Event-guided Unified Framework for Low-light Video Enhancement, Frame Interpolation, and Deblurring
–*Supplementary materials*–

Taewoo Kim
KAIST
intelpro@kaist.ac.kr

Kuk-Jin Yoon
KAIST
kjyoon@kaist.ac.kr

1. Results on the real-world low-light conditions.

To assess the generalization performance of the model trained on the RELBFI dataset in real-world low-light scenarios, such as nighttime environments without ND filters, we collected real-world data comprising real low-light videos, including low-light events. When capturing real-world low-light data, we used different settings for exposure time, gain, and white balance compared to those used for the RELBFI dataset, and we employed different cameras than those originally used. The visual results on the real-world low-light blurred images are presented in Fig. 1 and Fig. 2.

2. RELBFI dataset samples

When capturing the RELBFI dataset, the RGB camera captured images at an original resolution of 1440×1080, while the event camera recorded events at a resolution of 1280×720. To address issues arising from differing camera resolutions and coordinate transformations, we performed homography-based calibration. The RELBFI dataset consists of a total of 67 sequences, with 47 sequences in the training set and 20 sequences in the test set. More samples from the RELBFI dataset are illustrated in Fig. 3 (training set) and Fig. 4 (test set).

3. Details of IS-TFF encoder.

Due to space limitations in the main paper, the detailed network architecture of the IS-TFF encoder is provided in Fig. 5. As shown in the figure, the basic structure of the IS-TFF forward encoder at scale factor s consists of a 3×3 convolutional layer, a ResBlock, and an IS-TFF block. For the forward encoder, it additionally takes as input the feature information $\mathcal{G}(\mathcal{T})^{fw}_{s,k-1}$ from the previous timestamp $k-1$ to compute temporal dependencies. The backward IS-TFF encoder has the same structure but uses different weights.

4. Loss functions

As shown in Fig.3 in the main paper, we perform estimation for a total of N frames within the shutter interval $\{\mathcal{S}_0, \ldots, \mathcal{S}_{N-1}\}$. Subsequently, we apply the Charbonnier [1] loss function to these N frames.

$$L_{total} = \sum_{k=0}^{N-1} \sqrt{\|\mathcal{S}_k - \mathcal{S}_{k,gt}\| + \epsilon^2} \tag{1}$$

where $\mathcal{S}_k$ represents the estimated normal-light sharp image at time step k, and $\mathcal{S}_{k,gt}$ represents the normal-light ground truth sharp image at time step k. We empirically set ϵ to 10^{-3}.

5. Implementation details

We set the batch size to 8 and the learning rate to 2×10^{-4}. The model is optimized using the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We train our models over 200 epochs. The implementation is based on PyTorch, and all experiments are conducted on NVIDIA 3090 GPU. To augment the training data, we apply random cropping at the same positions for both low-light blurred videos and event voxel data, resulting in cropped blurred patches and voxel bins of size 256×256. For quantitative evaluation, we used standard metrics such as PSNR and SSIM [3].

6. More qualitative results on the RELBFI dataset.

In Fig. 6, we present additional qualitative results from the RELBFI dataset. The figure compares our approach with state-of-the-art event-guided blurry frame interpolation methods, REFID [2].

