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

In this supplementary document, we provide additional information and experimental results to complement the paper. First, we provide the detailed configuration of the proposed gated fusion network (GFN). Second, we show the variations of the proposed model we compared in Section 4.3 of the manuscript. Finally, we present more qualitative comparisons with the state-of-the-art methods.

A Network Configuration

We present the detailed configuration of the proposed network in Table 1, with respect to the four modules in the network: **deblurring module**, **SR feature extraction module**, **gate module**, and **reconstruction module**.

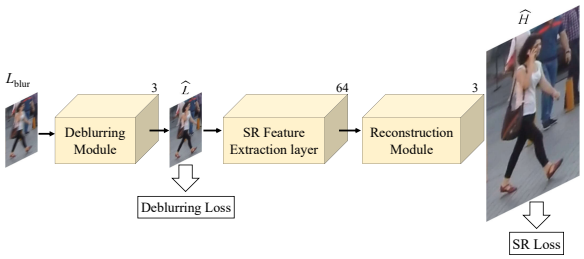
B Ablation Study

In Section 4.3 of the manuscript, we analyze the key components of the proposed method and conduct an ablation study. As shown in Figure 1, we illustrate the network schemes of the models in the ablation study. For the models using sequential strategies, i.e., Model-1 and Model-2, we first pre-train the deblurring and super-resolution modules separately and then finetune them on the LR-GOPRO dataset. All the other models are trained from scratch on the LR-GOPRO dataset using the same hyper-parameters.

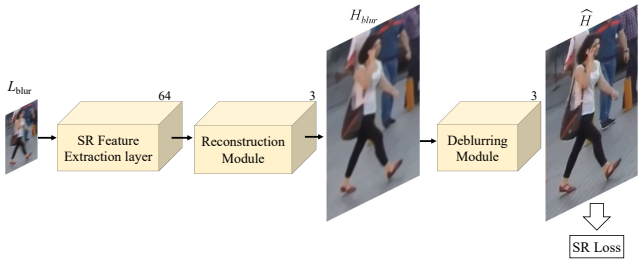
Table 1: **Configuration of the proposed network.** The values in the **skip** row are **layer** names, indicating whose outputs are added to the outputs of the corresponding layers.

Deblurring Module				
layer	**output size**	**kernel**	**LReLU**	**skip**
input_1	$3 \times h \times w$			
conv1	$64 \times h \times w$	7		
Resblock 1-6	$64 \times h \times w$	3		conv1
conv2	$128 \times \frac{h}{2} \times \frac{w}{2}$	3		
Resblock 7-12	$128 \times \frac{h}{2} \times \frac{w}{2}$	3		conv2
conv3	$256 \times \frac{h}{4} \times \frac{w}{4}$	3		
Resblock 13-18	$256 \times \frac{h}{4} \times \frac{w}{4}$	3		conv3
deconv1	$128 \times \frac{h}{2} \times \frac{w}{2}$	4	✓	
deconv2	$64 \times h \times w$	4	✓	
conv4	$64 \times h \times w$	7		conv1
conv5	$64 \times h \times w$	3	✓	
conv6	$3 \times h \times w$	3		

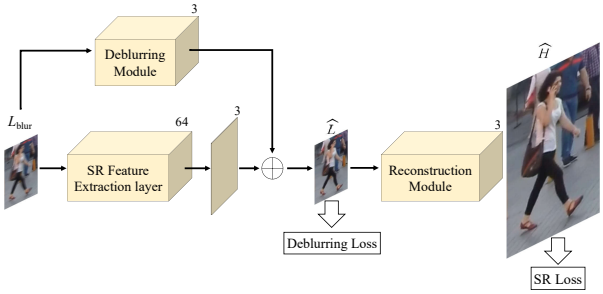
SR Feature Extraction Module				
layer	**output size**	**kernel**	**LReLU**	**skip**
input_1	$3 \times h \times w$			
conv7	$64 \times h \times w$	7		
Resblock 19-26	$64 \times h \times w$	3		
conv8	$64 \times h \times w$	3		conv7
Gate Module				
input_2	$131 \times h \times w$			
conv9	$64 \times h \times w$	3	✓	
conv10	$64 \times h \times w$	1		
element mul	$64 \times h \times w$			conv8
Reconstruction Module				
input_3	$64 \times h \times w$			
Resblock 27-34	$64 \times h \times w$	3		
conv11	$256 \times h \times w$	3		
pixel shuffle	$64 \times 2h \times 2w$		✓	
conv12	$256 \times 2h \times 2w$	3		
pixel shuffle	$64 \times 4h \times 4w$		✓	
conv13	$64 \times 4h \times 4w$	3	✓	
conv14	$3 \times 4h \times 4w$	3		



