IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 21, 2022, accepted May 14, 2022, date of publication May 18, 2022, date of current version June 3, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3176441

Frequency-Based Enhancement Network for Efficient Super-Resolution

PARICHEHR BEHJATI[1], PAU RODRIGUEZ[2], CARLES FERNÁNDEZ TENA[3], ARMIN MEHRI[1], F. XAVIER ROCA[1], SEIICHI OZAWA[4], (Senior Member, IEEE), AND JORDI GONZÀLEZ[1], (Member, IEEE)

[1]Computer Vision Center, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
[2]ServiceNow Research, Montreal, QC 95054, Canada
[3]Oxolo GmbH, 10965 Hamburg, Germany
[4]Center for Mathematical and Data Sciences, Kobe University, Kobe 657-8501, Japan

Corresponding author: Parichehr Behjati (pbehjati@cvc.uab.cat)

This work was supported in part by the Spanish Ministry of Economy and Competitiveness (MINECO), and in part by the European Regional Development Fund (ERDF) under Project PID2020-120311RB-I00/AEI/10.13039/501100011033.

ABSTRACT Recently, deep convolutional neural networks (CNNs) have provided outstanding performance in single image super-resolution (SISR). Despite their remarkable performance, the lack of high-frequency information in the recovered images remains a core problem. Moreover, as the networks increase in depth and width, deep CNN-based SR methods are faced with the challenge of computational complexity in practice. A promising and under-explored solution is to adapt the amount of compute based on the different frequency bands of the input. To this end, we present a novel Frequency-based Enhancement Block (FEB) which explicitly enhances the information of high frequencies while forwarding low-frequencies to the output. In particular, this block efficiently decomposes features into low- and high-frequency and assigns more computation to high-frequency ones. Thus, it can help the network generate more discriminative representations by explicitly recovering finer details. Our FEB design is simple and generic and can be used as a direct replacement of commonly used SR blocks with no need to change network architectures. We experimentally show that when replacing SR blocks with FEB we consistently improve the reconstruction error, while reducing the number of parameters in the model. Moreover, we propose a lightweight SR model — Frequency-based Enhancement Network (FENet) — based on FEB that matches the performance of larger models. Extensive experiments demonstrate that our proposal performs favorably against the state-of-the-art SR algorithms in terms of visual quality, memory footprint, and inference time. The code is available at https://github.com/pbehjatii/FENet

INDEX TERMS Deep learning, frequency-based methods, lightweight architectures, single image super-resolution.

I. INTRODUCTION

Single image super-resolution (SISR) has recently received a considerable amount of attention from both academia and industry. The purpose of SISR is to reconstruct a high-resolution (HR) image from its low-resolution observation (LR). This offers an opportunity for overcoming resolution limitations in various computer vision applications such as medical imaging [50], security and surveillance [47]. In general, SISR is an inverse ill-posed problem since multiple HR images can map to the same LR input. To tackle such an inverse problem, numerous image SR methods have been proposed [16] based on deep neural architectures [9], [34], [36], [61] and shown prominent performance.

Convolutional Neural Networks (CNNs) have recently achieved unprecedented success in various problems [18], [57]. The powerful feature representation and end-to-end training paradigm of CNNs make them a promising approach to SISR. Recently, most CNN-based SR methods focus on elaborate architecture designs such as residual learning [2], [5], [31], [32] and dense connections [25], [67]. Although significant progress has been made, as discussed in [21], [46], texture details of the LR images often tend to be smoothed in the super-resolved results since most existing

The associate editor coordinating the review of this manuscript and approving it for publication was Xianzhi Wang.

FIGURE 1. The visual comparison of SR results by the networks with different building modules for scale factor ×4. The residual block is used as building module for EDSR. In EDSR-FEB, we replace residual block with proposed FEB.

CNN-based SR methods do not pay enough attention to the limited high-frequency information in the LR images. In natural images, information is conveyed at different frequencies. The output feature maps of a convolutional layer can also be seen as a mixture of information at lower and higher frequencies. The lower frequency information is composed of global structures and textures that can directly be forwarded to the final HR output without substantial computations. The higher frequency information consists of fine details where more complex restoring functions are expected. At this point, leading CNN-based methods such as EDSR [37] and RDN [67] overlook the fact that most of the low-frequency information is already contained in the input. As a result, these models spend the same amount of computation treating low- and high-frequency information and lack flexible modulation ability in dealing with them, which ends up the representational ability of the network. Please note that, in this paper, the term *frequency* refers to low- and high-frequency features, and is not related to the frequency domain.

Previous works address this problem by incorporating attention mechanisms [9], [61], [66] into the networks to model interdependencies among spatial locations, channels, or both. The common idea behind attention-based SR methods is to adjust network architectures so that they produce rich feature representations. However, as SR networks are so diverse, the attention module is usually designed solely for a specific network structure [55]. Recently, various SR methods such as multi-branch networks [33], [60] and progressive reconstruction methods [35], [69] mainly focus on refining the high-frequency texture details. Although these methods delivered impressive results, they demand substantial memory and computational resources. Therefore, the efficient reconstruction of high-frequency details in SISR is still a challenge today.

In this paper, we address the aforementioned problems from a different perspective. Instead of designing deep and complex networks or adding various shortcut connections to strengthen feature representations, we introduce a novel Frequency-based Enhancement Block (FEB) which is able to separate features into low and high frequencies while also enabling efficient communication among them. Since low frequencies are preserved by downsampling operations and thus can be recovered directly from the input, FEB assigns more computational capacity to high frequencies. The proposed FEB gradually and iteratively enhances high-frequency feature maps during training while preserving low-frequency information, resulting in more accurate features that improve reconstruction quality.

The proposed FEB offers the following advantages. First, it is generic and can be easily applied to existing SR models without the need of modifying network architectures or requiring hyper-parameters tuning. Second, FEB reduces model parameters in the baseline SR models while simultaneously obtaining better SR performance. In Figure 1, we provide an example of visual quality of EDSR [37], which uses residual blocks [18] as its building module. It can be observed that, when we replace residual blocks with our blocks (EDSR-FEB), the network obtains better visual quality while reducing the number of parameters.

Based on FEB, we build a lightweight SR network named Frequency-based Enhancement Network (FENet), illustrated in Fig 2. Our network leads to significant improvements for single image SR, surpassing SR networks with complicated skip connections and concatenations. In summary, these are the main contributions of the paper:

- We propose a novel Frequency-based Enhancement Block (FEB) to perform frequency-based computation. Such a mechanism allocates more computation to high-frequency bands, allowing the network to focus on more informative features and improve its discriminative capabilities.
- The proposed block leads to the reduction of parameters by half in the baseline SR models while achieving better SR performance.
- We propose a lightweight Frequency-based Enhancement Network (FENet) for fast and accurate image super-resolution. Extensive experiments on a variety of public datasets demonstrate the superiority of the proposed architecture over state-of-the-art models, in terms of both quantitative and visual quality.

II. RELATED WORK

In recent years, the field of image SR has been dominated by CNNs, which achieve state-of-the-art performance [5], [36], [42], [43], [54], [61]. Here, we focus our discussion on the approaches that are most related to our work.

A. EVOLUTION OF ARCHITECTURES FOR SR

Recently, CNN-based methods have dramatically boosted the performance of image SR, due to their strong nonlinear representational power. They learn mappings between LR and HR images from large-scale paired datasets. Since the advent of SRCNN [13], a three layer CNN, a great number of CNN-based methods have been proposed to improve model representation ability by using more elaborate neural network architecture designs. Kim *et al.* [26] first pushed the depth

of SR networks to 20 with the help of residual learning, outperforming SRCNN by a large margin. Ledig *et al.* [31] employed residual blocks proposed in [18] to construct deeper network (SRResNet) for image SR, which was further improved by EDSR [37] and MDSR [37] by removing unnecessary modules (*e.g.,* batch normalization) from the residual blocks. By using effective building modules, image SR networks became deeper and yielded better performance. Later, in order to employ hierarchical features from all the convolutional layers in deep networks, dense blocks started being employed in SR architectures [3], [22], [25], [53], [63]. More recently, Zhang *et al.* [67] and Liu *et al.* [39] also used dense and residual connections in RDN and RFANet to utilize information from the whole feature hierarchy. In addition to residual and dense blocks, Li *et al.* [32] and Lan *et al.* [30] proposed a multi-scale block to explore the multi-scale information of LR images. Although these existing CNN-based SR approaches have provided outstanding performance, they devoted to designing deeper and wider network to enhance their representational learning capacity. Increase in depth and width has also raised computational demands and memory consumption. This makes modern architectures less applicable in practice.

