

# GT-Mean Loss: A Simple Yet Effective Solution for Brightness Mismatch in Low-Light Image Enhancement

Jingxi Liao, Shijie Hao\*, Richang Hong, Meng Wang,  
Hefei University of Technology, Hefei, China

{jingxi.ljx, hfut.hsj, hongrc.hfut, eric.mengwang}@gmail.com

## Abstract

Low-light image enhancement (LLIE) aims to improve the visual quality of images captured under poor lighting conditions. In supervised LLIE research, there exists a significant yet often overlooked inconsistency between the overall brightness of an enhanced image and its ground truth counterpart, referred to as brightness mismatch in this study. Brightness mismatch negatively impact supervised LLIE models by misleading model training. However, this issue is largely neglected in current research. In this context, we propose the GT-mean loss, a simple yet effective loss function directly modeling the mean values of images from a probabilistic perspective. The GT-mean loss is flexible, as it extends existing supervised LLIE loss functions into the GT-mean form with minimal additional computational costs. Extensive experiments demonstrate that the incorporation of the GT-mean loss results in consistent performance improvements across various methods and datasets.

## 1. INTRODUCTION

Low-light image enhancement (LLIE) is a crucial task in computer vision, aiming to improve the overall quality of images captured under poor lighting conditions [17, 19]. The primary objective of training a supervised LLIE model, denoted as  $f(\cdot)$ , is to map a low-light image  $x$  to an enhanced image  $f(x)$ , subjecting to the constraint that  $f(x)$  should resemble the ground truth (GT) image  $y$  as much as possible. While existing methods achieve reasonable brightness adjustment, they face a critical dilemma: models disproportionately prioritize global brightness alignment at the expense of suppressing other degradation factors like noise [20, 25, 37], color distortion [42, 48], and artifacts [49, 50]. This misalignment stems from three inter-related issues:

- **Inevitable Brightness Residual:** The inconsistency in overall brightness between the supervised low-light image enhancement (LLIE) function  $f(x)$  and the target image  $y$  is a common occurrence. The inherent nonlinear mapping from low-light to well-lit images inevitably leaves behind a small brightness residual, which can be simply expressed as  $\mathbb{E}[f(x)] \neq \mathbb{E}[y]$ . We refer to this phenomenon as brightness mismatch.
- **Sensitivity of Loss Functions:** Pixel-level losses (e.g., L1/L2) lack a perceptual saturation mechanism—they penalize brightness discrepancy regardless of whether these mismatches are visually negligible once  $\mathbb{E}[f(x)] \neq \mathbb{E}[y]$ .
- **Misalignment with Human Perception:** Under adequate brightness, human vision prioritizes noise suppression and color fidelity [2, 28], yet these factors contribute far less to pixel-level losses than brightness residuals.

As a result, the model disproportionately pursues marginal benefits in brightness alignment, entering a futile optimization cycle. This misguided effort directly contradicts the priorities of human vision—once basic brightness is met, noise and color accuracy dominate perceptual quality. Figure 1 illustrates a typical contradictory scenario.



Figure 1. (a) exhibits severe degradation (noise and blur), with overall brightness consistent with the GT, while (b) has mostly clear details and textures, with overall brightness at 0.9 times that of the GT. We use LPIPS(↓) [46], QALIGN’s [38] IQA(↑) and IAA(↑) modules to represent the image quality of (a) and (b). This example reveals an unexpected behavior of the L1 loss (a typical pixel-level loss): the L1 loss assigns a lower loss value to the low-quality image (a) and a higher loss value to the high-quality image (b).

\*Corresponding author

This work was supported by the National Natural Science Foundation of China under Grant 62172137.

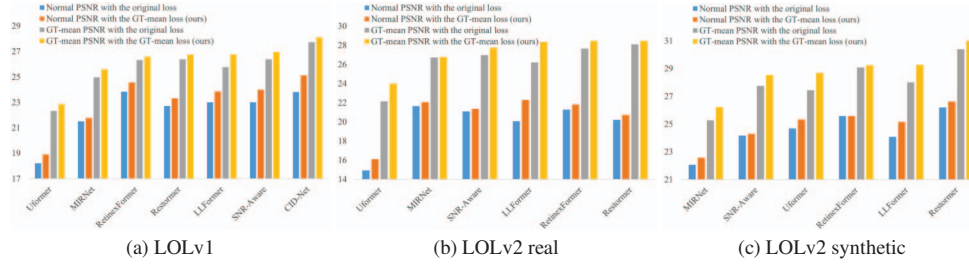


Figure 2. Performance across various supervised LLIE models trained by their original loss functions and our GT-mean loss functions. The performance is consistently improved when the GT-mean loss functions are adopted. Notably, this improvement is easily attainable, as the use of GT-mean loss functions is flexible and brings minimal additional computational costs during training.

The issue under brightness mismatch has largely been ignored in existing supervised LLIE research, despite some indirect solutions that do not primarily address this problem. For example, [23, 36, 39, 43] designed multiple sub-networks to decouple brightness from other factors and optimized them separately. Nevertheless, the divide-and-conquer roadmap inevitably complicates the model design, as well as introducing significant computational overhead. Others use perceptual losses [12] to bypass pixel-level alignment, sacrificing reconstruction fidelity. A unified, efficient solution remains elusive.

Inspired by human visual adaptation, we propose the GT-Mean Loss, which dynamically balances two optimization objectives to enable end-to-end training:

- **Original Loss Term.** Maintain the model’s baseline capabilities by directly optimizing  $L(f(x), y)$ .
- **Brightness-Aligned Loss Term.** This term is defined as  $L(\lambda f(x), y)$ , where  $\lambda = \mathbb{E}[y]/\mathbb{E}[f(x)]$  linearly scales  $f(x)$  such that the mean brightness of the scaled output  $\mathbb{E}[\lambda f(x)]$  aligns with  $\mathbb{E}[y]$ . This eliminates the interference of brightness differences during optimization.

To dynamically balance these two terms, we introduce  $W$ , which represents the Bhattacharyya distance between the brightness distributions of  $f(x)$  and  $y$ . Unlike multi-stage frameworks, GT-Mean Loss enables brightness-decoupled optimization without requiring architectural modifications or manually tuned thresholds.

Although perceptual losses can to some extent ignore the impact of brightness mismatch, almost all perceptual losses abandon corrections. In contrast, our GT-mean loss directly addresses the influence of brightness mismatch on pixel-level loss. The results in our experimental section validate this point in Section 4.2.

In summary, the GT-mean loss is highlighted in the following aspects:

- **Simplicity:** The construction of the GT-mean loss is both theoretically and practically straightforward. Its underlying mechanism is easy to understand, and its implementation is uncomplicated.

