Cascaded Deep Video Deblurring Using Temporal Sharpness Prior Supplemental Material

Jinshan Pan, Haoran Bai, and Jinhui Tang
Nanjing University of Science and Technology

Overview

In this document, we first present the network details in Section 1. Then, we further analyze the effectiveness of the proposed cascaded training and temporal sharpness prior in Section 2 and Section 3. In Section 4, we demonstrate that the proposed algorithm is able to generate clear images frame by frame and has a better temporal consistent property. Finally, we show more visual comparisons on both synthetic and real-world images in Section 5.

1. Network Details

As stated in Section 4 of the manuscript, the proposed algorithm contains the optical flow estimation module, latent frame restoration module, and the temporal sharpness prior. Figure 1 shows the flowchart of the proposed algorithm at one stage. The proposed network shares the same network parameters when handling every three adjacent frames. The network architecture for the latent image restoration is shown in Figure 2. For the optical flow estimation, we use the PWC-Net [5] to estimate optical flow. All the network modules are jointly trained in an end-to-end manner.

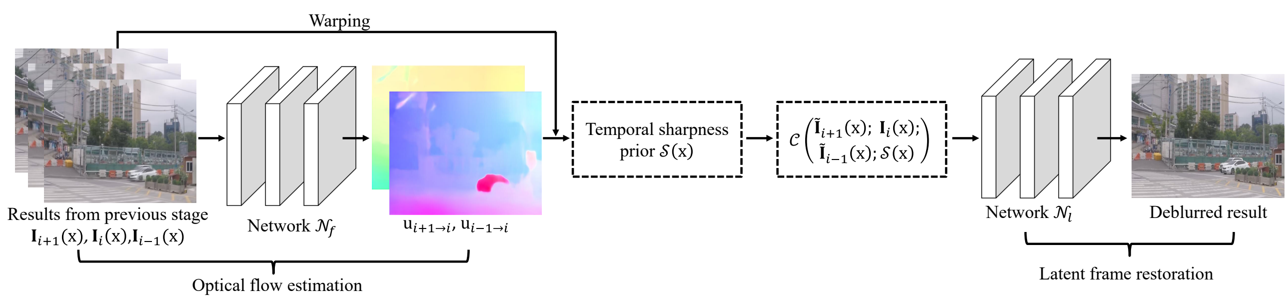


Figure 1. An overview of the proposed method at one stage. The proposed algorithm contains the optical flow estimation module, latent image restoration module, and the temporal sharpness prior. All the modules are jointly trained in an end-to-end manner. At each stage, it takes three adjacent frames estimated from the previous stage as the input and generates the deblurred results of the central frame. When handling every three adjacent frames, the proposed network shares the same network parameters. The variables $\tilde{\mathbf{I}}_{i+1}(\mathrm{x})$ and $\tilde{\mathbf{I}}_{i-1}(\mathrm{x})$ denote the warped results of $\mathbf{I}_{i+1}(\mathrm{x} + \mathbf{u}_{i+1\to i})$ and $\mathbf{I}_{i-1}(\mathrm{x} + \mathbf{u}_{i-1\to i})$, respectively.

2. Effectiveness of the Cascaded Training

As stated in Section 6.1 of the manuscript, we have analyzed the effectiveness of the cascaded training for video deblurring. In this supplemental material, we further show more visual comparisons to demonstrate the effectiveness of the cascaded training. Figure 3 shows that using the cascaded training method is able to generate much clearer images.

