TPCNet: A Low-Light Image Enhancement Network Inspired by Triple Physical Constraints

Jing-Yi Shi[1,2] Ming-Fei Li[1,2⋆] Ling-An Wu[1,2]

[1] Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China
[2] School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract. Low-light image enhancement is an essential computer vision task to improve image contrast and to decrease the effects of color bias and noise. Many existing interpretable deep-learning algorithms exploit the Retinex theory as the basis of model design. However, previous Retinex-based algorithms, that consider reflected objects as ideal Lambertian ignore specular reflection in the modeling process and construct the physical constraints in image space, limiting generalization of the model. To address this issue, we preserve the specular reflection coefficient and reformulate the original physical constraints in the imaging process based on the Kubelka-Munk theory, thereby constructing constraint relationship between illumination, reflection, and detection, the so-called triple physical constraints (TPCs) theory. Based on this theory, the physical constraints are constructed in the feature space of the model to obtain the TPC network (TPCNet). Comprehensive quantitative and qualitative benchmark and ablation experiments confirm that these constraints effectively improve the performance metrics and visual quality without introducing new parameters, and demonstrate that our TPCNet outperforms other state-of-the-art methods on 10 datasets. The code is available at https://github.com/2020shijingyi/TPCNet

Keywords: Low-Light Image Enhancement · Retinex theory · Physical constraints

1 Introduction

A major challenge for computer vision is how to overcome the low signal-to-noise ratio and inaccurate colors of images caused by underexposure [29, 44]. For this reason, low-light image enhancement (LLIE), which aims to improve image contrast while minimizing the effects of color bias, is an essential task [11]. During the past decades, conventional non-learning methods, like gamma correction [35] and histogram equalization [1, 15], have been widely applied for low-light enhancement, performing transformation on pixel values or adjusting

⋆ Corresponding author. Email: mf_li@iphy.ac.cn

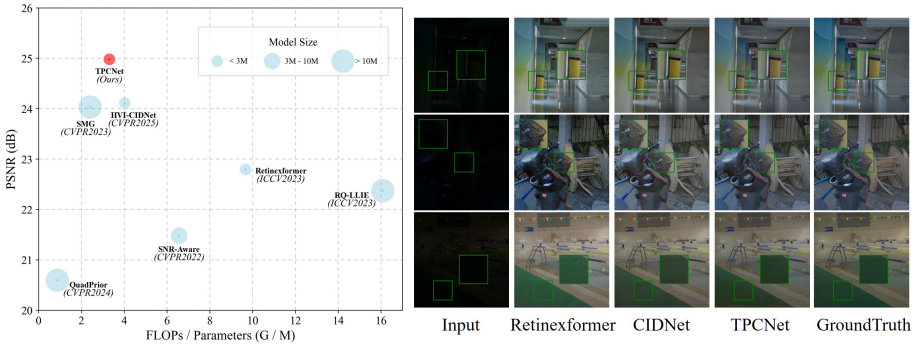

Fig. 1: Comparison with recent SOTA methods including CIDNet (HVI [51], CVPR2025), QuadPrior (Quadruple Priors [43], CVPR2024), SMG (SMG-LLIE [50], CVPR2023), Retinexformer [3] (ICCV2023), RQ-LLIE [25] (ICCV2023), and SNR-Aware (SNR-Net [49], CVPR2022) for Low-light Image enhancement on LOLv2-real [53] benchmarks. Our TPCNet demonstrates efficient network design and performance, and we exhibit the visualization of the top three methods.

their distribution to realize dynamic range expansion and contrast adjustments. In addition, various Retinex-based methods [11, 12] decompose low-light images into illuminated and reflected components and enhance the reflected component or optimize the illuminated features to realize enhancement. The above methods all exhibit interpretable physical processes, but these artificially designed constraints or priors may cause color bias and noise as their adaptability is insufficient for LLIE when dealing with complex illumination.

With the development of deep learning, convolutional neural networks (CNNs) and vision transformers (ViTs), have been applied in various tasks and have made significant advancements [22,34,39,45,55]. For LLIE, deep learning combined with an interpretable physical model has gradually evolved as an effective approach, rather than viewing the learnable modules as a simple black-box. Recently, COMO-ViT [46] constructed an illumination-aware gamma correction module whose gamma factor is calculated by the learnable modules to avoid the bias caused by manual design. Meanwhile, the deep learning framework inspired by the Retinex theory utilizes different CNNs (or ViTs) as the independent estimators for illumination $\hat{\mathbf{L}}$ and reflectivity $\hat{\mathbf{R}}$ and computes the enhanced images by $\hat{\mathbf{L}} \odot \hat{\mathbf{R}}$ ($\odot$ denotes the element-wise multiplication). Under the guidance of this theory, various methods [3,48,54,58] have been proposed; however, Retinex-based learnable methods typically regard the image as the invariant and decompose it into reflectivity and illumination components. To construct the explicit physical constraint ($\mathbf{I} = \hat{\mathbf{L}} \odot \hat{\mathbf{R}}$) between reflectivity and illumination, which is based on actual space ($\mathbf{R} \in \mathcal{R}^{3\times H\times W}$ and $\mathbf{L} \in \mathcal{R}^{1\times H\times W}$) and relies on input $\mathbf{I}$, it is necessary to supervise reflectivity (or illumination) features during training with a complex loss function [7, 34].

Inspired by physical quadruple priors [20, 43], which extract illumination invariant features from input, we exploit the intrinsic features in the imaging process to construct implicit constraints, avoiding the need for complex loss func-

tions to explicitly supervise process variables during training, thereby reducing the supervision complexity. Specifically, we reformulate the physical constraints in the imaging process based on the Kubelka-Munk (KM) theory [9, 10] and construct the triple physical constraint (TPC) among the detection $\hat{\mathbf{E}}$ (of the reflected light), illumination $\hat{\mathbf{e}}$ (by the light source), and reflectivity $\hat{\mathbf{R}}$ (of the material). Different from the previous Retinex theory, which only considers ideal Lambertian reflection, we retain the specular reflection coefficient and replace it with learnable variables, thus allowing the $\hat{\mathbf{e}}$ to characterize two complementary illumination components $\hat{\mathbf{L}}$ and $\hat{\tilde{\mathbf{L}}}$. Guided by this TPC theory, we design the reflectivity feature estimators (RFE) and light feature estimators (LFE) to estimate the initial $\hat{\mathbf{R}}$ and $\hat{\mathbf{L}}$, respectively, while the initial estimates are enhanced by utilizing the backbone structure, a Dual-Stream Cross-Guided Transformer (DCGT) whose core component is a Cross-Guided Multi-Head Self-Attention (CG-MSA). Based on the relationship between $\hat{\mathbf{E}}$, $\hat{\mathbf{e}}$, and $\hat{\mathbf{R}}$, we construct constraints in an implicit feature space ($\mathcal{R}^{C\times H\times W}$) rather than in the image space, which can be easily extended to different datasets and improves model robustness. Finally, we derive our method, TPCNet, by combining the core components RFE, LFE, and DCGT with the Color-Association Mechanism (CAM), which is exploited to decrease color bias. As shown in Fig. 1, our TPCNet outperforms recent state-of-the-art (SOTA) lightweight approaches on the real-image dataset LOL-v2-real [53].

Our contributions can be summarized as follows:

- We reformulate the origin of physical constraints in the imaging process, based on the KM theory, and construct the TPC among the detection, illumination, and reflectivity. This offers an effective and interpretable direction to design robust networks.
- We further propose the TPCNet, a novel LLIE CNN-Transformer framework, which establishes the physical constraints in an implicit feature space to improve its robustness and generalization. Although the TPCNet possesses not so many parameters (2.62M), it exhibits computationally-efficient capacity (only 8.68GFLOPs), benefiting from the efficient design of CG-MSA, which consumes only 25% of the computation of conventional multi-head self-attention (MSA) with the same input size.
- Quantitative and qualitative experiments results demonstrate that our TPCNet outperforms other SOTA approaches on different metrics across 10 datasets.

