

## Research Article

# Multilevel Feature Exploration Network for Image Superresolution

Xinbo Liu <sup>1</sup>, Ling Wang,<sup>2</sup> Xinyu Chen <sup>3</sup>, and Yuqing Liu <sup>3</sup>

<sup>1</sup>Key Laboratory of Security for Network and Data in Industrial Internet of Liaoning Province, Liaoning University of Technology, Jinzhou, China

<sup>2</sup>College of Chemistry and Materials Engineering, Hainan Vocational University of Science and Technology, Haikou, China

<sup>3</sup>School of Software, Dalian University of Technology, Dalian, China

Correspondence should be addressed to Yuqing Liu; liuyuqing@mail.dlut.edu.cn

Received 23 October 2021; Revised 22 November 2021; Accepted 13 January 2022; Published 7 February 2022

Academic Editor: Jianping Gou

Copyright © 2022 Xinbo Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Image superresolution (SR) is a classical issue in computer vision area. Recently, there are elaborated convolutional neural networks (CNNs) demonstrating remarkable effectiveness on image SR. However, most of the previous works lack effective exploration on the structural information, which plays a critical role for image quality. In this paper, we find that the hierarchical design can effectively restore the structural information and devise a multilevel feature exploration network for image SR (MFSR). Specially, we design an encoder-decoder architecture to concentrate on structural information from different levels and devise a spatial attention mechanism to address the inherent correlation among features for effective restoration. Experimental results show the proposed MFSR can restore more correct edges and lines and achieves both better objective and subjective performances than the state-of-the-art methods with higher PSNR/SSIM results, indicating the effectiveness on structural information restoration.

## 1. Introduction

Image superresolution (SR), as a traditional image processing issue, is widely considered in advanced vision applications, such as medical image enhancement, slope collapse detection, and art image classification [1]. Given a low-resolution (LR) image, the task of image SR is to find a corresponding high-resolution (HR) instance with satisfactory visual experience and more correct textures [2].

Recently, convolutional neural networks (CNNs) show the amazing performance on image restoration. As far as we know, superresolution convolutional neural network (SRCNN) [3] proposed by Dong et al. is the first CNN-based method for image SR. After SRCNN, there are elaborate network designs for effective image SR, such as enhanced deep residual superresolution network (EDSR) [4], residual dense network for image superresolution (RDN) [5, 6], deep

residual channel attention image superresolution network (RCAN) [7], deep Laplacian pyramid superresolution network (LapSRN) [8, 9], and cascading residual network (CARN) [10]. These works concentrate on elaborate designs and utilize different blocks for effective performance. However, very few of them specially concentrate on the structural information restoration.

There are also works considering the structural information as a prior. Fang et al. design a soft-edge assisted network for single image SR, which is termed as SeaNet [11]. Cheng et al. consider gradient guidance into structure-preserving superresolution (SPSR) [12] and utilize generative adversarial network for satisfactory performance. Sparse-mask superresolution network (SMSR) is also investigated by Zhang et al. for efficient inference. Yang et al. devise a deep edge guided recurrent residual network for image SR (DEGREE) [13]. However, the structural information is usually applied as a prior term, and they almost

neglect to focus on designing effective architecture to explore the edges and textures.

Multiscale structure proves to be an effective design pattern for structural information exploration. Multiscale residual network for image superresolution (MSRN) [14] designs a multiscale residual block for effective feature exploration. Recently, multiscale dense cross network (MDCN) [15] is also proposed for hierarchical exploration. However, these works lack finding the inherent correlations among features with restricted performance.

Attention mechanism has become a popular design in recent CNNs for different tasks. Recently, channel-wise attention [16, 17] has also been widely considered in image SR networks, such as residual-in-residual channel attention network for image SR (RCAN) [7] and channel-wise and spatial feature modulation network (CSFM) for single image SR [18]. However, the channel-wise attention usually utilizes a global pooling method to estimate the information of different channels, which misses the spatial diversity of feature maps.

In this paper, we design a multilevel feature exploration network for image superresolution, which is termed as MFSR. To effectively explore the structural information, we consider a UNet-style [19] to hierarchically process the image features and concatenate the multiscale features for joint restoration. Besides the UNet-like backbone, we also devise a spatial attention mechanism to consider the inherent correlation among spatial information and remove the global pooling operation in traditional channel-wise attention, which effectively avoids the information loss. Based on the spatial attention mechanism, we devise a spatial attention residual block (SARB) for effective structural information exploration. Experimental results show our MFSR can restore the HR images with more satisfactory textures and closer to the ground truth. Compared with the state-of-the-art methods, specially multiscale structure-based networks, our network achieves much higher peak signal-to-noise rate (PSNR) [20] and structural similarity (SSIM) [21] results with better visual quality.

For this paper, the main contributions are as follows:

- (i) We design a multilevel feature exploration network for image superresolution, which can effectively process the hierarchical structural information.
- (ii) We devise a spatial attention mechanism to explore the inherent correlation among spatial information without global pooling operation, which effectively avoids the information loss. Based on the spatial attention mechanism, we devise a spatial attention residual block (SARB) for effective structural information exploration.
- (iii) Experimental results show our MFSR can restore the HR images with more satisfactory textures and closer to the ground truth with higher PSNR/SSIM results.

In the following, we introduce the related work in Section 2. Then, the detail of MFSR is described in Section 3. Finally, the experiment results are shown in Section 4.

