MVSSM: Motion-aware Visual State Space Model for Efficient Video Deblurring –Supplementary Material–

Chen Zhou[1], Tao Wu[1], Wei Liu[2], Xi Wu[1], Ying Fu[1,*]

[1]Chengdu University of Information Technology, [2]Shanghai Jiao Tong University

3230609005@stu.cuit.edu.cn, wut@cuit.edu.cn, weiliucv@sjtu.edu.cn, wuxi@cuit.edu.cn, fuying@cuit.edu.cn

1. Overview

In the supplementary material, we provide a more detailed extension and explanation of the content in the main manuscript. First, in Sections 2 and 3, we conduct an in-depth analysis of the motion-aware scanning mechanism and the dynamic convolution module, respectively. Second, Section 4 investigates the influence of video length on deblurring performance. Section 5 discusses the limitations of the proposed method. Finally, Section 6 presents additional experimental results.

2. Motion-aware Scanning

We observe that the blur effect in a blurred frame manifests as pixel shifts along the blur trajectory. This makes features within the trajectory more important than those outside it. In basic visual state space models (such as VMamba [8]), the input image is flattened into a 1D sequence along the horizontal and vertical directions. However, such scanning mechanisms cannot adapt to all possible blur trajectories. Existing approaches [3, 7] also employ scanning methods along directions such as diagonal to enhance the ability of visual state space models in capturing local features. Meanwhile, some methods propose deformable scanning mechanisms for tasks like recognition and classification. However, in image restoration, the flattened 1D sequence needs to be reconstructed into a 2D image in subsequent steps, yet the aforementioned arbitrary deformable scanning mechanisms are generally irreversible. In light of this, we define four fixed scanning directions: horizontal, vertical, diagonal, and flipped diagonal scanning (Figure 1). These four scanning mechanisms are straightforward to invert. The optical flow estimator can compute pixel displacements between two blurred frames, and the offsets along the x- and y-axes can be conveniently used to estimate the pixel motion direction. As shown in Figure 2, the global pixel motion direction is largely consistent with the blur trajectory. Therefore, we adaptively assign scanning directions to blurred frames based on the overall motion angle to better capture spatial features along the blurred trajectories.

3. Dynamic Feature Fusion

The proposed MDFF consists of a gating-based feature fusion (Gate) module and a pixel-level dynamic convolution network ($\mathcal{DC}$). As shown in Table 1, the removal of either component leads to a significant decrease in both PSNR and SSIM, which occurs because Gate and $\mathcal{DC}$ adaptively explore and integrate spatial features. Furthermore, neither standard convolution modules (Conv) nor depthwise convolution (DConv) achieve the best results, as they are unable to distinguish the effectiveness of spatio-temporal features. Moreover, while single-scale dynamic convolution performs feature fusion only within fixed regions, multi-scale dynamic convolution overcomes this limitation.

4. Video Length

To analyze the effect of video length on deblurring, we set different lengths ($T = \{5, 10, 15, 20, 25, 30\}$) for ablation experiments. As shown in Figure 3, increasing the video length can improve the deblurring effect. However, when the video length $T \geq 30$, the improvement is not significant. This is because as the video length increases, the features contained in distant frames gradually decrease.

*Corresponding author

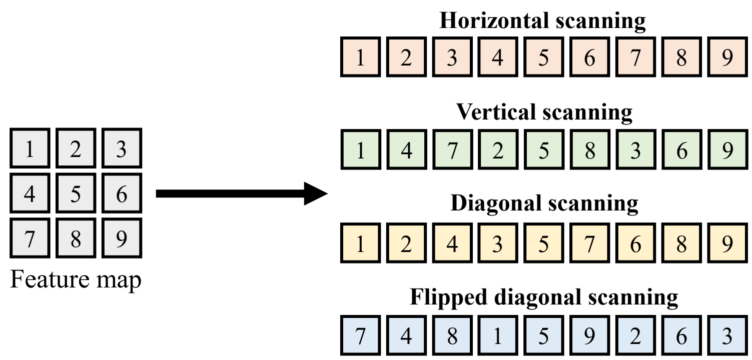

Figure 1. Horizontal, vertical, diagonal, and flipped diagonal scanning.

