Diffusion-based Event Generation for High-Quality Image Deblurring

Supplementary Material

1. Introduction

The supplementary materials primarily include the following content:

- The specific structure of the used networks.
- Additional visual results.

2. Details of RDNet

RDNet consists of three main modules: an motion prior encoder, an image encoder, and an image decoder, as shown in Figure 1.

The motion prior encoder. The primary function of the motion prior encoder is to encode the motion prior feature M (or M') into multi-scale features. The motion prior encoder consists of three levels, each containing two residual blocks and a down-sampling block composed of a convolutional layer and a strided convolution layer. The features output by the second residual block at each level are injected as F_m^k into the corresponding level of the image encoder. (k denotes the level number.)

The image encoder. The image encoder encodes the input blurry image B while integrating motion prior information F_m^k and image features F_i^k. The three levels of the image encoder correspond to the three levels of the motion prior encoder. Each level includes a residual block, two dual-attention channel fusion block (DACFB), and a down-sampling block. The two DACFB at each level operate in a sequential manner, receiving the corresponding F_m^k to incorporate motion information into the image features.

The image decoder. The image decoder receives the multi-level image features output by the image encoder and produces the final sharp image S. Each level in the image decoder contains an up-sampling block and two residual blocks, receiving the sum of the features processed by the previous level and the skip connection features as input.

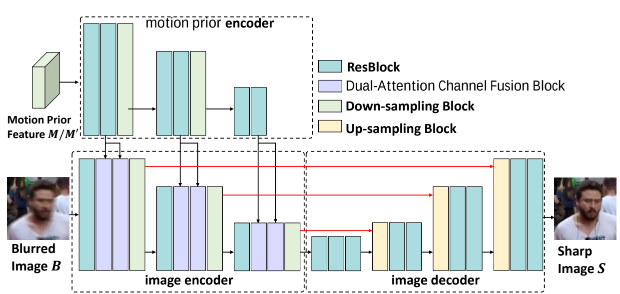

Figure 1. The architecture of the Regression Deblurring Network (RDNet).

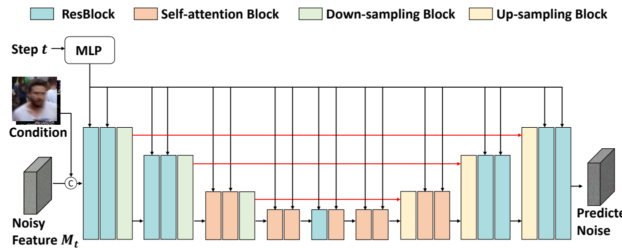

Figure 2. The architecture of the network ϵ_θ in the MPG-Diff.

3. Details of the network ϵ_θ

The network ϵ_θ in the Motion Prior Generation Diffusion Model (MPG-Diff) serves as a noise predictor used in the reverse process. The network ϵ_θ is based on a U-Net architecture, as shown in Figure 2. It takes as input the current iteration step t, the motion prior feature M_t at step t, and other conditions, and outputs the predicted noise for the subsequent reverse denoising process.

The step t is passed to a multilayer perceptron (MLP) to generate the corresponding embedding feature.This embedding feature is passed to the residual blocks and self-attention blocks within the U-Net to provide awareness of the current denoising step.

To balance performance and memory usage, we add self-attention blocks only at the two lowest levels of the U-Net. Each of these self-attention blocks is preceded by a residual block.

During the third training stage, only the network parameters of the self-attention blocks are updated. The gradients of the other network blocks are not recorded, which reduces the memory usage during training.

4. More Visual Comparison Results

We provide additional visual comparisons on benchmark datasets in Figure 3 - 8. We compare our EGDeblurring method with several recent state-of-the-art image deblurring methods, including HINet [1], FFTformer [2], AdaRevD [3], HI-Diff [2], Stripformer [6] and Restormer [7]. For visual results on the GoPro dataset [4], we prioritize using the results or models provided by each respective method. We train all models on the REBlur [5] dataset in a fair manner for comparison.

As shown in Figures 3–6, our method achieves superior detail restoration in complex blurry scenes. As shown in Figures 7–8, our method achieves improved deblurring performance in real-world deblurring scenarios.

Figure 3. Qualitative results on the GoPro dataset. From left to right: blurry image, results from HINet [1], FFTformer [2], AdaRevD [3], HI-Diff [2], Restormer [7], EGDeblurring (ours), and ground truth.

