

Supplementary Material for “Illumination Guided Attentive Wavelet Network for Low-light Image Enhancement”

Jingzhao Xu, Mengke Yuan, Dong-Ming Yan, and Tieu Wu

In this supplementary material, some experiments are provided to further demonstrate the effectiveness of our IGAWN model. In section I, we conduct ablation study to analyze the contribution of each loss function that adopted in our model. In section II, we investigate the influence of different training strategies on our model. In section III, we present extensive visual comparisons with representative state-of-the-art methods.

I. EFFECTIVENESS OF LOSS FUNCTIONS

Our adopted total loss function consists of three components: image content loss \mathcal{L}_{IMC} , image smooth loss \mathcal{L}_{IMS} , and illumination content loss \mathcal{L}_{ILC} . In this section, we conduct ablation experiments to analyze the contribution of each loss function.

Quantitative and visual results of our model trained by various combinations of loss functions are shown in Table I and Fig. 1. The image content loss \mathcal{L}_{IMC} directly measures the difference between the output and normal-light ground-truth image, and it greatly determines the quality of enhancement result. As shown in Fig. 1(b), the model without \mathcal{L}_{IMC} fails to enhance low-light image. Since MS-SSIM evaluates the luminance, contrast, and structure locally and globally, the brightness is uniformly enhanced in Fig. 1(c). Then the image smooth loss \mathcal{L}_{IMS} helps to refine edges and details as shown in Fig. 1(e). At last, the illumination content loss \mathcal{L}_{ILC} further improves the perceptual quality of enhancement result as displayed in Fig. 1(f), and our model trained with full loss function achieves the best NIQE values as reported in Table I. The above results demonstrate that each loss function plays an important role in producing the final results.

In Fig. 2, we present the results of the average NIQE values versus two loss function weights (ω_{IMS} and ω_{ILC}). Since the loss weights ω_{IMS} and ω_{ILC} are positive real numbers, there can be infinite choices of the values of ω_{IMS} and ω_{ILC} . For simplicity, we discretize the value ranges of ω_{IMS} and ω_{ILC} . When $\omega_{IMS} = 10$ and $\omega_{ILC} = 0.2$, it can be seen that our model achieves the lowest NIQE value, which indicates the best perceptual quality. Considering that the NIQE values vary slightly when $\omega_{IMS} \in \{9, 10, 11\}$ and $\omega_{ILC} \in \{0.1, 0.2, 0.3\}$, we empirically adopt $\omega_{IMS} = 10$ and $\omega_{ILC} = 0.2$ in our experiments.

The illumination content loss \mathcal{L}_{ILC} measures the difference between the output of ILE branch and the pseudo ground-truth illumination map estimated by RetinexNet [1]. Since it is

challenging to estimate illumination map from natural images, one may concern that the proposed method highly depends on the illumination map estimated from a specific approach. In our model, the ILE branch is integrated to alleviate the difficulty of enhancing low-light images. We use the ILE branch to learn reasonable guidance information to assist the IME branch in uniform exposure and noise suppression. To analyze the robustness of our model to the illumination map supervision, we train our model using different illumination maps. As shown in Table II, our model performs well even without the illumination map supervision. Each of the illumination estimation approaches [1]–[3] has its own advantages on varied datasets. Overall, our proposed model learns to extract beneficial illumination features by jointly optimizing the loss functions of ILE and IME branches, and thus does not rely heavily on the specific illumination estimation method [1].

R1
A2

II. ABLATION STUDY ON DIFFERENT TRAINING STRATEGIES

Our two-branch model is trained as a whole, illumination feature maps of ILE are fed to IME with FFT (frequency feature transform) layers, and actually ILE branch is supervised by all the three loss functions: illumination content loss \mathcal{L}_{ILC} , image content loss \mathcal{L}_{IMC} and image smooth loss \mathcal{L}_{IMS} . As indicated in Table I of Section I, each loss function plays an important role in achieving satisfactory low-light enhancement results.

We have conducted experiments to investigate the influence of other two different training strategies: sequential training with pretraining ILE until converging and alternating training ILE and IME. Specifically, for sequential training, we first train the ILE branch for 200 epochs, the initial learning rate is set as 10^{-4} for the first 100 epochs and linearly decreases to 0 for another 100 epochs. With the pretrained ILE branch, we train the IME branch for 200 epochs, the initial learning rate of IME branch is set as 10^{-4} for the first 100 epochs and linearly decreases to 0 for another 100 epochs. For alternating training, we alternately optimize these two branches for 200 epochs, the initial learning rate is set as 10^{-4} for the first 100 epochs and linearly decreases to 0 for another 100 epochs. The illumination guidance features are only constrained by the illumination content loss. After both branches are converged, we then jointly optimize the whole network with the total loss $\mathcal{L}_{total} = \mathcal{L}_{IMC} + 10 \times \mathcal{L}_{IMS} + 0.2 \times \mathcal{L}_{ILC}$ for 100 epochs, the learning rate linearly decreases from 10^{-4} to 0.

TABLE I
ABLATION STUDY ON THE LOSS FUNCTIONS. NIQE VALUES ON DIFFERENT DATASETS ARE REPORTED.

