_1} \\ c(\hat{I}, I)^\beta &= \frac{2\sigma_{\hat{I}}\sigma_I + C_2}{\sigma_{\hat{I}}^2 + \sigma_I^2 + C_2} \\ s(\hat{I}, I)^\gamma &= \frac{\sigma_{\hat{I}I} + C_3}{\sigma_{\hat{I}}\sigma_I + C_3} \end{aligned} \quad (6)$$

where $\mu_{\hat{I}}, \mu_I, \sigma_{\hat{I}}, \sigma_I$ denote mean and standard deviation of the image \hat{I}_{HR}, I_{HR} respectively. C_1, C_2, C_3 are three constants to avoid instability. α, β, γ are hyper-parameters to control the relative importance.

4.3 Implementation Details

Our FAN is implemented based on PyTorch framework with four NVIDIA 1080Ti GPUs. To avoid undesirable over-fitting behaviors due to the limited data, we use data augmentation in the process of training such as randomly cropping, flipping and rotation. The number of mini-batches was set as 8. We use Adam optimizer with a initial learning rate of 0.0001 which will decay by a factor 0.5 every 30 epochs. We spent almost 48 hours to train our FAN model.

For testing, we first crop the LR image into several 196×196 small patches and then fed them into our FAN model to obtain the HR patches. Finally we convert those patches into a complete HR image. Our model requires 0.19s on a single NVIDIA 1080-Ti GPU with an LR image of 196×196 for testing. And we all use the 130 epochs.

4.4 Comparison with the state-of-the-art methods

We compare our FAN model with seven state-of-the-art methods in same test settings. The first one is a traditional method, i.e, bicubic upsampling. And the

others are CNN-based methods which are retrained on the RealSR x4 dataset for fairness. DBPN [13] exploits iterative up and downsampling layers to self-correct features at each stage by error feedback mechanism. EDSR [23] modifies the conventional residual networks by removing some unnecessary modules and get significant performance improvement. RCAN [38] proposes the channel attention to capture the dependencies among deep residual feature maps. RDN [40] conducts a novel residual dense network to extract feature maps. SAN [5] incorporates the non-local attention mechanism to capture long-range spatial contextual information. VDSR [18] extends the depth of neural network which highly improves the accuracy.

Table 1. Quantitative results on RealSR x4 dataset. We compare our FAN with the state-of-the-art methods in terms of PSNR and SSIM.

Method	VDSR	EDSR	RDN	DBPN	RCAN	SAN	FAN(ours)	FAN+
PSNR	28.966	29.522	29.790	29.305	30.002	30.112	30.598	30.719
SSIM	0.8322	0.8444	0.8486	0.8388	0.8529	0.8557	0.8603	0.8621

Table 1 shows the comparisons of our FAN with the previous methods on RealSR x4 dataset. Compared to other methods, our FAN with self-ensemble ($\times 8$) termed as **FAN+** containing flipping ($\times 4$) and rotation ($\times 4$). FAN+ both performs the best scores in PSNR and SSIM, leading to a 0.607 dB and a 0.0064 dB increase respectively compared with SAN [5]. Even without self-ensemble, termed as **FAN(ours)**, can also achieve 0.486 dB and 0.0046 dB increase.

The qualitatively comparison is shown in Fig. 6. The traditional bicubic upsampling produces the most blurry results. And compared with the other cnn-based methods, our FAN model can generate sharper and clearer HR images without obvious artifacts or distortions which are caused by the enlarged DSLR camera lens in the RealSR dataset. For example, the building fences and leaf veins are more clearer and not anamorphic. All in all, our FAN model can solve the RealSR problem well while the previous SISR methods can not generalize to this problem.

4.5 Ablation study

To further demonstrate the effectiveness of our each module, we conduct several ablation studies to investigate the influence of the multi-frequency branches, channel attention for CA-GRDB, non-local module, channel attention mechanism and the number of RDBs in our FAN. The subjective performance comparisons in ablation study are shown in Fig. 7.

