Unrolled Decomposed Unpaired Learning for Controllable Low-Light Video Enhancement (Supplementary Material)

Lingyu Zhu[1], Wenhan Yang[2], Baoliang Chen[1], Hanwei Zhu[1], Zhangkai Ni[3], Qi Mao[4], and Shiqi Wang*[1]

[1] City University of Hong Kong, Hong Kong, China
[2] PengCheng Laboratory, Shenzhen, China
[3] Tongji University, Shanghai, China
[4] Communication University of China, Beijing, China
Corresponding author: shiqwang@cityu.edu.hk

In this supplementary material, we first provide the comparison benchmarks and training strategy in Sec. 1. Then, we provide the implementation details of the proposed **U**nrolled **D**ecomposed **U**npaired **Net**work (**UDU-Net**) and investigate the factors that influence the model performance in Sec. 2. Lastly, we provide more visual comparisons in terms of enhancement results and temporal warping error results in Sec. 3.

1 Experimental Details

Benchmarks. We evaluate the proposed network against non-deep-learning and deep-learning-based methods to ensure a fair comparison. The non-deep-learning enhancement methods include Bio-Inspired Multi-Exposure Fusion (BIMEF) [18], Light Image Enhancement via Illumination Map Estimation (LIME) [7], Multiscale Retinex (MR) [9], Dong *et al.*'s [3], Multiple Fusion (MF) [4], Naturalness Preserved Enhancement (NPE) [17], Simultaneous Reflectance and Illumination Estimation (SRIE) [5]. The deep-learning image enhancement methods include Zero-Reference Deep Curve Estimation (ZeroDCE) [6], EnlightenGAN [8], Self-Calibrated Illumination (SCI) [14], Progressive Self-Enhancement Network (PSENet) [15], CLIP-LIT [11] and Retinex-inspired Unrolling with Architecture Search (RUAS) [12] for image restoration. The deep-learning video enhancement methods include Self-supervised Adaptive Low-light Video Enhancement (SALVE) [1], SDSDNet [16], StableLLVE [19], DRVNet [2], Semantic-Guided Zero-Shot Learning (SGZSL) [20] and MBLLEN [13]. Furthermore, it should be noted that the proposed method is trained using a reference-free learning paradigm. Despite this, it can achieve comparable performance to supervised methods like MBLLVEN, DRVNet, StableLLVE, and SDSDNet in terms of both PSNR and SSIM.

Training Strategy. The proposed framework is implemented on a single GeForce RTX 3090 GPU using the PyTorch framework. During training, we utilize the Adam optimizer [10] with β_1=0.5 and β_2=0.99. The learning rate of $G_{\boldsymbol{\theta}}(\cdot)$, $D_{\boldsymbol{\theta}}(\cdot)$, $T_{\boldsymbol{\theta}}(\cdot)$ and $R_{\boldsymbol{\theta}}(\cdot)$ is set to 1×10^{-4}, 4×10^{-4}, 1×10^{-5}, 1×10^{-6}, respectively. The training process consists of 100 epochs with a batch size of 15. We also apply the data augmentation technique to increase the diversity of the training data, including random cropping, resizing, random horizontal flips, and vertical flips. All training video frames are cropped into 224×224.

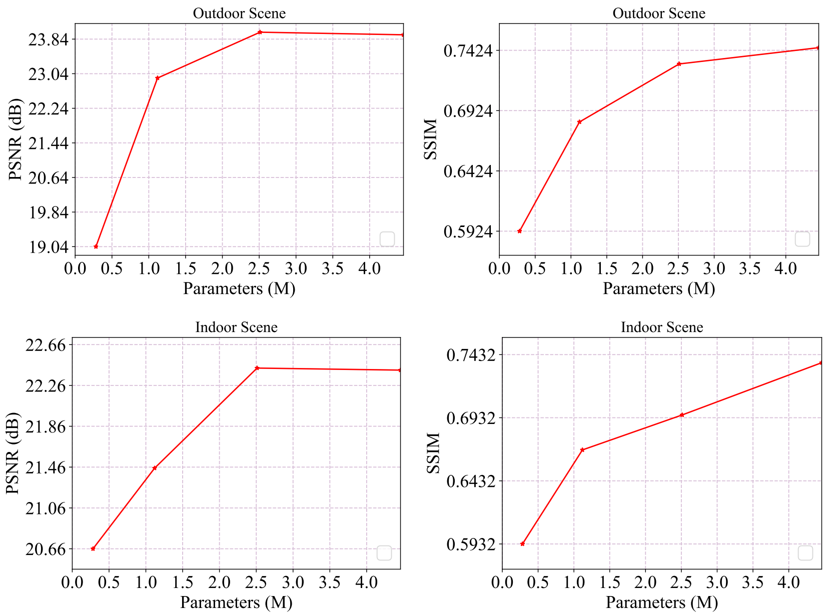

Fig. 1: Ablation study on model complexity.