2 Related Work

2.1 Low-Light Image Enhancement via Physical Modelling

Retinex-based model. The Retinex theory [18], a human visual simulation model widely applied in LLIE, is an effective enhancement algorithm. It regards the observed image as one that can be decomposed into illuminated and reflected components: $\mathbf{I} = \hat{\mathbf{L}} \odot \hat{\mathbf{R}}$, where $\hat{\mathbf{L}}$ and $\hat{\mathbf{R}}$ refer to the illumination and the reflectance, respectively. Early approaches based on Retinex viewed the

reflectance as the final enhanced result, usually causing the result to look unnatural and to be excessively enhanced, *e.g.*, single-scale Retinex [12]. Later, Guo *et al*. proposed the LIME algorithm [11], which enhances the image via illumination map estimation, improving the visual aesthetics and imaging efficiency [12, 32]. Recently, deep learning combined with interpretable physical modeling has been increasingly seen as an effective direction for LLIE. RetinexNet [48], which combines the Retinex model with data-driven schemes, realizes more accurate reconstruction and more efficient image decomposition than conventional Retinex algorithms by utilizing elaborately hand-crafted loss functions. The above Retinex Imaging model assumes images are corruption-free for decomposing. Corruption-free images cannot be achieved in the inadequate lighting scenes and flaws in the imaging system. To overcome this limitation, Li *et al*. [23] proposed a robust Retinex model and pointed out that for real environments the model should include a noise term $\mathbf{N}$ as follows: $\mathbf{I} = \hat{\mathbf{L}} \odot \hat{\mathbf{R}} + \mathbf{N}$. Then, Cai *et al*. [3] formulated a simple yet one-stage Retinex-based framework $\mathbf{I} = (\hat{\mathbf{L}} + \delta_{\hat{\mathbf{L}}}) \odot (\hat{\mathbf{R}} + \delta_{\hat{\mathbf{R}}})$ which introduced perturbation terms $\delta_{\hat{\mathbf{L}}}$ and $\delta_{\hat{\mathbf{R}}}$ for illumination and reflectance, respectively. However, these frameworks construct physical constraints in real space and rely on input features, so the constraints formed by learning generalize poorly to other datasets.

Kubelka–Munk-based model. The KM theory [10], which assumes that light within a material is isotropically scattered and characterizes the material layers by a wavelength-dependent scatter coefficient and absorption coefficient, is used for materials that both reflect and transmit light. As a general image formation model, it is adopted for studying the reflectance of light in real-world scenes, based on which Geusebroek *et al*. proposed a photometric reflectance model to formulate color invariance [9]. Recently, Wang *et al*. [43] devised illumination invariance, a physical quadruple prior, which originated from the KM theory, and can realize robust zero-reference LLIE. However, the KM model is mainly applied to calculate the invariances of the imaging process, while Retinex is a special case of this theory. The potential of the KM theory for LLIE to realize image decomposition or construct physical constraints has not yet been fully explored.

3 Method

Figure 2 displays the specific architecture of our method. As illustrated in Fig. 2(a), our TPCNet is composed of the LFE, RFE, DCGT, and CAM modules, where DCGT is designed as the core component of the enhancer, comprising the basis units, namely Cross-Guided Attention Blocks (CGABs). The latter, which include variants in skip connections, positional encoding, and feature crossover, such as, CGAB (V), CGAB (V-M), and CGAB (M), are composed of a normalization layer, a CG-MSA module, and an IEL module [51] similar to the Gated Forward Network. The CGAB(V) represents the variant that is different in positional encoding and skip connections, and CGAB (M) stands for the CG-MSA module with the pair-downsample removed.

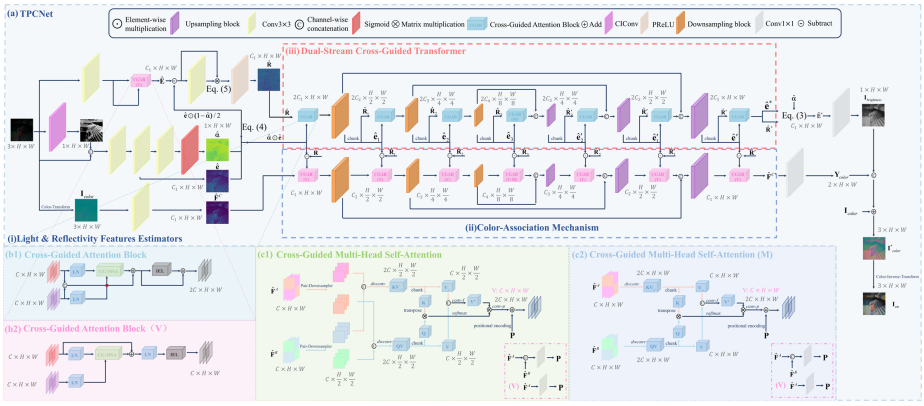

Fig. 2: (a) Overview of our proposed TPCNet, which contains a Light & Reflectivity Features Estimator (i), the Color-Association Mechanism (CAM) (ii), and the Dual-Stream Cross-Guided Transformer (iii). (b1) The Cross-Guided Attention Blocks (CGABs) are comprised of a normalization layer (LN), a Cross-Guided Multi-Head Self-Attention (CG-MSA) module, and an IEL module [51]. (b2) CGAB (V) with variation in skip connection. (c1)-(c2) Detailed structure of the CG-MSA and its variation.

3.1 Triple Physical Constraint Framework

To introduce our TPC concept, we start from the image formation model based on the Kubelka-Munk theory [10]

$$E(\lambda, \mathbf{x}) = e(\lambda, \mathbf{x})((1 - \rho_f(\mathbf{x}))^2 R_\infty(\lambda, \mathbf{x}) + \rho_f(\mathbf{x})), \tag{1}$$

where $E(\lambda, \mathbf{x})$ stands for the spectrum of light reflected from an object in the viewing direction, λ and $\mathbf{x}$ refer to the wavelength of the light and spatial location on the image plane, respectively, $e(\lambda, \mathbf{x})$ denotes the spectrum of the light source, $R_\infty(\lambda, \mathbf{x})$ is the material reflectivity, and $\rho_f(\mathbf{x})$ represents the specular reflection coefficient. Following the Retinex-based image formation approximation [48], we ignore the effect of the wavelength to simplify Eq. (1). But for $\rho_f(\mathbf{x})$, the Retinex theory [18,43] assumes the objects to be Lambertian and that $\rho_f(\mathbf{x}) \approx 0$, whereas we view $\rho_f(\mathbf{x})$ as a small quantity that cannot be ignored.

Then, by expanding $(1 - \rho_f(\mathbf{x}))^2$ as a Taylor series and omitting the higher-order term, we reformulate Eq. (1) as

$$E(\mathbf{x}) = (1 - 2\rho_f(\mathbf{x}))e(\mathbf{x})R_\infty(\mathbf{x}) + \rho_f(\mathbf{x})e(\mathbf{x}). \tag{2}$$

Subsequently, by setting $1 - 2\rho_f(\mathbf{x}) = \alpha(\mathbf{x})$, Eq. (2) can be written as

$$E(\mathbf{x}) = \underbrace{\alpha(\mathbf{x})e(\mathbf{x})}_{L(\mathbf{x})} R_\infty(\mathbf{x}) + \underbrace{(1 - \alpha(\mathbf{x}))e(\mathbf{x})}_{\bar{L}(\mathbf{x})}/2, \tag{3}$$

where $L(\mathbf{x}) = \alpha(\mathbf{x})e(\mathbf{x})$ can be regarded as an illumination map, while $\bar{L}(\mathbf{x}) = (1 - \alpha(\mathbf{x}))e(\mathbf{x})$ is its complementary illumination map, and both satisfy the constraint

$$L(\mathbf{x}) + \bar{L}(\mathbf{x}) = e(\mathbf{x}), \tag{4}$$

By transforming Eq. (3), we obtain another constraint as

$$R_{\infty}(\mathbf{x}) = \frac{E(\mathbf{x}) - \bar{L}(\mathbf{x})/2}{L(\mathbf{x})}, \tag{5}$$

Equations (3) to (5) constitute a TPC for detection of the reflected light, the illumination from the light source, and the material reflectivity characteristic, respectively. This TPC thus provides a heuristic guidance for network design.

