

Supplementary Materials

Architecture of Illumination Disentanglement Network

The detailed architecture of illumination disentanglement network is shown in Figure 9. It is made up of six modules, including one convolution layer, four ResBlocks and one activation layer. The first convolution layer calculates 16 feature maps with 7×7 kernels. The ResBlock is constructed by connecting ConnectUnit and ResidualUnit in series. The ConnectUnit in the first ResBlock consists of a convolution layer that can calculate 32 feature maps with 3×3 kernels. The ResidualUnit in the first ResBlock consists of two convolution layer that also can calculate 32 feature maps with 3×3 kernels. All the convolution layers in the second ResBlock can calculate 64 feature maps with 3×3 kernels. Analogously, the convolution layers in the third ResBlock can calculate 128 feature maps with 3×3 kernels. The convolution layers in the last ResBlock calculate 6 feature maps with 3×3 kernels. We use LeakyReLu to replace the conventional ReLu activation function in those four ResBlocks for avoiding vanishing gradient problem. The Sigmoid activation layer is adopted finally to avoid data overflow.

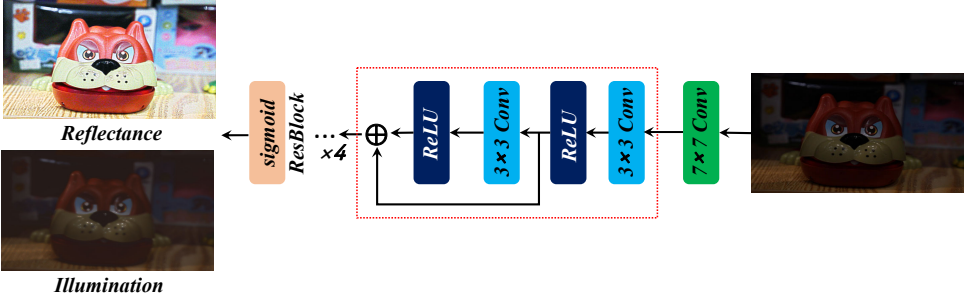


Figure 9: Structure of illumination disentanglement network

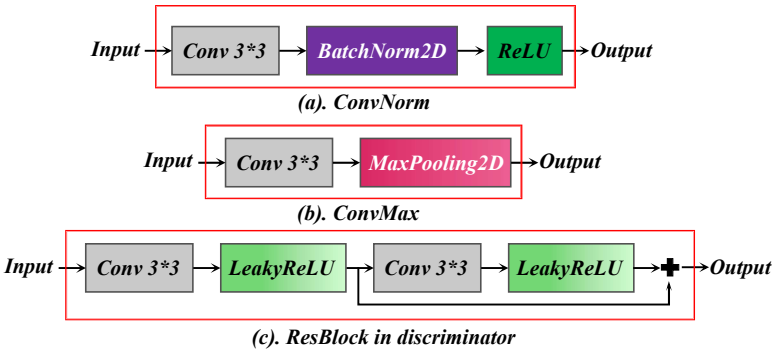


Figure 10: Structure of ConvMax, ConvNorm and Resblock in discriminators

Architecture of Global/Local Discriminator in DDFPN

The global discriminator is constructed for distinguishing reconstructed images from normal exposure images. It consists of four ResBlocks and final output can be obtained through

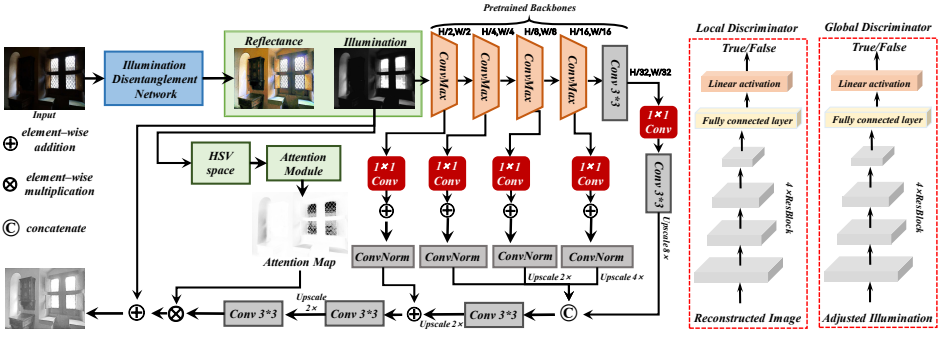


Figure 11: Structure of DDFPN based Illumination Brightening Network (We take the MobileNet pretrained backbone as an example): The detailed structures of those submodules (ConvMax, ConvNorm and ResBlock) are shown in **Supplementary Materials**.

this four ResBlocks and fully connected layers. We replace the sigmoid function with the least-square GAN[40] and the training loss of discriminator is:

$$J_D = (D_G(x_r) - 1)^2 + D_G(\mathcal{R} \otimes \mathcal{L}_a)^2 + \sum_{l_r \in \mathcal{P}_r} (D_L(l_r) - 1)^2 + \sum_{l_a \in \mathcal{P}} D_L(l_a)^2 \quad (15)$$

where \mathcal{P}_r stands for the set of image patches randomly cropped from the disentangled illumination \mathcal{L}_r . \mathcal{L}_r is disentangled from normal-exposed image x_r ($\mathcal{R}_r, \mathcal{L}_r = \mathcal{K}(x_r)$).

Tokenization in Transformer

To convert an image into tokens, a straightforward approach involves flattening the image into raw patches, as discussed in reference[41]. Given features of an image $\mathbf{F} \in \mathbb{R}^{H \times W \times C_F}$, it can be reshaped into a sequence of patches and treat them as tokens \mathbf{T} . We can control the dimension of tokens \mathbf{T} by downsampling the feature maps \mathbf{F} appropriately. Detail illustration can be seen in Figure 12.

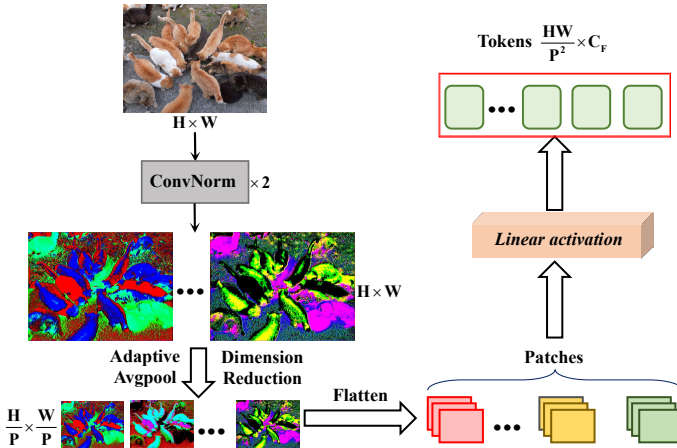


Figure 12: Detailed Introduction of Dimension Reduction in CRT

This tokenization strategy first use the two **ConvNorm** blocks to map the images into

feature maps. Then we exploit Adaptive AvgPooling operation to reduce the dimension of feature maps to alleviating memory usage, meanwhile reducing parameters for training. In our research, parameters $\frac{H}{P}$, $\frac{W}{P}$ and C_F are set to be 16, 16 and 36, respectively. The heads of transformer block are set to be 9. And the MLP dimension is set to be 64 as shown in Figure 3. Finally, we can derive the tokens $\mathbf{T} \in \mathbb{R}^{\frac{HW}{P^2} \times C_F}$ that will be passed as input value into the Transformer module shown in Figure 3.

