Retinex-Inspired Cooperative Game Through Multi-Level Feature Fusion for Robust, Universal Image Enhancement

Ruiqi Mao
maoruiqi95@mail.nwpu.edu.cn
Rongxin Cui*
r.cui@nwpu.edu.cn

School of Marine Science and Technology
Northwestern Polytechnical University
Xi'an, P.R.China

Abstract

Existing approaches to enhancing distorted images frequently grapple not only with the dual challenges of optimizing visual fidelity and computational efficiency but also tend to be ineffectual in uncharted and intricate scenarios. Herein, we present a Retinex-inspired cooperative game based image restoration technique termed **RICG** to address the difficulty of navigating model performance and efficiency in different kinds of environments within a unified model. Specifically, we propose a two-step pipeline, comprising self-supervised illumination disentanglement and adjustment. The zero-shot illumination disentanglement is trained through a novel camera response Transformer (CRT), followed by illumination adjustment using a dual-discriminator feature pyramid network (DDFPN) incorporating an self-attention regularization. It is worth mentioning that we devise a specialized training process to reconstruct the optimal restored image through cooperative game. We substantiate the diverse advantages of RICG over existing methods through a meticulous and comprehensive evaluation process, illustrating its versatility in unexplored and convoluted circumstances. (Implementation code can be accessed at https://github.com/Ruiqi-Mao/RICG.)

1 Introduction

Image restoration endeavors to enhance the visibility of concealed information within distorted imagery, thereby enhancing overall image quality. This subject has garnered significant attention across various emerging computer vision domains. However, existing models are typically tailored and trained for specific domains, whereas the causes of image distortions vary significantly across diverse environments. Consequently, it is impractical for a unified model to comprehensively restore distorted images in diverse environments.

Existing image restoration methods [1, 2, 3, 4, 5, 6, 7, 8, 9], whether supervised or unsupervised, mostly only work for distorted images collected in a specific environment. For instance, RetinexNet [1] is only suitable for training on low-light images, while methods such as TUDA [10] and USUIR [11] are trained on underwater image datasets labeled with synthetic images, as illustrated in Figure 1. However, deep learning based domain adaptation methods like CycleGAN [12, 13], DRIT [14] and MUNIT [15] are all considered to

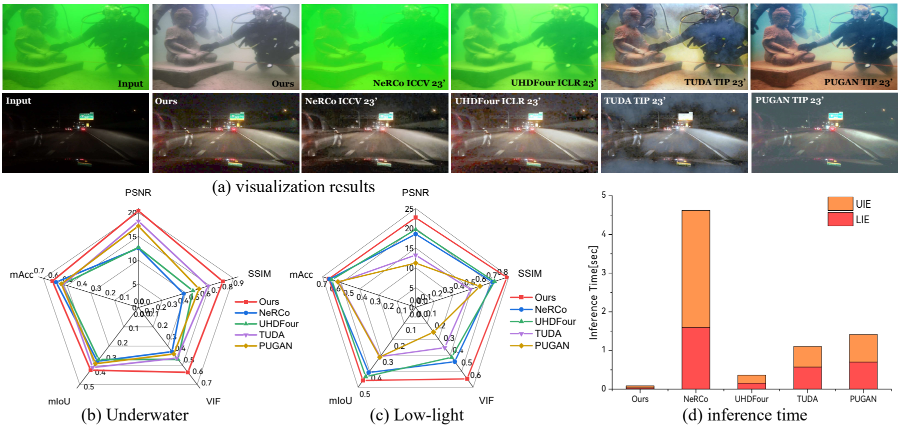

Figure 1: Comparison among recent state-of-the-art methods and our method in different environments. Visual quality comparison is shown in (a). Computational efficiency and numerical scores for five types of measurement metrics among three tasks including enhancement (PSNR, SSIM, and VIF) and segmentation (mIoU,mAcc) are shown in (b)-(d)

have significant potential for use in restoring distorted images in complex and variable environments, but they have some limitations. Firstly, they perform poorly when there are significant distribution differences between the target and source domains. Owing to the absence of ground-truth supervision information, they may generate artifacts and distortions, especially when the input images are of low quality or exhibit significant semantic differences. Most importantly, those models with a large number of trainable parameters and FLOPs that utilize cycle-consistency suffer from the hardship of significantly longer training time.