Numerous lightweight models have been proposed to alleviate the aforementioned computational burden. For example, DRCN [27] was the first to apply recursive algorithm to SISR to reduce the number of parameters by reusing them multiple times. Tai *et al.* [51] and Ahn *et al.* [2] improved DRCN by combining the recursive and residual network schemes in order to achieve better performance with even fewer parameters. Likewise, Behjati *et al.* [5] and Jiang *et al.* [25] also joined residual connections and recursive layers to reduce the computational cost. On the other hand, LapSRN [28] employed a pyramidal framework to increase the image size gradually. By doing so, LapSRN effectively performed SISR on extremely low-resolution cases. Chu *et al.* [11] and Ahn and Cho [1] introduced neural architecture search strategies to automatically build an SR model given certain constraints. Meanwhile, Hui *et al.* [24] proposed an information multi-distillation block (IMDB) that extracted features at a granular level with the channel splitting strategy. More recently, Luo *et al.* [40] proposed lattice blocks that applied so-called butterfly structures to combine residual blocks. Later, Xuehui Wang and Chen. Reference [58] proposed an attentive feature block to utilize auxiliary features of previous layers for facilitating features learning of the current layer. Li *et al.* [34] proposed a linearly-assembled pixel-adaptive regression network, which casts the direct LR to HR mapping learning into a linear coefficient regression task. Recently, to simplify the challenges of directly super-resolving details, some authors adopted the progressive structure to reconstruct HR images in a stage-by-stage upscaling manner [36], [38], [69].

By considering that there are different types of information within and across feature maps which have a different contribution for image SR, the aforementioned SR approaches cannot capture low- and high-frequency feature representations separately in the process of feature embedding, thus hindering their representational ability [21], [46], [66].

B. FREQUENCY BASED SR METHODS

It is well known that high-frequency information (*e.g.* texture, edges) is significant for SISR. Li *et al.* [35] proposed a super-resolution feedback network (SRFBN) based on a recurrent architecture design. The network is based on a feedback block that consists of several projection groups. Each projection group first finds high-resolution features (via deconvolution) and then generates low-resolution features (via convolution). As a result, this network is able to gradually recover high-frequency components. Later, Haris *et al.* [17] proposed a method to refine high-frequency texture details with a series of up and downsampling layers that are densely connected with each other to combine HR images from multiple depths in the network. More recently, Qiu *et al.* [46] and Yang and Lu [60] proposed multi-branch architectures. In these methods, one branch is responsible for capturing high-frequency features such as texture and edge, and another is to learn low-frequency features such as image outline and contour. Similarly, Li *et al.* [33] introduced the octave convolution to image SR which uses two branches to perform information update and frequency communication between low- and high-frequency features.

Although these existing SR approaches have made good efforts to improve SR performance, they tend to increase the amount of compute on high-frequency information by increasing the overall number of operations of the model, without paying attention to model complexity. The increase in complexity due to the independent treatment of multiple frequencies is a key issue that limits the performance of these deep CNN-based methods.

C. ATTENTION BASED SR METHODS

Attention mechanism has demonstrated great superiority in improving performance of CNNs for various computer vision tasks [20], [57]. Hu *et al.* [20] introduced squeeze-and-excitation (SE) block that models channel-wise relationships in a computationally efficient manner and enhances the representational ability of the network, showing its effectiveness on image classification. CBAM [57] modified the SE block to exploit both spatial and channel-wise attention. Zhang *et al.* [66] first incorporated SE [20] with SR and pushed the state-of-the-art performance of SISR. More recent works, such as [9], [21], [29], [43], [44], [58], [59], [61], extend this idea by adopting different spatial attention mechanisms or designing advanced attention blocks.

All above-mentioned approaches improve CNNs for image SR by either refining architectural designs or adding complexity to hand-designed blocks. Conversely, our proposal is able to efficiently restore textures at different frequencies. Such mechanism helps the network to explicitly allocate computation to high-frequency features, thus improving the discriminative capabilities of the network.

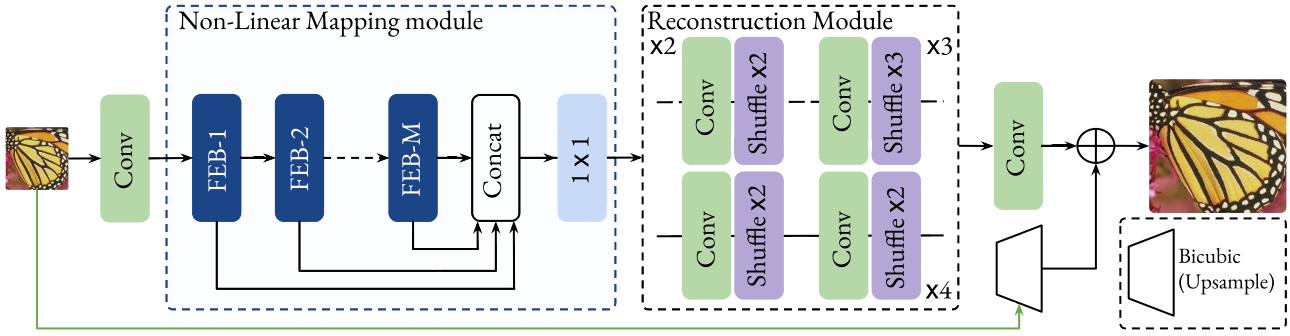


FIGURE 2. Proposed Frequency-based Enhancement Network (FENet) for SISR, which consists of non-linear mapping and reconstruction modules.

III. FREQUENCY-BASED ENHANCEMENT NETWORK

In this section, We first describe the overall network architecture. Next, we detail the proposed Frequency-based Enhancement Block (FEB). Finally, we discuss the differences between the proposed method and similar related works.

A. NETWORK OVERVIEW

As shown in Fig 2, the overall network architecture of Frequency-based Enhancement Network (FENet) consists of a non-linear mapping module and a reconstruction module. Let's denote as I_{LR} and I_{SR} the input and output of FENet, respectively. Following [2], [5], [40], [58], [68], we apply only one 3×3 convolutional layer ($\mathcal{H}$) to extract the initial features H_0 from the LR input image:

$$H_0 = \mathcal{H}(I_{LR}). \tag{1}$$

It is worth noting that only one convolutional layer is used here for lightweight design.

Then, we use the non-linear mapping module, which consists of several stacked FEBs to generate new powerful representations, which can be formulated as

$$H_k = \mathcal{B}_k(H_{k-1}), \quad k = 1, \ldots, M, \tag{2}$$

where $\mathcal{B}_k$ denotes mapping function of the k-th FEB. H_{k-1} represents the features from the previous adjacent FEB, and M is the total number of FEBs.

Inspired by [24], [32], [58], [67], we apply a feature fusion strategy to integrate the features from all the FEBs. This strategy helps to extract more hierarchical contextual information. The fusion operation is formulated as

$$H = \mathcal{F}([H_1, H_2, \ldots, H_M]) \tag{3}$$

where $[H_1, H_2, \ldots, H_M]$ refers to the concatenation of feature maps produced by FEBs and $\mathcal{F}$ is a 1×1 convolutional operation.

Finally, we utilize the reconstruction module that contains convolutional layers and pixelshuffle layers [49] to upsample the features to the HR size. In addition, we incorporate a global connection path (green line in the Fig 2) to grant access to the original LR information and facilitate the back-propagation of the gradients, in which only a bicubic interpolation is applied to the input I_{LR}. Therefore, we obtain:

$$I_{SR} = \mathcal{R}(H) + \text{Bicubic}^{\uparrow}(I_{LR}) \tag{4}$$

where $\mathcal{R}$ is the reconstruction module, and I_{SR} is the final output of the network.

To optimize the network parameters, we adopt L_1 loss as a cost function for training. Given a training set with N pairs of LR images and HR counterparts, denoted by $\{I^i_{LR}, I^i_{HR}\}_{i=1}^N$, the network is optimized to minimize the L_1 loss function:

$$L_1(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \|I_{SR} - I_{HR}\|_1 , \tag{5}$$

where $\boldsymbol{\theta}$ denotes the parameter set.

B. FREQUENCY-BASED ENHANCEMENT BLOCK (FEB)

A natural image can be decomposed into a low frequency component that describes smoothly changing structures and a high-frequency component that describes the rapidly changing fine details [8], [48]. Similarly, we argue that the output feature maps of a convolutional layer can also be decomposed into features of different frequencies, and propose an efficient Frequency-based Enhancement Block (FEB) which naturally decomposes low and high frequencies at feature level. The high-frequency information part is processed by higher-complexity operations (in number of parameters and non-linearities), whereas the lower-frequency part is processed by lower-complexity operations to compensate for the increase of computation. As a result, the proposed approach learns discriminative representations in order to efficiently achieve more accurate reconstructions.