Cases	\mathcal{L}_{IMC}	\mathcal{L}_{IMS}	\mathcal{L}_{ILC}	DICM	LIME	MEF	NPE	Fusion	Average
1	\times	\checkmark	\checkmark	7.6644	6.8120	7.6127	6.3940	6.1018	6.9970
2	\checkmark	\times	\times	4.1433	4.3991	4.2167	4.0969	3.9550	4.1622
3	\checkmark	\times	\checkmark	3.9941	3.9751	4.1274	3.7423	3.7321	3.9142
4	\checkmark	\checkmark	\times	3.7648	4.0113	3.8550	3.3488	3.3006	3.6561
5	\checkmark	\checkmark	\checkmark	3.7462	3.9385	3.6143	3.3295	3.2794	3.5816

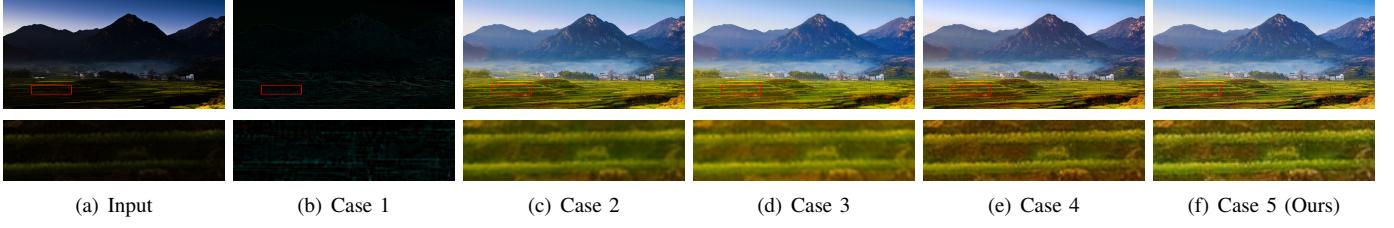


Fig. 1. Ablation study of the contribution of each loss function, where cases 1-5 correspond with the index numbers in Table I.

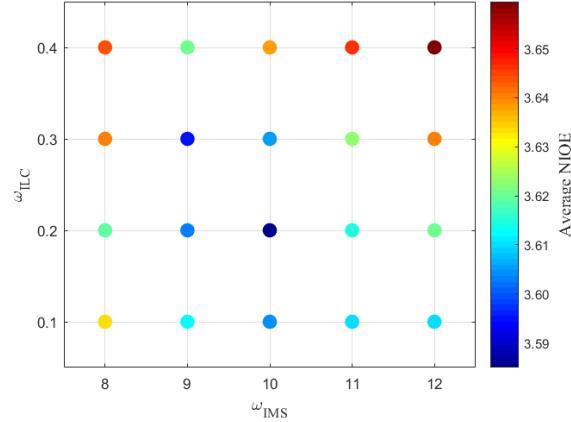


Fig. 2. Average NIQE values versus the values of ω_{IMS} and ω_{ILC} . The color of the point denotes the average NIQE value which is illustrated on the right bar. Lower NIQE value indicates better perceptual quality. Average NIQE values are evaluated on five real-scene datasets (*i.e.*, DICM [4], LIME [5], MEF [6], NPE [7], and Fusion [8]).

TABLE II
ABLATION STUDY ON THE ILLUMINATION MAP SUPERVISION. NIQE VALUES ON DIFFERENT DATASETS ARE REPORTED.

Method	DICM	MEF	NPE	Fusion	Average
w/o illumination map supervision	3.8300	3.7775	3.3166	3.3296	3.5634
Luminance map in YCbCr space	3.7749	3.8043	3.3295	3.2899	3.5497
Illumination from [2]	3.8094	3.7829	3.3201	3.2857	3.5495
Illumination from [3]	3.8015	3.5727	3.3291	3.3924	3.5239
Illumination from [1] (Ours)	3.7462	3.6143	3.3295	3.2794	3.4924

Quantitative results are shown in Table III. It can be seen that the model using sequential training strategy performs poorly on almost all datasets, which indicates that joint training stage is important for our two-branch architecture. Alternating training strategy and joint training strategy have their own advantages on different datasets. The model using alternating training strategy achieves the best performance on LOL-V2 Real dataset, but its NIQE values on real-scene

datasets are slightly worse than that of our model using joint training strategy. Visual comparisons on LOL-V2 dataset and real-scene images are shown in Fig. 3 and Fig. 4, respectively. Different training strategies have little difference on the LOL-V2 dataset, and results produced by the model acquired by alternating training strategy are the closest to the target ground truth. For real-scene low-light images, models using sequential training and alternating training strategies produce underexposed results in the red boxes in the two examples.

III. ADDITIONAL VISUAL COMPARISONS WITH STATE-OF-THE-ART METHODS

In this section, we provide more visual comparisons against state-of-the-art methods, including traditional methods (*i.e.*, SRIE [9] and LIME [5]), Retinex-based deep learning methods (*i.e.*, RetinexNet [1], KinD [3], and LPNet [10]), unsupervised deep learning methods (*i.e.*, EnlightenGAN [11] and ZeroDCE [12]).

Figs. 5-8 show the visual comparisons on the real captured images in the LOL-V2 dataset. Figs. 9-12 show visual comparisons on noise-free synthetic images. Figs. 13-17 demonstrate the image enhancement results on real-world images.

REFERENCES

- [1] C. Wei, W. Wang, W. Yang, and J. Liu, “Deep retinex decomposition for low-light enhancement,” in *Proc. Brit. Mach. Vis. Conf.*, 2018.
- [2] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, “Beyond brightening low-light images,” *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1013–1037, 2021.
- [3] Y. Zhang, J. Zhang, and X. Guo, “Kindling the darkness: A practical low-light image enhancer,” in *Proc. ACM Int. Conf. Multimedia*, 2019, p. 1632–1640.
- [4] C. Lee, C. Lee, and C.-S. Kim, “Contrast enhancement based on layered difference representation,” in *Proc. IEEE Int. Conf. Image Process.*, 2012, pp. 965–968.
- [5] X. Guo, Y. Li, and H. Ling, “LIME: Low-light image enhancement via illumination map estimation,” *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, 2017.
- [6] K. Ma, K. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, 2015.