In Figure 3, our designed Transformer contains a multi-head self-attention (MHSA) module and a Multi-Layer Perceptron (MLP) with skip connection. This modules adopt the GELU activation function and LayerNorm (LaN) as normalization. We can formulate the following equations and obtain the output of the Transformer \mathbf{y}_{out} as follows.

$$\begin{aligned} \mathbf{T}_0 &= \mathbf{T} \\ \tilde{\mathbf{T}}_i &= \text{MHSA}(\text{LaN}(\mathbf{T}_{i-1})) + \mathbf{T}_{i-1}, i = 1, 2, \dots, n \\ \mathbf{T}_i &= \text{MLP}(\text{LaN}(\tilde{\mathbf{T}}_{i-1})) + \tilde{\mathbf{T}}_{i-1}, i = 1, 2, \dots, n \\ \mathbf{y}_{\text{out}} &= \text{LaN}(\mathbf{T}_n) \end{aligned} \quad (16)$$

In MHSA, given a feature $X \in \mathbb{R}^{h \times w \times c}$ from LaN, the output of multiple heads attention are queue $Q = W_p^{qi} W_d^q X$, key $K = W_p^{ki} W_d^k X$ and value $V = W_p^{vi} W_d^v X$. The output of MHSA can be formulated as:

$$\tilde{X} = \text{softmax}(Q_r K_r / d) V_r + X \quad (17)$$

where $Q_r, V_r \in \mathbb{R}^{h \times w \times c}$ and $K_r \in \mathbb{R}^{h \times c \times w}$ are reshaped by Q, K, V . And we use Einstein's summation to calculate their product. d is a scale factor.

Analysis of CRT and Self-Supervised Training

Here we analyze how to derive our learning constraints and why embedding the composite curve in CRT is necessary for our self-supervised training strategy.

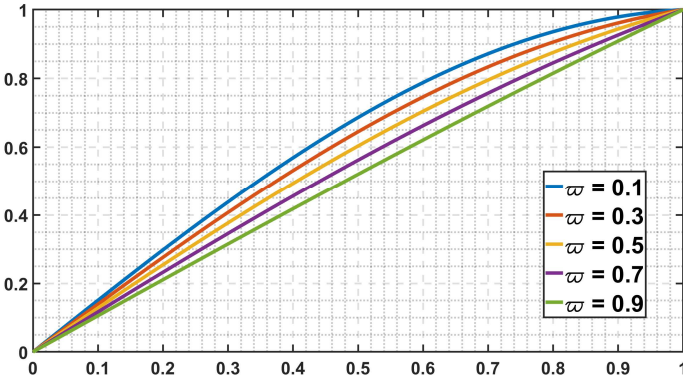


Figure 13: Composite curve embedding in CRT with different parameter ω

1) Composite Curve Embedding in CRT:

Assume that the parameter map and image irradiance are defined as ω and \mathbf{I} , the response value can be calculated as $\mathbf{I}' = (1 - \omega) \sin\left(\frac{\pi}{2} \mathbf{I}\right) + \omega \mathbf{I}$. The Taylor series of the sine function

at 0 can be expressed as

$$\begin{aligned} \sin\left(\frac{\pi}{2}\mathbf{I}\right) &= \frac{\pi}{2}\mathbf{I} - \left(\frac{\pi}{2}\right)^3 \frac{\mathbf{I}^3}{3!} + \left(\frac{\pi}{2}\right)^5 \frac{\mathbf{I}^5}{5!} + \\ &\dots + \left(\frac{\pi}{2}\right)^{2n+1} \frac{(-1)^n \mathbf{I}^{2n+1}}{(2n+1)!} + \mathcal{O}(\mathbf{I}^{2n+2}) \end{aligned} \quad (18)$$

Substituting Eq.(18) into our composite curve, thus we can derive that:

$$\begin{aligned} \mathbf{I}' &= \left(\frac{\pi(1-\varpi)}{2} + \varpi \right) \mathbf{I} + \chi \\ &= \left(\frac{\pi}{2} - 1 \right) (1-\varpi) \mathbf{I} + \chi \end{aligned} \quad (19)$$

where $\chi = (1-\varpi) \sum_{n=1}^{\infty} \left(\frac{\pi}{2}\right)^{2n+1} \frac{(-1)^n \mathbf{I}^{2n+1}}{(2n+1)!} + \mathbf{I}$.

Given an arbitrary image \mathbf{I} , the brightness and contrast adjustment formula can be expressed as follows.

$$\mathbf{I}' = \alpha (\mathbf{I} - \bar{\mathbf{I}}) + \beta \bar{\mathbf{I}} \quad (20)$$

where $\alpha > 0, \beta > 0$.

This adjustment formula is a linear equation which can be regarded as $\mathbf{I}' = \alpha \mathbf{I} + \gamma$ where $\gamma = (\beta - \alpha) \bar{\mathbf{I}}$. Assume that the mean value of \mathbf{I} can be calculated as $\mathbf{I} = \Gamma \bar{\mathbf{I}}$ and we can rewrite Eq.(19) as follows.

$$\begin{aligned} \mathbf{I}' &= \left(\frac{\pi}{2} - 1 \right) (1-\varpi) \mathbf{I} + \Gamma \bar{\mathbf{I}} \\ &+ (1-\varpi) \sum_{n=1}^{\infty} \left(\frac{\pi}{2} \right)^{2n+1} \frac{(-1)^n \Gamma^{2n+1} \bar{\mathbf{I}}^{2n+1}}{(2n+1)!} \end{aligned} \quad (21)$$

It is obvious that the role of Eq.(21) in adjusting image brightness and contrast is equivalent to linear equation $\mathbf{I}' = \alpha \mathbf{I} + \gamma$. In addition, we also present our designed composite curve with ϖ values ranging from -1 to 1 . It can be observed from Figure 13 that as the value of ϖ approaches -1 , the composite curve enhances the pixel brightness to a greater extent. When ϖ equals 1 , it is equivalent to making no adjustments to the image. Thus we can adjust the brightness and contrast of dim images by employing composite curves of varying orders. This enables the creation of a set of images with consistent content yet varying exposure levels. Moreover, the composite curves we devised obviate the requirement for computing the image mean value in each operation, a departure from conventional brightness adjustment defined in Eq.(20). Additionally, they enable adjustment of brightness and contrast through the modification of a single variable, leading to enhanced efficiency in terms of training parameters and computational performance. However, the pseudo-labeled images generated by the NRNB module, while achieving exposure level adjustments as demonstrated, often suffer from color shifts and discrepancies in contrast, saturation, and overall color fidelity when compared to real-world images, as illustrated in the Figure 3. Therefore, they can only serve as guiding pseudo-labels for self-supervised training.