## 2. Related Work

*2.1. Image Superresolution.* Image SR is a traditional topic in computer vision area, which is widely investigated in recent years. Classical SR methods can be generally separated into five types: interpolation-based methods, filter-based methods, patch-based methods [22, 23], optimization-based methods [24, 25], and sparse-coding-based methods [26, 27]. Interpolation aims to build a function that uses the neighbor pixel values to estimate the missing pixel value. Cao et al. bring low-rank matrix completion and recovery into image SR and investigate a novel representation of interpolation function [28]. Similar to interpolation methods, filter-based methods try to find a feasible filter to change the pixel value with the help of neighbor's information and generate pleasant visual results. Patch-based methods separate the image into several adjacent patches and optimize the performance of each patch separately to improve the whole image quality. Zhu et al. use deformable patches for single image SR [22]. Yang et al. convert the patches into sparse representation and perform the image restoration for raw data [23]. Optimization-based methods consider single image SR as minimizing the energy function and utilize iterative optimization to approximate the solution. Huang and Xia use matrix-variable optimization for fast blind image superresolution [24], and Shi et al. optimize the image quality with low-rank and total variation regularization [25]. Different from previous works, sparse-coding-based methods enable the learning step into image SR and derive a codebook for adaptive image restoration. Shi and Qi consider low-rank sparse representation with self-similarity for image SR [26]. Lu et al. also perform geometry constrained sparse coding to produce an acceptable result.

Convolutional neural networks (CNNs) demonstrate amazing performance on image SR task. To the best of our knowledge, superresolution convolutional neural network (SRCNN) [3] is the first CNN-based method for image SR, which brings a greater performance improvement than traditional methods. SRCNN follows a sparse-coding-like structure and uses three convolutional layers to denote the feature extraction, nonlinear mapping, and restoration operations separately. After SRCNN, there are numerous CNN-based works with elaborate designs for superior capacity. A fast version of SRCNN termed as FSRCNN [29] is proposed by Dong et al. to accelerate the inference phase and achieves better performance than SRCNN with a deeper but faster network. A very deep network with residual connection termed as VDSR [30] also increases the network depth for boosting the capacity. Kim et al. design a deep recursive network for image SR, which is termed as DRCN [31]. These works use deconvolution to upscale the feature map and generate the HR image or preupsample the image before the network processing. Shi et al. provide a new perspective to upscale the feature map by subpixel convolution and integrate the upsampling operation into the efficient subpixel convolutional neural network (ESPCN) [32]. After ESPCN, most of the image SR networks consider subpixel convolution to upscale the feature map instead of using deconvolution. Recent end-to-end networks can be

generally separated into single-scale methods and multiscale methods. Single-scale networks consider a deeper and wider design to seek higher quantitative performance. Enhanced deep residual network for image SR (EDSR) is a representative method to achieve the state-of-the-art performance, which stacks numerous residual blocks [33] and removes the batch normalization [34] layers. After EDSR, there are well-designed blocks for image SR with state-of-the-art performance. Dense net for image SR (SRDenseNet) [35] introduces the densely connection into image SR. Zhang et al. combine the superiority of dense connection and residual connection and propose a residual dense connection network (RDN) for image SR [5, 6]. Recently, residual-in-residual network with channel attention for image SR (RCAN) [7], residual feature aggregation deep network (RFDN) [36], and other works also achieve state-of-the-art performance. Deep network with component learning (DNCL) is recently presented for fast and accurate image SR. Inspired by traditional filter-based methods, Li et al. propose an adaptive filter learning network termed as FilterNet [37].

Besides the single-scale networks, there are also multiscale methods that concentrate on the hierarchical structural information exploration. Laplacian pyramid image SR network (LapSRN) [8, 9] designs a Laplacian pyramid inspired network to progressively restore the residual image and increases the image resolution. Multiscale residual network (MSRN) [14] utilizes a multiscale residual block for accurate image restoration. Multiscale dense cross network (MDCN) upgrades the MSRN and achieves state-of-the-art image SR performance. Resolution-aware network (RAN) [38] was recently considered for simultaneous SR of multiple factors. He et al. consider a multiscale design to accelerate the network speed, which is named as MRFN [39]. Recently, a deep inception-residual Laplacian pyramid network (IRLP) is also proposed for accurate single image SR [40].

**2.2. Attention Mechanism.** As far as we know, channel-wise attention [16, 17] is the first tiny but effective attention block for CNN design. The channel-wise attention omits the spatial information and only concentrates on the different importance of channels. Residual-in-residual channel attention network (RCAN) [7], information multiple distillation network (IMDN) [41], and other recent works consider the channel-wise attention for better restoration performance. Besides the channel-wise attention, there are also effective attention mechanisms for image SR. Residual feature information distillation network (RFDN) [42] proposes an enhanced spatial attention to overcome the shortcoming of channel-wise attention and considers the spatial information for better attention calculation.

### 3. Methodology

In this section, we introduce the MFSR in the following manner. Firstly, we describe the overall network design for restoring HR image from LR instance. Then, we specially discuss the blocks design of SARB. Finally, we show the hyperparameter settings for MFSR.

**3.1. Network Design.** We aim to restore the HR image  $\mathbf{I}^H$  from the given LR image  $\mathbf{I}^L$ . Let  $f^{\text{MFSR}}(\cdot)$  be the proposed MFSR; then, the restoration is

$$\mathbf{I}^H = f^{\text{MFSR}}(\mathbf{I}^L), \quad (1)$$

Figure 1 shows the network structure of proposed MFSR. There are three main modules in the MFSR: feature extraction module, multilevel feature exploration module, and restoration module. Before the restoration, one convolutional layer converts the image into feature space for further exploration as

$$\mathbf{F}^{\text{cvt}} = \text{Conv}(\mathbf{I}^L), \quad (2)$$

where  $\mathbf{F}^{\text{cvt}}$  is the converted feature maps from  $\mathbf{I}^L$ .

