Deep Blind Video Super-resolution Supplemental Material

Jinshan Pan[1] Haoran Bai[1] Jiangxin Dong[1] Jiawei Zhang[2] Jinhui Tang[1]
[1]Nanjing University of Science and Technology [2]SenseTime Research

Overview

In this document, we first present the detailed parameters of the networks in Section 1. Then we qualitatively evaluate our method on the datasets with realistic blur kernels in Section 2 and present both the qualitative and quantitative evaluation on the datasets with the commonly used degradation model (i.e., Bicubic) in Section 3. Next, we further demonstrate the effectiveness of the proposed blur kernel estimation method on the video super-resolution problem in Section 4. Section 5 provide the qualitative evaluations of the feature warping operation. Section 6 analyzes that the proposed method generates videos with better temporal consistency property. We report the running time of the proposed method against state-of-the-art methods in Section 7. Finally, we show more visual comparisons in Section 8.

1. Network Parameters

In the main manuscript, we have shown the architectures of the proposed network including $\mathcal{N}_k, \mathcal{N}_e, \mathcal{N}_f, \mathcal{N}_\gamma, \mathcal{N}_\beta, \mathcal{N}_d$, and $\mathcal{N}_I$. For the optical flow estimation network, we use the network architectures by the PWC-Net [12] with default parameters. In this supplemental material, we present the network parameters of $\mathcal{N}_k, \mathcal{N}_e, \mathcal{N}_f, \mathcal{N}_\gamma, \mathcal{N}_\beta, \mathcal{N}_d$, and $\mathcal{N}_I$ in Figures 1-3.

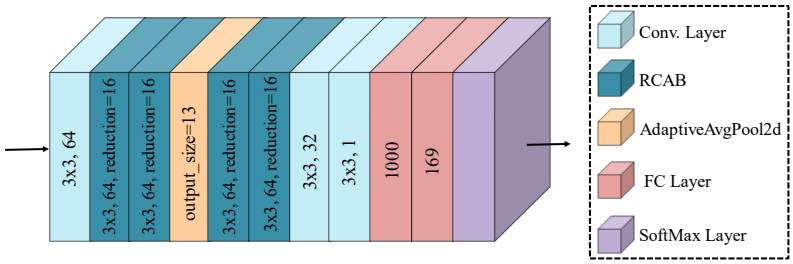


Figure 1. Network architectures and parameters of the blur kernel estimation network N_k. "FC" denotes a fully connected layer. "RCAB" denotes the Residual Channel Attention Block (RCAB) by [18]. The padding operation is used in all the convolutional layers. The first three numbers (e.g., 3 x 3, 64) in each unit denote the filter size and the number of the output feature channels.

2. Qualitative Evaluations on the Datasets with Realistic Blur Kernels

As shown in Table 2 of the main manuscript, our method is able to solve the LR videos with realistic blur kernels. In this document, we show the visual comparison results with realistic blur kernels from [1]. Some blur kernels that are used for evaluations are shown in Figure 4. Figures 5-7 show that the proposed method generates much clearer SR frames with finer detailed structures. All of these demonstrate that our method is able to solve the LR videos with realistic blur kernels.

3. Evaluations on the Datasets with Bicubic Kernels

We note that most existing video SR methods usually use the Bicubic downsampling as the approximation of the degradation process and use the training datasets generated by the Bicubic downsampling to train the deep CNN model. To examine whether the proposed method works well or not for such case, we follow the protocols of [3, 15, 16] to train the proposed

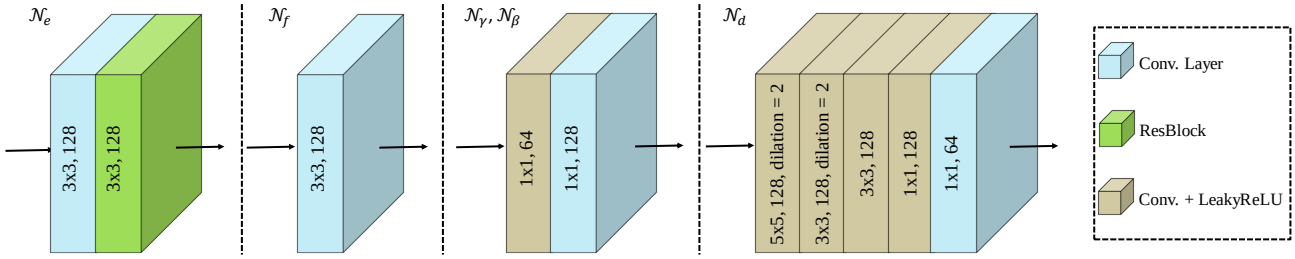


Figure 2. Network architectures and parameters of the networks $\mathcal{N}_e$, $\mathcal{N}_f$, $\mathcal{N}_d$, $\mathcal{N}_\gamma$, and $\mathcal{N}_\beta$. The negative slope value of the LeakyReLU is set to be 0.1. The padding operation is used in all the convolutional layers. When applying the network $\mathcal{N}_e$ to the color images, the number of the input feature channel in the first convolutional layer of $\mathcal{N}_e$ is set to be 3. When applying the network $\mathcal{N}_e$ to $\mathcal{S}(\tilde{I}_i)$, the number of the input feature channel in the first convolutional layer of $\mathcal{N}_e$ is the same as that of $\mathcal{S}(\tilde{I}_i)$. $\mathcal{S}$ denotes the spatial-to-depth transformation.