3. Effectiveness of the Temporal Sharpness Prior

As stated in Section 6.2 of the manuscript, we have analyzed the effectiveness of the temporal sharpness prior for video deblurring. In this supplemental material, we further show more visual comparisons to demonstrate the effectiveness of the

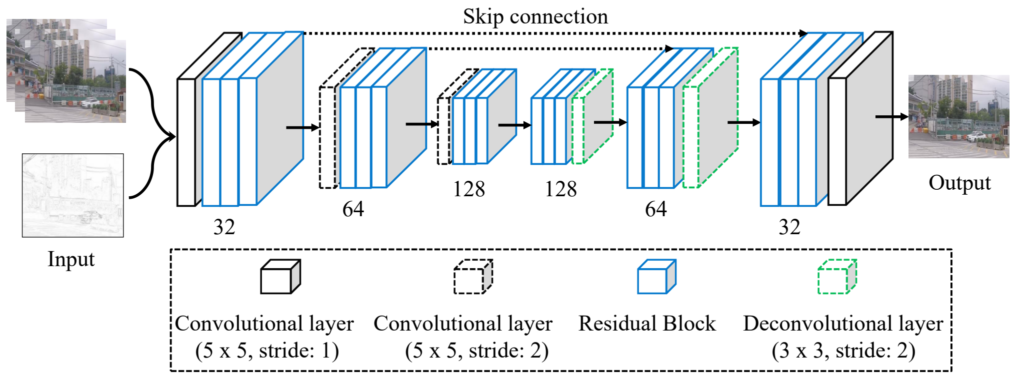


Figure 2. Detailed network architectures for the latent frame restoration. Each convolutional and deconvolutional layers are followed by a ReLU unit except the last one that outputs the latent frame. The number of feature channels in intermediate layers are 32, 64, 128, 128, 64, and 32, respectively.

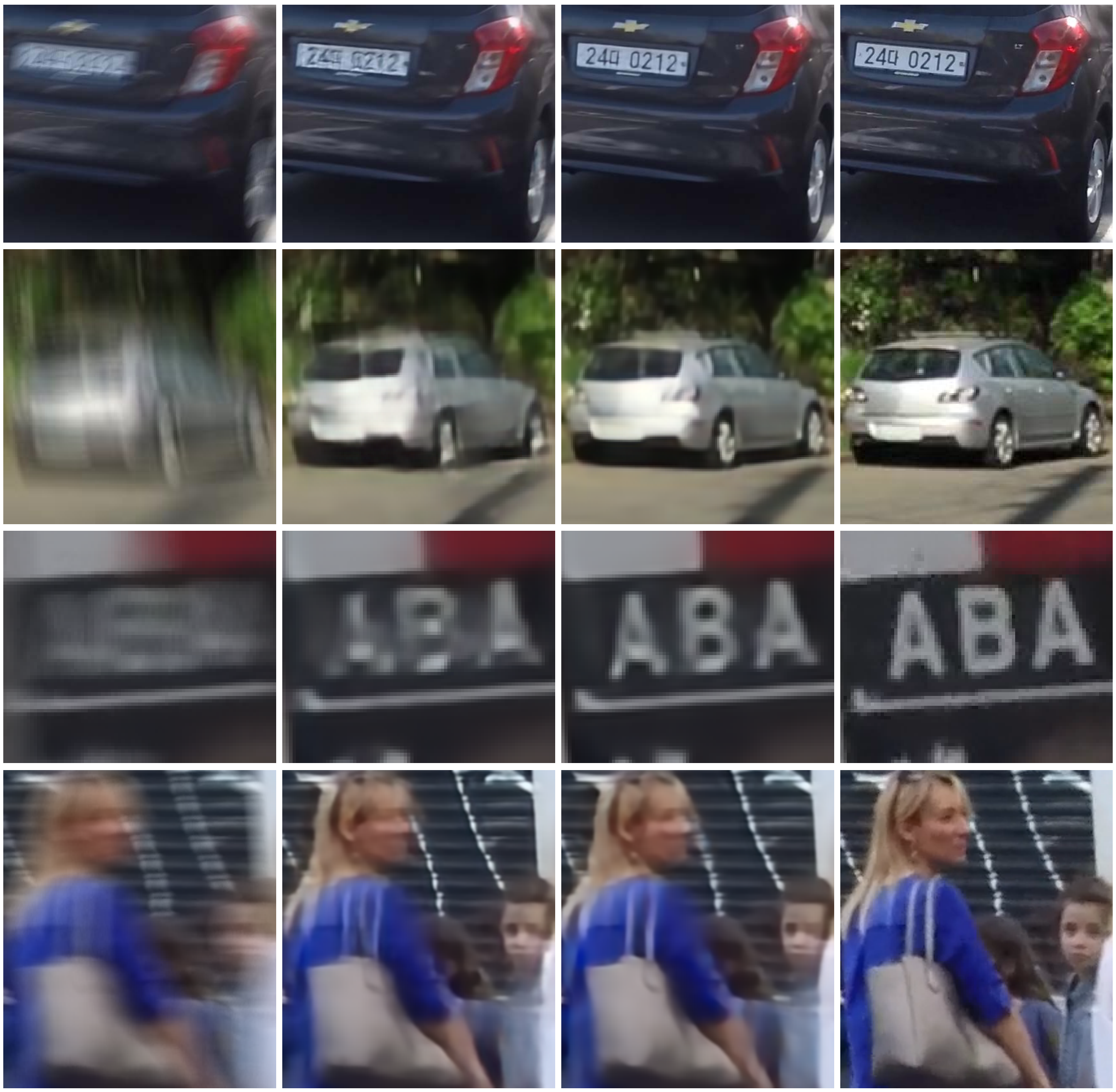
cascaded training. Figure 4 shows that using the temporal sharpness prior is able to distinguish the sharpness pixels and blurred pixels from adjacent frames and generates much clearer images.

