Multi-scale Grid Network for Image Deblurring with High-frequency Guidance

Yang Liu, Faming Fang, Tingting Wang, Juncheng Li, Yun Sheng, Guixu Zhang

***Abstract*—It has been demonstrated that the blurring process reduces the high-frequency information of the original sharp image, so the main challenge for image deblurring is to reconstruct high-frequency information from the blurry image. In this paper, we propose a novel image deblurring framework to focus on the reconstruction of high-frequency information, which consists of two main subnetworks: a high-frequency reconstruction subnetwork (HFRSN) and a multi-scale grid subnetwork (MSGSN). The HFRSN is built to reconstruct latent high-frequency information from multiple scale blurry images. The MSGSN performs deblurring processes with high-frequency guidance at different scales simultaneously. Besides, in order to better use high-frequency information to restore sharpening images, we designed a high-frequency information aggregation (HFAG) module and a high-frequency information attention (HFAT) module in MSGSN. The HFAG module is designed to fuse high-frequency features and image features at the feature extraction stage, and the HFAT module is built to enhance the feature reconstruction stage. Extensive experiments on different datasets show the effectiveness and efficiency of our method.**

***Index Terms*—Blind image deblurring, image processing, high-frequency guidance, convolutional neural networks, multi-scale.**

I. Introduction

MOTION blur caused by camera shake and object motion is one of the most common problems faced by photographers. Blurry images containing unreal artifacts will place an obstacle in application to autopilot systems, intelligent surveillance systems and so on. Thus image deblurring which aims to restore a sharp latent image from the blurry one has been a research hot spot in Computer Vision and Artificial Intelligence.

Over the past few decades, a large number of single image deblurring methods have been proposed, which can be roughly divided into two categories, i.e., optimization-based methods and recent learning-based methods. Optimization-based methods [1], [2], [3], [4], [5] follow the image degradation equation. The unknown blur kernel and latent image make it an ill-posed problem, so they need to estimate the blur kernel of blurry image based on different assumptions or priors. Xu et al. [2] develop an unnatural L_0 sparse expression, and Pan et al. [6] propose the L_0-regularized prior on both intensity and gradient for deblurring text images. Pan et al. [1] utilize the dark channel prior to increases the effect of deblurring, but when the images are bright pixels dominant, the dark channel prior losses its utilities. Yan et al. [4] introduced an extreme channel prior to solve this issue by combining the dark channel prior and the bright channel prior. Chen et al. [5] proposed a local maximum gradient prior to restore more high-frequency information. Nevertheless, these methods suffer from several drawbacks: (1) they need complex computational inference due to the iterative calculation process; (2) they hardly handle non-uniform dynamic blurs, that is, they can only deal with blur caused by camera shake, not by object motion.

Recently, deep learning technology drives the development of image restoration tasks [7], [8], [9], [10], [11]. There are lots of learning-based deblurring methods that have been proposed. Zhang et al. [12] trained a set of CNN denoisers and integrated them into the model-based optimization method as a prior. Liu et al. [13] designed two CNN modules, named Generator and Corrector, to extract the intrinsic image structures from the data-driven and knowledge-based perspectives, respectively. However, it is difficult for these methods to remove non-uniform dynamic blurs, and they are computationally inefficient due to the complex optimization process. Xu et al. [14] used a deconvolutional CNN to remove blur with the given blur kernel. Sun et al. [15] proposed a classification CNN to predict blur direction and strength in 30×30 image patches. Those methods still rely on the blur kernel to recover the sharp image because they are limited by assuming that the sources of blurs are only camera shake. To tackle such problems, recent end-to-end network proposed to learn the mapping between blurry images and clear images. Nah et al. [16] proposed a multi-scale method by designing a convolutional neural network for deblurring in a 'coarse-to-fine' manner. Kupyn et al. [17] presented a conditional generative adversarial network (GAN) to produce photorealistic deblurred images, with the assistance of the discriminator. Kupyn et al. [18] utilized the FPN architecture [19] and dual discriminator to improve the GAN network mentioned above. However, these GAN-based methods usually suffer from over smoothing. Zhang et al. [20] iteratively removed blur with a stackable

This work was supported by the Key Project of the National Natural Science Foundation of China under Grant 61731009, the NSFC-RGC under Grant 61961160734, the National Natural Science Foundation of China under Grant 61871185, the Shanghai Rising-Star Program under Grant 21QA1402500, the Science Foundation of Shanghai under Grant 20ZR1416200, and the Open Research Fund of KLATASDS-MOE,ECNU.(Corresponding author: Faming Fang.)

Y. Liu, T. Wang, J. Li, and G. Zhang are with the School of Computer Science & Technology, East China Normal University, Shanghai, China. E-mails: andy_corleone@outlook.com, ttwang@stu.ecnu.edu.cn, 51164500049@stu.ecnu.edu.cn, gxzhang@cs.ecnu.edu.cn.

Faming Fang is with the School of Computer Science & Technology, East China Normal University, Shanghai, China, and also with the Key Laboratory of Advanced Theory and Application in Statistics and Data Science - MOE, East China Normal University, Shanghai, China. Email:fmfang@cs.ecnu.edu.cn.

Yun Sheng is with Liverpool John Moores University. Email:y.sheng@ljmu.ac.uk.

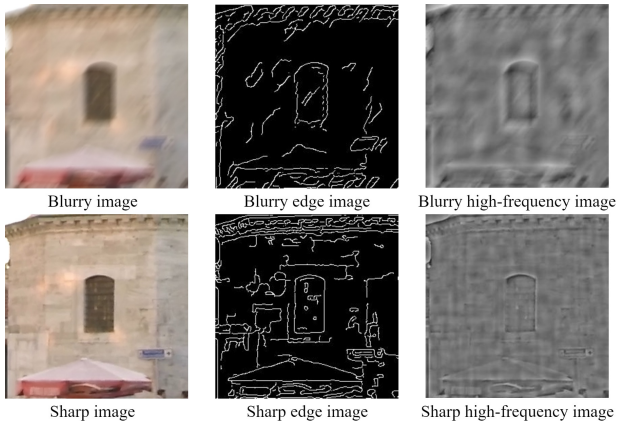


Fig. 1: Comparison of edge image and high-frequency image of the blurry image and sharp image. It is clear to find that the blurry edge was ruined by the blurring process, but the high-frequency information is still retained.

hierarchical method. Suin et al. [21] improved this method with a cross attention module. Aljadaany et al. [22] proposed a deconvolution network by learning both the image prior and data fidelity terms.

The key problem of image deblurring is to restore the high-frequency features in the original image because the high-frequency information of the original image is hidden after the blurring process. Therefore, some methods were proposed to introduce some priors to recover the high-frequency information from degraded images. For example, Fang et al. [23] introduce an edge prior to image super resolution problem which is able to extract shape edges from degraded image. Ma et al. [24] apply a gradient prior for the image super resolution task, they use a subnet to reconstruct HR gradient map from LR image which can preserve structure information. But these kinds of prior are not suitable for image deblurring because the blurring process will break edges or structures from the original sharp image. As shown in Fig. 1, the blurry image has different structures and edges from the sharp image, but some of the original high-frequency information is still retained in blurry image. This is because a blurred image can be regarded as a sharp image convolved by the blur kernel, and the high-frequency information in it is smoothed but not erased. Based on this finding, we propose a high-frequency guided framework for image deblurrring. First, we apply the high-frequency reconstruction subnetwork (HFRSN) to reconstruct latent high-frequency information from multiple scale blurry images. Second, we use high-frequency information as a prior to guide the deblurring process in different scale based on the multi-scale grid subnetwork (MSGSN). In each scale of MSGSN, we propose a high-frequency aggregation module (HFAG) to fuse image features and high-frequency features in the feature extraction stage, and we propose a high-frequency attention module (HFAT) to enhance the reconstruction stage. It is worth mentioning that the two sub-networks are jointly trained so that the final loss of deblurring process in MSGSN will influence the high-frequency reconstruction in HFRSN, which will help HFRSN to reconstruct the really useful high-frequency information. In this way, our method is capable of handling image deblurring task. Extensive experiments show that our method outperforms other state-of-the-art deblurring methods on two benchmarks.

