Fast Context-Based Low-Light Image Enhancement via Neural Implicit Representations

Tomáš Chobola[1,2], Yu Liu[1], Hanyi Zhang[1,2], Julia A. Schnabel[1,2,3], and Tingying Peng[1,2]

[1] School of Computation, Information and Technology, Technical University of Munich, Munich, Germany
[2] Helmholtz AI, Helmholtz Munich - German Research Center for Environmental Health, Neuherberg, Germany
[3] School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

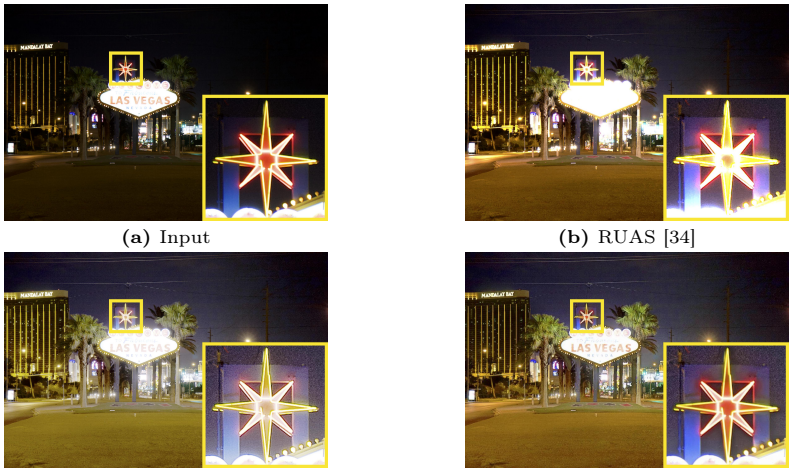
Abstract. Current deep learning-based low-light image enhancement methods often struggle with high-resolution images, and fail to meet the practical demands of visual perception across diverse and unseen scenarios. In this paper, we introduce a novel approach termed *CoLIE*, which redefines the enhancement process through mapping the 2D coordinates of an underexposed image to its illumination component, conditioned on local context. We propose a reconstruction of enhanced-light images within the HSV space utilizing an implicit neural function combined with an embedded guided filter, thereby significantly reducing computational overhead. Moreover, we introduce a single image-based training loss function to enhance the model's adaptability to various scenes, further enhancing its practical applicability. Through rigorous evaluations, we analyze the properties of our proposed framework, demonstrating its superiority in both image quality and scene adaptability. Furthermore, our evaluation extends to applications in downstream tasks within low-light scenarios, underscoring the practical utility of *CoLIE*. The source code is available at https://github.com/ctom2/colie.

Keywords: Low-light image · Illumination estimation · Neural implicit representation

1 Introduction

Images captured under low-light conditions frequently contain substantial dark areas alongside regions that are overly exposed, resulting in the distortion of visual content. This degradation not only makes certain parts of the image imperceivable to human observers but also detrimentally impacts the efficacy of computer vision algorithms [19, 20, 29, 49].

The task of Low-Light Image Enhancement (LLIE) has emerged as a prominent area of focus within the field. Besides early methods such as histogram equalization [42], subsequent approaches based on the Retinex theory [16, 18, 40, 43] and convolutional neural network-based techniques (CNNs) for illumination



(a) Input **(b)** RUAS [34] **(c)** SCI [37] **(d)** Ours

Fig. 1: An example of a challenging real night image from the MIT dataset [2] that illustrates the LLIE task. The intense light emitted by the sign results in significant under-exposure of other areas in the image. Our approach excels in recovering these regions with superior accuracy compared to state-of-the-art (SOTA) methods, which tend to overexpose well-lit areas and distort colors. We emphasize that the compared methods were pretrained on challenging low-light images, whereas our method restores the image in a zero-shot setting without any prior learning.

estimation and deep low-light image representation have garnered considerable attention [36, 48, 51, 54, 55].

Deep learning algorithms mostly aim to achieve image quality through supervised learning to learn direct mapping between **paired** low-light and normal-light images [4, 8, 22, 25, 56], while unsupervised methods rely only on training with low-light images [15, 21, 37]. Nevertheless, convolutional processing is typically constrained by its inductive biases [39]. Fully convolutional networks rely exclusively on local information, which hampers their ability to tackle large dark areas effectively. Therefore, CNN-based methods lack sensitivity in capturing precise representations required for accurate low-light image enhancement.

