A Universal Scale-Adaptive Deformable Transformer for Image Restoration across Diverse Artifacts

Supplementary Material

1. The Supplementary Introduction of the Fusion Module

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ denote the features from the encoder of above scale and from the current scale, respectively. The fusion operation is formulated as $\psi_{\text{fusion}}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u} + \Phi(\boldsymbol{u} \odot \boldsymbol{v})$, where $\Phi(\cdot)$ denotes the convolutional layer.

2. Extra Experiment on the RAW Data

We made minor revisions to our SADT to make it adapt to raw inputs, and the revised version is denoted as SADT*. Specifically, the input-output dimensions were adjusted and the final unsampling layers were added for each scale. For a fair comparison, we utlize the same loss function and training strategy as video-based RVDNet [1] to retrain SADT* on the raw image demoiréing dataset [2] and the raw video demoiréing dataset RawVDemoiré [1]. As shown in Tab. 1, SADT*, as an image-based method, maintains SOTA performance on both the raw image demoiréing dataset and even the raw video demoiréing dataset.

Dataset	Method	PSNR(dB)↑	SSIM↑	LPIPS↓
Raw image dataset [2]	RVDNet	27.26	**0.9346**	0.0748
	SADT*	**27.67**	**0.9346**	**0.0682**
RawVDemoiré [1]	RVDNet	28.71	**0.9201**	0.0904
	SADT*	**29.46**	**0.9201**	**0.0897**

Table 1. Comparison with the state-of-the-art raw data demoiréing method RVDNet.

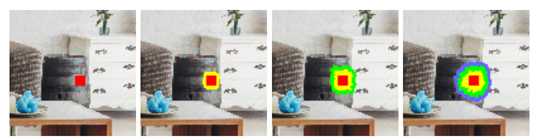
3. More Details on the Ablation Study

Analysis on the architecture design: In the proposed SADT framework, the encoders and decoders are constructed by the combination of SEDC and SADA. In this section, we examine how different architecture designs of the encoder and decoder affect the performance of SADT by experimenting with various combinations of SEDC and SADA. As indicated in Table 2, positioning the two SEDC modules at both ends of the block maximizes their ability to recovering genuine image details.

More details on the ablation study of our key designs: To further analyze the impact of SEDC, we replace it with a ResBlock while maintaining the same parameter count. The model shows 39.44dB/0.9704/0.0499 in PSNR/SSIM/LPIPS, demonstrating the cooperative effectiveness of SCMCs and DCNs. We also evaluate the contribution of two key components, SADA and SEDC, on the FHDMi, SPAD dataset for demoiréing, deraining task, respectively. Replacing SEDC with a single SCMC shows a PSNR drop of 0.32/0.24 (dB); replacing SEDC with a ResBlock shows a PSNR drop of 0.46/0.31 (dB); removing SADA shows a drop of 0.98/0.72 (dB), over the full model. It is consistent to debanding.

4. Visualization of SEDC's receptive field

SEDC is designed for local pattern analysis with expanded receptive fields that are morphologically adaptive to local patterns. Figure 1 visualizes the receptive field dynamics of each SCMC with SEDC. Notably, at deeper levels, the SCMCs exhibit progressively enhanced capabilities in capturing increasingly extensive and irregular receptive fields, enabling local pattern processing with both spatial scale adaptability and shape adaptivity.



0-th SCMC 1-st SCMC 2-nd SCMC 3-rd SCMC

Figure 1. Visual inspection of the receptive field of each SCMC within a window. The red area represents the pristine receptive field of each SCMC, while the yellow, green, purple areas sequentially depict the receptive fields that are expanded by the backpropagation through the 1-st, 2-nd, 3-rd DCNs, respectively.

5. More Visual Results

More visual comparisons for image demoiréing, debanding and deraining are shown in Fig. 2, 3 and 4. As shown, SADT consistently achieves superior visual quality compared to other methods. It effectively removes structured artifacts while preserving finer image details.