Our contributions are summarized in four aspects.

- We introduce a high-frequency prior to image deblurring by proposing a high-frequency guided framework.
- We propose a high-frequency reconstruction subnetwork called HFRSN, which can reconstruct latent high-frequency information from multiple scale blurry images.
- We propose a multi-scale grid subnetwork called MSGSN to perform the deblurring process in different scale with high-frequency information guidance.
- We propose two different modules named high-frequency aggregation module (HFAG) and high-frequency attention module (HFAT), and they guide the feature extraction stage and the feature reconstruction stage in each scale of MSGSN, respectively.

II. Related work

A. Image Prior in Learning-based Methods

The image prior has demonstrated its superpower on optimization-based methods which guided the equation to be solved to a solution domain closer to the real domain. In recent years, some deep learning methods based on image prior have gradually shown their capabilities. Cheng et al. [25] proposed a fusionnet with an edge prior for the semantic segmentation of remote sensing harbor images. Wang et al. [26] used the semantic segmentation probability map as a semantic prior to constrain the super-resolution solution space. Cho et al. [27] proposed a gradient prior-aided CNN denoiser to reduce the computational complexity while enhancing the denoising performance. These priors show remarkable performance in different area.

In the image deblurring field, some image priors were proposed. Shen et al. [28] propose a human-aware deblurring method which applys a semantic prior to solve human blur and background blur, respectively. But its performance is mediocre when it processes other blurred images that do not contain people. Yuan et al. [29] apply an optical-flow prior to guide the spatially variant deconvolution network. Zheng et al. [30] and Fu et al. [31] both proposed edge priors that extract edge information with a pretrained subnet before the deblurring process and use the extra edge information to improve the deblurring result. However, their methods are limited by the extraction of edge features, because if the wrong edge features is extracted, the result of deblurring will be affected. As mentioned above, the blurring process will break edges or structures from the original sharp image. In addition, the edge features are a kind of very sparse and high-frequency feature, and the effective information contained therein is limited. To address the drawbacks of previous methods, we proposed the high-frequency reconstruction subnetwork (HFRSN) to reconstruct latent high-frequency information from blurry images. As shown in Fig. 1, the high-frequency information contains more effective information than edges. What's more important is that our HFRSN adapts to the main deblurring subnetwork

(MSGSN), so as to avoid the generation of harmful information.

B. Multi-scale Methods

An image in different resolution has different performance. Large blur artifacts will become small after reducing the resolution, which is more suitable for repairing by convolution because the convolution kernel of the same size has a larger receptive field at a lower resolution. Therefore, the multi-scale method is very suitable for image deblurring. Nah et al. [16] exploited a multi-scale CNN to remove blurs in an end-to-end fashion, which applied successive coarse-to-fine strategy to recover the sharp image from coarser-scale to finer-scale in a successive manner. But this method didn't use downsampling to reduce the size of features, so that it consumes a lot of time to calculate the full size image of each scale. Following the same strategy, Tao et al. [32] applied encoder/decoder modules in each scale and also added an LSTM path from coarser-scale to finer-scale to promote the flow of information only at the middle level, and then Gao et al. [33] applied parameter sharing between different scales which greatly reduced the parameter number of the multi-scale network. Nevertheless, these multi-scale methods did not consider the transfer of information between different scales at different levels, which leads to the blockage in information. In order to tackle such an issue, we transform the grid-like architecture into a multi-scale form and then use its dense connection to fuse features between different scales.

C. Grid-like Architecture

The grid-like architecture was first proposed by [34] who used it in semantic segmentation, but its blockwise dropout is not suitable for image restoration work, and its batch normalization [35] costs a lot of time to calculate. Liu et al. [36] improved the grid network with an attention mechanism to better remove fog from foggy pictures. The grid-like architecture shows its superiority because it has dense connection between different levels, but the previous grid-like method mentioned above ignored the connection between levels and scales. We introduce the grid-like architecture to the image deblurring field and transform it into a multi-scale architecture. In detail, we abandon tricks from above grid networks and add multi-dimensional inputs and outputs to each level of the grid-like architecture, and design a multi-scale loss to guide each scale so that the grid-like network can gradually deblur the image and capture more information from different scales. It was demonstrated that convolution layers loss image details so that Ronneberger et al. [37] proposed the skip-connection in Unet to protect image details. Inspired by that, we introduce the skip-connection into our grid-like architecture which also protect image details from losing through a lot of convolutional layers and improved our deblurring result.

III. Our Framework

In this section, we propose a novel image deblurring method, the overall architecture of which is depicted in Fig. 2. We input multi-scale blurry images to HFRSN, which reconstruct high-frequency features directly from multi-scale blurry images, and we take the high-frequency features to guide the deblurring process (MSGSN). In MSGSN, we also take multi-scale blurry images as input and perform multi-scale deblurring processes simutaneously. During each scale's deblurring process, we apply the high-frequency aggregation module (HFAG) and high-frequency attention module (HFAT) to strengthen the feature extraction stage and feature reconstruction stage.

A. High-frequency Reconstruction Subnetwork (HFRSN)

As mentioned above, an image in different resolution has different high-frequency information. And the same convolutional network working on different resolution has different reception field. Inspired by this, we propose a subnetwork (called HFRSN) which possesses multi-scale inputs and outputs to reconstruct multi-scale high-frequency information simultaneously. The network architecture of HFRSN is illustrated in the upper part of Fig. 2. As you can see, the HFRSN, taking multi-scale blurry images as inputs, follows the popular encoder-decoder structure. The purpose is to obtain accurate high-frequency information with the multi-scale fusion capability of encoder and decoder. The operation of the HFRSN can be written as

$$\{HF_1, ..., HF_S\} = \mathcal{D}(\mathcal{E}(B_1, ..., B_S)), \tag{1}$$

where S denotes the number of scales, $\mathcal{E}$ and $\mathcal{D}$ denote the encoder and the decoder, respectively, $B_i, i \in 1, ..., S$ are the multi-scale blurry image and $HF_i, i \in 1, ..., S$ are the multi-scale high-frequency information reconstructed by HFRSN.

1) Encoder and Decoder: From a functional perspective, the encoder plays the role of feature extractor, while the decoder plays the role of feature reconstructor. As shown in Fig. 3, there are three main blocks in the encoder module and decoder module. ResGroup is composed of several resblocks [38]. The Downsampling Block/Upsampling Block consists of a convolutional/deconvolutional layer and a resblock.

Specifically, the encoder is represented as a top-to-bottom path containing ResGroup and Downsampling Blocks. The ResGroup is used to extract features in different scales. The extracted features are fused by the Downsampling Blocks from high-resolution to low-resolution. Conversely, the decoder has a bottom-to-top path containing ResGroup and Upsampling Blocks. It is designed to reconstruct images from low-resolution to high-resolution. Briefly, the encoder and decoder extract and then fuse multi-scale features.

