PSENet: Progressive Self-Enhancement Network for Unsupervised Extreme-Light Image Enhancement (Supplementary Material)

Hue Nguyen Diep Tran Khoi Nguyen Rang Nguyen
VinAI Research, Vietnam
{v.huent88, v.diepttn147, v.khoindm, v.rangnhm}@vinai.io

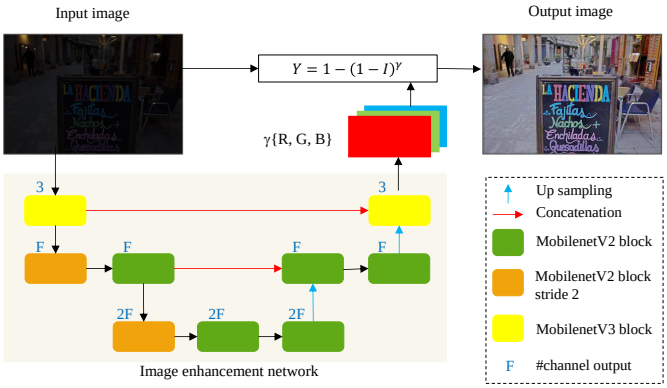

Figure 1. Image enhancement network in detail. A lightweight UNet architecture is employed to predict the gamma map γ for each channel. The enhanced image is obtained by applying the gamma mapping function with the predicted gamma map

In this supplementary material, we provide implementation details of our proposed method and additional results which are not included in the main paper due to the space limitation.

1. Implementation Details

Image Enhancement Network. As described in the main paper, we employ a lightweight UNet architecture [11] as illustrated in Figure 1 to build up our network. The specification of our network is given in Table 1.

Training Process. Our proposed approach is implemented using PyTorch framework. We train our image enhancement network on an NVIDIA A100 GPU from scratch, using the Adam optimizer with a batch size of 64. The learning rate is 0.0005 and is reduced by half on plateau with the patience of 5. The input images are resized to 256×256 without applying any augmentation techniques. For the SICE dataset, our model is trained for 140 epochs with the coefficient of the total variation loss α being 5. For the Afifi dataset, the number of training epochs is 30 and α is set to 500.

Table 1. Architecture detail of the image enhancement network. #Output denotes the number of output channels.

Input	Expand size	#Output	MobileNet	Stride
$256^2 \times 3$	6	3	V3	1
$256^2 \times 3$	24	16	V2	2
$128^2 \times 16$	24	16	V2	1
$128^2 \times 16$	48	32	V2	2
$64^2 \times 32$	48	32	V2	1
$64^2 \times 32$	48	16	V2	1
$128^2 \times 32$	48	16	V2	1
$128^2 \times 16$	24	3	V2	1
$256^2 \times 6$	9	3	V3	1

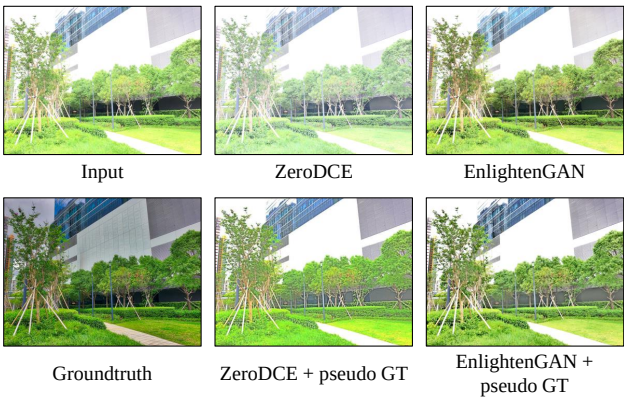

Figure 2. The influence of the pseudo GT generator on ZeroDCE [5] and EnlightenGAN [6]. Our proposed approach also improves these two networks' abilities of handling over-exposure cases

2. Ablation Study

The influence of pseudo GT image generator. As stated in the main paper, our training strategy also shows its effectiveness when combined with other image enhancement networks. Specifically, we apply our training strategy to the network architecture of ZeroDCE [5] and EnlightenGAN [6] with other settings kept unchanged. The results

Method	SICE		Afifi *et al.*	
	PSNR	SSIM	PSNR	SSIM
N = 1	17.74	0.704	19.36	0.869
N = 3	17.76	0.702	19.15	0.865
N = 5	17.84	0.706	18.61	0.856

Table 2. The impact of the number of randomly generated reference N to the final performance of our approach on SICE [2] and Afifi [1] datasets.

shown in Figure 2 demonstrate that our training strategy is robust to the network architecture selection when consistently improving the performance.