Dataset Since the ground truth HR images of validation dataset are not released, we randomly split the training dataset into two parts by the ratio 7:3 for training and testing respectively. Notice that this dataset setting is only for ablation study.

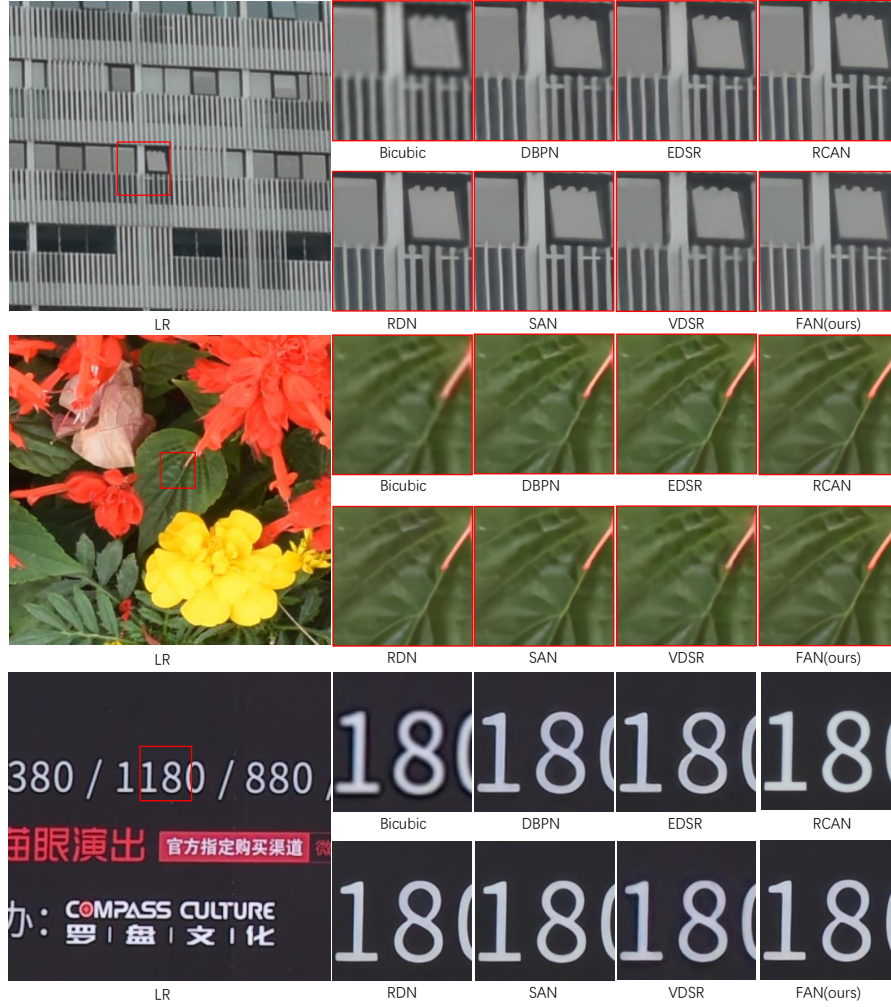


Fig. 6. The performance comparison of our FAN with the state-of-the-art methods performed on RealSR x4 dataset. The state-of-the-art methods are bicubic upsampling, DBPN [13], EDSR [23], RCAN [38], RDN [40], SAN [5] and VDSR [18] respectively.