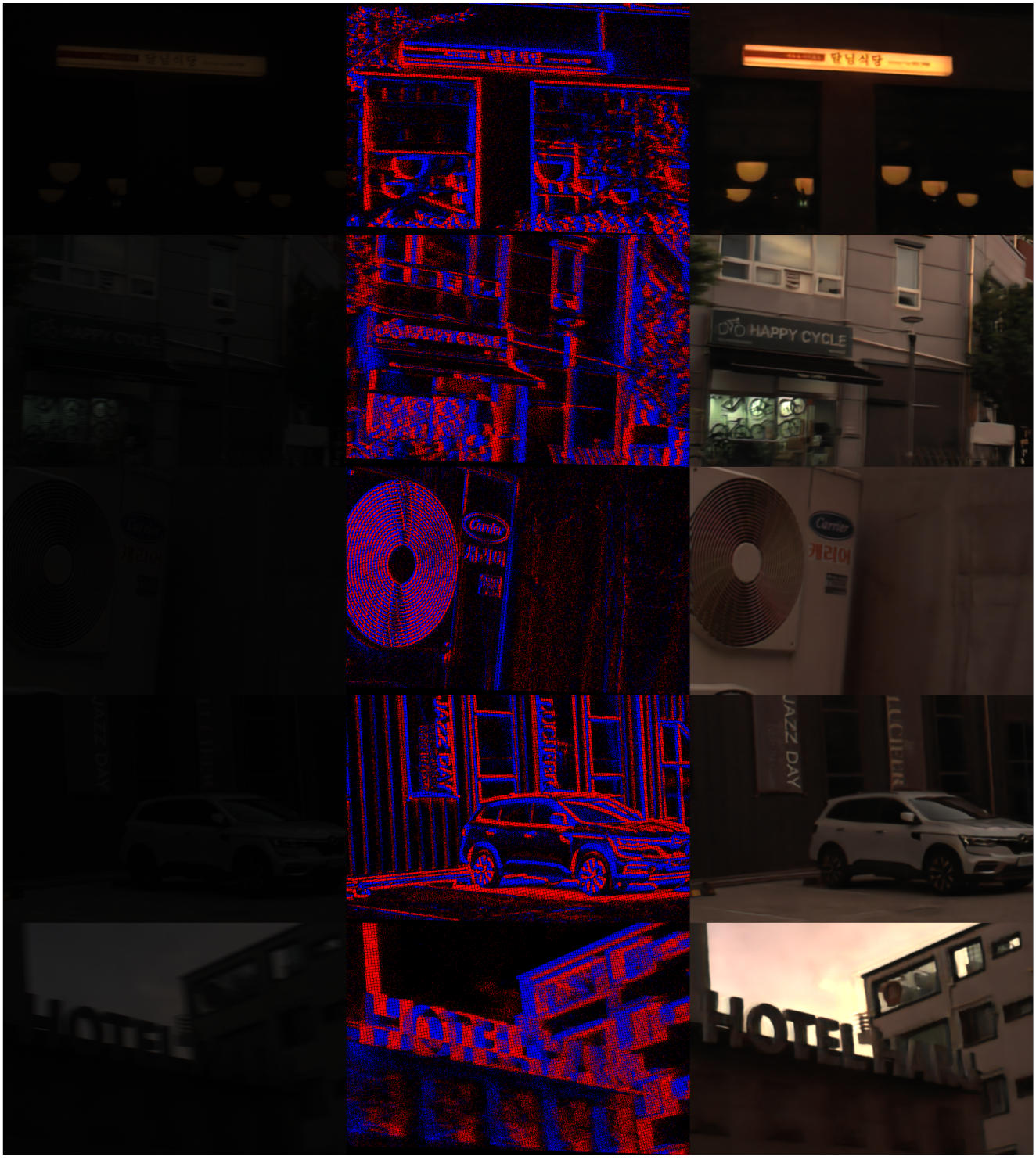
References

[1] Pierre Charbonnier, Laure Blanc-Feraud, Gilles Aubert, and Michel Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of*

1st international conference on image processing, pages 168–172. IEEE, 1994. 1

[2] Lei Sun, Christos Sakaridis, Jingyun Liang, Peng Sun, Jiezhang Cao, Kai Zhang, Qi Jiang, Kaiwei Wang, and Luc Van Gool. Event-based frame interpolation with ad-hoc deblurring. *arXiv preprint arXiv:2301.05191*, 2023. 1, 8

[3] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 1



(a) (b) (c)

Figure 1. Qualitative results on the real-world low-light blurred videos. (a) Low-light blurred image, (b) Low-light events, and (c) Ours.

Figure 2. Qualitative results on the real-world low-light blurred videos. (a) Low-light blurred image, (b) Low-light events, and (c) Ours.

