Gated Fusion Network for Joint Image Deblurring and Super-Resolution

Xinyi Zhang[1]
http://xinyizhang.tech
Hang Dong[1]
dhunter@stu.xjtu.edu.cn
Zhe Hu[2]
zhe.hu@hikvision.com
Wei-Sheng Lai[3]
wlai24@ucmerced.edu
Fei Wang[1]
wfx@mail.xjtu.edu.cn
Ming-Hsuan Yang[34]
mhyang@ucmerced.edu

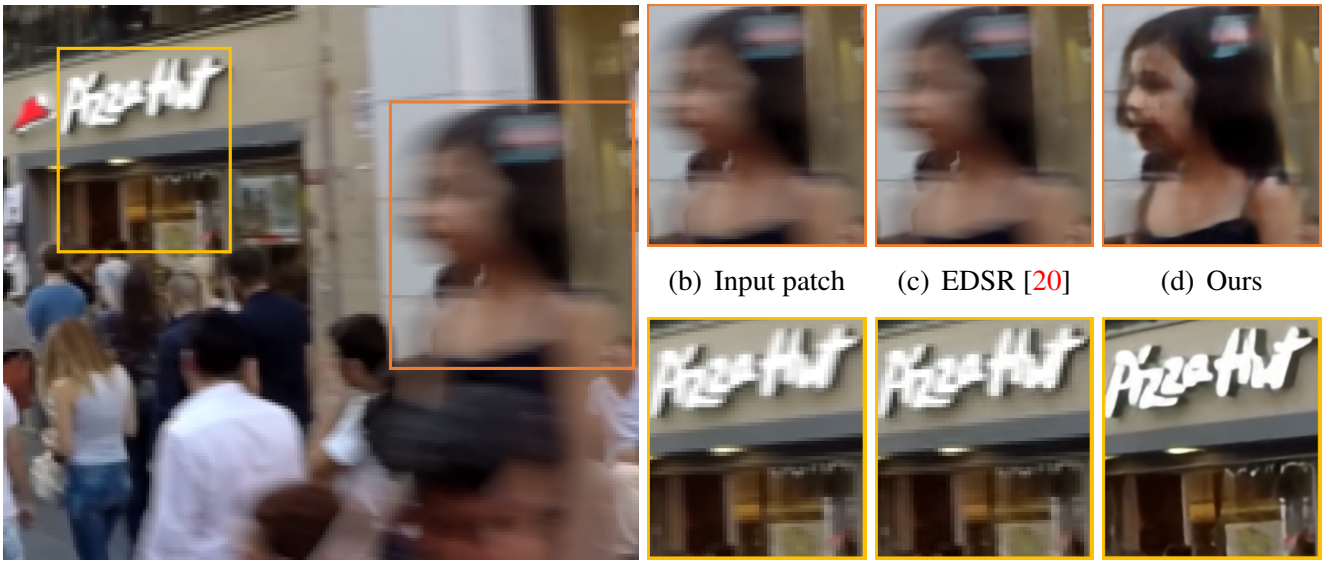
[1] National Engineering Laboratory for Vision Information Processing and Application, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[2] Hikvision Research
[3] University of California, Merced
[4] Google Cloud

Abstract

Single-image super-resolution is a fundamental task for vision applications to enhance the image quality with respect to spatial resolution. If the input image contains degraded pixels, the artifacts caused by the degradation could be amplified by super-resolution methods. Image blur is a common degradation source. Images captured by moving or still cameras are inevitably affected by motion blur due to relative movements between sensors and objects. In this work, we focus on the super-resolution task with the presence of motion blur. We propose a deep gated fusion convolution neural network to generate a clear high-resolution frame from a single natural image with severe blur. By decomposing the feature extraction step into two task-independent streams, the dual-branch design can facilitate the training process by avoiding learning the mixed degradation all-in-one and thus enhance the final high-resolution prediction results. Extensive experiments demonstrate that our method generates sharper super-resolved images from low-resolution inputs with high computational efficiency.

1 Introduction

Single-image super-resolution (SISR) aims to restore a high-resolution (HR) image from one low-resolution (LR) frame, such as those captured from surveillance and mobile cameras. The generated HR images can improve the performance of the numerous high-level vision tasks, e.g., object detection [43] and face recognition [47]. However, image degradation is inevitable during the imaging process, and would consequently harm the following high-level tasks. Image blur is a common image degradation in real-world scenarios, e.g., camera or object motion blur during the exposure time. Furthermore, the blur process becomes



(a) Blurry low-resolution input (b) Input patch (c) EDSR [20] (d) Ours (e) Input patch (f) DeepDeblur [21] (g) Ours

Figure 1: **Examples of joint image deblurring and super-resolution.** While the state-of-the-art SR algorithm [20] cannot reduce the non-uniform blur in the input image due to the simple assumption of the bicubic downsampling, the top-performing non-uniform deblurring algorithm [21] generates clear results but with few details. In contrast, the proposed method generates a clear and high-resolution image with more details.

more complicated due to the depth variations and occlusions around motion boundaries. Therefore, real-world images often undergo the effects of non-uniform motion blur.