2) Learning Constraints in Self-Supervised Training:

Consider the objective of illumination disentanglement. Given a sequence of images $\{\mathbf{I}_i\}_{i=1,2,\dots,m}$ with the same content but different exposure levels, we aim to learn a model \mathcal{K}

to disentangle the illumination from those images by minimizing the objective:

$$\min \sum_{i=1}^m \mathbb{E}_{\mathbf{I}_i \sim \mathcal{P}_I} \|\mathcal{K}^R(\mathbf{I}_i) \otimes \mathcal{K}^L(\mathbf{I}_i) - \mathbf{I}_i\|_1 \quad (22)$$

where $\mathcal{K}^R, \mathcal{K}^L$ are submodels in \mathcal{K} and \mathcal{P}_I is the distribution of arbitrary images.

This is the basic rule of illumination disentanglement that the reflectance and illumination maps of a specific image can reconstruct this image. Besides, there exists an vital property of illumination disentanglement that the reflectance of an image does not change due to changes in its exposure level. The minimization problem defined in Eq.(22) can be reformulated as follows.

$$\begin{aligned} \min \sum_{i=1}^m \mathbb{E}_{\mathbf{I}_i \sim \mathcal{P}_I} \|\mathcal{K}^R(\mathbf{I}_i) \otimes \mathcal{K}^L(\mathbf{I}_i) - \mathbf{I}_i\|_1 + \lambda_1 \min \sum_{i,j \in [1,m]}^{i \neq j} \mathbb{E}_{\mathbf{I}_i, \mathbf{I}_j \sim \mathcal{P}_I} \|\mathcal{K}^R(\mathbf{I}_i) \otimes \mathcal{K}^L(\mathbf{I}_j) - \mathbf{I}_j\|_1 \\ \text{s.t. } \|\mathcal{K}^R(\mathbf{I}_i) - \mathcal{K}^R(\mathbf{I}_j)\|_1 = 0 \end{aligned} \quad (23)$$

In that case, the objective function can be formulated as:

$$\begin{aligned} J = \sum_{i=1}^m \mathbb{E}_{\mathbf{I}_i \sim \mathcal{P}_I} \|\mathcal{K}^R(\mathbf{I}_i) \otimes \mathcal{K}^L(\mathbf{I}_i) - \mathbf{I}_i\|_1 + \lambda_1 \sum_{i,j \in [1,m]}^{i \neq j} \mathbb{E}_{\mathbf{I}_i, \mathbf{I}_j \sim \mathcal{P}_I} \|\mathcal{K}^R(\mathbf{I}_i) \otimes \mathcal{K}^L(\mathbf{I}_j) - \mathbf{I}_j\|_1 \\ + \lambda_2 \sum_{i,j \in [1,m]}^{i \neq j} \mathbb{E}_{\mathbf{I}_i, \mathbf{I}_j \sim \mathcal{P}_I} \|\mathcal{K}^R(\mathbf{I}_i) - \mathcal{K}^R(\mathbf{I}_j)\|_1 + \lambda_4 J_{tv} \end{aligned} \quad (24)$$

where $\lambda_4 = 0.0001$ and J_{tv} represents the total variation loss defined in traditional Retinex decomposition[1, 21]. The first term $\sum_{i=1}^m \mathbb{E}_{\mathbf{I}_i \sim \mathcal{P}_I} \|\mathcal{K}^R(\mathbf{I}_i) \otimes \mathcal{K}^L(\mathbf{I}_i) - \mathbf{I}_i\|_1$ corresponds to the \mathcal{L}_{sr} in our self-supervised training that is the most basic and indispensable.

When using Penalty or Augmented Largangian methods to solve minimization problems in Eq.(23), if the constraint factor λ_1, λ_2 is applied too large, it will make the \mathcal{K}_v^R and \mathcal{K}_v^L converge to a trivial solution as:

$$\mathcal{K}^{L*}(\mathbf{I}_i) = \mathbf{I}_i, \mathcal{K}^{R*}(\bullet) = \mathbf{1} \quad (25)$$

The illumination model \mathcal{K}^{L*} becomes an identity function and reflectance model converges to constants. Since we model \mathcal{K}^{R*} and \mathcal{K}^{L*} by end-to-end CNNs like autoencoders, it is extremely easy to optimize them to trivial solutions as long as the constraint weights are slightly unreasonable.

To overcome the above problem, we establish a learning constraint to facilitate our self-supervised training. We employ the Max-RGB method in[42, 43] to regularize the learning trajectory. We hope the solution of minimization problems in Eq.(23) should satisfy the following constraint:

$$\|\mathcal{K}^L(\mathbf{I}_i) - \max_{y \in \Omega} \max_{c \in \{R, G, B\}} \mathbf{I}_i\|_1 = 0 \quad (26)$$

where Ω stands for the 7×7 regions in the \mathbf{I}_i . And this constraint corresponds to the \mathcal{L}_{ig} in our self-supervised training.

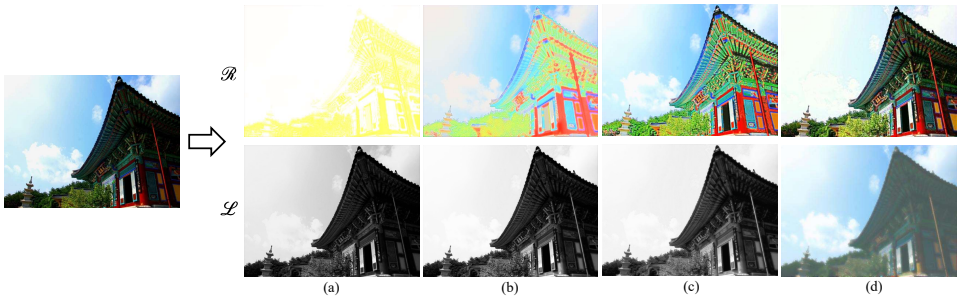


Figure 14: Optimization failures caused by unreasonable constraint weight settings. (a) $w/o \mathcal{L}_{ig}$, $\lambda_1 = 0.3$, $\lambda_2 = 0.01$. (b) $w/o \mathcal{L}_{ig}$, $\lambda_1 = 0.1$, $\lambda_2 = 0.001$. (c) $with \mathcal{L}_{ig}$, $\lambda_1 = 0.3$, $\lambda_2 = 0.01$. (d) Ours ($with \mathcal{L}_{ig}$, $\lambda_1 = 0$, $\lambda_2 = 0$)

Finally, according to [43], an effective solution for the illumination map must exhibit a smooth texture while retaining the overall structural boundaries. Therefore, the total variation constraint \mathcal{L}_{is} is also applied in our self-supervised training to preserve the reflectance structures while restraining marginal noise. Different to previous works [1, 21], we utilize the underexposed image directly, replacing the reflectance, to calculate the total variation loss.