The impact of the number of random reference images. We further evaluate our model's performance when adjusting the number of random reference images. The results are presented in Table 2. We empirically find that increasing the number of random references improves the quality of the output images in the SICE dataset. However, with the Afifi dataset, it might have a negative impact on network performance. Thus, this hyper-parameter is dataset-specific.

Impact of the range for sampling reference images. In terms of brightness modification, we found that the best range to sample the reference images is from 0 to 3 for darker image generation and from -2 to 0 for synthesizing brighter images. If we narrow the range for under-exposure to $(0, 2)$, our model's performance decrease noticeably. The reason is that the produced gamma map is then limited, thus, our model could not increase the brightness of the input image to a proper value in extreme cases, as demonstrated in the two last rows of Fig. 2. On the other hand, regarding the range of sampling brighter images, extending this range from $(-2, 0)$ to $(-3, 0)$ might create undesired artifacts in overexposed areas. Due to image clipping, the color information in these areas is not well preserved. Therefore, when reducing image brightness, instead of producing a vivid image, our model tends to modify the color tone of the input image to gray, which is visually unpleasant.

The impact of the network size. We examine how our image enhancement network performs when the number of trainable parameters is increased or decreased. The quantitative results are shown in Table 3 and qualitative examples are visualized in Figure 4. Although the quantitative results vary slightly, we do not observe any obvious failure cases when visually comparing the output images. The difference in quantitative results appears to be caused by the shift in the brightness level of the output images compared to the ground truths. However, such output images are still acceptable when analyzed by humans.

Comparison with an image fusion method. Although our pseudo GT generator's design are inspired by the high level idea of the work introduced in [9], there are some noticeable differences between ours and theirs including our new quality score and our image combination strategy. We present

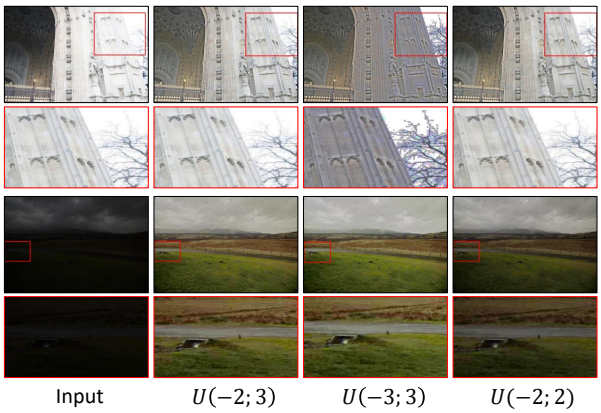

Figure 3. The impact of the range for sampling random reference images. $U(a;b)$ indicates the range to sample brightness factor X. The first and third rows show the whole images, while the second and last rows show the corresponding close-ups.

Method	SICE		Afifi *et al.*	
	PSNR	SSIM	PSNR	SSIM
# channels $\times$ 0.5	17.58	0.703	19.36	0.869
# channels $\times$ 1	17.74	0.704	19.36	0.869
# channels $\times$ 2	17.59	0.702	19.29	0.868

Table 3. The performance in PSNR and SSIM with different network parameters. The higher the better. # channels represents the number of channels in each layer of the proposed network (except the first and last layers)

Method	SICE		Afifi *et al.*	
	PSNR	SSIM	PSNR	SSIM
$\mu = 0.4$	16.02	0.690	17.97	0.844
$\mu = 0.5$	17.74	0.704	19.36	0.869
$\mu = 0.6$	16.60	0.6923	18.37	0.855

Table 4. The performance in PSNR and SSIM with different well-exposed levels μ . The higher the better

Method	PSNR	SSIM
[9]'s quality score + [9]'s RCS	15.14	0.652
[9]'s quality score + our RCS	16.78	0.702
Our quality score + our RCS	17.74	0.704

Table 5. Comparison with the quality score and reference combination strategy (RCS) proposed in [9]

the comparison between our method and the method in [9] when they are used inside pseudo GT generator module in Table 5. The results suggest that our proposed design are more effective than the prior work.