3.2 Overview of the Triple Physical Constraint Network

To endow the network with TPC (TPCNet), the computational graph can be formulated as:

$$\begin{aligned} &\psi(\mathbf{I}) \rightarrow (\hat{\mathbf{e}}, \hat{\boldsymbol{\alpha}}) \xrightarrow{\text{Eq. (4)}} (\hat{\mathbf{L}}, \hat{\bar{\mathbf{L}}}) \rightarrow \zeta(\hat{\mathbf{L}}, \hat{\bar{\mathbf{L}}}) \rightarrow \hat{\mathbf{R}} \\ &\hat{\mathbf{R}} \rightarrow \mathcal{E}(\hat{\mathbf{R}}, \hat{\mathbf{L}}) \rightarrow (\hat{\mathbf{R}}^{*}, \hat{\mathbf{e}}^{*}) \xrightarrow[\text{add } \hat{\boldsymbol{\alpha}}]{\text{Eq. (3)}} \hat{\mathbf{E}}, \end{aligned} \tag{6}$$

where ψ refers to the LFE, ζ denotes the RFE and $\mathcal{E}$ stands for the enhancer. Taking low-light images $\mathbf{I} \in \mathcal{R}^{3\times H\times W}$ as input, ψ estimates the initial light features $\hat{\mathbf{e}} \in \mathcal{R}^{C\times H\times W}$ and corresponding weight $\hat{\boldsymbol{\alpha}} \in \mathcal{R}^{1\times H\times W}$ with value [0-1]. Subsequently, substituting the $\hat{\mathbf{e}}$ and $\hat{\boldsymbol{\alpha}}$ into Eq. (4) as the first constraint, we can calculate the complementary illumination map $\hat{\bar{\mathbf{L}}} \in \mathcal{R}^{C\times H\times W}$. Then $\mathbf{I}$ and $\hat{\bar{\mathbf{L}}}$ are fed into ζ with the Eq. (5) constraint to produce the reflectivity feature $\hat{R} \in \mathcal{R}^{C\times H\times W}$. Ultimately, the enhanced $\hat{\mathbf{R}}^{*} \in \mathcal{R}^{C\times H\times W}$ and $\hat{\mathbf{e}}^{*} \in \mathcal{R}^{C\times H\times W}$ are calculated by a four-scale U-shaped architecture $\mathcal{E}$ and we substitute $\hat{\mathbf{R}}^{*}$, $\hat{\mathbf{e}}^{*}$ and $\hat{\boldsymbol{\alpha}}$ into Eq. (3) to obtain the reflected-light feature $\hat{\mathbf{E}}^{*} \in \mathcal{R}^{C\times H\times W}$, thus realizing the third physical constraint on the network. It is worth noting that the variables in Eq. (6), such as $\hat{\mathbf{e}}$, $\hat{\mathbf{L}}$, $\hat{\bar{\mathbf{L}}}$, $\hat{\mathbf{R}}$, and $\hat{\mathbf{E}}$, are feature-space representations rather than direct physical quantities in the image domain. Therefore, TPC is used as a physically inspired principle to guide the network design, instead of assigning strict physical meanings to each individual channel.

However, since the effect of wavelength has been ignored in our model, directly mapping $\hat{\mathbf{E}}^{*}$ to $\mathbf{I}_{en} \in \mathcal{R}^{3\times H\times W}$ may result in color deviation or overexposure. To overcome this limitation, we introduce a multi-scale color-association mechanism. First, we convert the RGB color space of $\mathbf{I}$ into another color space $\mathbf{I}_{color} \in \mathcal{R}^{3\times H\times W}$ (*e.g.*, YCbCr [14], HSV [36], HVI [51], *etc.*) to realize the separation of brightness and color information. Subsequently, the color features $\hat{\mathbf{F}}^{C} \in \mathcal{R}^{C\times H\times W}$ are extracted from $\mathbf{I}_{color}$, and are then fed into the DCGTs to obtain the enhanced color feature $\hat{\mathbf{F}}^{C^{*}} \in \mathcal{R}^{C\times H\times W}$ with the multi-scale $\hat{\mathbf{R}}^{*} \odot \hat{\mathbf{e}}^{*}$ information as feature guidance. Next, the $\mathbf{F}^{C^{*}}$ and $\hat{\mathbf{E}}^{*}$ are mapped to the color information $\mathbf{Y}_{color} \in \mathcal{R}^{2\times H\times W}$ and brightness $\mathbf{I}_{brightness} \in \mathcal{R}^{1\times H\times W}$, respectively, and we concatenate $\mathbf{Y}_{color}$ and $\mathbf{I}_{brightness}$ for each pixel along the channel dimension to acquire the $\mathbf{I}^{*}{}_{color} \in \mathcal{R}^{3\times H\times W}$ and introduce the residual connection $\mathbf{I}^{*}{}_{color} = \mathbf{I}^{*}{}_{color} + \mathbf{I}_{color}$. Finally, the $\mathbf{I}^{*}{}_{color}$ is transformed back into the $\mathbf{I}_{\text{en}}$.

The architecture of ψ is shown in Fig. 2, where ψ first extracts the invariance properties W of the illumination intensity from $\mathbf{I}$ by a CIConv [20], and a $conv3\times3$ (convolution with kernel size = 3) is used to fuse the concatenation of $\mathbf{I}$ and W. Since the fused feature possesses rich semantic contextual information, we utilize a $conv3\times3$ operation to generate $\hat{\mathbf{e}}$, and the corresponding weight $\hat{\boldsymbol{\alpha}}$ is calculated by another $conv3\times3$ followed by Sigmoid activation.

The composition of ζ is exhibited in Fig. 2(a). First, ζ employs a $conv3\times3$ followed by the CGAB to estimate $\hat{\mathbf{E}}$ from $\mathbf{I}$ and calculates $\hat{\mathbf{E}}-\frac{\hat{\mathbf{L}}}{2}$. Subsequently, $\hat{\mathbf{E}}-\frac{\hat{\mathbf{L}}}{2}$ is fed into another $conv3\times3$ to produce $\hat{\mathbf{L}}'$ where $\hat{\mathbf{L}}'\odot\hat{\mathbf{L}}=1$. To enhance the physical constraints, we substitute $\hat{\mathbf{E}}-\frac{\hat{\mathbf{L}}}{2}$ and $\hat{\mathbf{L}}'$ into Eq. (5) to compute $\hat{\mathbf{R}}$.

Discussion. (i) Compared with the previous deep learning methods based on the Retinex theory [7,11,48,54,58] that view the $\hat{\mathbf{R}}$ and $\hat{\mathbf{L}}$ as independent estimations and employ deep convolutional neural networks to estimate them separately or to estimate $\hat{\mathbf{L}}$ only, the TPC framework regards the $\hat{\mathbf{R}}$ and $\hat{\mathbf{L}}$ as the estimations with mathematical connection, and then the $\hat{\mathbf{R}}$ can be calculated by $(\hat{\mathbf{E}}-\frac{\hat{\mathbf{L}}}{2})\odot\hat{\mathbf{L}}'$. This framework utilizes the inherent relationships between $\hat{\mathbf{R}}$ and $\hat{\mathbf{L}}$ for self-constraint without requiring complicated loss functions, which can be effective in decreasing training difficulty. Moreover, the calculation process $\hat{R}=(\hat{\mathbf{E}}-\frac{\hat{\mathbf{L}}}{2})\odot\hat{\mathbf{L}}'$ possesses some advantages. From one perspective, $(\hat{\mathbf{E}}-\frac{\hat{\mathbf{L}}}{2})$ can decrease the estimation bias to a certain extent by computing the difference between $\hat{\mathbf{E}}$ and $\frac{\hat{\mathbf{L}}}{2}$). From another perspective, $\hat{R}=(\hat{\mathbf{E}}-\frac{\hat{\mathbf{L}}}{2})\odot\hat{\mathbf{L}}'$ avoids directly dividing $(\hat{\mathbf{E}}-\frac{\hat{\mathbf{L}}}{2})$ by $\hat{\mathbf{L}}$, which can enlarge the deviation [3] and can produce a more robust $\hat{\mathbf{R}}$.

