Blind Video Super-Resolution based on Implicit Kernels
——— Supplementary Materials ———

In the supplementary materials, we first present additional experimental results for a comprehensive comparison in Section A. Then, we provide more analysis for our implicit modules in Section B.

A. Additional Experimental Results

A.1. Results for perceptual quality

To further evaluate the perceptual quality of the BVSR models, we apply three representative perceptual quality metrics, i.e., LPIPS [17], FID [4], and DISTS [5], to assess four competitive BVSR methods (DBVSR [11], BSVSR [12], Self-BVSR [1], FMA-Net [15]) and our BVSR-IK model on two degradation scenarios (Gaussian blur and realistic motion blur). Their experimental results are provided in Table 1. It can be found from Table 1 that our BVSR-IK model offers best perceptual quality results in most cases and offers second-best results in a few cases. In particular, our BVSR-IK model offers the best perceptual quality results in terms of LPIPS on three testing datasets for two degradation scenarios. These results demonstrate that our BVSR-IK model can achieve a state-of-the-art perceptual quality.

A.2. Results for real-world scenario

It is noted that different environmental factors like object/camera motions, depth of field variation, and camera defocusing can potentially result in spatially varying degradations in the real world [6]. Several real-world super-resolution (SR)/video super-resolution (VSR) datasets have been proposed to simulate these degradations, such as RealSR [2], VideoLQ [3], and RealVSR [13].

We provide the results of our BVSR-IK model for real-world complex blur kernels scenario. To evaluate our proposed implicit kernels in real-world VSR methods, we choose a recent state-of-the-art real-world VSR method, i.e., Realviformer [18], as a basic compared method. We first replace the channel attention fusion (CAF) module of Realviformer [18] with our implicit kernel dictionary to construct the Realviformer-IK model, evaluating our implicit kernels in existing real-world VSR method. Secondly, we trained our full BVSR-IK model for the real-world scenario to evaluate real-world VSR performance.

We evaluate three models, i.e., original Realviformer, Realviformer-IK and our full BVSR-IK models, in terms of NRQM [8], ILNIQE [16], BRISQUE [9] on on two real-world VSR datasets, VideoLQ [3] and RealVSR [13]. The corresponding results are reported in Table 2. We can also observe from Table 2 that evident gains achieved by two our proposed models over Realviformer. Additionally, our BVSR-IK model achieves more significant real-world VSR performance gain.

B. More Analysis

B.1. Relevance of implicit modules

To showcase the relevance of each implicit module, we provide visualization examples of implicit kernel atoms and coefficient weights $\boldsymbol{\mu}$ and $\boldsymbol{\omega}$ of our Implicit Spatial Correction (ISC) and Implicit Temporal Alignment (ITA) modules in Figure 1. Specifically, kernels with three different scales, i.e., 5×5, 9×9, and 13×13, were first visualized in Figure 1 (left). It can be found from Figure 1 (left) that the kernel atoms within the same scale contain diverse patterns, with similar shapes across scales, which shows that implicit kernel atoms are associated with strong isotropy. Additionally, we provide three pairs of coefficient weights of our ISC and ITA (forward propagation) modules, i.e., μ_1, μ_3, μ_5 and ω_1, ω_3, ω_5 in Figure 1 (right). It illustrates the adaptation to scene depth ($\boldsymbol{\mu}$) and image ($\boldsymbol{\omega}$). These results effectively present the advantage of our implicit modules.

References

[1] Haoran Bai and Jinshan Pan. Self-supervised deep blind video super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1, 2

[2] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3086–3095, 2019. 1

[3] Kelvin CK Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. Investigating tradeoffs in real-world video super-resolution. In *Proceedings of the IEEE/CVF*

Table 1. The results in terms of LPIPS, FID, and DISTS on three testing datasets.

Scenarios	Methods	REDS4 [10] LPIPS↓ / FID↓ / DISTS↓	Vid4 [7] LPIPS↓ / FID↓ / DISTS↓	UDM10 [14] LPIPS↓ / FID↓ / DISTS↓
Gaussian Blur	DBVSR [11]	0.2665 / 16.34 / 0.1296	0.3285 / 76.64 / 0.1877	0.1624 / 28.67 / 0.1327
	BSVSR [12]	0.2692 / 23.13 / 0.1200	0.3142 / 74.95 / 0.1821	0.1763 / 25.42 / 0.1295
	Self-BVSR [1]	0.2679 / 16.72 / 0.1279	0.3099 / 82.43 / 0.1743	0.1187 / 23.34 / 0.0893
	FMA-Net [15]	0.2527 / 15.49 / **0.1122**	0.3326 / **69.40** / 0.1718	0.1958 / 29.47 / 0.1106
	BVSR-IK (ours)	**0.2402 / 12.09** / 0.1238	**0.2976** / 73.14 / **0.1682**	**0.1164 / 19.87 / 0.0812**
Realistic Motion Blur	DBVSR [11]	0.3509 / 22.31 / 0.1692	0.3254 / 72.98 / 0.1895	0.1508 / 20.76 / 0.0975
	BSVSR [12]	0.3113 / 20.12 / 0.1441	0.3279 / 78.31 / 0.1772	0.1442 / 19.63 / 0.0892
	Self-BVSR [1]	0.2750 / 17.70 / 0.1313	0.3165 / 78.20 / 0.1694	0.1289 / 19.27 / 0.0843
	FMA-Net [15]	0.2726 / 16.90 / 0.1290	0.3052 / **61.04 / 0.1490**	0.1266 / 18.93 / **0.0747**
	BVSR-IK (ours)	**0.2466 / 12.61 / 0.1266**	**0.2884** / 66.52 / 0.1683	**0.1127 / 18.35** / 0.0811

Table 2. The results in terms of NRQM, ILNIQE, and BRISQUE on two testing datasets for real-world video super-resolution.