Well-exposed level. We conduct additional experiments to evaluate the effect of well-exposed level μ in the Equation (2) of our main paper on the performance of our approach. As shown in Figure 5 our model trained with a well-exposed level of 0.4 does not work effectively on under-exposed images while increasing this value to 0.6 makes our model fail to recover the detail of over-exposed images. Training with

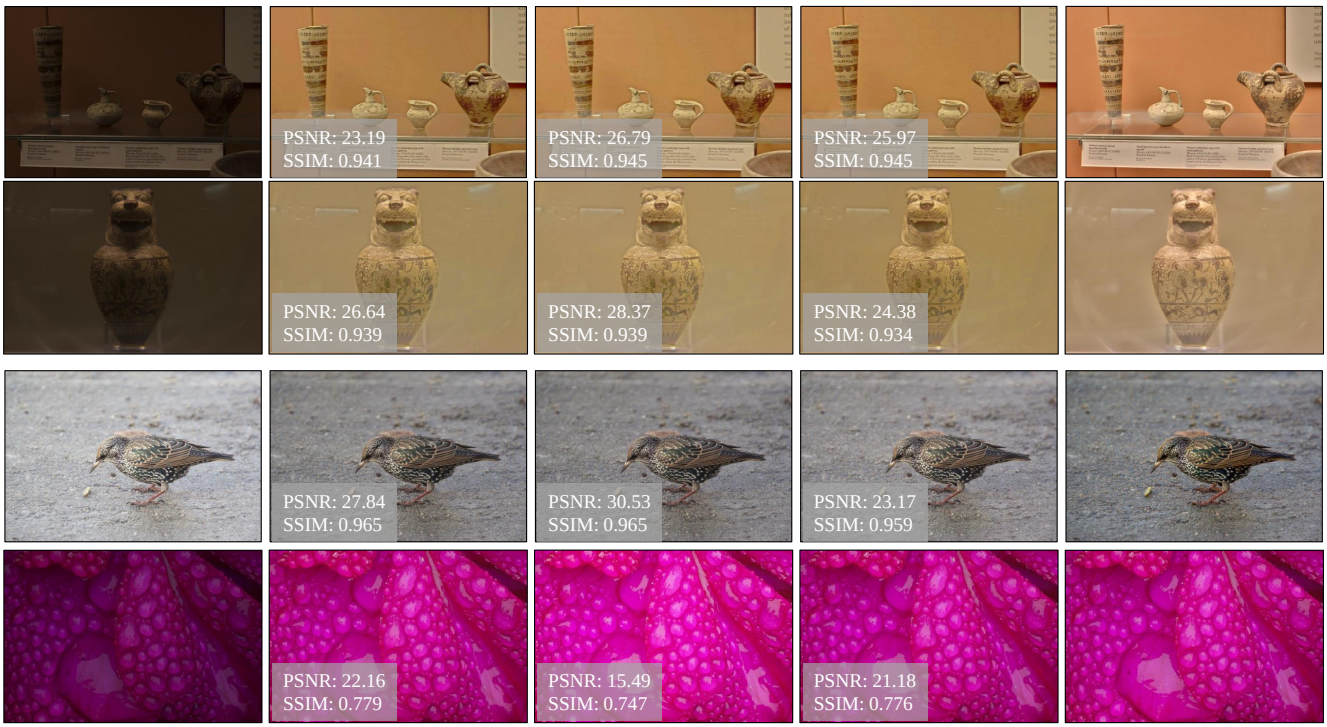

Figure 4. Samples whose PSNR values varied most when the number of network parameters is changed. # channels represents the number of feature maps in each layer of the proposed network (except the first and last layers). It seems that the change in PSNR value is mostly caused by the shift in the brightness level of the output images compared to the ground truths

Figure 5. Visual comparison among outputs of our model trained with different well-exposed level μ. Training with a well-exposed level of 0.5 seems to balance our model in handling both under-exposed and over-exposed images

a well-exposed level of 0.5 seems to balance our model between these two cases, yielding the highest quantitative results, as shown in Table 4.