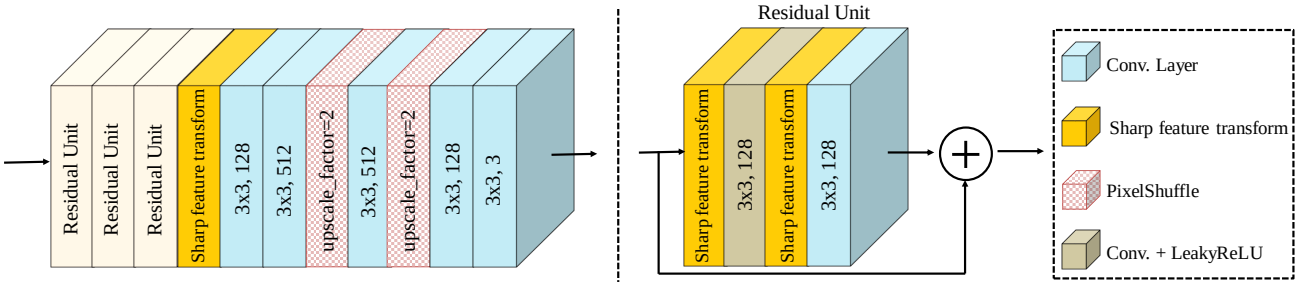


Figure 3. Network architectures and parameters of the restoration network $\mathcal{N}_I$. The negative slope value of the LeakyReLU is set to be 0 (i.e., ReLU). The padding operation is used in all the convolutional layers.

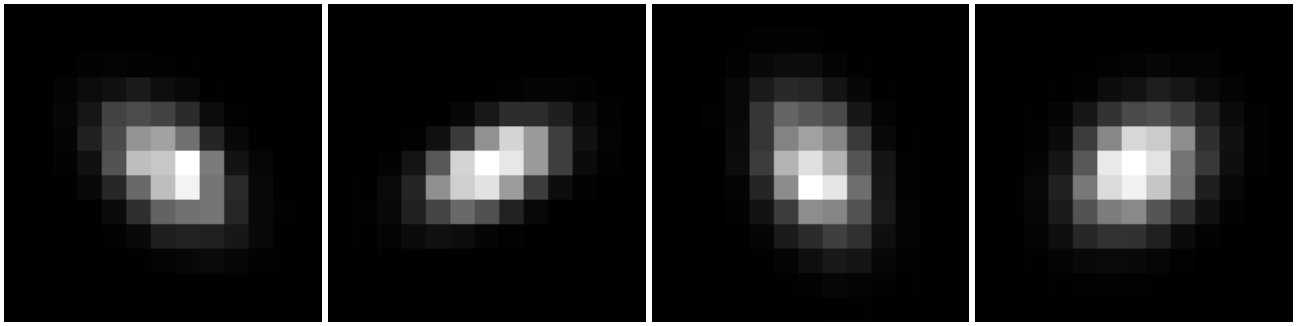


Figure 4. Realistic blur kernels that are used for evaluations. The size of the blur kernels are resized into 13×13 pixels for evaluations.

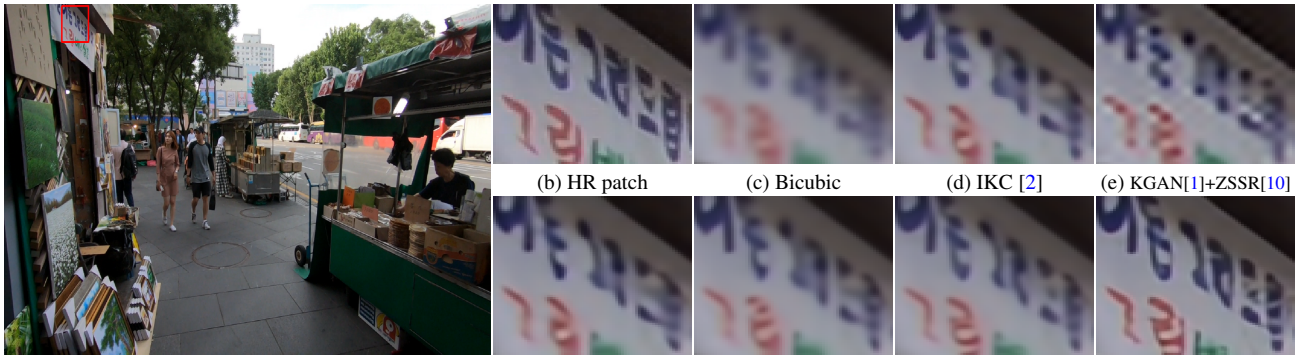
Table 1. Quantitative evaluations of the state-of-the-art video SR methods on the REDS4 dataset, where the degradation process is approximated by the Bicubic downsampling.

Algorithms	Bicubic	RCAN [18]	SPMC [13]	DUF [5]	TOFlow [16]	RBPN [3]	Ours-L
REDS4 [15]	25.59 0.7077	28.71 0.8184	27.74 0.7915	28.60 0.8254	27.77 0.7949	29.82 0.8537	**30.28** **0.8643**

method on the REDS dataset [8]. Table 1 shows the quantitative evaluations on the REDS4 dataset. Although our algorithm is designed for the blind video SR problem, it performs favorably against state-of-the-art methods when using the Bicubic downsampling as the approximation of the degradation process, suggesting the effectiveness of the proposed method.