TABLE III
ABLATION STUDY ON THE TRAINING STRATEGIES. BEST RESULTS ARE HIGHLIGHTED IN **BOLD**.

Method	LOL-V2 Real		LOL-V2 Synthetic		DICM	LIME	MEF	NPE	Fusion	Average
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	NIQE↓	NIQE↓	NIQE↓	NIQE↓	NIQE↓	NIQE↓
Sequential training	20.82	0.8304	21.91	0.8808	3.8441	3.9427	3.8464	3.2985	3.4566	3.6777
Alternating training	21.31	0.8423	22.04	0.8939	3.7573	3.9055	3.8345	3.3242	3.3424	3.6328
Joint training (Ours)	21.21	0.8400	22.14	0.9001	3.7462	3.9385	3.6143	3.3295	3.2794	3.5816

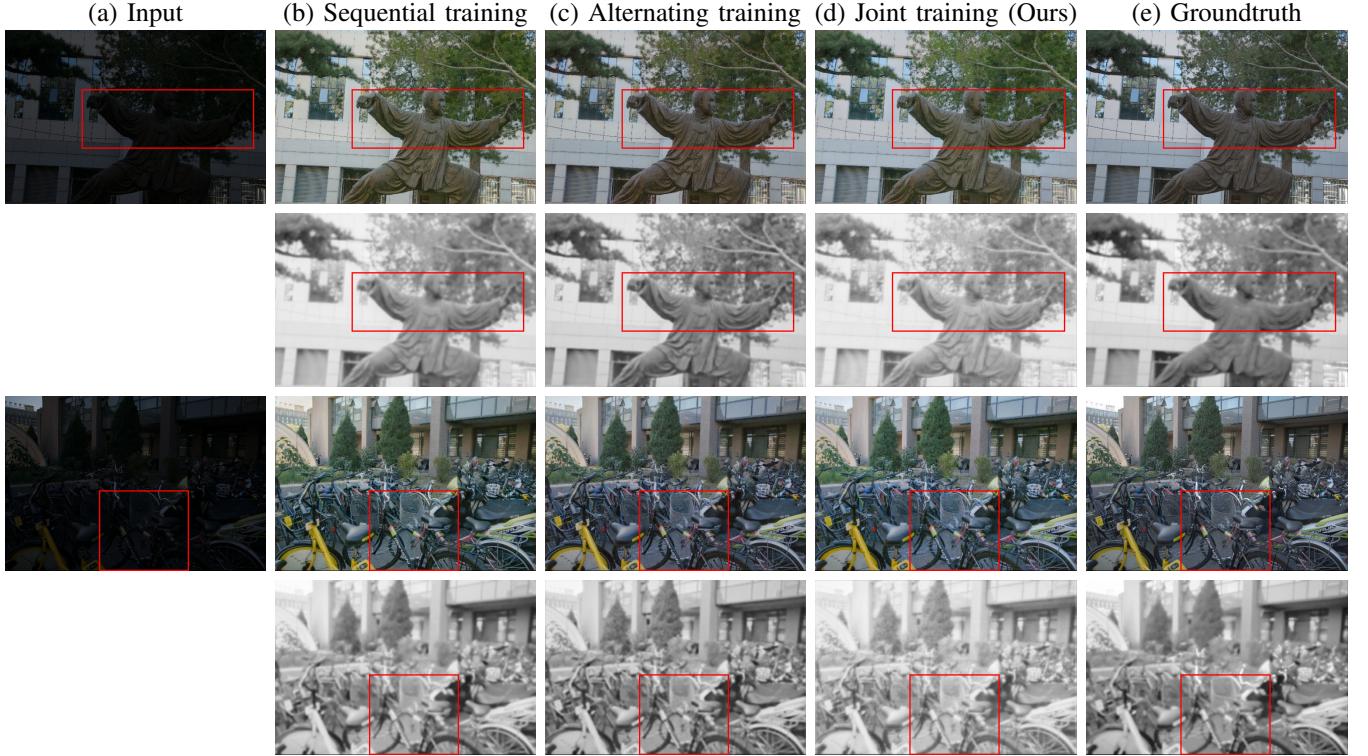


Fig. 3. Visual comparisons from the ablation study of using different training strategy on LOL-V2 real dataset. (a) Input low-light images. (b) Results from the model using sequential training strategy. (c) Results from the model using alternating training strategy. (d) Results from our model using joint training strategy. (e) Groundtruth and the illumination map estimated from the groundtruth image. Rows 1 and 3 show the enhanced results. Rows 2 and 4 show the estimated illumination maps.

- [7] S. Wang, J. Zheng, H.-M. Hu, and B. Li, “Naturalness preserved enhancement algorithm for non-uniform illumination images,” *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, 2013.
- [8] Q. Chen and D. Wu, “Image denoising by bounded block matching and 3D filtering,” *Signal Process.*, vol. 90, no. 9, pp. 2778–2783, 2010.
- [9] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, “A weighted variational model for simultaneous reflectance and illumination estimation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2782–2790.
- [10] J. Li, J. Li, F. Fang, F. Li, and G. Zhang, “Luminance-aware pyramid network for low-light image enhancement,” *IEEE Trans. Multimedia*, vol. 23, pp. 3153–3165, 2021.
- [11] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, “EnlightenGAN: Deep light enhancement without paired supervision,” *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.
- [12] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, “Zero-reference deep curve estimation for low-light image enhancement,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1777–1786.
- [13] W. Yang, W. Wang, H. Huang, S. Wang, and J. Liu, “Sparse gradient regularized deep retinex network for robust low-light image enhancement,” *IEEE Trans. Image Process.*, vol. 30, pp. 2072–2086, 2021.