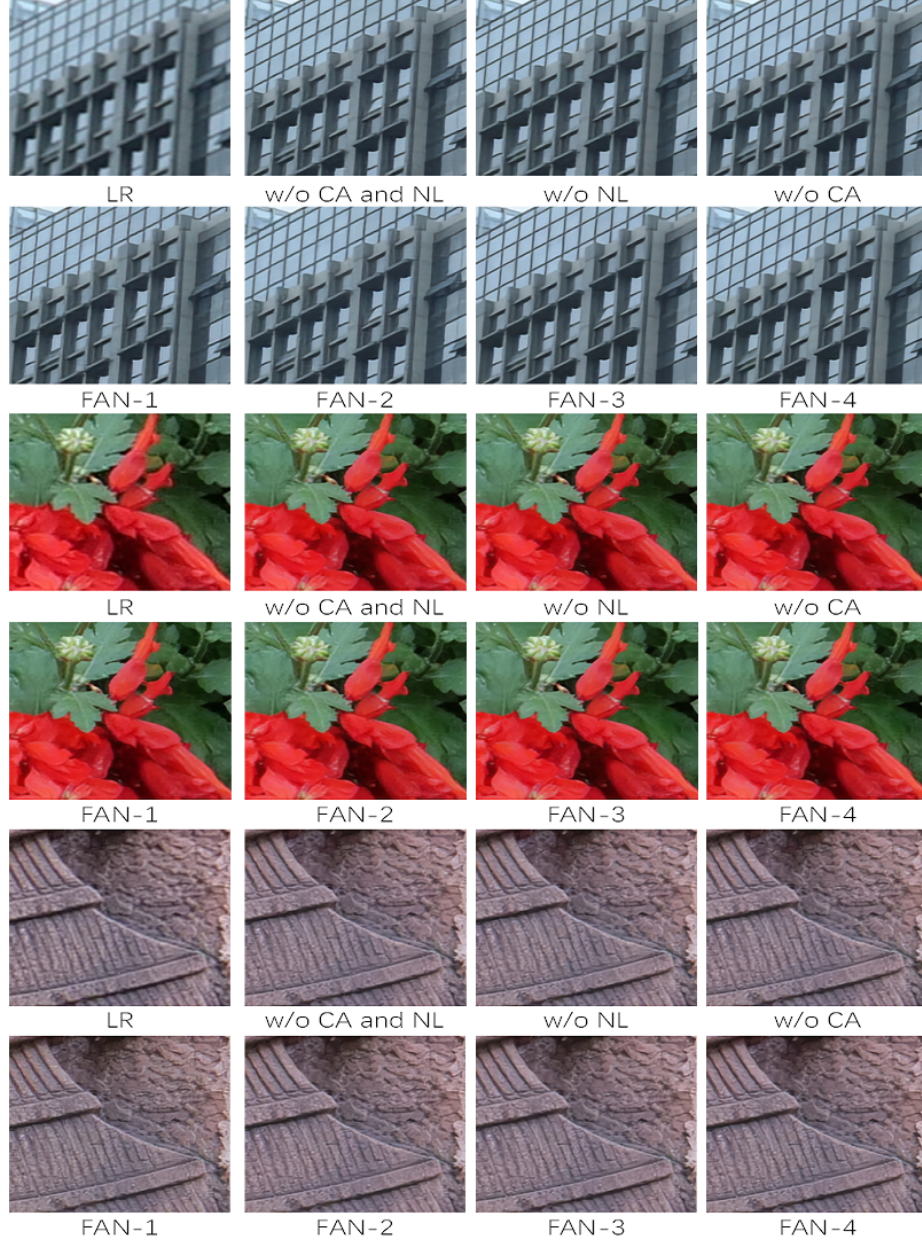


Fig. 7. The subjective performance comparisons in ablation studies of FAN.

Multi-frequency branches We set the number of branches from 1 to 4 to figure out the effects of different branches, denoted as FAN-1, FAN-2, FAN-3, FAN-4 shown in Table 2. with the number of branches increasing, we can observe the improvement of performance in general. However, the gain gets smaller and the number of parameters increase significantly when the number of branches is more than 3. Therefore, we select 3 as a trade-off to balance the complexity of network and performance.