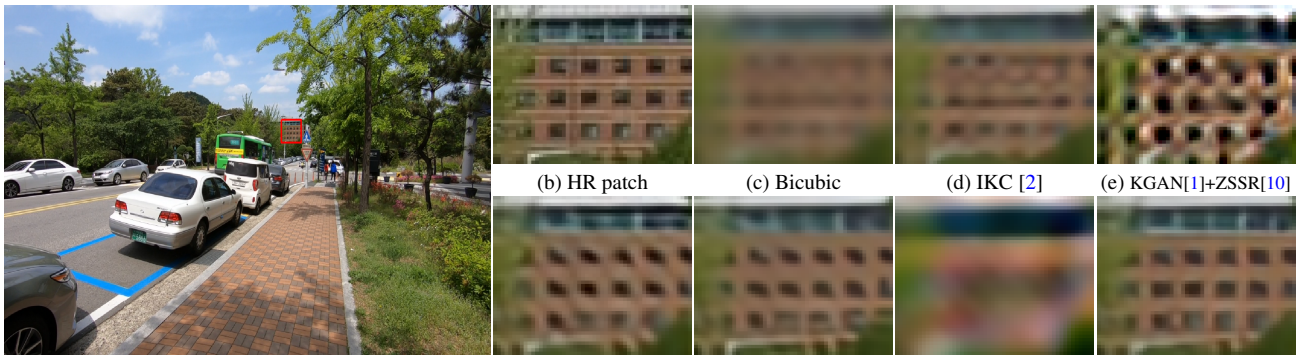
4. Effectiveness of the Blur Kernel Estimation on Blind Video Super-resolution

In Figure 1, Table 4, and Figure 7 of the main manuscript, we have shown that using the blur kernel estimation is able to generate high-quality frames with clearer structural details. In this supplemental material, we further demonstrate the effect of the blur kernel estimation on video super-resolution. We compare the proposed method without using the blur kernel



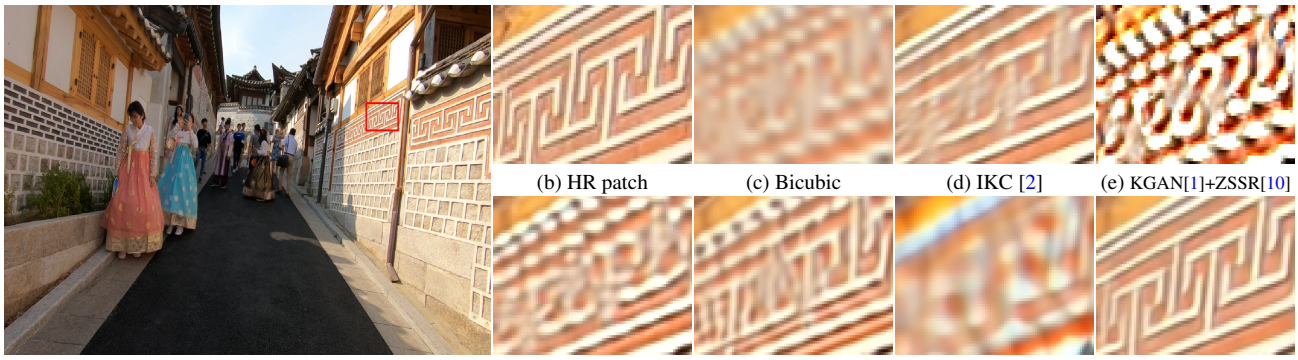
(a) Ground truth HR frame (b) HR patch (c) Bicubic (d) IKC [2] (e) KGAN[1]+ZSSR[10]
(f) DUF [5] (g) RBPN [3] (h) EDVR [15] (i) Ours-L

Figure 5. Video SR result ($\times 4$) on the REDS dataset [8], where the realistic blur kernels [1] are used in the degradation process. The proposed algorithm recovers high-quality frames with clearer structures.



(a) Ground truth HR frame (b) HR patch (c) Bicubic (d) IKC [2] (e) KGAN[1]+ZSSR[10]
(f) DUF [5] (g) RBPN [3] (h) Deblur[9]+EDVR[15] (i) Ours-L

Figure 6. Video SR result ($\times 4$) on the REDS dataset [8], where the realistic blur kernels [1] are used in the degradation process. The proposed algorithm recovers high-quality frames with clearer structures.



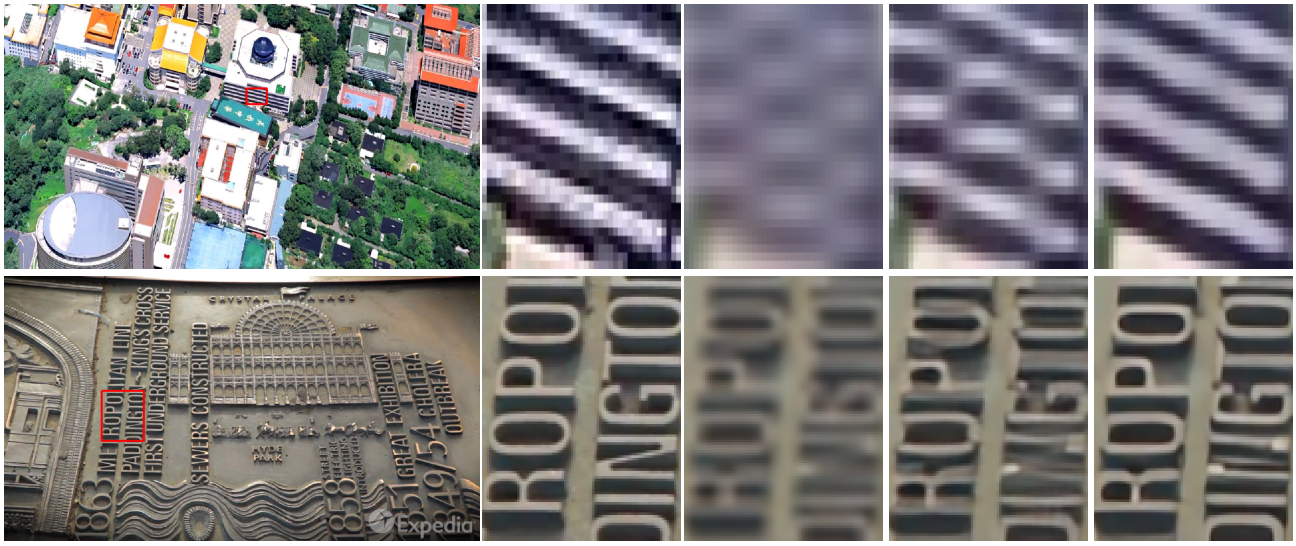
(a) Ground truth HR frame (b) HR patch (c) Bicubic (d) IKC [2] (e) KGAN[1]+ZSSR[10]
(f) DUF [5] (g) RBPN [3] (h) Deblur[9]+EDVR[15] (i) Ours-L