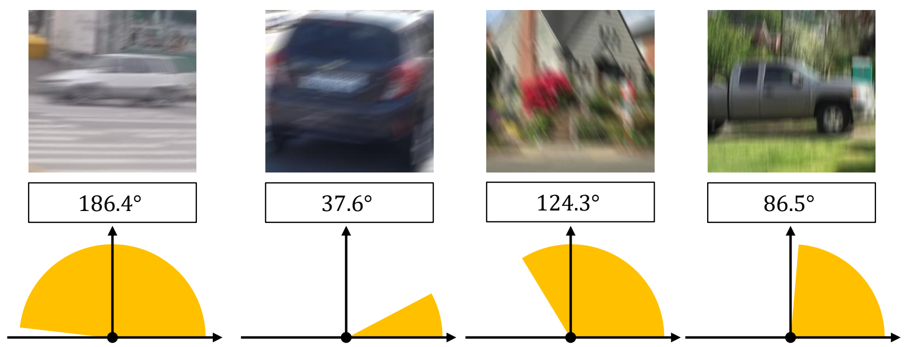

Figure 2. Results of using optical flow to compute the angle of blurred trajectories.

Table 1. Effectiveness of proposed MDFF.

Methods	w/o Gate	w/o $\mathcal{DC}$	w/ Conv in $\mathcal{DC}$	w/ DConv in $\mathcal{DC}$	w/ 3×3 in $\mathcal{DC}$	w/ 5×5 in $\mathcal{DC}$	w/ 7×7 in $\mathcal{DC}$	Ours
PSNRs	34.31	33.12	33.42	33.39	34.52	34.57	34.58	**34.66**
SSIMs	0.9679	0.9605	0.9622	0.9620	0.9697	0.9700	0.9700	**0.9702**

5. Limitations

The proposed method applies different scanning strategies to blurred frames based on the global motion direction, thereby reducing the computational burden of VSSM. However, only a few simple scanning mechanisms are employed in the current implementation, which are insufficient to fully model complex blur trajectories (e.g., mixed blur). Future work will explore more sophisticated scanning mechanisms to better approximate motion trajectories.

6. More Experimental Results

We provide more deblurred results on GOPRO [9], DVD [12], and BSD [15] datasets.

- Deblurred results on the GOPRO dataset [9] (the comparison methods are STFAN [16], CDVDTSP [10], FGST [6], DSTNet [11], EAMamba [7], and Shift-Net [4]) : Figure 4, 5, 6, 7, and 8.
- Deblurred results on the DVD dataset [12] (the comparison methods are CDVDTSP [10], MPRNet [13], RNN-MBP [2], FGST [6], DSTNet [11], and RVRT [5]) : Figure 9, 10, 11, and 12.
- Deblurred results on the BSD dataset [15] (the comparison methods are DVD [12], STFAN [16], CDVDTSP [10], ESTRNN [14], VDTR [1], and DSTNet [11]) : Figure 13, 14, 15, and 16, and 17.