The problems of SR and deblurring are often dealt separately, as each of them is known to be ill-posed. In this work, we aim at addressing the joint problem of generating clear HR images from given blurry LR inputs. Figure 1 shows one blurry LR image that is affected by non-uniform blur. As existing SR algorithms [12, 17, 18, 20] cannot reduce motion blur well, the resulting HR image is blurry (see Figure 1(b) and 1(c)). On the other hand, the state-of-the-art non-uniform deblurring methods [4, 16, 21, 23] generate clear images but cannot restore fine details and enlarge the spatial resolution (see Figure 1(e) and 1(f)).

With the advances of deep CNNs, state-of-the-art image super-resolution [17, 18, 20] and deblurring [16, 21] methods are built on end-to-end deep networks and achieve promising performance. To jointly solve the image deblurring and super-resolution problems, a straightforward approach is to solve the two problems sequentially, i.e., performing image deblurring followed by super-resolution, or vice versa. However, there are numerous issues with such approaches. First, a simple concatenation of two models is a sub-optimal solution due to error accumulation, i.e., the estimated error of the first model will be propagated and magnified in the second model. Second, the two-step network does not fully exploit the correlation between the two tasks. For example, the feature extraction and image reconstruction steps are performed twice and result in computational redundancy. As both the training and inference processes are memory and time-consuming, these approaches cannot be applied to resource-constrained applications. Several recent methods [40, 42, 45] jointly solve the image deblurring and super-resolution problems using end-to-end deep networks. However, these methods either focus on domain-specific applications, e.g., face and text [40, 42] images, or address uniform Gaussian blur [45]. In contrast, we aim to handle natural images degraded by non-uniform motion blur, which is more challenging.

In this work, we propose a Gated Fusion Network (GFN) based on a dual-branch architecture to tackle this complex joint problem. The proposed network consists of two branches:

one branch aims to deblur the LR input image, and the other branch generates a clear HR output image. In addition, we learn a gate module to adaptively fuse the features from the deblurring and super-resolution branches. We fuse the two branches on the feature level rather than the intensity level to avoid the error accumulation from imperfect deblurred images. The fused features are then fed into an image reconstruction module to generate clear HR output images. Extensive evaluations demonstrate that the proposed model performs favorably against the combination of the state-of-the-art image deblurring and super-resolution methods as well as existing joint models.

The contributions of this work are summarized as follows:

- To the best of our knowledge, this is the first CNN-based method to jointly solve non-uniform motion deblurring and generic single-image super-resolution.
- We decouple the joint problem into two sub-tasks for better regularization. We propose a dual-branch network to extract features for deblurring and super-resolution and learn a gate module to adaptively fuse the features for image reconstruction.
- The proposed model is efficient and entails low computational cost as most operations are executed in the LR space. Our model is much faster than the combination of the state-of-the-art SR and image deblurring models while achieving significant performance gain.

2 Related Work

We aim to solve the image super-resolution and non-uniform motion deblurring problems jointly in one algorithm. Both motion deblurring and image super-resolution are fundamental problems in computer vision. In this section, we focus our discussion on recent methods for image super-resolution and motion deblurring closest to this work.

Image super-resolution. Single image super-resolution (SISR) is an ill-posed problem as there are multiple HR images correspond to the same LR input image. Conventional approaches learn the LR-to-HR mappings using sparse dictionary [38], random forest [32] or self-similarity [9]. In recent years, CNN-based methods [3, 12] have demonstrated significant performance improvement against conventional SR approaches. Several techniques have been adopted, including the recursive learning [11, 37], pixel shuffling [20, 34], and Laplacian pyramid network [17]. In addition, some approaches use the adversarial loss [18], perceptual loss [10] and texture loss [29] to generate more photo-realistic SR images. As most SR algorithms assume that the LR images are generated by a simple downsampling kernel, e.g., bicubic kernel, they do not perform well when the input images suffer from unexpected motion blur. In contrast, the proposed model performs super-resolution on LR blurry images.