3. Visual Comparison Results

This section presents additional qualitative results on other different public datasets including DICM [7], MEF [8], TMDIED[1]. We compare our method with two non-learning methods: CLAHE [10], IAGCWD [3], an unpaired method EnlightenGAN [6], two unsupervised methods: ZeroDCE [5], Zheng and Gupta[12], and two supervised methods: HDRNet [4], Afifi *et al*. [1]. The results are presented in Figures 6, 8, 9 and 7. It is worth noting that all the learning-based methods are trained on the SICE dataset except the Afifi *et al*. [1] due to its Matlab license issue.

[1]https://sites.google.com/site/vonikakis/datasets

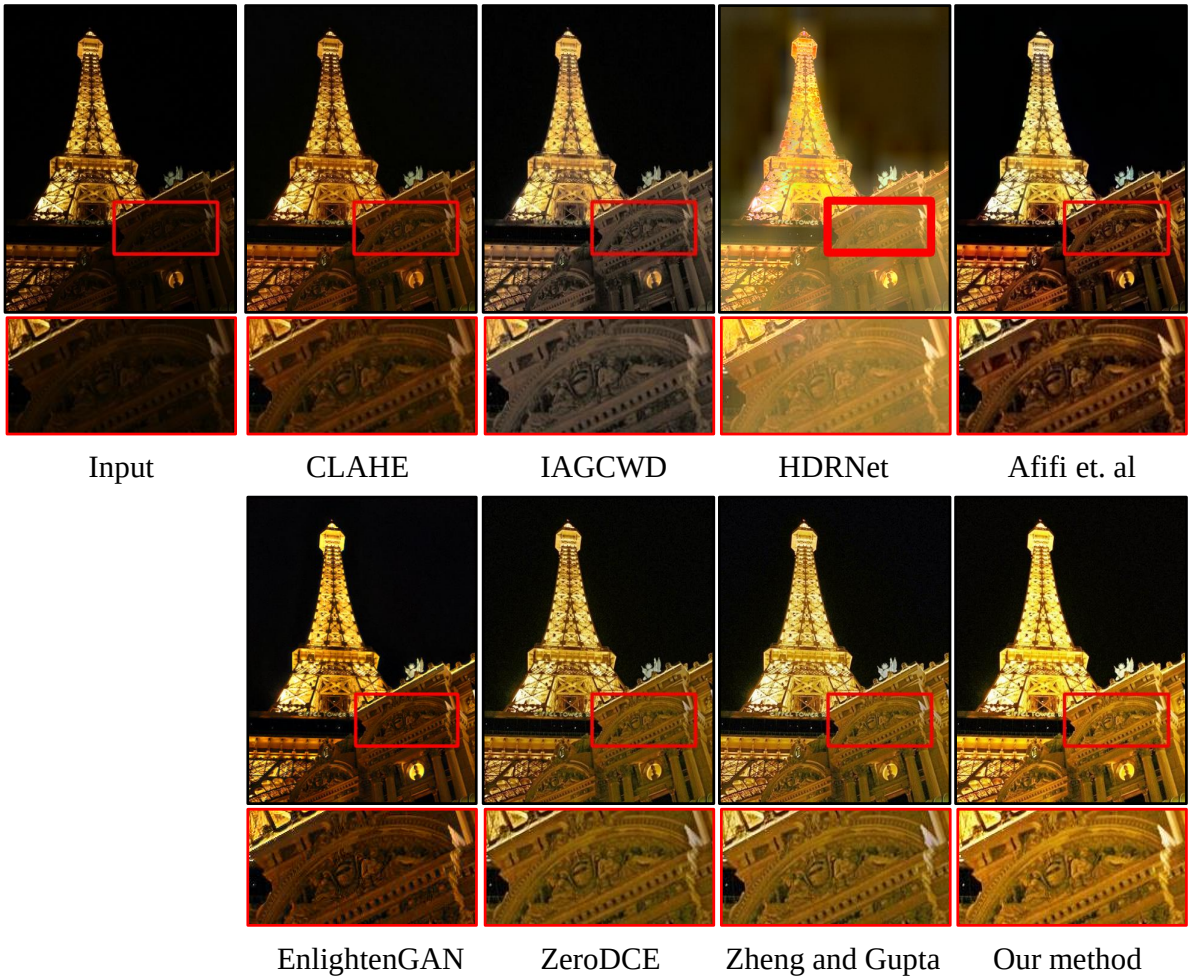