The feature extraction module extracts the feature maps from  $\mathbf{F}^{\text{cvt}}$ . Let  $f^{\text{FE}}(\cdot)$  denote the feature extraction module; then, there is

$$\mathbf{F}^L = f^{\text{FE}}(\mathbf{F}^{\text{cvt}}), \quad (3)$$

where  $\mathbf{F}^L$  represents the extracted features from LR image.

After feature extraction, the multilevel feature exploration module processes  $\mathbf{F}^L$  in a hierarchical manner for effectively restoring the structural information. Firstly, we build the multilevel features form  $\mathbf{F}^L$  for further processing. We use max pooling to downscale  $\mathbf{F}^L$  and generate feature maps with different resolutions as

$$\begin{cases} \mathbf{F}_{1 \times 1}^L = (\mathbf{F}^L) \downarrow_{1 \times 1}, \\ \mathbf{F}_{2 \times 2}^L = (\mathbf{F}^L) \downarrow_{2 \times 2}, \\ \mathbf{F}_{4 \times 4}^L = (\mathbf{F}^L) \downarrow_{4 \times 4}, \end{cases} \quad (4)$$

where  $(\cdot) \downarrow_{s \times s}$  denotes the max pooling with scaling factor  $s \times s$ . The downsampled features with different resolutions are explored by several spatial attention residual blocks (SARB) for generating the feature maps with HR information. There are  $m_s$  SARBs for processing the feature maps with scaling factor  $s \times s$ . After SARBs, a padding structure with residual connection is devised to keep the original information and improve the gradient transmission. The padding structure consists of two convolutional layers and one ReLU activation. Thus, the multilevel exploration is

$$\begin{cases} \mathbf{F}_{1 \times 1}^H = f_{1 \times 1}^{\text{Pad}}(f_{m_1}^{\text{SARB}}(\mathbf{F}_{1 \times 1}^L)) + \mathbf{F}_{1 \times 1}^L, \\ \mathbf{F}_{2 \times 2}^H = f_{2 \times 2}^{\text{Pad}}(f_{m_2}^{\text{SARB}}(\mathbf{F}_{2 \times 2}^L)) + \mathbf{F}_{2 \times 2}^L, \\ \mathbf{F}_{4 \times 4}^H = f_{4 \times 4}^{\text{Pad}}(f_{m_4}^{\text{SARB}}(\mathbf{F}_{4 \times 4}^L)) + \mathbf{F}_{4 \times 4}^L, \end{cases} \quad (5)$$

where  $f_{s \times s}^{\text{Pad}}(\cdot)$  is the padding structure for scaling factor  $s \times s$  and  $f_{m_s}^{\text{SARB}}(\cdot)$  means the processing procedure with  $m_s$  SARBs.  $\mathbf{F}_{s \times s}^H$  is the HR feature with scaling factor  $s \times s$ .

After exploration, the multilevel features should be concatenated jointly for hierarchical consideration. In this paper, we use bilinear interpolation to upscale the resolution. Bilinear interpolation considers the image interpolation from two directions and finds a suitable value for missing pixels [43]. The concatenation is demonstrated as

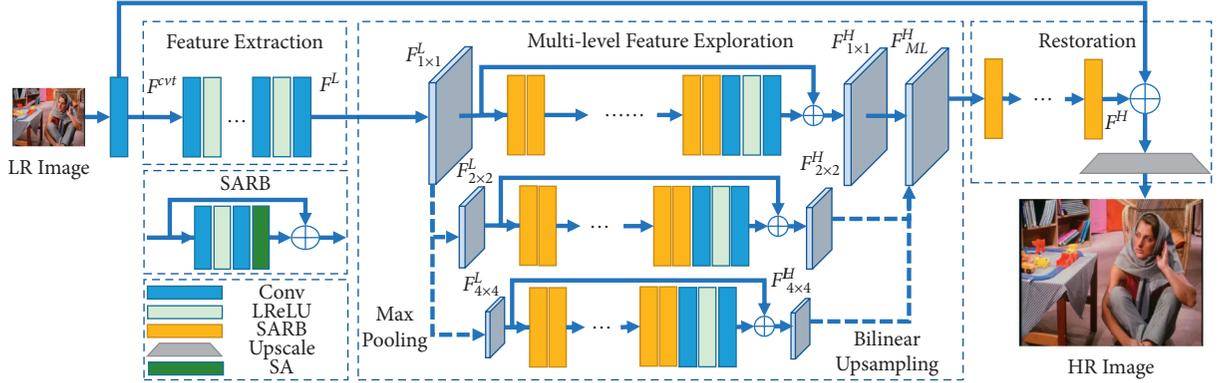


FIGURE 1: The network structure of the proposed multilevel feature exploration image superresolution network (MFSR).

$$\mathbf{F}_{ML}^H = \text{Concat}([\mathbf{F}_{1 \times 1}^H, (\mathbf{F}_{2 \times 2}^H) \uparrow_{2 \times 2}, (\mathbf{F}_{4 \times 4}^H) \uparrow_{4 \times 4}]), \quad (6)$$

where  $\text{Concat}(\cdot)$  is the concatenation operation,  $(\cdot) \uparrow_{s \times s}$  is the bilinear upsampling operation, and the concatenated multilevel HR feature is  $\mathbf{F}_{ML}^H$ .