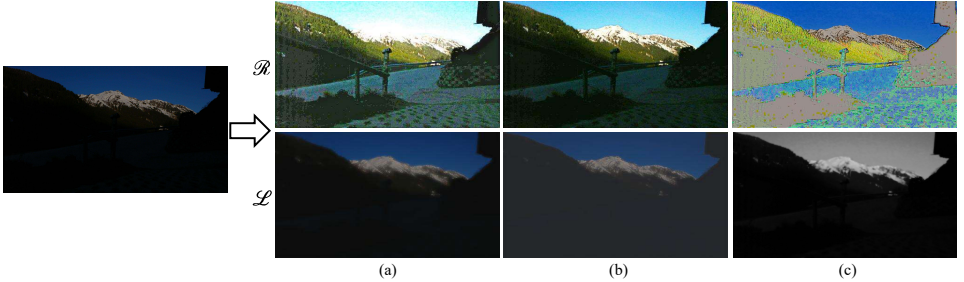


Figure 15: Optimization failures caused by improper total variation constraint. (a) $w/o \mathcal{L}_{is}$. (b) Ours with \mathcal{L}_{is} defined in Eq.(8). (c) \mathcal{L}_{is} defined in [1].

To verify our proposed learning constraints, we specially conduct two experiments. The results of the first experiment are shown in Figure 14. The examples in Figure 14 (a)-(c) are obtained through self-supervised training using objective function defined in Eq.(24). The superiority of the learning constraint proposed in our work is confirmed by comparing various values of λ_1 and λ_2 , as well as examining the impact of the new constraint established in Eq.(26) on the disentangled results.

According to the results in Figure 14, we can find that the self-supervised training easily converges to unintentional solutions when hyperparameters λ_1 , λ_2 are improper. After implementing the Max-RGB constraint, even when using inappropriate hyperparameter settings, the impact on the final Retinex decomposition result is minimized, ensuring consistently high visual quality. In simpler terms, the designed learning constraints alleviate the challenge associated with hyperparameter adjustments. Furthermore, to mitigate the model training’s reliance on pseudo labels, we eliminate two losses by setting both variables, λ_1 and λ_2 , to zero. Observing the results depicted in Figure. 14, we find a significant reduction

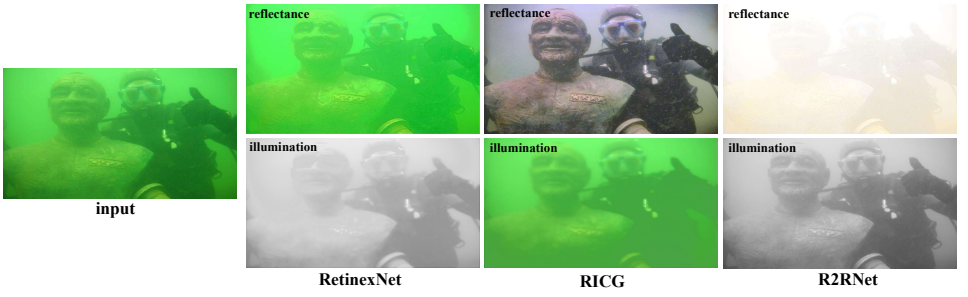


Figure 16: Performance comparison of illumination disentanglement (Part 1)

in artifacts in the disentangled reflectance when these two terms are eliminated, resulting in a more aesthetically pleasing and realistic output.

Another experiment is conducted to test the total variation constraint in our self-supervised training strategy. The specific results are shown in Figure. 15. From Figure. 15, it is evident that the total variation constraint designed in our paper ensures that there are no artifacts in the disentangled reflectance and reduces the noise within it.

Consequently, we use the linear combination of \mathcal{L}_{sr} , \mathcal{L}_{ig} and \mathcal{L}_{is} to guide self-supervised training. This improvement boosts the robustness of illumination disentanglement in RICG and reduces its vulnerability to unsuitable hyperparameters.

The pseudo codes of our self-supervised training and unsupervised training are shown in **Algorithm. 1**

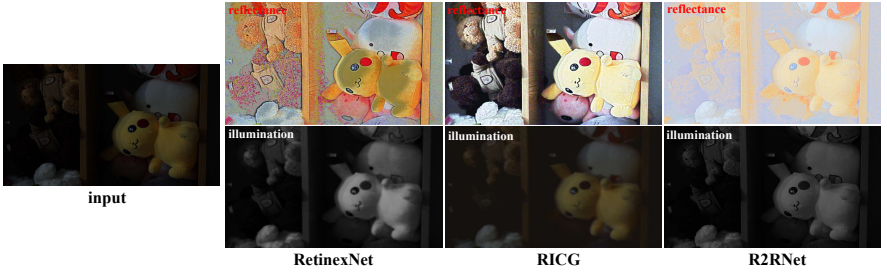


Figure 17: Performance comparison of illumination disentanglement (Part 2)

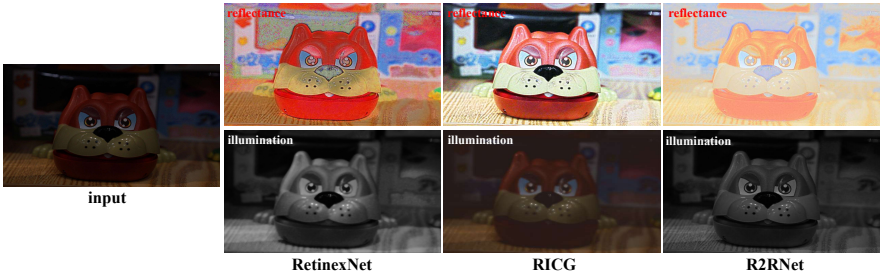


Figure 18: Performance comparison of illumination disentanglement (Part 3)

Additionally, we compared the illumination disentanglement performance of our proposed RICG, RetinexNet[1], and R2RNet[21]. Experimental results are illustrated in the Figure. 16-Figure. 19. Upon examination of each reflectance component depicted in the

Figure. 16-Figure. 19, it reveals that R2RNet[21] significantly reduces the contrast of reflectance components. Although the brightness of the images is increased, the visual quality still remains unsatisfactory. Conversely, RetinexNet[1] presents numerous artifacts in its disentangled results, making the reflectance more unauthentic. It is noteworthy that R2RNet[21] and RetinexNet[1] are ineffective in processing underwater distorted images. (Refer to Figure. 16 for further details.) Conversely, our approach demonstrates robust performance in achieving illumination disentanglement for underwater distorted images.

Implementation Specifications of Experiments

Parameter Settings: We implement our framework with Pytorch on two NVIDIA RTX 3090 GPUs. The batch size in training is set to be 32. The kernel weights and bias of each layer in the models (except for the downsampling in DDFPN) are initialized with standard zero mean and 0.1 standard deviation. Downsampling operations in DDFPN are initialized by different pretrained backbones. The optimizer used in our framework are all Adam optimizers with default parameters and learning rate 0.001 for illumination disentanglement network, 0.0001 for DDFPN. The weight parameters are set to be $\lambda_{is} = 0.1$, $\lambda_{ig} = 1$, $\lambda_{sr} = 1$, $\kappa_f = 1$, $\kappa_{adv} = 0.5$. The stages m of the self-supervised training in illumination disentanglement is set to be 6.

Compared Algorithms: We compare our RICG with many other mainstream image restoration algorithms: Restormer [6], UHDFour [2], IAT [44], URetinex-Net [24], ZeroDCE++ [20], EnlightenGAN [7], SCI [22], CLIT-LIP [28], RUAS [45], UNIE [29], NeRCO [5], Neural Preset [30], TUDA [10], USUIR [11] and PUGAN [33].