4. Temporal Consistency Property

In the manuscript, we only show some deblurred frames. One may wonder whether the proposed algorithm is able to deblur all the frames and has a better temporal consistent property or not. To demonstrate the effectiveness of the proposed algorithm on temporal consistency, we show a real deblurred video in Figure 5. The restored video results show that the proposed algorithm deblurs each frame well and generates a much clear video. Thus, it has a better temporal consistency property than other methods.

5. More Experimental Results

In this section, we provide more visual comparisons with state-of-the-art methods on both synthetic (Figures 6-11) and real-world (Figures 12-14) images. The proposed algorithm generates much clearer frames.



(a) Blurred frames (b) w/o cascaded training (c) Ours (d) GT

Figure 3. Effectiveness of the cascaded training for video deblurring. Using the cascaded training is able to generate much clearer images while the method without using cascaded training method is less effective for blur removal.

(a) Cropped blurred patches (b) TSP (c) w/o TSP (d) Ours

Figure 4. Effectiveness of the temporal sharpness prior (TSP) for video deblurring. Using the TSP is able to generate much clearer images while the method without TSP is less effective for blur removal.

(a) Blurred video

(b) Kim et al. [2]

(c) Su et al. [4]

(d) EDVR [7]

(e) STFAN [8]

(f) Ours

Figure 5. Temporal consistency property. State-of-the-art methods do not deblur each frame well and there exist significant blur residual in some frames. In contrast, the proposed algorithm deblurs each frame well, where the characters and structures of the buildings are much clearer. Thus, it has a better temporal consistency property than other methods. (*Please use the Adobe Acrobat Reader to view the video with high-resolution display. For better visualization comparisons, the FPS is set to be 2.*)

(a) Cropped blurred patches
(b) Tao et al. [6]
(c) Su et al. [4]
(d) EDVR [7]
(e) STFAN [8]
(f) Ours