Inspired by Retinex theory, we specially develop an illumination disentanglement module that estimate illumination and reflectance of distorted images without any supervision from ground-truth images through a multi-stage zero-shot training process. That help us remove the influence of ambient illumination on distorted images. On the other hand, we utilize multi-level feature fusion method, based on DDFPN, to leverage the advantages of much more flexible, robust illumination adjustment. On this basis, a more robust image restoration under intricate scenarios is realized through a cooperative game between the illumination disentanglement and multi-level feature fusion.

Our contribution could be summarized as follow:

1): Motivated by the properties of nonlinear camera response models, we **first** design a novel data-driven camera response function based on a lightweight Transformer, named CRT, and successfully apply it to self-supervised training for multi-stage illumination disentanglement.

2): To reconcile the restoration of large and small objects in images, we propose a DDFPN motivated by multi-level feature fusion approaches. The global discriminator is employed to discern adjusted illumination, ensuring overall restoration, while a local discriminator operates on randomly sampled image patches within the reconstructed image, ensuring restoration of small-scale objects.

3): We propose a training strategy based on cooperative games, enabling the collaboration of two vital modules within the RICG framework to achieve optimal image restoration results. And comprehensive experimental results validate our method's robustness across various scenes including many application scenarios like underwater image enhancement, nighttime image enhancement and backlit images, etc.

2 Self-Supervised Illumination Disentanglement

In this section, we will introduce the data-driven illumination disentanglement module and its multi-stage zero-shot training process without any supervision from ground-truth images.

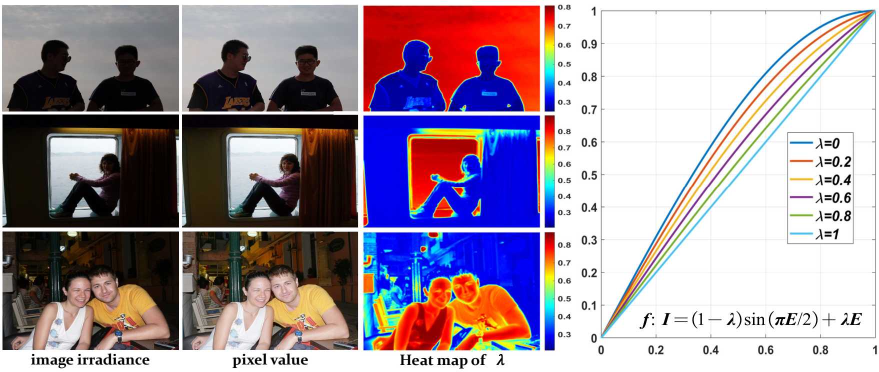

Figure 2: Here we show our designed CRF, image irradiance, their response pixel values and the heat map of parameters in our designed CRF.

2.1 CRT for Pseudo Label

Due to the lack of ground-truth images and no strong form of external supervision is available, we need to develop a module to generate pseudo labels for regularized self-supervised training. Motivated by camera manufacturers and their nonlinear in-camera processes termed camera response function (CRF) [16, 17, 18], we can develop a data-driven CRF to adjust exposure levels without altering the original image content, thereby effectively guiding self-supervised training. The design of such a function f in data-driven CRF needs to satisfy three vital properties as follows: 1) f is the same for all pixels on the sensor. 2) It is crucial that each pixel value in the pseudo labels should fall within the normalized range of $(0,1)$ that can be represented as $f \in [0,1]$. 3) f monotonically increases.