Motion deblurring. Conventional image deblurring approaches [2, 24, 30, 31, 33, 39] assume that the blur is uniform and spatially invariant across the entire image. However, due to the depth variation and object motion, real-world images typically contain non-uniform blur. Several approaches solve the non-uniform deblurring problem by jointly estimating blur kernels with scene depth [8, 26] or segmentation [13]. As the kernel estimation step is computationally expensive, recent methods [7, 21, 22, 23] learn deep CNNs to bypass the kernel estimation and efficiently solve the non-uniform deblurring problem. Orest et al. [16] adopt the Wasserstein generative adversarial network (GAN) to generate realistic deblurred

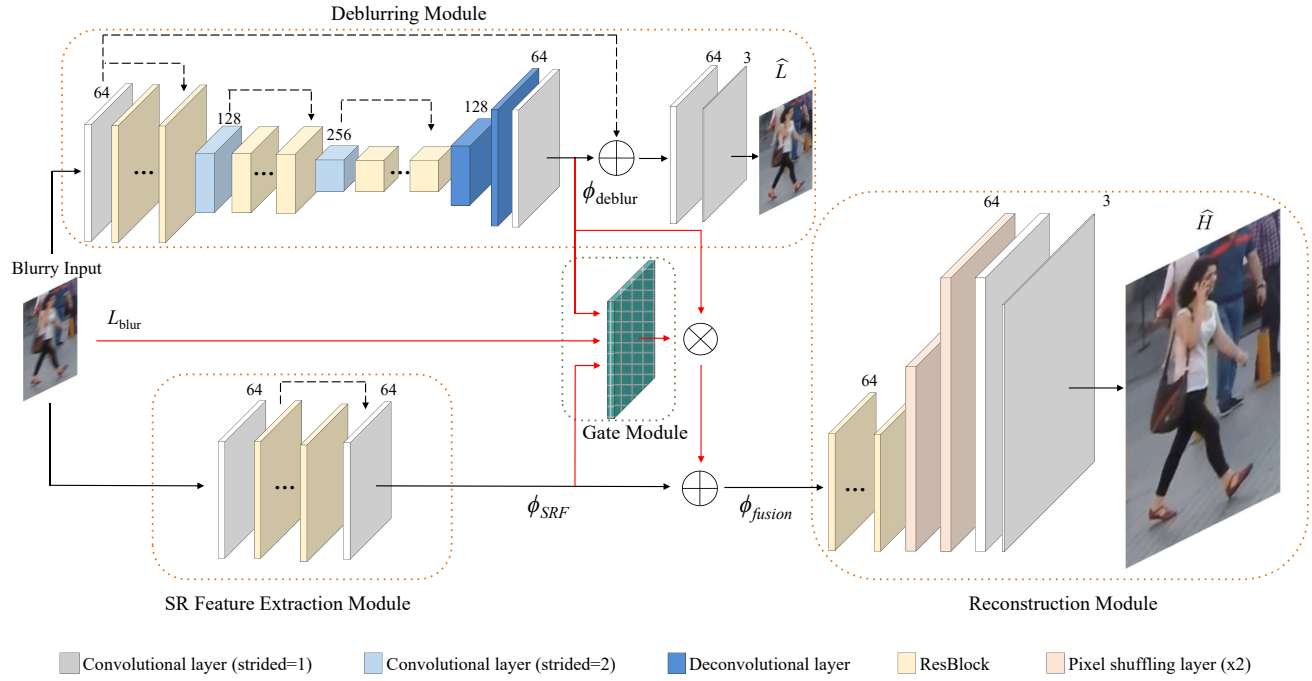


Figure 2: **Architecture of the proposed GFN.** Our model consists of four major modules: deblurring module G_{deblur}, SR feature extraction module G_{SRF}, gate module G_{gate}, and reconstruction module G_{recon}. The features extracted by G_{deblur} and G_{SRF} are fused by G_{gate} and then fed into G_{recon} to reconstruct the HR output image.

images and facilitate the object detection task. When applying to LR blurry images, the above approaches cannot produce sufficient details and upscale the spatial resolution.

Joint image super-resolution and motion deblurring. The joint problem is more challenging than the individual problem as the input images are extremely blurry and of low-resolution. Some approaches [1, 27, 41] reconstruct sharp and HR images from a LR and blurry video sequence. However, these methods rely on the optical flow estimation, which cannot be applied to the case where the input is a single image. Xu et al. [40] train a generative adversarial network to super-resolve blurry face and text images. As face and text images have relative small visual entropy, i.e., high patch recurrence within an image [35], a small model can restore these category-specific images well. Our experimental results show that the model of Xu et al. [40] does not have sufficient capacity to handle natural images with non-uniform blur. Zhang et al. [45] propose a deep encoder-decoder network (ED-DSRN) for joint image deblurring and super-resolution. However, they focus on LR images degraded by uniform Gaussian blur. In contrast, our model can handle complex non-uniform blur and use much fewer parameters to effectively restore clear HR images.

3 Gated Fusion Network

In this section, we describe the architecture design, training loss functions, and implementation details of the proposed GFN for joint image deblurring and super-resolution.