Figure 7. Video SR result ($\times 4$) on the REDS dataset [8], where the realistic blur kernels [1] are used in the degradation process. The proposed algorithm recovers high-quality frames with clearer structures.

estimation. The results in Figure 8 demonstrate that using the motion blur estimation generates high-resolution images with finer structural details.

5. Qualitative Evaluations of the Image-space Warping and Feature Space Warping Operations

In Table 6 of the main manuscript, we have shown that using the warping operation in the deep feature space generates the results with higher PSNR and SSIM values. In supplemental material, we provide some visual comparisons. Figure 9 shows



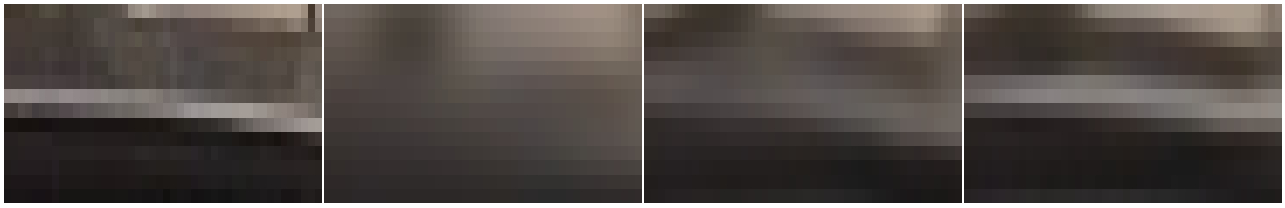
(a) Ground truth high-resolution frame (b) HR patch (c) Bicubic (d) w/o kernel modeling (e) Ours

Figure 8. Effectiveness of the blur kernel estimation on video super-resolution ($\times 4$). Using the blur kernel estimation is able to generate the results with much clearer structural details.

Table 2. Comparisons of running time (in second). The results are tested on the videos with 720×1280 pixels.

Methods	RCAN [18]	RBPN [3]	DAN [6]	USRNet [17]	Ours
Running time (/s)	0.945	0.709	0.544	1.36	0.523

the visual comparisons of the results by the image space warping and feature space warping. Using feature space warping can generate high-quality frames with better structural details.



(a) GT (b) Bicubic (c) Image space warping (d) Feature space warping

Figure 9. Comparisons of the results ($\times 4$) by the image space warping and feature space warping.

6. Temporal Consistency

To demonstrate the effectiveness of the proposed algorithm on temporal consistency, we show the super-resolved videos in Figure 10. The video results show that the proposed algorithm has a better temporal consistency property than other methods.

7. Running Time Comparisons

In the main manuscript, we show that the proposed method has relatively fewer model parameters and the lowest floating point operations (FLOPs) in Table 7. In this document, we further report the running time of the proposed against state-of-the-art methods. Table 2 shows that the proposed method is much more efficient. In contrast to the methods [6, 17] that need alternative optimization processes to generate HR images, our network directly estimates latent HR frames, which is 2.6x faster than [17].

8. More Experimental Results

In this section, we show more experimental results to demonstrate the effectiveness of the proposed methods.

Results on the unknown realistic blur kernels. In addition to Figures 5-7, we further show the evaluation results on the realistic blur kernels in Figure 11. The results show that the proposed algorithm is able to super-resolve videos and generates much clearer results than those by state-of-the-art methods.

Results on the unknown Gaussian blur kernels. Figures 12-15 show evaluation results from the SPMCS dataset [13], where the blur kernel in the degradation process are unknown Gaussian kernels. The results show that the proposed algorithm generates much clearer frames with finer detailed structures than those by state-of-the-art methods.

Results on real-world videos with complex motion blur. Figures 16-19 show evaluation results on real-world videos which contains complex motion blur [7]. The results show that the proposed algorithm generates much clearer frames with finer detailed structures than those by state-of-the-art methods.