After multilevel exploration, the restoration module generates the HR image from the feature maps. There are two steps in the restoration module. Firstly, we explore the multilevel HR feature for further integrating the multilevel information. There are  $n$  SARBs for the information integration as

$$\mathbf{F}^H = f_n^{\text{SARB}}(\mathbf{F}_{ML}^H), \quad (7)$$

where  $\mathbf{F}^H$  is the explored hierarchical HR feature. Then, the upscale block generates the image from the feature maps, which is composed of one convolutional layer and one subpixel convolution. Skip connection is designed between the LR and HR features for better gradient transmission. Finally, there is

$$\mathbf{I}^H = \text{Upscale}(\mathbf{F}^H + \mathbf{F}^{\text{cvt}}), \quad (8)$$

where  $\text{Upscale}(\cdot)$  is the upscale block.

The overall proposed MFRN can be concluded as in Algorithm 1.

(1) MFSR ( $\mathbf{I}^L$ )

Input: The input LR image  $\mathbf{I}^L$

Output: The restored HR image  $\mathbf{I}^H$

- (2) Convert LR image into feature space by convolution as  $\mathbf{F}^{\text{cvt}} = \text{Conv}(\mathbf{I}^L)$ .
- (3) Extract LR features by the feature extraction module as  $\mathbf{F}^L = f^{\text{FE}}(\mathbf{F}^{\text{cvt}})$ .
- (4) Generate the hierarchical HR feature according equation (4)–(7).
- (5) Produce the HR image from the feature as  $\mathbf{I}^H = \text{Upscale}(\mathbf{F}^H + \mathbf{F}^{\text{cvt}})$
- (6) Return  $\mathbf{I}^H$ .

Algorithm 1: The proposed MFSR for generating HR image from the LR instance

**3.2. Spatial Attention Residual Block.** Spatial attention residual block (SARB) is the main component for the MFSR, which is generally utilized in the multilevel feature

exploration module and the restoration module. The left-middle box in Figure 1 shows the block design of SARB. SARB enjoys a residual design with two convolutional layers, one Leaky ReLU activation, and one spatial attention (SA) layer. The skip connection in SARB helps effective gradient transmission and information preservation. Specially, we mainly introduce the spatial attention in the SARB that effectively explore the spatial correlation among features.

Figure 2 shows the architecture of SA layer, which is composed of four convolutional layers, three Leaky ReLU activation, and a Sigmoid activation. Let  $\mathbf{F}_{in}^{sa}$  be the input feature of SA, there is one convolution with ReLU activation to process the feature as

$$\mathbf{F}_1^{sa} = \text{LReLU}(\text{Conv}(\mathbf{F}_{in}^{sa})). \quad (9)$$

After convolution, the feature maps are downsampled by max pooling with scaling factor  $3 \times 3$  to increase the receptive field. Then, one convolutional layer and one ReLU activation explore the correlation among features as

$$\mathbf{F}_2^{sa} = \text{LReLU}(\text{Conv}((\mathbf{F}_1^{sa}) \downarrow_{3 \times 3})). \quad (10)$$

The processed feature will be upsampled by bilinear upsampling to the original size for calculating the attention. After upscaling, two convolutional layers with a ReLU activation explore the feature to find the attention map. Sigmoid activation is used to maintain the nonnegativity. Finally, the output feature of SA layer is

$$\mathbf{F}_{out}^{sa} = \sigma(\text{Conv}(\text{LReLU}(\text{Conv}((\mathbf{F}_2^{sa}) \uparrow_{3 \times 3})))) * \mathbf{F}_{in}^{sa}, \quad (11)$$

where  $\sigma(\cdot)$  denotes the Sigmoid activation and  $\uparrow_{3 \times 3}$  denotes the bilinear interpolation with scaling factor  $3 \times 3$ .

It is worth noting that SA considers the spatial correlation of features. Different from channel-wise attention which utilizes global pooling to embed the channel information [7, 16, 18], SA considers the spatial correlation by exploring the attention with the same size of the input feature maps. By utilizing SA, every pixel of feature map will get a specific attention value.

**3.3. Settings.** In this paper, all layers are set with filter number as 64 except for the upscale block. The filter number of the upscale block follows the same design as IMDN [41].

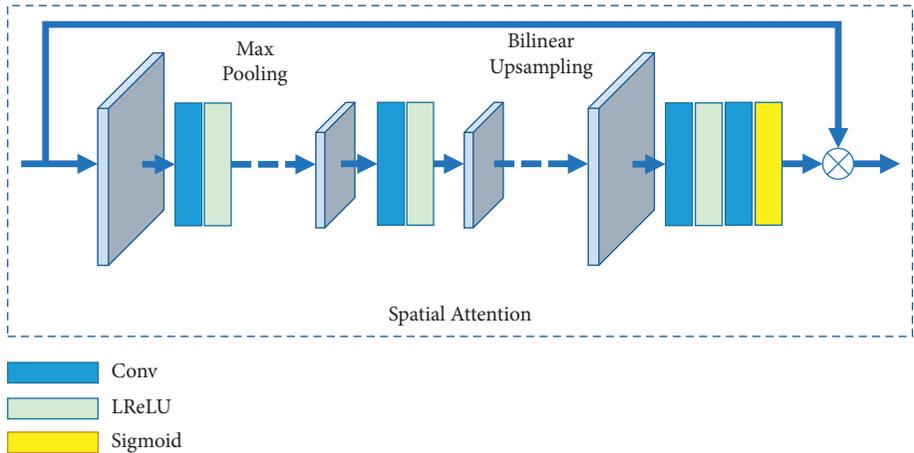


FIGURE 2: The structure of spatial attention (SA) mechanism.