3.1 Network architecture

Given a blurry LR image L_{blur} as the input, our goal is to recover a sharp HR image $\widehat{H}$. In this work, we consider the case of $4\times$ SR, i.e., the spatial resolution of L_{blur} is $4\times$ smaller than that of $\widehat{H}$. The proposed model is based on the dual-branch architecture [5, 19, 25, 44, 46] and consists of four major modules: (i) a deblurring module G_{deblur} for extracting deblurring features and predicting a sharp LR image $\widehat{L}$, (ii) an SR feature extraction module G_{SRF} to extract features for image super-resolution, (iii) a gate module G_{gate} to estimate a weight map for blending the deblurring and super-resolution features, and (iv) a reconstruction module G_{recon} to reconstruct the final HR output image. An overview of the proposed model is illustrated in Figure 2.

Debluring module. The deblurring module aims to restore a sharp LR image $\widehat{L}$ from the input blurry LR image L_{blur}. Different from the encoder-decoder architecture in [36], we adopt an asymmetric residual encoder-decoder architecture in our deblurring module to enlarge the receptive field. The encoder consists of three scales, where each scale has six ResBlocks [20] followed by a strided convolutional layer to downsample the feature maps by $1/2\times$. The decoder has two deconvolutional layers to enlarge the spatial resolution of feature maps. Finally, we use two additional convolutional layers to reconstruct a sharp LR image $\widehat{L}$. We denote the output features of the decoder by ϕ_{deblur}, which are later fed into the gate module for feature fusion.

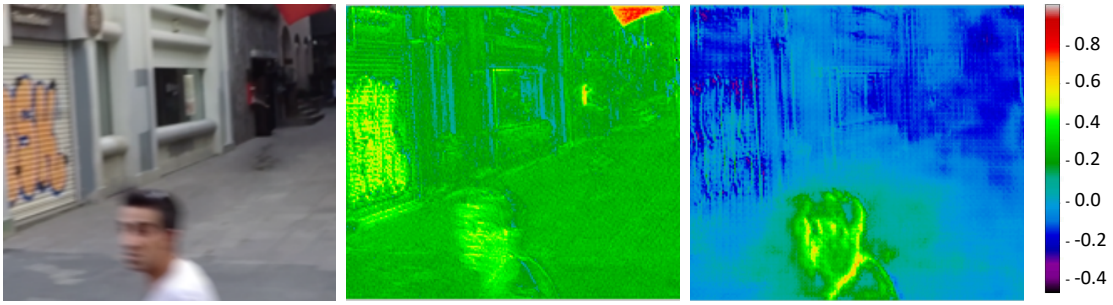
Super-resolution feature extraction module. We use eight ResBlocks [20] to extract high-dimensional features for image super-resolution. To maintain the spatial information, we do not use any pooling or strided convolutional layers. We denote the extracted features as ϕ_{SRF}.

Gate module. In Figure 3, we visualize the responses of ϕ_{deblur} and ϕ_{SRF}. While the SR features contain spatial details of the input image (as shown on the wall of Figure 3(b)), the deblurring features have high response on regions of large motion (as shown on the moving person of Figure 3(c)). Thus, the responses of ϕ_{deblur} and ϕ_{SRF} complement each other, especially on blurry regions. Inspired by the fact that gate structures can be used to discover feature importance for multimodal fusion [6, 28], we propose to learn a gate module to adaptively fuse the two features. Unlike [28] that predicts the weight maps of three derived inputs, our gate module enables local and contextual feature fusion from two independent branches, and thus it is expected to selectively merge sharp features ϕ_{fusion} from ϕ_{deblur} and ϕ_{SRF} in this work. Our gate module, G_{gate}, takes as input the set of ϕ_{deblur}, ϕ_{SRF}, and the LR blurry image L_{blur}, and generate a pixel-wise weight maps for blending ϕ_{deblur} and ϕ_{SRF}:

$$\phi_{fusion} = G_{gate}(\phi_{SRF}, \phi_{deblur}, L_{blur}) \otimes \phi_{deblur} + \phi_{SRF}, \tag{1}$$

where $\otimes$ denotes the element-wise multiplication. The gate module consists of two convolutional layers with the filter size of 3×3 and 1×1, respectively.

Reconstruction module. The fused features ϕ_{fusion} from the gate module are fed into eight ResBlocks and two pixel-shuffling layers [34] to enlarge the spatial resolution by $4\times$. We then use two final convolutional layers to reconstruct an HR output image $\widehat{H}$. We note that most of the network operations are performed in the LR feature space. Therefore, the proposed model entails low computational cost in both training and inference phases.



(a) Input image (b) ϕ_{SRF} (c) ϕ_{deblur}

Figure 3: **Feature response of the super-resolution features ϕ_{SRF} and the deblurring features ϕ_{deblur}.** While the super-resolution features contain more spatial details, the deblurring features have higher response on large motion.