References

[1] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In *NeurIPS*, pages 284–293, 2019. 1, 3, 7, 8, 9

[2] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *CVPR*, pages 1604–1613, 2019. 3, 7, 8, 9

[3] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Recurrent back-projection network for video super-resolution. In *CVPR*, pages 3897–3906, 2019. 1, 2, 3, 4, 7, 8, 9

[4] Takashi Isobe, Songjiang Li, Xu Jia, Shanxin Yuan, Gregory G. Slabaugh, Chunjing Xu, Ya-Li Li, Shengjin Wang, and Qi Tian. Video super-resolution with temporal group attention. In *CVPR*, pages 8005–8014, 2020. 7

[5] Younghyun Jo, Seoung Wug Oh, Jaeyeon Kang, and Seon Joo Kim. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation. In *CVPR*, pages 3224–3232, 2018. 2, 3, 6, 7, 9

[6] Zhengxiong Luo, Yan Huang, Shang Li, Liang Wang, and Tieniu Tan. Unfolding the alternating optimization for blind super resolution. In *NeurIPS*, 2020. 4

[7] Ziyang Ma, Renjie Liao, Xin Tao, Li Xu, Jiaya Jia, and Enhua Wu. Handling motion blur in multi-frame super-resolution. In *CVPR*, pages 5224–5232, 2015. 5

[8] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. NTIRE 2019 challenge on video deblurring and super-resolution: Dataset and study. In *CVPR Workshops*, pages 1996–2005, 2019. 2, 3, 7

[9] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Deblurring images via dark channel prior. *IEEE TPAMI*, 40(10):2315–2328, 2018. 3, 7, 8

[10] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *CVPR*, pages 3118–3126, 2018. 3, 7, 8, 9

[11] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *CVPR*, pages 3513–3522, 2020. 8, 9

[12] 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

[13] Xin Tao, Hongyun Gao, Renjie Liao, Jue Wang, and Jiaya Jia. Detail-revealing deep video super-resolution. In *ICCV*, pages 4482–4490, 2017. 2, 5, 7, 8, 9

[14] Yapeng Tian, Yulun Zhang, Yun Fu, and Chenliang Xu. TDAN: temporally-deformable alignment network for video super-resolution. In *CVPR*, pages 3357–3366, 2020. 7

[15] 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. 1, 2, 3, 7, 8

[16] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. Video enhancement with task-oriented flow. *IJCV*, 127(8):1106–1125, 2019. 1, 2, 7, 8, 9

[17] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *CVPR*, pages 3214–3223, 2020. 4

[18] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, pages 294–310, 2018. 1, 2, 4, 6, 8, 9

(a) Input video

(b) Bicubic

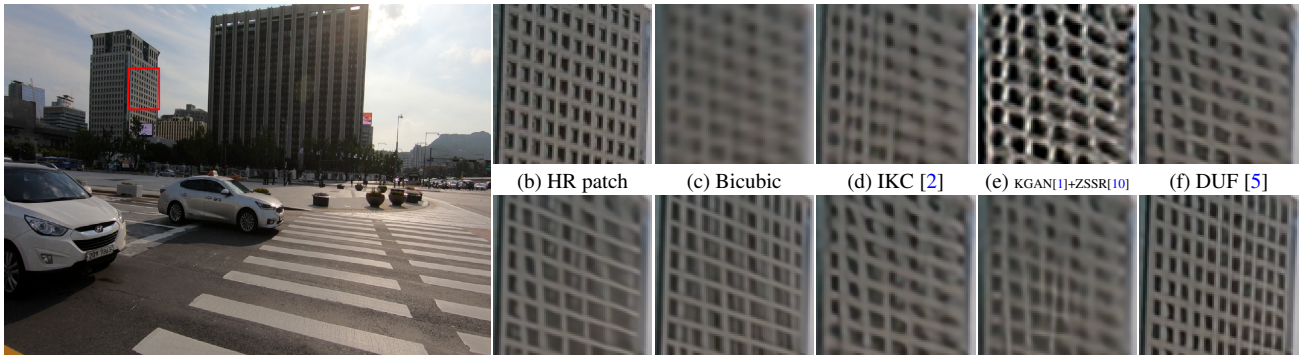
(c) RCAN [18]

(d) DUF [5]

(e) w/o kernel modeling

(f) Ours

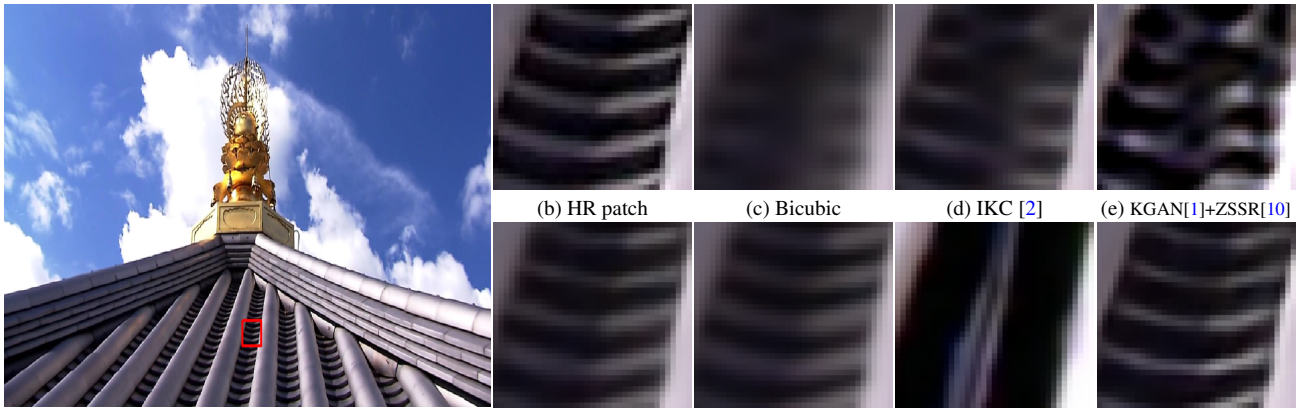
Figure 10. Temporal consistency property. The proposed algorithm generates the high-resolution videos with a better temporal consistency property. *Please view this figure using the Adobe Acrobat Reader as it contains videos.*