There are  $m_1 = 16$  SARBs in the  $1 \times 1$  scale of multilevel feature exploration,  $m_2 = 8$  SARBs in the  $2 \times 2$  scale, and  $m_4 = 4$  SARBs in the  $4 \times 4$  scale. It is worth noting that the block numbers are set to maintain the same receptive field for different scaling factors. There are four convolutions in feature extraction module, and five SARB blocks in restoration module. The filter sizes of all convolutional layers are set as  $3 \times 3$ .

### 4. Experiment

In this paper, we train our network with DIV2K [44] dataset. DIV2K is firstly proposed in NTIRE 2017 challenge [44] and contains numerous high-resolution instances for image SR. There are numerous SR works trained on this dataset [4–7]. We choose 800 images from DIV2K for training, and 5 images for validation. For testing, we use five widely used benchmarks to evaluate the performance: Set5 [45], Set14 [46], B100 [47], Urban100 [48], and Manga109 [49]. Peak signal-to-noise ratio (PSNR) [20] and structural similarity [21] are selected to estimate the objective performance of restored images. We update our model by Adam [50] optimizer with learning rate as  $lr = 10^{-4}$  and halve the learning rate for every 200 epochs. The loss function of this paper is set as L1-loss. We use PyTorch 0.4.1 to train the network. The CUDA version is 10.0. The batch size is set as 32. The experimental environment is Ubuntu 18.04, and we train the model on one NVIDIA GTX-1080Ti GPU.

#### 4.1. Ablation Study

4.1.1. *Investigation on the Multilevel Exploration.* In this paper, we set different number of SARBs in the three scales of multilevel exploration module to keep the receptive field. To investigate the effectiveness of the number settings, we consider different combinations for comparison. Table 1 shows the performances of different block settings. Herein,  $(m_1, m_2, m_4)$  denotes the different number of SARBs for the scaling factor as  $1 \times 1$ ,  $2 \times 2$ , and  $4 \times 4$ . In the table, we find that, with the increase of SARBs, the PSNR/SSIM results are higher. According to the third and fourth rows, the

TABLE 1: Investigation on PSNR/SSIM performances of different multilevel settings. Herein,  $(m_1, m_2, m_4)$  denotes the different number of SARB on the three pathways.

Setting	Set5	Set14	B100
(16, 0, 0)	32.24/0.8954	28.64/0.7834	27.62/0.7380
(16, 8, 0)	32.26/0.8958	28.63/0.7834	27.62/0.7378
(16, 4, 2)	32.23/0.8952	28.63/0.7835	27.62/0.7379
(8, 4, 2)	32.17/0.8948	28.60/0.7822	27.59/0.7368
<b>(16, 8, 4)</b>	<b>32.26/0.8961</b>	<b>28.65/0.7838</b>	<b>27.63/0.7381</b>

PSNR of  $m_1 = 16$  on Set5 is 32.23 dB while the PSNR of  $m_1 = 8$  is 32.17 dB. The decrease of  $m_1$  leads to near 0.06 dB PSNR drop on Set5. However, the PSNR results of the second, third, and last rows on Set5 are 32.26 dB, 32.23 dB, and 32.26 dB separately, which shows the decrease of  $m_2$  and  $m_4$  leads to less than 0.03 dB drop on PSNR. In this point of view,  $m_1$  influences the performance more than blocks from other levels. We believe this is because the resolution of  $m_1$  is larger than other levels, and the computation complexity is much higher.

4.1.2. *Investigation on Spatial Attention.* We also investigate the effectiveness of spatial attention. Table 2 shows the performance comparisons on spatial attention. The model with spatial attention enjoys the SARB design proposed in the paper. The model without spatial attention holds the same block design as EDSR [4]. In the table, we can find that the spatial attention mechanism brings 0.04 dB, 0.02 dB, and 0.04 dB PSNR gain on Set5, Set14, and B100 datasets separately. Similarly, the model with spatial attention achieves 0.8961, 0.7838, and 0.7381 SSIM results on different benchmarks, which are all higher than the model without attention mechanism. In this point of view, attention mechanism is an effective component for better performance.

4.2. *Comparison with State-of-the-Art Methods.* We mainly compare our model with several classical and recent works, including SRCNN [3], FSRCNN [29], VDSR [30], DRCN [31], LapSRN [8, 9], RAN [38], DNCL [51], FilterNet [37],

TABLE 2: Investigation on PSNR/SSIM performances of spatial attention (SA) mechanism.

Col Count:4F0E0Attention	Set5	Set14	B100
w/o	32.22/0.8954	28.63/0.7831	27.59/0.7373
w	<b>32.26/0.8961</b>	<b>28.65/0.7838</b>	<b>27.63/0.7381</b>

TABLE 3: Average PSNR/SSIM with degradation model BI  $\times 2$ ,  $\times 3$ , and  $\times 4$  on five benchmarks. The best performances are shown in **bold**. The dash line means the paper does not report their performance.