Under these assumptions, define $\mathscr{G}$ as the theoretical space of f:

$$\mathscr{G} := \{f | f(0) = 0, f(1) = 1, a > b, f(a) > f(b)\} \tag{1}$$

The most representative CRF is an empirical model called EMoR[18] by analyzing the real-world camera response curves. This approach applies Principal Component Analysis to the DoRF[17] database, which comprises 201 real-world response curves, and derives the eigenvectors of these curves. Assume that $E \in \mathbb{R}^{m \times n}$ is the light reaching the camera, i.e. the image irradiance. The EMoR can be represented as:

$$f(E) : P = f_0(E) + \sum_{n=1}^{M} c_n h_n(E) \tag{2}$$

where f_0 is the average curve of the DoRF and h_n is the n-th eigenvector. P represent the pixel value.

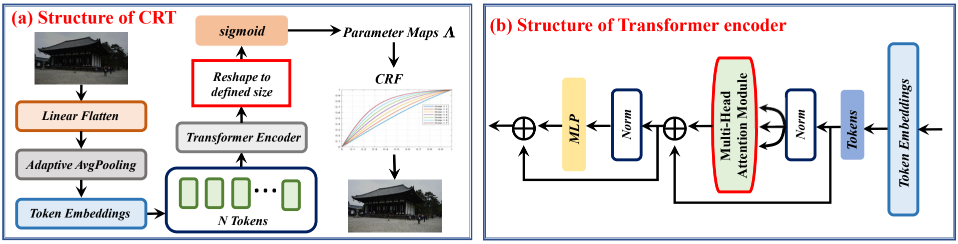

Figure 3: (a) The basic architecture of the proposed CRT that is established based on Transformer module and a specially designed composite curve. (b) The detailed structure of Transformer encoder is illustrated in this subfigure.

Compared to traditional CRFs, our proposed CRF focuses more on mapping image irradiance to pixel values of varying intensities rather than fitting the photodetectors of a specific camera model. Thus, we can design a novel CRF as follows: $f(E): P = (1-\lambda)\sin(\pi E/2) + \lambda E, \lambda \in (0,1)$

The introduction of the sine function term is intended to facilitate pixel value adjustments, while the value of λ determines the degree of irradiance preservation. As shown in Figure 2, a larger λ indicates a diminished influence of the sine function term, resulting in lesser changes to the irradiance. When $\lambda = 1$, the CRF is equivalent to an identity transformation.

Considering the refinement requirements of high-resolution images, λ should be a global parameter map of the same size as the image $f(E): P = (1-\Lambda)\sin(\pi E/2) + \Lambda E, \Lambda \in \mathbb{R}^{m\times n}$. The response of each pixel can be expressed as $p_{i,j} = (1-\lambda_{i,j})\sin(\pi e_{i,j}/2) + \lambda_{i,j} e_{i,j}$. $p_{i,j}$, $e_{i,j}$ and $\lambda_{i,j}$ are the elements in row i and column j of P, E and Λ respectively.

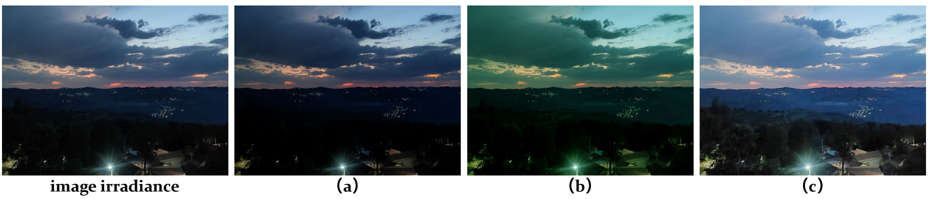

Figure 4: Visualization results of different CRFs: (a) EMoR* [18] (b) EMoR [17] (c) Our CRT

To enhance the flexibility of the CRF across various scenarios, we integrate our designed CRF with a Transformer encoder called camera response Transformer (CRT). The detailed structure of CRT is shown in Figure 3(a). Due to the adoption of globally adaptive parameters, our proposed CRT demonstrates a superior ability to maintain image semantics and color consistency compared to traditional CRFs like EMoR [17] and EMoR* [18], as illustrated in Figure 4. As observed in the heatmap in Figure 2, regions with lower irradiance exhibit smaller $\lambda_{i,j}$, indicating a lower degree of preservation of the original pixel values. Our proposed CRT emphasizes enhancing the mean pixel values while maintaining semantic and texture consistency, thereby generating pseudo labels for illumination disentanglement training.