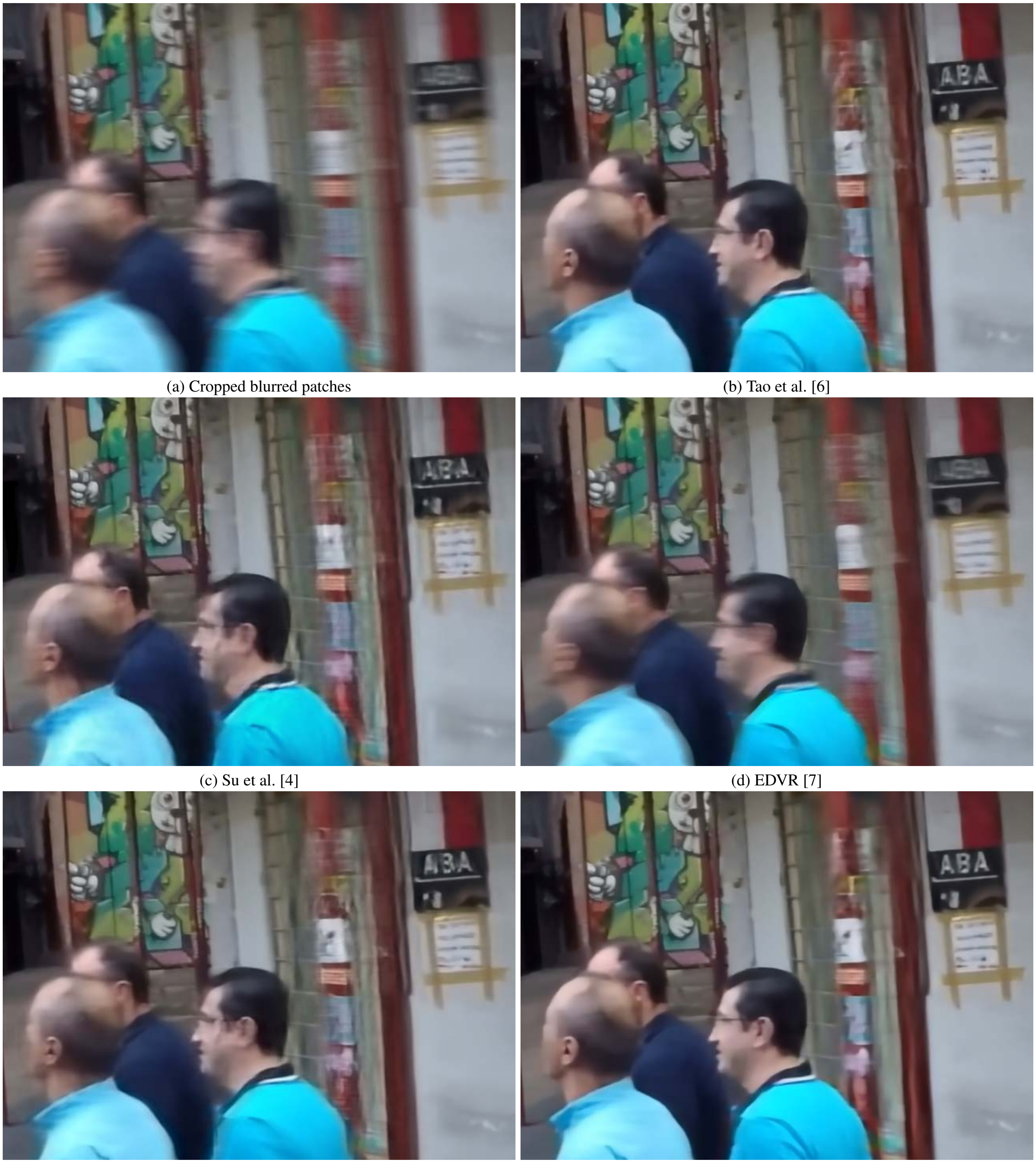
Figure 6. Deblurred results on the test dataset [4]. Our method generates the image with much clearer characters.

(a) Cropped blurred patches (b) Tao et al. [6]

(c) Su et al. [4] (d) EDVR [7]

(e) STFAN [8] (f) Ours

Figure 7. Deblurred results on the test dataset [4]. State-of-the-art methods generate the deblurred images with significant blur residual. In contrast, our method generates the image with clearer structural details.