Scale	Model	Set5 [45] PSNR/SSIM	Set14 [46] PSNR/SSIM	B100 [47] PSNR/SSIM	Urban100 [48] PSNR/SSIM	Manga109 [49] PSNR/SSIM
$\times 2$	SRCNN [3]	36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946	35.74/0.9661
	FSRCNN [29]	37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020	36.67/0.9694
	VDSR [30]	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9729
	DRCN [31]	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.63/0.9723
	LapSRN [9]	37.52/0.9590	33.08/0.9130	31.80/0.8950	30.41/0.9100	37.27/0.9740
	RAN [38]	37.58/0.9592	33.10/0.9133	31.92/0.8963	—	—
	DNCL [51]	37.65/0.9599	33.18/0.9141	31.97/0.8971	30.89/0.9158	—
	FilterNet [37]	37.86/0.9610	33.34/0.9150	32.09/0.8990	31.24/0.9200	—
	MRFN [39]	37.98/0.9611	33.41/0.9159	32.14/0.8997	31.45/0.9221	38.29/0.9759
	SeaNet-baseline [11]	37.99/0.9607	33.60/0.9174	32.18/0.8995	32.08/0.9276	38.48/0.9768
	DEGREE [13]	37.58/0.9587	33.06/0.9123	31.80/0.8974	—	—
	IRLP [40]	37.83/0.9692	33.18/0.9247	32.01/0.9021	30.86/0.9192	37.71/0.9861
	FSN [52]	37.68/0.9065	33.51/0.9180	32.09/0.9015	31.68/0.9248	—
	DSRLN [53]	38.05/—	33.56/—	32.38/—	31.79/—	—
MFSR	<b>38.07/0.9608</b>	<b>33.69/0.9191</b>	<b>32.22/0.9002</b>	<b>32.35/0.9307</b>	<b>38.75/0.9768</b>	
$\times 3$	SRCNN [3]	32.75/0.9090	29.28/0.8209	28.41/0.7863	26.24/0.7989	30.59/0.9107
	FSRCNN [29]	33.16/0.9140	29.43/0.8242	28.53/0.7910	26.43/0.8080	30.98/0.9212
	VDSR [30]	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9310
	DRCN [31]	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.31/0.9328
	DRRN [31]	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.74/0.9390
	RAN [38]	33.71/0.9223	29.84/0.8326	28.84/0.7981	—	—
	DNCL [51]	33.95/0.9232	29.93/0.8340	28.91/0.7995	27.27/0.8326	—
	FilterNet [37]	34.08/0.9250	30.03/0.8370	28.95/0.8030	27.55/0.8380	—
	MRFN [39]	34.21/0.9267	30.03/0.8363	28.99/0.8029	27.53/0.8389	32.82/0.9396
	SeaNet-baseline [11]	34.36/0.9280	30.34/0.8428	29.09/0.8053	28.17/0.8527	33.40/0.9444
	DEGREE [13]	33.76/0.9211	29.82/0.8326	28.74/0.7950	—	—
	IRLP [40]	34.22/0.9437	30.16/0.8472	29.21/0.8163	27.58/0.8327	32.61/0.9492
	DSRLN [53]	34.56/—	30.36/—	29.29/—	27.88/—	—
	MFSR	<b>34.49/0.9280</b>	<b>30.42/0.8442</b>	<b>29.16/0.8068</b>	<b>28.39/0.8577</b>	<b>33.72/0.9457</b>
$\times 4$	SRCNN [3]	30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221	27.66/0.8505
	FSRCNN [29]	30.71/0.8657	27.59/0.7535	26.98/0.7150	24.62/0.7280	27.90/0.8517
	VDSR [30]	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8809
	DRCN [31]	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.98/0.8816
	LapSRN [9]	31.54/0.8850	28.19/0.7720	27.32/0.7280	25.21/0.7560	29.09/0.8845
	RAN [38]	31.43/0.8847	28.09/0.7691	27.31/0.7260	—	—
	DNCL [51]	31.66/0.8871	28.23/0.7717	27.39/0.7282	25.36/0.7606	—
	FilterNet [37]	31.74/0.8900	28.27/0.7730	27.39/0.7290	25.53/0.7680	—
	MRFN [39]	31.90/0.8916	28.31/0.7746	27.43/0.7309	25.46/0.7654	29.57/0.8962
	SeaNet-baseline [11]	32.18/0.8948	28.61/0.7822	27.57/0.7359	26.05/0.7896	30.44/0.9088
	DEGREE [13]	31.47/0.8837	28.10/0.7669	27.20/0.7216	—	—
	IRLP [40]	31.94/0.9021	28.49/0.7910	27.70/0.7479	25.61/0.7764	29.49/0.9025
	FSN [52]	32.10/0.8959	28.57/0.7874	27.53/0.7438	25.76/0.7817	—
	MFSR	<b>32.26/0.8961</b>	<b>28.65/0.7838</b>	<b>27.63/0.7381</b>	<b>26.25/0.7919</b>	<b>30.62/0.9103</b>

MRFN [39], SeaNet [11], DEGREE [13], IRLP [40], FSN [52], and DSRLN [53]. Table 3 shows the PSNR/SSIM performance on five benchmarks with scaling factor  $\times 2$ ,  $\times 3$ , and  $\times 4$ .

In the table, we can find the proposed MFSR achieves the best performance on all testing benchmarks. It is worth noting that our model is mainly designed for structural

information restoration. In this point of view, we achieve significant PSNR/SSIM improvement on the benchmark with plentiful lines and edges, such as Urban100 and Manga109. Particularly, our method achieves near 0.3 dB higher on Manga109 and Urban100 than other works with scaling factor  $\times 2$  and  $\times 3$ . When scaling factor is  $\times 4$ , we achieve near 0.2 dB higher than other works. Urban100 is