- **Flexibility:** The GT-mean loss is highly flexible. For instance, the  $L_1$  loss can be directly extended into the  $L_1$  GT-mean version. This character makes adopting the GT-mean loss a universal choice for supervised LLIE models to upgrade their loss functions.
- **Low Cost:** Using the GT-mean loss introduces minimal overhead during training (approximately doubling the original loss computation). It is negligible compared to the overall model optimization process.
- **Effectiveness:** Extensive experiments have demonstrated that the GT-mean loss consistently improves model performance across a wide range of supervised LLIE methods (as shown in Figure 2).

## 2. Related works

Loss function is essential in LLIE tasks, as it directs model training. We categorize the loss functions in LLIE into two groups based on their purposes.

**Fidelity Losses.** These losses are designed to ensure that the enhanced image  $f(x)$  closely resembles the ground truth  $y$ . For example, the  $L_1$ -like loss functions directly ensure the pixel-level fidelity [17, 19]. Loss functions focused on color representation often use color histogram-based metrics [42], while others preserve fidelity at the frequency domain [10, 30]. Among these, pixel-level loss functions are essential for image reconstruction, but their inherent point-to-point computation can easily misinterpret slight brightness variations as image degradations. Additionally, some loss functions aim to maintain fidelity at higher representation levels, such as the perceptual loss [12]. Recently, some novel loss functions have emerged that subtly utilize fine-grained semantic information [18, 40]. While these high-level losses demonstrate effectiveness in perceptual enhancement, they cannot fully substitute pixel-level. Current implementations must be combined with pixel-based losses to ensure functional model performance.

While these high-level losses demonstrate effectiveness in semantic-aware enhancement, they cannot fully substitute pixel-level supervision.

**Prior-Based Losses.** These losses integrate domain-specific prior knowledge into LLIE models, aiming to maximize the use of available information. Typical examples are the Retinex-based methods [7, 23, 36, 39, 43, 47], which are founded on the Retinex theory that an image is composed of illumination map and reflectance map. Specific loss functions are employed to penalize the locally smooth properties of illumination map and the lightness-insensitive properties of reflectance map. While providing clear physical interpretability, the ill-posed nature of Retinex decomposition, for instance, can propagate estimation errors in illumination and reflectance maps. This inherent ambiguity may lead to suboptimal results in practical applications. Furthermore, for unsupervised LLIE models that lack GT images for training, prior-based loss functions are essential for guiding model optimization [11, 16, 22, 33]. However, these loss functions can be less flexibility and robust. For example, [11, 22, 33] heavily rely on specific network designs. ZeroDCE [16] builds an exposure control loss function with a hard threshold, which may result in over-exposure.

To pursue comprehensive visual quality enhancement, LLIE models tend to incorporate multiple loss functions. While this strategy can lead to improved results, it also increases the burden of model design and hyperparameter tuning. More importantly, as the existing loss functions overlook the brightness mismatch factor, the fundamental challenges brought by this factor remain unaddressed.

### 3. METHOD

In Section 3.1 present the formulation of the GT-mean loss. In Section 3.2 detail the crucial component  $W$  of this loss function. Section 3.3 summarizes the features of the proposed loss.

#### 3.1. GT-mean loss

In general, the GT-mean loss  $L_{GT}(f(x), y)$  can be regarded as an extension of the existing loss  $L(f(x), y)$  used for LLIE. To deal with the issues arising from brightness mismatch, the key to constructing  $L_{GT}(f(x), y)$  is matching the average brightness of  $f(x)$  and  $y$ . Furthermore,  $L_{GT}(f(x), y)$  is designed to retain the form and effectiveness of the original loss  $L(f(x), y)$ . Based on this,  $L_{GT}(f(x), y)$  is formulated as follows:

$$L_{GT}(f(x), y) = W \cdot \underbrace{L(f(x), y)}_{\text{original loss}} + (1 - W) \cdot \underbrace{L\left(\frac{\mathbb{E}[y]}{\mathbb{E}[f(x)]}f(x), y\right)}_{\text{brightness-aligned loss}}, \quad (1)$$

where  $\frac{\mathbb{E}[y]}{\mathbb{E}[f(x)]}$  is a scaling factor for aligning the average brightness of  $f(x)$  and  $y$ . To clarify,  $\mathbb{E}[y]$  and  $\mathbb{E}[f(x)]$  de-

note the average brightness of the ground truth image  $y \in \mathbb{R}^{1 \times 3 \times H \times W}$  and the enhanced image  $f(x) \in \mathbb{R}^{1 \times 3 \times H \times W}$ , respectively. The weight  $W \in [0, 1]$  balances the two terms in  $L_{GT}$ . It is noted that the choice of  $L(\cdot)$  is arbitrary, provided that it accepts  $f(x)$  and  $y$  as inputs.

#### 3.2. Weight Design

##### 3.2.1. Probabilistic Modeling on Average Brightness

According to the Contrast Sensitivity Function [3, 26], human vision is sensitive to local contrasts (e.g., edges, textures, and intensity variations) but less sensitive to minor shifts in average brightness  $\mathbb{E}[\cdot]$ . Moreover, research by Mathys [5, 6] has shown that uncertainty in human perceptual behavior is often well modeled by Gaussian distributions, making them a suitable choice for representing brightness perception variability.

This perceptual characteristic motivates our probabilistic modeling of brightness distributions: by treating  $\mathbb{E}[\cdot]$  as a random variable rather than a fixed value, we represent the inherent uncertainty in human perception of brightness. This modeling approach enhances the controllability of the loss function during training, as it allows us to estimate the weight  $W$  using common metrics like the Kullback-Leibler (KL) divergence or Wasserstein distance. These metrics ensure a smooth transition between the two terms in Eq. 1, avoiding abrupt changes that could destabilize the optimization process. Consequently, the loss function exhibits good continuity with respect to parameter changes and features flat regions around its minima. These flat regions indicate that small perturbations in brightness do not significantly affect the loss value, making the optimization process more robust (as exemplified in Section 4.5).