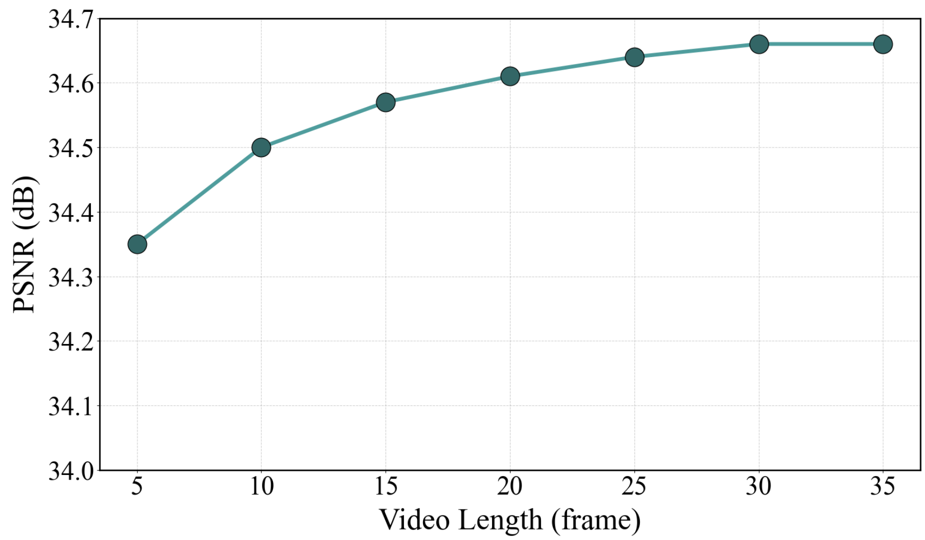

Figure 3. Effect of video length on deblurring on the GOPRO dataset [9].

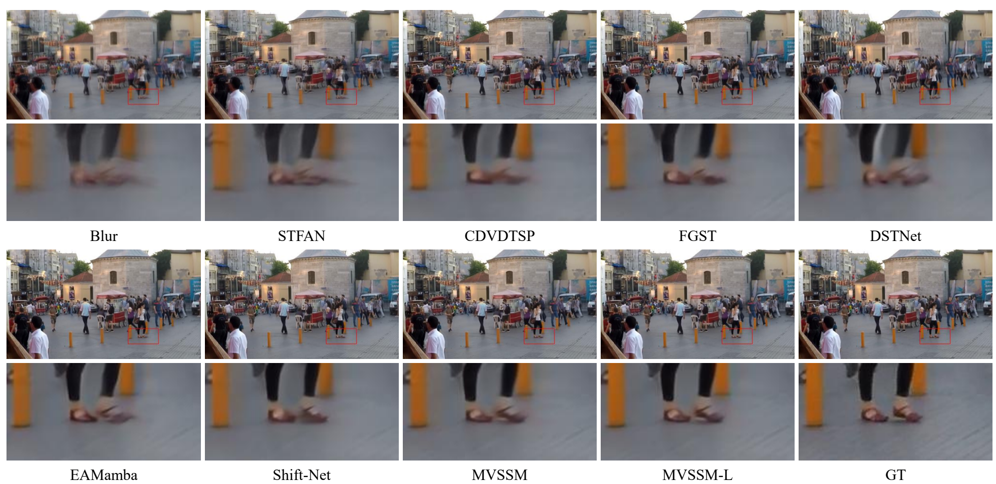

Figure 4. Deblurred results on the GOPRO dataset [9].

References

[1] Mingdeng Cao, Yanbo Fan, Yong Zhang, Jue Wang, and Yujiu Yang. Vdtr: Video deblurring with transformer. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1):160–171, 2023. 2

[2] Zhu Chao, Dong Hang, Pan Jinshan, Liang Boyang, Huang Yuhao, Fu Lean, and Wang Fei. Deep recurrent neural network with multi-scale bi-directional propagation for video deblurring. In *AAAI*, 2022. 2

[3] Boyun Li, Haiyu Zhao, Wenxin Wang, Peng Hu, Yuanbiao Gou, and Xi Peng. Mair: A locality- and continuity-preserving mamba for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7491–7501, 2025. 1

[4] Dasong Li, Xiaoyu Shi, Yi Zhang, Ka Chun Cheung, Simon See, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. A simple baseline for video restoration with grouped spatial-temporal shift. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9822–9832, 2023. 2

[5] Jingyun Liang, Yuchen Fan, Xiaoyu Xiang, Rakesh Ranjan, Eddy Ilg, Simon Green, Jiezhang Cao, Kai Zhang, Radu Timofte, and Luc Van Gool. Recurrent video restoration transformer with guided deformable attention. *arXiv preprint arXiv:2206.02146*, 2022. 2