(ii) Inspired by the Retinex theory [18], we employ $\hat{\mathbf{R}}\odot\hat{\mathbf{e}}$, a rough light-up feature, as an initial guidance to produce the corresponding color features in our CAM. However, multiplying two estimated values directly might intensify the deviation due to the existence of estimation bias, and the expanded deviation will further contaminate the subsequent output by forward propagation, affecting model performance. To address these limitations, we adopt multi-scale light-up features and recalculate each scale-guided feature by employing enhanced light-up instead of the skip connections from the initial $\hat{\mathbf{R}}\odot\hat{\mathbf{e}}$ to the next scale, thus diminishing the deviation accumulation.

3.3 Dual-Stream Cross-Guided Transformer

Recently, due to their superiority in capturing long-range dependencies, transformers have been increasingly exploited in image enhancement or restoration. Their computing mechanism can be flexibly used to calculate the correlation between features and to realize cross guidance. Some transformer-based guided mechanisms, like Retinexformer [3], realize feature guidance by value elements $\mathbf{V}$ multiplied with $\mathbf{F}_{lu}$; others, like CIDNet [51], calculate $\mathbf{KV}$ elements and $\mathbf{Q}$ from two different features using separate learnable layers, and then realize cross guidance based on a self-attention mechanism (SAM). These guidance

mechanisms have inherent limitations. Facing different features, the separate learnable layers with distinct learning abilities can cause attention bias, and the output can be affected by the portion that possesses stronger learning capabilities. To address this limitation, we design a DCGT as the enhancer $\mathcal{E}$.

Network Structure. As illustrated in Fig. 2(a), DCGT exploits a four-scale U-Structure. Initial $\hat{\mathbf{L}}$, $\hat{\mathbf{R}}$ and $\hat{\mathbf{F}}^C$ are input into DCGT. During downsampling, $\hat{\mathbf{L}}$ and $\hat{\mathbf{R}}$ are processed by a CGAB and a downsampling block which consists of $conv3\times3$ (for channel conversion), bilinear interpolation, and PReLU activation, to generate $\hat{\mathbf{F}}_i^{RL} \in \mathcal{R}^{2C_i\times\frac{H}{2^i}\times\frac{W}{2^i}}$ where i = 1, 2, 3. Then $\hat{\mathbf{F}}_i^{RL}$ is partitioned into $\hat{\mathbf{R}}_i$ and $\hat{\mathbf{e}}_i$ along the channel dimension; $\hat{\mathbf{R}}_i$ and $\hat{\mathbf{e}}_i$ as the new inputs are fed into the CGAB. For the downsampling color-association portion, $\hat{\mathbf{F}}^C$ and $\hat{\mathbf{R}}\odot\hat{\mathbf{L}}$ undergo a CGAB and the downsampling block to produce the $\hat{\mathbf{F}}_i^C \in \mathcal{R}^{C_i\times\frac{H}{2^i}\times\frac{W}{2^i}}$. As the guidance for $\hat{\mathbf{F}}_i^C$, $\hat{\mathbf{R}}_i\odot\hat{\mathbf{e}}_i$ is applied in the next CGAB. Then $\hat{\mathbf{R}}_3$ and $\hat{\mathbf{e}}_3$ pass through CGAB (M) while $\hat{\mathbf{F}}_3^C$ and $\hat{\mathbf{R}}_3\odot\hat{\mathbf{e}}_3$ are fed into CGAB (V-M), respectively. Subsequently, a symmetrical structure is designed for upsampling. The upsampling block is utilized to fuse skip connections and to upscale the features, which is composed of $conv3\times3$ (for channel conversion), $conv1\times1$ (for connection fusion) bilinear interpolation, and PReLU activation. The color space of TPCNet is Ycbcr, and the results of ablation experiments shown in Table 3 verify the effect of different color spaces.

CG-MSA. As displayed in Fig. 2(c1-c2), $\hat{\mathbf{F}}^A \in \mathcal{R}^{C\times H\times W}$ and $\hat{\mathbf{F}}^B \in \mathcal{R}^{C\times H\times W}$ are fed into the CG-MSA which is a core component for all variants of CGAB. To calculate the cross-correlation features between $\hat{\mathbf{F}}^A$ and $\hat{\mathbf{F}}^B$, we first utilize the pair downsampler [30] to divide $\hat{\mathbf{F}}$ into non-overlapping 2×2 patches, and obtain $\hat{\mathbf{F}}_i^j \in \mathcal{R}^{C\times\frac{H}{2}\times\frac{W}{2}}$, where i=1,2; j=A, B. Subsequently, we concatenate $\hat{\mathbf{F}}_1^A, \hat{\mathbf{F}}_2^B$ and $\hat{\mathbf{F}}_2^A, \hat{\mathbf{F}}_1^B$ along the channel dimension to generate combined features $\hat{\mathbf{F}}_i^M \in \mathcal{R}^{2C\times\frac{H}{2}\times\frac{W}{2}}$ where i=1, 2. Then, utilizing the SAM, the combined features are further fused to produce the cross-guided attention map. In our designed MSA, the query ($\mathbf{Q}$), key ($\mathbf{K}$) and value ($\mathbf{V}$) projections are calculated by a projective $conv1\times1$ followed by a 3×3 depth-wise convolution ($dwconv3\times3$). The $conv1\times1$ is used to polymerize pixel-wise cross-channel information and then we utilize $dwconv3\times3$ to enhance the polymerized information. To avoid attention bias caused by different learning capability, the formulas for calculating the projections are as follows:

$$\mathbf{QV}^* = W_d^{KV^*} W_p^{KV^*} \hat{\mathbf{F}}_1^M, KV = W_d^{KV} W_p^{KV} \hat{\mathbf{F}}_2^M, \tag{7}$$

where $W_p^{(\cdot)}$ and $W_d^{(\cdot)}$ represent the learnable parameters of the $conv1\times1$ and $dwconv3\times3$, respectively. Subsequently, splitting $\mathbf{QV}^*$ and $\mathbf{KV}$ into $\mathbf{Q}, \mathbf{K}, \mathbf{V}, \mathbf{V}^* \in \mathcal{R}^{C\times\frac{H}{2}\times\frac{W}{2}}$, we combine the $\mathbf{V}^*$ and $\mathbf{V}$ to produce $\mathbf{V}^{'}$ by using a $conv3\times3$ and divide the number of channels into each "head". Then, the cross-guided attention map for each head can be formulated as

$$\text{Attention}(\mathbf{Q}_j, \mathbf{K}_j, \mathbf{V}_j^{'}) = \text{Softmax}(\frac{\mathbf{Q}_j\mathbf{K}_j^T}{\alpha_H}) \otimes \mathbf{V}_j^{'}, \tag{8}$$

where $\mathbf{Q}_j, \mathbf{K}_j, \mathbf{V}_j^{'} \in \mathcal{R}^{h_k \times \frac{HW}{4}}$, $h_k = \frac{C}{k}$, $j = 1, 2, ..., k$; k is the head number and α_H is a learnable scaling parameter that adaptively scales the magnitude of the dot product of $\mathbf{Q}_j$ and $\mathbf{K}_j^T$ [6, 26]. The k-head features are concatenated along the channel dimension and reshaped to $\hat{\mathbf{F}}_{out}^{M} \in \mathcal{R}^{2C \times \frac{H}{2} \times \frac{W}{2}}$; $\hat{\mathbf{F}}_{out}^{M}$ undergoes the pixel-shuffle $conv1\times1$ to upscale its resolution, and finally adds a positional encoding calculated by a $conv1\times1$ that concatenates $\hat{\mathbf{F}}^{A}$ and $\hat{\mathbf{F}}^{B}$ as input, to generate the cross-guided features $\hat{\mathbf{F}}_{out}^{AB} \in \mathcal{R}^{2C \times H \times W}$.