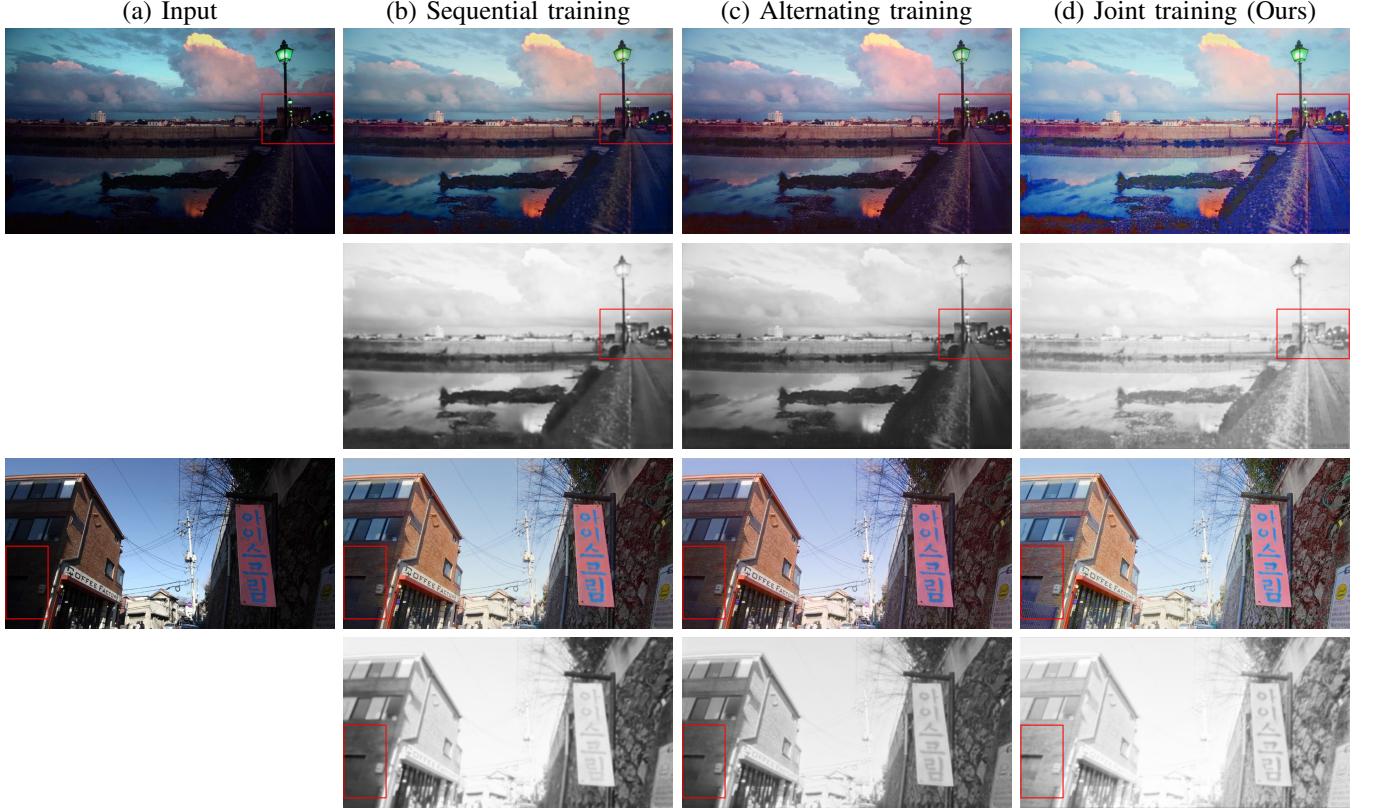


Fig. 4. Visual comparisons from the ablation study of using different training strategies. (a) Input low-light images. (b) Results from the model using sequential training strategy. (c) Results from the model using alternating training strategy. (d) Results from our model using joint training strategy. Rows 1 and 3 show the enhanced results. Rows 2 and 4 show the estimated illumination maps.

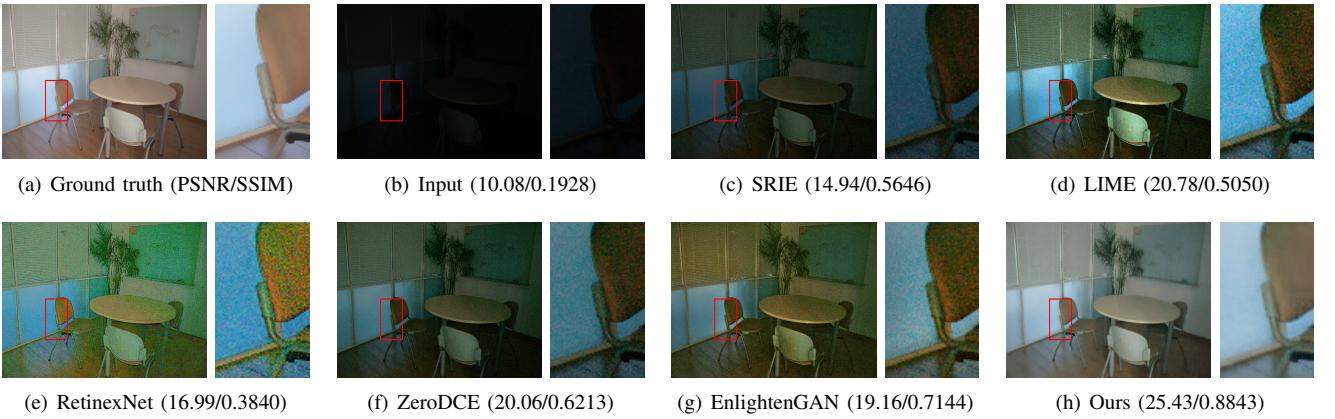


Fig. 5. Comparisons with state-of-the-art low-light image enhancement methods on a real image from LOL-V2 test dataset [13].