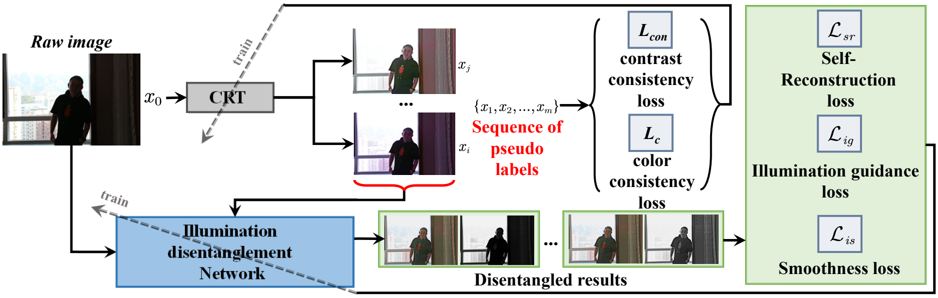

Figure 5: Our Self-Supervised Training Strategy

2.2 Self-Supervised Training Strategy

We specially develop a self-supervised training strategy to train the illumination disentanglement network, as shown in Figure. 5. Due to the absence of ground-truth images as reference, we transform the CRT designed in the previous section into a sequential form and it can generate pseudo label images with different exposure level. Thus, we can establish the following multi-stage training model:

$$\left\{ \begin{array}{c} x_t = f_{t-1}\left(x_{t-1}\right), f_{t-1} : x_t = \left(1-\Lambda_{t-1}\right)\sin\left(\frac{\pi}{2}x_{t-1}\right)+\Lambda_{t-1}, \Lambda_{t-1} = \mathrm{CRT}\left(x_{t-1}\right) \\ \mathscr{R}_t, \mathscr{L}_t = \mathcal{K}\left(x_t\right), t = 1,2,\cdots,m \end{array} \right. \tag{3}$$

where $\mathcal{K}(\bullet)$ stands for the forward function of illumination disentanglement network.

We devise several loss terms to regularize the learning. First, we need to restrict the response pixel value from CRT. A reasonable CRT need to maintain consistency with the contrast of raw images. Inspired by prior works [19, 20], we have the following constraint:

$$\boldsymbol{L}_{con} = \frac{1}{N_p}\sum_{t=1}^{m}\sum_{i=1}^{N_p}\sum_{j\in\Omega_i}\left[\left(P_{x_0}^{i}-P_{x_0}^{j}\right)-\left(P_{x_t}^{i}-P_{x_t}^{j}\right)\right]^2 \tag{4}$$

where N_p represent the number of pixels in the image. Ω_i denote the set of eight pixels around pixel i (up, down, left, right, upper left, upper right, lower left and lower right). P_{x_0} and P_{x_t} represent the pixels of x_0 and x_t.

The color distribution of response pixel values also need to be consistent with that of raw images. Thus we need to make learning constraints as follows:

$$\boldsymbol{L}_c = \sum_{t=1}^{m}\sum_{c\in\xi}\left(1-\mathcal{C}\left(x_0^c, x_t^c\right)\right) \tag{5}$$

where $\mathcal{C}$ represent the cosine similarity function. And ξ represent the RGB channels of pseudo labels.