To simplify the analysis of computational complexity, we focus on computation of our SAM in Eq. (8), which involves two matrix multiplications $\mathcal{R}^{h_k \times \frac{HW}{4}} \times \mathcal{R}^{\frac{HW}{4} \times h_k}$ and $\mathcal{R}^{h_k \times h_k} \times \mathcal{R}^{h_k \times \frac{HW}{4}}$. The formulas of the CG-MSA computational complexity are thus as follows,

$$\begin{aligned} \mathcal{O}(\mathbf{CG{-}MSA}) &= k{\cdot}(\frac{HW}{4}{\cdot}h_k{\cdot}h_k) + k{\cdot}(h_k{\cdot}h_k{\cdot}\frac{HW}{4}) \\ &= k\frac{HW}{2}h_k^2 = k\frac{HW}{2}(\frac{C}{k})^2 = \frac{HWC^2}{2k}. \end{aligned} \tag{9}$$

Compared with the previous CNN-Transformer methods [3, 49], the computational complexity of CG-MSA is linear in the spatial size, and only 25% of the conventional MSA computation is required for the same input size. Therefore, CG-MSA can be effective in improving our TPCNet's inference speed, and can serve as a plug-and-play lightweight design module.

4 Experiment

4.1 Datasets and Settings Details

We evaluated our method on commonly-used LLIE benchmark datasets including LOL v2 [53], DICM [19], LIME [11], MEF [28], NPE [41], and VV [38]. We also conducted further experiments on a real-world LLIE dataset VILNC [34], an extreme dataset SID (Sony-Total-Dark) [5], and an LCDP [40] dataset containing over/under exposure images.

LOL-v2. LOL-v2 is divided into two subsets, LOL-v2-real and LOL-v2-Synthetic. The training and testing parts of LOL-v2-Real and LOL-v2-Synthetic are split proportionally into 689:100 and 900:100, respectively.

VILNC. VILNC is a new LLIE dataset [34], which comprises 115 indoor scenes and 20 outdoor scenes. Each indoor scene possesses three different brightness levels to simulate varying degrees of low-light images, with a normal brightness image used as a reference, while each outdoor scene only contains a pair of low/normal brightness images. We randomly extract 20% of the images from the indoor scene as the testing set, and the rest as the training set. To compare the performance of various methods on the VILNC dataset fairly, we retrain all methods in VILNC-Indoor by using their open-source code with default training parameters. Except for the method QuadPrior [43], which is trained with large-scale datasets, we directly use its pre-training weight for inference.

Table 1: Quantitative comparison with LOL v2 [53] and VILNC-indoor [34] datasets using different evaluation methods (PSNR↑, SSIM↑, and LPIPS↓). The best and second-best performances are highlighted in red and cyan, respectively. The FLOPs is calculated on a single 256×256 image. Runtime is measured as the average inference time for a single 512×512 image with batch size 1.

Methods	Complexity			LOL-v2-Real			LOL-v2-Synthetic			VILNC-Indoor		
	Params/M	FLOPs/G	Runtime/ms	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
KinD [57]	8.02	34.99	19.3	17.544	0.669	0.375	18.320	0.796	0.252	9.453	0.583	0.411
RUAS [24]	0.003	0.83	4.1	15.506	0.491	0.314	13.880	0.676	0.291	8.836	0.570	0.735
RetinexNet [48]	0.84	584.47	26.6	15.011	0.397	0.501	19.750	0.824	0.192	16.573	0.456	0.547
ZeroDCE++ [21]	0.01	0.61	3.9	12.338	0.443	0.488	15.871	0.804	0.217	15.992	0.434	0.579
SNR-Net [49]	4.01	26.35	15.6	21.480	0.849	0.163	24.140	0.928	0.056	23.367	0.765	0.202
LLFormer [42]	24.55	22.52	260.3	20.056	0.792	0.211	24.038	0.909	0.066	23.593	0.769	0.313
Retinexformer [3]	1.61	15.57	61.4	22.794	0.840	0.171	25.670	0.930	0.059	23.735	0.782	0.295
PairLIE [8]	0.33	20.81	14.5	19.885	0.778	0.317	19.074	0.794	0.230	9.255	0.570	0.681
HVI-CIDNet [51]	1.88	7.57	59.4	24.111	0.868	0.116	25.705	0.942	0.045	23.656	0.785	0.303
QuadPrior [43]	1252.75	1103.2	82774.0	20.592	0.811	0.202	16.108	0.758	0.114	11.293	0.552	0.487
LANet [52]	0.58	43.23	51.4	18.074	0.735	0.364	18.088	0.807	0.183	18.410	0.734	0.484
ZERO-IG [34]	0.087	11.37	8.0	16.943	0.738	0.395	17.610	0.743	0.410	12.002	0.595	0.504
SMG [50]	17.8	42.93	258.8	24.032	0.818	0.169	24.979	0.891	0.092	23.440	0.688	0.212
RQ-LLIE [25]	13.48	216.78	703.3	22.371	0.854	0.142	25.937	0.941	0.044	23.540	0.791	0.288
TPCNet(Ours)	2.62	8.76	35.1	24.978	0.882	0.102	26.032	0.943	0.041	24.634	0.785	0.194

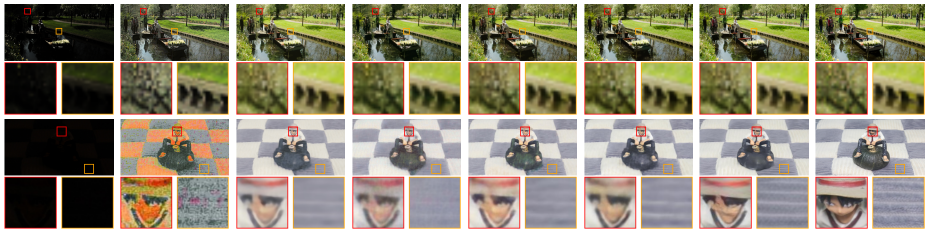

Fig. 3: Comparison of the enhanced images with various SOTA methods on LOL-v2-Synthetic (top row) and VILNC-Indoor (bottom row).

SID. The official SID dataset [5] contains 5094 raw short-exposure images, each with a corresponding long-exposure reference image. We adopt its customized version subset (Sony-Total-Dark) [51], which transfers the raw format images to sRGB images without gamma correction.

LCDP. The LCDP [40] contains images with both over-exposed and under-exposed regions, comprising 1,733 images, which are split into 1,415 for training, 100 for validation, and 218 for testing.

Implementation Details. Our model TPCNet is implemented by PyTorch and is trained with the Adam optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) for 1500 epochs by using a single NVIDIA 3090 or 4090 GPU. The initial learning rate is set to 2.5×10^{-4} and then steadily decreased to 1×10^{-7} by the cosine annealing scheme [27] during the training process. Patches of size 320×320 are randomly cropped from the low-/normal-light image pairs and the batch size is 8. The training data is augmented with flipping and random rotation. Runtime is measured using the official open-source code on 512×512 inputs with batch size 1 on a NVIDIA 3090 GPU; diffusion-based methods follow their default sampling procedures.