2) Discrete Cosine Transform: There are many transformations that can be used to extract high-frequency information of the image, such as Discrete Cosine Transform (DCT) [39], Wavelet [40], and Framelet [41]. Here we select DCT to transforms images from spatial representation to frequency representation. The high-frequency coefficients representing image edges or shapes are regarded as high-frequency information of the image and they are used as the ground-truth of our HFRSN.

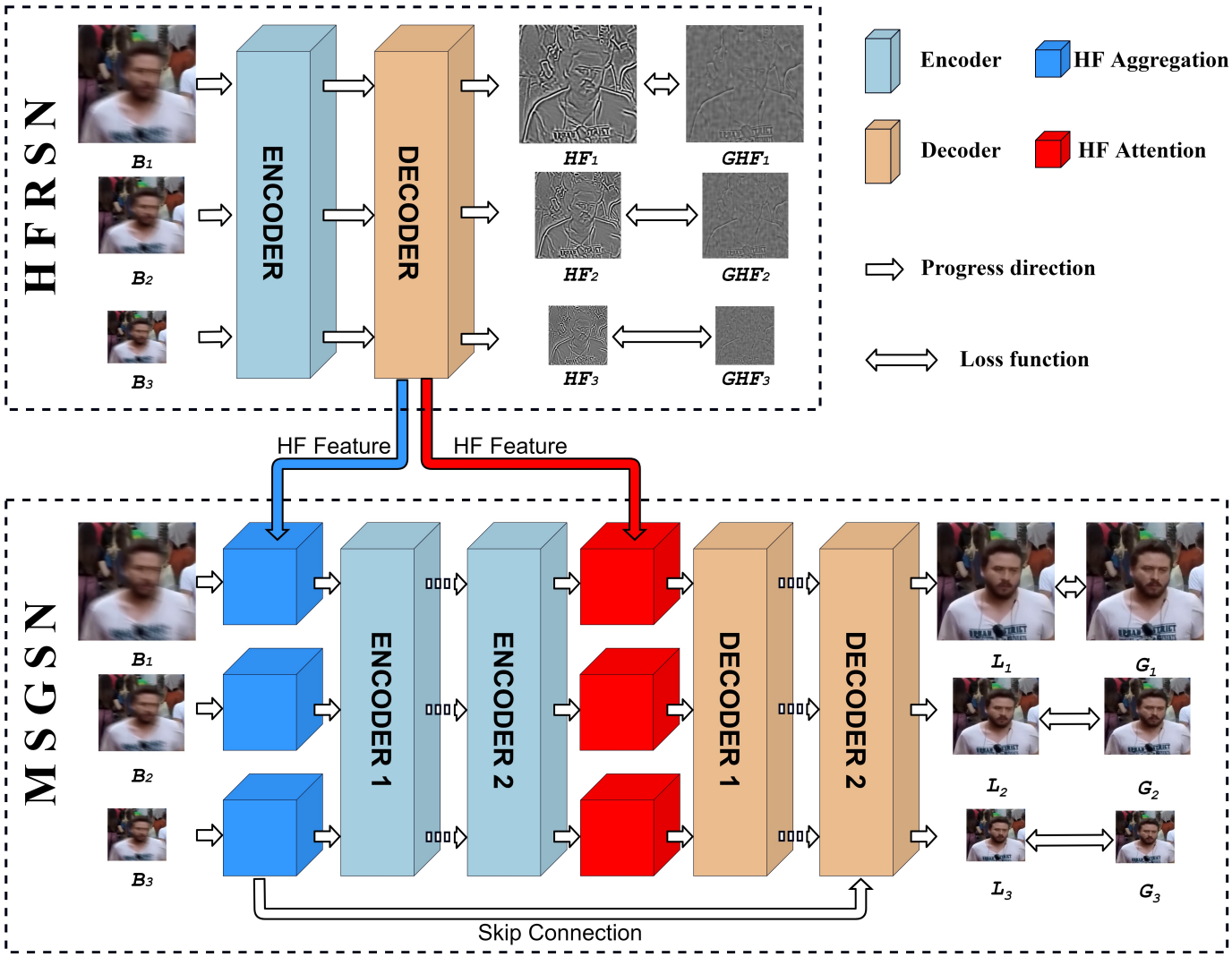


Fig. 2: The whole network of our method. The upper part is our high-frequency reconstruction subnetwork, and the lower part is our multi-scale grid subnetwork.

B. Multi-scale Grid Subnetwork

It has been demonstrated that the multi-scale approaches [16], [32], [33] are capable of recovering the sharp image from the blurry one in the field of single image deblurring. Benefitting from the different reception field, the multi-scale approaches are able to handle different sizes of blurs. Besides HFRSN, our MSGSN also follows the multi-scale manner.

MSGSN applies the grid-like architecture to perform multi-scale deblurring in parallel. The main architecture of the MSGSN is shown in the lower part of Fig. 2. It takes multi-scale blurry images as input, processes multi-scale feature extraction and feature reconstruction in parallel and finally outputs multi-scale deblurred results. Compared with HFRSN, MSGSN also uses the same encoder and decoder to extract and aggregate image features from different scales despite of the doubled number. Besides, we apply high-frequency aggregation module (HFAG) and high-frequency attention module (HFAT) to enhance the capability of the encoder and decoder, respectively. In the feature extraction stage, the HFAG is used to aggregate image features and high-frequency features together which enriches the feature diversity. While in the feature reconstruction stage, the high-frequency attention map calculated by HFAT is used to strengthen the reconstruction of high-frequency regions. The details of the HFAG and the HFAT will be described later. In addition, to preserve image details, we introduce the skip connection from U-net [37] to improve the performance of the grid-like architecture.

Given the multi-scale blurry images $B_i, i \in \{1, ..., S\}$ as input, the MSGSN executes HFAG $\mathcal{G}$ to aggregate image features and image high-frequency features in each scale, it can be formulated as

$$F_i^G = \mathcal{G}(B_i, F^{HF}), i \in \{1, ..., S\}, \tag{2}$$

where F^{HF} denotes the high-frequency features reconstructed by HFRSN, i.e., the output of the last convolutional layer in HFRSN.

Then, the aggregated features F_i^G are fed into cascaded encoders $\mathcal{E}_1$ and $\mathcal{E}_2$ to extract deeper image features,