[6] Jing Lin, Yuanhao Cai, Xiaowan Hu, Haoqian Wang, Youliang Yan, Xueyi Zou, Henghui Ding, Yulun Zhang, Radu Timofte, and Luc Van Gool. Flow-guided sparse transformer for video deblurring. In *ICML*, 2022. 2

[7] Yu-Cheng Lin, Yu-Syuan Xu, Hao-Wei Chen, Hsien-Kai Kuo, and Chun-Yi Lee. Eamamba: Efficient all-around vision state space

Figure 5. Deblurred results on the GOPRO dataset [9].

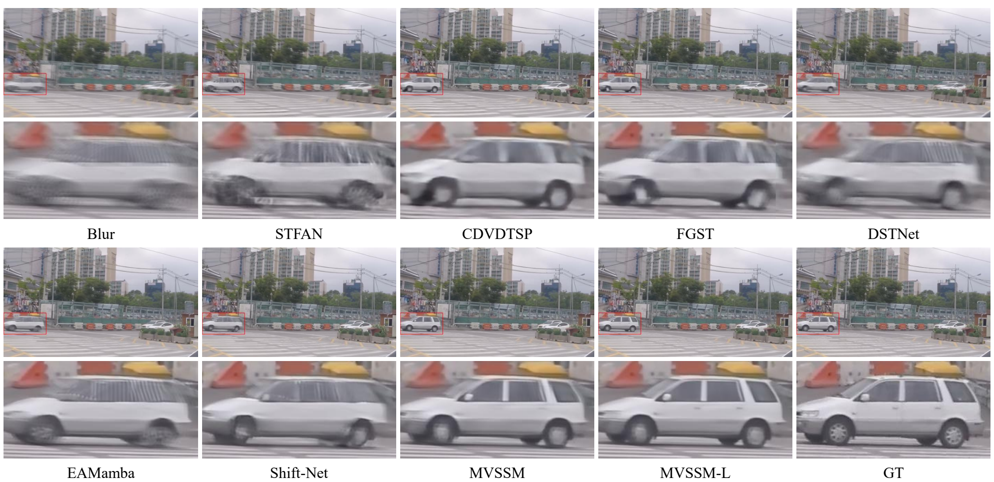

Figure 6. Deblurred results on the GOPRO dataset [9].

model for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11708–11719, 2025. 1, 2

[8] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, and Yunfan Liu. Vmamba: Visual state space model. *arXiv preprint arXiv:2401.10166*, 2024. 1

[9] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2, 3, 4, 5

[10] Jinshan Pan, Haoran Bai, and Jinhui Tang. Cascaded deep video deblurring using temporal sharpness prior. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

[11] Jinshan Pan, Boming Xu, Jiangxin Dong, Jianjun Ge, and Jinhui Tang. Deep discriminative spatial and temporal network for efficient video deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2023. 2

[12] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1279–1288, 2017. 2, 6, 7

[13] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021. 2

[14] Zhihang Zhong, Ye Gao, Yinqiang Zheng, and Bo Zheng. Efficient spatio-temporal recurrent neural network for video deblurring. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, pages 191–207. Springer, 2020. 2

Figure 7. Deblurred results on the GOPRO dataset [9].

Figure 8. Deblurred results on the GOPRO dataset [9].

[15] Zhihang Zhong, Ye Gao, Yinqiang Zheng, Bo Zheng, and Imari Sato. Real-world video deblurring: A benchmark dataset and an efficient recurrent neural network. *International Journal of Computer Vision*, 131(1):284–301, 2023. 2, 7, 8

[16] Shangchen Zhou, Jiawei Zhang, Jinshan Pan, Haozhe Xie, Wangmeng Zuo, and Jimmy Ren. Spatio-temporal filter adaptive network for video deblurring. In *Proceedings of the IEEE International Conference on Computer Vision*, 2019. 2

