Supplementary Material
Stripformer: Strip Transformer for Fast Image Deblurring

Fu-Jen Tsai[1], Yan-Tsung Peng[2], Yen-Yu Lin[3], Chung-Chi Tsai[4], and Chia-Wen Lin[1]

[1] National Tsing Hua University, Taiwan
cwlin@ee.nthu.edu.tw, fjtsai@gapp.nthu.edu.tw
[2] National Chengchi University, Taiwan
ytpeng@cs.nccu.edu.tw
[3] National Yang Ming Chiao Tung University, Taiwan
lin@cs.nycu.edu.tw
[4] Qualcomm Technologies, Inc., San Diego
chuntsai@qti.qualcomm.com

This document provides additional materials to supplement our main submission. First, we further analyze the effectiveness of the proposed strip-wise attention regarding its directions. Next, we demonstrate the effect of pre-training models with a larger dataset. Lastly, we show more deblurring results and comparisons on real-world blurred scenes.

1 Further Analysis on strip-wise attention

We propose intra/inter-strip attention (SA) to deal with motion blur, which decomposes blur motions into horizontal and vertical directions to catch blur orientations and magnitudes by strip-wise tokens. Figure A demonstrates some attention maps obtained with the last horizontal and vertical SA after multiple interlaced Intra-SA and Inter-SA blocks for two blurred images. Here, the last horizontal and vertical SA refer to Intra-SA-H and Intra-SA-V of the last stacked interlaced Intra-SA and Inter-SA blocks, as shown in Figure 3. Generally, the maps in Figure A show that larger values represent largely blurred regions while smaller for sharper regions.

Besides, we conduct an additional ablation on only horizontal or vertical strips used to analyze them further. Table A shows that Stripformer with both horizontal and vertical strips performs better than using single directional strips alone, which again attests to the effectiveness of our design for deblurring. We did not test the setting with no intra-SA and inter-SA since it would reduce our model to a simple CNN architecture, which has proven insufficient to achieve comparable deblurring performances in the literature.

2 Pre-training Data Analysis.

Compared to IPT [1], which utilizes ImageNet (1M images) as pre-training data to achieve comparable performance to SOTA methods. Our Stripformer can

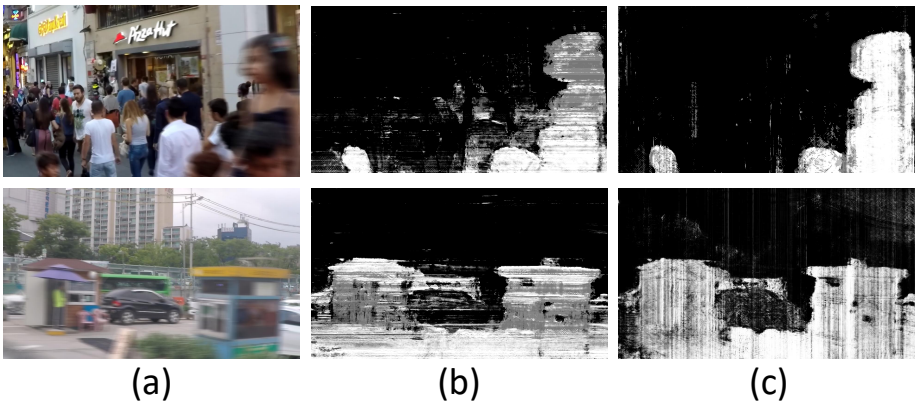

Fig. A. Attention maps obtained after multiple interlaced Intra-SA and Inter-SA blocks (referred to Figure 3). (a) Input blurred images on GoPro test set, and their attention maps obtained from (b) horizontal SA and (c) vertical SA.

Table A. Ablation on the directions for Intra-SA/Inter-SA on GoPro test set.

Strip tokens	Horizontal	Vertical	Both
GoPro (PSNR)	32.84	32.95	33.08

be trained on only GoPro (with 2,103 images) to outperform previous SOTA approaches. To further leverage the pre-training data, we utilize the REDS dataset [4] (with 24,000 images) as the pre-training data and then fine-tune the model on the GoPro training set. Stripformer trained only on the GoPro training set can exceed IPT by 0.5dB in PSNR, and with the help of REDS pre-training data, our method denoted as Stripformer* can outperform IPT by 1.13 dB (achieving 33.71 on the GoPro test set), as shown in Table B. Figure B shows qualitative comparisons among our method and the SOTA methods [2,7], where Stripformer produces sharper deblurred images than the compared SOTA methods. Moreover, Stripformer* presents even cleaner deblurred results than Stripformer. Note that IPT [1]'s results were not provided since the work did not release the code and deblurred results but quantitative scores, reported in Table B.