Ablation Study

Table 4: Influence of Different Backbones on Model Efficiency and Performance (Test images are selected from BAID dataset.)

	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FLOPs(G)	Pa(M)	IT(s)
MobileNet	23.45	0.802	0.308	49.46	1.18	0.0478
SENet	22.83	0.765	0.340	229.98	76.10	0.4208
DenseNet	23.04	0.784	0.368	134.16	10.21	0.1635
Inception-ResNet-v2	23.88	0.829	0.285	249.63	5.19	0.2192

Impact of Pretrained Backbones in FPN. The proposed RICG is a flexible image enhancement scheme since we can choose different pretrained backbones for the FPN in illumination generator. Here we conduct experiments where we retrain our model with different pretrained backbones in illumination generator. We present the results of the impact on the selection of pretrained backbones in Table 4. The MobileNet and Inception-ResNet-v2 are the most representative choices among which we can choose Inception-ResNet-v2 in applications with higher performance requirements, while in cases with high real-time requirements, we can choose MobileNet to serve as the backbones in FPN.

Impact of Stages in Self-Supervised Training. To investigate the impact of stages m in self-supervised training, we conduct an experiment where we retrained our model with different setting of m . In this ablation experiment, we adopt a new image quality evaluation

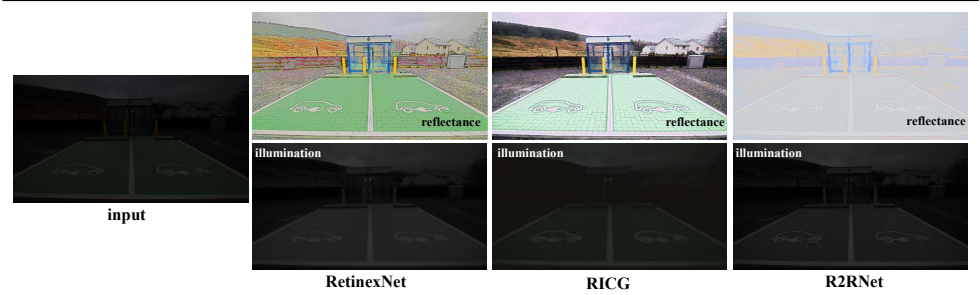


Figure 19: Performance comparison of illumination disentanglement (Part 4)

index termed visual information fidelity (VIF) [46]. The higher its value, the closer the target image is to the reference image. The experimental results are shown in Figure 20 and Table. 5. They also indicate that our RICG can obtain better performance with more stages of self-supervised training that means that more pseudo label images with different exposure levels generated by CRT help our RICG improve its ability to learn the mapping rules of illumination disentanglement.

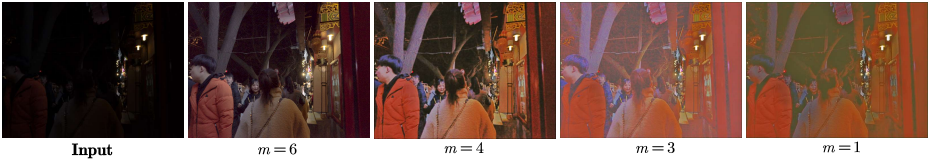


Figure 20: Ablation study on the stages m in self-supervised training

Table 5: Influence of stages m in self-supervised training (Test images are selected from LSRW datasets.)

	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	VIF \uparrow
Ours($m = 6$)	19.88	0.802	0.311	0.592
$m = 5$	19.32	0.795	0.386	0.563
$m = 4$	18.44	0.790	0.392	0.486
$m = 3$	18.32	0.775	0.423	0.465
$m = 2$	16.14	0.711	0.462	0.441
$m = 1$	16.14	0.711	0.607	0.392

Impact of local discriminator in DDFPN: We conduct the ablation study on impact of local discriminator in our DDFPN. The local discriminator operates on random image patches of the adjusted illumination, primarily aimed at preventing local underexposure during adjustment of illumination, thereby averting the occurrence of similar issues in the final reconstructed image. The visual comparison is given in Figure 21. It is obvious that some regions of images in row (c) suffer from severe color distortion and overexposure/underexposure issues. These results show that the local discriminator for disentangled illumination proposed in our research has excellent performance for model adaptation to underexposed image enhancement under various lighting conditions.



Figure 21: Visual comparison from the ablation study. (contribution of local discriminator) (a): Input. (b): Ours with local discriminator. (c): *w/o* local discriminator. Please zoom in view to see the details.

More Visualization Results

Visualization Results on OceanDark Datasets: The testing results are illustrated in Figure 22. Although the brightness of the enhanced version from SCI[22] and Neural Preset[30] has been restored, the overall color appears dimmed, leading to a reduction in visual quality. Unfortunately, enhanced results from IAT[44] method suffer from severe color deviation. Notwithstanding EnlightenGAN[7] also adopts a self attention mechanism, we are still able to observe in the enhanced images certain areas where the restoration of exposure levels is unreasonable, resulting in the presence of black patches. Similar issues also arise in the results of the STAR[47], with the difference being that the enhancement results of the Retinex method deviate more from the ideal enhancement results. The brightness of regions originally characterized by high luminosity decreases significantly, whereas areas initially possessing low luminosity exhibit abnormal increases in brightness. The CLIT-LIP method[28] and NeRCo approach[5] exhibit limited image restoration capabilities when applied to OceanDark datasets. This suggests that the post-enhancement brightness remains relatively low. Our RICG yields the most visually favorable results, whether it is applied to low-light images on land or underwater.

Visualization Results on Night Scene Image: Figure 23 and Figure 24 show more visual comparisons between the enhanced version of night scene images generated by our RICG and the compared methods. These findings demonstrate that the RICG we propose effectively enhances low-light images, avoiding issues related to overexposure or underexposure. In comparison to the methods employed for comparison, our approach achieves a superior level of naturalness in the enhanced images.

Visualization Results on LSRW dataset: Figure 25 show more visual comparisons between the enhanced results of low-light images in LSRW dataset. The results show that our RICG stores the color and the content of details in the underexposed regions most clearly and realistically, and the enhanced details have best and natural color contrast while keeping the normal-exposed background remain unchanged.

Visualization Results on DCIM dataset: Figure 26 show more visual comparisons

between the enhanced results of a backlit image in DCIM dataset. The illumination disparity among different regions within this backlit image is substantial, thereby increasing the complexity of the restoration process. Compared to other mainstream methods, the results achieved by RICG not only preserve the brightness in well-lit regions but also enhance the backlight areas moderately. This enhancement elevates the underexposed regions while simultaneously improving the overall contrast, thereby enhancing the visual quality.

Visualization Results on VV dataset: Figure 27 provides some enhanced performance comparisons between the enhanced results of underexposed images in VV dataset. Our RICG produces the most visually favorable results among those state-of-the-art algorithms.