Based on these considerations, we treat  $\mathbb{E}[\cdot]$  as an observation from a random variable  $\tilde{\mathbb{E}}[\cdot]$  that follows a Gaussian distribution. Specifically, for a single ground truth  $y \in \mathbb{R}^{1 \times 3 \times H \times W}$ , it can be regarded as an observation from  $\mathbb{E}[y]$ , represented as:

$$\tilde{\mathbb{E}}[y] = \alpha \mathbb{E}[y], \quad \alpha \sim \mathcal{N}(1, \sigma_\alpha^2), \quad (2)$$

where  $\alpha$  determines the probability distribution type of  $\tilde{\mathbb{E}}[y]$ , and  $\sigma_\alpha^2$  defines its spread. Therefore, we have average brightness distribution  $p(\tilde{\mathbb{E}}[y]) = \mathcal{N}(\mu_y, \sigma_y^2)$ , with  $\mu_y = \mathbb{E}[y]$  and  $\sigma_y = \sigma_\alpha \mathbb{E}[y]$ . Similarly, for a single enhanced image  $f(x) \in \mathbb{R}^{1 \times 3 \times H \times W}$ , its average brightness  $\mathbb{E}[f(x)]$  can also be regarded as an observation from  $\tilde{\mathbb{E}}[f(x)]$ , represented as:

$$\tilde{\mathbb{E}}[f(x)] = \beta \mathbb{E}[f(x)], \quad \beta \sim \mathcal{N}(1, \sigma_\beta^2), \quad (3)$$

where  $\beta$  determines the probability distribution type of  $\tilde{\mathbb{E}}[f(x)]$ , and  $\sigma_\beta^2$  defines its spread. Consequently, we have average brightness distribution  $q(\tilde{\mathbb{E}}[f(x)]) = \mathcal{N}(\mu_{fx}, \sigma_{fx}^2)$ , with  $\mu_{fx} = \mathbb{E}[f(x)]$  and  $\sigma_{fx} = \sigma_\beta \mathbb{E}[f(x)]$ .

### 3.2.2. Estimation of $W$

To quantify the divergence between the two distributions, we employ the Bhattacharyya distance [13], which offers a closed-form solution for Gaussian distributions, significantly enhancing computational efficiency. For Gaussian distributions  $p(\mathbb{E}[y]) = \mathcal{N}(\mu_y, \sigma_y^2)$  and  $q(\mathbb{E}[f(x)]) = \mathcal{N}(\mu_{fx}, \sigma_{fx}^2)$ , the closed-form expression for the Bhattacharyya distance  $D_B$  is given by:

$$D_B(p||q) = \frac{1}{4} \frac{(\mu_y - \mu_{fx})^2}{\sigma_y^2 + \sigma_{fx}^2} + \frac{1}{2} \ln \left( \frac{\sigma_y^2 + \sigma_{fx}^2}{2\sigma_y\sigma_{fx}} \right), \quad (4)$$

where  $\mu_y \triangleq \mathbb{E}[y]$  and  $\mu_{fx} \triangleq \mathbb{E}[f(x)]$  denote the expected brightness values, while  $\sigma_y^2 = (\sigma_\alpha \mathbb{E}[y])^2$  and  $\sigma_{fx}^2 = (\sigma_\beta \mathbb{E}[f(x)])^2$  characterize distribution spreads derived from Eq. (2) and Eq. (3).

To enforce  $W$  as a normalized weighting coefficient, we apply value clipping:

$$W \leftarrow \text{clip}(D_B(p||q), 0, 1) \quad (5)$$

This operation induces three interpretable regimes:

- $W = 1$ : Maximum emphasis on general quality when  $\mathbb{E}[y]$  and  $\mathbb{E}[f(x)]$  exhibit strong divergence ( $D_B \geq 1$ ).
- $W = 0$ : Exclusive focus on brightness-aligned refinement when distributions coincide ( $D_B = 0$ ).
- Smooth transition via  $W \in (0, 1)$  during intermediate phases.

In supervised LLIE, the enhanced image  $f(x)$  should closely resemble the ground truth image  $y$ . Consequently, we assume that the shape of  $p(\mathbb{E}[y])$  and  $q(\mathbb{E}[f(x)])$  are similar. Based on this assumption, we equate  $\sigma_\alpha^2$  and  $\sigma_\beta^2$ , setting them both to  $\sigma^2$ . In our experiments, we empirically set  $\sigma$  as 0.1 for all the comparisons. In Section 4.4, we delve deeper into the influence of  $\sigma$  on the performance of LLIE.

### 3.3. Discussion of the GT-mean loss

The primary strength of the GT-mean loss lies in its ability to dynamically balance the model’s focus during training. In the early stages, when the difference between  $p(\mathbb{E}[y])$  and  $q(\mathbb{E}[f(x)])$  is significant, the weight  $W$  is expected to approach 1, causing the first term  $L(f(x), y)$  to dominate the overall loss function. This behavior ensures that the GT-mean loss closely resembles the original loss  $L(f(x), y)$ , thereby preserving the baseline model’s capabilities.

As training progresses and  $p(\mathbb{E}[y])$  and  $q(\mathbb{E}[f(x)])$  become closer,  $W$  decreases to a smaller value. This shift in weight emphasizes the second term  $L\left(\frac{\mathbb{E}[y]}{\mathbb{E}[f(x)]}, f(x), y\right)$ , ensuring that  $f(x)$  and  $y$  are compared under the condition of average brightness alignment. At this stage, the GT-mean loss operates under the assumption of consistent

image brightness, compelling the model to focus on addressing degradation factors beyond brightness enhancement, such as noise, contrast, or structural details. This dynamic adjustment enables the model to refine its performance progressively, further mitigating degradation while preserving the capabilities of the baseline model.

In application, flexibility is the primary advantage of the GT-mean loss, as it is independent of the model architecture and only requires the enhanced image  $f(x)$  and the ground truth  $y$  as inputs. Therefore, the GT-mean loss can be used in any supervised LLIE method by simply extending its supervised loss function into the GT-mean version.

The GT-mean loss is also efficient. Compared to the original loss, it doubles the computation, and  $W$  is estimated with an analytical solution. Despite there is extra computation overhead, the overall increase introduced by the GT-mean loss is negligible throughout the training process.

## 4. Experiment

### 4.1. Datasets and Settings

**Baselines.** We selected a total of seven supervised LLIE models as baselines. To validate the combined effect of the GT-Mean loss and perceptual loss, the baselines include SNR-aware [41] and CID-Net [42], both of which incorporate perceptual loss. All baselines and their original loss functions are listed in Table 1:

Table 1. Baselines and Their Loss Functions

Method	Loss Function
Restormer [45]	$L_1$ loss
RetinexFormer [4]	$L_1$ loss
LLFormer [32]	Smooth $L_1$ loss [8]
MIRNet [44]	Charbonnier loss [1]
Uformer [35]	Charbonnier loss [1]
SNR-Aware [41]	Charbonnier loss [1], perceptual loss [12]
CID-Net [42]	$L_1$ loss, edge loss [27], perceptual loss [12]

**Datasets.** We conducted experiments on both paired and unpaired datasets to evaluate our loss. For paired datasets, we used LOLv1 [36], LOLv2-real [43], and LOLv2-syn [43]. Specifically, the LOLv1 dataset includes 485 training images and 15 testing images. The LOLv2-real dataset includes 689 training images and 100 testing images. The LOLv2-synthetic dataset includes 900 training images and 100 testing images. For unpaired datasets, we chose DICM [15], VV [29], NPE [31], MEF [21], and LIME [9] as the test sets.