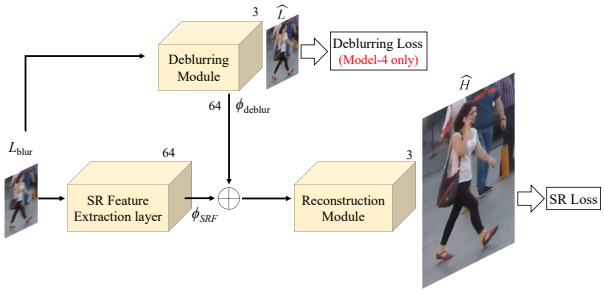
(a) Model-1



(b) Model-2



(c) Model-3

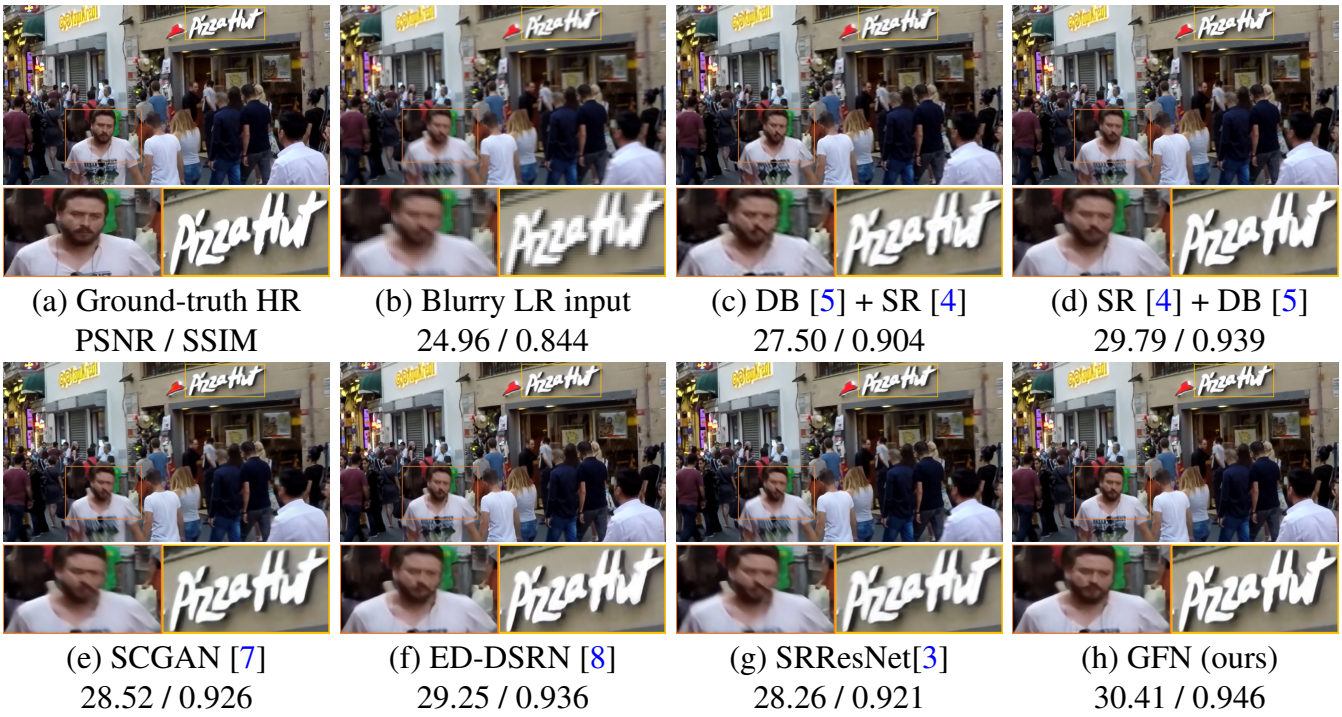


(d) Model-4 & 5

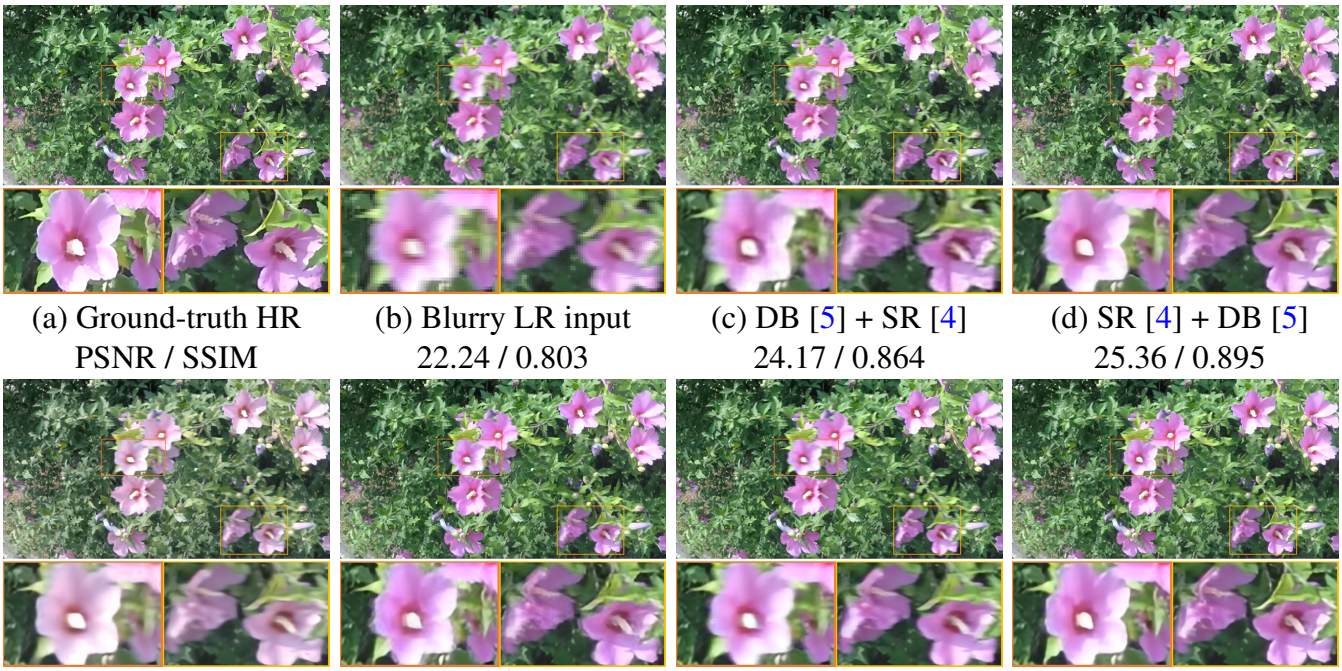
Figure 1: **Network schemes of the models in the ablation study of the manuscript.**

C Qualitative Comparisons

In this section, we present more qualitative comparisons with the state-of-the-art SR methods [3, 4], the joint image deblurring and SR approaches [7, 8], and the combinations of SR algorithms [3, 4] and non-uniform deblurring algorithms [2, 5], on the LR-GOPRO, LR-Köhler dataset [1], and real dataset [6].