Different to conventional Retinex decomposition[1, 21], we abandon the invariable reflectance constraint and do not directly use pseudo labels as supervision information for training. Thus we can derive the self-reconstruction loss as:

$$\mathcal{L}_{sr} = \sum_{t=1}^{m}\left|\left|\mathscr{R}_t\otimes\mathscr{L}_t - x_t\right|\right|_1 \tag{6}$$

The illumination guidance loss can be formulated on the basis of Max-RGB illumination:

$$\mathcal{L}_{ig} = \sum_{t=1}^{m} \left(||\mathscr{L}_t - \max_{p \in \mathbf{\Omega}} \max_{c \in \{R,G,B\}} x_t^c(p)||_1 \right) \tag{7}$$

where $\mathbf{\Omega}$ stands for the 7×7 regions in the x_t^c.

To preserve the structures of reflectance meanwhile restraining marginal noise, we propose an smoothness loss[1, 22].

$$\mathcal{L}_{is} = \sum_{t=1}^{m} \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_i} \varpi_{ij} |\mathscr{L}_t^i - \mathscr{L}_t^j|, \varpi_{ij} = \exp\left(\frac{\Delta_{ij}}{2\sigma^2}\right), \Delta_{ij} = \sum_c (x_i - x_j)^2 \tag{8}$$

where N denotes the total number of pixels of the image and $i-$ represents the ith pixel in the image. $\mathcal{N}_i$ is the adjacent pixels of i in 7×7 region.

The final loss of illumination disentanglement network is a linear combination of Eq.(4)-Eq.(8), as:

$$\boldsymbol{L}_D = \boldsymbol{L}_{con} + \boldsymbol{L}_c + \lambda_{is}\mathcal{L}_{is} + \lambda_{ig}\mathcal{L}_{ig} + \lambda_{sr}\mathcal{L}_{sr} \tag{9}$$

where positive constants $\lambda_{is}, \lambda_{sr}, \lambda_{ig}$ stand for weighting factors. Please refer to the analysis in **Supplementary Material** on why proposed self-supervised training strategy can guarantee a robust enhancement performance and superiority of our algorithm over RetinexNet [1] and R2RNet [21].

3 Flexible Illumination Adjustment

3.1 Unsupervised Training Loss

As shown the illustration of DDFPN in **Supplementary Material**, the disentangled results denoted as $\mathscr{R}$ (reflectance) and $\mathscr{L}$ (illumination). And $\mathscr{L}$ is the input of DDFPN. The adjusted disentangled illumination (output of DDFPN) is denoted as $\mathscr{L}_a$. The $\mathscr{L}_a$ is used to reconstruct the enhanced image. The downsampling of DDFPN can be flexibly switched by using a variety of pretrained backbones, and users can choose according to their different needs.

To ensure coherence between the enhanced semantic feature information of the image and the original content, we introduced a illumination map perception loss, guiding the model's learning process. Thus the loss term can be expressed as follows.

$$\mathcal{J}_f = ||\phi_4(\mathscr{R} \otimes \mathscr{L}_a) - \phi_4(x)||_1 + \sum_{i=1}^{4} ||\phi_i(\mathcal{K}(\mathscr{R})) - \phi_i(\mathscr{L}_a)||_1 \tag{10}$$

Following [23] we instantiate $\{\phi_i\}_{i=1}^4$ as *relu1-1*, *relu2-1*, *relu3-1* and *relu4-1* layers in VGG19.

Finally, we employ an adversarial loss to guide the training of DDFPN, allowing the synthesis of a counterfeit image that attains a level of realism comparable to real images. The global discriminator evaluates enhanced results to guide the DDFPN in generating more authentic enhanced images, while the local discriminator receives randomly cropped patches from the adjusted illumination, ensuring that their distribution closely matches that of the disentangled illumination from normal exposure image. It can be formulated as follows.

$$\mathcal{J}_{adv} = \left(D_G\left(\mathscr{R} \otimes \mathscr{L}_a\right) - 1\right)^2 + \sum_{l_a \in \mathcal{P}} \left(D_L\left(l_a\right) - 1\right)^2 \tag{11}$$

where D_G and D_L denote the global discriminator and local discriminator. And $\mathcal{P}$ denotes the set of randomly cropped patches from adjusted illumination $\mathscr{L}_a$.