(a) Ground truth HR frame (b) HR patch (c) Bicubic (d) IKC [2] (e) KGAN[1]+ZSSR[10] (f) DUF [5]

(g) RBPN [3] (h) EDVR [15] (i) TGA [4] (j) TDAN [14] (k) Ours-L

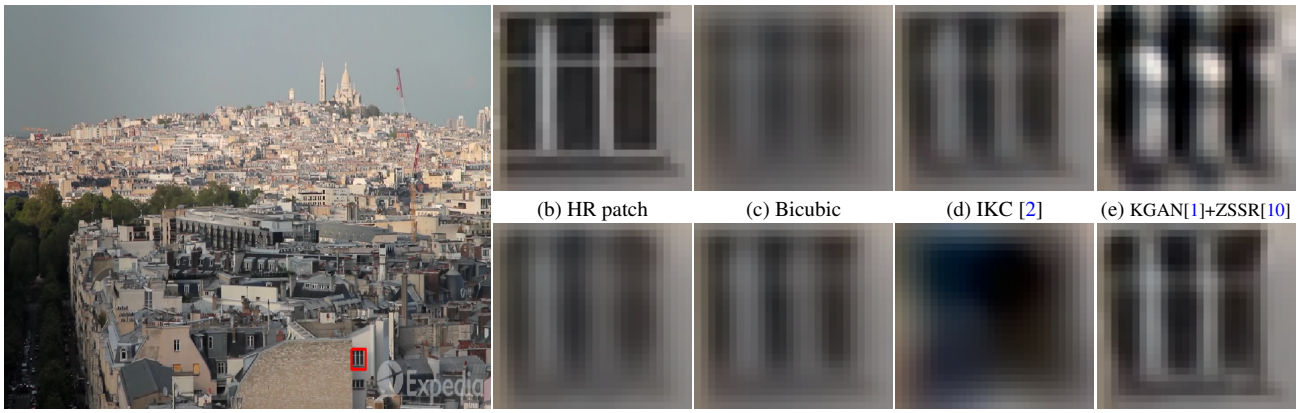
Figure 11. Video SR result ($\times 4$) on the REDS dataset [8], where the realistic blur kernels [1] are used in the degradation process. The proposed algorithm recovers high-quality frames with clearer structures.



(a) Ground truth HR frame (b) HR patch (c) Bicubic (d) IKC [2] (e) KGAN[1]+ZSSR[10]

(f) DUF [5] (g) RBPN [3] (h) Deblur[9]+EDVR[15] (i) Ours

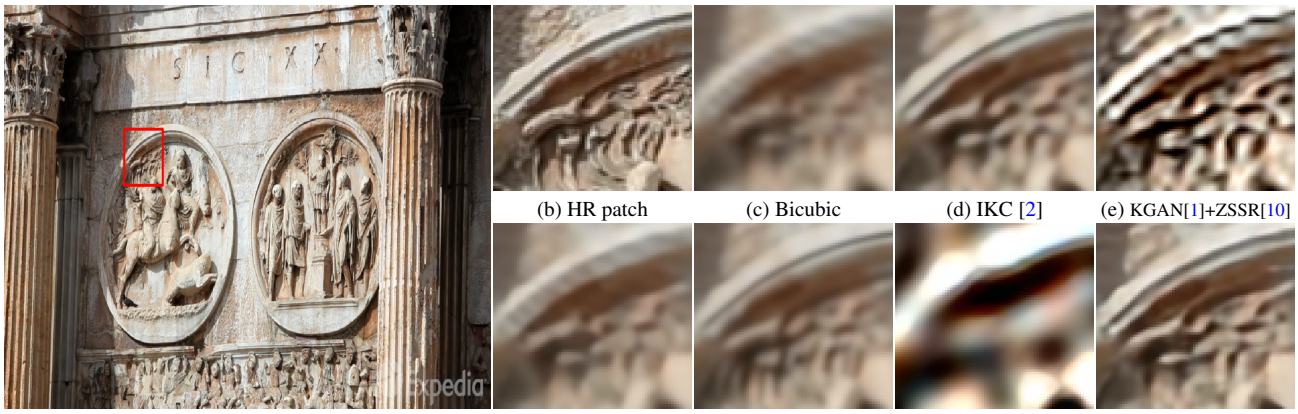
Figure 12. Video super-resolution result ($\times 4$) on the SPMCS dataset [13], where the blur kernels in the degradation process are unknown Gaussian kernels. The proposed algorithm recovers high-quality frames with clearer structures.



(a) Ground truth HR frame (b) HR patch (c) Bicubic (d) IKC [2] (e) KGAN[1]+ZSSR[10]