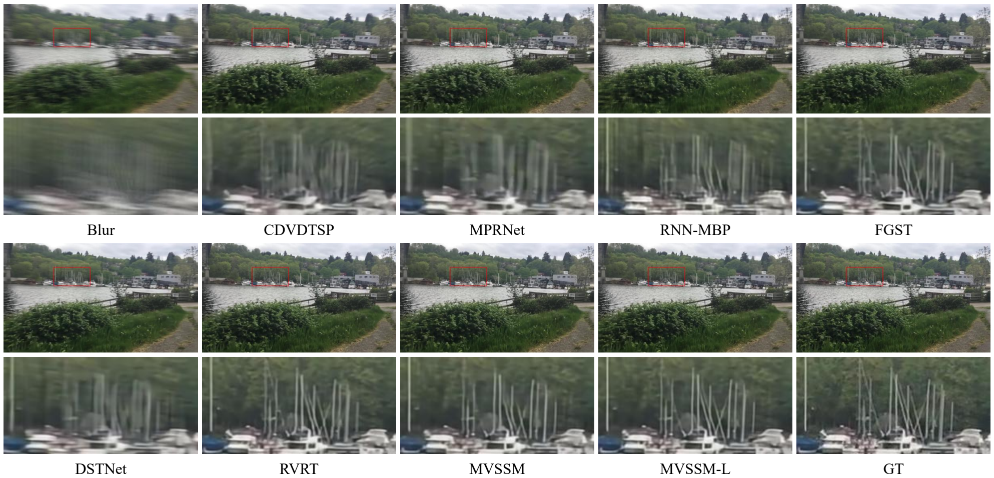

Figure 9. Deblurred results on the DVD dataset [12].

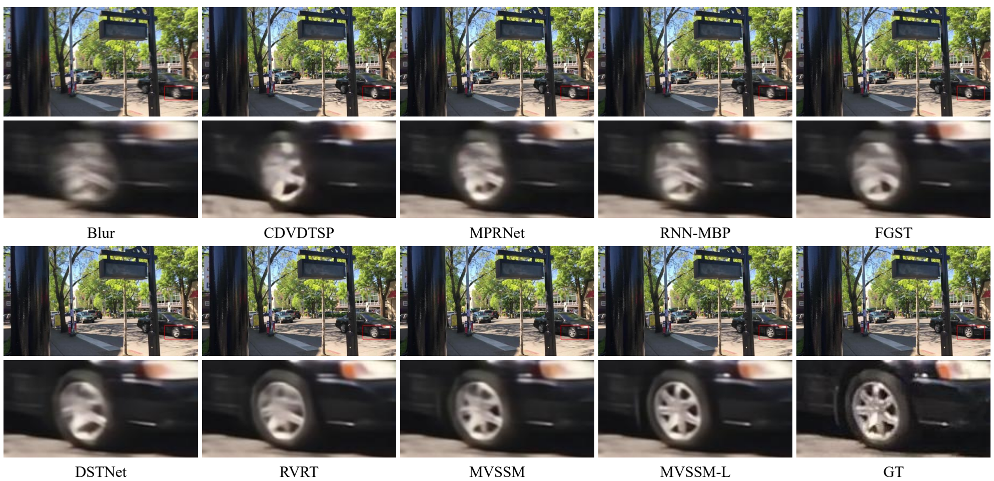

Figure 10. Deblurred results on the DVD dataset [12].

Figure 11. Deblurred results on the DVD dataset [12].