(a) Cropped blurred patches (b) Tao et al. [6]

(c) Su et al. [4] (d) EDVR [7]

(e) STFAN [8] (f) Ours

Figure 8. Deblurred results on the test dataset [3]. State-of-the-art methods generate the deblurred images with significant blur residual and artifacts. In contrast, our method generates a clearer image with recognizable characters.


(a) Cropped blurred patches


(b) Tao et al. [6]


(c) Su et al. [4]


(d) EDVR [7]


(e) STFAN [8]


(f) Ours

Figure 9. Deblurred results on the test dataset [3]. State-of-the-art methods generate the deblurred images with significant blur residual. In contrast, our method generates a clearer image, where the face is much clearer.

(a) Cropped blurred patches

(b) Tao et al. [6]

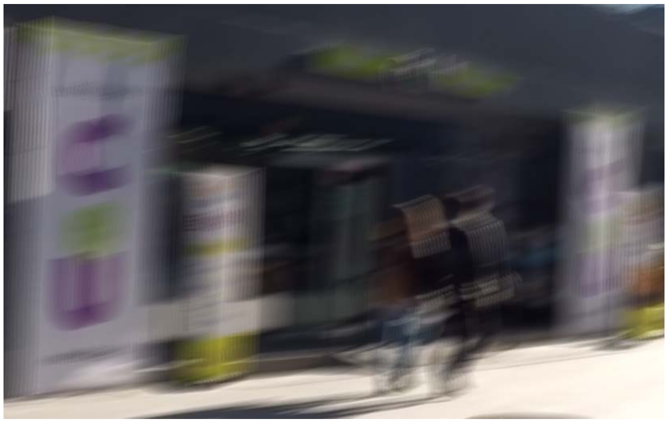
(c) Su et al. [4]

(d) EDVR [7]

(e) STFAN [8]

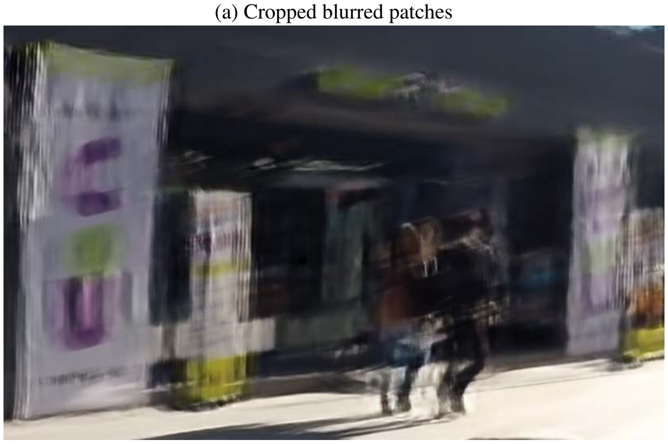
(f) Ours

Figure 10. Deblurred results on the test dataset [3]. State-of-the-art methods generate the deblurred images with significant blur residual. In contrast, our method generates a clearer image with recognizable characters.



(a) Cropped blurred patches

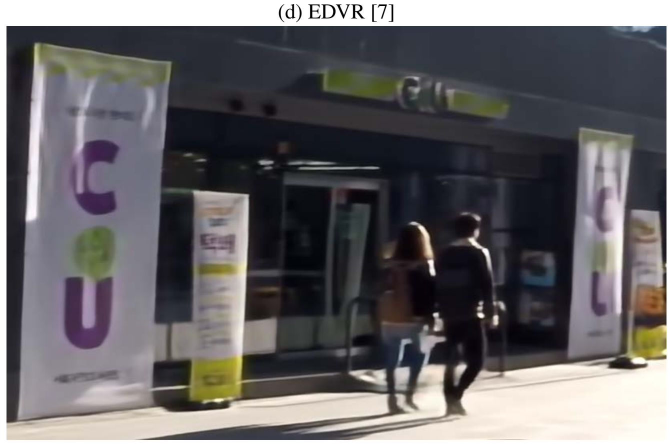
(b) Tao et al. [6]