Algorithm 1: Cooperative Game in RICG

Input: Unpaired Training dataset X , total number of paired training samples N , training steps N_{ep} , batch size n_b

Output: Image restoration model \mathcal{M}

- 1 calculate the number of times a sample needs to be traversed in each training step
 $n_{bs} = N // n_b$;
- 2 **for** $i = 1; i \leq N_{ep}; i++$ **do**
- 3 **for** $j = 1; j \leq n_{bs}; j++$ **do**
- 4 **if** $i < 40$ **then**
- 5 $d_{iter} = 3, \beta = 5$;
- 6 **else**
- 7 $d_{iter} = 6, \beta = 0.1$;
- 8 **if** $j // d_{iter} \neq 0$ **then**
- 9 $\mathcal{K}.\text{eval}(), \mathcal{T}.\text{eval}()$ (fix the parameters in \mathcal{K}, \mathcal{T} and denote them as α_d^*, ω^*);
- 10 Get a batch of distorted images x from unpaired image dataset X ;
- 11 Calculate the cooperative loss $\mathcal{L}_{\text{game}} = \mathcal{J}(\alpha_f) + \beta \mathcal{L}_D(\alpha_d^*, \omega^*)$;
- 12 Update the trainable parameters of the DDFPN using *Adam* optimizer
 $\alpha_f \leftarrow \alpha_f - \eta \nabla_{\alpha_f} \mathcal{L}_{\text{game}}(\alpha_f, \alpha_d^*, \omega^*)$;
- 13 **Continue**;
- 14 **else**
- 15 $\mathcal{K}.\text{train}(), \mathcal{T}.\text{train}()$;
- 16 Get a batch of distorted images x from unpaired image dataset X ;
- 17 Calculate the cooperative loss $\mathcal{L}_{\text{game}} = \mathcal{J}(\alpha_f) + \beta \mathcal{L}_D(\alpha_d, \omega)$;
- 18 Update the trainable parameters of the illumination disentanglement network, CRT and DDFPN using *Adam* optimizer;
- 19 $\alpha_f \leftarrow \alpha_f - \eta \nabla_{\alpha_f} \mathcal{L}_{\text{game}}(\alpha_f, \alpha_d, \omega)$;
- 20 $\alpha_d \leftarrow \alpha_d - \eta \nabla_{\alpha_d} \mathcal{L}_{\text{game}}(\alpha_d, \alpha_f, \omega)$;
- 21 $\omega \leftarrow \omega - \eta \nabla_{\omega} \mathcal{L}_{\text{game}}(\omega, \alpha_f, \alpha_d)$;
- 22 Update the global/local discriminators according to J_D defined in Eq.(15).

23 **Return** Image restoration model \mathcal{M} ;



Figure 22: Visualization results of our method and other state-of-the-art algorithms on OceanDark dataset[48]

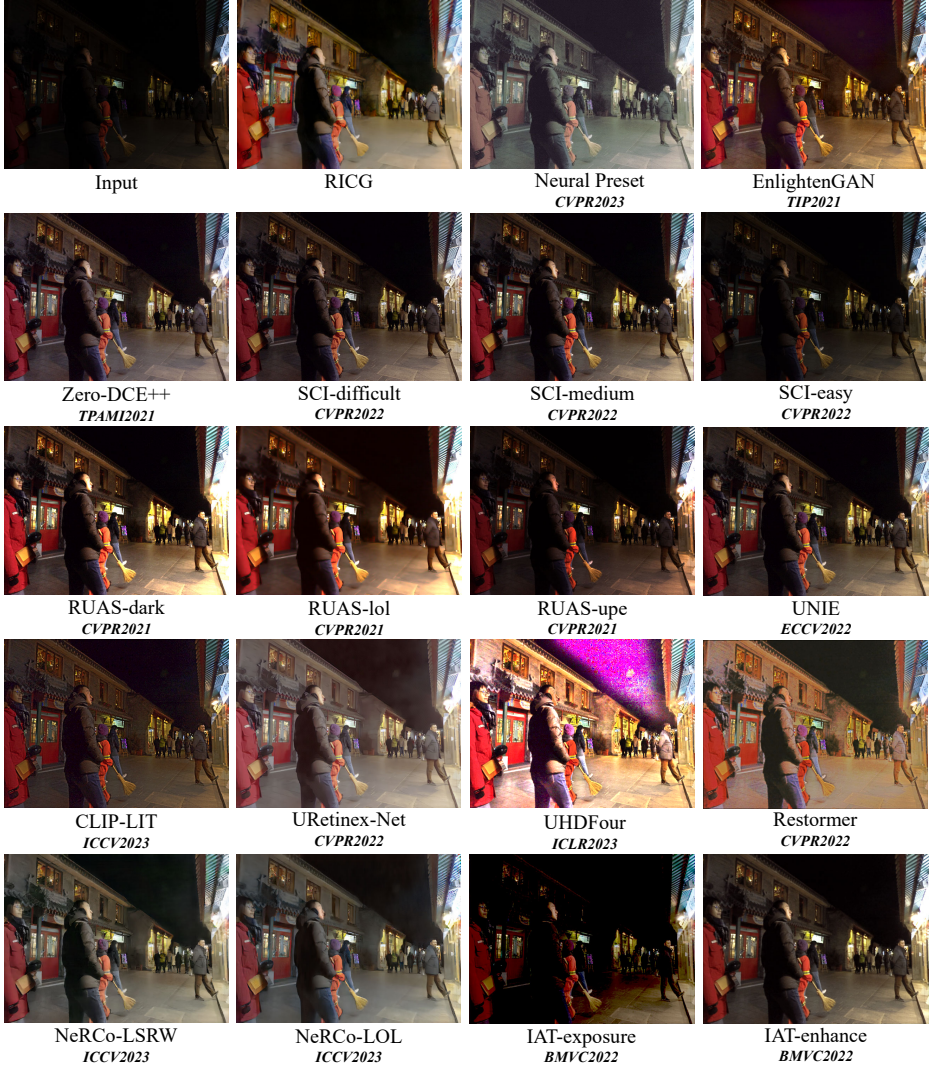


Figure 23: Visualization results of our method and other state-of-the-art algorithms on Dark-Face dataset[49]



Figure 24: Visualization results of our method and other state-of-the-art algorithms on Dark-Face dataset[49]