Figure 6. Visual comparison on a lowlight image taken from the DICM dataset. The unsupervised methods including ZeroDCE [5], Zheng and Gupta [12], and our method produce more compelling results than others. Among them, our method's result is arguably the best in terms of contrast and color preservation as shown in boxed regions

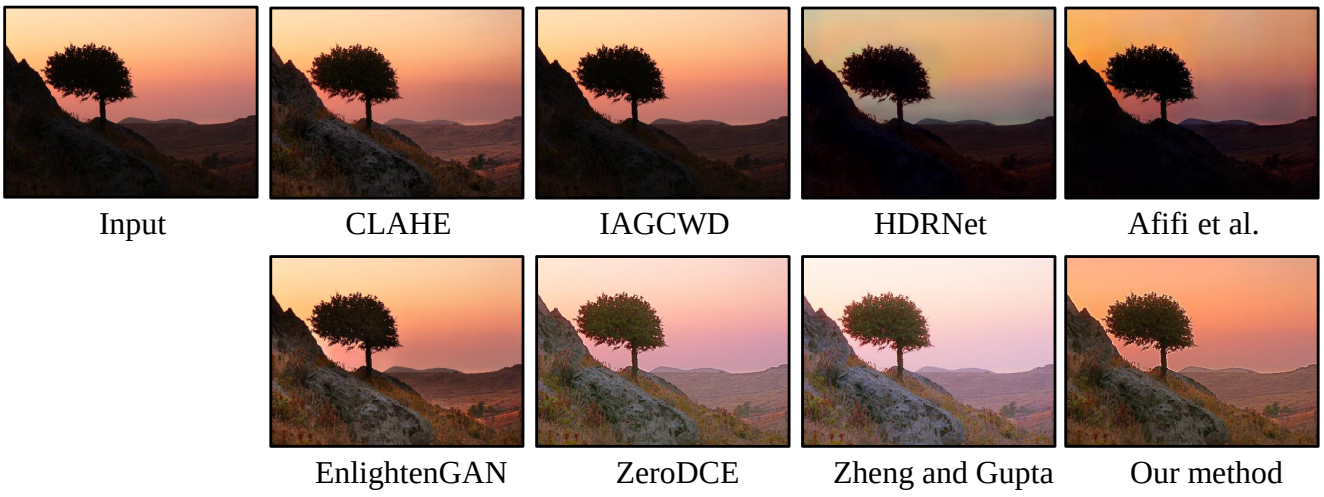

Figure 7. Visual comparison on an image taken from the TMDIED dataset. Our result image seems to be more lively

Figure 8. Visual comparison on an image taken from the MEF dataset. Our model gives a better result in terms of enhancing under-exposed areas and preserving the original color temperature