2 Detailed Information of the Proposed UDU-Net

Model complexity. We provide detailed information about the network components in Table. 2. By varying the channel number, we examine how it affects the model complexity (*i.e.,* model parameters), as depicted in Fig. 1. Based on our findings, the structural performance might be further improved if computational resources (*i.e.,* memory usage) are allowed.

Activation Function. Based on our experimental observations, we choose to use the $\mathrm{ReLU}(\cdot)$ instead of $\mathrm{LeakyReLU}(\cdot)$ in the feature representation of $T_{\boldsymbol{\theta}}(\cdot)$ to preserve texture information. The comparison results are listed in Table. 1.

Table 1: The impact of activation in $T_{\boldsymbol{\theta}}(\cdot)$.

Network	Outdoor				Indoor			
	PSNR↑	SSIM↑	$warp$↓	MABD ↓	PSNR↑	SSIM ↑	$warp$↓	MABD ↓
Activation (LeakyReLU)	20.41	0.5849	0.17	0.11	19.92	0.4513	0.38	0.99
Activation (Relu)	23.94	0.7446	0.24	0.21	22.41	0.7368	0.41	1.05

3 More Visual Results

In this section, we illustrate more visual comparisons with the state-of-the-art enhancement method results in Fig. 2∼ Fig.6. The proposed method stands out with its superior performance compared to other methods. It is observed that the proposed method could recover the illumination, maintain structural details and produce fewer artifacts. The temporal warping results in Fig. 7∼ Fig.10 could be adopted to analyze the temporal smoothness. By examining the residual maps, it can be concluded that the proposed method surpasses other enhancement methods, demonstrating superior performance in maintaining temporal consistency.

Fig. 2: Qualitative comparison results on SDSD indoor dataset [16], which are enhanced using different enhancement methods. The supervised method (*) is provided for reference.

Table 2: Implementation details of our network structure. $G_{\boldsymbol{\theta}}(\cdot)$, $D_{\boldsymbol{\theta}}(\cdot)$ and $T_{\boldsymbol{\theta}}(\cdot)$ denote the generator, discriminator and the temporal model, respectively. ch denotes the channel number.