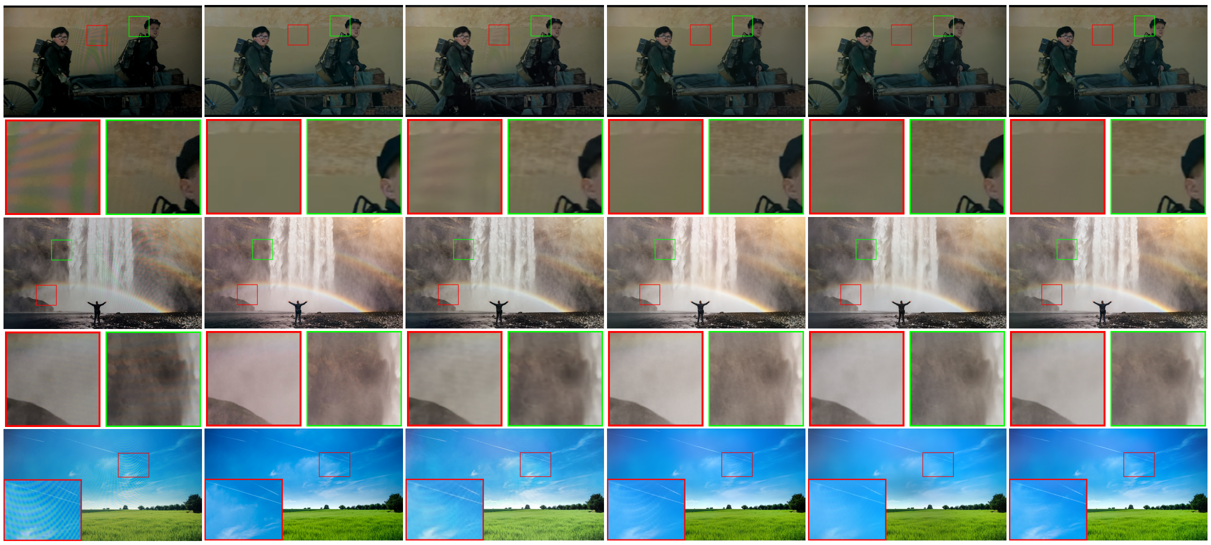
References

[1] Yijia Cheng, Xin Liu, and Jingyu Yang. Recaptured raw screen image and video demoiréing via channel and spatial modulations. *Advances in Neural Information Processing Systems*, 36, 2024. 1

[2] Huanjing Yue, Yijia Cheng, Yan Mao, Cong Cao, and Jingyu Yang. Recaptured screen image demoiréing in raw domain. *IEEE Transactions on Multimedia*, 25:5589–5600, 2022. 1

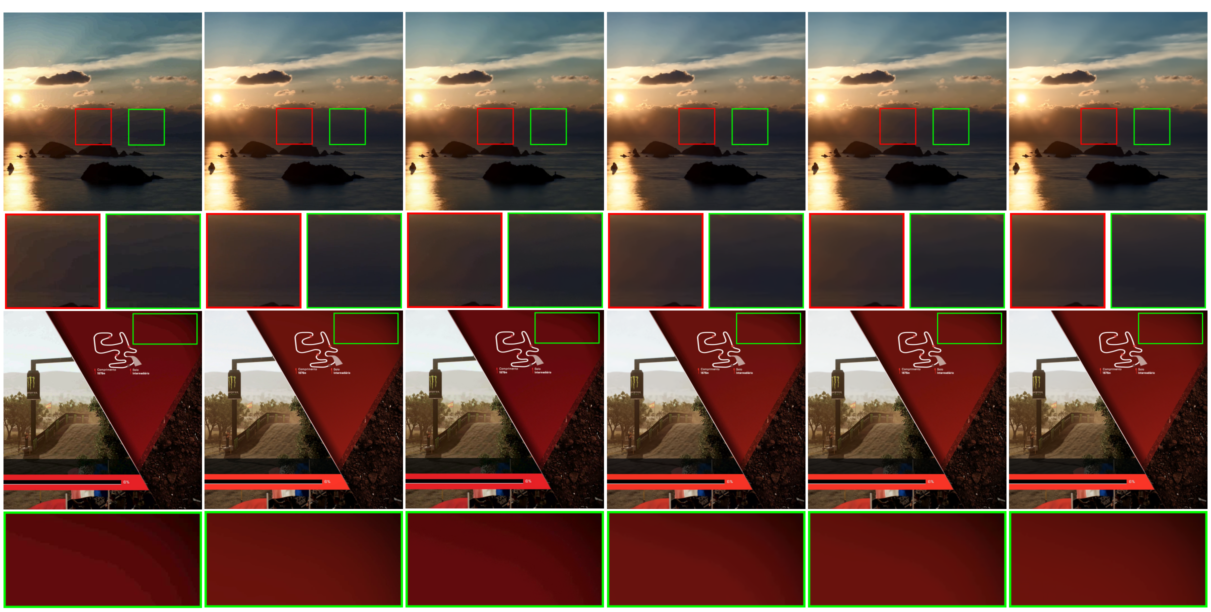
M_s	PSNR(dB)↑	SSIM↑	LPIPS↓
SEDC-SEDC-SADA($\times N_i$)	39.67	0.9717	0.0484
SADA($\times N_i$)-SEDC-SEDC	39.68	0.9716	0.0482
SADA($\times \lfloor N_i/2 \rfloor$)-SEDC-SEDC-SADA($\times \lfloor N_i/2 \rfloor$)	39.72	0.9722	0.0470
SEDC-SADA($\times N_i$)-SEDC	39.78	0.9729	0.0453