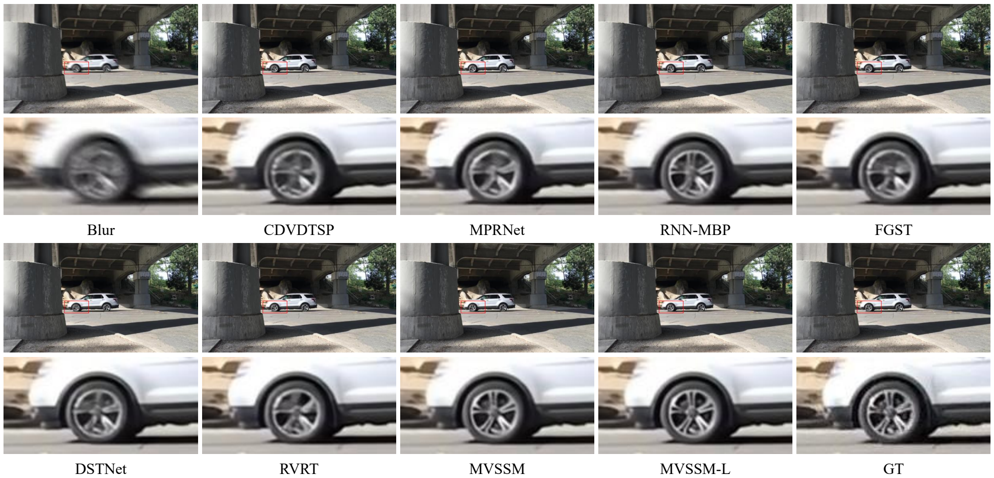

Figure 12. Deblurred results on the DVD dataset [12].

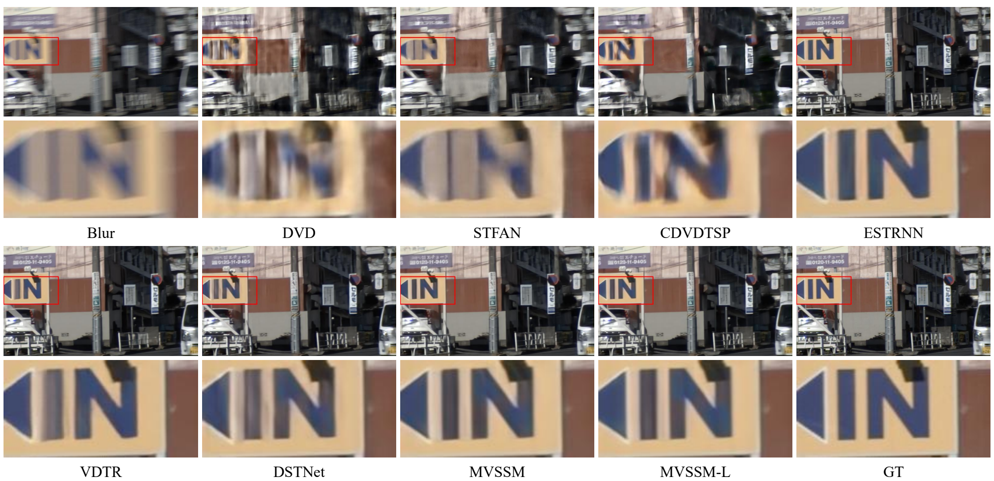

Figure 13. Deblurred results on the BSD dataset [15].

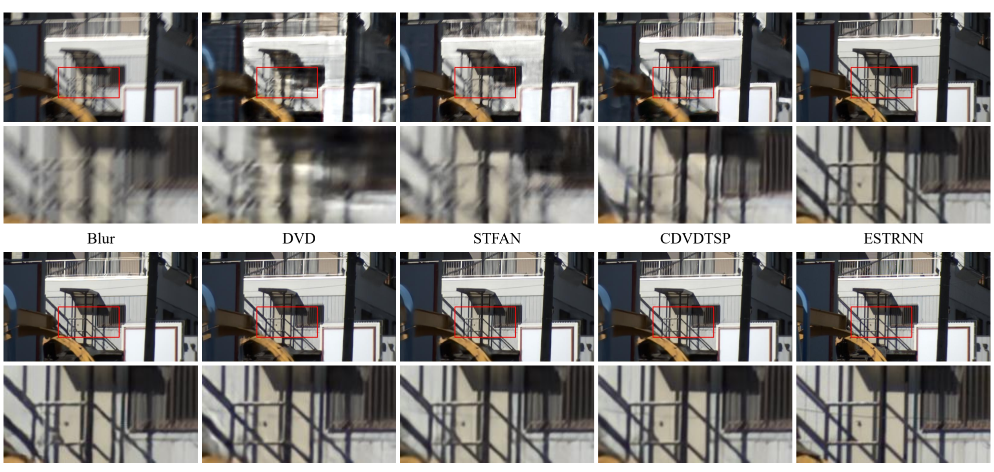

Figure 14. Deblurred results on the BSD dataset [15].