In this paper, we present a novel approach for enhancing low-light images by leveraging the neural implicit representation (NIR) of illumination, inferred directly from single input images. Unlike conventional methods that aim to learn direct mappings between RGB color spaces, our proposed method encodes spatial coordinates to generate corresponding image representations in HSV space. Inspired by recent advancements in implicit function modelling for tasks such as 3D shape reconstruction [7, 12, 13] and image processing [6, 44, 50], we use a network as an implicit function to map 2D image coordinates to the illumination field conditioned on original image values. The estimated illumination field is then used to enhance the low-light image according to principles of Retinex theory [26]. This innovative approach allows for zero-shot network training, re-

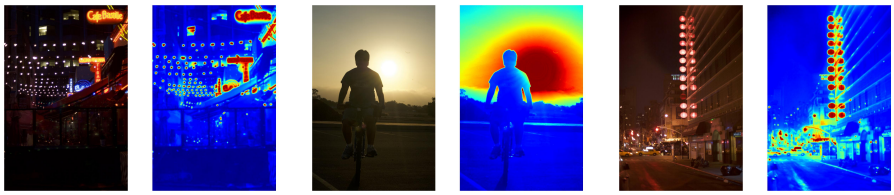
moving the reliance on paired images or specific training sets of low-light images. Consequently, our approach offers enhanced flexibility, circumventing issues related to domain gaps and the need for retraining under varying under-exposure levels. Our method has the potential to significantly advance low-light image enhancement techniques, offering a promising avenue for real-world applications in computer vision. The contributions of this paper are as follows:

- We present a novel formulation of the low-light image enhancement challenge by mapping 2D image coordinates to the illumination component using a network as the neural implicit function, conditioned on the original image intensity values in a zero-shot setting. This novel approach circumvents the need for prior training, thus mitigating domain gap concerns.
- We incorporate guided filtering into our framework, significantly reducing computational overhead for image restoration. This integration ensures that the runtime remains largely unaffected by increases in the input image size, thereby maintaining efficiency without compromising performance.
- We conducted extensive evaluations of our method across three diverse datasets containing natural images as well as microscopy images. The consistent and superior performance across these varied datasets underscores the robustness and versatility of our approach, showcasing its effectiveness in enhancing low-light images in both general and specialized imaging contexts.

2 Related Work

Classic Methods for LLIE. Conventional LLIE methods often rely on the Retinex theory [26], which formulates image formation as a multiplication of two key elements: illumination and reflectance (Equation 1). Some approaches adopt a variational framework to estimate reflectance and illumination, and subsequently refining illumination to enhance the input image [10, 23, 29]. Other methods utilize the Maximum-a-Posteriori (MAP) framework to restore low-light images by imposing priors on reflectance and illumination components [16,18,33], such as total variation or structure-aware regularization. However, the drawback of these methods is their reliance on assumed priors, controlled by numerous parameters. Consequently, this can lead to distorted colors or inadequate balance between heavily over- and under-exposed regions, limiting their practical applicability in real-world scenarios.

Network-based Methods for LLIE. The advancement of LLIE has been significantly driven by the use of CNNs [31], improving the restoration accuracy. While some approaches incorporate the Retinex theory into their models [5, 48, 54, 55], others opt for directly predicting enhanced images from low-light inputs through supervised learning strategies [4, 8, 22, 36]. However, the required paired training data for these supervised approaches are often difficult to obtain in practice [37]. Recent methods have sought to train restoration models exclusively on low-light images to address this limitation. For instance,



(a) (b) (c)

Fig. 2: Instances of inferred illumination components represented by heatmaps based on images within the MIT dataset [2]. These estimated illumination components are subsequently used to improve low-light images according to the principles of the Retinex theory. Leveraging this approach, the enhancement process aims to faithfully reproduce the perceived visual richness and detail even in challenging lighting conditions.

ZeroDCE [15] utilizes a deep network to establish image-specific curve estimations. Liu *et al.* [34] introduced a Retinex-based unrolling framework employing architecture search, while Ma *et al.* [37] proposed a self-calibrating module for image brightening through cascaded illumination learning. Additionally, Fu *et al.* [11] presented an approach leveraging adaptive priors learned from pairs of low-light images to assist the Retinex decomposition. Despite being framed as zero-reference and self-supervised methods, they still depend on prior training on low-light images, making them less stable and challenging to use in real-world scenarios where the level of under-exposure may not align with that of the training data. Furthermore, the CNNs in these methods are highly computationally demanding, especially when dealing with high-resolution images.

Neural Implicit Representations. Although neural implicit representations (NIR) haven been adapted for a variety of image processing tasks, such as super-resolution [6, 28], image warping into continuous shapes [27], tomographic reconstruction [46], their application in low-light image enhancement (LLIE) is relatively unexplored. Closest to our proposed approach is NeRCo introduced by Yang *et al.* [50], which uses a neural representation module to normalize the degradation condition of the input low-light image before enhancing it with a ResNet-based module. Our proposed approach significantly diverges from NeRCo, which uses NIR merely as a pre-processing step, relying on conventional CNNs for the enhancement. In contrast, our method leverages NIR directly to extract the illumination component, utilizing it to enhance the low-light image according to the Retinex theory.