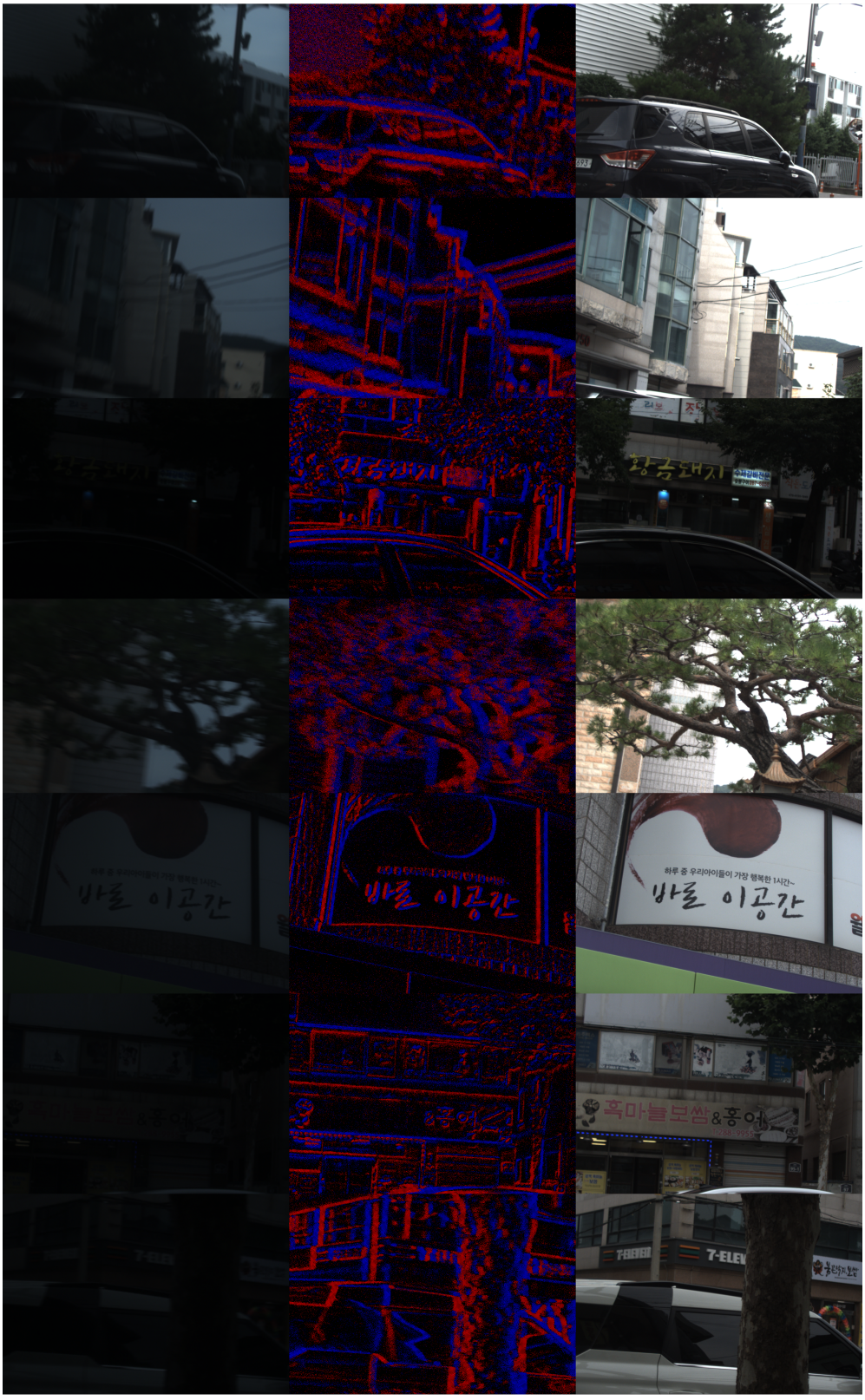


Figure 3. The examples of our RELBFI dataset.(train sets). Our dataset captures a wide variety of scenes with diverse motions, dynamic objects, and rich textures.

Figure 4. The examples of our RELBFI dataset.(tests sets). Our dataset captures a wide variety of scenes with diverse motions, dynamic objects, and rich textures.