Finally, the total loss function can be expressed as:

$$\mathcal{J} = \kappa_f \mathcal{J}_f + \kappa_{adv} \mathcal{J}_{adv} \tag{12}$$

where κ_f and κ_{adv} are positive constant which serve as the weights of the loss terms.

3.2 Cooperative Game

The illumination disentanglement module and DDFPN are both vital component in our image restoration scheme. Our primary objective is to investigate how two modules collaborate to decouple ambient illumination and autonomously adjust illumination, aiming to achieve more robust and flexible image restoration in unknown complex scenarios. Specifically, we formulate the training process of these two modules as a cooperative game and aim to solve the following optimization model:

$$\min_{\alpha_f} \left\{ \min_{\alpha_d, \omega} \mathcal{L}_{\texttt{game}}\left(\alpha_f, \alpha_d, \omega\right) \right\} \tag{13}$$

We denote $\mathcal{L}_{\texttt{game}}$ as a cooperative loss as follow:

$$\mathcal{L}_{\texttt{game}} := \mathcal{J}\left(\alpha_f\right) + \beta \boldsymbol{L}_D\left(\alpha_d, \omega\right) \tag{14}$$

where α_f are trainable parameters of DDFPN, α_d, ω are parameters of illumination disentanglement network and CRT. $\beta \geq 0$ denotes a trade-off parameter.

Our training strategy and logic are illustrated in the form of pseudo code shown in **Supplementary Materials**.

Table 1: Quantitative Comparison With State-of-the-Arts on the BAID, LSRW, UHD-LL. (The best result is in red whereas the second best one is in blue under each case. And green indicates the third best.)

Datasets	BAID			LSRW			UHD-LL			IT[sec]
Metrics	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	
URetinex-Net[24]	18.68	0.773	0.3081	14.78	0.661	0.4197	13.43	0.739	0.495	0.5785
UHDFour[2]	18.71	0.801	0.3176	18.20	0.656	0.3883	**18.33**	**0.855**	0.420	0.1088
LightenDiff [25]	19.88	0.855	0.3381	15.89	0.694	0.3563	16.23	0.789	0.447	0.0826
Wang et al. [26]	20.73	0.870	**0.2682**	16.05	0.706	0.3776	12.08	0.795	0.502	**0.0028**
GlobalDiff [27]	19.82	0.854	0.2986	13.82	0.685	**0.3279**	14.01	0.811	0.425	0.0976
CLIP-LIT [28]	22.35	0.862	0.3098	15.62	0.691	0.4087	13.12	0.651	0.470	0.1376
UNIE [29]	14.48	0.689	0.4628	10.35	0.562	0.4913	9.58	0.682	0.554	0.3675
NeRCo[5]	20.45	0.849	0.3281	14.20	0.653	0.4680	12.75	0.722	0.483	3.9918
Neural Preset[30]	18.05	0.726	0.3369	15.12	0.646	0.5091	12.36	0.708	0.582	0.0279
RICG	**22.65**	**0.886**	0.2927	**19.46**	**0.716**	0.3671	15.25	0.804	**0.413**	0.0306

4 Experiments

We assemble a mixture of 1000 distorted images and 1000 normal images from several datasets released in [1, 9, 31] to train our RICG model. All those training images are resized to the size of $400 \times 600 \times 3$. Our framework is implemented with Pytorch on two NVIDIA RTX 3090 GPUs. The model undergoes training for the initial 100 epochs using a learning rate of 0.0001, after which it proceeds with an additional 300 epochs during which the learning rate linearly decays to 0. We use the Adam optimizer and the batch size is set to be 32.

We evaluate our proposed RICG on benchmark datasets for both low-level and high-level vision tasks under different illumination conditions. Two low-level vision tasks include: (1) low-light image enhancement. (2) underwater image enhancement. One high-level visions tasks include: (3) semantic segmentation.