**Evaluation Metrics.** For the paired datasets, we employed standard evaluation metrics, including PSNR, SSIM [34], and LPIPS [46]. To further align with human perception, we utilized the large multi-modality model QALIGN [38], specifically its image quality assessment (IQA) and image aesthetic assessment (IAA) modules, for more perceptually

Table 2. Comparison on the Paired Datasets. Red(Green) denotes the improvement(reduction) of performance.

Datasets	Methods	Normal			GT-mean			IQA $\uparrow$	IAA $\uparrow$
		PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$		
LOLv1	Uformer	18.218	0.771	0.212	22.325	0.81	0.195	3.317	1.959
	Uformer with GT-mean loss (ours)	18.915 (+0.697)	0.795 (+0.023)	0.168 (-0.044)	22.854 (+0.529)	0.830 (+0.019)	0.157 (-0.038)	3.331( $\uparrow$ )	1.971( $\uparrow$ )
	MIRNet	21.512	0.788	0.222	24.968	0.8	0.216	2.917	1.745
	MIRNet with GT-mean loss (ours)	21.780 (+0.268)	0.804 (+0.016)	0.196 (-0.026)	25.596 (+0.628)	0.818 (+0.018)	0.189 (-0.027)	3.039( $\uparrow$ )	1.758( $\uparrow$ )
	RetinexFormer	23.83	0.832	0.141	26.312	0.844	0.134	3.027	1.800
	RetinexFormer with GT-mean loss (ours)	24.561 (+0.731)	0.834 (+0.003)	0.138 (-0.003)	26.586 (+0.274)	0.849 (+0.005)	0.132 (-0.002)	3.373( $\uparrow$ )	1.956( $\uparrow$ )
	Restormer	22.718	0.83	0.128	26.375	0.848	0.122	3.567	2.032
	Restormer with GT-mean loss (ours)	23.313 (+0.595)	0.837 (+0.007)	0.122 (-0.006)	26.743 (+0.368)	0.855 (+0.007)	0.117 (-0.005)	3.672( $\uparrow$ )	2.054( $\uparrow$ )
	LLFormer	23.007	0.805	0.183	25.762	0.823	0.178	3.087	1.946
	LLFormer with GT-mean loss (ours)	23.847 (+0.840)	0.830 (+0.025)	0.138 (-0.045)	26.769 (+1.007)	0.846 (+0.023)	0.133 (-0.045)	3.419( $\uparrow$ )	2.049( $\uparrow$ )
SNR-Aware	23.005	0.824	0.164	26.373	0.843	0.158	3.330	1.893	
SNR-Aware with GT-mean loss (ours)	23.992 (+0.988)	0.836 (+0.012)	0.158 (-0.006)	26.942 (+0.569)	0.853 (+0.009)	0.153 (-0.005)	3.509( $\uparrow$ )	1.913( $\uparrow$ )	
CID-Net	23.809	0.857	0.086	27.715	0.876	0.079	4.087	2.157	
CID-Net with GT-mean loss (ours)	25.122 (+1.313)	0.865 (+0.008)	0.081 (-0.005)	28.108 (+0.393)	0.878 (+0.002)	0.075 (-0.004)	4.074( $\downarrow$ )	2.161( $\uparrow$ )	
LOLv2-real	Uformer	14.941	0.76	0.228	22.148	0.831	0.199	3.478	2.009
	Uformer with GT-mean loss (ours)	16.103 (+1.162)	0.792 (+0.032)	0.180 (-0.048)	23.989 (+1.841)	0.858 (+0.026)	0.156 (-0.043)	3.778( $\uparrow$ )	2.048( $\uparrow$ )
	MIRNet	21.648	0.81	0.313	26.712	0.827	0.303	2.598	1.520
	MIRNet with GT-mean loss (ours)	22.050 (+0.402)	0.830 (+0.021)	0.214 (-0.099)	26.769 (+0.057)	0.846 (+0.019)	0.208 (-0.101)	2.924( $\uparrow$ )	1.702( $\uparrow$ )
	RetinexFormer	21.272	0.841	0.163	27.65	0.877	0.152	2.714	1.590
	RetinexFormer with GT-mean loss (ours)	21.810 (+0.538)	0.852 (+0.011)	0.143 (-0.020)	28.437 (+0.787)	0.879 (+0.002)	0.134 (-0.018)	3.206( $\uparrow$ )	1.884( $\uparrow$ )
	Restormer	20.235	0.841	0.162	28.159	0.88	0.147	3.478	1.987
	Restormer with GT-mean loss (ours)	20.717 (+0.482)	0.845 (+0.004)	0.149 (-0.013)	28.440 (+0.281)	0.884 (+0.004)	0.135 (-0.012)	3.554( $\uparrow$ )	2.020( $\uparrow$ )
	LLFormer	21.308	0.803	0.248	27.052	0.828	0.236	2.882	1.827
	LLFormer with GT-mean loss (ours)	22.291 (+0.983)	0.844 (+0.041)	0.166 (-0.082)	28.334 (+1.282)	0.870 (+0.042)	0.156 (-0.080)	3.104( $\uparrow$ )	1.880( $\uparrow$ )
SNR-Aware	21.103	0.839	0.169	26.971	0.866	0.161	3.354	1.879	
SNR-Aware with GT-mean loss (ours)	21.350 (+0.247)	0.844 (+0.005)	0.164 (-0.005)	27.740 (+0.770)	0.875 (+0.010)	0.154 (-0.007)	3.468( $\uparrow$ )	1.889( $\uparrow$ )	
LOLv2-synthetic	Uformer	24.693	0.932	0.060	27.438	0.941	0.055	3.148	2.114
	Uformer with GT-mean loss (ours)	25.319 (+0.626)	0.940 (+0.007)	0.049 (-0.011)	28.683(+1.245)	0.948(+0.007)	0.045 (-0.010)	3.191( $\uparrow$ )	2.144( $\uparrow$ )
	MIRNet	22.059	0.894	0.122	25.274	0.908	0.114	2.956	2.145
	MIRNet with GT-mean loss (ours)	22.576 (+0.517)	0.906 (+0.011)	0.104 (-0.018)	26.215(+0.941)	0.918(+0.010)	0.094(-0.020)	3.064( $\uparrow$ )	2.187( $\uparrow$ )
	RetinexFormer	25.281	0.928	0.064	28.827	0.939	0.057	3.102	2.099
	RetinexFormer with GT-mean loss (ours)	25.583 (+0.299)	0.933 (+0.005)	0.063 (-0.001)	29.261(+0.434)	0.944(+0.005)	0.056(-0.001)	3.197( $\uparrow$ )	2.130( $\uparrow$ )
	Restormer	26.288	0.944	0.045	30.57	0.955	0.039	3.350	2.187
	Restormer with GT-mean loss (ours)	26.630 (+0.342)	0.946 (+0.002)	0.041 (-0.004)	31.001(+0.431)	0.957(+0.002)	0.036 (-0.003)	3.404( $\uparrow$ )	2.218( $\uparrow$ )
	LLFormer	24.195	0.918	0.07	27.862	0.930	0.064	3.176	2.137
	LLFormer with GT-mean loss (ours)	25.152 (+0.957)	0.932 (+0.014)	0.058 (-0.012)	29.266(+1.404)	0.945(+0.015)	0.051 (-0.013)	3.283( $\uparrow$ )	2.177( $\uparrow$ )
SNR-Aware	24.173	0.924	0.064	27.756	0.937	0.058	3.275	2.209	
SNR-Aware with GT-mean loss (ours)	24.301 (+0.128)	0.933(+0.009)	0.057(-0.007)	28.525(+0.769)	0.945(+0.008)	0.050 (-0.008)	3.326( $\uparrow$ )	2.207( $\downarrow$ )	