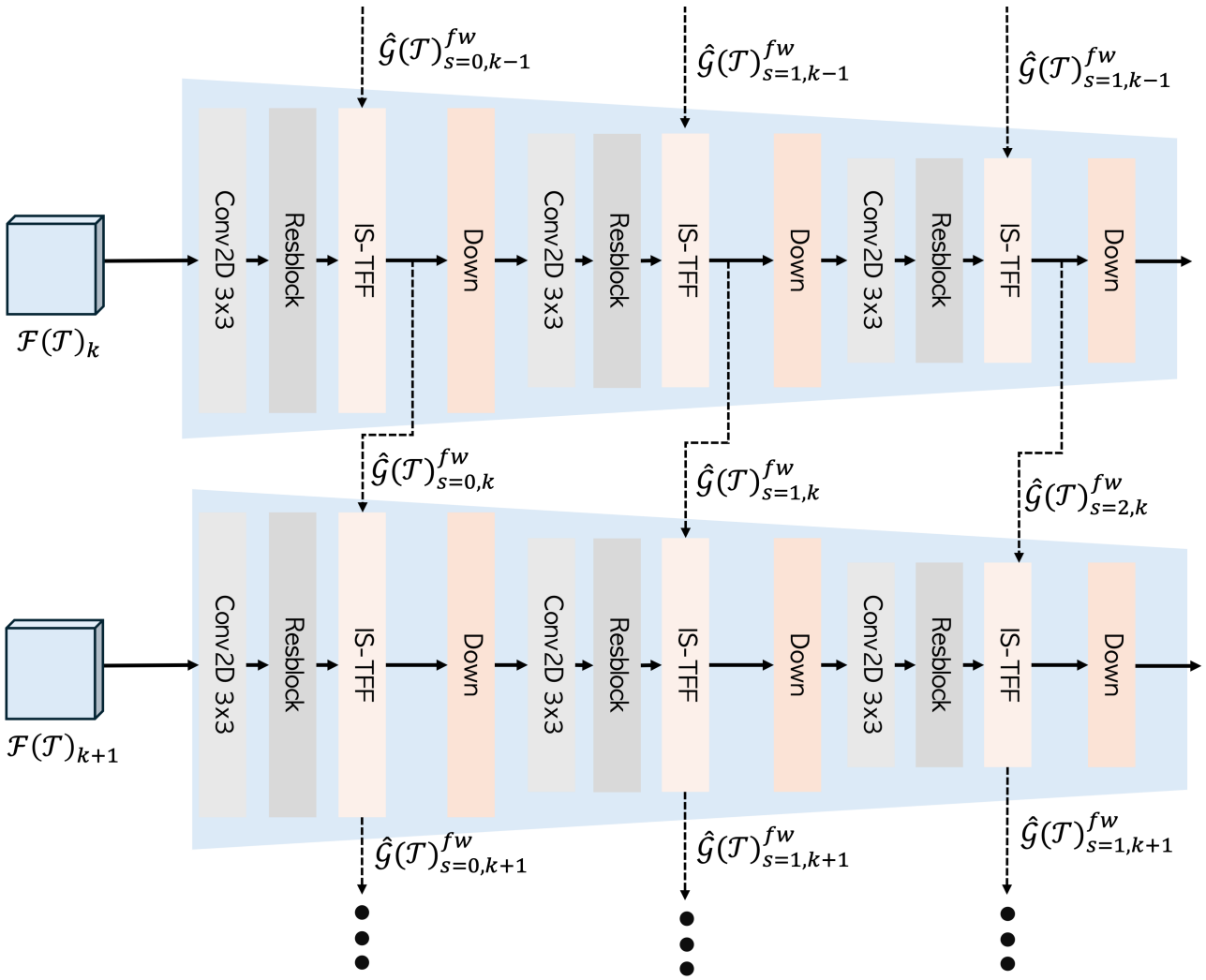


Figure 5. Network details of IS-TFF forward encoder.