As demonstrated in Fig 3, the proposed FEB contains two pathways, each of which is responsible for a different functionality. Each pathway has a 1×1 convolutional layer at the beginning. Given the input $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$, where C denotes the number of channels and $H \times W$ the spatial dimensions, we have

$$\mathbf{X}_1 = \mathcal{F}'_{split}(\mathbf{X}) \tag{6}$$

$$\mathbf{X}_2 = \mathcal{F}''_{split}(\mathbf{X}) \tag{7}$$

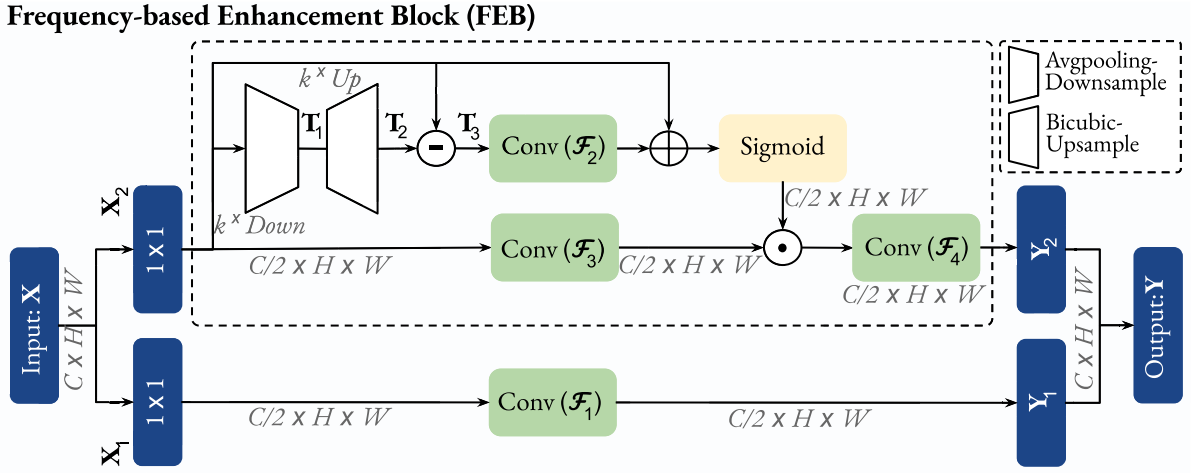


FIGURE 3. Schematic illustration of the proposed Frequency-based Enhancement Block (FEB). As it can be seen, the original filters are separated into two processing lines, each of which is in charge of a different functionality. More details in Section III-B.

where $\{\mathbf{X}_1, \mathbf{X}_2\}$ only have half of the channel number of $\mathbf{X}$. $\mathcal{F}'_{split}$ and $\mathcal{F}''_{split}$ are two 1×1 convolutional operations, respectively.

Then, the described operations are separately sent into a dedicated pathway for collecting different types of information (*i.e.* low- and high-frequency information). The first pathway targets at retaining the original information (low-frequency). To save computation, we perform only a simple 3×3 convolutional operation to capture the global layout and coarse details as follows:

$$\mathbf{Y}_1 = \mathcal{F}_1(\mathbf{X}_1), \tag{8}$$

where $\mathbf{Y}_1$ is the output of the 3×3 convolutional layer ($\mathcal{F}_1$).

In the second pathway, we first apply an average pooling layer upon $\mathbf{X}_2$, yielding $\mathbf{T}_1$

$$\mathbf{T}_1 = \text{AvgPool}^{\downarrow}(\mathbf{X_2}, k), \tag{9}$$

where k denotes the kernel size of the pooling layer and the size of the intermediate feature map $\mathbf{T}_1$ is $\frac{C}{2} \times \frac{H}{k} \times \frac{W}{k}$. Each value in $\mathbf{T}_1$ can be viewed as the average intensity of each specified small area of $\mathbf{X}_2$. After that, $\mathbf{T}_1$ is upsampled via a bicubic interpolation operator to produce a new tensor $\mathbf{T}_2$ of the same size as $\mathbf{X}_2$

$$\mathbf{T}_2 = \text{Bicubic}^{\uparrow}(\mathbf{T}_1, k), \tag{10}$$

where $\mathbf{T}_2$ contains averaged information and it can be regarded as a smoother version of the original $\mathbf{X}_2$. Then, in order to obtain the high-frequency information, $\mathbf{T}_2$ is element-wise subtracted from $\mathbf{X}_2$:

$$\mathbf{T}_3 = \mathbf{X}_2 - \mathbf{T}_2, \tag{11}$$

The visual activation maps of $\mathbf{X}_2$, $\mathbf{T}_2$ and high-frequency information ($\mathbf{T}_3$) are also shown in Fig 4. It can be observed that $\mathbf{T}_2$ is smoother than $\mathbf{X}_2$ as it is the average information of $\mathbf{X}_2$. Meanwhile, $\mathbf{T}_3$ retains the details and edges. Now, the

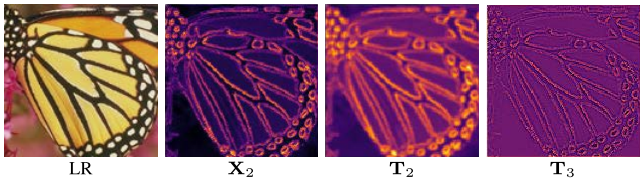


FIGURE 4. Visual activation feature maps of input $\mathbf{X_2}$, $\mathbf{T_2}$, and obtained high-frequency information ($\mathbf{T_3}$).

high-frequency enhancement operation can be formulated as follows

$$\mathbf{Y}'_2 = \sigma(\mathcal{F}_2(\mathbf{T}_3) + \mathbf{X}_2) \cdot \mathcal{F}_3(\mathbf{X}_2), \tag{12}$$

where σ is the sigmoid function, and $\mathcal{F}_2$ and $\mathcal{F}_3$ are two 3×3 convolutional layers, respectively. As shown in (12), we use $\mathbf{X}_2$ as residuals to form the weights, which is found beneficial. Then the output of the second pathway can be written as

$$\mathbf{Y}_2 = \mathcal{F}_4(\mathbf{Y}'_2), \tag{13}$$

where $\mathcal{F}_4$ is a 3×3 convolutional operation. Finally, both intermediate outputs of the first and second pathways $\{\mathbf{Y}_1, \mathbf{Y}_2\}$ are concatenated together as the output $\mathbf{Y} \in \mathbb{R}^{C \times H \times W}$ to obtain a rich feature representation.

Compared to other works such as [33], [60], which require a considerably large amount of computations for decomposing features of different frequencies, FEB can separate the low- and high-frequency feature representations in an efficient way and focus on reconstructing the high-frequency ones.

C. DISCUSSION

1) DIFFERENCE TO PROMINENT SR BLOCKS

Prominent SR blocks such as residual blocks [37] or dense blocks [53] process low- and high-frequency information

simultaneously by the same convolution operations and do not discriminate the computation of features by their frequencial components. Therefore, some local details of LR images cannot be effectively utilized for HR reconstruction, leading to blurry super-resolved results [33]. In contrast, our proposal treats different frequencies in a heterogeneous way and also models inter-channel dependencies, which consequently enrich the output feature. Moreover, FEB benefits SR approaches by reducing the number of parameters while achieving superior SR performance.

2) DIFFERENCE TO ATTENTION-BASED METHODS

Our work is quite different from existing methods such as [12], [21], [43], [66] which rely on supplementary attention blocks and require additional learnable parameters. In contrast our approach internally changes the way of exploiting convolutional filters of convolutional layers, and hence require no additional learnable parameters. In the following experiment section, we will demonstrate without any extra learnable parameters, FEB can yield significant improvements over baselines and other attention-based SR approaches. Moreover, it is complementary to attention mechanisms, and also benefit from their inclusion into the pipeline.

IV. EXPERIMENTAL RESULTS

In this section, we first conduct an ablation study to validate the effectiveness of the proposed FEB. Then, we systematically compare FENet with state-of-the-art SISR algorithms on five commonly used benchmark datasets.

A. SETTINGS

1) DATASETS AND METRICS

Following [67], we use 800 high-quality images from the DIV2K dataset [52] for training. We evaluate our model on several benchmark datasets: Set5 [6], Set14 [62], B100 [4], Urban100 [23], and Manga109 [41], each with diverse characteristics. To keep the consistency with previous works [11], [25], [29], [33], [34], [36], [58], [61], we use Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [56] as the assessment methods to evaluate image reconstruction accuracy. PSNR evaluates the image by statistically measuring distortion values between the reconstructed image and the ground-truth image. The higher the PSNR, the better the quality of the reconstructed image. SSIM measures the structural similarity between two images based on luminance, contrast, and structure. The SSIM values range between 0 to 1, 1 means perfect matching the reconstructed image with the original one. All results are evaluated on the luminance channel (Y). In addition to PSNR and SSIM, we adopt Perceptual Index (PI) [7] to evaluate reconstructed image perceptual quality accurately. PI has a high correlation with human-opinion scores and can avoid the situation where over-smoothed images may present a higher PSNR and SSIM when the performances of the two methods are similar. The lower PI value denotes the better perceptual quality.

2) DEGRADATION MODELS

To fairly compare against existing works, we adopt bicubic downsampling (denoted as **BI**) as our standard degradation model for generating LR images from ground truth HR images, at $\times 2$, $\times 3$, and $\times 4$ scales. Moreover, to comprehensively illustrate the efficacy of the proposed FEB, we further adopt two other multi-degradation models as in [67]. We define **BD** as a degradation model that performs bicubic downsampling on HR images at $\times 3$ scale, and then blurs them with a Gaussian kernel of size 7×7 and standard deviation 1.6. Additionally, we further produce LR images in a more challenging way: we first bicubic downsample HR images with scaling factor $\times 3$ and then add Gaussian noise with noise level 30 (denoted as **DN**).

3) IMPLEMENTATION DETAILS

During training, data augmentation is carried out by means of random horizontal flips and 90° rotation. At each training mini-batch, 64 LR RGB patches of size 64×64 are provided as inputs. We train FENet using an ADAM optimizer with learning rate 10^{-3}. The learning rate is halved every 2×10^5 iterations. We set the number of FEB to 12 in our FENet. Our network has been implemented using PyTorch, and trained on a NVIDIA RTX 3090 GPU.