Inspired by reference [51], we utilize the loss function (L1loss, perceptual loss [13], edge loss [33], and SSIM loss [47]) to supervise the enhanced result simultaneously in RGB space and specific color space.

Evaluation Metrics. For datasets with reference images, we exploit the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [47] as the pixel-level evaluation and learned perceptual image patch similarity (LPIPS) [56] with AlexNet [17] as the perceptual quality evaluation. For the no-reference datasets, we evaluate the generalization of models trained on LOL-v2-Synthetic using various evaluation methods, such as natural image quality evaluator (NIQE) [31], multi-scale image quality transformer (MUSIQ) [16], and perceptual index (PI) [2]. We exploit PyTorch Toolbox [4] for calculating no-reference metrics.

4.2 Main Results

Quantitative Results. The quantitative results between our TPCNet and a wide range of SOTA enhancement algorithms are shown in Tables 1 and 2. Our TPCNet outperforms SOTA methods on LOL-v2 and VILNC-Indoor datasets in terms of almost all PSNR, SSIM, and LPIPS metrics, and possesses advantages in various metrics, such as NIQE, MUSIQ, and PI on no-reference datasets (MEF [28], NPE [41], LIME [11], DICM [19], and VV [38]), while requiring less computation.

For the reference datasets shown in Table 1, we compared our results with the SOTA lightweight method HVI-CIDNet. Our TPCNet achieves 0.867 dB, 0.327 dB, and 0.978 dB improvements for approximately the same computational cost. Compared to RetinexFormer, a SOTA Retinex-based method, TPCNet realizes higher quality image enhancement and further expands its parameters 1.6 times (2.62/1.61), but only costs 56% (8.76/15.57) FLOPs, demonstrating that our CG-MSA is efficiently designed. Especially, in the real-world dataset VILNC-Indoor, TPCNet achieves over 15 dB improvement compared to the previous Retinex-based methods.

To verify TPCNet generalization on the no-reference datasets shown in Table 2, we compare our model with a recent SOTA unsupervised method. Compared to QuadPrior, which is trained on the large-scale datasets, TPCNet exhibits an obvious improvement in the NIQE and PI metric but only requires 0.8% (8.76/1103.2) of its FLOPs. Compared with other approaches, TPCNet shows the advantages in NIQE, MUSIQ, and PI metrics without a large computational consumption, achieving a balance between performance and efficiency. These results demonstrate that the TPC constructed on an implicit feature space can improve the generalization and robustness of our model.

Qualitative Results. The visual results for comparing TPCNet to other SOTA algorithms are shown in Figs. 3 and 4. From the Fig. 3 visualization, we can observe that TPCNet achieves stable brightness enhancement and maintains color consistency while other algorithms exhibit a certain flaw, especially in real-world datasets (VILNC-Indoor). For example, the earliest Retinex-based DL,

Table 2: Quantitative results for the LCDP [40] , Sony-Total-Dark [5], MEF [28], NPE [41], LIME [11], DICM [19], and VV [38] datasets. The top-ranking score is in **bold**, and the second-ranking is underlined; (·) represents the rank order.

Methods	Sony-Total-Dark PSNR↑	SSIM↑	LCDP PSNR↑	SSIM↑	Methods	Unpaired NIQE↓	MUSIQ↑	PI↓	FLOPs(G)	Overall rank
RUAS [24]	10.456	0.086	13.981	0.633	RUAS [24]	5.248(6)	51.275(5)	3.955(6)	0.83(2)	5
PairLIE [8]	15.340	0.532	14.748	0.686	ZERO-IG [34]	3.987(4)	47.618(6)	3.542(5)	11.37(4)	5
RetinexNet [48]	18.230	0.596	19.626	0.689	LANet [52]	**3.502**(1)	56.218(3)	**2.804**(1)	43.23(5)	2
ZeroDCE++ [21]	12.683	0.082	12.223	0.668	ZeroDCE++ [21]	3.959(3)	52.172(4)	3.231(3)	0.61(1)	3
HVI-CIDNet [51]	22.904	0.676	22.840	0.846	QuadPrior [43]	4.666(5)	**62.241**(1)	3.455(4)	1103.2(6)	4
TPCNet(Ours)	**23.577**	**0.691**	**23.398**	**0.874**	**TPCNet(Ours)**	3.625(2)	58.070(2)	2.982(2)	8.76(3)	**1**

Fig. 4: Visual comparison on the MEF [28], NPE [41], LIME [11], DICM [19], and VV [38] datasets.

RetinexNet, produces serious color deviation during enhancement when facing extreme darkness, while RetinexFormer recovers the correct color, but still with a little color deviation. Compared to a recent SOTA algorithm, HVI-CIDNet can recover accurate color due to its HVI color space advantage, but some details are blurred.

In addition, Fig. 4 presents representative visual comparisons between TPCNet and competing supervised and unsupervised LLIE methods on five unpaired datasets without ground truth. These examples are selected to illustrate typical enhancement behaviors across different scenes and degradation patterns, while comprehensive quantitative results for all evaluated methods are reported in Table 2. As observed from Fig. 4, existing methods may suffer from over-/under-exposure, color deviation, or noise amplification during enhancement. In contrast, TPCNet more effectively enhances low-light regions while preserving natural colors and suppressing obvious artifacts. Together with the quantitative results, these qualitative examples further demonstrate the superior performance of TPCNet over previous SOTA methods in the lightweight LLIE field, and the effectiveness of the implicit-space TPC constraints in improving model generalization. Additional visual results are available in the `Results` directory of our released repository.

4.3 Ablation Study

We perform all ablation experiments on LOL-v2-Real for good convergence and stable performance of TPCNet. To verify the key modules, physical constraint, and the CAM with different color spaces in TPCNet, we use quantitative comparisons. The results are reported in Table 3.

Table 3: Ablation on the LOLv2-real dataset. PSNR, SSIM, LPIPS, Params, and FLOPS (size = 256×256) are reported. The best results are marked in **bold** and the second best are underlined.

Baseline-1	Baseline-2	CAM	LFE	RFE	TPC	PSNR↑	SSIM↑	LPIPS↓	Params (M)	FLOPs (G)
✓						21.637	0.854	0.119	1.65	5.72
	✓					22.279	0.849	0.128	0.71	5.72
✓		✓				23.729	0.871	0.111	2.61	8.32
✓		✓	✓			24.232	0.876	0.111	2.61	8.46
✓		✓		✓		24.263	0.874	0.109	2.62	8.55
✓		✓	✓	✓		24.357	0.881	**0.099**	2.62	8.68
✓		✓	✓	✓	✓	**24.641**	**0.881**	0.100	2.62	8.76

(a) Decomposed ablation of each component.

Metrics		PSNR↑	SSIM↑	LPIPS↓	Params (M)	FLOPs (G)
Physical Constraint	w/o Eqs. (3-5)	24.357	0.881	**0.099**	2.62	8.68
	w/o Eq. (3)	24.598	0.881	0.101	2.62	8.76
	w/o Eqs. (4-5)	24.575	0.876	0.110	2.62	8.68
	Enforced Eq. (5)	24.188	0.878	0.105	2.62	8.68
	$\hat{\boldsymbol{\alpha}} = 1$	24.523	0.877	0.106	2.62	8.76
Color Space	HVI	24.641	0.881	0.100	2.62	8.76
	LAB	24.588	0.873	0.114	2.62	8.76
	YCBCR	**24.978**	**0.882**	0.102	2.62	8.76

(b) Ablation studies on physical constraints and color space.