3 Image Deblurring on Real-world Scenes

To demonstrate more how the proposed method performs on real-world blurry images, we compare it with state-of-the-art deblurring approaches, including DeblurGAN-v2 [3], SRN [6], MPRNet [7], and MIMO [2] on two real-world blur datasets, RealBlur-J [5] testing set and RWBI [8]. All the methods are trained on the RealBlur-J training set.

Fig. B. Qualitative comparisons on the GoPro test set. The deblurred results from left to right are produced by MIMO [2], MPRNet [7], Stripformer, Stripformer*. MIMO, MPRNet, and Stripformer are trained on the GoPro training set. Stripformer* is pre-trained on the REDS dataset and then fine-tuned on the GoPro training set. The last column shows the ground truth.

Table B. Ablation study on the amount of pre-training data used. The listed models are IPT [1], Stripformer, and Stripformer*, where IPT uses ImageNet (with 1M images) to pre-train, Stripformer uses only the GoPro training set (with 2,103 images), and Stripformer* is pre-trained on the REDS dataset (with 24,000 images) and then fine-tuned on the GoPro training set.

Method	IPT [1]	Stirpformer	Stripformer*
GoPro (PSNR)	32.58 (+0)	33.08 (+0.5)	33.71 (+1.13)

From Figure C to Figure F, we show comparison results on images from RealBlur-J testing set. Figure G to Figure J show some deblurring results on RWBI. As can be seen, these additional qualitative demonstrations further verify that our method performs favorably against the compared state-of-the-arts.

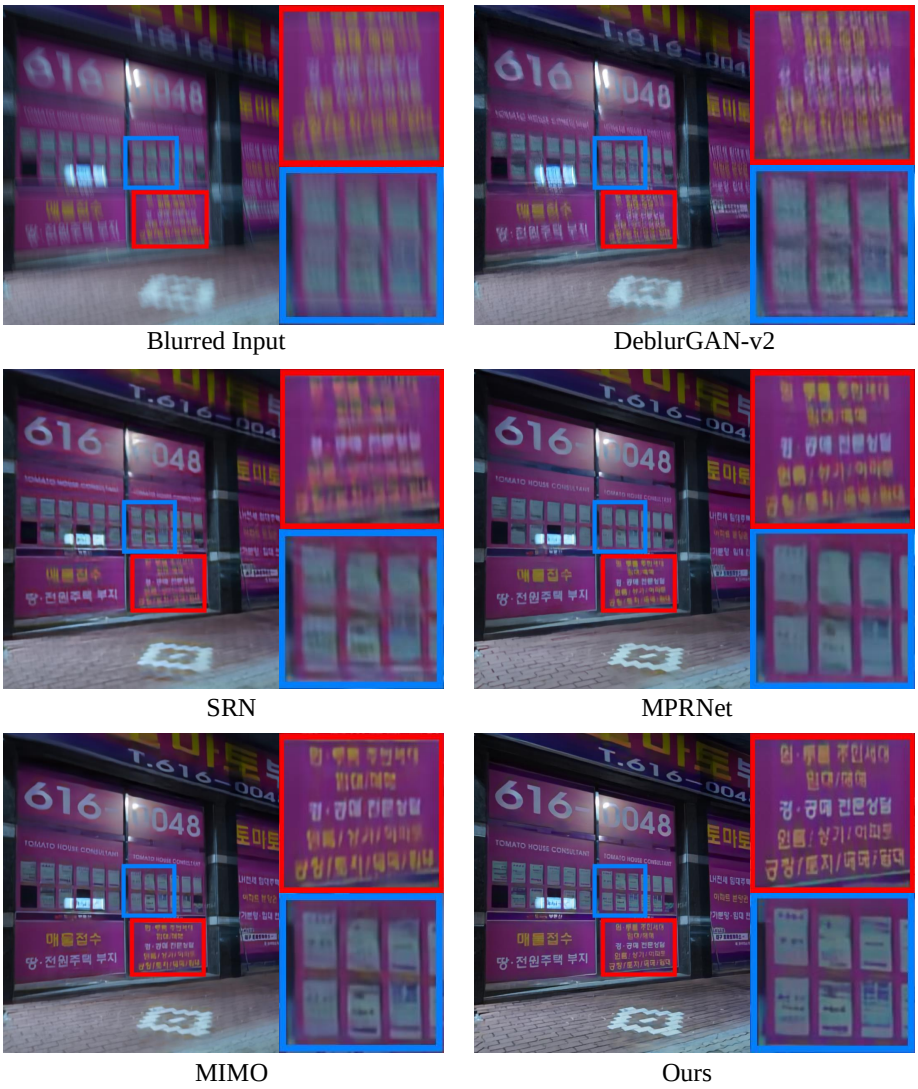

Fig. C. Qualitative comparisons on the RealBlur dataset. The deblurred results are produced by DeblurGan-v2 [3], SRN [6], MPRNet [7], MIMO [2] and our method.