3.2 Loss functions

The proposed network generates two output images: a deblurred LR image $\widehat{L}$ and a sharp HR image $\widehat{H}$. In our training data, each blurry LR image L_{blur} has a corresponding ground truth HR image H and a ground truth LR image L, which is the bicubic downsampled image of H. Therefore, we train our network by jointly optimizing a super-resolution loss and a deblurring loss:

$$\min \ \mathcal{L}_{SR}(\widehat{H},H)+\alpha\mathcal{L}_{deblur}(\widehat{L},L), \tag{2}$$

where α is a weight to balance the two loss terms. We use the pixel-wise MSE loss function for both $\mathcal{L}_{SR}$ and $\mathcal{L}_{deblur}$, and empirically set $\alpha = 0.5$.

3.3 Implementation and training details

In the proposed network, the filter size is set as 7×7 in the first and the last convolutional layers, 4×4 in the deconvolutional layers, and 3×3 in all other convolutional layers. We initialize all learnable parameters from a Gaussian distribution. We use the leaky rectified linear unit (LReLU) with a negative slope of 0.2 as the activation function. As suggested in [20], we do not use any batch normalization layers in order to retain the range flexibility of features. More implementation details of the network can be found in the supplementary due to space limitation.

To facilitate the network training, we adopt a two-step training strategy. First, we disable the gate module and fuse the deblurring feature and SR feature directly. We train the deblurring/super-resolution/reconstruction modules from scratch for 60 epochs. We initialize the learning rate to 1e-4 and multiply by 0.1 for every 30 epochs. Second, we include the gate module and train the entire network for another 50 epochs. The learning rate is set to 5e-5 and multiplied by 0.1 for every 25 epochs. We use a batch size of 16 and augment the training data by random rotation and horizontal flipping. We use the ADAM solver [14] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ to optimize our network. The training process takes about 4 days on an NVIDIA 1080Ti GPU card. The source code can be found at https://github.com/jacquelinelala/GFN .

4 Experimental Results

In this section, we first describe the dataset for training our network, present quantitative and qualitative comparisons with state-of-the-art approaches, and finally analyze and discuss several design choices of the proposed model.

4.1 Training dataset

We use the GOPRO dataset [21] to generate the training data for the joint SR and deblurring problem. The GOPRO dataset contains 2103 blurry and sharp HR image pairs for training. To generate more training data, we resize each HR image pair with three random scales within $[0.5, 1.0]$. We then crop HR image pairs into 256×256 patches with a stride of 128 to obtain blurry HR patches H_{blur} and sharp HR patches H. Finally, we use the bicubic downsampling to downscale H_{blur} and H by $4\times$ to generate blurry LR patches L_{blur} and sharp LR patches L, respectively. In total, we obtain 10,7584 triplets of $\{L_{blur}, L, H\}$ for training (the blurry HR patches H_{blur} are discarded during training). We refer the generated dataset as LR-GOPRO in the following content.

4.2 Comparisons with state-of-the-arts

We compare the proposed GFN with several algorithms, including the state-of-the-art SR methods [18, 20], the joint image deblurring and SR approaches [40, 45], and the combinations of SR algorithms [18, 20] and non-uniform deblurring algorithms [16, 21]. For fair comparisons, we re-train the models of SCGAN [40], SRResNet [18], and ED-DSRN [45] on our training dataset.

We use the bicubic downsampling to generate blurry LR images from the test set of the GOPRO dataset [21] and the Köhler dataset [15] for evaluation. Table 1 shows the quantitative evaluation in terms of PSNR, SSIM, and average inference time. The tradeoff between image quality and efficiency is better visualized in Figure 4. The proposed GFN performs favorably against existing methods on both datasets and maintains a low computational cost and execution time. While the re-trained SCGAN and SRResNet perform better than their pre-trained models, both methods cannot handle the complex non-uniform blur well due to their small model capacity. The ED-DSRN method uses more network parameters to achieve decent performance. However, the single-branch architecture of ED-DSRN is less effective than the proposed dual-branch network.

The concatenations of SR [18, 20] and deblurring methods [16, 21] are typically less effective due to the error accumulation. We note that the concatenations using the SR-first strategy, i.e., performing SR followed by image deblurring, typically have better performance than that using the DB-first strategy, i.e., performing image deblurring followed by SR. However, the SR-first strategy entails heavy computational cost as the image deblurring step is performed in the HR image space. Compared with the best-performing combination of EDSR [20] and DeepDeblur [21] methods, the proposed GFN is $116\times$ faster and has 78% fewer model parameters.

We present the qualitative results on the LR-GOPRO dataset in Figure 5 and the results on a real blurry image in Figure 6. The methods using the concatenation scheme, e.g., DeepDeblur [21] + EDSR [20] and EDSR [20] + DeepDeblur [21], often introduce undesired artifacts due to the error accumulation problem. Existing joint SR and deblurring meth-

Table 1: **Quantitative comparison with state-of-the-art methods.** The methods with a $\star$ sign are trained on our LR-GOPRO training set. Red texts indicate the best performance. The proposed GFN performs favorably against existing methods while maintaining a small model size and fast inference speed.