Channel attention(CA) and Non-local module(NL) We exploit CA to model the feature inter-dependencies, which enforces our model to concentrate on more informative features. Like [31] claimed, NL tends to capture the long-term dependencies of features to compensate the lost information due to the local neighborhood filtering in convolutional layers. From Table 3, we can see that whether removing CA (w/o CA) or removing NL (w/o NL) results in the drop of performance. At the same time, adding CA and NL both would achieve best performance while the number of parameters only has slightly increasing.

Table 2. Comparisons between different number of branches in FAN. Tested on partitioned dataset, which is different from validation dataset.

Branches	FAN-1	FAN-2	FAN-3	FAN-4
Parameters (MB)	5.81	27.95	50.30	72.74
PSNR (dB)	28.5210	28.8029	28.9464	28.9598
SSIM	0.8153	0.8194	0.8231	0.8233

Table 3. Ablation study on different combinations of channel attention and non-local attention. Tested on partitioned dataset.

CA.	NL.	Parameters(MB)	PSNR(dB)	SSIM
✓	✓	50.30	28.9464	0.8231
×	✓	50.28	28.8073	0.8195
✓	×	50.12	28.7931	0.8194
×	×	50.10	28.7571	0.8190

5 Conclusions

In this paper, we proposed FAN, a frequency aggregation network, to solve the real-world image super-resolution problem which suffers from more complicated degradation process rather than simple bicubic downsampling. Our model is designed in an end-to-end manner, in which HFE extract three feature maps contained different frequency information firstly. And then pass them into CA-GRDB module independently to output corresponding residual dense feature

map in each frequency. Finally, we apply fusion module to integrate them to recover the HR image. We validate the effectiveness of our FAN in RealSR Dataset with extensive experiments. In addition, our model performs well on the AIM 2020 challenge which came the fourth place on the X4 track with PSNR of 31.1735 and SSIM of 0.8728.

Acknowledgement

This work was supported in part by NSFC under Grant U1908209, 61632001 and the National Key Research and Development Program of China 2018AAA0101400.

References

1. Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: Toward real-world single image super-resolution: A new benchmark and a new model. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3086–3095 (2019)
2. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. vol. 1, pp. I–I. IEEE (2004)
3. Chen, Y., Fan, H., Xu, B., Yan, Z., Kalantidis, Y., Rohrbach, M., Yan, S., Feng, J.: Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3435–3444 (2019)
4. Dahl, R., Norouzi, M., Shlens, J.: Pixel recursive super resolution. In: Proceedings of the IEEE international conference on computer vision. pp. 5439–5448 (2017)
5. Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L.: Second-order attention network for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 11065–11074 (2019)
6. Dong, C., Chen, C.L., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution (2014)
7. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network (08 2016)
8. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-Based Super-Resolution. IEEE Computer Society Press (2002)
9. Fritsche, M., Gu, S., Timofte, R.: Frequency separation for real-world super-resolution. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 3599–3608. IEEE (2019)
10. Gohshi, S.: Real-time super resolution algorithm for security cameras. In: 2015 12th International Joint Conference on e-Business and Telecommunications (ICETE). vol. 5, pp. 92–97. IEEE (2015)
11. Greenspan, H.: Super-resolution in medical imaging. The computer journal **52**(1), 43–63 (2009)
12. Han, W., Chang, S., Liu, D., Yu, M., Witbrock, M., Huang, T.S.: Image super-resolution via dual-state recurrent networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1654–1663 (2018)
13. Haris, M., Shakhnarovich, G., Ukita, N.: Deep back-projection networks for super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1664–1673 (2018)
14. Ji, X., Cao, Y., Tai, Y., Wang, C., Li, J., Huang, F.: Real-world super-resolution via kernel estimation and noise injection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 466–467 (2020)
15. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution (2016)
16. Kalarot, R., Li, T., Porikli, F.: Component attention guided face super-resolution network: Cagface. In: The IEEE Winter Conference on Applications of Computer Vision. pp. 370–380 (2020)
17. Kim, D.W., Ryun Chung, J., Jung, S.W.: Grdn: Grouped residual dense network for real image denoising and gan-based real-world noise modeling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–0 (2019)