meaningful evaluations.

For unpaired datasets, we evaluated the model performance using QALIGN’s IQA and IAA modules, NIQE [24], and MUSIQ [14].

**Implementation Details.** To retrain the baselines equipped with GT-mean loss functions, we followed their official settings, which can be found in Appendix.

## 4.2. Quantitative Results

**Paired Datasets.** As shown in Table 2, we evaluated the performance of the GT-Mean loss on three paired datasets, covering reference-based evaluation metrics (PSNR, SSIM, LPIPS) and their GT-Mean versions (GT-Mean PSNR, GT-Mean SSIM, GT-Mean LPIPS), along with QALIGN for image quality assessment (IQA) and image alignment assessment (IAA).

The experimental results demonstrate that, across the seven baseline models, regardless of the type of original

loss function, extending them with the GT-Mean loss consistently improves performance across all evaluation metrics. This outcome validates the universal effectiveness of the GT-Mean loss in LLIE tasks. Notably, both SNR-aware and CIDNet methods exhibit significant performance improvements after incorporating the GT-Mean loss, indicating that the GT-Mean loss effectively complements perceptual loss. One possible explanation is that the GT-Mean loss compensates for the lack of pixel-level perception in perceptual loss. Furthermore, since the weight  $W$  is computed analytically during training and does not require gradient propagation, its cost is negligible. Compared to the model’s forward and backward propagation, the original loss computation is minimal. The second term in the GT-Mean Loss is essentially a brightness-adjusted version of the original loss, sharing the same computational complexity. Thus, the GT-Mean Loss incurs only twice the computational overhead of the original loss. For example,

using GT-Mean Loss with Retinexformer increases training time by just 1%. Therefore, this loss function can be seamlessly integrated into existing supervised LLIE methods, achieving nearly zero-cost performance improvements.

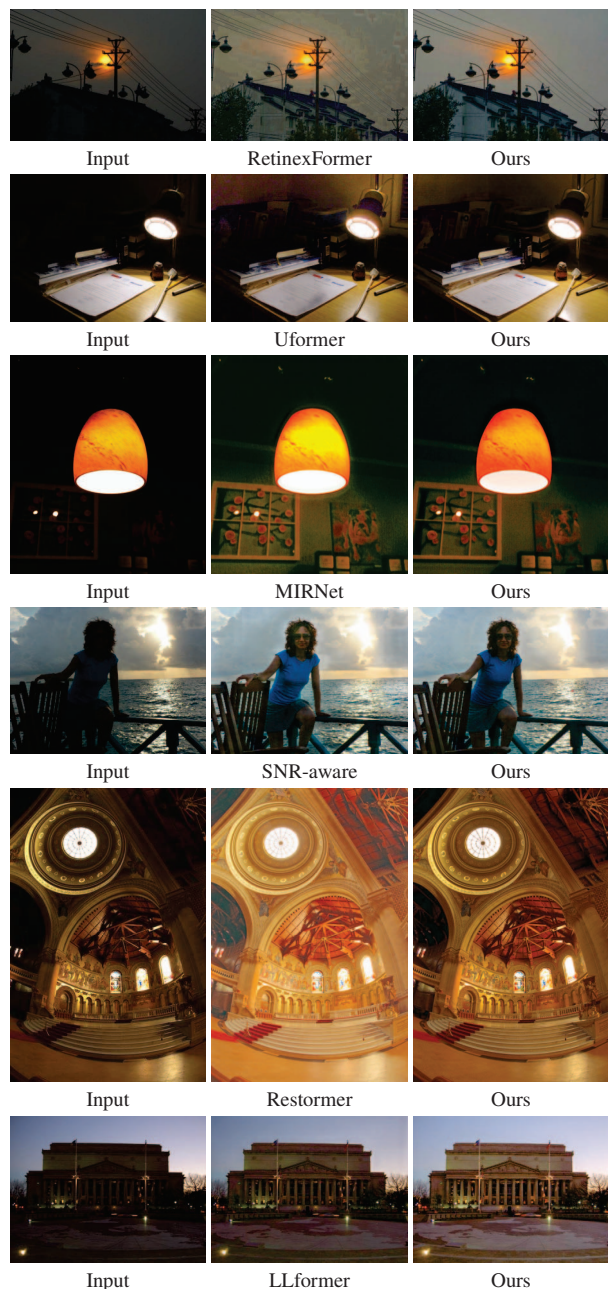


Figure 3. Visual comparison of results on the five unpaired datasets: original image (left), results from baselines (middle), and results from their improved versions based on our loss (right).

**Unpaired Datasets.** Table 3 presents the performance

of the models on five unpaired datasets. Compared with the baseline performance, using GT-mean loss demonstrates superior or comparable results in most cases in terms of the non-reference evaluation metrics. These results indicate that the GT-Mean loss possesses strong generalization capabilities: even when tested on unseen images, performance improvements are still observed, proving that its success on supervised datasets is not due to overfitting. As shown in Figure 3, the examples based on the GT-Mean loss demonstrate improvements over the baseline results. Specifically, the baseline results of Retinexformer and Uformer show artifacts in sky regions and paper textures respectively. For MIRNet and SNR-Aware, the baseline outputs exhibit noticeable halo artifacts around object contours. Restormer and LLformer represent typical cases of over-exposure and under-exposure in their baseline results. Despite the diverse failure patterns observed across these baseline models, all these issues are effectively mitigated through the integration of our GT-Mean loss. This comprehensive improvement demonstrates the robustness and adaptability of our approach in enhancing image quality across diverse model architectures.