Datasets	Realviformer [18] NRQM↑/ILNIQE↓/BRISQUE↓	Realviformer-IK NRQM↑/ILNIQE↓/BRISQUE↓	BVSR-IK (ours) NRQM↑/ILNIQE↓/BRISQUE↓
VideoLQ [3]	6.338 / 25.94 / 25.21	6.368 / 25.76 / 24.36	**6.417 / 25.49 / 24.23**
RealVSR [13]	6.588 / 28.61 / 14.41	6.608 / 28.49 / 14.26	**6.693 / 28.32 / 14.10**

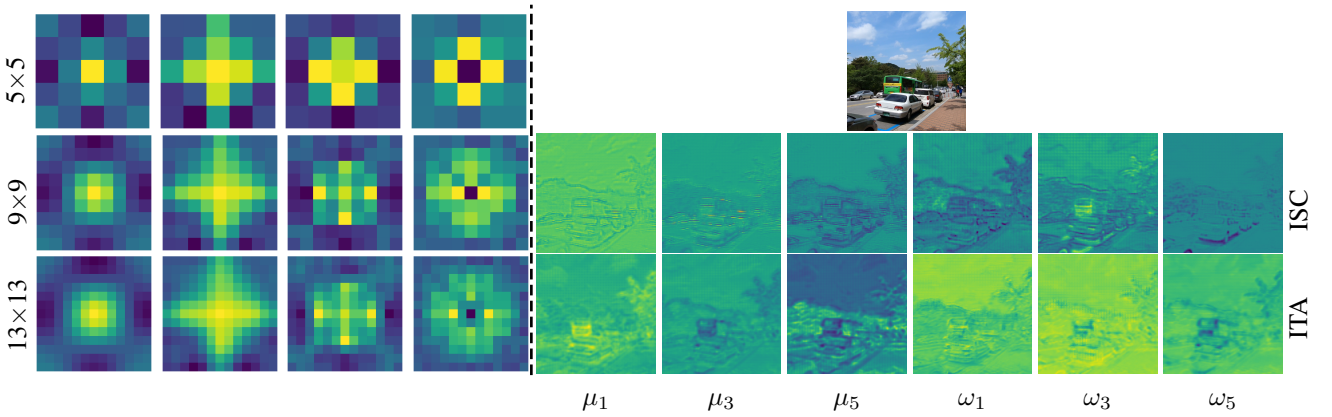

Figure 1. The visualization of implicit kernels (Left) and predicted weights (Right).

Conference on Computer Vision and Pattern Recognition, pages 5962–5971, 2022. 1, 2

[4] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE transactions on pattern analysis and machine intelligence*, 44(5):2567–2581, 2020. 1

[5] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 1

[6] Jingyun Liang, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Mutual affine network for spatially variant kernel estimation in blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4096–4105, 2021. 1

[7] Ce Liu and Deqing Sun. On bayesian adaptive video super resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(2):346–360, 2013. 2

[8] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158:1–16, 2017. 1

[9] Anish Mittal, Anush K Moorthy, and Alan C Bovik. Blind/referenceless image spatial quality evaluator. In *2011 Conference Record of the Forty Fifth Asilomar Conference*

on Signals, Systems and Computers, pages 723–727. IEEE, 2011. 1

[10] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 2

[11] Jinshan Pan, Haoran Bai, Jiangxin Dong, Jiawei Zhang, and Jinhui Tang. Deep blind video super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4811–4820, 2021. 1, 2

[12] Yi Xiao, Qiangqiang Yuan, Qiang Zhang, and Liangpei Zhang. Deep blind super-resolution for satellite video. *IEEE Transactions on Geoscience and Remote Sensing*, 2023. 1, 2

[13] Xi Yang, Wangmeng Xiang, Hui Zeng, and Lei Zhang. Real-world video super-resolution: A benchmark dataset and a decomposition based learning scheme. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4781–4790, 2021. 1, 2

[14] Peng Yi, Zhongyuan Wang, Kui Jiang, Junjun Jiang, and Jiayi Ma. Progressive fusion video super-resolution network via exploiting non-local spatio-temporal correlations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3106–3115, 2019. 2

[15] Geunhyuk Youk, Jihyong Oh, and Munchurl Kim. Fma-net: Flow-guided dynamic filtering and iterative feature refinement with multi-attention for joint video super-resolution and deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 44–55, 2024. 1, 2

[16] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015. 1

[17] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 1

[18] Yuehan Zhang and Angela Yao. Realviformer: Investigating attention for real-world video super-resolution. In *European Conference on Computer Vision*, pages 412–428. Springer, 2024. 1, 2