Fig. 6. Comparisons with state-of-the-art low-light image enhancement methods on a real image from LOL-V2 test dataset [13].

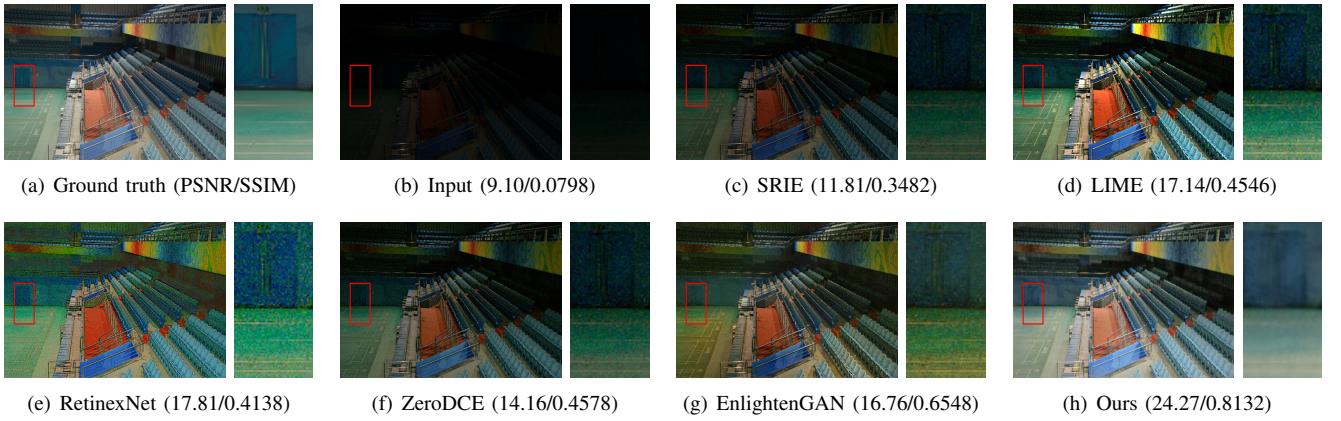


Fig. 7. Comparisons with state-of-the-art low-light image enhancement methods on a real image from LOL-V2 test dataset [13].

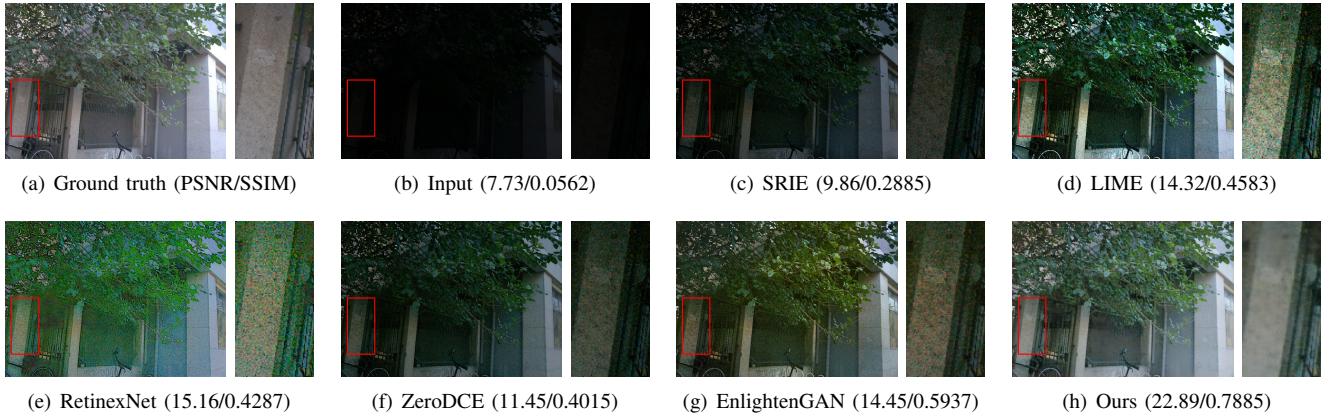


Fig. 8. Comparisons with state-of-the-art low-light image enhancement methods on a real image from LOL-V2 test dataset [13].

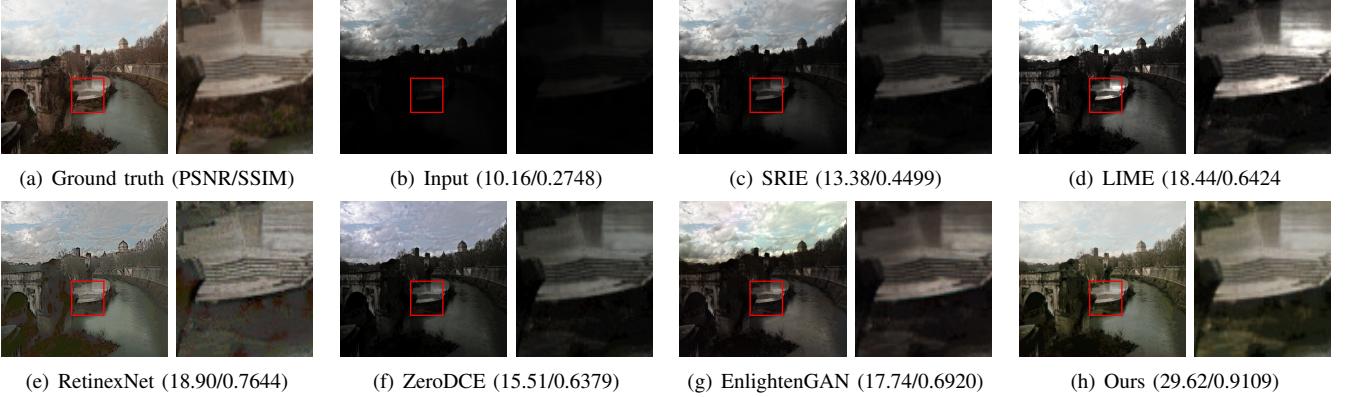


Fig. 9. Comparisons with state-of-the-art low light image enhancement methods on a synthetic image from LOL-V2 test dataset [13].