Method	#Params	LR-GOPRO 4× PSNR / SSIM / Time (s)	LR-Köhler 4× PSNR / SSIM / Time (s)
SCGAN [40]	1.1M	22.74 / 0.783 / 0.66	23.19 / 0.763 / 0.45
SRResNet [18]	1.5M	24.40 / 0.827 / 0.07	24.81 / 0.781 / 0.05
EDSR [20]	43M	24.52 / 0.836 / 2.10	24.86 / 0.782 / 1.43
SCGAN* [40]	1.1M	24.88 / 0.836 / 0.66	24.82 / 0.795 / 0.45
SRResNet* [18]	1.5M	26.20 / 0.818 / 0.07	25.36 / 0.803 / 0.05
ED-DSRN* [45]	25M	26.44 / 0.873 / 0.10	25.17 / 0.799 / 0.08
DB [21] + SR [18]	13M	24.99 / 0.827 / 0.66	25.12 / 0.800 / 0.55
SR [18] + DB [21]	13M	25.93 / 0.850 / 6.06	25.15 / 0.792 / 4.18
DB [16] + SR [18]	13M	21.71 / 0.686 / 0.14	21.10 / 0.628 / 0.12
SR [18] + DB [16]	13M	24.44 / 0.807 / 0.91	24.92 / 0.778 / 0.54
DB [16] + SR [20]	54M	21.53 / 0.682 / 2.18	20.74 / 0.625 / 1.57
SR [20] + DB [16]	54M	24.66 / 0.827 / 2.95	25.00 / 0.784 / 1.92
DB [21] + SR [20]	54M	25.09 / 0.834 / 2.70	25.16 / 0.801 / 2.04
SR [20] + DB [21]	54M	26.35 / 0.869 / 8.10	25.24 / 0.795 / 5.81
GFN (ours)	12M	27.74 / 0.896 / 0.07	25.72 / 0.813 / 0.05

Table 2: **Analysis on key components in the proposed GFN.** All the models are trained on the LR-GOPRO dataset with the same hyper-parameters.

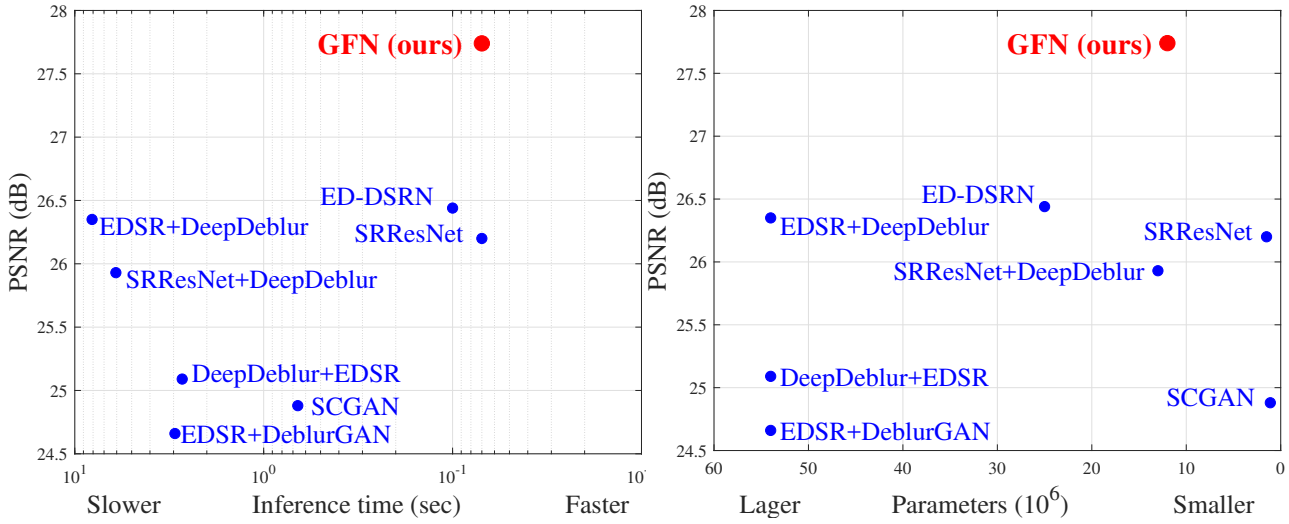
Modifications	SRResNet	Model-1	Model-2	Model-3	Model-4	Model-5	GFN
deblurring loss	✓	✓		✓		✓	✓
deblurring module		✓	✓	✓	✓	✓	✓
dual-branch				✓	✓	✓	✓
feature level					✓	✓	✓
gate module							✓
SR-first			✓				
PSNR	26.20	26.82	27.00	27.01	27.39	27.53	27.74
Time (s)	0.05	0.09	0.57	0.10	0.05	0.05	0.05

ods [18, 40, 45] cannot reduce the non-uniform blur well. In contrast, the proposed method generates clear HR images with more details.