Fig. D. Qualitative comparisons on the RealBlur dataset. The deblurred results are produced by DeblurGan-v2 [3], SRN [6], MPRNet [7], MIMO [2] and our method.

Fig. E. Qualitative comparisons on the RealBlur dataset. The deblurred results are produced by DeblurGan-v2 [3], SRN [6], MPRNet [7], MIMO [2] and our method.

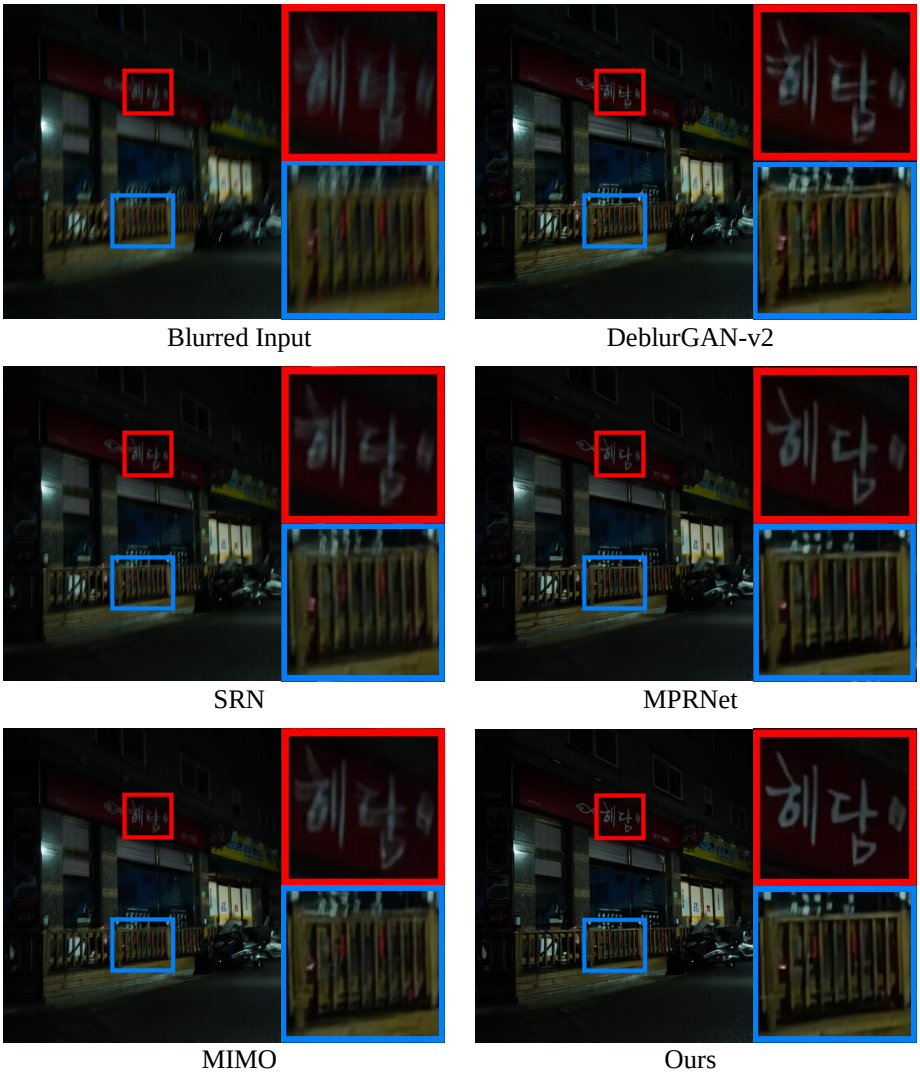

Fig. F. Qualitative comparisons on the RealBlur dataset. The deblurred results are produced by DeblurGan-v2 [3], SRN [6], MPRNet [7], MIMO [2] and our method.