Module	Layer / Operation	Output
	SpectralNorm, Conv2d(3, ch * 1, 7, 1, 0), LeakyReLU	$\mathbf{GX_1}$
	SpectralNorm, Conv2d(ch * 1, ch * 2, 3, 2, 0), LeakyReLU	$\mathbf{GX_2}$
	SpectralNorm, Conv2d(ch * 2, ch * 4, 3, 2, 0), LeakyReLU	$\mathbf{GX_3}$
	SpectralNorm, Conv2d(ch * 4, ch * 8, 3, 2, 0), LeakyReLU	$\mathbf{GX_4}$
	SpectralNorm, Conv2d(ch * 8, ch * 16, 3, 2, 0), LeakyReLU	$\mathbf{GX_5}$
	Upsample, SpectralNorm, Conv2d(ch * 16, ch * 8, 1, 1, 0)	$\mathbf{Gy_1}$
	Concat ($\mathbf{Gy_1}$, Attention($\mathbf{GX_4}$))	-
	SpectralNorm, Conv2d(ch * 16, ch * 8, 3, 1, 0), LeakyReLU	$\mathbf{GY_1}$
	Upsample, SpectralNorm, Conv2d(ch * 8, ch * 4, 1, 1, 0)	$\mathbf{Gy_2}$
	Concat ($\mathbf{Gy_2}$, Attention($\mathbf{GX_3}$))	-
Illumination Learning (G)	SpectralNorm, Conv2d(ch * 8, ch * 4, 3, 1, 0), LeakyReLU	$\mathbf{GY_2}$
	Upsample, SpectralNorm, Conv2d(ch * 4, ch * 2, 1, 1, 0)	$\mathbf{Gy_3}$
	Concat ($\mathbf{Gy_3}$, Attention($\mathbf{GX_3}$))	-
	SpectralNorm, Conv2d(ch * 4, ch * 2, 3, 1, 0), LeakyReLU	$\mathbf{GY_3}$
	Upsample, SpectralNorm, Conv2d(ch * 2, ch * 1, 1, 1, 0)	$\mathbf{Gy_4}$
	Concat ($\mathbf{Gy_3}$, Attention($\mathbf{GX_2}$))	-
	SpectralNorm, Conv2d(ch * 2, ch * 1, 3, 1, 0), LeakyReLU	$\mathbf{GY_4}$
	Mul ($\mathbf{GX_1}$, Attention($\mathbf{GY_4}$))	-
	SpectralNorm, Conv2d(ch * 1, ch * 1, 3, 1, 0)	-
	SpectralNorm, Conv2d(ch * 1, 3, 1, 1, 0), Tanh	$\mathbf{GY_5}$
	Add ($\mathbf{GX_5}$, x)	-
	SpectralNorm, Conv2d(3, ch, 7, 2, 3), LeakyReLU	-
	Conv2d(ch, 1, 7, 1, 3), Tanh	$\mathbf{DX_1}$
	SpectralNorm, Conv2d(ch, ch * 2, 7, 2, 3), LeakyReLU	-
	Conv2d(ch * 2, 1, 7, 1, 3), Tanh	$\mathbf{DX_2}$
Illumination Learning (D)	SpectralNorm, Conv2d(ch * 2, ch * 4, 7, 2, 3), LeakyReLU	-
	Conv2d(ch * 4, 1, 7, 1, 3), Tanh	$\mathbf{DX_3}$
	SpectralNorm, Conv2d(ch * 4, ch * 8, 5, 2, 2), LeakyReLU	-
	Conv2d(ch * 8, 1, 5, 1, 2,), Tanh	$\mathbf{DX_4}$
	SpectralNorm, Conv2d(ch * 8, ch * 16, 5, 2, 2), LeakyReLU	-
	Conv2d(ch * 16, 1, 5, 1, 2), Tanh	$\mathbf{DX_5}$
	Conv3d(3, ch, 3, 1, 1), ReLU	$\mathbf{TF_1}$
	Conv3d(ch, ch, 3, 1,1), ReLU	-
	Conv3d(ch, ch, [2, 3, 3], 1, [0, 1, 1]), ReLU	$\mathbf{TF_2}$
	Conv3d(ch, ch, 3, 1,1), ReLU	$\mathbf{TF_3}$
	Conv3d(ch, ch, 3, 1,1), ReLU	-
Smoothness Learning (T)	Conv3d(ch, ch, [2, 3, 3], 1, [0, 1, 1]), ReLU	$\mathbf{TF_4}$
	Conv3d(ch, ch, 3, 1, 1), ReLU	$\mathbf{TF_5}$
	Conv3d(ch, ch, 3, 1, 1), ReLU	-
	Conv3d(ch, ch, [2, 3, 3], 1, [0, 1, 1]), ReLU	$\mathbf{TF_6}$
	Conv3d(ch, ch, 3, 1, 1), ReLU	$\mathbf{TF_7}$
	Conv3d(ch, ch, 3, 1, 1), ReLU	-
	Conv3d(ch, ch, [2, 3, 3], 1, [0, 1, 1]), ReLU	$\mathbf{TF_8}$
	Conv3d(ch, 3, 3, 1, 1)	$\mathbf{TF_9}$