TABLE 1. Average PSNR obtained when either low- or high-frequency path is deactivated inside the FEB on five benchmark datasets with scale factor ×4.

Configurations	1	2	3
Low-frequency path	✓	×	✓
High-frequency path	×	✓	✓
Params	706K	771K	675K
Set5	32.02	32.06	**32.24**
Set14	28.46	28.43	**28.61**
B100	27.42	27.44	**27.63**
Urban100	25.84	25.86	**26.20**
Manga109	30.12	30.20	**30.46**

B. ABLATION STUDY

1) THE IMPORTANCE OF FEB

In this section, we conduct ablation experiments to explore the influence of each pathway (low- and high-frequency paths) inside the proposed FEB on the reconstruction performance. Therefore, we use FENet as the basic network and run the following experiments: (1) deactivating low-frequency path (Y_1) in FEB; (2) deactivating high-frequency path (Y_2) in FEB, and (3) activating both low- and high-frequency paths. To keep the number of parameters similar, we use 8 and 6 FEBs in the first two experiments respectively without channel reduction.

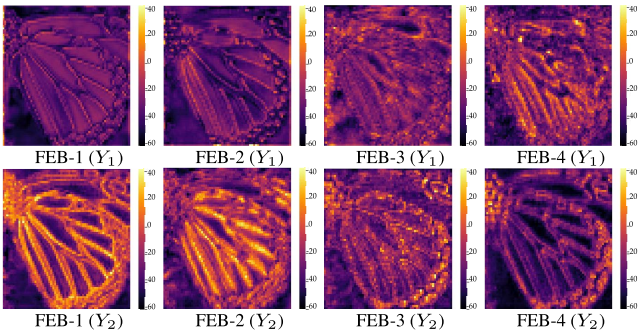


FIGURE 5. Average feature maps of low-frequency (Y_1 in (8)) and high-frequency (Y_2 in (13)) paths.

As reported in Table 1, we observe a significant performance drop when either low- or high-frequency path is deactivated in FEB. This is mainly because: 1) when low-frequency path (Y_1) is deactivated in FEB the high-frequency path (Y_2) focuses too strongly on high-frequency details, smoothing other important aspects of the input that should be preserved by the low-frequency path; 2) the network without high-frequency path (Y_2) processes low- and high-frequency information simultaneously by the same convolution operations, and do not explicitly extrude the high frequencies from image features. Thus, some local details of the LR image cannot be effectively utilized for HR reconstruction.

In Fig 5, we additionally visualize the average feature maps of low-frequency (Y_1 in (8)) and high-frequency (Y_2 in (13)) paths within the first four FEBs. It can be observed that low-frequency feature maps describe the overall outline of the butterfly, while high-frequency ones represent the edges and textures of the butterfly. This visualization shows how FEB is able to efficiently restore textures at different frequencies and can potentially improve performance.

2) THE EFFECTIVENESS OF FEB

To demonstrate the effectiveness of our proposed FEB scheme, we use FENet as the basic network. To keep the number of parameters similar, we replace the 12 FEBs with 8 residual blocks (RB) [37], 5 dense blocks (DB) [53], 6 information multi-distillation blocks (IMDB) [24], or 4 multi-scale residual blocks (MSRB) [32]. In Table 2, we compare the number of parameters and the performance in PSNR for all methods for scale factor $\times 4$.

As reported in Table 2, the method with FEB outperforms all the methods with different SR blocks with fewer number of parameters. The reason is the proposed block treats different frequencies in a heterogeneous way and thus it improves the performance of super-resolution. This experiments justify that the proposed FEB results more helpful for image SR. We additionally provide visual comparisons (Fig 6) of FENet using different SR blocks for scale factor $\times 4$. It can be observed that the network using FEB obtains better visual quality and represents more diverse structure patterns.

TABLE 2. Average PSNR obtained with FENet when using different SR blocks on five benchmark datasets (scale factor $\times 4$).

Name	RB	DB	IMDB	MSRB	FEB
Params	707K	714K	727K	739K	675K
Set5	32.02	32.06	32.07	32.10	**32.24**
Set14	28.46	28.53	28.53	28.57	**28.61**
B100	27.42	27.51	27.54	27.55	**27.63**
Urban100	25.84	25.87	25.89	25.97	**26.20**
Manga109	30.12	30.20	30.21	30.17	**30.46**

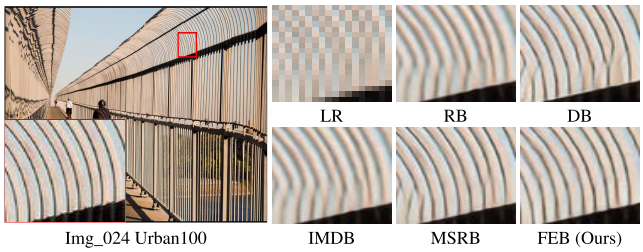


FIGURE 6. Visual comparisons of SR results using FENet with different SR blocks (scale factor $\times 4$).

TABLE 3. Average PSNR obtained with FENet when using different attention mechanisms on five benchmark datasets (scale factor $\times 4$).

Methods	Params	Set14	B100	Urban100
ResNet	707K	28.46	27.42	25.84
ResNet-CA	733K	28.50	27.46	25.89
ResNet-CSAR	782K	28.53	27.50	25.93
FENet	675K	28.61	27.63	26.20
FENet-CA	701K	28.68	27.70	26.27
FENet-CSAR	750K	28.70	27.73	26.30

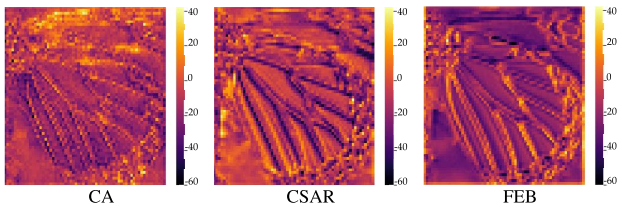


FIGURE 7. Average feature maps of the output Y_2' in (12), the outputs of CA and CSAR in residual blocks, respectively.

3) ATTENTION MECHANISMS VS FEB

To further verify the effectiveness of FEB, we use a ResNet architecture, i.e., a regular architecture composed of 8 stacked residual blocks. Then, we integrate two commonly used attention mechanism namely channel attention (CA) [66] (*ResNet-CA*) and channel-wise and spatial attention residual [21] (*ResNet-CSAR*) into residual blocks as done in [66], respectively. Furthermore, we replace 8 residual blocks with 12 FEBs (*FENet*) and integrated the two mentioned attention mechanisms into FEBs and named them as *FENet-CA* and *FENet-CSAR*.

As reported in Table 3, ResNet-CSAR and ResNet-CA obtain better performance than ResNet but they require additional learnable parameters. Quite differently, FENet does not rely on any extra learnable parameters since it heterogeneously exploits the convolutional filters and thus achieves better performance than ResNet-CSAR and ResNet-CA. It should also be mentioned that the proposed FEB is also compatible with the above mentioned attention mechanisms. For example, when adding CA blocks to each FEB of FENet (FENet-CA), we can further gain another 0.07dB in average. This also indicates that our approach is orthogonal to this kind of supplementary attention modules.

To dig deeper into difference between the proposed block and attention-based approaches, we visualize the average feature map of the output of $\mathbf{Y}_2'$ in (12), the outputs of CA, and CSAR attentions in the residual blocks in Fig 7. Our network should focus on high-frequency components (*i.e.* edges and contours) and suppress the smooth area of the original input image. Compared with the CA and CSAR, feature maps acquired from $\mathbf{Y}_2'$ in (12) contain more negative values, showing a stronger effect of suppressing the smooth area of the input image as well as directing computations towards edges and details. This visualization indicates that the network with FEB can generate richer and more discriminative feature representations than the different attention mechanisms.

4) GENERALIZATION ABILITY

To demonstrate the generalization ability of the proposed structure, we select two state-of-the-art SR networks with different model sizes, called EDSR [37] and RCAN [66]. The EDSR contains 32 stacked residual blocks with 256×256 filters. The RCAN consists of 200 residual channel attention blocks with 64×64 filter sizes. We replace their building blocks with FEBs. The corresponding networks with FEB are named as *EDSR-FEB* and *RCAN-FEB*, respectively. For fair comparison, all networks are trained on their default settings.

As shown in Table 4, EDSR-FEB has an improvement of 0.08dB in average with almost $\times 2$ fewer number of parameters (parameters: 28M) compared to the original EDSR (parameters: 43M). Moreover, the improvement by RCAN-FEB is also higher than RCAN with approximately half amount of parameters. From these comparisons, we can easily find that (1) the proposed FEB perform much better than channel attention, (2) for deeper networks, a similar phenomenon can also be observed, (3) FEB reduces the number of parameters by half while achieving better performance.

Fig 1 and 8 additionally show visual comparisons for scale factor $\times 4$. It can be observed that EDSR-FEB and RCAN-FEB can reconstruct sharper and more natural-looking images. This is mainly because FEB can extract high-frequency features and use them for reconstruction.