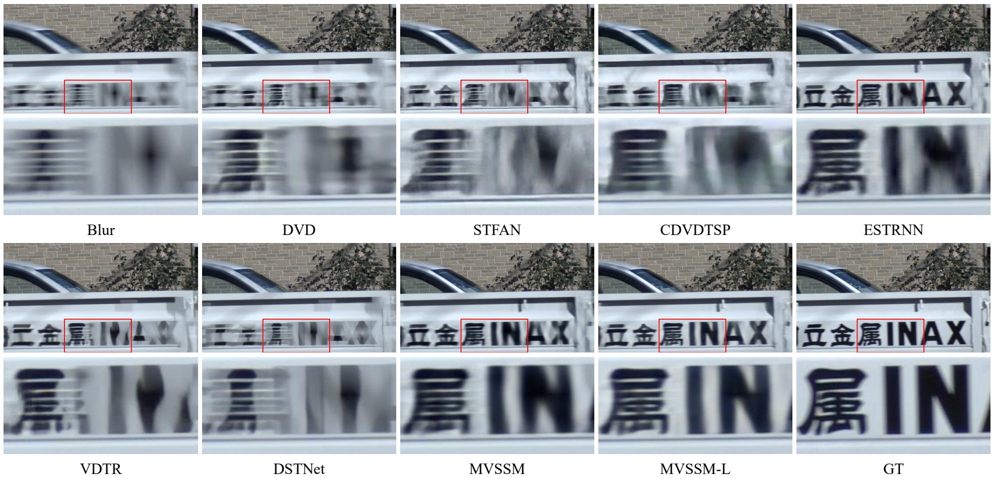

Figure 15. Deblurred results on the BSD dataset [15].

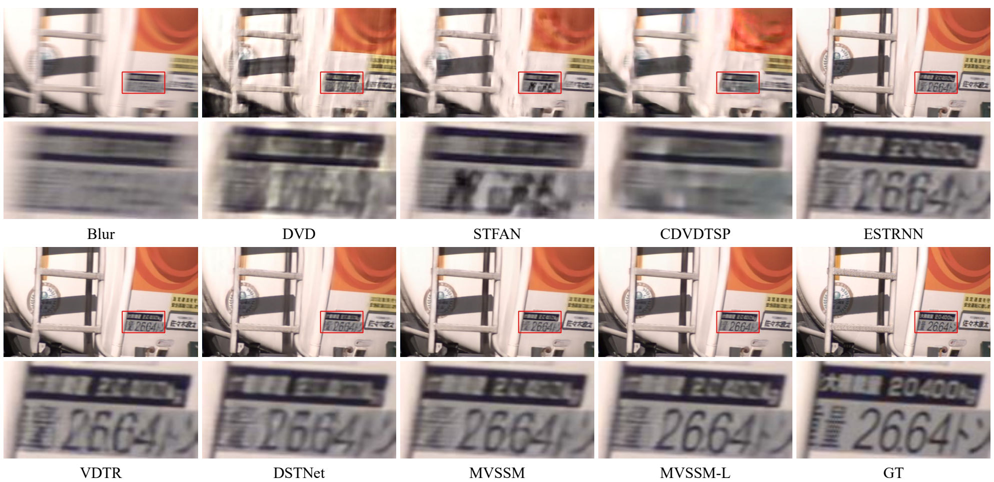

Figure 16. Deblurred results on the BSD dataset [15].

Figure 17. Deblurred results on the BSD dataset [15].