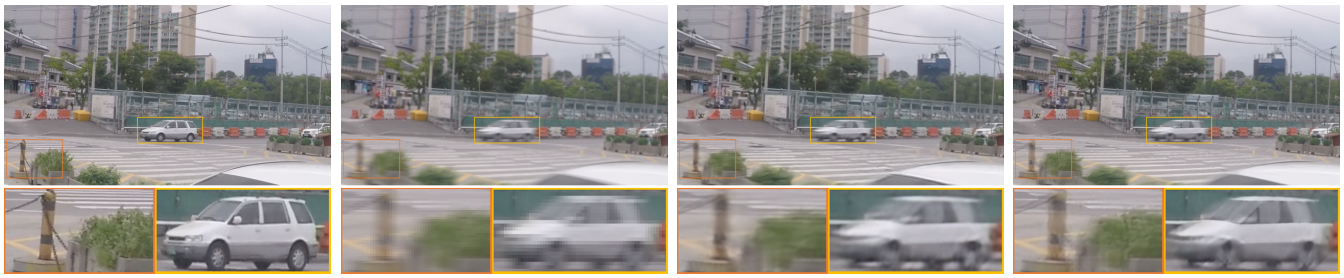


(a) Ground-truth HR PSNR / SSIM
(b) Blurry LR input 24.96 / 0.844
(c) DB [5] + SR [4] 27.50 / 0.904
(d) SR [4] + DB [5] 29.79 / 0.939
(e) SCGAN [7] 28.52 / 0.926
(f) ED-DSRN [8] 29.25 / 0.936
(g) SRResNet[3] 28.26 / 0.921
(h) GFN (ours) 30.41 / 0.946



(a) Ground-truth HR PSNR / SSIM
(b) Blurry LR input 22.24 / 0.803
(c) DB [5] + SR [4] 24.17 / 0.864
(d) SR [4] + DB [5] 25.36 / 0.895
(e) SCGAN [7] 23.13 / 0.820
(f) ED-DSRN [8] 24.93 / 0.887
(g) SRResNet[3] 24.63 / 0.882
(h) GFN (ours) 25.64 / 0.898

Figure 2: **Visual comparisons on the LR-GOPRO dataset.**

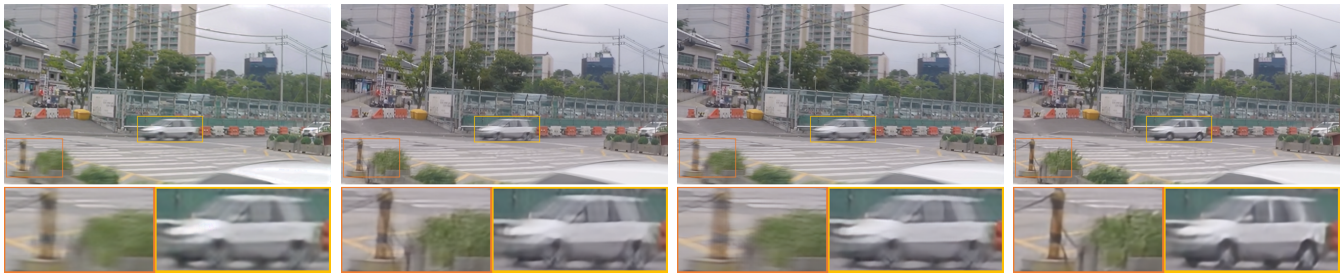


(a) Ground-truth HR
PSNR / SSIM

(b) Blurry LR input
26.60 / 0.768

(c) DB [5] + SR [4]
28.22 / 0.831

(d) SR [4] + DB [5]
28.49 / 0.852



(e) SCGAN [7]
27.83 / 0.827

(f) ED-DSRN [8]
28.18 / 0.845

(g) SRResNet[3]
28.16 / 0.840

(h) GFN (ours)
29.23 / 0.862

(a) Ground-truth HR
PSNR / SSIM

(b) Blurry LR input
24.24 / 0.738

(c) DB [5] + SR [4]
25.14 / 0.764

(d) SR [4] + DB [5]
27.26 / 0.836

(e) SCGAN [7]
25.28 / 0.775

(f) ED-DSRN [8]
26.82 / 0.826

(g) SRResNet[3]
26.76 / 0.828

(h) GFN (ours)
28.55 / 0.863

Figure 3: **Visual comparisons on the LR-GOPRO dataset.**

(a) Ground-truth HR PSNR / SSIM | (b) Blurry LR input 20.12 / 0.645 | (c) DB [5] + SR [4] 21.99 / 0.699 | (d) SR [4] + DB [5] 23.33 / 0.765