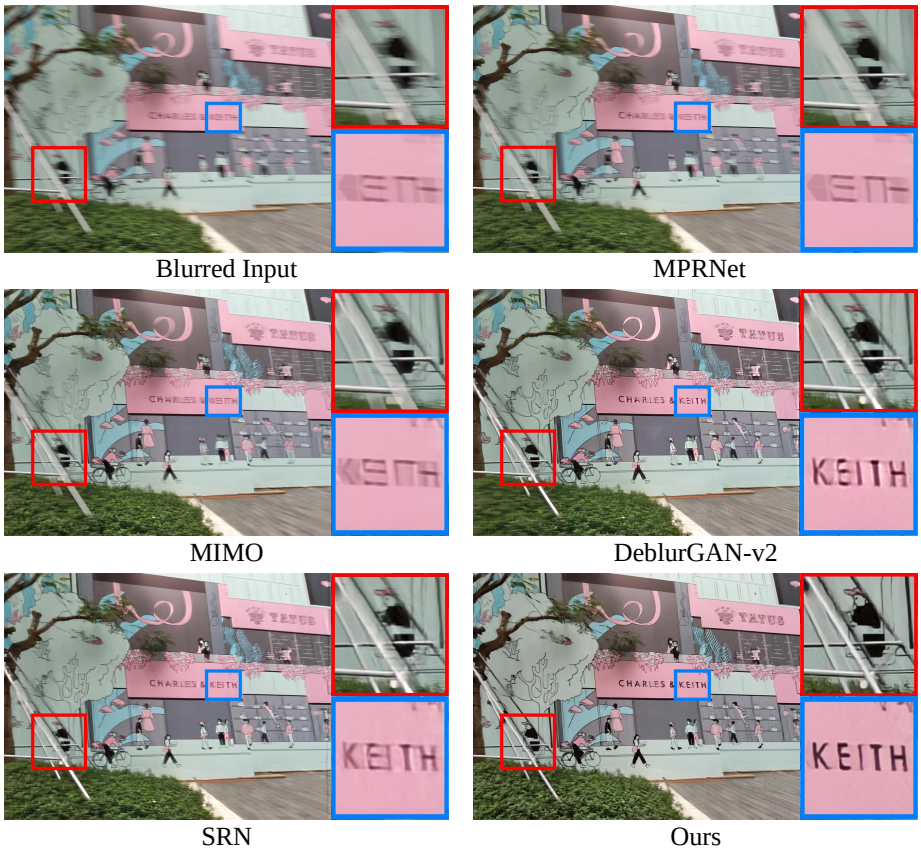

Fig. G. Qualitative comparisons on the RWBI dataset. The deblurred results are produced by MPRNet [7], MIMO [2]DeblurGan-v2 [3], SRN [6] and our method.

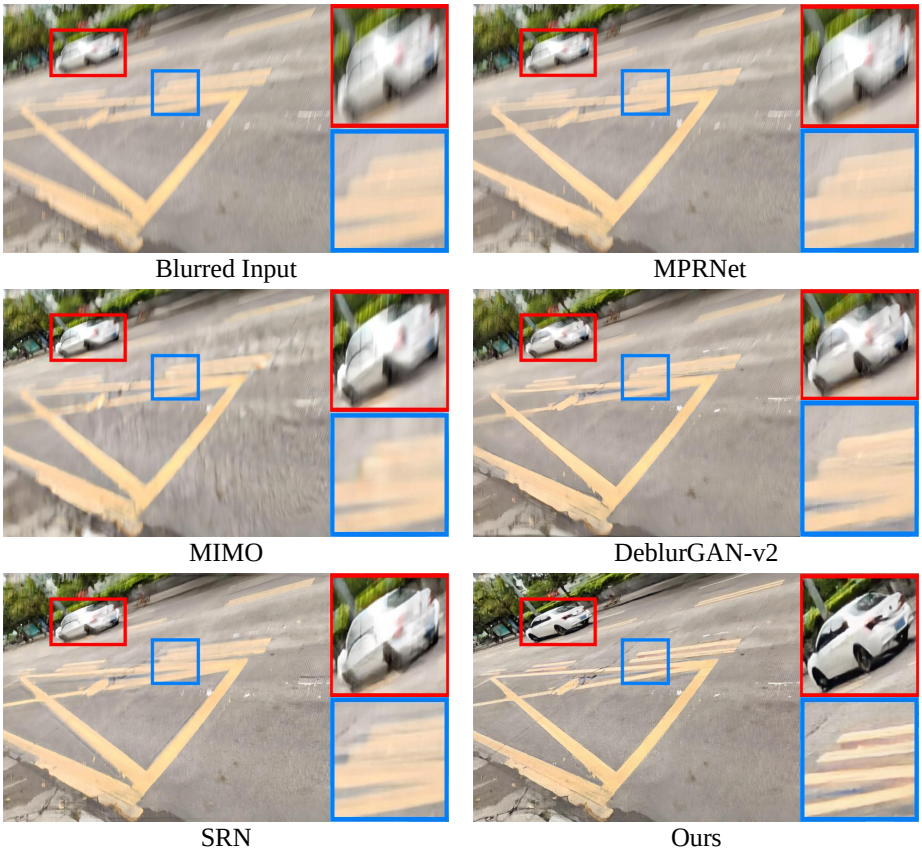

Fig. H. Qualitative comparisons on the RWBI dataset. The deblurred results are produced by MPRNet [7], MIMO [2]DeblurGan-v2 [3], SRN [6] and our method.