18. Kim, J., Lee, J.K., Lee, K.M.: Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
19. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)
20. Lan, R., Sun, L., Liu, Z., Lu, H., Su, Z., Pang, C., Luo, X.: Cascading and enhanced residual networks for accurate single-image super-resolution. *IEEE Transactions on Cybernetics* (2020)
21. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4681–4690 (2017)
22. Li, J., Fang, F., Mei, K., Zhang, G.: Multi-scale residual network for image super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 517–532 (2018)
23. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution (2017)
24. Lu, L., Shin, Y., Su, Y., Karniadakis, G.E.: Dying relu and initialization: Theory and numerical examples (2019)
25. Lugmayr, A., Danelljan, M., Timofte, R.: Ntire 2020 challenge on real-world image super-resolution: Methods and results. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 494–495 (2020)
26. Lugmayr, A., Danelljan, M., Timofte, R., Fritsche, M., Gu, S., Purohit, K., Kandula, P., Suin, M., Rajagoapalan, A., Joon, N.H., et al.: Aim 2019 challenge on real-world image super-resolution: Methods and results. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 3575–3583. IEEE (2019)
27. Maeda, S.: Unpaired image super-resolution using pseudo-supervision. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
28. Ren, H., El-Khamy, M., Lee, J.: Image super resolution based on fusing multiple convolution neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 54–61 (2017)
29. Shocher, A., Cohen, N., Irani, M.: zero-shot super-resolution using deep internal learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3118–3126 (2018)
30. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3147–3155 (2017)
31. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
32. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 0–0 (2018)
33. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4), 600–612 (2004)

34. Wei, P., Lu, H., Timofte, R., Lin, L., Zuo, W., Pan, Z., Xi, T., Fan, Y., Zhang, G., Liu, J., Han, J., Ding, E., Xie, T., Shen, Y., Zhang, J., Jia, Y., Cheng, K., Wu, C., Lin, Y., Liu, C., Peng, Y., Zou, X., Luo, Z., Yao, Y., Xu, Z., Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Ahn, K.H., Kim, J.H., Choi, J.H., Lee, J.S., Zhao, T., Zhao, S., Han, Y., Kim, B.H., Baek, J., Wu, H., Xu, D., Zhou, B., Guan, W., Li, X., Ye, C., Li, H., Zhong, H., Shi, Y., Yang, Z., Yang, X., Zhong, H., Li, X., Jin, X., Wu, Y., Pang, Y., Liu, S., Liu, Z.S., Wang, L.W., Li, C.T., Cani, M.P., Siu, W.C., Zhou, Y., Umer, R.M., Micheloni, C., Cong, X., Gupta, R., Ahn, K.H., Kim, J.H., Choi, J.H., Lee, J.S., Almasri, F., Vandamme, T., Debeir, O.: Aim 2020 challenge on real image super-resolution: Methods and results (2020)
35. Wei, P., Lu, H., Timofte, R., Lin, L., Zuo, W., et al.: AIM 2020 challenge on real image super-resolution. In: European Conference on Computer Vision Workshops (2020)
36. Zhang, L., Zhang, H., Shen, H., Li, P.: A super-resolution reconstruction algorithm for surveillance images. *Signal Processing* **90**(3), 848–859 (2010)
37. Zhang, W., Liu, Y., Dong, C., Qiao, Y.: Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3096–3105 (2019)
38. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 286–301 (2018)
39. Zhang, Y., Li, K., Li, K., Zhong, B., Fu, Y.: Residual non-local attention networks for image restoration. *arXiv preprint arXiv:1903.10082* (2019)
40. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2472–2481 (2018)