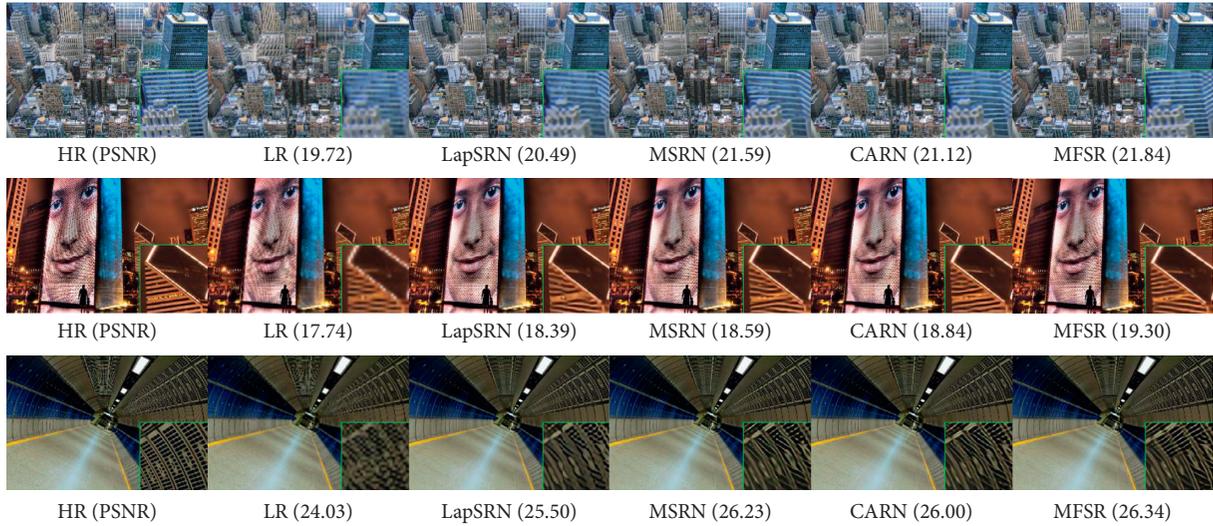


FIGURE 3: Visualization comparison on Urban100 dataset.

composed of high-resolution building images, and Manga109 is composed of comic cover pages. Thus, the high PSNR/SSIM performance demonstrates our effectiveness on structural information restoration.

We also achieve superior performance than other edge preserving and multiscale SR works. FilterNet, DEGREE, and SeaNet are aiming at structure information preserving, and MRFN is a modern effective multiscale network for image SR. In Table 3, our model gains more than 0.2 dB on PSNR, which is a significant improvement.

Besides quantitative comparisons, we also compare the visualization performance with different methods. We mainly compare our method with LapSRN [9], MSRN [14], and CARN [10]. LapSRN is a representative progressive structure inspired by Laplacian pyramid, and MSRN is a state-of-the-art multiscale design. Thus, we compare with them to show the effectiveness of our multilevel design. Figure 3 shows the visualization comparison. In the first and second rows of the figure, our method recovers the lines on the building more accurately than MSRN and LapSRN. In the last row, our method can restore more correct small lines, while both LapSRN and MSRN fail to recover the straight lines.

## 5. Conclusion

In this paper, we proposed a multilevel feature exploration network (MFSR) to restore the structural information for better visualization experience, which is designed with a UNet-like architecture for hierarchical exploration. To furthermore explore the structural information, spatial attention is considered in the MFSR for better performance. Spatial attention residual blocks (SARBs) are specially designed based on the proposed spatial attention mechanism. Experiment results show MFSR achieves better PSNR/SSIM performance than other multiscale and structural-preserving image SR works with scaling factor  $\times 2$ ,  $\times 3$ , and  $\times 4$ . When the scaling factor is larger, there is near 0.2 dB PSNR improvement on the benchmarks with plentiful edges

and lines, indicating the effectiveness on structural information restoration. The visual comparison result also demonstrates the superiority of MFSR on producing more accurate lines and edges.

## Data Availability

The image data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was partially supported by the Program for Liaoning Innovation Talents in University (no. LR2019034) and the Overseas Training Foundation of Liaoning (no. 2019GJWYB015).

## References

- [1] Y. Wu, "Application of improved boosting algorithm for art image classification," *Scientific Programming*, vol. 2021, Article ID 3480414, 11 pages, 2021.
- [2] K. Cao, Y. Liu, L. Duan, and X. Tian, "Adaptive residual channel attention network for single image super-resolution," *Scientific Programming*, vol. 2020, Article ID 8877851, 10 pages, 2020.
- [3] C. Dong, C. Change Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.
- [4] B. Lim, S. Son, H. Kim, S. Nah, and K. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1132–1140, Honolulu, HI, USA, July 2017.
- [5] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of*