$$\{F_1, ..., F_S\} = \mathcal{E}_2(\mathcal{E}_1(F_1^G, ..., F_S^G)). \tag{3}$$

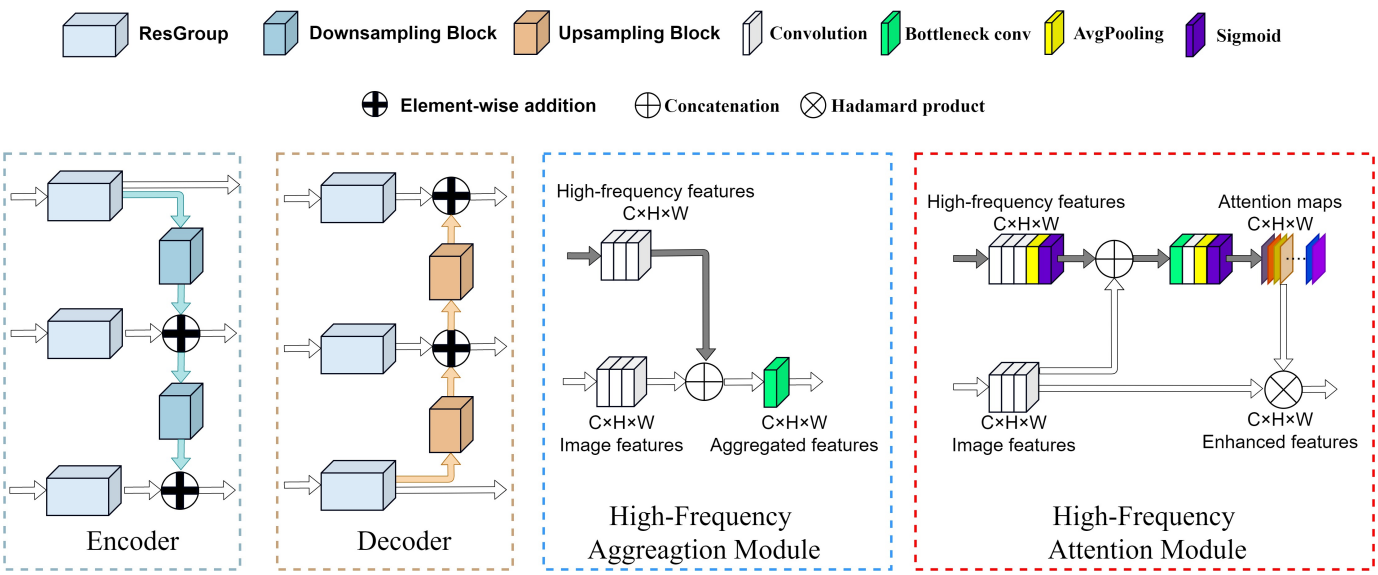


Fig. 3: The illustration of the encoder module, the decoder module, the high-frequency aggregation module and the high-frequency attention module.

After that, the MSGSN utilizes HFAT $\mathcal{T}$ to enhance the reconstruction of the high-frequency area,

$$F_i^T = \mathcal{T}(F_i, F^{HF}), i \in \{1, ..., S\}. \tag{4}$$

Then, the enhanced features F_i^T are used to reconstruct clear image L_i with cascaded decoders $\mathcal{D}_1$ and $\mathcal{D}_2$,

$$\begin{aligned} \{F_1^{IM}, ..., F_S^{IM}\} &= \mathcal{D}_1(F_1^T, ..., F_S^T), \\ \{L_1, ..., L_S\} &= \mathcal{D}_2((F_1^{IM} + F_1^G), ..., (F_S^{IM} + F_S^G)), \end{aligned} \tag{5}$$

where the F_i^{IM} are the intermediate outputs of the first decoder. We input the sum of F_i^{IM} and F_i^G to the second decoder to preserve the low-level details of the input image.

1) High-frequency Aggregation module: Inspiring by [24], we select concatenation to fuse image features and high-frequency features in each scale. However, the two types of features are independent of each other, which is not conducive to effective feature fusion. Thus we use convolutional layers to refine the features separately and make them have the same number of channels before concatenation. Then we apply a bottleneck convolution layer to extract useful features. The network details of HFAG module are illustrated in Fig. 3. With the help of HFAG, our deblurring method make full use of extra high-frequency information to better recover sharp images.

2) High-frequency Attention module: To better take advantage of high-frequency information, HFAT module is introduced to enhance the deblurring process. The HFAT network is shown in Fig. 3, which takes the concatenation of high-frequency features and image features as input and adaptively learn high-frequency attention maps. Then the attention maps multiply image features so as to pay more attention to high-frequency area.

C. Loss Function

Our loss function contains three terms: pixel consistency loss $\mathcal{L}_C$, perceptual loss $\mathcal{L}_P$, and high-frequency loss $\mathcal{L}_{HF}$. The overall loss function $Loss$ is defined as follows:

$$Loss = \mathcal{L}_C + \lambda_P \mathcal{L}_P + \lambda_{HF} \mathcal{L}_{HF}, \tag{6}$$

where λ_P and λ_{HF} control the weight of the perceptual loss and the high-frequency loss, respectively. Specifically, $\mathcal{L}_C$ controls the pixel level accuracy, $\mathcal{L}_P$ measures the high level feature similarity between the reconstructed image and ground-truth image and $\mathcal{L}_{HF}$ guarantees the HFRSN to reconstruct real high-frequency information from blurry images. It is worth noting that we calculate all the loss functions in different scales separately, which helps our method better handle multi-scale process.

1) Pixel Consistency Loss: As shown in Fig. 2, we let L_i and G_i respectively denote the deblurred sharp image and the ground-truth sharp image in ith scale. The pixel consistency loss can be expressed as

$$\mathcal{L}_C = \frac{1}{2S} \sum_{i=1}^{S} \|L_i - G_i\|_2^2 . \tag{7}$$

2) Perceptual Loss: The perceptual loss is defined as the l_2-norm between the VGG-19 features of the deblurred sharp image L_i and the ground-truth sharp image G_i in each scale:

$$\mathcal{L}_P = \sum_{i=1}^{S} \frac{1}{2SC_jH_jW_j} \|\phi_j(L_i) - \phi_j(G_i)\|_2^2 , \tag{8}$$

where $\phi_j(L_i)$ and $\phi_j(G_i)$ denote the aforementioned VGG19 feature maps from the jth level associated with the deblurred sharp image and the ground-truth sharp image in each scale, and C_j, H_j and W_j are dimensions of the feature. In our work, we use the feature from the conv3_3 layer ($j = 15$).

3) High-frequency Loss: The high-frequency loss function can be defined as

$$\mathcal{L}_{HF} = \frac{1}{2S}\sum_{i=1}^{S}\|HF_i - GHF_i\|_2^2, \tag{9}$$

where GHF_i denotes the high-frequency information extracted from ground-truth image.

IV. Experiments

In this section, we conduct a series of experiments to evaluate the effectiveness and efficiency of our method. First, we explain the implementation details of our method and the datasets we used for training and validation. Then we show the comparisons on synthetic datasets and a real-world dataset with start-of-the-art methods. Next, we explore our method on high-level task. Finally, we present an ablation study to investigate the effect of each component in our methods. Note that we use Matlab to calculate the PSNR and SSIM values on the RGB color space.

A. Experimental Settings

1) Implementation: We implement our method in PyTorch on a single NVIDIA V100 GPU. During training, we randomly crop the blurry and sharp images to 512×512 in pixel size and then downsample them several times for different scales. We use the DCT to get ground-truth high-frequency information maps. The batch size is set to 1 for training. We use the Xavier method [42] to initialize network parameters. The network is optimized by the Adam solver [43] for 600 epochs. Initial learning rate is set to 0.0001, and then exponentially decayed to 0 using power 0.3. We use 3×3 convolution kernel for convolutional layers all over the networks, and for different scales in HFRSN and MSGSN, we set the numbers of filters as 40, 80, 160 from large scale to small scale. We set the numbers of scale (S) to 3 in order to balance the parameters and performance. In the total loss function, we set $\lambda_P = 0.01$ and $\lambda_{HF} = 0.01$ for the best result which will be discussed in section IV.C.

2) Datasets: We choose the GoPro dataset [16] to train our method, and use the GoPro dataset [16] and the HIDE dataset [28] to evaluate our method. Moreover, we choose blurry images from [44] for comparisons with real-world images.

The GoPro dataset [16] consists of 3214 pairs of blurry images and corresponding sharp images, and each blurry image was synthesized from multiple continuous sharp images, which can simulate the way of real blur generation. The GoPro dataset was captured at several different scenarios by a high-speed action camera, whose image size is 720×1280. For a fair comparison, we follow the same protocol in [16], which uses 2103 image pairs for training and 1111 image pairs for testing.