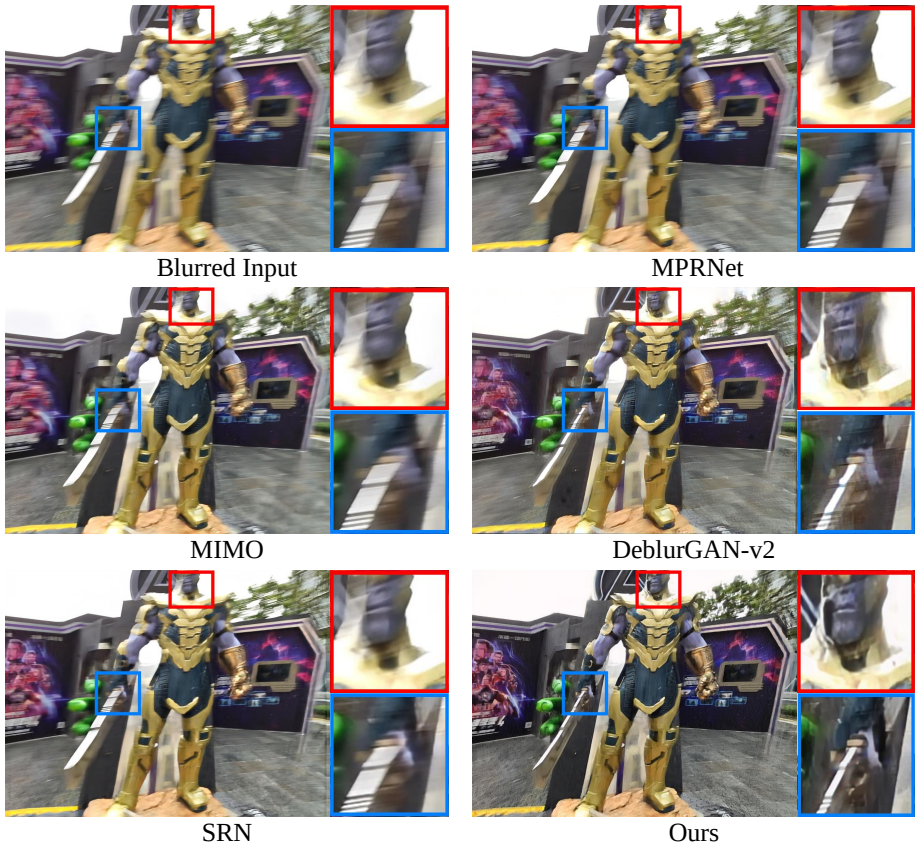

Fig. I. Qualitative comparisons on the RWBI dataset. The deblurred results are produced by MPRNet [7], MIMO [2]DeblurGan-v2 [3], SRN [6] and our method.

Fig. J. Qualitative comparisons on the RWBI dataset. The deblurred results are produced by MPRNet [7], MIMO [2]DeblurGan-v2 [3], SRN [6] and our method.

References

1. Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W.: Pre-trained image processing transformer. In: Proc. Conf. Computer Vision and Pattern Recognition (2021)
2. Cho, S.J., Ji, S.W., Hong, J.P., Jung, S.W., Ko, S.J.: Rethinking coarse-to-fine approach in single image deblurring. In: Proc. Int'l Conf. Computer Vision (2021)
3. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: Proc. Int'l Conf. Computer Vision (2019)
4. Nah, S., Baik, S., Hong, S., Moon, G., Son, S., Timofte, R., Lee, K.M.: Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In: Proc. Conf. Comput. Vis. Pattern Recognit Workshops (2019)
5. Rim, J., Lee, H., Won, J., Cho, S.: Real-world blur dataset for learning and benchmarking deblurring algorithms. In: Proc. Euro. Conf. Computer Vision (2020)
6. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: Proc. Conf. Computer Vision and Pattern Recognition (2018)
7. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proc. Conf. Computer Vision and Pattern Recognition (2021)
8. Zhang, K., Luo, W., Zhong, Y., Ma, L., Stenger, B., Liu, W., Li, H.: Deblurring by realistic blurring. In: Proc. Conf. Computer Vision and Pattern Recognition (2020)