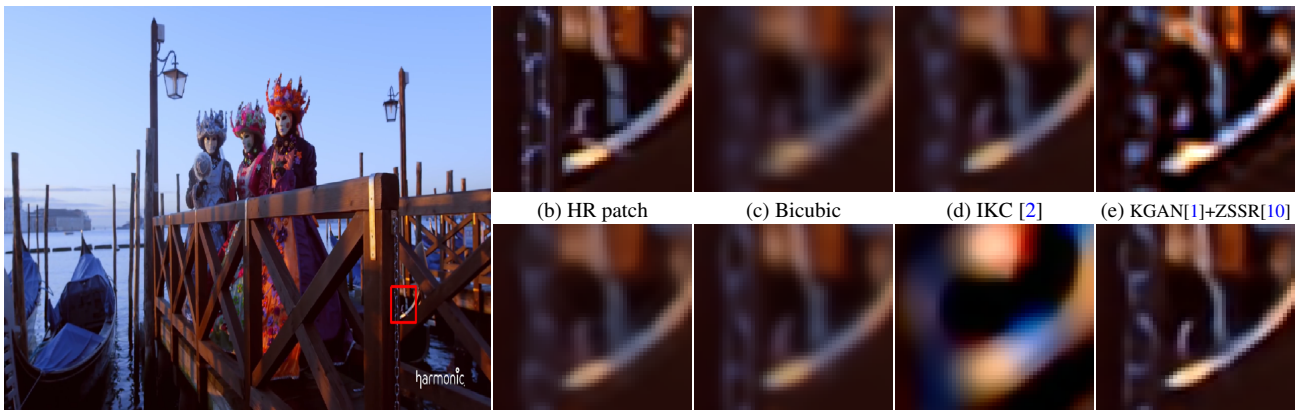
(f) TOFlow [16] (g) RBPN [3] (h) Deblur[9]+EDVR[15] (i) Ours

Figure 13. Video super-resolution result ($\times 4$) on the SPMCS dataset [13], where the blur kernels in the degradation process are unknown Gaussian kernels. The proposed algorithm recovers high-quality frames with clearer structures.



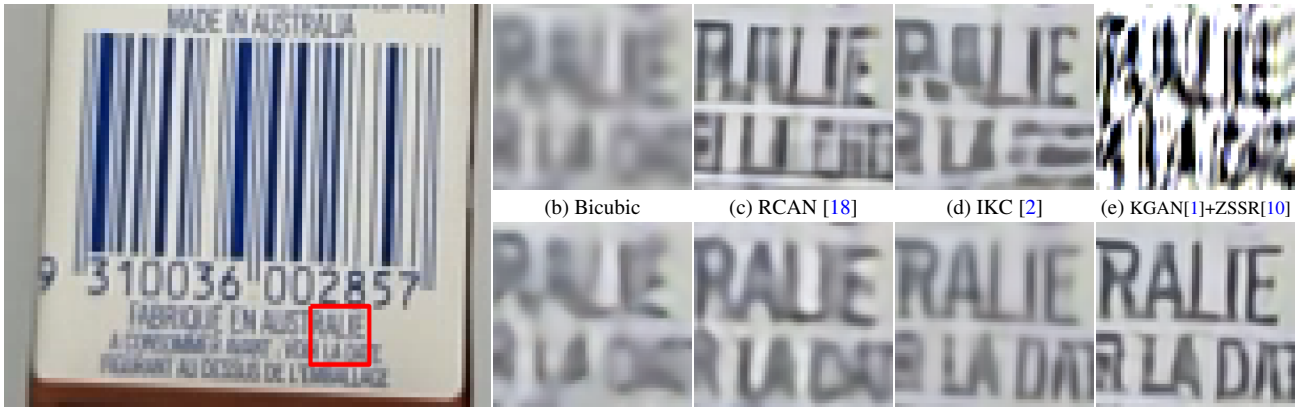
(a) Ground truth HR frame (b) HR patch (c) Bicubic (d) IKC [2] (e) KGAN[1]+ZSSR[10]
(f) TOFlow [16] (g) RBPN [3] (h) Deblur[9]+EDVR[15] (i) Ours

Figure 14. Video super-resolution result ($\times 4$) on the SPMCS dataset [13], where the blur kernels in the degradation process are unknown Gaussian kernels. The proposed algorithm recovers high-quality frames with clearer structural details.



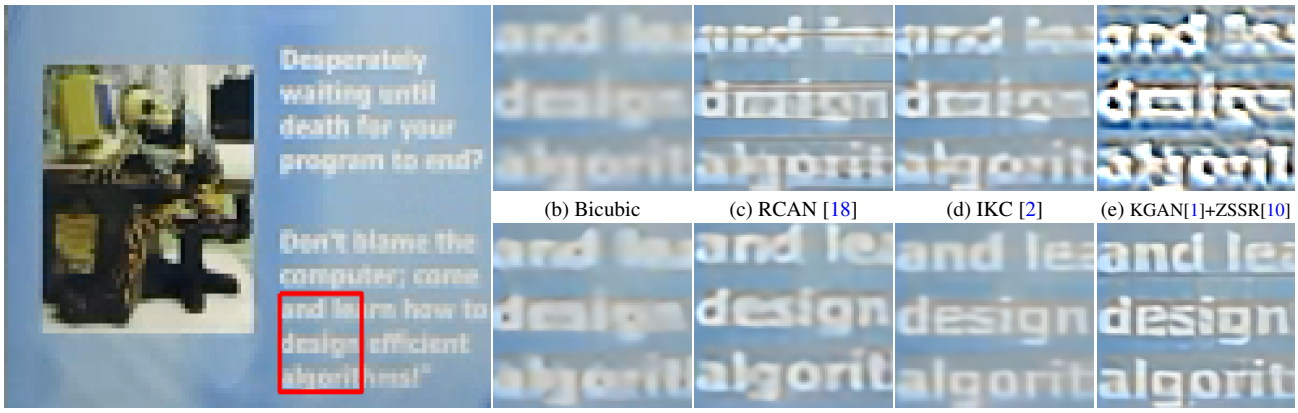
(a) Ground truth HR frame (b) HR patch (c) Bicubic (d) IKC [2] (e) KGAN[1]+ZSSR[10]
(f) TOFlow [16] (g) RBPN [3] (h) Deblur[9]+EDVR[15] (i) Ours

Figure 15. Video super-resolution result ($\times 4$) on the SPMCS dataset [13], where the blur kernels in the degradation process are unknown Gaussian kernels. The proposed algorithm recovers high-quality frames with clearer structures.