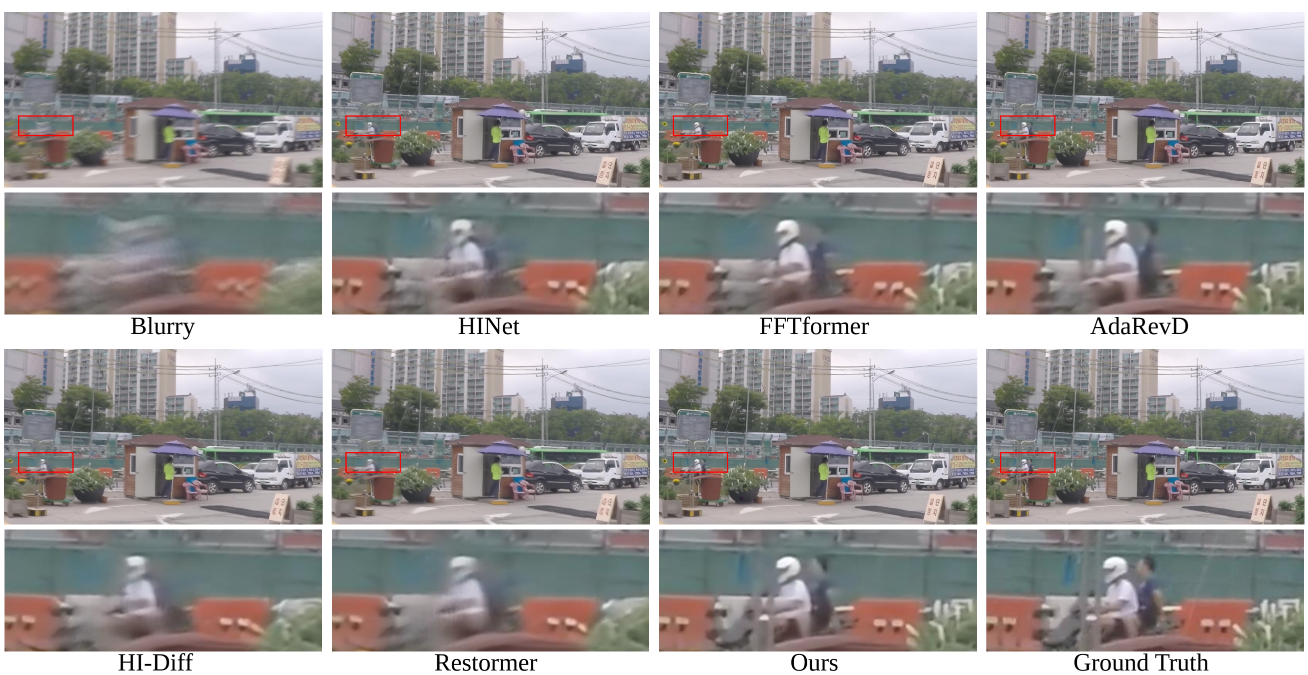

Figure 4. Qualitative results on the GoPro dataset. From left to right: blurry image, results from HINet [1], FFTformer [2], AdaRevD [3], HI-Diff [2], Restormer [7], EGDeblurring (ours), and ground truth.

Figure 5. Qualitative results on the GoPro dataset. From left to right: blurry image, results from HINet [1], FFTformer [2], AdaRevD [3], HI-Diff [2], Restormer [7], EGDeblurring (ours), and ground truth.