Table 2. Analysis for sequence of SEDCs and SADAs within a block.



Degraded | Reference | FHDe^{2}Net | ESDNet | RVDNet | Ours

Figure 2. Visual inspection of the results from different image demoiréing methods on sample images.



Degraded | Reference | FFmpeg | MWCNN | Restormer | Ours

Figure 3. Visual inspection of the results from different image debanding methods on sample images; see zoom-in box for details inspection.

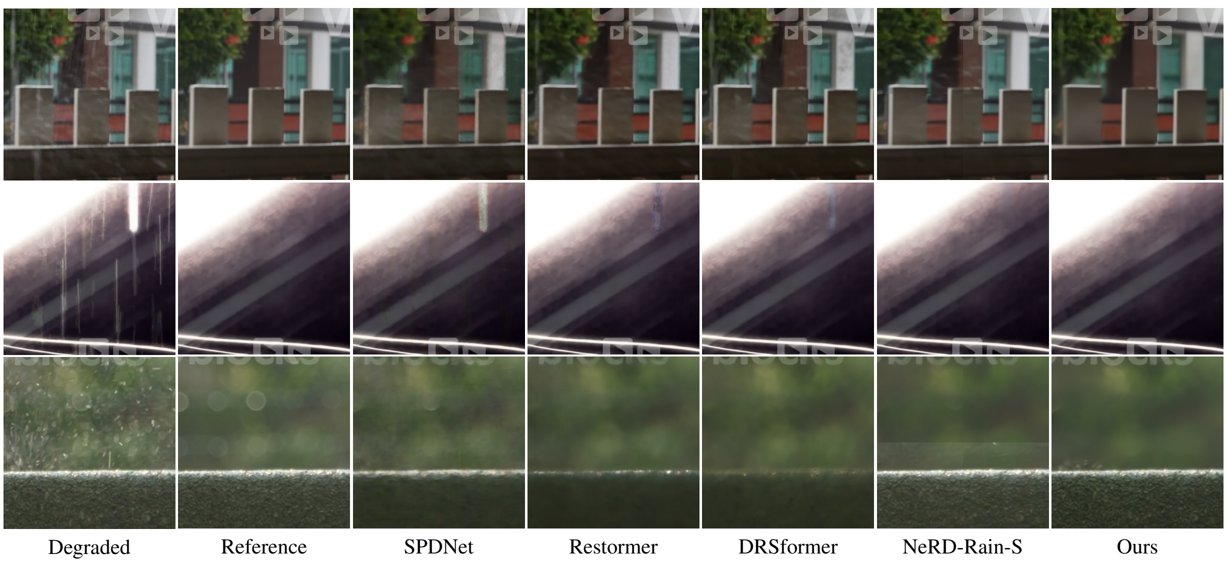


Figure 4. Visual inspection of the results from different image deraining nethods on sampled images.