3 Method

3.1 Illumination Learning

The Retinex theory [26] establishes a foundational connection between the visual observation of images captured under low-light conditions, denoted as $\mathbf{y}$,

and their corresponding high-quality representations under ideal lighting conditions, represented as $\mathbf{z}$. This connection is tied to the concept of an illumination component, denoted by $\mathbf{x}$, which indicates local light intensities in the image, thus yielding the observed low-light image as follows,

$$\mathbf{y} = \mathbf{x} \odot \mathbf{z}. \tag{1}$$

Within the domain of low-light image enhancement, the illumination component $\mathbf{x}$ assumes a central role, wielding significant influence over the overall perceived visual quality, and therefore the accurate estimation and manipulation of this illumination component is crucial to effectively enhance low-light images while simultaneously preserving essential details and mitigating the presence of undesirable artifacts.

We introduce a mapping $\mathcal{M}_\theta$ with parameters θ that aims to learn the illumination component to enhance the low-light image as follows,

$$\mathcal{M}_\theta(\mathbf{y}) : \begin{cases} \hat{\mathbf{x}} = \mathbf{y} + f_\theta(\cdot), \\ \hat{\mathbf{z}} = \mathbf{y} \oslash \hat{\mathbf{x}} \end{cases}. \tag{2}$$

This process is inspired by the consensus that the illumination and low-light observation are similar, and there exists a linear connection in most areas. By learning the residual with function $f_\theta(\cdot)$, we reduce the computational difficulty and improve the stability and robustness for exposure control.

Unlike the recent work [15, 34, 37, 50], we deliberately departure from RGB image representations $\mathbf{y}_C \in \mathbb{R}^{H \times W}$ where $C \in \{R, G, B\}$ in favor of the Hue-Saturation-Value (HSV) cylindrical coordinate system, $\mathbf{y}_{\overline{C}} \in \mathbb{R}^{H \times W}$ where $\overline{C} \in \{H, S, V\}$. This choice is motivated by the inherent advantages of HSV in disentangling color and brightness components, and thereby allowing for improved manipulation capabilities within low-light representations. A plausible conjecture within our framework is that the Hue and Saturation components exhibit a degree of similarity between corresponding regions within both the low-light and enhanced images, as seen in Fig. 3. Therefore, our proposed mapping is done directly on the Value component (brightness) of the HSV image representation,

$$\mathcal{M}_\theta(\mathbf{y}_V) : \begin{cases} \hat{\mathbf{x}}_V = \mathbf{y}_V + f_\theta(\cdot), \\ \hat{\mathbf{z}}_V = \mathbf{y}_V \oslash \hat{\mathbf{x}}_V \end{cases}. \tag{3}$$

The culmination of our image enhancement process involves the combination of the original Hue and Saturation components alongside the enhanced Value component, resulting in composite representation $\hat{\mathbf{z}} \in \{\mathbf{y}_H, \mathbf{y}_S, \hat{\mathbf{z}}_V\}$, which integrates the color attributes extracted from the input image and refined brightness.

3.2 Problem Formulation With Implicit Function

We construct a model, represented by f_θ, aimed at extracting the illumination component $\mathbf{x}$ by employing a parameterized MLP to embed the input image.

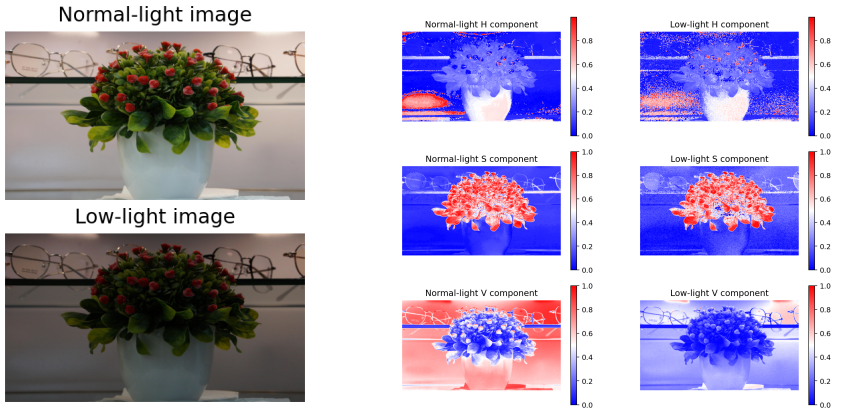


Fig. 3: The motivation behind the design decisions of our method stems from a comparative analysis of the HSV components of normal-light and low-light images from the UHD-LL dataset [30]. This overview shows that, while the Hue and Saturation components display only minimal differences, the primary distinction between the two images lies in the Value component.