TABLE 4. Average PSNR obtained with state-of-the-art SR methods when using FEB on five benchmark datasets (scale factor $\times 4$).

Name	EDSR	EDSR-FEB	RCAN	RCAN-FEB
Params	43M	28M	16M	9M
Set5	32.50	**32.58**(+0.08dB)	32.63	**32.70**(+0.07dB)
Set14	28.72	**28.80**(+0.08dB)	28.87	**28.96**(+0.06dB)
B100	27.72	**27.81**(+0.09dB)	27.77	**27.85**(+0.08dB)
Urban100	26.67	**26.76**(+0.09dB)	26.82	**26.89**(+0.07dB)
Manga109	31.02	**31.09**(+0.07dB)	31.22	**31.30**(+0.08dB)

FIGURE 8. The visual comparison of SR results by the networks with different building modules for $\times 4$ scale factor. The residual blocks followed by channel attentions are used as building modules for RCAN. In RCAN-FEB, we replace its blocks with proposed FEBs.

TABLE 5. Average PSNR obtained when FEB using different pooling methods on five benchmark datasets (scale factor $\times 4$). The best performance is shown highlighted and the second best underlined.

Configurations	1	2	3
Max-Bicubic	✓		
Avg-Bicubic		✓	
Conv-Deconv			✓
Params	675K	675K	2325K
Set5	32.17	32.24(+0.07dB)	**32.35 (+0.11dB)**
Set14	28.53	28.61(+0.08dB)	**28.72 (+0.11dB)**
B100	27.54	27.63(+0.09dB)	**27.74 (+0.11dB)**
Urban100	26.12	26.20(+0.08dB)	**26.30 (+0.10dB)**
Manga109	30.38	30.46(+0.08dB)	**30.56 (+0.10dB)**

5) COMPARING POOLING METHODS

In this section, we investigate the influence of different pooling types on the performance. The proposed block adopts average pooling for downsampling and bicubic interpolation for upsampling. In our experiments, we use FENet as the basic network and then replace average pooling operators in all FEBs with maximum pooling operators. As shown in Table 5, using the average pooling operator while keeping the rest of configurations unchanged yields a performance increase of about 0.08dB in average. We argue that this may be due to the fact that, unlike maximum pooling, average pooling builds connections among locations within the whole

TABLE 6. Average PSNR to show the effect of downsampling rate on the performance on Set5 dataset. We record the results in 10×10^4 iterations.

Downsampling Rate	Scales		
	×2	×3	×4
2	37.89	34.22	32.08
3	37.91	34.24	32.10
4	37.94	34.29	32.14
5	**37.95**	**34.31**	**32.15**

pooling window, which can better capture local contextual information.

In addition, we further investigate the behavior of the proposed block (FEB), when the average pooling for downsampling in (9) and bicubic interpolation used for upsampling in (10) are replaced with a convolutional layer and a deconvolutional layer, respectively. As reported in Table 5, it can be observed that the performance as well as the number of parameters of the network increase, when we replace average pooling and bicubic interpolation with learnable operations. Although the performance of the network increases by 0.11dB on average, this leads to a more complex network with more parameters. While weighing the network performance and network complexity, we finally use average pooling and bicubic interpolation for the rest of the experiments, the results are close to the network with conv-deconv operations, but the number of model parameters is only one fourth of it.

6) THE EFFECT OF DOWNSAMPLING RATE

We also investigate how the downsampling rate in FEB influences the image SR performance. In Table 6, we show the performance with different downsampling rates used in FEB. It can be observed that as the downsampling rate increases, slightly better performance is achieved. However, we do not use larger downsampling rates due to two reasons: (1) the resolution of the input features is already very small; (2) higher downsampling rates lead to performance improvements at the expense of more computations due to bicubic operation. Therefore, for the rest of experiments, we set the downsampling rate to 4 for all scale factors, as it still provides significant improvements with a lower computational cost than ×5.

7) THE EFFECT OF INCREASING THE NUMBER OF FEB

As discussed in [37], increasing the depth of the network can effectively improve the performance. In this work, adding the number of FEBs is the simplest way to gain excellent result. For better balancing the model size and performance, we compare the proposed model with the different numbers of FEBs, i.e., 6, 8, 10, and 12. As shown in Table 7, our FENet performance improves rapidly with the growth in number of FEBs. Although the performance of the network would further improve by using more FEBs, we found it leads to diminishing returns with respect to the number of parameters.

TABLE 7. Average PSNR obtained with FENet when using different number of FEBs on five benchmark datasets (scale factor ×4).

Blocks	6	8	10	12	12
Global path	✓	✓	✓	×	✓
Params	379K	477K	572K	675K	675K
Set5	31.98	32.15	32.19	32.21	**32.24**
Set14	28.52	28.54	28.57	28.59	**28.61**
B100	27.51	27.53	27.55	27.58	**27.63**
Urban100	25.90	25.97	26.05	26.16	**26.20**
Manga109	30.07	30.20	30.35	30.39	**30.46**

Therefore, we use 12 FEBs in our experiments. Furthermore, we find by adding a global connection path (green line in Fig 2) to grant the output access to the original LR input is beneficial for reconstruction performance. Discarding this connection decreases performance (-0.04dB on average).

C. COMPARISON WITH STATE-OF-THE-ART METHODS

1) RESULTS WITH **BI** DEGRADATION MODEL

In this section, we compare the proposed FENet with state-of-the-art lightweight models: VDSR [26], DRCN [27], SRDenseNet [53], SEINet [10], SRResNet [31], CARN [2], IMDN [24], SRFBN-S [35], A2F-S [58], CBPN [69], LAPAR-A [34], MADNet [29], FALSR-A [11], DPN [36], HDRN [25], and OISR-RK2 [19]. We have also listed the performance of state-of-the-art large SR methods including EDSR [37], FSN [33], and CASGCN [61] for reference.

Table 8 shows quantitative results when evaluating PSNR and SSIM on five benchmark dataset with different algorithms for scale factors ×2, ×3, and ×4. For a more informative comparison, the number of parameters is also given. From Table 8, we find that FENet only has less than 0.7M parameters but performs favorably against other compared approaches on most datasets. For example, in comparison with SRDenseNet [53] and OISR-RK2 [19], FENet achieves better or competitive results, while only needing 30% and 40% of their parameters, respectively. On the other hand, thanks to the FEB, FENet achieves competitive or better results when compared to the large SR methods. Specifically, FENet outperforms FSN [33] by a large margin at all scales in all datasets with 18× fewer parameters.

In Fig 9, we present some qualitative visual comparisons for the ×4 scale factor. It can be observed that SR images reconstructed by FENet have more refined details, especially in the edges and lines. This further validates the effectiveness of the proposed FEB.

2) RESULTS WITH **BD** AND **DN** DEGRADATION MODELS

Following [67], we also show the SR results with **BD** degradation model and further introduce **DN** degradation model. The proposed FENet is compared with state-of-the-art methods including SPMSR [45], SRCNN [14], FSRCNN [15], VDSR [26], IRCNN_G [64], IRCNN_C [64], and

TABLE 8. Average PSNR/SSIM values for models with the same order of magnitude of parameters. Performance is shown for scale factors ×2, ×3 and ×4 with BI degradation model. The best and second best results are highlighted in red and blue respectively.