(c) Su et al. [4]

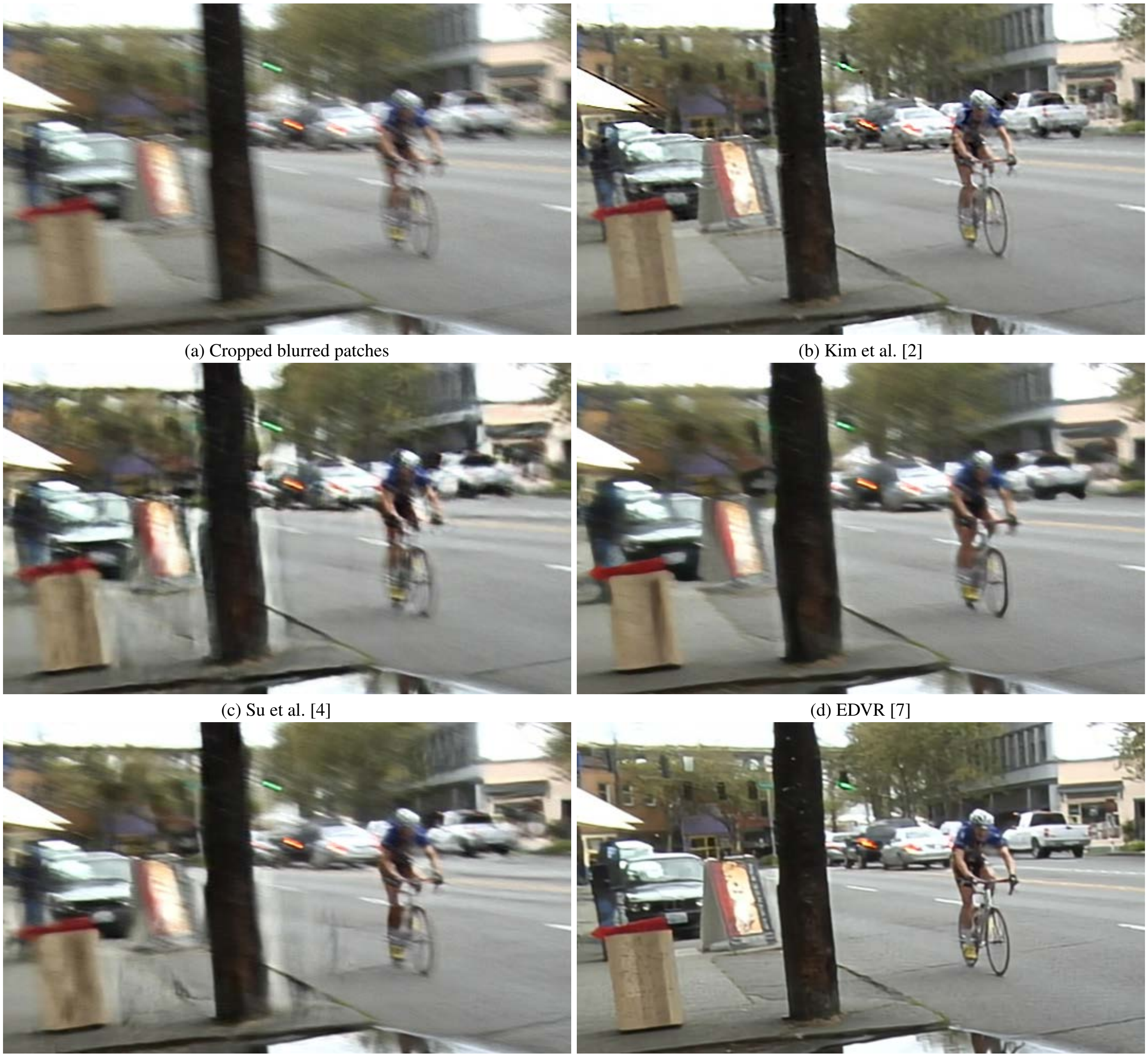
(d) EDVR [7]