- the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2472–2481, Salt Lake City, UT, USA, June 2018.
- [6] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image restoration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2480–2495, 2021.
  - [7] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 294–310, Springer, Munich, Germany, September 2018.
  - [8] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Fast and accurate image super-resolution with deep laplacian pyramid networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2599–2613, 2019.
  - [9] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Deep laplacian pyramid networks for fast and accurate super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5835–5843, Honolulu, HI, USA, July 2017.
  - [10] N. Ahn, B. Kang, and K.-A. Sohn, “Fast, accurate, and lightweight super-resolution with cascading residual network,” in *Proceedings of the European Conference on Computer Vision*, vol. 11214, pp. 256–272, Munich, Germany, September 2018.
  - [11] F. Fang, J. Li, and T. Zeng, “Soft-edge assisted network for single image super-resolution,” *IEEE Transactions on Image Processing*, vol. 29, pp. 4656–4668, 2020.
  - [12] M. Cheng, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, “Structure-preserving super resolution with gradient guidance,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7766–7775, Seattle, WA, USA, June 2020.
  - [13] W. Yang, J. Feng, J. Yang et al., “Deep edge guided recurrent residual learning for image super-resolution,” *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 5895–5907, 2017.
  - [14] J. Li, F. Fang, K. Mei, and G. Zhang, “Multi-scale residual network for image super-resolution,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 527–542, Springer, Munich, Germany, September 2018.
  - [15] J. Li, F. Fang, J. Li, K. Mei, and G. Zhang, “Mdcn: multi-scale dense cross network for image super-resolution,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 7, pp. 2547–2561, 2021.
  - [16] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-excitation networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.
  - [17] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132–7141, Salt Lake City, UT, USA, June 2018.
  - [18] Y. Hu, J. Li, Y. Huang, and X. Gao, “Channel-wise and spatial feature modulation network for single image super-resolution,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 3911–3927, 2020.
  - [19] W. Wang, K. Yu, J. Hugonot, F. Pascal, and M. Salzmann, “Recurrent u-net for resource-constrained segmentation,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2142–2151, Seoul, Korea, November 2019.
  - [20] Video Quality Experts Group (Vqeg), “Final report from the video quality experts group on the validation of objective models of video quality assessment,” 2000, <http://www.vqeg.org/>.
  - [21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
  - [22] Yu Zhu, Y. Zhang, and L. Alan, “Single image super-resolution using deformable patches,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2917–2924, Columbus, OH, USA, June 2014.
  - [23] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution as sparse representation of raw image patches,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, Anchorage, AK, USA, June 2008.
  - [24] L. Huang and Y. Xia, “Fast blind image super resolution using matrix-variable optimization,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 3, pp. 945–955, 2021.
  - [25] F. Shi, J. Cheng, L. Wang, P.-T. Yap, and D. Shen, “Lrtv: mr image super-resolution with low-rank and total variation regularizations,” *IEEE Transactions on Medical Imaging*, vol. 34, no. 12, pp. 2459–2466, 2015.
  - [26] J. Shi and C. Qi, “Low-rank sparse representation for single image super-resolution via self-similarity learning,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 1424–1428, Phoenix, AZ, USA, September 2016.
  - [27] X. Lu, H. Yuan, P. Yan, Y. Yuan, and X. Li, “Geometry constrained sparse coding for single image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1648–1655, Providence, RI, USA, June 2012.
  - [28] F. Cao, M. Cai, and Y. Tan, “Image interpolation via low-rank matrix completion and recovery,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 8, pp. 1261–1270, 2015.
  - [29] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 391–407, Springer, Amsterdam, The Netherlands, October 2016.
  - [30] J. Kim, K. Lee, and K. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1646–1654, Las Vegas, NV, USA, June 2016.
  - [31] J. Kim, K. Lee, and K. Lee, “Deeply-recursive convolutional network for image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1637–1645, Las Vegas, NV, USA, June 2016.
  - [32] W. Shi, J. Caballero, F. Huszár et al., “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1874–1883, Las Vegas, NV, USA, June 2016.
  - [33] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.
  - [34] S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” in *Proceedings of the International Conference on Machine Learning (ICML)*, vol. 37, pp. 448–456, Lille, France, July 2015.

- [35] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 4809–4817, Venice, Italy, October 2017.
- [36] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2356–2365, Seattle, WA, USA, June 2020.
- [37] F. Li, H. Bai, and Y. Zhao, "Filternet: adaptive information filtering network for accurate and fast image super-resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 6, pp. 1511–1523, 2020.
- [38] Y. Wang, L. Wang, H. Wang, and P. Li, "Resolution-aware network for image super-resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 5, pp. 1259–1269, 2019.
- [39] Z. He, Y. Cao, L. Du et al., "MRFN: multi-receptive-field network for fast and accurate single image super-resolution," *IEEE Transactions on Multimedia*, vol. 22, no. 4, pp. 1042–1054, 2020.
- [40] Y. Tang, W. Gong, X. Chen, and W. Li, "Deep inception-residual laplacian pyramid networks for accurate single-image super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1514–1528, 2020.
- [41] H. Zheng, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proceedings of the ACM International Conference on Multimedia*, pp. 2024–2032, Nice, France, October 2019.
- [42] J. Liu, J. Tang, and G. Wu, "Residual feature distillation network for lightweight image super-resolution," 2020, <https://arxiv.org/abs/2009.11551>.
- [43] S. Yang, "Improved bilinear interpolation method for image fast processing," in *Proceedings of the 2014 7th International Conference on Intelligent Computation Technology and Automation*, pp. 308–311, Changsha, China, October 2014.
- [44] E. Agustsson and T. Radu, "Ntire 2017 challenge on single image super-resolution: dataset and study," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1122–1131, Honolulu, HI, USA, July 2017.
- [45] M. Bevilacqua, A. Roumy, C. Guillemot, and M. Alberi Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proceedings of the British Machine Vision Conference*, vol. 10, Surrey, UK, September 2012.
- [46] Z. Roman, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces*, pp. 711–730, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [47] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *IEEE International Conference on Computer Vision*, vol. 2, pp. 416–423, 2001.
- [48] J. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5197–5206, Boston, MA, USA, June 2015.
- [49] Y. Matsui, K. Ito, Y. Aramaki et al., "Sketch-based manga retrieval using manga109 dataset," *Multimedia Tools and Applications*, vol. 76, no. 20, pp. 21811–21838, 2017.
- [50] P. Diederik and B. Jimmy, "Adam: a method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations*, San Diego, CA, USA, May 2015.
- [51] C. Xie, W. Zeng, and X. Lu, "Fast single-image super-resolution via deep network with component learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 12, pp. 3473–3486, 2019.
- [52] 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.
- [53] T. Geng, X.-Y. Liu, X. Wang, and G. Sun, "Deep shearlet residual learning network for single image super-resolution," *IEEE Transactions on Image Processing*, vol. 30, pp. 4129–4142, 2021.