The HIDE dataset [28] consists of 8422 pairs of blurry images and corresponding sharp images, and the images are divided into two categories, i.e. , long-shot (HIDE I) and close-shot (HIDE II). Evaluating each group can capture different aspects of the multi-motion blurring problem. In addition, we evaluate the model trained on the GoPro dataset [16] on the HIDE dataset to evaluate the generalization ability of our methods.

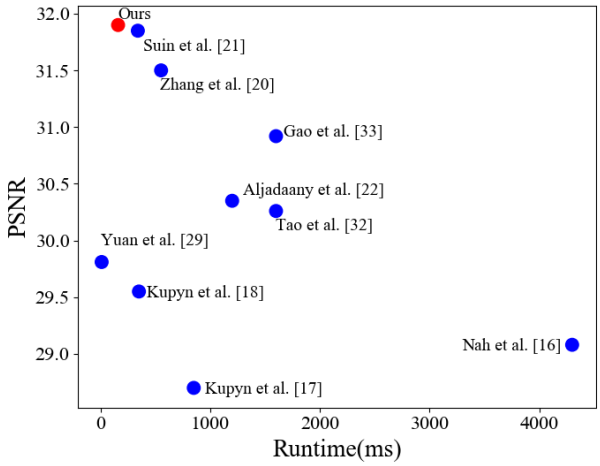


Fig. 4: The PSNR vs Runtime of our methods and the state-of-the-art deblurring methods.

Models	PSNR	SSIM	Time(ms)	Params(Mb)
Nah et al. [16]	29.08	0.913	4300	21
Kupyn et al. [17]	28.70	0.958	850	null
Tao et al. [32]	30.26	0.934	1600	6.4
Gao et al. [33]	30.92	0.942	1600	**2.8**
Aljadaany et al. [22]	30.35	**0.961**	1200	6.7
Zhang et al. [20]	31.50	0.948	552	27.6
Kupyn et al. [18]	29.55	0.934	350	3.3
Suin et al. [21]	31.85	0.948	340	null
Yuan et al. [29]	29.81	0.936	10	3.1
Jiang et al. [45]	31.79	0.949	null	null
ours	**31.90**	0.951	160	25.7

TABLE I: Comparison with other deblurring methods on GoPro dataset [16].

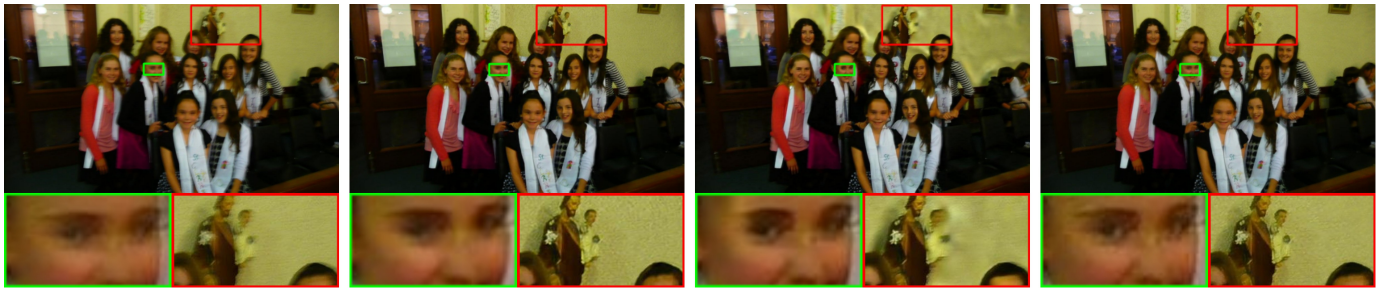
B. Performance Comparisons

We compare our method with other state-of-the-art deblurring methods [16], [17], [32], [33], [22], [20], [18], [21], [29], [45] on three datasets as mentioned above. Unless stated otherwise, all the reported results are directly copied from the original paper and the null stand for the result cannot be found in the original paper and its open source code cannot be found.

1) Comparision on the GoPro dataset: We first compare our method with other deblurring methods on the GoPro evaluation dataset. The quantitative results are listed in TABLE I. Time and Params in the tables refer to inference time and the number of network parameters. Our method performs better than previous multi-scale methods [16], [32], [33] because our method has high frequency guidance. Method [29] achieves the fastest speed on processing 720×1280 image with a small model. Our proposed method achieves a competitive result (31.90 dB in PSNR). Furthermore, we evaluated the computational efficiency of the aforementioned state-of-the-art methods as shown in Fig. 4, our method achieves good balance between speed and PSNR. It is worth noting that the

(a) Comparison on the GoPro dataset.

(b) Comparison on the HIDE dataset.



(c) Comparison on the real-world blurry image.

Fig. 5: Visual comparison on the synthetic datasets and real-world blurry image, respectively. From left to right: the blurry image, the deblurred result of [32], [20] and ours, respectively. The zoomed-in area is shown below each result image.

Models	HIDE I		HIDE II	
	PSNR	SSIM	PSNR	SSIM
Nah et al. [16]	27.11	0.897	26.01	0.870
Tao et al. [32]	28.01	0.920	26.97	0.901
Gao et al. [33]	29.98	0.943	28.14	0.919
Kupyn et al. [18]	26.28	0.879	25.09	0.858
Shen et al. [28]	29.60	0.941	28.12	0.919
Zhang et al. [20]	29.79	0.942	28.33	0.924
ours	**30.33**	**0.943**	**28.61**	**0.925**

TABLE II: Comparison with other deblurring methods on HIDE dataset [28].

runtime of network is not directly related to the number of parameters. It is affected by many factors, such as network architecture, convolution kernel size, different neural network modules, etc.

Visual comparison on the GoPro evaluation dataset is shown in Fig. 5. No matter it is a small blur caused by object motion or a large blur caused by camera shake, our method can obtain a promising deblurring result. Looking at the enlarged image area, it can be seen that our method reconstructs more high-frequency information. Please zoom in for more details.

2) Comparision on the HIDE dataset: In order to verify the generalization ability, we further evaluate our model on the HIDE test set. TABLE II shows a quantitative evaluation in terms of PSNR and SSIM, where our model achieves the competitive performance, which proves that our method can remove blurs in different scenes well, even though our model is trained on the GoPro dataset. Fig. 5 presents the visual comparison on the HIDE dataset, our method deblurs better and reconstructs more high-frequency details.

3) Comparision on the real-world images: We further compare our method against previous deblurring methods on the real-world dataset [44]. Since there is no ground-truth for these real-world blurry images, we can only make qualitative comparisons. As shown in Fig. 5, method [32] can remove blur and strengthen the edges well, but with the loss of details. Method [20] causes many artifacts. Our method removes blur without generating artifacts and reconstructs more high-frequency information.

4) Comparision on High-level Task: The quality of the deblurred images will affect the performance of the high-level computer vision tasks, such as object detection and image classification. To further demonstrate the effectiveness of our method, we conducted a comparative experiment on high-level computer vision tasks. Specifically, we used the VGG19 network to train a flower classifier on the flower classification dataset [46]. The dataset, containing 102 types of flower images and are divided into a training set and a test set. Eight random blur kernels generated by the same method as [47] are used to blur each image in the test set. We train the flower classifier on the clear training set, and test it using clear test images, blurred test images, deblurred images by method [20], method [33] and our method, respectively. The classification accuracy is presented in Fig. 6, One can see that blurry images severely reduce the accuracy of high-level image classification task. The deblurred images reconstructed by our method get

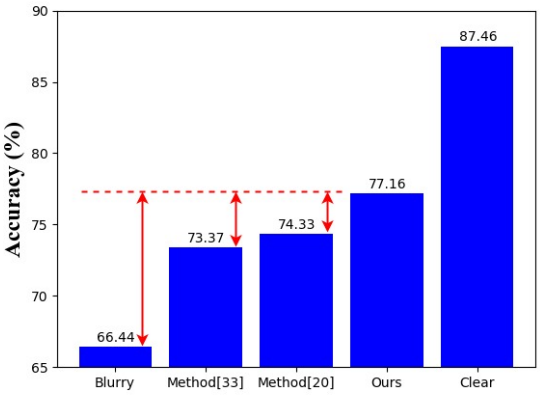


Fig. 6: Exploration on image classification.