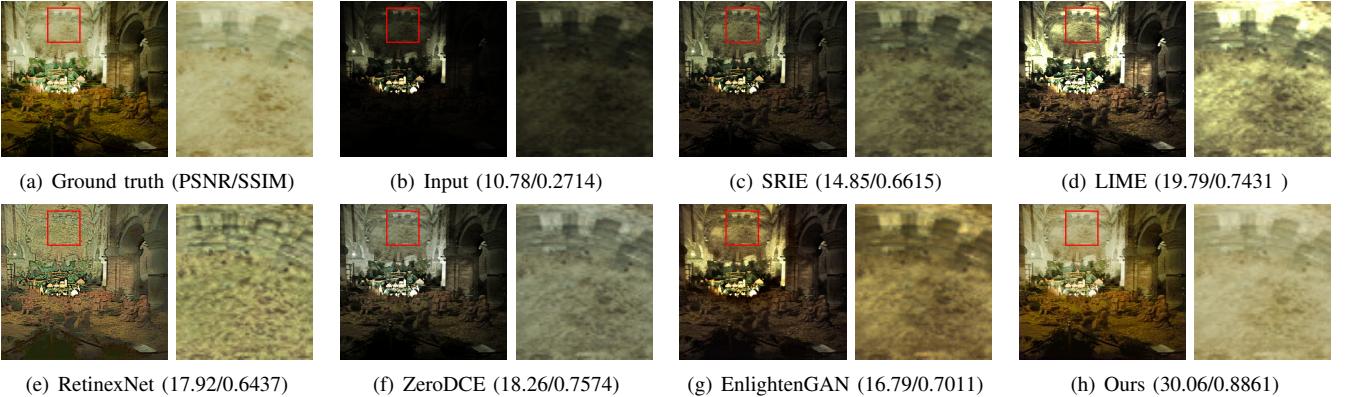


Fig. 10. Comparisons with state-of-the-art low light image enhancement methods on a synthetic image from LOL-V2 test dataset [13].

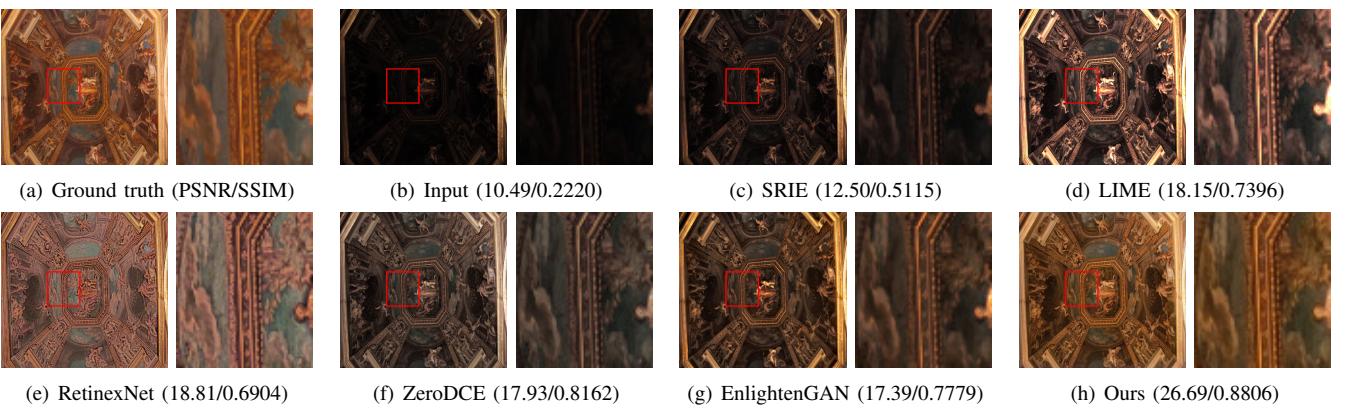


Fig. 11. Comparisons with state-of-the-art low light image enhancement methods on a synthetic image from LOL-V2 test dataset [13].