Decomposed ablation. We utilize a decomposed ablation to analyze the effect of each component on improving performance, as displayed in Table 3a. To obtain Baseline-1, which only comprises the DCGT framework, we remove the CAM from TPCNet and replace its LFE and RFE with $conv3\times3$. Baseline-2 is constructed as an attention-operator ablation by replacing CG-MSA with standard MSA while maintaining FLOPs comparable to Baseline-1. Meanwhile, Eq. (3) is replaced by a $conv3\times3$ mapping the output from the last CGAB to the enhanced result, and Eqs. (4) and (5) are removed in TPCNet, to avoid the impact of physical constraint. When we exploit the CAM with HVI [51] as the color space, the mapping $conv3\times3$ is utilized to produce enhanced brightness instead of the final enhanced result. Baseline-1 realizes an obvious improvement, gaining 2.102 dB after applying CAM. Subsequently, we employ the LFE and RFE, respectively, and Baseline-1 achieves 2.605 and 2.636 dB improvements, respectively. When jointly applying these two estimators, the Baseline-1 achieves an increase of 2.730 dB. Subsequently, the TPC is introduced to obtain our final TPCNet version, and Baseline-1 realizes a 3.004 dB improvement. These decomposed experiments prove the effectiveness of CAM, LFE, RFE, and physical constraint in our TPCNet.

Physical constraint. To analyze the effect of TPC, we administer an ablation experiment to study each physical constraint. The quantitative results are exhibited in Table 3b We first remove TPC from TPCNet, and initially obtain an PSNR of 24.357 dB. Then, we separately apply Eqs. (4-5) and Eq. (3) to form constraints on the inner network, and the model achieves a gain of 0.241 dB and 0.218 dB, respectively. Finally, we construct the complete TPC for further improvement. These results validate that exploiting physical constraints within the inner network can optimize performance without requiring additional parameters, simultaneously providing an interpretable method for network design. In addition, we conducted experiments to explicitly enforce Eq (5) by directly dividing $(\hat{\mathbf{E}} - \frac{\hat{\mathbf{L}}}{2})$ by $\hat{\mathbf{L}}$. However, introducing division operations into the model leads to numerical instability, resulting in degraded PSNR and SSIM performance.

Learnable $\hat{\boldsymbol{\alpha}}$ Analysis. $\hat{\boldsymbol{\alpha}}$ serves as an intermediate variable within the TPC framework. Given our approximation that $1 - 2\rho_f = \hat{\boldsymbol{\alpha}}$, when $\hat{\boldsymbol{\alpha}} = 1$, it follows that $\rho_f = 0$. Under this condition, the TPC framework degenerates into a Retinex-based formulation, and the triple physical constraints become equivalent to a single physical constraint. As shown in Table 3(b), when $\hat{\boldsymbol{\alpha}} = 1$, the model performance degrades compared to the full TPC model, indicating that

introducing ρ_f (i.e., $\hat{\boldsymbol{\alpha}} \neq 1$) contributes to performance improvement. We further estimate our approximation error induced by the omitted higher-order term on LOL-v2-Real as error $= \hat{\rho}_f^2 = \left(\frac{1-\hat{\alpha}}{2}\right)^2$. The mean and P95 errors are 7.57% and 7.92%, respectively, where P95 denotes the threshold below which 95% of pixel-wise errors fall, indicating that the approximation adopted in the TPC framework is empirically justified.

Table 4: Cross-dataset evaluation on LOL-v2 datasets and LCDP. The leading dataset denotes the source dataset used for training and conventional evaluation, while the dataset in parentheses denotes the target dataset for direct testing without fine-tuning. Conventional and cross-dataset results are marked in black and cyan, respectively.

Methods	LOL-v2-Real (LCDP)			LOL-v2-Real (LOL-v2-Synthetic)			LOL-v2-Synthetic (LOL-v2-Real)		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
RetinexNet [48]	15.011 (15.003)	0.397 (0.635)	0.397 (0.256)	15.011 (16.349)	0.397 (0.759)	0.397 (0.236)	19.750 (14.951)	0.824 (0.458)	0.192 (0.432)
Retinexformer [3]	22.794 (11.985)	0.840 (0.531)	0.171 (0.374)	22.794 (13.998)	0.840 (0.638)	0.171 (0.332)	25.670 (13.124)	0.930 (0.489)	0.059 (0.369)
TPCNet(Ours)	24.978 (15.198)	0.882 (0.726)	0.102 (0.231)	24.978 (17.058)	0.882 (0.801)	0.102 (0.208)	26.032 (15.213)	0.943 (0.530)	0.041 (0.335)

Color spaces. We conduct ablation experiments to study the robustness of CAM with different color spaces and further explore its impact on model performance. Based on our image formation model, we chose a color conversion that can transform RGB images into a color space where brightness and color information can be separated, such as HVI [51], YCbCr [14], and LAB [37]. The corresponding quantitative results are presented in Table. 3b. Our results demonstrate the great robustness of CAM in various color spaces, and can achieve superior performance (the PSNR is 24.641 / 24.588 / 24.978 with Color space HVI / LAB / Ycbcr), better than the SOTA methods [3, 49, 51] in recent years.

4.4 Cross-dataset Study

To investigate the impact of the space in which physical constraints are imposed on the model's generalization performance, we conducted a cross-dataset experiment, where the model is trained on dataset A without fine-tuning and directly evaluated on dataset B. In contrast to representative methods such as RetinexFormer and RetinexNet, which impose physical constraints in the real space, our method constructs the TPC in the feature space. As shown in Table 4, our method achieves superior performance and exhibits stronger generalization capability.

5 Conclusion

In this work, we have reformulated the influence of physical constraints in the imaging process based on the Kubelka-Munk theory, and have constructed the physical constraint among illumination, reflection, and detection. Our TPC theory addresses the limitations of previous Retinex-based algorithms, which view the reflected objects as ideal Lambertian and ignore the specular reflection coefficient in the modeling process. Extensive quantitative and qualitative experiments

verify the effectiveness of our model in improving the robustness and LLIE performance by constructing physical constraints in feature space, and demonstrate that TPCNet outperforms SOTA LLIE methods on 10 datasets, thus expanding the interpretable network-design direction for low-light enhancement.

Acknowledgements

We gratefully acknowledge the anonymous reviewers and area chair for their valuable comments and suggestions.

References

1. Banik, P.P., Saha, R., Kim, K.D.: Contrast enhancement of low-light image using histogram equalization and illumination adjustment. In: 2018 International Conference on Electronics, Information, and Communication (ICEIC). pp. 1–4 (2018). https://doi.org/10.23919/ELINFOCOM.2018.8330564
2. Blau, Y., Michaeli, T.: The perception-distortion tradeoff. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6228–6237 (2018)
3. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 12504–12513 (October 2023)
4. Chen, C., Mo, J.: IQA-PyTorch: Pytorch toolbox for image quality assessment. [Online]. Available: https://github.com/chaofengc/IQA-PyTorch (2022)
5. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3291–3300 (2018). https://doi.org/10.1109/CVPR.2018.00347
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale (2020)
7. Fu, H., Zheng, W., Meng, X., Wang, X., Wang, C., Ma, H.: You do not need additional priors or regularizers in retinex-based low-light image enhancement. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 18125–18134 (2023). https://doi.org/10.1109/CVPR52729.2023.01738
8. Fu, Z., Yang, Y., Tu, X., Huang, Y., Ding, X., Ma, K.K.: Learning a simple low-light image enhancer from paired low-light instances. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 22252–22261 (2023). https://doi.org/10.1109/CVPR52729.2023.02131
9. Geusebroek, J.M., van den Boomgaard, R., Smeulders, A., Geerts, H.: Color invariance. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(12), 1338–1350 (2001). https://doi.org/10.1109/34.977559
10. Geusebroek, J.M.: Color and geometrical structure in images. Appl. Microsc **3** (2000)
11. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing **26**(2), 982–993 (2017). https://doi.org/10.1109/TIP.2016.2639450