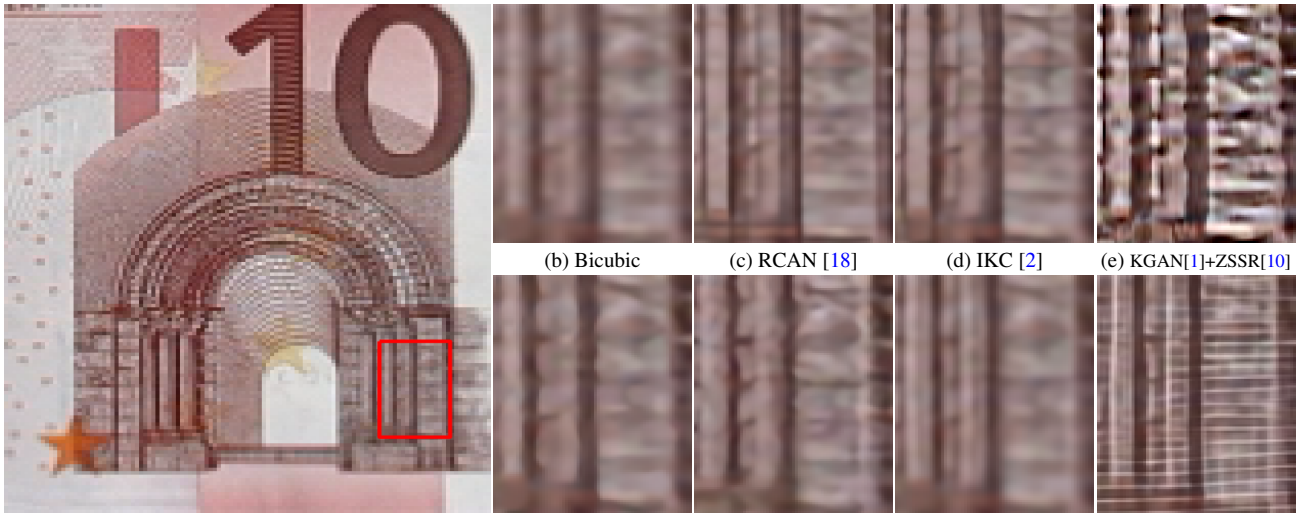
(a) LR frame (b) Bicubic (c) RCAN [18] (d) IKC [2] (e) KGAN[1]+ZSSR[10]
(f) MZSR [11] (g) TOFlow [16] (h) RBPN [3] (i) Ours

Figure 16. Video super-resolution results ($\times 4$) on a real low-resolution video. The proposed algorithm generates the frame with clearer characters.



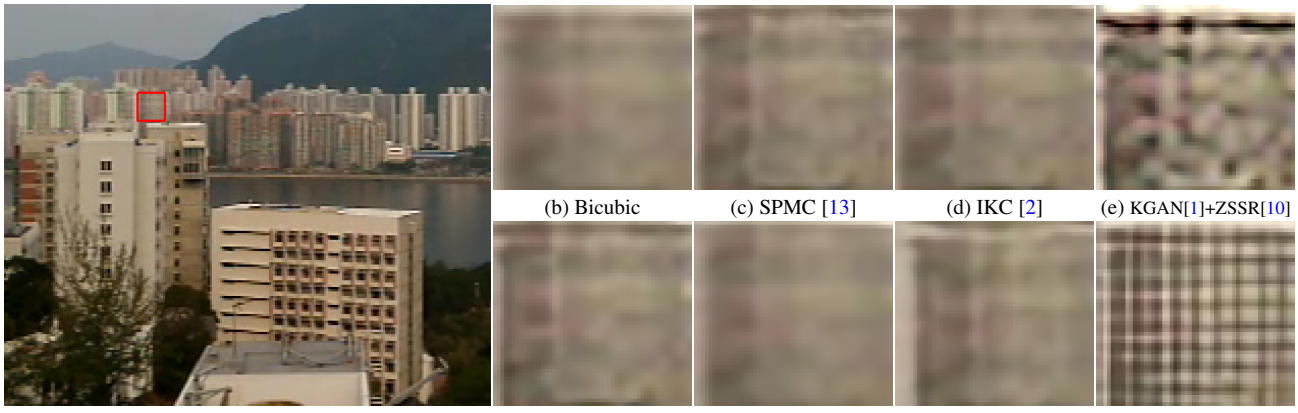
(a) LR frame (b) Bicubic (c) RCAN [18] (d) IKC [2] (e) KGAN[1]+ZSSR[10]
(f) MZSR [11] (g) TOFlow [16] (h) RBPN [3] (i) Ours

Figure 17. Video super-resolution results ($\times 4$) on a real low-resolution video. The proposed algorithm generates the frame with clearer characters.



(a) LR frame (b) Bicubic (c) RCAN [18] (d) IKC [2] (e) KGAN[1]+ZSSR[10]
(f) MZSR [11] (g) TOFlow [16] (h) RBPN [3] (i) Ours

Figure 18. Video super-resolution results ($\times 4$) on a real low-resolution video. The proposed algorithm generates much clearer frames with images with finer detailed structures.



(a) Input (b) Bicubic (c) SPMC [13] (d) IKC [2] (e) KGAN[1]+ZSSR[10]
(f) DUF [5] (g) TOF [16] (h) RBPN [3] (i) Ours

Figure 19. Results ($\times 4$) on a real low-resolution video. The proposed algorithm generates much clearer frames with finer structural details.