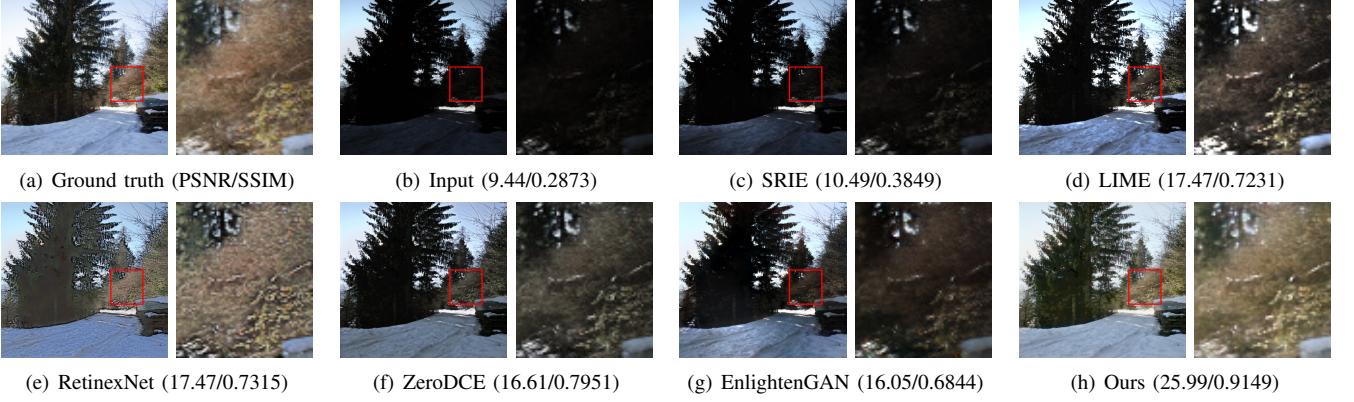


Fig. 12. Comparisons with state-of-the-art low light image enhancement methods on a synthetic image from LOL-V2 test dataset [13].

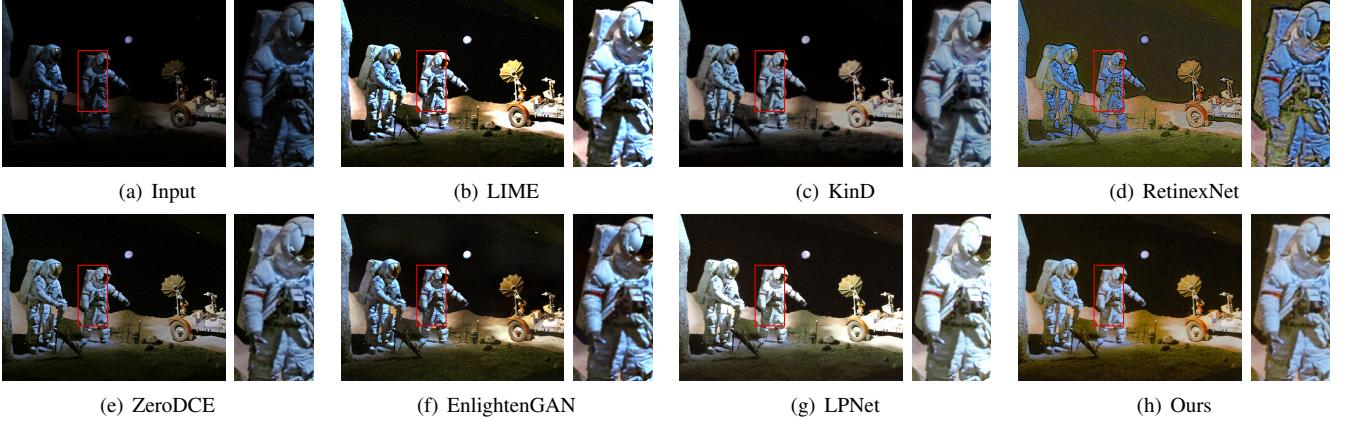


Fig. 13. Comparisons with state-of-the-art low-light image enhancement methods on a real-scene image.

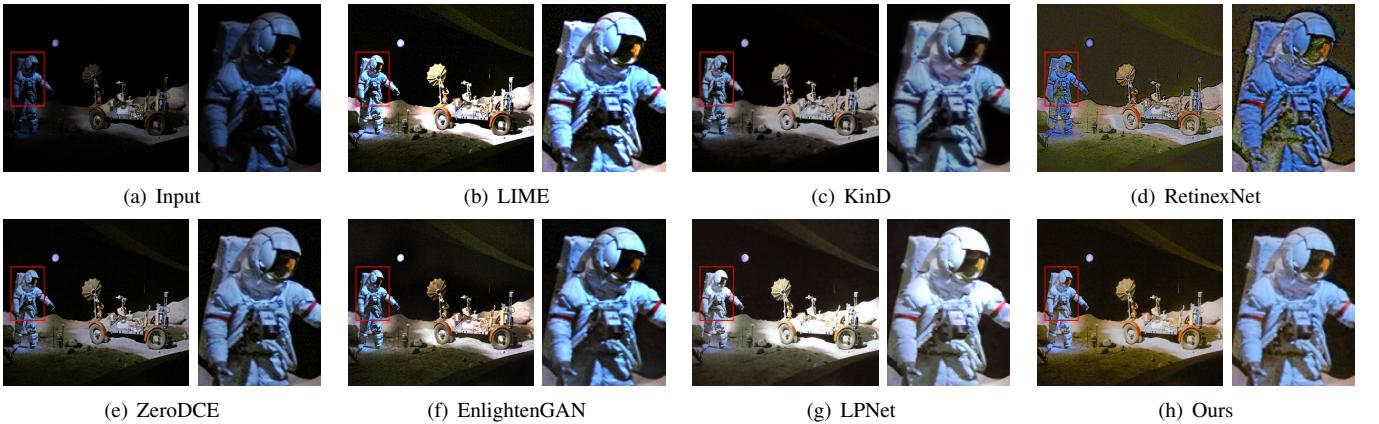


Fig. 14. Comparisons with state-of-the-art low-light image enhancement methods on a real-scene image.