More details and hyperparameters settings please refer to **Supplementary Materials**.

Figure 6: Visualization results on our NCampus dataset.

4.1 Low-Light Image Enhancement

To assess the robustness of the proposed RICG across diverse real-world scenarios, we employ our self-collected NCampus dataset that are captured in wild and harsh environment to demonstrate the enhancement results obtained by different algorithms. The enhancement results are shown in Figure 6. We observe that LightenDiff [25], Wang et al. [26] and GlobalDiff [27] have made some progress in enhancing the brightness of low-light images. However, these methods still suffer from low visibility and subpar visual quality. The enhanced results generated by those algorithms, including NeRCo[8], URetinex-Net[24] and UNIE[29], manifest subpar visual quality, marked by notable deficiencies in brightness and clarity. This deficiency stems from the severely restricted capacity of these three methods to accurately restore authentic nighttime images captured in the wild environment. By observing the zoomed-in-view region, it can be noted that RICG yields clearer details and higher restoration quality compared to CLIT-LIP[28].

To assess the quantitative comparison of our experimental results, we employ three full-reference image quality evaluation (IQA) metrics including PSNR, SSIM and LPIPS to compare the performance of our RICG with many mainstream algorithms on BAID[28], LSRW[21] and UHD-LL[2]. In addition, we use the inference time (IT) to measure efficiency of proposed RICG and other algorithms. The detailed comparison results with respect to IQA metrics is presented in Table. 1. As shown in Table. 1, our method achieves the best results for the SSIM, PSNR metrics on the BAID[28], LSRW[21] datasets. Particularly, our method exhibits the best average values for SSIM and PSNR metrics across those three datasets. Regarding LPIPS metrics[33], our method also demonstrates competitive performance compared to state-of-the-art alternatives. It remains highly competitive and effectively balances model performance and efficiency.

4.2 Underwater Image Enhancement

In underwater image restoration tasks, we use Neural Preset[30], TUDA [10], USUIR [11] and PUGAN [34] to compare with our RICG. Testing images are selected from **URPC dataset**[1] and **Color-Checker7** dataset [35].

As depicted in Fig. 7, USUIR [11] exhibits poor performance when faced with severely distorted underwater images. Its ability to correct color shifts in such conditions is relatively weak. In contrast, PUGAN [34] demonstrates a significantly more comprehensive improvement in severe color shifts. However, its enhancement of contrast is not pronounced, leading to suboptimal visual quality. The performance of TUDA [10] is comparable to that of RICG. However, upon closer examination, some minor artifacts can be observed in TUDA's handling of certain image details. Meanwhile, in the upper right corner of the image results, we annotate the evaluation metric scores for each image. The first row is derived from the **Color-Checker7** dataset [35], where we assessed using the CIEDE2000 [36]. The second row represents quantitative comparisons conducted through UCIQE [37]. Our method exhibits the lowest CIEDE2000 [36] and the largest UCIQE [37] scores, indicating that the enhanced images from RICG have the smallest deviation from the reference image. Table. 2 lists the results of quantitative comparison and our method has competitive performance with state-of-the-art alternatives.

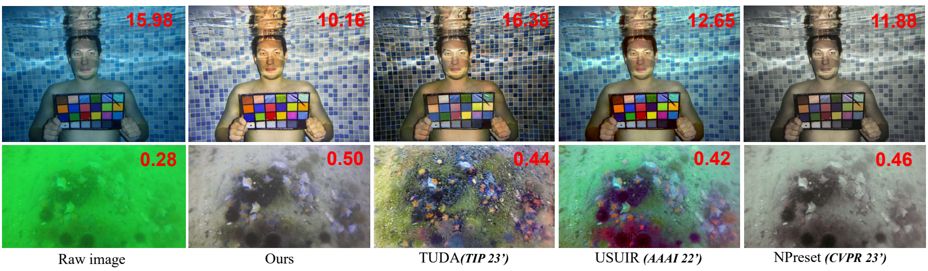

Figure 7: Ablation study on learning constraints in self-supervised training.