Scale	Method	Params	Set5		Set14		B100		Urban100		Manga109	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×2	VDSR [26]	0.7M	37.53	0.9587	33.03	0.9124	31.90	0.8960	30.76	0.9140	37.22	0.9750
	DRCN [27]	1.8M	37.53	0.9587	33.03	0.9124	31.90	0.8960	30.76	0.9140	37.22	0.9750
	SEINet [10]	1M	37.89	0.9598	33.61	0.9160	32.08	0.8984	–	–	–	–
	CARN [2]	1.6M	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256	38.36	0.9765
	SRFBN-S [35]	0.3M	37.78	0.9597	33.35	0.9156	32.00	0.8970	31.41	0.9207	38.06	0.9757
	A2F-S [58]	0.3M	37.79	0.9597	33.32	0.9152	31.99	0.8972	31.44	0.9211	38.11	0.9757
	CBPN [69]	1M	37.90	0.9590	33.60	0.9171	32.17	0.8989	32.14	0.9279	–	–
	MADNet [29]	0.9M	37.94	0.9604	33.46	0.9167	32.10	0.8988	31.74	0.9246	–	–
	FALSR-A[11]	1M	37.82	0.9595	33.55	0.9168	32.12	0.8987	31.93	0.9256	–	–
	HDRN [25]	0.9M	37.75	0.9590	33.49	0.9150	32.03	0.8980	31.87	0.9250	38.07	0.9770
	DPN [36]	0.8M	37.52	0.9586	33.08	0.9129	31.89	0.8958	30.82	0.9144	–	–
	LAPAR-A [34]	0.5M	38.01	0.9605	33.62	0.9183	32.19	0.8999	32.10	0.9283	38.67	0.9772
	IMDN [24]	0.7M	38.00	0.9605	33.63	0.9177	32.19	0.8996	32.17	0.9283	38.88	0.9774
	OISR-RK2 [19]	1.4M	38.02	0.9605	33.62	0.9178	32.20	0.9000	32.21	0.9290	–	–
	FENet (Ours)	0.6M	38.08	0.9608	33.70	0.9184	32.20	0.9001	32.18	0.9287	38.89	0.9775
	EDSR [37]	43M	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.93	0.9351	39.10	0.9773
	CASGCN [61]	14M	38.26	0.9615	34.02	0.9213	32.36	0.9020	33.17	0.9377	39.41	0.9785
	FSN [33]	7.3M	37.68	0.9605	33.51	0.9180	32.09	0.9015	31.68	0.9248	–	–
×3	VDSR [26]	0.7M	33.66	0.9213	29.77	0.8314	28.82	0.7976	27.14	0.8279	37.22	0.9750
	DRCN [27]	1.7M	33.82	0.9226	29.76	0.8311	28.80	0.7963	27.15	0.8276	32.24	0.9343
	CARN [2]	1.6M	34.29	0.9255	30.29	0.8407	29.06	0.8034	28.06	0.8493	33.50	0.9440
	SRFBN-S [35]	0.4M	34.20	0.9255	30.10	0.8372	28.96	0.8010	27.66	0.8415	33.02	0.9404
	A2F-S [58]	0.3M	34.06	0.9241	30.08	0.8370	28.92	0.8006	27.57	0.8392	32.86	0.9394
	MADNet [29]	0.9M	34.26	0.9262	30.29	0.8410	29.04	0.8033	27.91	0.8464	–	–
	HDRN [25]	0.9M	34.24	0.9240	30.23	0.8400	28.96	0.8040	27.93	0.8490	33.17	0.9420
	DPN [36]	0.8M	33.71	0.9222	29.80	0.8320	28.84	0.7981	27.17	0.8282	–	–
	LAPAR-A [34]	0.5M	34.36	0.9267	30.34	0.8421	29.11	0.8054	28.15	0.8523	33.51	0.9441
	IMDN [24]	0.7M	34.36	0.9270	30.32	0.8417	29.09	0.8046	28.17	0.8519	33.61	0.9445
	OISR-RK2 [19]	1.6M	34.39	0.9272	30.35	0.8420	29.11	0.8058	28.24	0.8544	–	–
	FENet (Ours)	0.6M	34.40	0.9273	30.36	0.8422	29.12	0.8060	28.17	0.8524	33.52	0.9444
	EDSR [37]	43M	34.65	0.9280	30.52	0.8462	29.25	0.8093	28.80	0.8653	34.17	0.9476
	CASGCN [61]	14M	34.75	0.9300	30.59	0.8476	29.33	0.8114	28.93	0.8671	34.36	0.9494
×4	VDSR [26]	0.7M	31.35	0.8838	28.01	0.7674	27.29	0.7251	25.18	0.7524	28.83	0.8809
	DRCN [27]	1.8M	31.54	0.8850	29.19	0.7720	27.32	0.7280	25.12	0.7560	29.09	0.8845
	SEINet [10]	1.4M	32.05	0.8934	28.49	0.7783	27.44	0.7325	–	–	–	–
	SRDenseNet [53]	2M	32.00	0.8931	28.50	0.7782	27.53	0.7337	26.05	0.7819	30.41	0.9071
	SRResNet [31]	1.5M	32.05	0.8910	28.53	0.7804	27.57	0.7354	26.07	0.7839	–	–
	CARN [2]	1.6M	32.13	0.8937	28.60	0.7806	27.58	0.7349	26.07	0.7837	30.47	0.9084
	SRFBN-S [35]	0.5M	31.98	0.8923	28.45	0.7779	27.44	0.7313	25.71	0.7719	29.91	0.9008
	A2F-S [58]	0.3M	31.87	0.8900	28.36	0.7760	27.41	0.7305	25.58	0.7685	29.77	0.8987
	CBPN [69]	1.2M	32.21	0.8944	28.63	0.7813	27.58	0.7356	26.14	0.7869	–	–
	MADNet [29]	1M	32.11	0.8939	28.52	0.7799	27.52	0.7340	25.89	0.7782	–	–
	HDRN [25]	0.9M	32.23	0.8960	28.58	0.7810	27.53	0.7370	26.09	0.7870	30.43	0.9080
	DPN [36]	0.8M	31.42	0.8849	28.07	0.7688	27.30	0.7256	25.25	0.7546	–	–
	LAPAR-A [34]	0.7M	32.15	0.8944	28.61	0.7818	27.61	0.7366	26.14	0.7871	30.42	0.9074
	IMDN [24]	0.7M	32.21	0.8948	28.58	0.7811	27.56	0.7353	26.04	0.7838	30.45	0.9075
	OISR-RK2 [19]	1.5M	32.14	0.8947	28.63	0.7819	27.60	0.7369	26.17	0.7888	–	–
	FENet (Ours)	0.6M	32.24	0.8961	28.61	0.7818	27.63	0.7371	26.20	0.7890	30.46	0.9083
	EDSR [37]	43M	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	31.02	0.9148
	CASGCN [61]	14M	32.60	0.9002	28.88	0.7890	27.70	0.7416	26.79	0.8086	31.18	0.9169
	FSN [33]	8M	32.10	0.8959	28.57	0.7874	27.53	0.7438	25.76	0.7817	–	–

TABLE 9. Quantitative results with **BD** and **DN** degradation models. The best and second best results are highlighted in **red** and **blue** respectively.

Methods	Degrad.	Set5		Set14		B100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SPMSR [45]	BD	32.21	0.9001	28.89	0.8105	28.13	0.7740	25.84	0.7856	29.64	0.9003
	DN	–	–	–	–	–	–	–	–	–	–
SRCNN [14]	BD	32.05	0.8944	28.80	0.8074	28.13	0.7736	25.70	0.7770	29.47	0.8924
	DN	25.01	0.6950	23.78	0.5898	23.76	0.5538	21.19	0.5737	23.75	0.7148
FSRCNN [15]	BD	26.23	0.8124	24.44	0.7106	24.86	0.6832	22.04	0.6745	23.04	0.7927
	DN	24.18	0.6932	32.02	0.5856	23.41	0.5556	21.15	0.5682	22.39	0.7111
VDSR [26]	BD	33.25	0.9150	29.46	0.8244	28.57	0.7893	26.61	0.8136	31.06	0.9234
	DN	25.20	0.7183	24.00	0.6112	24.00	0.5749	22.22	0.6096	24.20	0.7525
IRCNN_G [64]	BD	33.38	0.9182	29.63	0.8281	28.65	0.7922	26.77	0.8154	31.15	0.9245
	DN	25.70	0.7379	24.45	0.6305	24.28	0.5900	22.90	0.6429	24.88	0.7765
IRCNN_C [64]	BD	29.55	0.8246	27.33	0.7135	26.46	0.6572	24.89	0.7172	28.68	0.7701
	DN	26.18	0.7430	24.68	0.6300	24.52	0.5850	22.63	0.6205	24.74	0.7701
SRMDNF [65]	BD	34.09	0.9242	30.11	0.8364	28.98	0.8009	27.50	0.8370	32.97	0.9391
	DN	27.74	0.8026	26.13	0.6924	25.64	0.6495	24.28	0.7092	26.72	0.8590
FENet (Ours)	BD	34.60	0.9277	30.57	0.8433	29.22	0.8060	28.39	0.8539	34.03	0.9459
	DN	28.57	0.8164	26.29	0.6945	26.01	0.6611	24.99	0.7369	28.26	0.8611

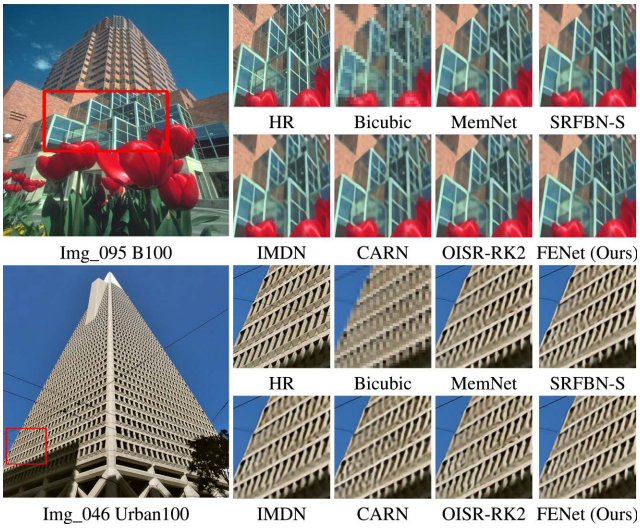


FIGURE 9. Visual results of **BI** degradation model (×4).

FIGURE 10. Visual results of **BD** and **DN** degradation models (×3).

SRMDNF [65]. As shown in Table 9, FENet performs the best on all datasets with **BD** and **DN** degradation models. The significantly better results of our method indicate that FENet adapts well to scenarios with multiple degradation models.

In Fig 10, we show two sets of visual results with **BD** and **DN** degradation models from the standard benchmark datasets. For **BD** degradation model, the proposed FENet suppresses the blurring artifacts and recovers sharper edges. For **DN** degradation model, FENet can not only handle the noise efficiently, but also recover details more accurately. These comparisons further showcase the robustness and effectiveness of our method in handling **BD** and **DN** degradation models.