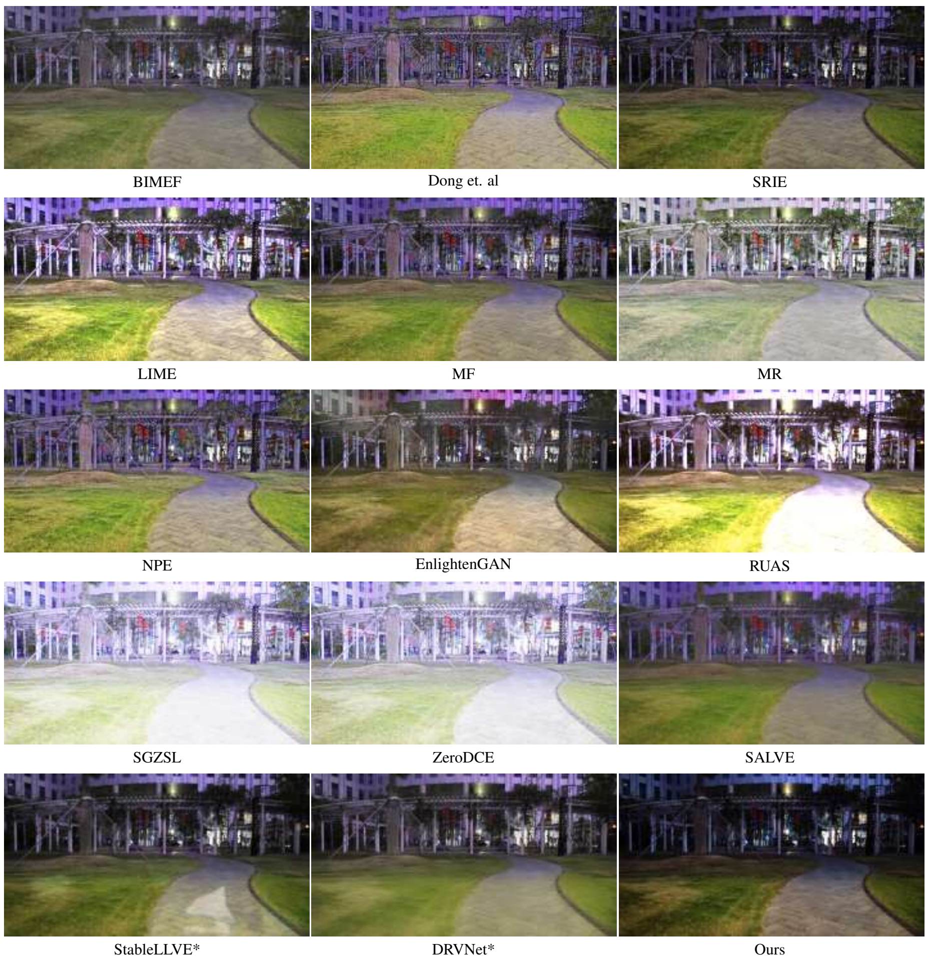

Fig. 3: Qualitative comparison results on SDSD outdoor dataset [16], which are enhanced using different enhancement methods. The supervised method (*) is provided for reference.

Fig. 4: Qualitative comparison results on SDSD indoor dataset [16], which are enhanced using different enhancement methods. The supervised method (*) is provided for reference.