4.3 High-Level Vision Tasks

We utilize the PSPNet[38] as the benchmark to assess segmentation performance by employing the "pre-train + fine-tune" pattern, analogous to the methodology utilized in [22].

[1]https://aistudio.baidu.com/datasetdetail/228251

Table 2: Quantitative Comparison With State-of-the-Arts.

Metrics / Methods	UCIQE↑	UIQM↑	UIQM			CCF↑	CCF		
			UICM	UISM	UICONM		Colorfulness	Contrast	Fogdensity
USUIR [11]	0.59	5.0744	2.78	7.39	0.787	29.2048	20.10	37.38	7.60
PUGAN [34]	0.58	4.7900	2.19	7.27	0.722	28.8607	14.97	38.02	**8.08**
Neural Preset [30]	0.54	5.0795	1.98	7.75	0.765	27.4928	14.01	36.53	7.26
TUDA [10]	**0.62**	5.0952	2.75	7.33	**0.798**	30.1717	**20.13**	38.86	7.74
Ours	0.60	**5.2013**	**2.86**	**7.80**	0.788	**30.8307**	18.21	**40.54**	7.62

We conduct image segmentation test using the ADE20K dataset[39] and ACDC dataset[40]. Specifically, we employ an image rendering model[30] to render images from the ADE20K dataset[39] as underexposed images. Subsequently, these underexposed images are restored using image restoration techniques and then input them into the PSPNet[38] to obtain segmentation results. Figure. 8 and Table. 3 demonstrate the results of quantitative and qualitative comparison among different methods. Our performance surpasses that of other state-of-the-art methods by a significant margin. It can be seen from Figure. 8 that our RICG can restore the image with the highest visual quality from the distorted image, so it has the highest accuracy of segmentation results.

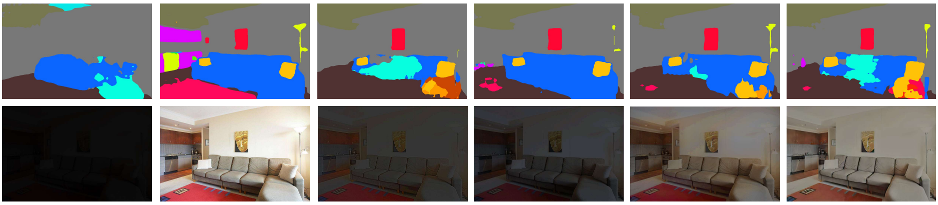

Figure 8: Visual results of semantic segmentation on the ADE20K dataset[39].

Table 3: Quantitative results of semantic segmentation

Methods	**RICG**	NeRCo[5]	TUDA [10]	CLIP-LIT [28]	SCI [22]	UHDFour[2]	PUGAN[34]
mIoU	**0.4667**	0.3920	0.3728	0.3896	0.4343	0.4533	0.3804
mAcc	0.6067	0.5541	0.5446	0.5729	0.6011	**0.6093**	0.5259
aAcc	**0.7925**	0.7623	0.7015	0.7419	0.7535	0.7336	0.7126

5 Conclusion

We have developed a versatile image restoration framework trained on unpaired data, which demonstrates enhanced robustness and faster performance in complex and changeable environments. The primary innovation of the RICG method involves a cooperative game between CRT-assisted multi-stage illumination disentanglement through self-supervised training and multi-level feature fusion-driven DDFPN. Sufficient experimental results on various kinds of distorted images demonstrate that our approach outperforms multiple state-of-the-art methods across both subjective and objective metrics in wild environment. In our future endeavors, we will explore methods to control and adjust image restoration style based on user preference within a unified model.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant U22A2066, Grant 61733014 and Grant U21B2047.

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, Zhexin 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] Zhexin 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] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE Trans. Image Process.*, 24(12):6062–6071, 2015.

[38] 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.

[39] 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.

[40] 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.