RT-X NET: CROSS ATTENTION TRANSFORMER USING RGB AND THERMAL IMAGES FOR LOW-LIGHT IMAGE ENHANCEMENT

This supplementary material presents the following details which we could not include in the main paper due to space constraints. The additional references for this elaboration are also added here.

Contents

1. EXPERIMENTAL SETTINGS

1.1. V-TIEE dataset

In the V-TIEE dataset, we have systematically captured RGB and thermal images under varying gain and exposure conditions. Specifically, images were recorded at two distinct gain settings, each with five different exposure values. Furthermore, in four distinct scenes, we have expanded the exposure range to include ten different values for each gain setting. The gain values in our dataset span from 0 dB to 44.99 dB. Scenes include lower gain settings at 0, 4.99, 19.99, 23.99, and 24.99 dB, as well as higher gain settings at 23.99 and 44.99 dB. The exposure times range from as short as 1/20000 seconds to as long as 10 seconds. This extensive range of exposure settings can also facilitate the dataset's utility in generating High Dynamic Range (HDR) [1] images, providing a robust resource for HDR imaging research and applications.

1.2. Mutiple exposure of V-TIEE dataset in high and low gain conditions.

This extensive range of exposure settings can also facilitate the dataset's utility in generating High Dynamic Range (HDR) [1] research and applications in lot light image datasets. Since this dataset is not used for training, performance on this evaluation dataset would demonstrate the generalization of the models. In practical scenarios, noise levels increase as the exposure of a scene decreases. Our V-TIEE dataset captures images under various gain conditions, thus incorporating noise characteristics that are similar to real-world settings.