Differing from prior approaches that establish mappings between color spaces (e.g., from RGB to RGB), we reframe the problem as a mapping from 2D coordinates to the elements of the Value component, conditioned on the original Value component,

$$f : (p, \mathcal{N}(p)) \rightarrow r, \tag{4}$$

where $p_i = (a_i, b_i)$ are the coordinates of the Value component elements, and r_i is the corresponding pixel intensity. The conditioning is done based on a window of elements of the original Value component with a size $W \times W$ centered around the pixel indexed by p_i, specifically,

$$f_\theta : (p_i, \mathcal{N}(p_i)) \rightarrow r_i. \tag{5}$$

We depart from conventional neural implicit representation functions by introducing a novel incorporation of local context information as input. This is motivated by the intention to provide richer information about the input data, thus enhancing the quality of output representations. By integrating local context alongside the coordinates, our model gains a more comprehensive understanding of the spatial relationships and contextual connections within the scene. This strategy enables our network to capture intricate details and nuanced variations, resulting in more informative and robust output representations.

Similarly to [32], we build the MLP structure with SIREN layers [44], where the pixel coordinates p_i and the context-window values $\mathcal{N}(p_i)$ are separated into two branches and then concatenated into a single output branch. We use 256 channels for the hidden layers and the output of the two input branches is reduced in half before being concatenated and passed to the output layers.

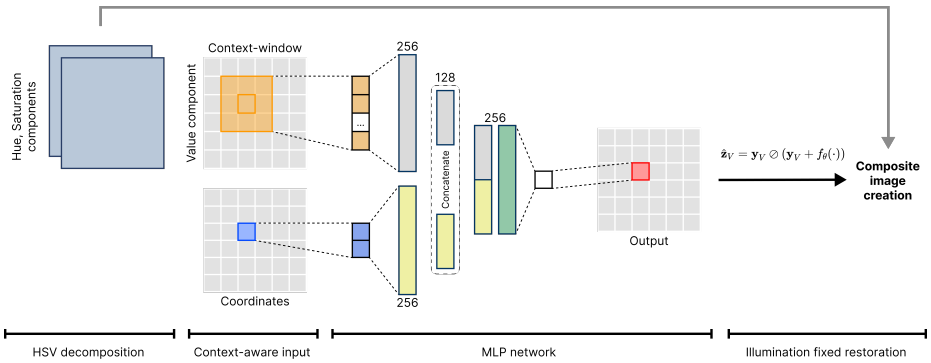


Fig. 4: Our proposed framework begins with the extraction of the Value component from the HSV image representation. Subsequently, we employ a NIR model to infer the illumination component which is an essential part for effective enhancement of the input low-light image. This refined Value component is then reintegrated with the original Hue and Saturation components, forming a comprehensive representation of the enhanced image. The architecture of CoLIE involves dividing the inputs into two distinct parts: the elements of the Value component and the coordinates of the image. Each of these components is subject for regularization with unique parameters within their respective branches. By adopting this structured approach, our framework ensures precise control over the enhancement process.

3.3 Zero-Shot Training

Considering the inaccuracy of existing paired training data and the necessity of model retraining for each image domain and degree of under-exposure, we propose a no-reference learning to mitigate those concerns. We define the total loss as

$$\mathcal{L}_{total} = \alpha\mathcal{L}_f + \beta\mathcal{L}_s + \gamma\mathcal{L}_{exp} + \delta\mathcal{L}_{spa}, \tag{6}$$

where $\mathcal{L}_f, \mathcal{L}_s, \mathcal{L}_{exp}$ and $\mathcal{L}_{spa}$ represent fidelity, smoothness, exposure and sparsity loss, respectively, and $\alpha, \beta, \gamma, \delta$ are loss balancing parameters.

We use the mean squared error (MSE) for the fidelity loss $\mathcal{L}_f$ to maintain pixel-level consistency between the estimated illuminates component and the low-light observation. The smoothness, which is one of the defining factors of the illumination field according to the Retinex theory, and is a broad consensus in this task, is enforced by the total variation (TV) as follows,

$$\mathcal{L}_s = \left(\|\nabla_i \hat{\mathbf{x}}_V\|_2 + \|\nabla_j \hat{\mathbf{x}}_V\|_2\right)^2, \tag{7}$$

where ∇_i and ∇_j represent the vertical and horizontal gradient operations, respectively.