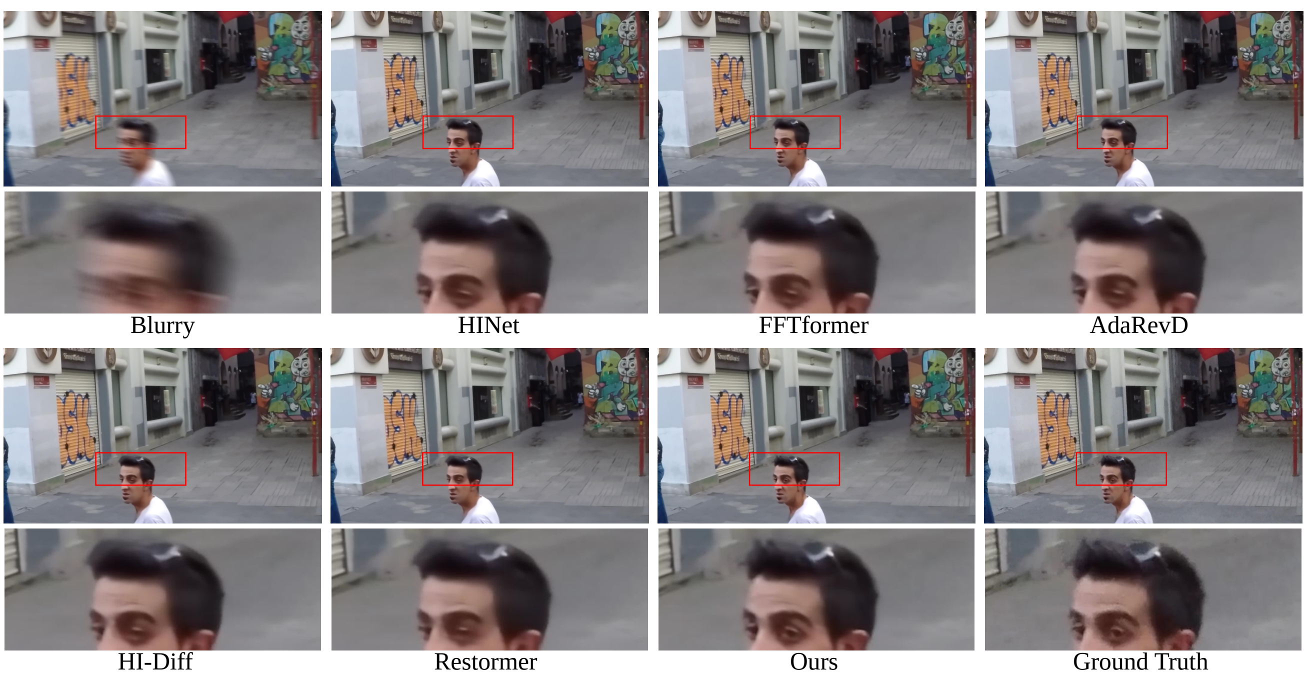

Figure 6. Qualitative results on the GoPro dataset. From left to right: blurry image, results from HINet [1], FFTformer [2], AdaRevD [3], HI-Diff [2], Restormer [7], EGDeblurring (ours), and ground truth.