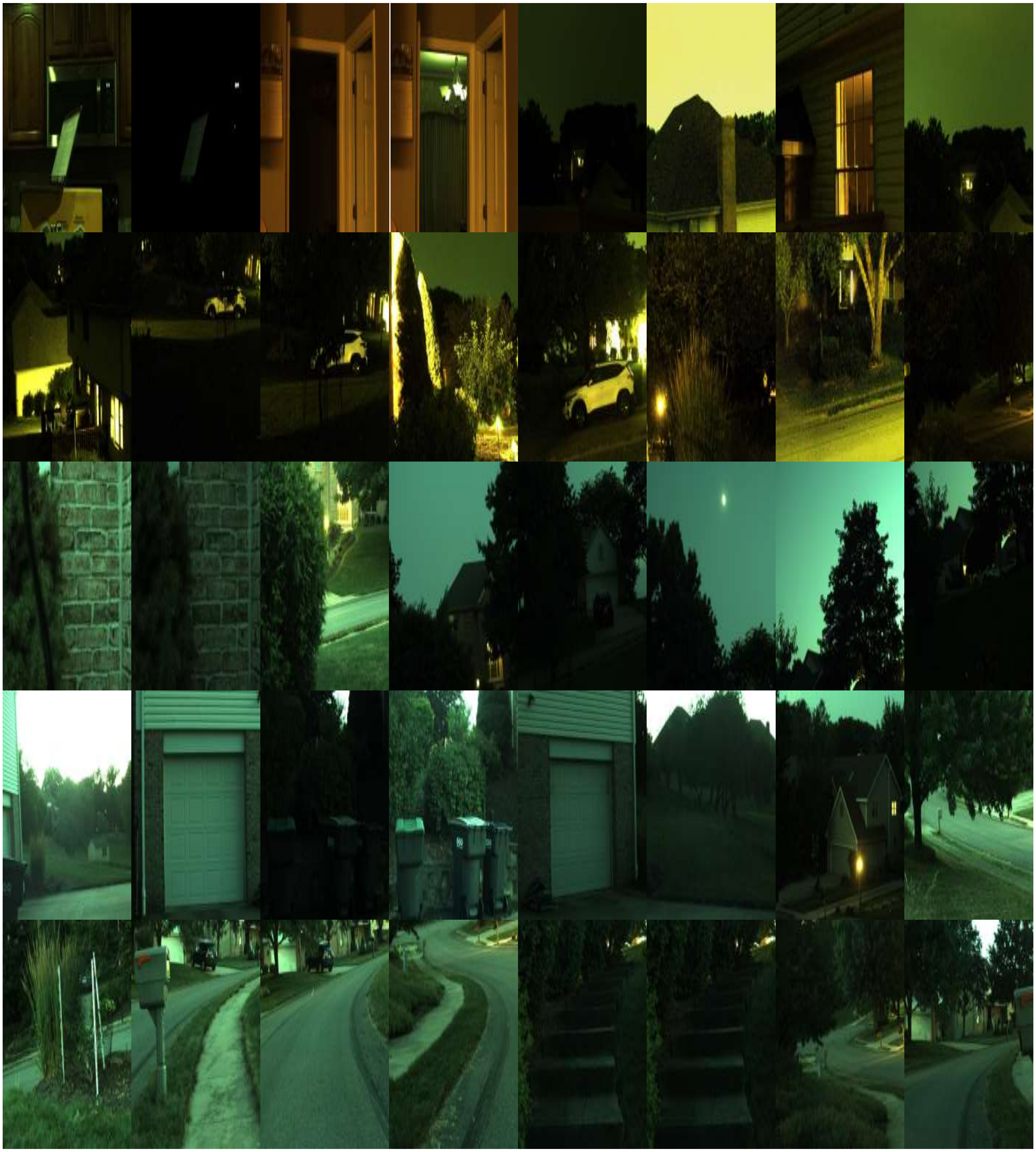

Fig. 1. The figure shows all the indoor and outdoor scenes of RGB and Thermal images in the V-TIEE dataset which we captured in real-time for the low light image enhancement. For space constraints, we are only showing RGB images.

1.3. Visual representation of real-world V-TIEE dataset

In practical scenarios, noise levels increase as the exposure of a scene decreases. Our V-TIEE dataset captures images under various gain conditions, thus incorporating noise characteristics typical of real-world settings. The low-light input images, depicted in Fig. 2, were utilized in our experiments. For enhanced visual comprehension, amplified versions of these inputs are also presented. Additionally, we provide a well-exposed image of the same scene with the thermal image.

1.4. Noise incorporation in simulated low-light LLVIP dataset

Noise model: [2] To encapsulate the fundamental characteristics of noise, it is considered a variable with a mean of zero and variance from two independent sources. Specifically, the following representation applies to pixels below the saturation threshold:

$$Var(n) = \phi t/g^2 + \sigma_{read}^2/g^2 \tag{1}$$

Here, g represents the sensor gain. The initial component describes the Poisson distribution of photon arrival, directly proportional to the accumulated photon count ϕt. The final component, representing the pre-amplification stage, accounts for noise from the sensor's readout process. We applied this method to add noise to the low-light synthetic LLVIP [3] dataset, ensuring real-world conditions. The amplified image in Fig.3 is provided solely for visual representation, of the noise simulation.

2. ADDITIONAL QUALITATIVE RESULTS

3. REFERENCES

[1] Paul E. Debevec and Jitendra Malik, "Recovering high dynamic range radiance maps from photographs," USA, 1997, ACM Press/Addison-Wesley Publishing Co.

[2] Samuel W. Hasinoff, Frédo Durand, and William T. Freeman, "Noise-optimal capture for high dynamic range photography," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 553–560.

[3] Xinyu Jia, Chuang Zhu, Minzhen Li, Wenqi Tang, and Wenli Zhou, "Llvip: A visible-infrared paired dataset for low-light vision," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3496–3504.

Fig. 2. The figure shows noise in our real-world V-TIEE dataset, showing low-light input, amplified input, well-lit, and corresponding thermal images. Amplified images, enhanced by a factor of 10, highlight the noise in low-light images.

Fig. 3. The figure shows noise incorporation in the simulated low-light LLVIP dataset, low-light input, amplified input, well-lit, and corresponding thermal images. Amplified images, enhanced by a factor of 10, highlight the noise in low-light images.