(e) SCGAN [7] 20.57 / 0.700 | (f) ED-DSRN [8] 22.17 / 0.738 | (g) SRResNet[3] 23.97 / 0.779 | (h) GFN (ours) 25.39 / 0.828

(a) Ground-truth HR PSNR / SSIM | (b) Blurry LR input 22.60 / 0.744 | (c) DB [5] + SR [4] 23.96 / 0.781 | (d) SR [4] + DB [5] 26.24 / 0.857

(e) SCGAN [7] 23.40 / 0.791 | (f) ED-DSRN [8] 26.05 / 0.850 | (g) SRResNet[3] 26.27 / 0.858 | (h) GFN (ours) 28.05 / 0.888

Figure 4: **Visual comparisons on the LR-GOPRO dataset.**

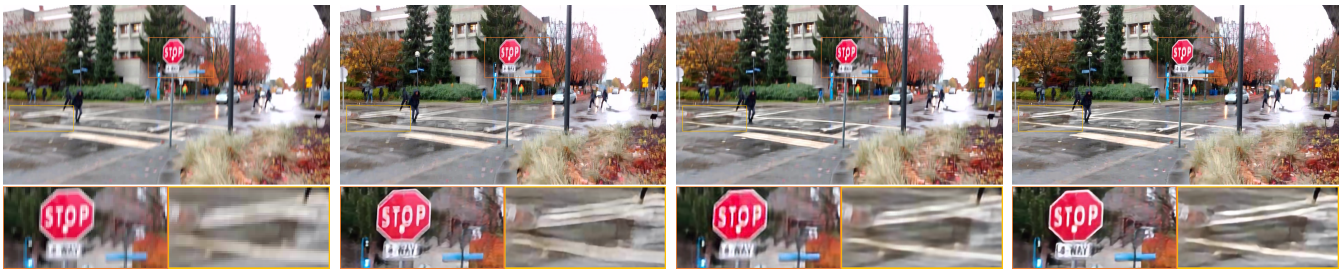
(a) Blurry LR input (b) ED-DSRN [8] (c) SRResNet[3] (d) GFN (ours)

(a) Blurry LR input (b) ED-DSRN [8] (c) SRResNet[3] (d) GFN (ours)

(a) Blurry LR input (b) ED-DSRN [8] (c) SRResNet[3] (d) GFN (ours)

Figure 5: **Visual comparisons on the LR-Köhler dataset [1].**

(a) Blurry LR input (b) ED-DSRN [8] (c) SRResNet[3] (d) GFN (ours)



(a) Blurry LR input (b) ED-DSRN [8] (c) SRResNet[3] (d) GFN (ours)

(a) Blurry LR input (b) ED-DSRN [8] (c) SRResNet[3] (d) GFN (ours)

(a) Blurry LR input (b) ED-DSRN [8] (c) SRResNet[3] (d) GFN (ours)

Figure 6: **Visual comparisons on the real blurry dataset [6].**

References

[1] 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. 3, 6

[2] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. DeblurGAN: Blind motion deblurring using conditional adversarial networks. In CVPR, 2018. 3

[3] 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. 3, 4, 5, 6, 7

[4] 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. 3, 4, 5

[5] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In CVPR, 2017. 3, 4, 5

[6] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring. In CVPR, 2017. 3, 7

[7] 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. 3, 4, 5

[8] Xinyi Zhang, Fei Wang, Hang Dong, and Yu Guo. A deep encoder-decoder networks for joint deblurring and super-resolution. In ICASSP, 2018. 3, 4, 5, 6, 7