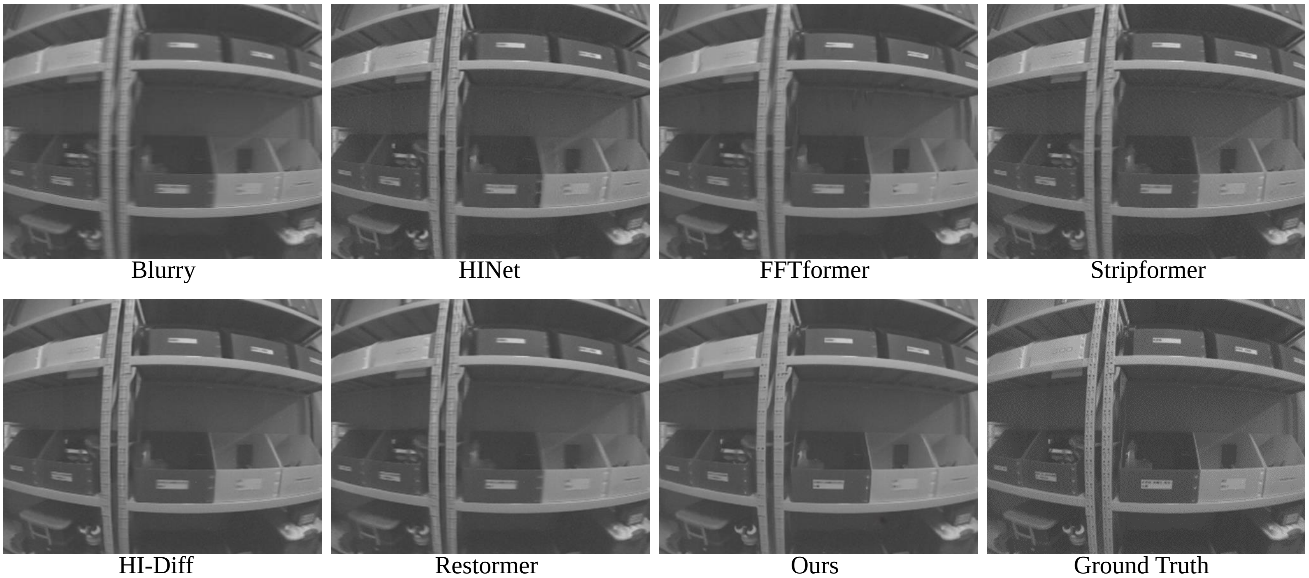

Figure 7. Qualitative results on the REBlur dataset. From left to right: blurry image, results from HINet [1], FFTformer [2], Stripformer [6], HI-Diff [2], Restormer [7], EGDeblurring (ours), and ground truth.

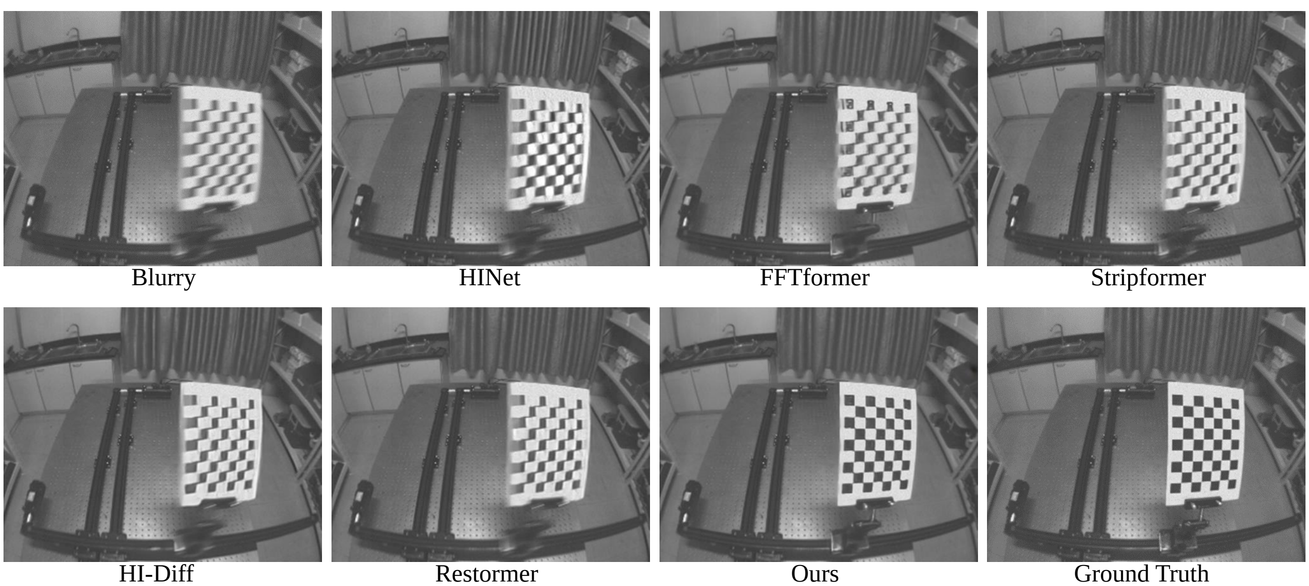

Figure 8. Qualitative results on the REBlur dataset. From left to right: blurry image, results from HINet [1], FFTformer [2], Stripformer [6], HI-Diff [2], Restormer [7], EGDeblurring (ours), and ground truth.

References

[1] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *CVPR*, pages 182–192, 2021. 1, 2, 3, 4

[2] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *CVPR*, pages 5886–5895, 2023. 1, 2, 3, 4

[3] Xintian Mao, Qingli Li, and Yan Wang. Adarevd: Adaptive patch exiting reversible decoder pushes the limit of image deblurring. In *CVPR*, pages 25681–25690, 2024. 1, 2, 3

[4] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, pages 3883–3891, 2017. 1

[5] Lei Sun, Christos Sakaridis, Jingyun Liang, Qi Jiang, Kailun Yang, Peng Sun, Yaozu Ye, Kaiwei Wang, and Luc Van Gool. Event-based fusion for motion deblurring with cross-modal attention. In *ECCV*, pages 412–428. Springer, 2022. 1

[6] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *ECCV*, pages 146–162. Springer, 2022. 1, 4

[7] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pages 5728–5739, 2022. 1, 2, 3, 4