To restrain issues stemming from under- and over-exposed areas, we regulate the intensity levels within the illumination field. We control the distance between the gamma-controlled region and the optimally-intense threshold denoted as L. In our experimental setting, unless explicitly stated otherwise, we set L as 0.5

to achieve most visually pleasing results based on our empirical evidence (see Section 5). The loss function $\mathcal{L}_{exp}$ can be expressed as follows,

$$\mathcal{L}_{exp} = \frac{1}{N} \sum_{k=1}^{N} \left\| \sqrt{\mathcal{T}_k} - L \right\|_2 , \tag{8}$$

where N represents the number of non-overlapping local regions, and $\mathcal{T}$ denotes the average intensity value of a given local region within the illumination field.

In order to preserve the fidelity of low-light areas and mitigate the risk of over-exposure in dim regions, we introduce a sparsity term into the loss function. This sparsity loss penalizes high-intensity values, thus promoting a more balanced illumination across the image.

$$\mathcal{L}_{spa} = \frac{1}{M} \sum_{l=1}^{M} \left| (\hat{\mathbf{z}}_V)_l \right| , \tag{9}$$

where M denotes the total number of pixels in the image. By incorporating the sparsity loss term, we encourage the preservation of dark regions while discouraging the occurrence of excessively bright areas, thereby enhancing the overall fidelity of the reconstructed image.

To address the computational demands inherent in handling (ultra) high-definition images, we process the image in its low-resolution representation $\mathbf{y}^{LR}$ to obtain $\hat{\mathbf{z}}^{LR}$, and then adopt the guided filter [47] to restore the original resolution. This strategic utilization of guided filtering not only mitigates computational costs and scalability challenges but also ensures the preservation of image fidelity throughout the upscaling process.

4 Experiments

Implementation Details. During training, we use the Adam optimizer [24] with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate was set to 10^{-5} to initialize the training for 100 epochs. The regularisation parameters were set as $\alpha = 1, \beta = 20, \gamma = 8, \delta = 5$, and the size of the context window was set to 7×7. Both input branches for the context window and coordinates comprised two layers each, with their outputs passed to two separate output layers for the final prediction. Prior to training, the input images were downscaled to the resolution of 256×256 pixels to enhance processing speed. After training, the outputs were upscaled using the guided filter [47] to restore the original dimensions of the image. To utilize the context window even for pixels located at the border of the image, we employ reflection padding to extend the image.

Compared Methods. We include 10 state-of-the-art (SOTA) methods with different pretrained weights (16 models in total) in our benchmarking study and performance evaluation. The methods cover various approaches, including

supervised methods such as DeepLPF [9], HWMNet [9], RetinexNet [5], Zhao *et al.* [57], unsupervised methods like ZeroDCE [15], EnGAN [21], RUAS [34], SCI [37], PairLIE [11], and one optimization-based method, LIME [16]. The methods have been trained with diverse paired and unpaired low-light images including the MIT [2], LOL [5], LOL-v2 [52], SICE [3], LSRW [17] and DarkFace [53] datasets.

For quantitative benchmarking with two full-reference metrics, SSIM and PSNR, we use the testing set of the UHD-LL dataset [30] and randomly select 200 images from the MIT [2] dataset. For qualitative analysis and downstream task evaluation, we use the DarkFace dataset [53].

Performance Evaluation. We conduct a reference-based quantitative analysis between our approach and the SOTA on the UHD-LL and MIT datasets, as presented in Tables 1 and 2, respectively. Our method demonstrates competitive performance, achieving the highest PSNR score on the UHD-LL dataset. Furthermore, our method surpasses the performance of the compared methods on the MIT dataset, showcasing its effectiveness in enhancing low-light images.

Additionally, we provide a qualitative comparison using images from the MIT and DarkFace datasets in Figures 5 and 6, respectively. Our proposed method exhibits superior performance by preserving the original colors of the image while efficiently enhancing the under-exposed regions. We also include quantitative no-reference evaluation of the enhancement measured by the entropy using the EMEE metric [1] in Table 3 given a random subset of 100 images from the DarkFace dataset.

DarkFace Detection. To assess the effectiveness of our method, we utilized Grounding DINO [35] to detect individuals in images, utilizing the term "person" as the detection prompt. Our evaluation reveals that images enhanced by CoLIE exhibit heightened detection efficacy, resulting in an increased number of correctly detected objects as shown in Fig. 7. This indicates the enhanced performance of the detection model when integrating our method into image pre-processing pipelines.

5 Ablation Study

Context Window. We explore the impact of different context window sizes on the output of the model and show the results in Fig. 8. It is evident that the model with a small window size of only one pixel encounters difficulties in accurately capturing image fidelity, especially in severely under-exposed regions. However, as the window size increases, the model captures more contextual information, resulting in finer, smoother, and more representative results. Moreover, the disparity in illumination strength between well-lit and dim regions becomes more pronounced with larger window sizes.