### 4.3. Impact of GT-Mean Loss on Training

#### 4.3.1. Experimental Configuration.

We evaluate training dynamics using Retinexformer [4] on the LOLv1 training dataset with four distinct strategies:

- *Baseline*: Original Retinexformer loss (L1 loss).
- *Hybrid*: A phased approach transitioning from the baseline loss (0–70K iterations) to a brightness-aligned loss in Eq. 1 (80K–150K iterations), with linear interpolation between 70K and 80K iterations.
- *Full Brightness Alignment*: Exclusive use of the brightness-aligned loss throughout training.
- *GT-Mean Loss*: Our proposed dynamic combination of the baseline and brightness-aligned losses.

Metrics were recorded every 5K iterations over a total of 150K iterations, including perceptual quality (LPIPS), pixel fidelity (PSNR), and brightness discrepancy  $|\mathbb{E}[y] - \mathbb{E}[f(x)]|$ .

#### 4.3.2. Quantitative Analysis.

As shown in Figure 4A, three key findings are revealed: **Does brightness alignment improve perception?** We first compare the hybrid strategy with baseline. When brightness alignment is introduced at 80K iterations, the hybrid approach achieves 0.129 LPIPS (vs. baseline 0.155), indicating enhanced perceptual quality. However, this comes at the cost of increased mean brightness discrepancy (from 0.028 to 0.075) and 4.6dB PSNR drop. These results suggest that while the introduction of brightness alignment in the hybrid strategy benefits human perception, its linear implementation disrupts the learning of brightness distribution.

Table 3. Comparison on the Unpaired Datasets. Red(Green) denotes the improvement(reduction) of performance. Note that all the models were trained on the LOLv2-synthetic dataset.

Method	DICM				MEF				LIME				NPE				VV				AVG			
	NIQE↓	MUSIQ↑	IQA↑	IAA↑	NIQE↓	MUSIQ↑	IQA↑	IAA↑	NIQE↓	MUSIQ↑	IQA↑	IAA↑	NIQE↓	MUSIQ↑	IQA↑	IAA↑	NIQE↓	MUSIQ↑	IQA↑	IAA↑	NIQE↓	MUSIQ↑	IQA↑	IAA↑
Restormer	3.220	58.525	3.885	2.800	3.660	56.528	3.215	2.415	3.660	58.461	3.267	2.466	3.470	61.031	3.781	2.735	3.290	37.919	3.710	2.264	3.460	54.493	3.572	2.536
Restormer with GT-mean loss (ours)	3.180	58.604	3.913	2.821	3.630	56.522	3.208	2.431	3.630	58.124	3.380	2.521	3.450	60.971	3.820	2.769	3.300	38.290	3.712	2.255	3.440(+)	54.502(+)	3.607(+)	2.559(+)
MIRNET	3.820	52.467	3.111	2.237	3.670	47.399	2.742	2.066	4.230	54.837	2.860	2.088	3.470	58.641	3.285	2.274	3.640	54.566	2.955	2.162	3.770	53.582	2.991	2.203
MIRNET with GT-mean loss (ours)	3.200	53.188	3.295	2.375	3.600	47.611	2.817	2.067	4.330	55.776	2.747	2.058	3.500	58.718	3.366	2.428	3.710	54.891	3.120	2.215	3.670(-)	54.037(+)	3.069(+)	2.229(+)
Retinexformer	3.230	57.398	3.800	2.740	3.860	56.170	3.137	2.373	3.880	57.262	3.111	2.323	3.380	60.507	3.673	2.699	2.730	37.513	3.471	2.154	3.420	53.77	3.438	2.458
Retinexformer with GT-mean loss (ours)	3.210	57.247	3.805	2.773	3.820	56.633	3.195	2.412	3.840	57.374	3.273	2.423	3.370	60.682	3.706	2.719	2.770	37.654	3.517	2.166	3.400(-)	53.918(+)	3.499(+)	2.498(+)
SNR	6.070	47.025	3.379	2.526	4.270	48.685	2.531	1.943	6.060	49.216	2.836	2.102	6.470	46.441	3.445	2.551	11.520	23.186	3.067	1.955	6.880	42.911	3.052	2.215
SNR with GT-mean loss (ours)	6.120	47.430	3.521	2.571	4.260	48.780	2.571	1.957	6.110	49.008	2.946	2.152	6.460	46.602	3.580	2.605	11.550	23.853	3.137	1.997	6.900(+)	43.135(+)	3.151(+)	2.256(+)
Uformer	3.080	58.084	2.971	2.144	3.720	56.177	3.177	2.403	3.660	57.698	2.646	1.967	3.400	61.310	2.938	2.131	2.700	36.235	2.904	1.839	3.310	53.901	2.927	2.097
Uformer with GT-mean loss (ours)	3.120	58.981	3.067	2.180	3.690	56.641	3.229	2.426	3.640	58.200	2.712	1.973	3.380	61.704	2.956	2.113	2.700	36.695	3.001	1.848	3.300(-)	54.444(+)	2.993(+)	2.108(+)
LLformer	3.260	56.642	3.832	2.788	3.750	53.335	2.667	1.993	4.010	55.671	3.040	2.343	3.320	59.824	3.657	2.716	3.160	60.885	3.557	2.249	3.500	57.271	3.351	2.418
LLformer with GT-mean loss (ours)	3.050	57.038	3.910	2.837	3.650	53.842	2.706	2.017	4.070	55.830	3.118	2.416	3.330	60.044	3.707	2.731	2.990	60.858	3.563	2.231	3.410(-)	57.522(+)	3.401(+)	2.446(+)

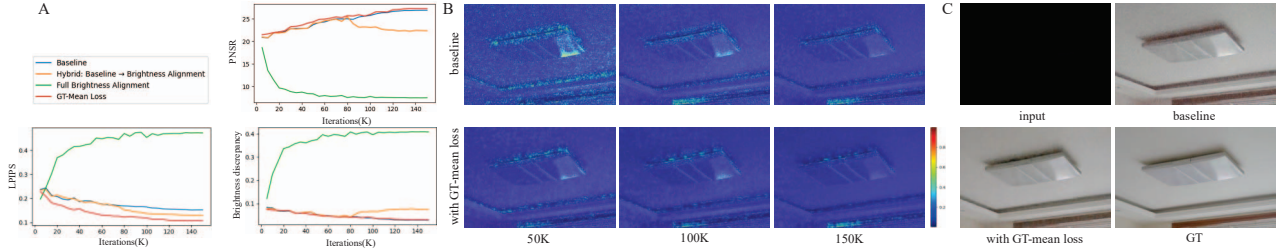


Figure 4. Impact of GT-Mean Loss integration on training dynamics and enhancement quality. (A) Evolution of LPIPS (perceptual quality), PSNR (pixel fidelity), and brightness discrepancy ( $|\mathbb{E}[y] - \mathbb{E}[f(x)]|$ ) across four training strategies. (B) Absolute residual heatmaps ( $|y - f(x)|$ ) at critical training stages. (C) Final output comparison showing improvements in color fidelity and detail preservation.