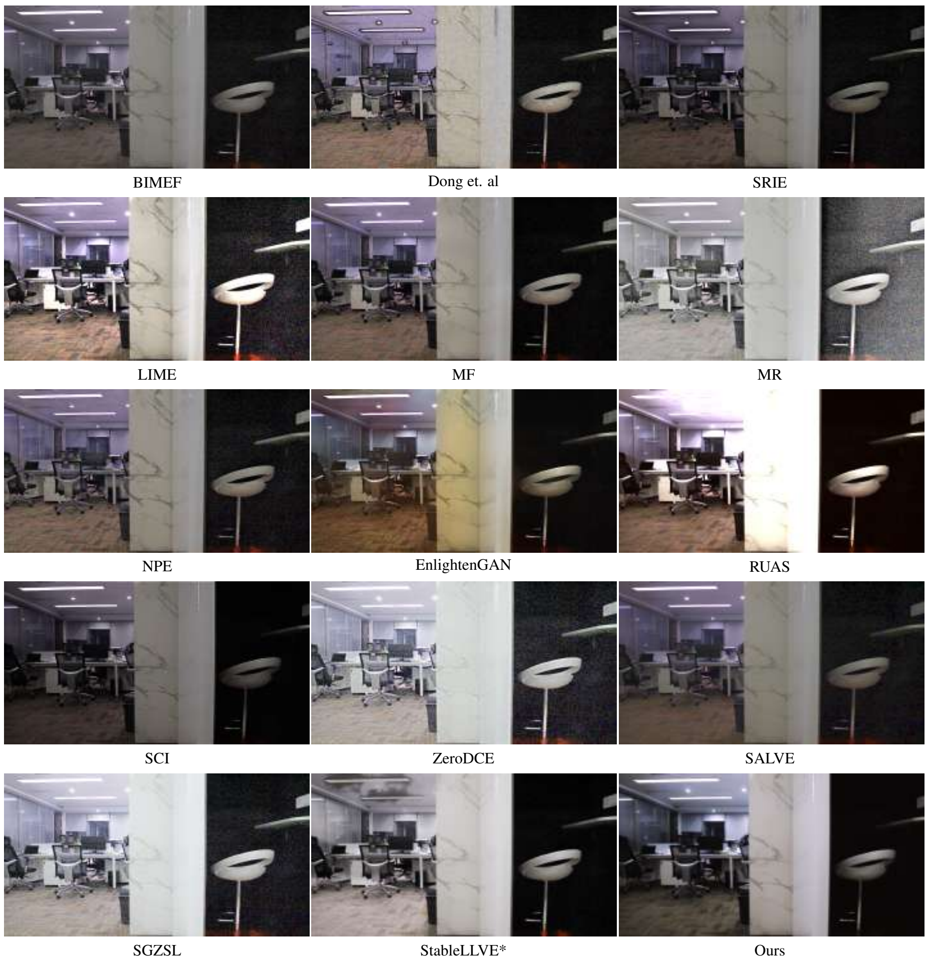

Fig. 5: Qualitative comparison results on SDSD indoor dataset [16], which are enhanced using different enhancement methods. The supervised method (*) is provided for reference.

Fig. 6: Qualitative comparison results on SDSD outdoor dataset [16], which are enhanced using different enhancement methods. The supervised method (*) is provided for reference.

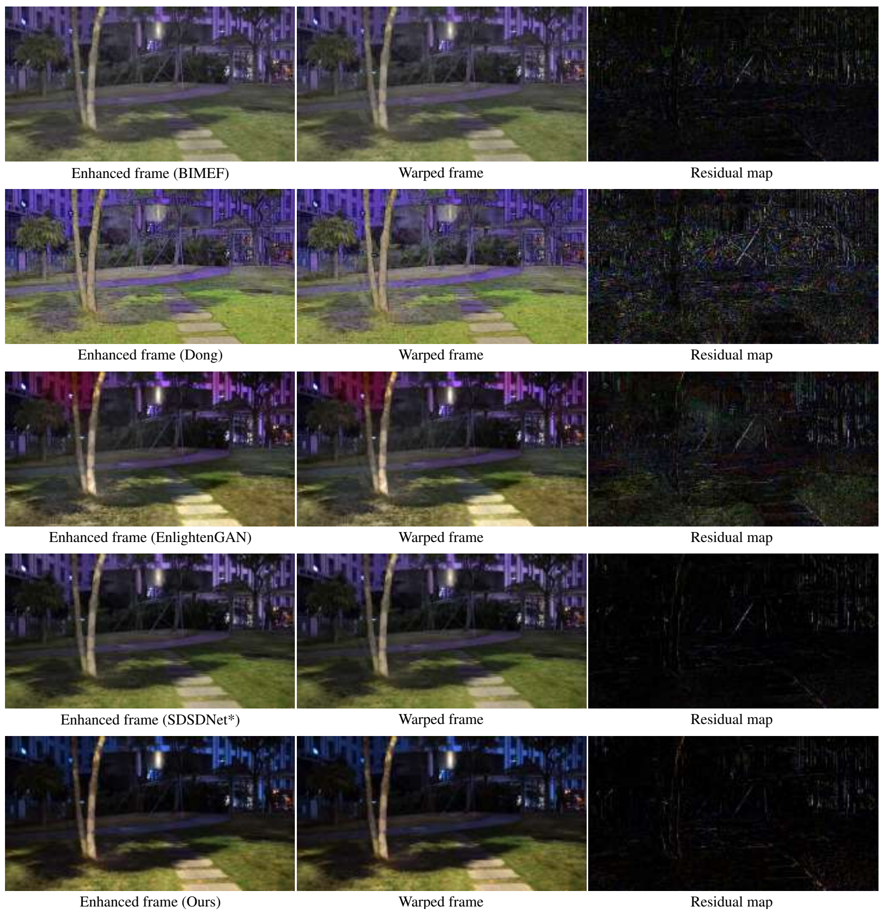

Fig. 7: Qualitative comparison results on SDSD outdoor dataset [16]. The first column shows the enhanced results. The second column visualizes the warped result based on the fine-tuned $R_{\boldsymbol{\theta}}(\cdot)$. The third column visualizes the per-pixel warping error.