3) MODEL COMPLEXITY ANALYSIS

In this section, we compare the trade-off between performance, number of parameters and the number of multiplications and additions (Multi-Adds) for our method and existing lightweight SR networks. The Multi-Adds are calculated corresponding to a 1280 × 720 HR image.

Fig 11 shows the PSNR performances of several existing lightweight models, namely VDSR [26], DRCN [27], SRDenseNet [53], SEINet [10], SRResNet [31], CARN [2], IMDN [24], SRFBN-S [35], A2F-S [58], CBPN [69], LAPAR-A [34], MADNet [29], FALSR-A [11], DPN [36], HDRN [25], and OISR-RK2 [19] versus the number of parameters and Multi-Adds with results evaluated on Urban100 for ×4. As shown in Fig 11, FENet achieves

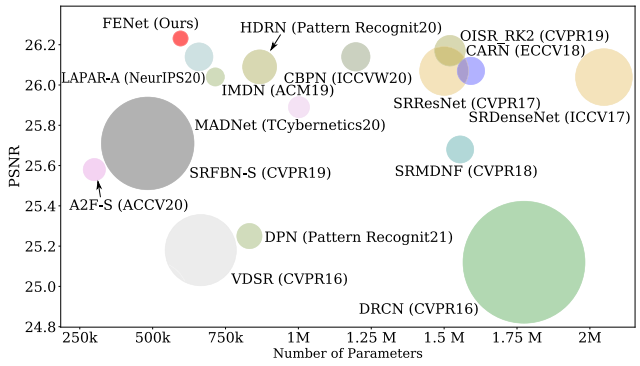


FIGURE 11. Comparing capacity vs. performance for lightweight state-of-the-art SISR models on Urban100 (×4). Circle sizes are set proportional to the number of multiplications and additions (multi-adds).

TABLE 10. Average running time(s) and memory consumption (MB) comparison on Urban100 for ×4.

Methods	Params	Memory	Running Time(s)	PSNR
CARN [2]	1.5M	1,116	0.032	26.07
SRFBN-S [35]	0.5M	2,154	0.031	25.71
SRDenseNet [53]	2.0M	5,531	0.221	26.05
IMDN [24]	0.7M	871	0.028	26.04
A2F-S [58]	0.3M	915	0.032	25.58
LAPAR-A [34]	0.7M	1,240	0.053	26.14
MSRN [32]	8.0M	2,731	0.070	26.04
RCAN [66]	16M	1,531	0.087	26.82
EDSR [37]	43M	2,731	0.035	26.64
FENet (Ours)	0.6M	850	0.018	26.20

state-of-the-art results with less parameters and Multi-Adds operations. This demonstrates that our proposal achieves a better trade-off between model size and reconstruction performance.

4) MEMORY COMPLEXITY AND RUNNING TIME ANALYSIS

Table 10 illustrates the superiority of the proposed FENet in terms of Inference Time (s) and Memory Consumption (MB), when compared to recent light- and heavy-weight state-of-the-art approaches on Urban100 with scale factor ×4. For a fair comparison, we use a single NVIDIA RTX 3090 GPU for evaluation, and their official source code implementations. It can be observed that our model achieves dominant performance in terms of memory usage and time consumption, reflecting its efficiency.

5) PERCEPTUAL METRICS

Perceptual metrics better reflect the human judgment of image quality. In this paper, Perceptual Index (PI) [7] is chosen as the perceptual metric. Table 11 shows the PI for those works with publicly available source code, and the same order of magnitude in terms of parameters. We observe that our proposed model obtains better results than all the compared baselines. This demonstrates the ability of the proposed FENet for generating realistic images.

TABLE 11. Perceptual index comparison of the proposed method with recent lightweight state-of-the-art methods on five datasets for ×4. The lower is better. All of the output SR images are provided officially.

Methods	Params	Set5	Set14	B100	Urban 100	Manga 109
DRCN [27]	1.7M	6.451	5.945	5.897	5.791	5.563
CARN [2]	1.5M	6.297	5.775	5.700	5.540	5.132
SRFBN-S [26]	0.6M	6.451	5.775	5.702	5.549	5.010
SRDenseNet [53]	2M	6.128	5.615	5.653	5.526	4.762
IMDN [24]	0.7M	6.124	5.644	5.659	5.531	4.810
LAPAR-A [34]	0.7M	6.084	5.499	5.532	5.179	4.771
FENet (Ours)	0.6M	5.598	5.495	5.447	5.175	4.761

V. LIMITATIONS AND FUTURE WORK

Although our method is the fastest compared to other SR approaches we have identified the bicubic interpolation operation in (10) as one of the main computational bottlenecks. Thus, we hypothesize that substituting it for a more efficient operation or implementation would effectively speed up our model. Furthermore, the loss function adopted by our method is the distortion-oriented rather than perception-oriented metric, which also limits obtaining better perceptual quality HR images.

In future work, we will explore the extensions of the proposed framework on other image restoration applications, such as deblocking, inpainting, and low-light image enhancement. We also wish to further develop this work by applying our technique to video data. Many streaming services require a large storage to provide high-quality videos. In conjunction with our approach, one may devise a service that stores low-quality videos that go through our SR system to produce high-quality videos on the fly.

VI. CONCLUSION

This paper presents a novel Frequency-based Enhancement Block (FEB). This block is able to naturally decompose features into low and high frequencies and explicitly allocate more computational capacity to high-frequency ones thus improving the discriminative capabilities of the network. The proposed FEB can be easily replaced with commonly used SR blocks. We proved that when replacing SR blocks with FEB we consistently improve the reconstruction error (PSNR: +0.08dB on average) while reducing the number of parameters by half in the model. Furthermore, We showed that the proposed block is orthogonal and complementary to attention-based SR methods. Based on FEB, we proposed a lightweight Frequency-based Enhancement Network (FENet) for accurate image SR. Experimental results on several benchmark datasets demonstrate that our method can achieve superior performance at a moderate size. We hope that the idea of decomposing low- and high-frequency information at the feature level for adaptive computation can

provide the computer vision community with a different perspective on network architecture design.

APPENDIX

A. ABBREVIATIONS

The abbreviations and acronyms used in this paper are first introduced in the text, and for convenience, the list of abbreviation is summarized in Table 12.

TABLE 12. Summary of abbreviations.

List of abbreviations and their associated meanings	
CNN	Convolutional neural network
CA	Channel attention
CSAR	Channel-wise and spatial attention residual
DB	Dense block
FEB	Frequency-based enhancement block
FENet	Frequency-based enhancement network
HR	High-resolution
IMDB	Information multi-distillation block
LR	Low-resolution
MB	Memory consumption
MSRB	Multi-scale residual block
PI	Perceptual index
PSNR	Peak signal-to-noise ratio
RB	Residual block
SISR	Single image super-resolution
SR	Super-resolution
SSIM	Structural similarity index measure

B. DATA DESCRIPTION

In Table 13, we list a number of image datasets commonly used by the SR community and this work. We specifically indicate their amount of HR images, average resolution, image formats, and category keywords.

TABLE 13. List of public image datasets for super-resolution benchmarks.

Dataset	Amount	Avg. Resolution	Format	Category Keywords
Set5	5	(313, 336)	PNG	Baby, bird, butterfly, head, woman
Set14	14	(492, 446)	PNG	Humans, animals, insects, flowers, vegetables, comic, slides, etc.
B100	100	(435, 367)	JPG	Animal, building, food, landscape, people, plant, etc.
Urban100	100	(984, 797)	PNG	Architecture, city, structure, urban, etc.
Manga109	109	(826, 1169)	PNG	Manga volume
DIV2K	1000	(1972, 1437)	PNG	Environment, flora, fauna, handmade object, people, scenery, etc.

REFERENCES

[1] J. Y. Ahn and N. I. Cho, "Multi-branch neural architecture search for lightweight image super-resolution," *IEEE Access*, vol. 9, pp. 153633–153646, 2021.

[2] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 252–268.

[3] Z. An, J. Zhang, Z. Sheng, X. Er, and J. Lv, "RBDN: Residual bottleneck dense network for image super-resolution," *IEEE Access*, vol. 9, pp. 103440–103451, 2021.

[4] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2010.

[5] P. Behjati, P. Rodriguez, A. Mehri, I. Hupont, C. F. Tena, and J. Gonzalez, "OverNet: Lightweight multi-scale super-resolution with overscaling network," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2694–2703.

[6] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L.-A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, p. 135.

[7] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, Sep. 2018, pp. 334–355.

[8] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Physiol.*, vol. 197, no. 3, pp. 551–566, 1968.

[9] R. Chen, H. Zhang, and J. Liu, "Multi-attention augmented network for single image super-resolution," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108349.

[10] J.-S. Choi and M. Kim, "A deep convolutional neural network with selection units for super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 154–160.

[11] X. Chu, B. Zhang, H. Ma, R. Xu, and Q. Li, "Fast, accurate and lightweight super-resolution with neural architecture search," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 59–64.

[12] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11065–11074.

[13] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2014, pp. 184–199.

[14] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.

[15] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 391–407.