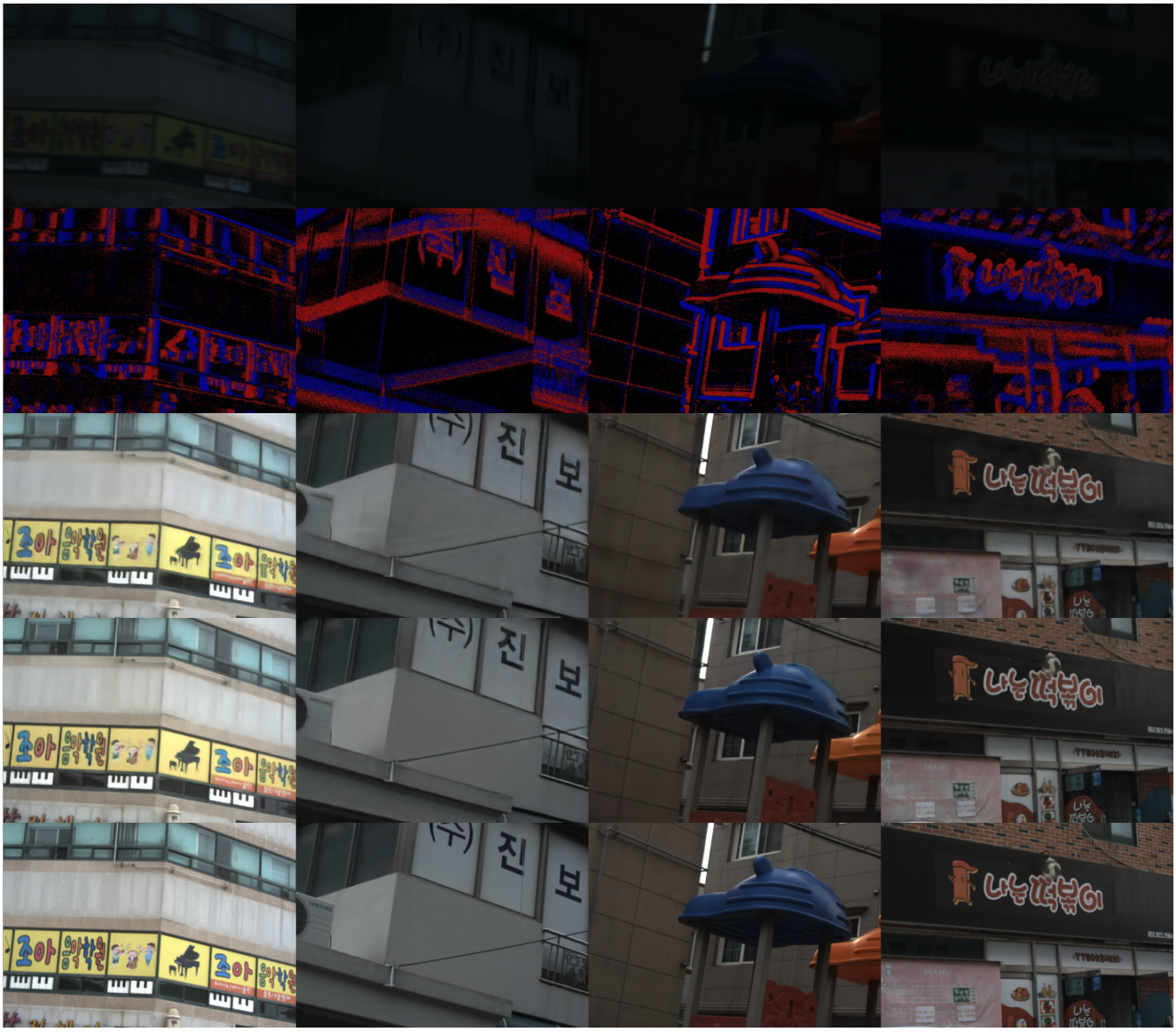


Figure 6. Qualitative results on the RELBFI dataset. From top to bottom: low-light blurred image, low-light events, REFID [2], and Ours, GT. Zoom in for better visualization.