Table 1: Benchmarking study on the testing set of the UHD-LL dataset. We highlight the best and the second and third best results.

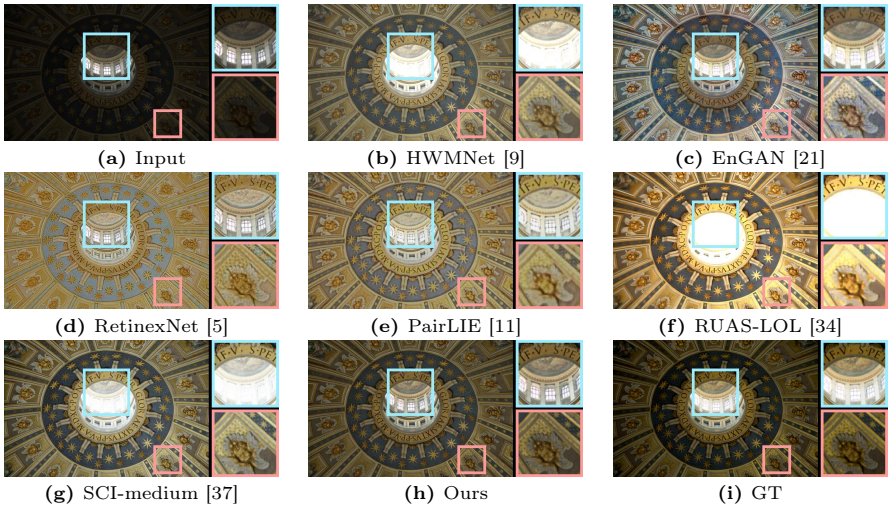
Method	Training Setting	PSNR↑	SSIM↑	Training Set
DeepLPF [38]	Supervised	16.3087	0.5322	MIT-Adobe FiveK
HWMNet [9]	Supervised	14.3063	0.5441	MIT-Adobe FiveK
RetinexNet [5]	Supervised	16.0355	0.5796	LOL
Zhao *et al.* [57]	Supervised	15.1646	0.5465	MIT-Adobe FiveK
ZeroDCE [15]	Unsupervised	12.6484	0.5961	DarkFace
SCI-easy [37]	Unsupervised	15.5433	0.6078	MIT-Adobe FiveK
SCI-medium [37]	Unsupervised	15.4666	0.6218	LOL+LSRW
SCI-difficult [37]	Unsupervised	17.8476	0.5757	DarkFace
RUAS-LOL [34]	Unsupervised	11.6932	0.6970	LOL-v2
RUAS-MIT5K [34]	Unsupervised	14.2146	0.5894	MIT-Adobe FiveK
RUAS-DARK [34]	Unsupervised	11.2424	0.5737	DarkFace
EnGAN [21]	Unsupervised	13.1644	0.5964	Assembled
PairLIE [11]	Unsupervised	17.5349	0.6847	LOL+SICE
LIME [16]	Zero-shot	17.3496	0.5418	–
Ours	Zero-shot	17.8927	0.6252	–

Table 2: Benchmarking evaluation on a subset of the MIT dataset. We highlight the best and the second and third best results.

Method	Training Setting	PSNR↑	SSIM↑	Training Set
HWMNet [9]	Supervised	17.0284	0.8190	LOL
RetinexNet [5]	Supervised	12.0806	0.6653	LOL
ZeroDCE [15]	Unsupervised	15.2448	0.7509	SICE
SCI-medium [37]	Unsupervised	12.2299	0.7385	LOL+LSRW
SCI-difficult [37]	Unsupervised	15.0544	0.7941	DarkFace
RUAS-LOL [34]	Unsupervised	7.3771	0.4903	LOL-v2
RUAS-DARK [34]	Unsupervised	7.0460	0.4787	DarkFace
EnGAN [21]	Unsupervised	12.5576	0.7662	Assembled
PairLIE [11]	Unsupervised	12.8361	0.7677	LOL+SICE
LIME [16]	Zero-shot	14.9763	0.5418	–
Ours	Zero-shot	18.6703	0.8317	–

Exposure Control. The brightness level of the enhanced low-light image is primarily controlled by the parameter L within the exposure term $\mathcal{L}_{exp}$ of the loss function. Lower values of this parameter force the model to refine the illumination field more intensely, resulting in increased brightness and enhanced visibility of objects within the image. However, if the parameter is set too low, it may lead to over-exposure of the image, revealing noisy structures or introducing unnatural visual characteristics. Conversely, maintaining the parameter at high values can diminish the enhancement effect, retaining illumination levels similar to those of the input image. We present the results corresponding to various values of the parameter L in Fig. 9.