[16] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, vol. 30, no. 2, pp. 1–11, Apr. 2011.

[17] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[19] X. He, Z. Mo, P. Wang, Y. Liu, M. Yang, and J. Cheng, "ODE-inspired network design for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1732–1741.

[20] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[21] Y. Hu, J. Li, Y. Huang, and X. Gao, "Channel-wise and spatial feature modulation network for single image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3911–3927, Nov. 2020.

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[23] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5197–5206.

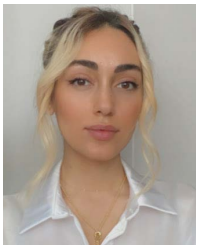
[24] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2024–2032.

[25] K. Jiang, Z. Wang, P. Yi, and J. Jiang, "Hierarchical dense recursive network for image super-resolution," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107475.

[26] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[27] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1637–1645.

[28] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 624–632.
[29] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, and X. Luo, "MADNet: A fast and lightweight network for single-image super resolution," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1443–1453, Mar. 2021.
[30] R. Lan, L. Sun, Z. Liu, H. Lu, Z. Su, C. Pang, and X. Luo, "Cascading and enhanced residual networks for accurate single-image super-resolution," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 115–125, Jan. 2021.
[31] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.
[32] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 517–532.
[33] S. Li, Q. Cai, H. Li, J. Cao, L. Wang, and Z. Li, "Frequency separation network for image super-resolution," *IEEE Access*, vol. 8, pp. 33768–33777, 2020.
[34] W. Li, K. Zhou, L. Qi, N. Jiang, J. Lu, and J. Jia, "LAPAR: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 20343–20355.
[35] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3867–3876.
[36] Y. Liang, R. Timofte, J. Wang, S. Zhou, Y. Gong, and N. Zheng, "Single-image super-resolution—When model adaptation matters," *Pattern Recognit.*, vol. 116, Aug. 2021, Art. no. 107931.
[37] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.
[38] H. Lin and J. Yang, "Light weight IBP deep residual network for image super resolution," *IEEE Access*, vol. 9, pp. 93399–93408, 2021.
[39] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2359–2368.
[40] X. Luo, Y. Xie, Y. Zhang, Y. Qu, C. Li, and Y. Fu, "LatticeNet: Towards lightweight image super-resolution with lattice block," in *Proc. Eur. Conf. Comput. Vis.*, Glasgow, U.K.: Springer, Aug. 2020, pp. 272–289.
[41] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, "Sketch-based Manga retrieval using Manga109 dataset," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21811–21838, 2017.
[42] A. Mehri, P. B. Ardakani, and A. D. Sappa, "LiNet: A lightweight network for image super resolution," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 7196–7202.
[43] A. Mehri, P. B. Ardakani, and A. D. Sappa, "MPRNet: Multi-path residual network for lightweight image super resolution," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2704–2713.
[44] Y. K. Ooi, H. Ibrahim, and M. N. Mahyuddin, "Enhanced dense space attention network for super-resolution construction from single input image," *IEEE Access*, vol. 9, pp. 126837–126855, 2021.
[45] T. Peleg and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2569–2582, Jun. 2014.
[46] Y. Qiu, R. Wang, D. Tao, and J. Cheng, "Embedded block residual network: A recursive restoration model for single-image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4180–4189.
[47] P. Rasti, T. Uiboupin, S. Escalera, and G. Anbarjafari, "Convolutional neural network super resolution for face recognition in surveillance monitoring," in *Proc. Int. Conf. Articulated Motion Deformable Objects*. Cham, Switzerland: Springer, 2016, pp. 175–184.
[48] R. L. DeValois, K. K. D. Valois, and K. K. DeValois, *Spatial Vision*, vol. 14, 1988.
[49] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.
[50] W. Shi, J. Caballero, C. Ledig, X. Zhuang, W. Bai, K. Bhatia, A. M. S. M. de Marvao, T. Dawes, D. O'Regan, and D. Rueckert, "Cardiac image super-resolution with global correspondence using multi-atlas patchmatch," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2013, pp. 9–16.
[51] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3147–3155.
[52] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 114–125.
[53] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4799–4807.
[54] F. Wang and M. Gong, "Single image super-resolution by residual recovery based on an independent deep convolutional network," *IEEE Access*, vol. 9, pp. 43701–43710, 2021.
[55] F. Wang, H. Hu, and C. Shen, "BAM: A balanced attention mechanism for single image super resolution," 2021, *arXiv:2104.07566*.
[56] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[57] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.
[58] X. Wang, Q. Wang, Y. Zhao, J. Yan, L. Fan, and L. Chen, "Lightweight single-image super-resolution network with attentive auxiliary feature learning," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, Nov. 2020, pp. 268–285.
[59] Y. Yan, X. Xu, W. Chen, and X. Peng, "Lightweight attended multi-scale residual network for single image super-resolution," *IEEE Access*, vol. 9, pp. 52202–52212, 2021.
[60] C. Yang and G. Lu, "Deeply recursive low- and high-frequency fusing networks for single image super-resolution," *Sensors*, vol. 20, no. 24, p. 7268, Dec. 2020.
[61] Y. Yang and Y. Qi, "Image super-resolution via channel attention and spatial graph convolutional network," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107798.
[62] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.*, Springer, 2010, pp. 711–730.
[63] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3194–3203.
[64] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.
[65] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3262–3271.
[66] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 286–301.
[67] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.
[68] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong, "Efficient image super-resolution using pixel attention," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2020, pp. 56–72.
[69] F. Zhu and Q. Zhao, "Efficient single image super-resolution via hybrid residual feature learning with compact back-projection network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 2453–2460.



PARICHEHR BEHJATI received the B.Sc. and M.Sc. degrees in computer science from Eastern Mediterranean University. She is currently pursuing the Ph.D. degree in deep learning and computer vision with the Computer Vision Center, Universitat Autònoma de Barcelona. She had previously worked as a Research Assistant with the Computer Science Department, Eastern Mediterranean University, from 2014 to 2016. Her research interests include deep learning and computer vision.

PAU RODRIGUEZ received the Master of Artificial Intelligence degree from KU Leuven and the Ph.D. degree in deep learning and computer vision from the Computer Vision Center, Universitat Autònoma de Barcelona (CVC-UAB). He is a Research Scientist with ServiceNow Research and an Adjunct Professor with UAB. His main research interest includes machine learning methods that generalize with fewer labeled data, closer to how humans learn. He would like AI to solve the most important problems of humanity. He is a member of ELLIS.

CARLES FERNÁNDEZ TENA received the Ph.D. degree *(cum laude)* in computer vision and AI from the Universitat Autònoma de Barcelona, in 2010. He worked for the face recognition company Herta, from 2010 to 2021, where he led the Research and Development Department as the Director of Research and the CTO. He is currently the Computer Vision Lead with Oxolo, working on AI-based realistic face and body synthesis. He has published more than 60 scientific articles in peer-reviewed international journals and conferences and participated in several FP7 and H2020 projects. His research interests include deep learning, computer vision, and HPC for face and video analytics. He received the 2010 Extraordinary Ph.D. Award.

ARMIN MEHRI received the B.S. degree in computer engineering and the M.S. degree from Eastern Mediterranean University, in 2014 and 2017, respectively. Currently, he is currently the Ph.D. degree with the Computer Vision Center, Universitat Autònoma de Barcelona. His main research interests include computer vision, image processing, image enhancement, and image colorization under cross-modal frameworks resulting in cross-spectral domains.

F. XAVIER ROCA received the Ph.D. degree in computer science from the Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès, Spain, in 1990. He is an Associate Professor and the Director of the Department of Computer Science, UAB. He is also a Research Fellow with the Computer Vision Center. He has been a Principal Researcher in several projects (public and private funds). He is working in technological transfer computer vision. His research interests include active vision, biometrics, and tracking.

SEIICHI OZAWA (Senior Member, IEEE) received the Dr.Eng. degree in computer science from Kobe University, Japan. He is currently the Deputy Director of the Center for Mathematical and Data Sciences and a Full Professor with the Department of Electrical and Electronic Engineering, Graduate School of Engineering, Kobe University. He has published more than 160 journal and conference papers, and book chapters/monographs. His current research interests include deep learning, machine learning, pattern recognition, incremental learning, big data analytics, cybersecurity, text mining, computer vision, and privacy preserving data mining. He is a member of the Neural Networks TC and Smart World TC of IEEE CI Society. He is the Vice-President of the Membership of the International Neural Network Society, the Vice-President of Finance of the Asia–Pacific Neural Network Society, and the Board of Governor of the Japan Neural Network Society. He is currently an Associate Editor of IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, and two international journals.

JORDI GONZÀLEZ (Member, IEEE) received the Ph.D. degree in computer engineering from the Universitat Autònoma de Barcelona (UAB), in 2004. Currently, he is an Associate Professor in computer science with the Computer Science Department, UAB. He is also a Research Fellow with the Computer Vision Center, where he has cofounded three spin-offs (Cloud Size Services, Visual Tagging, and Care Respite) and the Image Sequence Evaluation (ISE Laboratory) Research Group. His research interests include machine learning techniques for the computational interpretation of social images or visual hermeneutics.