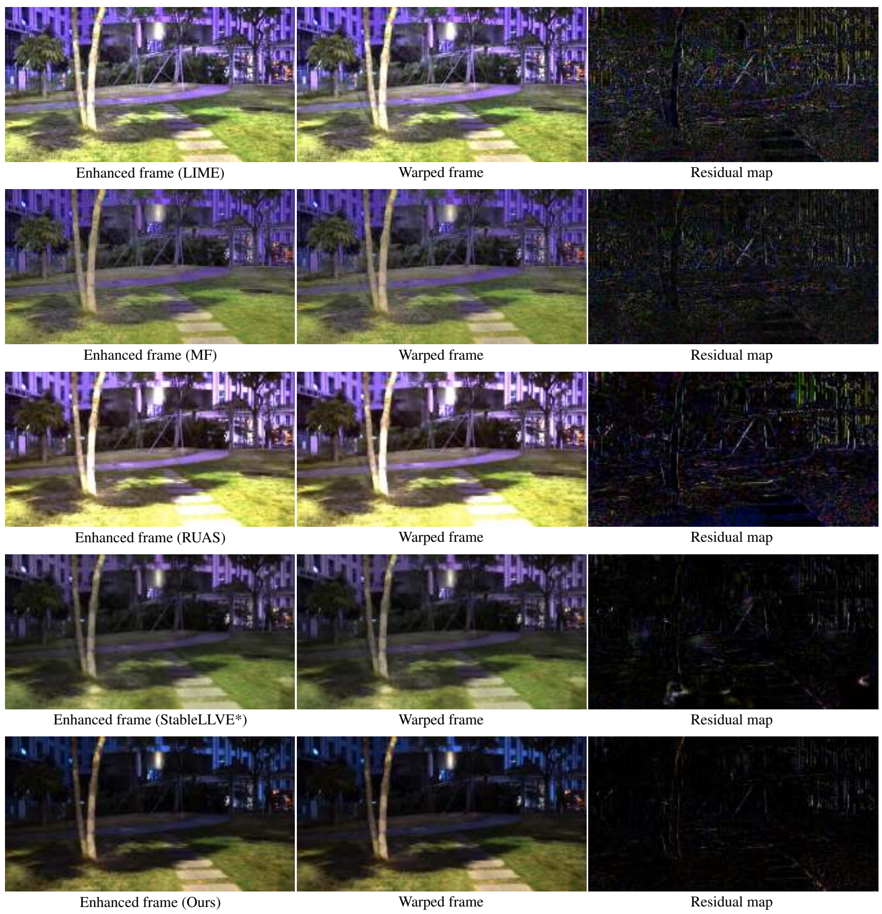

Fig. 8: Qualitative comparison results on SDSD outdoor dataset [16]. The first column shows the enhanced results. The second column visualizes the warped result based on the fine-tuned $R_{\boldsymbol{\theta}}(\cdot)$. The third column visualizes the per-pixel warping error.

Fig. 9: Qualitative comparison results on SDSD outdoor dataset [16]. The first column shows the enhanced results. The second column visualizes the warped result based on the fine-tuned $R_{\boldsymbol{\theta}}(\cdot)$. The third column visualizes the per-pixel warping error.