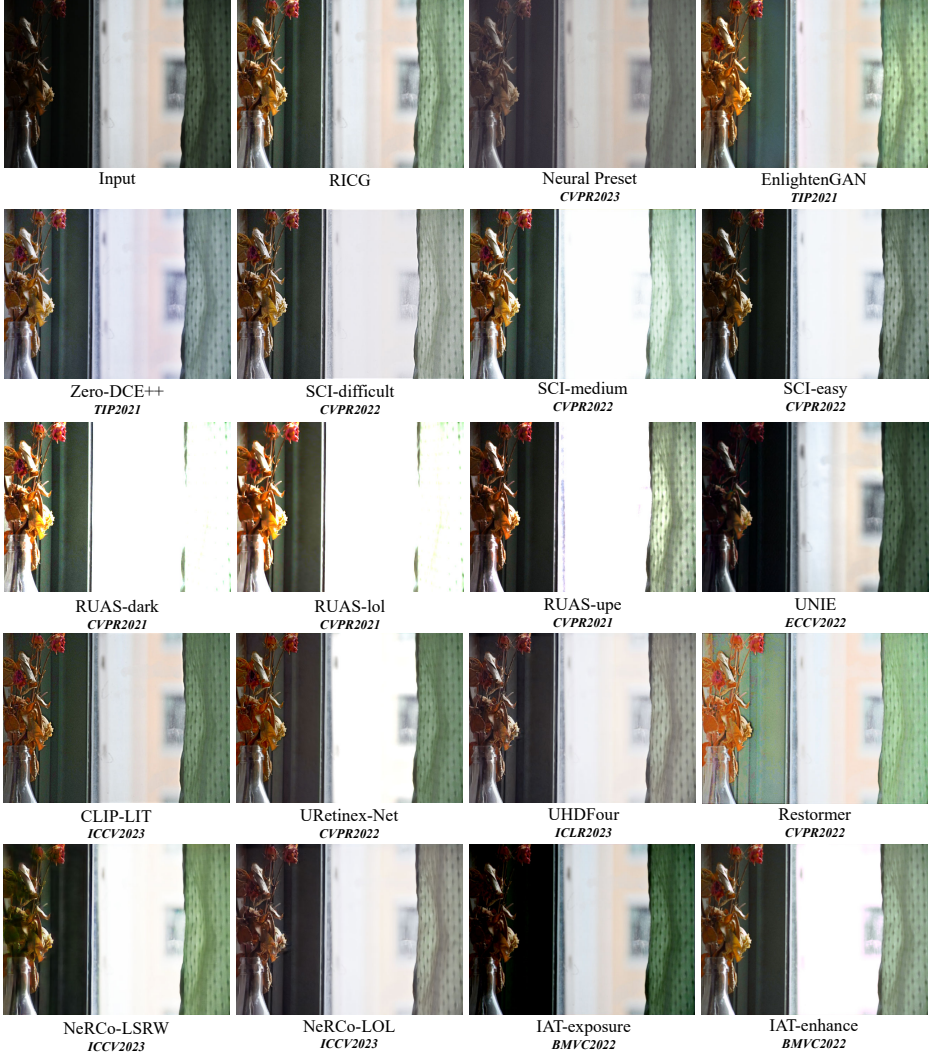


Figure 25: Visualization results of our method and other state-of-the-art algorithms on LSRW dataset[21]

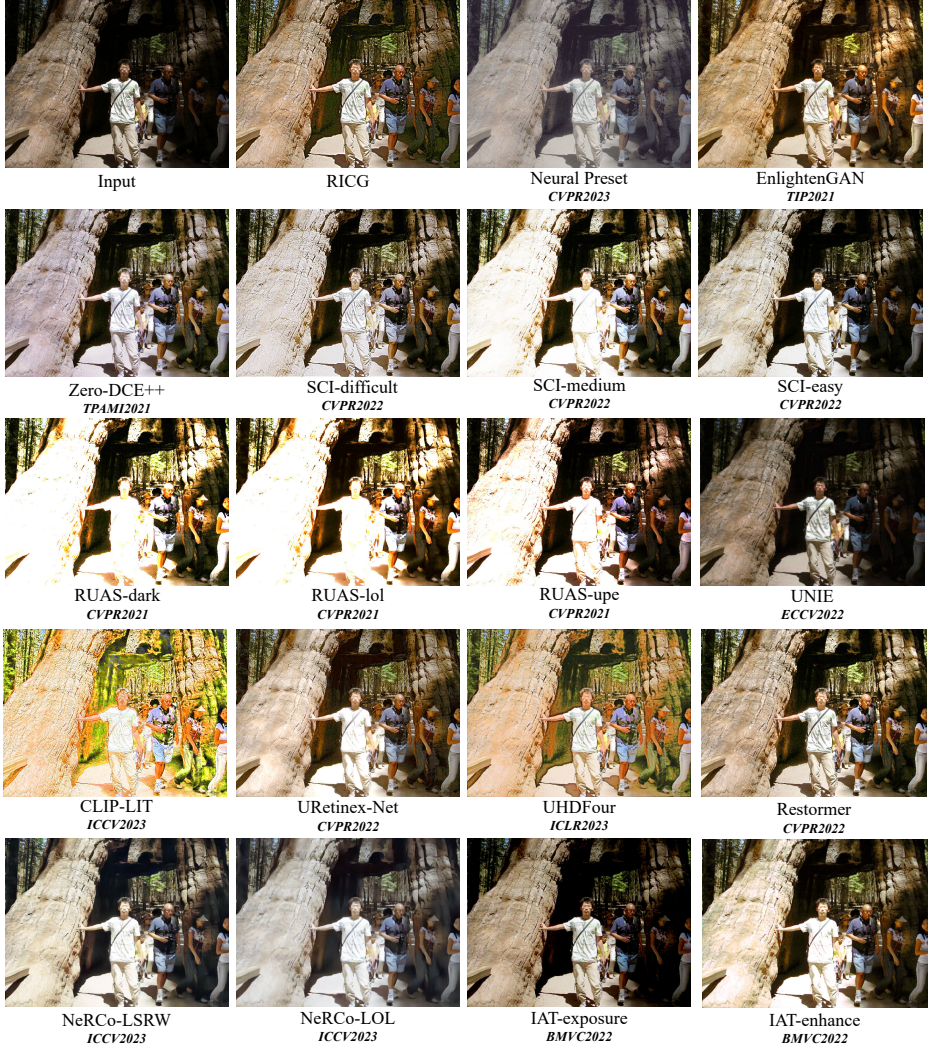


Figure 26: Visualization results of our method and other state-of-the-art algorithms on DCIM dataset[50]