Method	PSNR	SSIM	Params(Mb)
(a) w/o HFRSN	31.21	0.944	16.5
(b) w/o joint training	31.71	0.949	25.7
(c) our method	31.90	0.951	25.7
(d) DMPHN [20]	30.21	0.935	7.2
(e) DMPHN [20] with HFRSN	30.49	0.939	9.7

TABLE III: The impact of HFRSN and joint training strategy on model performance.

77% classification accuracy, which is 11% higher than the blurry images, 3% higher than that using deblurred images by method [20], 4% higher than that using deblurred images by method [33]. This is a huge improvement, demonstrating that our method can reconstruct clearer images than other deblur methods.

C. Ablation Analysis

We conduct a series of ablation studies to demonstrate the efficiency of our method. We use GoPro dataset for evaluation.

1) Ablation Analysis on High-frequency Prior: In recent years, image priors in deep learning-based methods have gradually received researcher's attention. In this paper, we propose a high-frequency prior to the single image deblurring field. To demonstrate its effectiveness, we designed a series of experiments.

(i) As mentioned in section III, our high-frequency features is reconstructed by our HFRSN and we use it as a prior in MSGSN. In order to verify the importance of our high-frequency prior, we separately trained a MSGSN without HFRSN and a MSGSN with HFRSN for comparison experiments. As shown in Fig. 7, high-frequency prior can help our MSGSN network recover sharper edges and structures. We also performed quantitative analysis on the GoPro test set, as shown in (a) and (c) of TABLE III, our method improved 0.69dB in PSNR with the assistance of high-frequency prior. It is clearly showed that our HFRSN can reconstruct high-frequency features and help the MSGSN to get better deblur results.

(ii) Our proposed high-frequency prior is independent of the backbone. To further demonstrate the effectiveness of the

Fig. 7: Visual comparison of our method w/o HFRSN and our method with HFRSN

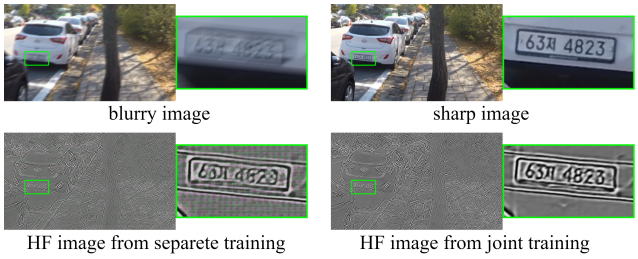


Fig. 8: Visual comparison of separate training and joint training

proposed high-frequency prior, we apply it to DMPHN [20], a classic multi-patch deblurring network. As shown in (d) and (e) of TABLE III, our high-frequency prior can improve the deblurring performance of DMPHN. It is worth mentioning that we only use single-level high-frequency guidance on the DMPHN network, the performance can be further improved by using multi-level high-frequency guidance.

(iii) Previous methods [30], [31] use the edge prior to enhance the high-frequency feature extraction by utilizing a pretrained network to extract edges from blurry images. But wrong edge features may be extracted. In our method, we train the two subnets (MSGSN and HFRSN) jointly, thus the final error gradient will propagate to the HFRSN to avoid the reconstruction of harmful high-frequency information. The difference between joint training and separate training can be seen in (b) and (c) of TABLE III. Compared with separate training, joint training brings about 0.2dB improvement in PSNR. In addition, we show the visual comparison of joint training and separate training in Fig. 8. It is obvious that the high-frequency image reconstructed by joint training has more sharper edge than separate training does, which further demonstrates the importance of joint training.

(iv) The edge features are one of the most important component of image features, which has been widely used in image reconstruction tasks [23], [24]. Nevertheless, the edge features is seriously destroyed by the blurring process, so it is hard to recover sharp edges from the blurry image. In addition, the edge features are a kind of very sparse and high-frequency feature, so the effective information contained therein is limited. By contrast, high-frequency prior has more information than edge prior so it is easier to extract useful information by networks. Therefore, we propose the high-frequency prior to guide the deblurring process.

In order to verify the effectiveness of our high-frequency prior, we trained a model with (a) the Canny edge map; (b) the Laplace edge map; (c) the gradient map; (d) the Wavelet high-frequency map; (e) the DCT high-frequency map. The visual and quantitative comparison are shown in Fig. 9 and TABLE IV, respectively. It can be seen that the performance of the high-frequency priors (c, d, e) is better than that of the edge (a, b), and our DCT high-frequency prior achieves the best performance.

Image prior	PSNR	SSIM
(a) Canny edge	31.72	0.949
(b) Laplace edge	31.67	0.948
(c) Gradient	31.80	0.950
(d) Wavelet Transform	31.79	0.950
(e) our method (DCT)	31.90	0.951

TABLE IV: Comparision of different image prior.

Method	PSNR	SSIM	Params(Mb)
(a) w/o HFAG	31.40	0.947	25.2
(b) directly concatenating	31.78	0.950	25.4
(c) w/o HFAT	31.69	0.949	22.3
(d) single-level guidance	31.62	0.948	22.4
(e) our method	31.90	0.951	25.7

TABLE V: Comparision of high-frequency guidance.

2) Ablation Analysis on High-frequency Guidance: As mentioned above, high-frequency features have been reconstructed by our HFSRN, but it is also a challenge to apply the high-frequency to guide the deblurring process. Different from other prior guidance methods [23], [31], our high-frequency prior guidance is implemented by two different modules HFAG and HFAT and we apply them at different scales. We designed several sets of comparison models to demonstrate the effectiveness of our high-frequency guidance: (a) a model without HFAG; (b) a model directly concatenating the high-frequency features and image features; (c) a model without HFAT; (d) a model with single-level guidance; (e) our method with multi-level guidance. The result is shown in TABLE V. The model without HFAG (a) is nearly 0.5 dB worse than our complete model (e) and the (c) is 0.2 dB worse than the model (e), which demonstrates our HFAG and HFAT module is effective. Moreover, the model (b) directly concatenating the high-frequency features and image features shows worse result than our complete model (e), which also verifies the effectiveness of the proposed HFAG. The result of single-level guidance (d) is 0.28 dB lower than our complete model (e), which demonstrates the effectiveness of the multi-level guidance.

Besides, we showed the visualization of our high-frequency attention map in Fig. 10. We can observe the high correlation between estimated attention weights and the high-frequency regions presented in the image. Thus, our decoders will pay more attention to feature reconstruction of high-frequency regions after multiplying image features with our attention maps.