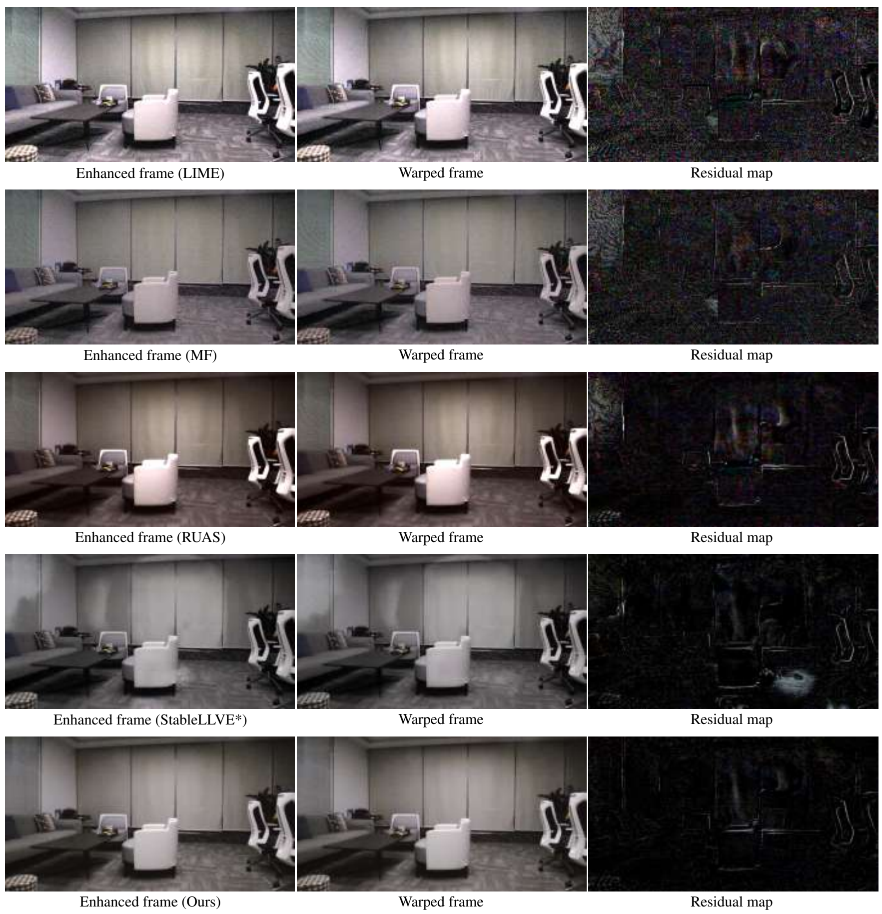

Fig. 10: Qualitative comparison results on SDSD outdoor dataset [16]. The first column shows the enhanced results. The second column visualizes the warped result based on the fine-tuned $R_{\boldsymbol{\theta}}(\cdot)$. The third column visualizes the per-pixel warping error.

References

1. Azizi, Z., Kuo, C.C.J., et al.: Salve: Self-supervised adaptive low-light video enhancement. APSIPA Transactions on Signal and Information Processing **12**(4) (2022) 1
2. Chen, C., Chen, Q., Do, M.N., Koltun, V.: Seeing motion in the dark. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3185–3194 (2019) 1
3. Dong, X., Wang, G., Pang, Y., Li, W., Wen, J., Meng, W., Lu, Y.: Fast efficient algorithm for enhancement of low lighting video. In: 2011 IEEE International Conference on Multimedia and Expo. pp. 1–6. IEEE (2011) 1
4. Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., Paisley, J.: A fusion-based enhancing method for weakly illuminated images. Signal Processing **129**, 82–96 (2016) 1
5. Fu, X., Zeng, D., Huang, Y., Zhang, X.P., Ding, X.: A weighted variational model for simultaneous reflectance and illumination estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2782–2790 (2016) 1
6. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1780–1789 (2020) 1
7. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing **26**(2), 982–993 (2016) 1
8. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: EnlightenGAN: Deep light enhancement without paired supervision. IEEE Transactions on Image Processing **30**, 2340–2349 (2021) 1
9. Jobson, D.J., Rahman, Z.u., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Transactions on Image Processing **6**(7), 965–976 (1997) 1
10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR (2014) 1
11. Liang, Z., Li, C., Zhou, S., Feng, R., Loy, C.C.: Iterative prompt learning for unsupervised backlit image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8094–8103 (2023) 1
12. Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10561–10570 (2021) 1
13. Lv, F., Lu, F., Wu, J., Lim, C.: Mbllen: Low-light image/video enhancement using cnns. In: British Machine Vision Conference. vol. 220 (2018) 1
14. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward fast, flexible, and robust low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5637–5646 (2022) 1
15. Nguyen, H., Tran, D., Nguyen, K., Nguyen, R.: Psenet: Progressive self-enhancement network for unsupervised extreme-light image enhancement. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1756–1765 (2023) 1
16. Wang, R., Xu, X., Fu, C.W., Lu, J., Yu, B., Jia, J.: Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9700–9709 (2021) 1, 3, 5, 6, 7, 8, 9, 10, 11, 12
17. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing **22**(9), 3538–3548 (2013) 1
18. Ying, Z., Li, G., Gao, W.: A bio-inspired multi-exposure fusion framework for low-light image enhancement. arXiv preprint arXiv:1711.00591 (2017) 1

19. Zhang, F., Li, Y., You, S., Fu, Y.: Learning temporal consistency for low light video enhancement from single images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4967–4976 (2021) 1
20. Zheng, S., Gupta, G.: Semantic-guided zero-shot learning for low-light image/video enhancement. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 581–590 (2022) 1