**Can brightness alignment alone achieve performance improvement?** We found that the full brightness alignment strategy causes a monotonic increase in LPIPS from 0.191 to 0.473, and PSNR decreases from 18.6 dB to 7.9 dB. These results demonstrate that the original loss term is crucial to the model’s training process. Completely removing the original loss term severely impairs the model’s performance.

**Does GT-Mean Loss achieve simultaneous improvement in perception and fidelity?** GT-Mean Loss achieves optimal LPIPS, PSNR, and mean brightness discrepancy. Its dynamic weighting mechanism  $W$  adaptively adjusts the original loss term and brightness alignment term, avoiding the limitations of a single objective. Unlike the phase fluctuations of hybrid strategies or the collapse associated with full brightness alignment, GT-Mean Loss maintains continuous positive optimization of both perception and fidelity throughout training, leading to a transition from conflict to mutual enhancement.

### 4.3.3. Qualitative Analysis.

In Figure 4B, we recorded the absolute difference  $|y - f(x)|$  as residual heatmaps at 50K, 100K, and 150K iterations. The GT-Mean Loss demonstrates significantly smaller residuals in both texture and noise regions compared to the baseline, suggesting its superior capability in detail refinement.

In Figure 4C, the final results generated using GT-Mean Loss exhibit excellent alignment with the ground truth (GT) in terms of color fidelity and structural clarity. In con-

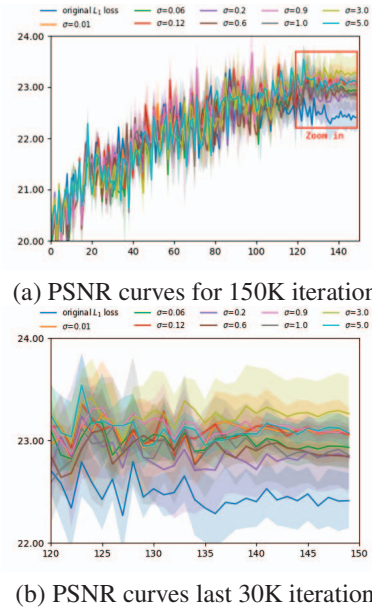


Figure 5. The effect of different  $\sigma$  on model performance.

trast, the baseline method shows noticeable color distortions, such as red artifacts, and fails to fully suppress noise.

### 4.4. Effect of the parameter $\sigma$

To investigate the influence of  $\sigma$ , we conducted experiments on LOLv1 using RetinexFormer trained with GT-mean  $L_1$

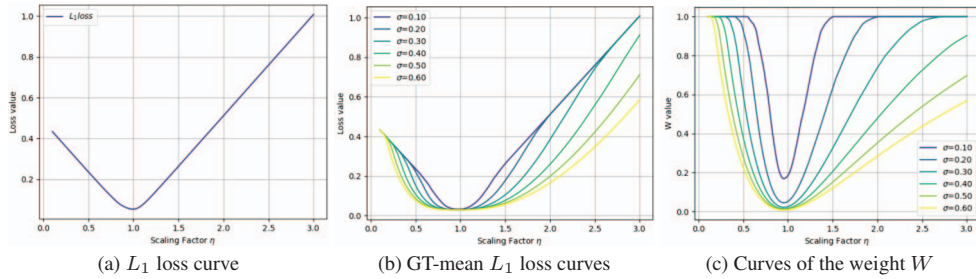


Figure 6. Loss curves and weight curves for analyzing the effectiveness of the GT-mean loss.

loss under different  $\sigma$  values. We selected 10 different  $\sigma$  values, running each setting three times for consistency. Notably,  $\sigma = 0$  represents a special case where the GT-mean  $L_1$  loss degrades to the original  $L_1$  loss. For every 1,000 (1K) iterations in the 150K iterations, we calculated mean and variance of PSNR, shown in Figure 5 for demonstrating the trend during training. In the early stages (as can be seen in Figure 5 (a)), the curve tendencies under different  $\sigma$  settings are similar. Considering the curve with  $\sigma = 0$  is the original  $L_1$  loss, we can verify that the GT-mean loss at the early stage behaves like the original  $L_1$  loss. In contrast, as shown in the zoomed-in views of the last 30K iterations (Figure 5 (b)), we observe that all settings become stable, and the settings with non-zero  $\sigma$  consistently perform better than  $\sigma = 0$ . This observation shows that the GT-mean loss diverges significantly from the original  $L_1$  loss in the late training stage. The second term in Eq.1 allows the GT-mean loss to continuously improve model performance. In addition, the experiment shows that the choice of  $\sigma$  value is open. As  $\sigma$  measures the spread of the random variable  $\tilde{\mathbb{E}}[\cdot]$  deviating from the observed average image brightness  $\mathbb{E}[\cdot]$  in our modeling, we recommend using a small value, such as  $\sigma = 0.1$  for real world application.

#### 4.5. Further analysis on the GT-mean loss

In this experiment, we further investigate the difference between the original loss and the GT-mean loss. Specifically, we randomly selected a batch of low-light images (batchsize = 8) and their corresponding ground truth images. These images were enhanced using RetinexFormer to produce enhanced outputs  $f(x)$ . To simulate the varying brightness, we introduced a unified scaling factor  $\eta$  ranging from 0 to 3, simulating the progression of the enhanced images  $\eta \cdot f(x)$  from dark to bright. This experiment setting simulates how the loss value varies under different brightness levels, facilitating us to investigate the loss curve with respect to the brightness variation only.

Based on the above experimental design, we present the curve of the original  $L_1$  loss (Figure 6(a)), and the curves of the GT-mean  $L_1$  loss under different  $\sigma$  values (Figure 6(b)). The difference between them is that the use of the

GT-mean loss clearly produces basins around  $\eta = 1$ . In another word, the GT-mean loss produces small-gradient region with regard to brightness around  $\eta = 1$ , of which the range is controlled by  $\sigma$ . From an optimization perspective, since the gradients with respect to brightness become smaller, the optimization along the direction of brightness adjustment is in turn slowed down. Based on this characteristic, in real-world model training, the GT-mean  $L_1$  loss enables LLIE models to focus on other important degeneration factors, when  $p(\tilde{\mathbb{E}}[y])$  and  $q(\tilde{\mathbb{E}}[f(x)])$  become closer. In contrast, the original  $L_1$  loss is less capable of decoupling the optimization with respect to brightness and other visual quality factors.