3) Ablation Analysis on Network Structure: Both HFRSN and MSGSN use the encoder-decoder structure to extract and reconstruct multiple scale features in parallel. In this part,

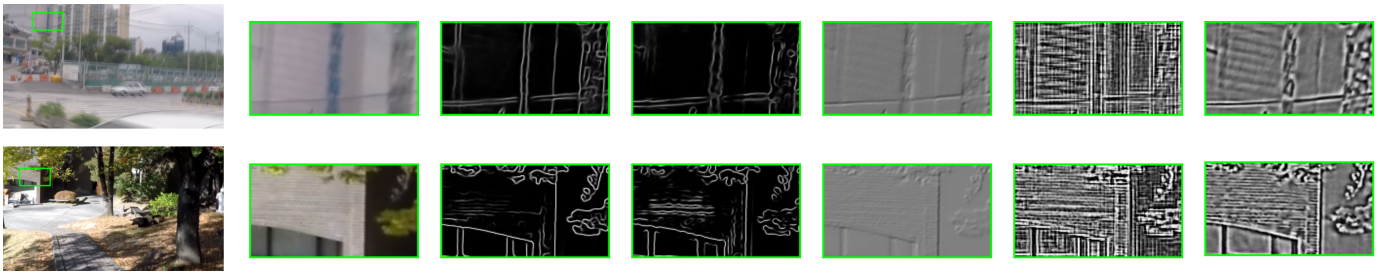


Fig. 9: Visualization of different image priors used by our method.

Fig. 10: Visualization of high-frequency attention map from HFAT.

we will discuss the effectiveness of the multi-scale strategy, encoder and decoder in HFRSN and MSGSN. Moreover, we will discuss the effectiveness of our MSGSN backbones.

(i) HFRSN: here we train the HFRSN with single scale and with different numbers of encoder/decoder, respectively. The result is shown in TABLE VI. The single scale (a) is 0.2dB lower than the multi-scale (b), which demonstrates that our multi-scale HFRSN can extract more high-frequency information from multi-scale blurry images. It is worth noting that our single scale model (a) has the same architecture and the same number of parameters of our complete method (b) except the multi-scale inputs and outputs. The model (c) to (f) represent the results of different numbers of encoders and decoders. We set the numbers of encoders and decoders of MSGSN to 1 for convenience. It can be seen that increasing the numbers of encoders and decoders cannot bring much performance improvement but will take a long time to converge when the numbers of parameters increased. So we choose the E1D1 as our proposed HFRSN.

(ii) MSGSN: similar experiments are also designed for MSGSN. The experiment result is showed in TABLE VII. The model (b) performs better than the model (a) about 0.57dB but does not bring additional parameters, which demonstrates that multi-scale features are very important for our deblurring method. The model (c) to (f) shows that the performance of the different numbers of encoders and decoders. We can observe that $E2D2$ showed the best performance. Different from HFRSN, we found that increasing the numbers of encoders and decoders in MSGSN will bring an improvement in deblurring result and the reason may lie in that MSGSN has more image information to be processed. However, when we continue to increase the numbers of encoders and decoders, we found the improvement is limited. So we choose E2D2 as our proposed MSGSN.

Method	PSNR	SSIM	Params(Mb)
(a) w/o multi-scale	31.70	0.949	25.7
(b) our method	31.90	0.951	25.7
(c) E1D1	31.23	0.945	17.5
(d) E2D1	31.26	0.945	21.6
(e) E1D2	31.25	0.945	21.6
(f) E2D2	31.29	0.945	25.7

TABLE VI: Results of the impact of each component in HFRSN. The $ExDy$ means the model has x encoder and y decoder.

Method	PSNR	SSIM	Params(Mb)
(a) w/o multi-scale	31.33	0.946	25.7
(b) our method	31.90	0.951	25.7
(c) E1D1	31.23	0.945	17.5
(d) E2D1	31.55	0.948	21.6
(e) E1D2	31.61	0.949	21.6
(f) E2D2	31.90	0.951	25.7

TABLE VII: Results of the impact of each component in MSGSN. The $ExDy$ means the model has x encoder and y decoder.

(iii) Our MSGSN backbone is inspired by the grid-like network [34] and the multi-scale network [16] and aims to combine their advantages. In detail, we use the dense connections of the grid network to fuse information from multiple scales and use the skip-connection to preserve low-level features. We compared our MSGSN with other three networks: previous grid-like network [34], previous multi-scale network [33], and our MSGSN backbone without skip-connection. For a fair comparison, we guarantee the four networks to have similar parameters and numbers of convolutional layers, and trained them on GoPro dataset [16] with the same training strategy. As illustrated in TABLE VIII, our MSGSN backbone achieves the best results, which demonstrates that our MSGSN backbone is able to combine the advantage of both the grid-like network and the multi-scale network.

Method	PSNR	SSIM	Params(Mb)
(a) GN	30.96	0.942	16.3
(b) MSN	31.03	0.942	16.5
(c) MSGN w/o skip-connection	31.13	0.943	16.5
(d) MSGN	31.21	0.944	16.5

TABLE VIII: Comparison with different backbones in MSGSN. GN denotes the previous grid-like network, MSN denotes the previous multi-scale network, and MSGN presents the backbone network of our MSGSN.

Method	PSNR	SSIM
(a) $\mathcal{L}_C$	29.53	0.929
(b) $\mathcal{L}_P$	27.06	0.881
(c) $\mathcal{L}_C + \lambda_P \mathcal{L}_P$	30.94	0.940
(d) $\mathcal{L}_C + \lambda_P \mathcal{L}_P + \lambda_{HF} \mathcal{L}_{HF}$	31.23	0.945

TABLE IX: Ablation study on loss function

4) Ablation Analysis on Loss Function: Our loss function contains three terms: pixel consistency loss $\mathcal{L}_C$, perceptual loss $\mathcal{L}_P$, and high-frequency loss $\mathcal{L}_{HF}$. Specifically, in order to show the importance of each loss term, several cases are considered: (a) a model trained only with pixel consistency loss $\mathcal{L}_C$; (b) a model trained only with perceptual loss $\mathcal{L}_P$; (c) a model trained with pixel consistency loss $\mathcal{L}_C$ and high-frequency loss $\mathcal{L}_{HF}$; (d) a model trained with all three loss terms. The result is shown in TABLE IX. Note that case (c) is used in many deblurring literatures [17], [18], and it can indeed improve the performance compared with cases only using single loss, i.e., cases (a) and (b). On the basis, we add another loss term called high-frequency loss to further boosting the performance. It is worth reminding that high-frequency loss $\mathcal{L}_{HF}$ is only applied to HFRSN, so it cannot be used alone to train the entire network.

5) Ablation Analysis on Hyper-parameters: Our total loss function Eq. (6) contains three terms so that we set two hyper-parameters λ_P and λ_{HF} to control the balance. The λ_P controls the weight of the perceptual loss Eq. (8) and the λ_{HF} controls the weight of the high frequency loss Eq. (9). In order to find the most suitable hyper-parameters, we designed following experiments. (1) We fixed the λ_{HF} to 0.01, and then use different values of λ_P to train our model. The result is shown in Fig.11, we found that the best value of λ_P is 0.01. (2) We fixed the λ_P to 0.01 and keep it constant, and then use different values of λ_{HF} to train our model. The result is shown in Fig.11. It is clearly shown that if the λ_{HF} is too small, the effect of high-frequency guidance is limited. In contrast, if λ_{HF} is too big, the deblurred images will be over-sharpened. We found that the best λ_{HF} is 0.01 in our experiments.

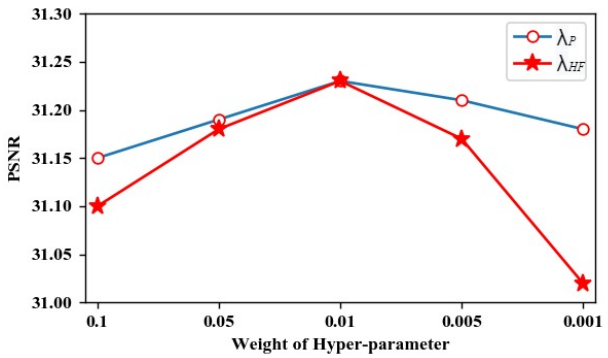


Fig. 11: Ablation study on λ_P and λ_{HF} values.