Contribution of Each Loss. We present the quantitative results of our proposed approach applied to the MIT dataset based on various combinations of losses in Table 4. Notably, our results demonstrate a substantial impact when the term enforcing sparsity, denoted as $\mathcal{L}_{spa}$, is excluded. This particular term plays a critical role in ensuring that the enhanced images avoid over-exposure, thereby



(a) Input **(b)** HWMNet [9] **(c)** EnGAN [21]
(d) RetinexNet [5] **(e)** PairLIE [11] **(f)** RUAS-LOL [34]
(g) SCI-medium [37] **(h)** Ours **(i)** GT

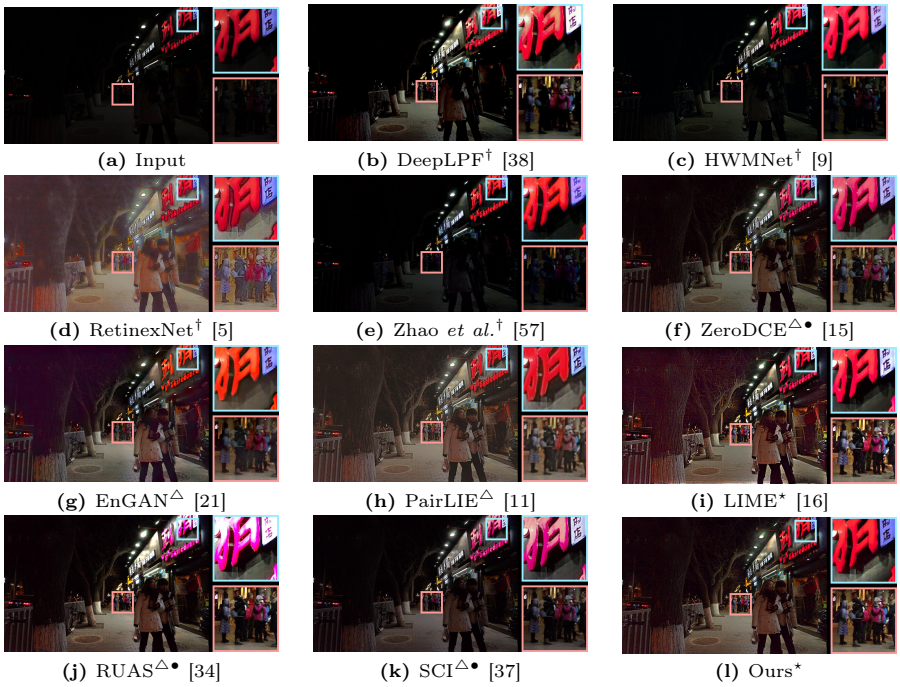
Fig. 5: Visual quality comparison with SOTA methods on a real-world low-light image from the MIT dataset. Our method does not over-expose well-lit regions and preserves the original colors.

Table 3: Benchmarking evaluation on a subset of images from the DarkFace dataset using a no-reference metric.

Method	Training Setting	EMEE↑
Input	–	2.6286
DeepLPF [38]	Supervised	4.0142
HWMNet [9]	Supervised	4.3744
ZeroDCE [15]	Unsupervised	5.5686
SCI-difficult [37]	Unsupervised	6.3856
RUAS-DARK [34]	Unsupervised	6.2555
EnGAN [21]	Unsupervised	2.9794
PairLIE [11]	Unsupervised	1.7985
Ours	Zero-shot	6.7473

preserving high perceptual quality. On the contrary, the influence of the term $\mathcal{L}_{exp}$, which regulates the level of brightness enhancement in the image, is less pronounced in quantitative evaluations. However, its inclusion remains important for maintaining image quality, particularly in low-light scenarios, as depicted in Fig. 9. Lastly, the omission of the illumination smoothness loss $\mathcal{L}_s$ disrupts the correlations among neighboring regions, thereby violating the assumptions of the Retinex theory and consequently leading to a decline in performance.

Microscopy Image Application. Optical microscopy data, especially fluorescence imaging, is often severely affected by shading or vignetting, as detailed by Goldman [14]. This issue is typically manifested as an attenuation of brightness intensity from the center of the optical axis to the edges. This not only degrades the visual quality of the image but, more critically, distorts the quantification of