(e) STFAN [8]



(f) Ours

Figure 11. Deblurred results on the test dataset [3]. State-of-the-art methods generate the deblurred images with significant blur residual and artifacts. In contrast, our method generates a much clearer image.



(a) Cropped blurred patches (b) Kim et al. [2]

(c) Su et al. [4] (d) EDVR [7]

(e) STFAN [8] (f) Ours

Figure 12. Deblurred results on a real video from [1]. State-of-the-art methods generate the deblurred images with significant blur residual. In contrast, our method generates a much clearer image with finer structural details.

(a) Cropped blurred patches (b) Tao et al. [6]

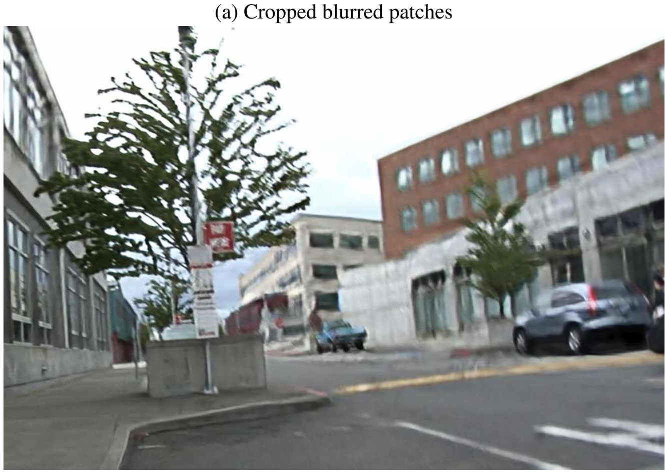
(c) Su et al. [4] (d) EDVR [7]

(e) STFAN [8] (f) Ours

Figure 13. Deblurred results on a real video from [1]. State-of-the-art methods generate the deblurred images with significant blur residual. In contrast, our method generates a clearer image with recognizable characters.

(a) Cropped blurred patches

(b) Tao et al. [6]



(c) Su et al. [4]

(d) EDVR [7]

(e) STFAN [8]

(f) Ours

Figure 14. Deblurred results on a real video from [1]. State-of-the-art methods generate the deblurred images with significant blur residual. In contrast, our method generates a clearer image with recognizable characters and finer structural details.

References

[1] Sunghyun Cho, Jue Wang, and Seungyong Lee. Video deblurring for hand-held cameras using patch-based synthesis. *ACM TOG*, 31(4):64:1–64:9, 2012. 12, 13, 14

[2] Tae Hyun Kim and Kyoung Mu Lee. Generalized video deblurring for dynamic scenes. In *CVPR*, pages 5426–5434, 2015. 5, 12

[3] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, pages 257–265, 2017. 8, 9, 10, 11

[4] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *CVPR*, pages 237–246, 2017. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

[5] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In *CVPR*, pages 8934–8943, 2018. 1

[6] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *CVPR*, pages 8174–8182, 2018. 6, 7, 8, 9, 10, 11, 13, 14

[7] Xintao Wang, Kelvin C.K. Chan, Ke Yu, Chao Dong, and Chen Change Loy. EDVR: Video restoration with enhanced deformable convolutional networks. In *CVPR Workshops*, 2019. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

[8] Shangchen Zhou, Jiawei Zhang, Jinshan Pan, Haozhe Xie, Wangmeng Zuo, and Jimmy Ren. Spatio-temporal filter adaptive network for video deblurring. In *ICCV*, 2019. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14