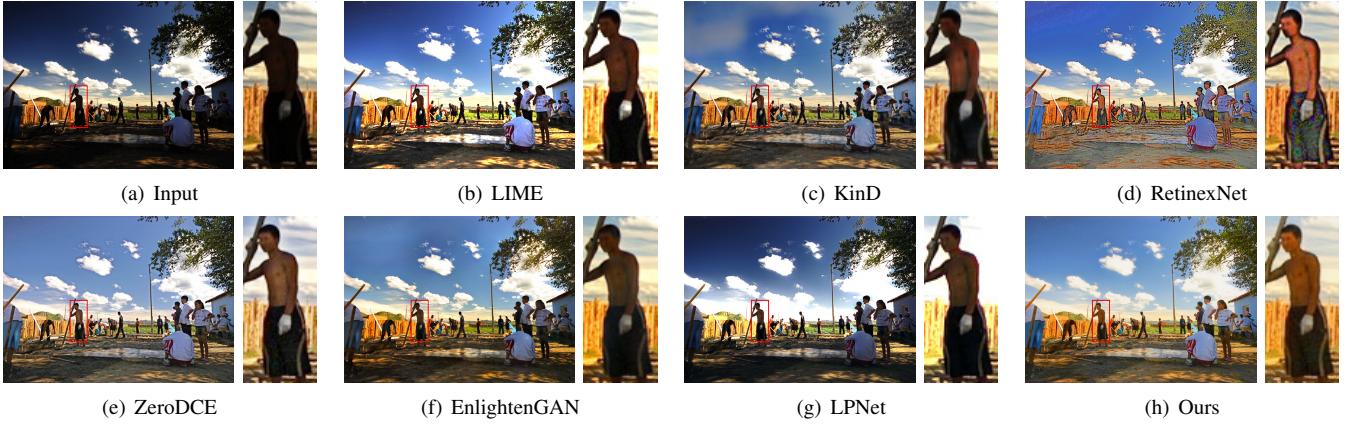


Fig. 15. Comparisons with state-of-the-art low-light image enhancement methods on a real-scene image.

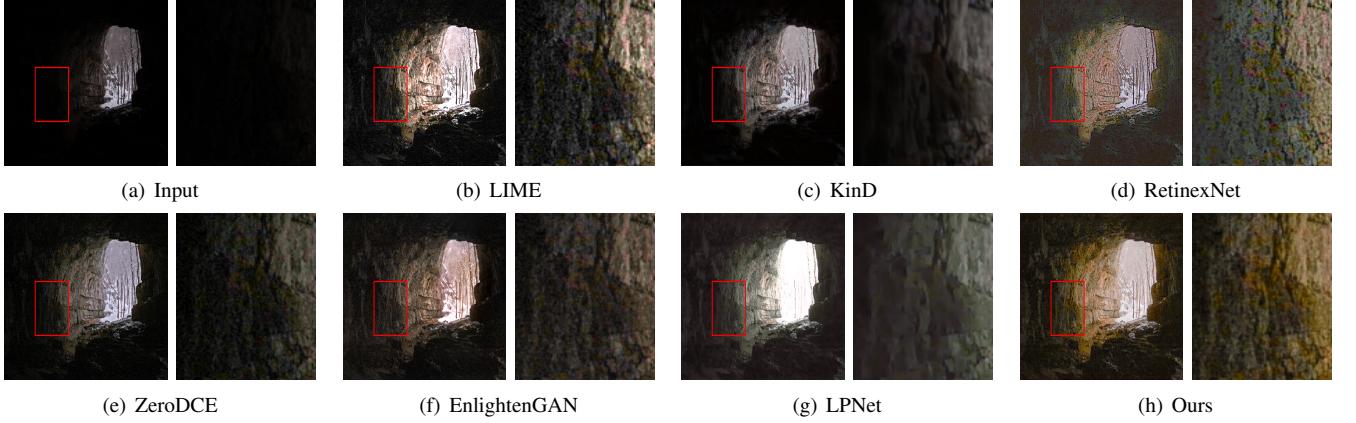


Fig. 16. Comparisons with state-of-the-art low-light image enhancement methods on a real-scene image.

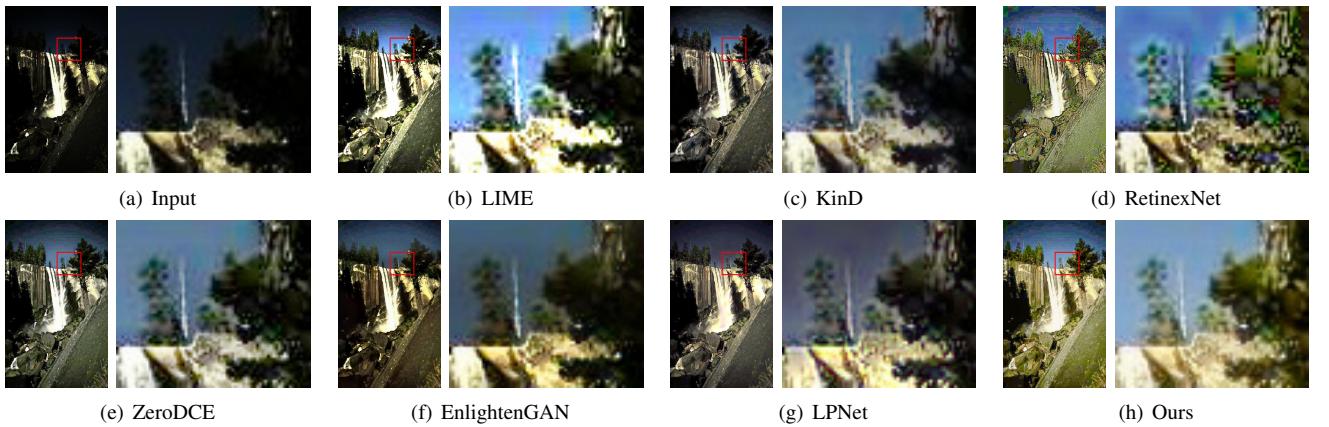


Fig. 17. Comparisons with state-of-the-art low-light image enhancement methods on a real-scene image.