4.3 Analysis and discussion

The proposed GFN consists of several key components: (1) using a deblurring loss for regularization; (2) using a deblurring module to extract the deblurring features; (3) using a dual-branch architecture instead of a concatenation of SR and deblurring models; (4) fusing features in the feature level instead of the intensity level; (5) using a gate module to adaptive fuse SR and deblurring features. The training schemes of these models are presented in the supplementary material. Here we discuss the performance contribution of these components.

We first train a baseline model by introducing a deblurring loss to the SRResNet. We then include the deblurring module to the baseline model using two sequential strategies: DB-first (Model-1) and SR-first (Model-2). In Table 2, the deblurring module shows significant performance improvement over the baseline SRResNet model. The SR-first model achieves better performance but has a slower execution speed. However, the concatenation of SR and



(a) PSNR vs. inference time (b) PSNR vs. number of parameters

Figure 4: **Performance versus inference time and model parameters.** The results are evaluated on the LR-GOPRO dataset.

deblurring modules might be sub-optimal due to the error accumulation. Next, we employ the dual-branch architecture and fuse the features in the intensity level (Model-3) and the feature level (Model-5). Compared with Model-3, Model-5 achieves 0.5 dB performance improvement and runs faster. We also train a model without the deblurring loss (Model-4). Experimental results show that the deblurring loss helps the network converge faster and obtain improved performance (0.14 dB). Finally, we show that the learned gate module can further boost the performance by 0.2 dB.

5 Conclusions

In this paper, we propose an efficient end-to-end network to recover sharp high-resolution images from blurry low-resolution input. The proposed network consists of two branches to extract blurry and sharp features more effectively. The extracted features are fused through a gate module, and are then used to reconstruct the final results. The network design decouples the joint problem into two restoration tasks, and enables efficient training and inference. Extensive evaluations show that the proposed algorithm performs favorably against the state-of-the-art methods in terms of visual quality and runtime.

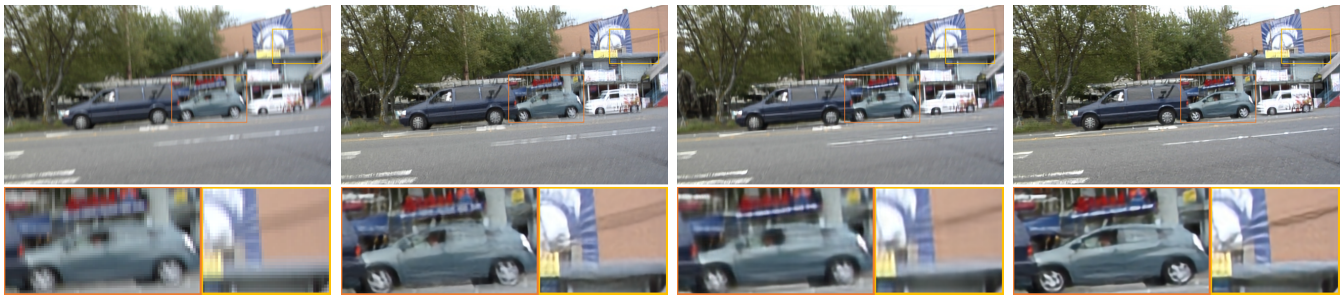
Acknowledgement

This work is partially supported by National Science and Technology Major Project (No. 2018ZX01008103), NSF CARRER (No.1149783), and gifts from Adobe and Nvidia.

(a) Ground-truth HR PSNR / SSIM | (b) Blurry LR input 21.04 / 0.787 | (c) DB [21] + SR [20] 24.58 / 0.846 | (d) SR [20] + DB [21] 25.04 / 0.876

(e) SCGAN [40] 23.00 / 0.835 | (f) ED-DSRN [45] 26.07 / 0.896 | (g) SRResNet[18] 25.63 / 0.881 | (h) GFN (ours) 29.02 / 0.929

Figure 5: **Visual comparison on the LR-GOPRO dataset.** The proposed method generates sharp HR images with more details.



(a) Blurry LR input | (b) ED-DSRN [45] | (c) SRResNet[18] | (d) GFN (ours)

Figure 6: **Visual comparison on the real blurry image dataset [36].** Our GFN is more robust to outliers in real images and generates a sharper result than state-of-the-art methods [18, 45].

References

[1] Benedicte Bascle, Andrew Blake, and Andrew Zisserman. Motion deblurring and super-resolution from an image sequence. In ECCV, 1996.

[2] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. ACM TOG, 28(5):145, 2009.

[3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. TPAMI, 38(2):295–307, 2016.

[4] Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian Reid, Chunhua Shen, AVD Hengel, and Qinfeng Shi. From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur. In CVPR, 2017.