V. Conclusions

In this paper, we propose a novel deblurring framework for single image deblurring task by introducing a high-frequency prior to convolutional networks. Speciffically, we built a high-frequency reconstruction subnetwork (HFRSN) to reconstruct high-frequency features directly from multiple-resolution blurry images. And then we built a multi-scale grid subnetwork (MSGSN) to fuse high-frequency features and image features at multiple scales. In order to make better use of the extracted high-frequency features, we designed two modules named HFAG and HFAT. The HFAG is built to better fuse high-frequency features and image features to strengthen the feature extraction. While the HFAT calculates attention maps to enhance the reconstruction of the high-frequency information. Through extensive evaluations of both qualitative and quantitative criteria, it is demonstrated that our approach has a competitive advantage over the state-of-the-art methods.

References

[1] J. Pan, D. Sun, H. Pfister, and M. Yang, "Deblurring images via dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2315–2328, 2018.
[2] L. Xu, S. Zheng, and J. Jia, "Unnatural L0 sparse representation for natural image deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1107–1114.
[3] T. F. Chan and C. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 370–375, 1998.
[4] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao, "Image deblurring via extreme channels prior," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6978–6986.
[5] L. Chen, F. Fang, T. Wang, and G. Zhang, "Blind image deblurring with local maximum gradient prior," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1742–1750.
[6] J. Pan, Z. Hu, Z. Su, and M. Yang, "Deblurring text images via l0-regularized intensity and gradient prior," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2901–2908.
[7] C. Li, C. Guo, J. Guo, P. Han, H. Fu, and R. Cong, "Pdr-net: Perception-inspired single image dehazing network with refinement," *IEEE Trans. Multim.*, vol. 22, no. 3, pp. 704–716, 2020.
[8] R. Furuta, N. Inoue, and T. Yamasaki, "Pixelrl: Fully convolutional network with reinforcement learning for image processing," *IEEE Trans. Multim.*, vol. 22, no. 7, pp. 1704–1719, 2020.
[9] Y. Du, G. Han, Y. Tan, C. Xiao, and S. He, "Blind image denoising via dynamic dual learning," *IEEE Trans. Multim.*, pp. 1–1, 2020.
[10] Y. Wang, D. Gong, J. Yang, Q. Shi, A. van den Hengel, D. Xie, and B. Zeng, "Deep single image deraining via modeling haze-like effect," *IEEE Trans. Multim.*, pp. 1–1, 2020.
[11] X. Yang, H. Mei, J. Zhang, K. Xu, B. Yin, Q. Zhang, and X. Wei, "DRFN: deep recurrent fusion network for single-image super-resolution with large factors," *IEEE Trans. Multim.*, vol. 21, no. 2, pp. 328–337, 2019.
[12] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2808–2817.
[13] R. Liu, Y. He, S. Cheng, X. Fan, and Z. Luo, "Learning collaborative generation correction modules for blind image deblurring and beyond," in *ACM Multimedia Conference on Multimedia Conference*, 2018, pp. 1921–1929.

[14] L. Xu, J. S. J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014*, 2014, pp. 1790–1798.

[15] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 769–777.

[16] S. Nah, T. H. Kim, and K. M. Lee, "Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 257–265.

[17] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8183–8192.

[18] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *IEEE International Conference on Computer Vision*, 2019, pp. 8877–8886.

[19] T. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature pyramid networks for object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 936–944.

[20] H. Zhang, Y. Dai, H. Li, and P. Koniusz, "Deep stacked hierarchical multi-patch network for image deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5978–5986.

[21] M. Suin, K. Purohit, and A. N. Rajagopalan, "Spatially-attentive patch-hierarchical network for adaptive motion deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3603–3612.

[22] R. Aljadaany, D. K. Pal, and M. Savvides, "Douglas-rachford networks: Learning both the image prior and data fidelity terms for blind image deconvolution," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 235–10 244.

[23] F. Fang, J. Li, and T. Zeng, "Soft-edge assisted network for single image super-resolution," *IEEE Trans. Image Process.*, vol. 29, pp. 4656–4668, 2020.

[24] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, "Structure-preserving super resolution with gradient guidance," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7766–7775.

[25] D. Cheng, G. Meng, S. Xiang, and C. Pan, "Fusionnet: Edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, pp. 5769–5783, 2017.

[26] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 606–615.

[27] S. I. Cho and S. Kang, "Gradient prior-aided CNN denoiser with separable convolution-based optimization of feature dimension," *IEEE Trans. Multim.*, vol. 21, no. 2, pp. 484–493, 2019.

[28] Z. Shen, W. Wang, X. Lu, J. Shen, H. Ling, T. Xu, and L. Shao, "Human-aware motion deblurring," in *IEEE International Conference on Computer Vision*, 2019, pp. 5571–5580.

[29] Y. Yuan, W. Su, and D. Ma, "Efficient dynamic scene deblurring using spatially variant deconvolution network with optical flow guided training," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3552–3561.

[30] S. Zheng, Z. Zhu, J. Cheng, Y. Guo, and Y. Zhao, "Edge heuristic GAN for non-uniform blind deblurring," *IEEE Signal Process. Lett.*, vol. 26, no. 10, pp. 1546–1550, 2019.

[31] Z. Fu, Y. Zheng, H. Ye, Y. Kong, J. Yang, and L. He, "Edge-aware deep image deblurring," *CoRR*, vol. abs/1907.02282, 2019. [Online]. Available: http://arxiv.org/abs/1907.02282

[32] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8174–8182.

[33] H. Gao, X. Tao, X. Shen, and J. Jia, "Dynamic scene deblurring with parameter selective sharing and nested skip connections," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3848–3856.

[34] D. Fourure, R. Emonet, É. Fromont, D. Muselet, A. Trémeau, and C. Wolf, "Residual conv-deconv grid network for semantic segmentation," in *British Machine Vision Conference*, 2017.

[35] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.

[36] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *IEEE International Conference on Computer Vision*, 2019, pp. 7313–7322.

[37] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European Conference on Computer Vision*, 2016, pp. 630–645.

[39] K. R. Rao and P. C. Yip, *Discrete Cosine Transform - Algorithms, Advantages, Applications*, 1990.

[40] M. Farge, "Wavelet transforms and their applications to turbulence," in *Annual review of fluid mechanics, Vol. 24*, 1992, pp. 395–457.

[41] P. V. Hough, "Method and means for recognizing complex patterns," 12 1962. [Online]. Available: https://www.osti.gov/biblio/4746348

[42] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.

[44] W. Lai, J. Huang, Z. Hu, N. Ahuja, and M. Yang, "A comparative study for single image blind deblurring," in *IEEE Computer Vision and Pattern Recognition*, 2016, pp. 1701–1709.

[45] Z. Jiang, Y. Zhang, D. Zou, J. S. J. Ren, J. Lv, and Y. Liu, "Learning event-based motion deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3317–3326.

[46] M. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *Sixth Indian Conference on Computer Vision, Graphics & Image Processing, ICVGIP 2008, Bhubaneswar, India, 16-19 December 2008*. IEEE Computer Society, 2008, pp. 722–729.

[47] G. Boracchi and A. Foi, "Modeling the performance of image restoration from motion blur," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3502–3517, 2012.