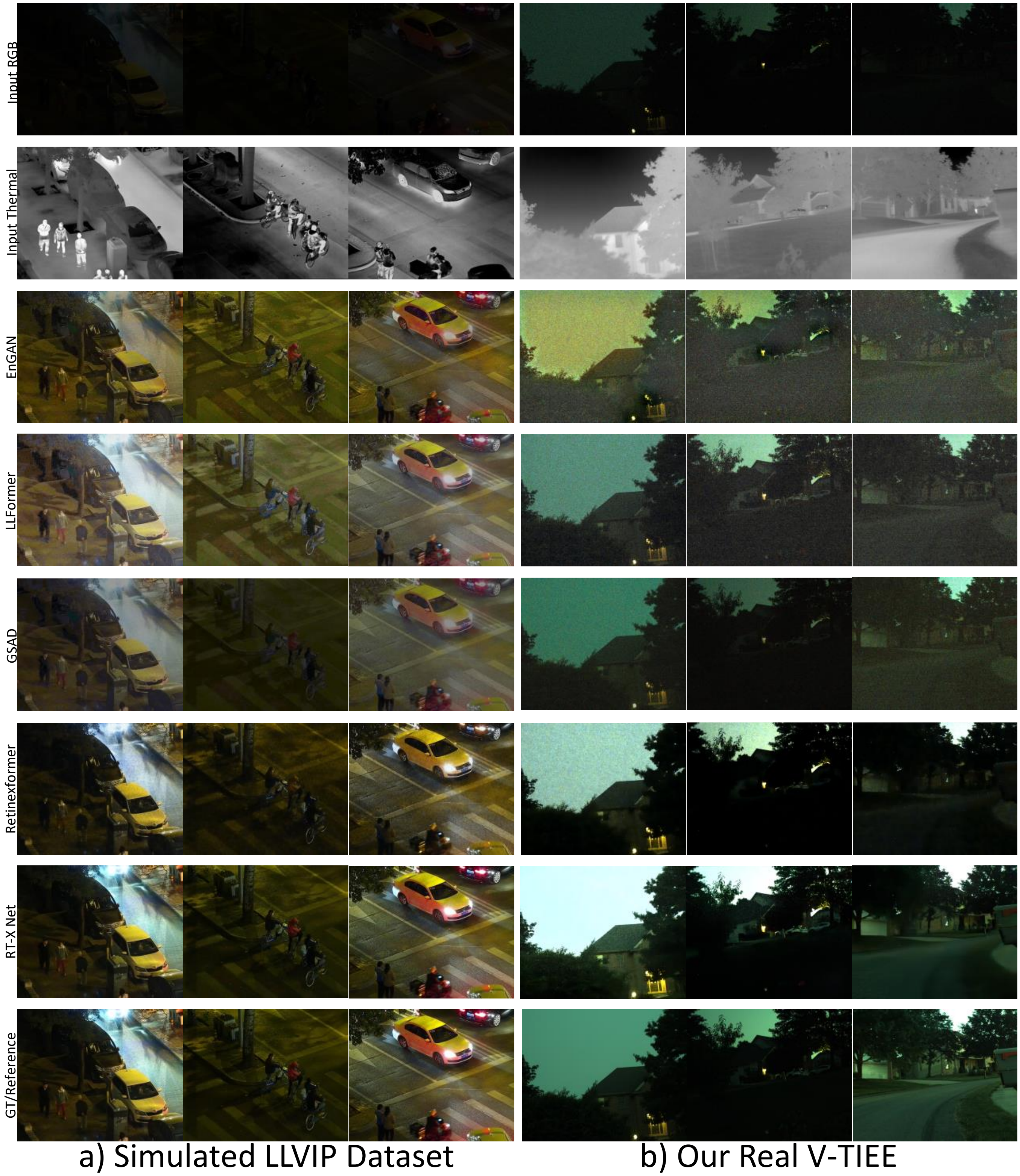

Fig. 4. Qualitative results on the synthetic LLVIP dataset and real-world V-TIEE dataset. Columns denote different scenes. The first two rows show the input visible and thermal images. The next five rows are the outputs from RT-X Net and state-of-the-art visible image enhancement algorithms. The last row shows the reference well-exposed image.