12. Jobson, D., Rahman, Z., Woodell, G.: Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing **6**(3), 451–462 (1997). https://doi.org/10.1109/83.557356
13. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. pp. 694–711. Springer (2016)
14. Kaur, A., Kranthi, B.: Comparison between ycbcr color space and cielab color space for skin color segmentation. International Journal of Applied Information Systems **3**(4), 30–33 (2012)
15. Kaur, M., Kaur, J., Kaur, J.: Survey of contrast enhancement techniques based on histogram equalization. International Journal of Advanced Computer Science and Applications **2**(7) (2011). https://doi.org/10.14569/IJACSA.2011.020721, http://dx.doi.org/10.14569/IJACSA.2011.020721
16. Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: Musiq: Multi-scale image quality transformer. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5128–5137 (2021). https://doi.org/10.1109/ICCV48922.2021.00510
17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (May 2017). https://doi.org/10.1145/3065386, https://doi.org/10.1145/3065386
18. Land, E.H.: The retinex theory of color vision. Scientific American **237**(6), 108–129 (1977)
19. Lee, C., Lee, C., Kim, C.S.: Contrast enhancement based on layered difference representation of 2d histograms. IEEE Transactions on Image Processing **22**(12), 5372–5384 (2013). https://doi.org/10.1109/TIP.2013.2284059
20. Lengyel, A., Garg, S., Milford, M., van Gemert, J.C.: Zero-shot day-night domain adaptation with a physics prior. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4399–4409 (October 2021)
21. Li, C., Guo, C., Loy, C.C.: Learning to enhance low-light image via zero-reference deep curve estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(8), 4225–4238 (2022). https://doi.org/10.1109/TPAMI.2021.3063604
22. Li, J., Li, B., Tu, Z., Liu, X., Guo, Q., Juefei-Xu, F., Xu, R., Yu, H.: Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15205–15215 (2024)
23. Li, M., Liu, J., Yang, W., Sun, X., Guo, Z.: Structure-revealing low-light image enhancement via robust retinex model. IEEE Transactions on Image Processing **27**(6), 2828–2841 (2018). https://doi.org/10.1109/TIP.2018.2810539
24. Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10556–10565 (2021). https://doi.org/10.1109/CVPR46437.2021.01042
25. Liu, Y., Huang, T., Dong, W., Wu, F., Li, X., Shi, G.: Low-light image enhancement with multi-stage residue quantization and brightness-aware attention. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 12106–12115 (2023). https://doi.org/10.1109/ICCV51070.2023.01115
26. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9992–10002 (2021). https://doi.org/10.1109/ICCV48922.2021.00986
27. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)

28. Ma, K., Zeng, K., Wang, Z.: Perceptual quality assessment for multi-exposure image fusion. IEEE Transactions on Image Processing **24**(11), 3345–3356 (2015). https://doi.org/10.1109/TIP.2015.2442920
29. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward fast, flexible, and robust low-light image enhancement. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5627–5636 (2022). https://doi.org/10.1109/CVPR52688.2022.00555
30. Mansour, Y., Heckel, R.: Zero-shot noise2noise: Efficient image denoising without any data. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14018–14027 (2023). https://doi.org/10.1109/CVPR52729.2023.01347
31. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal Processing Letters **20**(3), 209–212 (2013). https://doi.org/10.1109/LSP.2012.2227726
32. Rahman, Z., Jobson, D., Woodell, G.: Multi-scale retinex for color image enhancement. In: Proceedings of 3rd IEEE International Conference on Image Processing. vol. 3, pp. 1003–1006 vol.3 (1996). https://doi.org/10.1109/ICIP.1996.560995
33. Seif, G., Androutsos, D.: Edge-based loss function for single image super-resolution. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1468–1472 (2018). https://doi.org/10.1109/ICASSP.2018.8461664
34. Shi, Y., Liu, D., Zhang, L., Tian, Y., Xia, X., Fu, X.: Zero-ig: Zero-shot illumination-guided joint denoising and adaptive enhancement for low-light images. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3015–3024 (2024). https://doi.org/10.1109/CVPR52733.2024.00291
35. Singh, G., Mittal, A., et al.: Various image enhancement techniques-a critical review. International Journal of Innovation and Scientific Research **10**(2), 267–274 (2014)
36. Smith, A.R.: Color gamut transform pairs. SIGGRAPH Comput. Graph. **12**(3), 12–19 (Aug 1978). https://doi.org/10.1145/965139.807361, https://doi.org/10.1145/965139.807361
37. Tkalcic, M., Tasic, J.: Colour spaces: perceptual, historical and applicational background. In: The IEEE Region 8 EUROCON 2003. Computer as a Tool. vol. 1, pp. 304–308 vol.1 (2003). https://doi.org/10.1109/EURCON.2003.1248032
38. Vonikakis, V., Kouskouridas, R., Gasteratos, A.: On the evaluation of illumination compensation algorithms. Multimedia Tools and Applications **77**(8), 9211–9231 (2018)
39. Wang, C.Y., Yeh, I.H., Mark Liao, H.Y.: Yolov9: Learning what you want to learn using programmable gradient information. In: European conference on computer vision. pp. 1–21. Springer (2024)
40. Wang, H., Xu, K., Lau, R.W.: Local color distributions prior for image enhancement. In: European conference on computer vision. pp. 343–359. Springer (2022)
41. 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). https://doi.org/10.1109/TIP.2013.2261309
42. Wang, T., Zhang, K., Shen, T., Luo, W., Stenger, B., Lu, T.: Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 2654–2662 (2023)
43. Wang, W., Yang, H., Fu, J., Liu, J.: Zero-reference low-light enhancement via physical quadruple priors. In: 2024 IEEE/CVF Conference on Computer Vision

and Pattern Recognition (CVPR). pp. 26057–26066 (2024). https://doi.org/10.1109/CVPR52733.2024.02462
44. Wang, W., Yang, W., Liu, J.: Hla-face: Joint high-low adaptation for low light face detection. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16190–16199 (2021). https://doi.org/10.1109/CVPR46437.2021.01593
45. Wang, X., Ma, K., Liu, Q., Zou, Y., Fu, Y.: Multi-object tracking in the dark. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 382–392 (2024). https://doi.org/10.1109/CVPR52733.2024.00044
46. Wang, Y., Liu, Z., Liu, J., Xu, S., Liu, S.: Low-light image enhancement with illumination-aware gamma correction and complete image modelling network. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 13082–13091 (2023). https://doi.org/10.1109/ICCV51070.2023.01207
47. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
48. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: British Machine Vision Conference. British Machine Vision Association (2018)
49. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 17693–17703 (2022). https://doi.org/10.1109/CVPR52688.2022.01719
50. Xu, X., Wang, R., Lu, J.: Low-light image enhancement via structure modeling and guidance. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9893–9903 (2023). https://doi.org/10.1109/CVPR52729.2023.00954
51. Yan, Q., Feng, Y., Zhang, C., Pang, G., Shi, K., Wu, P., Dong, W., Sun, J., Zhang, Y.: Hvi: A new color space for low-light image enhancement. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5678–5687 (2025). https://doi.org/10.1109/CVPR52734.2025.00533
52. Yang, K.F., Cheng, C., Zhao, S.X., Yan, H.M., Zhang, X.S., Li, Y.J.: Learning to adapt to light. International Journal of Computer Vision **131**(4), 1022–1041 (2023)
53. Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE Transactions on Image Processing **30**, 2072–2086 (2021). https://doi.org/10.1109/TIP.2021.3050850
54. Yi, X., Xu, H., Zhang, H., Tang, L., Ma, J.: Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12302–12311 (2023)
55. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022)
56. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
57. Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: A practical low-light image enhancer. In: Proceedings of the 27th ACM international conference on multimedia. pp. 1632–1640 (2019)

58. Zhao, Z., Xiong, B., Wang, L., Ou, Q., Yu, L., Kuang, F.: Retinexdip: A unified deep framework for low-light image enhancement. IEEE Transactions on Circuits and Systems for Video Technology **32**(3), 1076–1088 (2022). https://doi.org/10.1109/TCSVT.2021.3073371