Additionally, Figure 6(c) presents the weight curves that correspond to Figure 6(b), demonstrating how the GT-mean  $L_1$  loss behaves with regard to weight variation. As  $\eta$  approaches 1, the weight  $W$  rapidly decreases, indicating that the second term in Eq.1 begins to dominate the loss function, confirming the mechanism of our loss. Notably, as  $\sigma$  increases,  $W$  starts to drop at smaller values of  $\eta$ , meaning that the second term in Eq.1 takes over earlier in the optimization process. This behavior aligns with the design of  $\sigma$ , which controls the spread of  $\tilde{\mathbb{E}}[\cdot]$ .

## 5. Conclusion

In this paper, we propose the GT-mean loss to advance research on supervised low-light image enhancement (LLIE) methods. The GT-mean loss enables the model training process to circumvent the misleading issue caused by brightness mismatch, thereby comprehensively addressing the various degeneration factors in low-light images. Due to its simple construction, the GT-mean loss can be easily adopted by a wide range of supervised LLIE methods, imposing minimal additional computational overhead during training. Experiments across various supervised LLIE methods consistently demonstrate the effectiveness of the proposed loss. While the estimation of the weight  $W$  remains an open problem, we plan to explore various  $W$ -estimation strategies to potentially unlock even greater performance gains in LLIE models.

## References

- [1] Jonathan T. Barron. A general and adaptive robust loss function. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 4
- [2] W. R. J. Brown and D. L. MacAdam. Visual sensitivities to combined chromaticity and luminance differences. *J. Opt. Soc. Am.*, 39(10):808–834, 1949. 1
- [3] Jens Bühren. *Contrast Sensitivity Function, General*, pages 510–511. Springer Berlin Heidelberg, Berlin, Heidelberg, 2018. 3
- [4] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 4, 6
- [5] Mathys et al. A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, 2011. 3
- [6] Mathys et al. Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience*, 2014. 3
- [7] Huiyuan Fu, Wenkai Zheng, Xiangyu Meng, Xin Wang, Chuanming Wang, and Huadong Ma. You do not need additional priors or regularizers in retinex-based low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 3
- [8] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015. 4
- [9] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017. 4
- [10] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2022. 2
- [11] Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In *European Conference on Computer Vision*, 2024. 3
- [12] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 694–711, 2016. 2, 4
- [13] T. Kailath. The divergence and bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1):52–60, 1967. 4
- [14] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 5148–5157, 2021. 5
- [15] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE Transactions on Image Processing*, 22(12):5372–5384, 2013. 4
- [16] Chongyi Li, Chunle Guo, and Change Loy Chen. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page 1–1, 2021. 3
- [17] C. Li, C. Guo, L. Han, J. Jiang, M. Cheng, J. Gu, and C. Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 44(12):9396–9416, 2022. 1, 2
- [18] Zhixin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8094–8103, 2023. 2
- [19] J. Liu, D. Xu, W. Yang, et al. Benchmarking low-light image enhancement and beyond. *Proceedings of International Journal of Computer Vision (IJCV)*, 129:1153–1184, 2021. 1, 2
- [20] Yucheng Lu and Seung-Won Jung. Progressive joint low-light enhancement and noise removal for raw images. *IEEE Transactions on Image Processing*, 31:2390–2404, 2022. 1
- [21] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, page 3345–3356, 2015. 4
- [22] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 3
- [23] Qianting Ma, Yang Wang, and Tiejiong Zeng. Retinex-based variational framework for low-light image enhancement and denoising. *IEEE Transactions on Multimedia*, 25:5580–5588, 2023. 2, 3
- [24] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, page 209–212, 2013. 5
- [25] Ben Moseley, Valentin Bickel, Ignacio G. Lopez-Francos, and Loveneesh Rana. Extreme low-light environment-driven image denoising over permanently shadowed lunar regions with a physical noise model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1
- [26] J. G. Robson. Spatial and temporal contrast-sensitivity functions of the visual system. *J. Opt. Soc. Am.*, 56(8):1141–1142, 1966. 3
- [27] George Seif and Dimitrios Androutsos. Edge-based loss function for single image super-resolution. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018. 4
- [28] M. Treisman. Noise and weber’s law: the discrimination of brightness and other dimensions. *Psychological Review*, 71:314–330, 1964. 1
- [29] Vassilios Vonikakis, Rigas Kouskouridas, and Antonios Gasteratos. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, page 9211–9231, 2018. 4
- [30] Chenxi Wang, Hongjun Wu, and Zhi Jin. Fourllie: Boosting low-light image enhancement by fourier frequency informa-

- tion. In *Proceedings of the 31st ACM International Conference on Multimedia (ACM MM)*, 2023. 2
- [31] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, page 3538–3548, 2013. 4
- [32] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of Association for the Advancement of Artificial Intelligence (AAAI)*, 2023. 4
- [33] Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. In *IEEE conference on computer vision and pattern recognition (CVPR)*, 2024. 3
- [34] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, page 600–612, 2004. 4
- [35] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 4
- [36] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *Proceedings of British Machine Vision Conference (BMVC)*, 2018. 2, 3, 4
- [37] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1
- [38] Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, Qiong Yan, Xionghuo Min, Guangtao Zhai, and Weisi Lin. Q-align: Teaching LMMs for visual scoring via discrete text-defined levels. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 54015–54029. PMLR, 2024. 1, 4
- [39] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2, 3
- [40] Yuhui Wu, Chen Pan, Guoqing Wang, Yang Yang, Jiwei Wei, Chongyi Li, and Heng Tao Shen. Learning semantic-aware knowledge guidance for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1662–1671, 2023. 2
- [41] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 4
- [42] Qingsen Yan, Yixu Feng, Cheng Zhang, Pei Wang, Peng Wu, Wei Dong, Jinqiu Sun, and Yanning Zhang. Hvi: A new color space for low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 1, 2, 4
- [43] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, page 2072–2086, 2021. 2, 3, 4
- [44] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2020. 4
- [45] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 4
- [46] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018. 1, 4
- [47] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia (ACM MM)*, 2019. 3
- [48] Zhao Zhang, Huan Zheng, Richang Hong, Mingliang Xu, Shuicheng Yan, and Meng Wang. Deep color consistent network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1
- [49] Chu Zhou, Mingguo Teng, Jin Han, Chao Xu, and Boxin Shi. Delieve-net: Deblurring low-light images with light streaks and local events. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2021. 1
- [50] Shangchen Zhou, Chongyi Li, and Chen Change Loy. Led-net: Joint low-light enhancement and deblurring in the dark. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2022. 1