[5] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In ICCV, 2017.

[6] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.

[7] Michal Hradiš, Jan Kotera, Pavel Zemcík, and Filip Šroubek. Convolutional neural networks for direct text deblurring. In BMVC, 2015.

[8] Zhe Hu, Li Xu, and Ming-Hsuan Yang. Joint depth estimation and camera shake removal from single blurry image. In CVPR, 2014.

[9] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In CVPR, 2015.

[10] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In ECCV, 2016.

[11] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In CVPR, 2016.

[12] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In CVPR, 2016.

[13] Tae Hyun Kim, Byeongjoo Ahn, and Kyoung Mu Lee. Dynamic scene deblurring. In ICCV, 2013.

[14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In ICLR, 2015.

[15] Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In ECCV, 2012.

[16] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. DeblurGAN: Blind motion deblurring using conditional adversarial networks. In CVPR, 2018.

[17] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In CVPR, 2017.

[18] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In CVPR, 2017.

[19] Yijun Li, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep joint image filtering. In ECCV, 2016.

[20] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In CVPR Workshops, 2017.

[21] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In CVPR, 2017.

[22] T M Nimisha, Akash K Singh, and Ambasamudram N Rajagopalan. Blur-invariant deep learning for blind-deblurring. In ICCV, 2017.

[23] Mehdi Noroozi, Paramanand Chandramouli, and Paolo Favaro. Motion deblurring in the wild. In GCPR, 2017.

[24] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In CVPR, 2016.

[25] Jinshan Pan, Sifei Liu, Deqin Sun, Jiawei Zhang, Yang Liu, Jimmy Ren, Zechao Li, Jinhui Tang, Huchuan Lu, Yu-Wing Tai, and Ming-Hsuan Yang. Learning dual convolutional neural networks for low-level vision. In CVPR, 2018.

[26] Chandramouli Paramanand and Ambasamudram N Rajagopalan. Non-uniform motion deblurring for bilayer scenes. In CVPR, 2013.

[27] Haesol Park and Kyoung Mu Lee. Joint estimation of camera pose, depth, deblurring, and super-resolution from a blurred image sequence. In ICCV, 2017.

[28] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In CVPR, 2018.

[29] Mehdi SM Sajjadi, Bernhard Schölkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In ICCV, 2017.

[30] Uwe Schmidt, Kevin Schelten, and Stefan Roth. Bayesian deblurring with integrated noise estimation. In CVPR, 2011.

[31] Uwe Schmidt, Carsten Rother, Sebastian Nowozin, Jeremy Jancsary, and Stefan Roth. Discriminative non-blind deblurring. In CVPR, 2013.

[32] Samuel Schulter, Christian Leistner, and Horst Bischof. Fast and accurate image upscaling with super-resolution forests. In CVPR, 2015.

[33] Qi Shan, Jiaya Jia, and Aseem Agarwala. High-quality motion deblurring from a single image. In ACM TOG, volume 27, page 73. ACM, 2008.

[34] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In CVPR, 2016.

[35] Assaf Shocher, Nadav Cohen, and Michal Irani. "Zero-Shot" super-resolution using deep internal learning. In CVPR, 2018.

[36] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring. In CVPR, 2017.

[37] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In CVPR, 2017.

[38] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In ACCV, 2014.

[39] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural l0 sparse representation for natural image deblurring. In CVPR, 2013.

[40] Xiangyu Xu, Deqing Sun, Jinshan Pan, Yujin Zhang, Hanspeter Pfister, and Ming-Hsuan Yang. Learning to super-resolve blurry face and text images. In ICCV, 2017.

[41] Takuma Yamaguchi, Hisato Fukuda, Ryo Furukawa, Hiroshi Kawasaki, and Peter Sturm. Video deblurring and super-resolution technique for multiple moving objects. In ACCV, 2010.

[42] Xin Yu, Basura Fernando, Richard Hartley, and Fatih Porikli. Super-resolving very low-resolution face images with supplementary attributes. In CVPR, 2018.

[43] Haichao Zhang, Jianchao Yang, Yanning Zhang, Nasser M Nasrabadi, and Thomas S Huang. Close the loop: Joint blind image restoration and recognition with sparse representation prior. In ICCV, 2011.

[44] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Joint transmission map estimation and dehazing using deep networks. arXiv preprint arXiv:1708.00581, 2017.

[45] Xinyi Zhang, Fei Wang, Hang Dong, and Yu Guo. A deep encoder-decoder networks for joint deblurring and super-resolution. In ICASSP, 2018.

[46] Shizhan Zhu, Sifei Liu, Chen Change Loy, and Xiaoou Tang. Deep cascaded bi-network for face hallucination. In ECCV, 2016.

[47] Wilman WW Zou and Pong C Yuen. Very low resolution face recognition problem. TIP, 21(1):327–340, 2012.