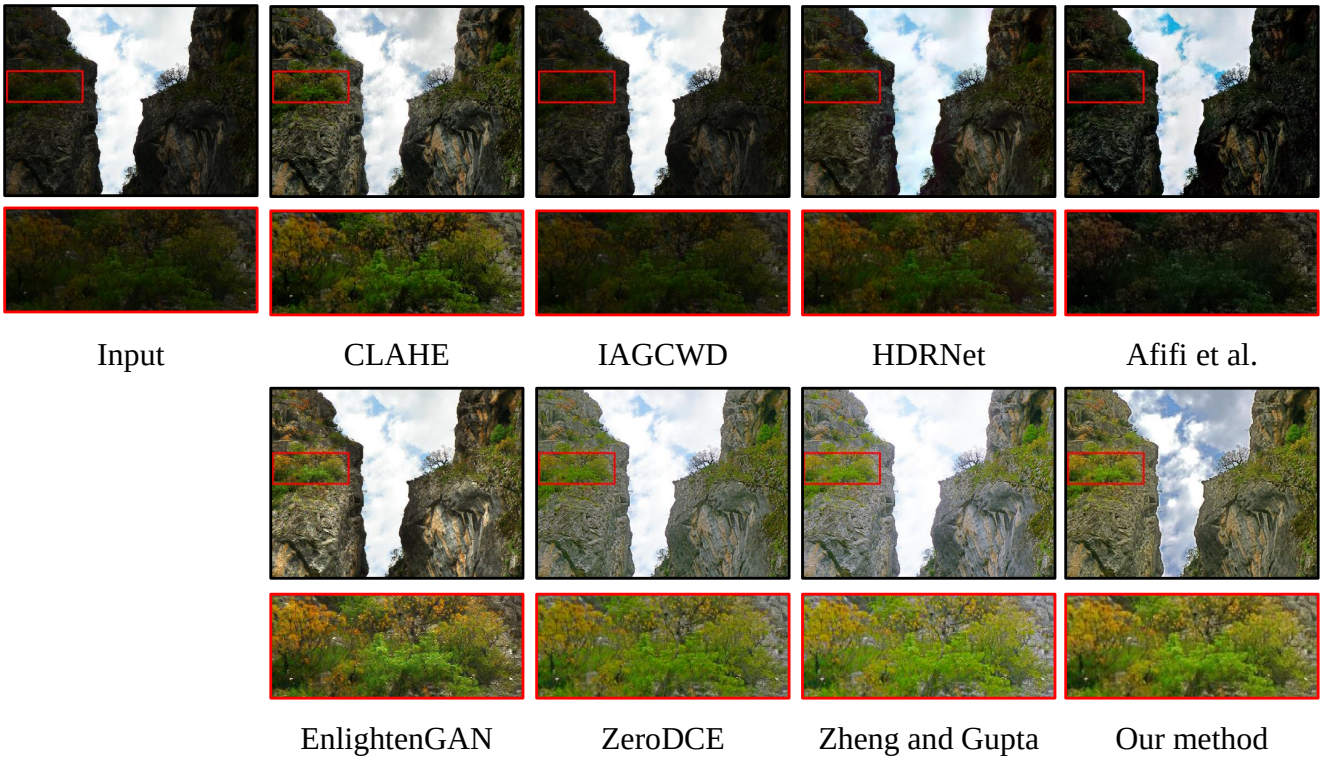

Figure 9. Visual comparison on an image taken from the TMDIED dataset. Our method gives the best balance in contrast between the dark and the bright regions

References

[1] Mahmoud Afifi, Konstantinos G. Derpanis, Björn Ommer, and Michael S. Brown. Learning multi-scale photo exposure correction. In *CVPR*, 2021.

[2] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *TIP*, 2018.

[3] Gang Cao, Lihui Huang, Huawei Tian, Xianglin Huang, Yongbin Wang, and Ruicong Zhi. Contrast enhancement of brightness-distorted images by improved adaptive gamma correction. *Computers & Electrical Engineering*, 2018.

[4] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *TOG*, 2017.

[5] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*, 2020.

[6] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *TIP*, 2021.

[7] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation. In *ICIP*, 2012.

[8] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *TIP*, 24(11):3345–3356, 2015.

[9] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *Pacific Conference on Computer Graphics and Applications*, 2007.

[10] Stephen M Pizer, R Eugene Johnston, James P Ericksen, Bonnie C Yankaskas, and Keith E Muller. Contrast-limited adaptive histogram equalization: speed and effectiveness. In *Conference on Visualization in Biomedical Computing*, 1990.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241, 2015.

[12] Shen Zheng and Gaurav Gupta. Semantic-guided zero-shot learning for low-light image/video enhancement. In *WACVW*, 2022.