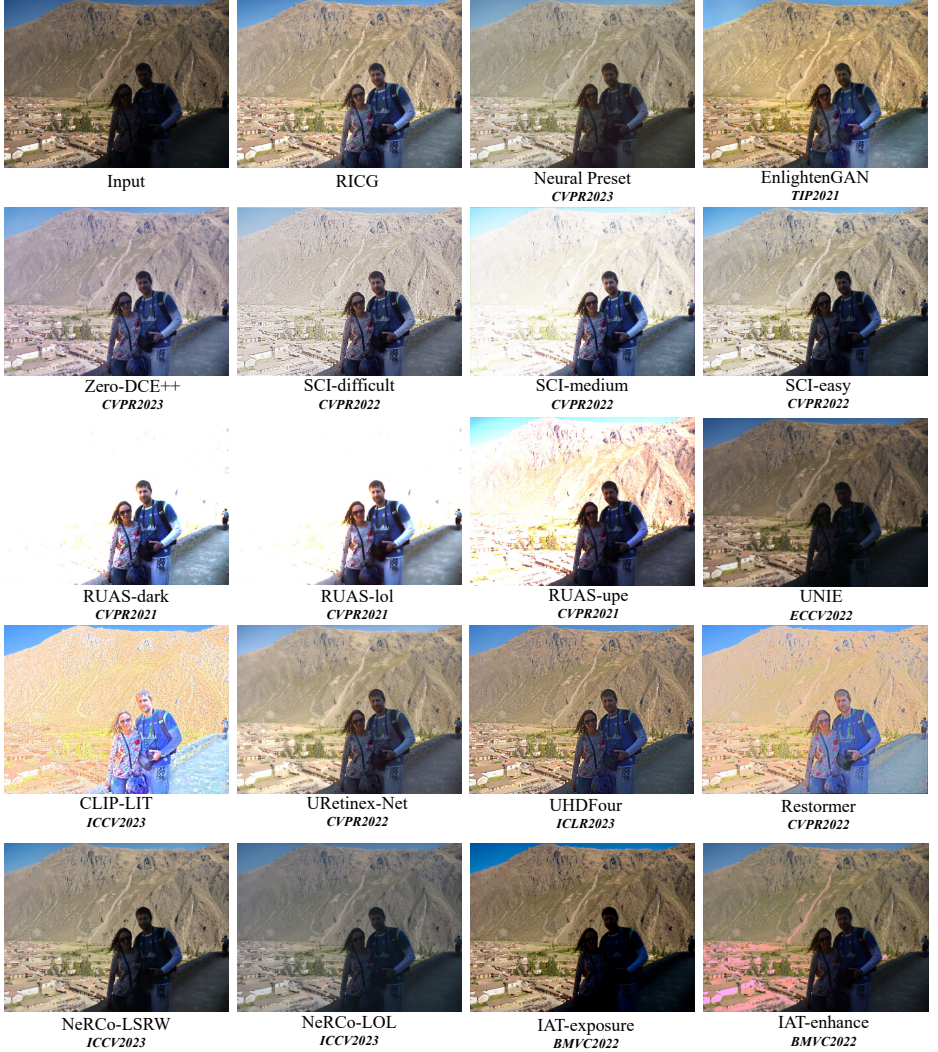


Figure 27: Visualization results of our method and other state-of-the-art algorithms on VV dataset

References

- [1] Wenhan Yang Jiaying Liu Chen Wei, Wenjing Wang. Deep retinex decomposition for low-light enhancement. In *Proc. Brit. Mach. Vis. Conf. (BMVC)*. British Machine Vision Association, 2018.
- [2] Chongyi Li, Chun-Le Guo, Man Zhou, Zhixin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. In *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2023.
- [3] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proc. AAAI Conf. (AAAI)*, volume 36, pages 2604–2612, 2022.
- [4] Haoyuan Wang, Xiaogang Xu, Ke Xu, and Rynson WH Lau. Lighting up nerf via unsupervised decomposition and enhancement. In *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pages 12632–12641, 2023.
- [5] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pages 12918–12927, October 2023.
- [6] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 5728–5739, 2022.
- [7] Y. Jiang, X. Gong, D. Liu, Y. Cheng, and Z. Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.*, 30:2340–2349, 2021.
- [8] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans. Image Process.*, 30:4985–5000, 2021.
- [9] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.*, 29:4376–4389, 2019.
- [10] Zhengyong Wang, Liquan Shen, Mai Xu, Mei Yu, Kun Wang, and Yufei Lin. Domain adaptation for underwater image enhancement. *IEEE Trans. Image Process.*, 32:1442–1457, 2023.
- [11] Zhenqi Fu, Huangxing Lin, Yan Yang, Shu Chai, Liyan Sun, Yue Huang, and Xinghao Ding. Unsupervised underwater image restoration: From a homology perspective. In *Proc. AAAI Conf. Artif. Intell.*, volume 36, pages 643–651, 2022.
- [12] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 30, 2017.
- [13] Jun Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. IEEE Int. Conf. Comput. Vision. (ICCV)*, pages 2223–2232, 2017.

- [14] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pages 35–51, 2018.
- [15] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018.
- [16] Michael D Grossberg and Shree K Nayar. Determining the camera response from images: What is knowable? *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(11):1455–1467, 2003.
- [17] Michael D Grossberg and Shree K Nayar. Modeling the space of camera response functions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(10):1272–1282, 2004.
- [18] Yurui Ren, Zhenqiang Ying, Thomas H Li, and Ge Li. Lecarm: Low-light image enhancement using the camera response model. *IEEE Trans. Circuits Syst. Video Technol.*, 29(4):968–981, 2018.
- [19] 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 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 1780–1789, 2020.
- [20] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(8):4225–4238, 2021.
- [21] Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. R2rnet: Low-light image enhancement via real-low to real-normal network. *J. Vis. Commun. Image Represent.*, 90:103712, 2023.
- [22] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 5627–5636, 2022.
- [23] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pages 1501–1510, 2017.
- [24] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 5901–5910, June 2022.
- [25] Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2024.
- [26] Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. *arXiv preprint arXiv:2403.12933*, 2024.

- [27] Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 36, 2024.
- [28] Zhixin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pages 8094–8103, 2023.
- [29] Yeying Jin, Wenhan Yang, and Robby T Tan. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pages 404–421. Springer, 2022.
- [30] Zhanghan Ke, Yuhao Liu, Lei Zhu, Nanxuan Zhao, and Rynson WH Lau. Neural preset for color style transfer. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 14173–14182, 2023.
- [31] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Trans. Image Process.*, 27(4):2049–2062, 2018.
- [32] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 586–595, 2018.
- [33] Runmin Cong, Wenyu Yang, Wei Zhang, Chongyi Li, Chun Le Guo, Qingming Huang, and Sam Kwong. Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators. *IEEE Trans. Image Process.*, 2023.
- [34] Codruta O Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Philippe Bekaert. Color balance and fusion for underwater image enhancement. *IEEE Trans. Image Process.*, 27(1):379–393, 2018.
- [35] Gaurav Sharma, Wencheng Wu, and Edul N Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Res. Appl.*, 30(1):21–30, 2005.
- [36] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE Trans. Image Process.*, 24(12):6062–6071, 2015.
- [37] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 2881–2890, 2017.
- [38] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 633–641, 2017.
- [39] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pages 10765–10775, 2021.

- [40] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proc. IEEE/CVF Int. Conf. Comput. Vis.(ICCV)*, pages 2794–2802, 2017.
- [41] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 12299–12310, 2021.
- [42] Edwin H Land. The retinex theory of color vision. *Sci. Am.*, 237(6):108–129, 1977.
- [43] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.*, 26(2):982–993, 2016.
- [44] Ziteng Cui, Kunchang Li, Lin Gu, Shenghan Su, Peng Gao, ZhengKai Jiang, Yu Qiao, and Tatsuya Harada. You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. In *Proc. Brit. Mach. Vis. Conf.(BMVC)*. BMVA Press, 2022.
- [45] Risheng Liu, Long Ma, Jiao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 10556–10565, 2021.
- [46] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Trans. Image Process.*, 15(2):430–444, 2006.
- [47] Jun Xu, Yingkun Hou, Dongwei Ren, Li Liu, Fan Zhu, Mengyang Yu, Haoqian Wang, and Ling Shao. Star: A structure and texture aware retinex model. *IEEE Trans. Image Process.*, 29:5022–5037, 2020.
- [48] Tunai Porto Marques and Alexandra Branzan Albu. L2uwe: A framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 538–539, 2020.
- [49] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 5525–5533, 2016.
- [50] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation. In *Proc. IEEE Int. Conf. Image Process.*, pages 965–968. IEEE, 2012.