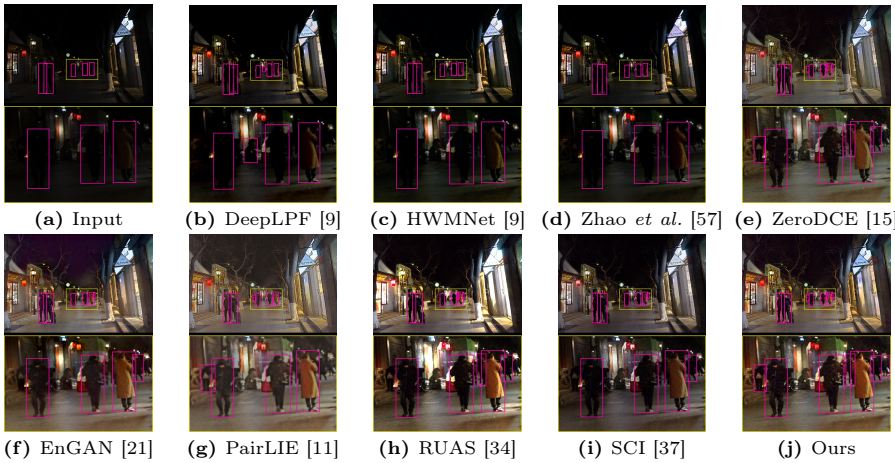


(a) Input **(b)** DeepLPF† [38] **(c)** HWMNet† [9]
(d) RetinexNet† [5] **(e)** Zhao *et al.*† [57] **(f)** ZeroDCE△• [15]
(g) EnGAN△ [21] **(h)** PairLIE△ [11] **(i)** LIME⋆ [16]
(j) RUAS△• [34] **(k)** SCI△• [37] **(l)** Ours⋆

Fig. 6: Visual comparison on a real-world night image from the DarkFace dataset. Labels †, △ and ⋆ denote supervised, unsupervised and zero-shot methods, respectively. Moreover, • indicates the method has been trained directly on images from the DarkFace dataset. Our method achieves the best perceptual quality with more stable improvement of under- and over-exposed regions without any color distortions.

fluorescence intensities that are indicative of biological or molecular properties. While the existing microscopy intensity correction methodologies [41, 45] predominantly hinges on utilizing a series of images with a consistent illumination profile, their applicability to individual images remains limited. In this context, we apply our proposed CoLIE to a singular fluorescence image. Our findings underscore a significant enhancement in image intensity, with a pronounced improvement observable at the peripheries where shading effects are most acute (see Fig. 10).

Limitations & Discussion. While our proposed method exhibits superior performance compared to previous approaches, it is important to note that there remain some open challenges it does not entirely address. Our method primarily targets the enhancement of the Value component within the HSV image representation. However, this focus introduces a vulnerability to noise, as each component of the HSV representation can be affected differently (see Fig. 3). While our approach effectively improves the overall quality of images by manipulating



(a) Input **(b)** DeepLPF [9] **(c)** HWMNet [9] **(d)** Zhao *et al.* [57] **(e)** ZeroDCE [15]

(f) EnGAN [21] **(g)** PairLIE [11] **(h)** RUAS [34] **(i)** SCI [37] **(j)** Ours

Fig. 7: An overview of the downstream detection task given an image from the DarkFace dataset. Our proposed CoLIE stands out as the only method capable of revealing all individuals present in the background, which are subsequently accurately detected.

Table 4: Ablation study of loss variations based on the MIT dataset.

PSNR↑	SSIM↑	$\mathcal{L}_f$	$\mathcal{L}_s$	$\mathcal{L}_{exp}$	$\mathcal{L}_{spa}$
17.3028	0.7891	✓		✓	✓
15.5348	0.8132	✓	✓		✓
7.2470	0.4915	✓	✓	✓	
18.6703	0.8317	✓	✓	✓	✓

the Value component, it does not fully mitigate noise interference originating from variations in the Hue and Saturation components.

6 Conclusion

In this paper, we introduce CoLIE, a novel zero-shot low-light image enhancement model utilizing a neural implicit representation function (NIR). Our approach focuses on restoring the illumination component in accordance with the Retinex theory, aiming to enhance image quality and perceptual fidelity. Particularly, we integrate information from both image coordinates and input pixel values, beyond conventional NIR methods. Experimental results demonstrate superior performance on real-world night and under-exposed images, supported by both quantitative metrics and qualitative evaluations.

Acknowledgements

Tomáš Chobola is supported by the Helmholtz Association under the joint research school "Munich School for Data Science - MUDS".



(a) Input **(b)** 1×1 **(c)** 3×3 **(d)** 5×5 **(e)** 7×7

Fig. 8: Instances of predicted illumination components based on the size of the context window used for prediction (based on an image from the DarkFace dataset). Larger contextual windows output predictions of higher fidelity, particularly in areas with severe under-exposure.



(a) Input **(b)** $L = 0.9$ **(c)** $L = 0.7$ **(d)** $L = 0.5$ **(e)** $L = 0.3$

Fig. 9: Overview of the effect of the parameter L in the exposure loss term on the brightness of the output image (based on an image from the DarkFace dataset).

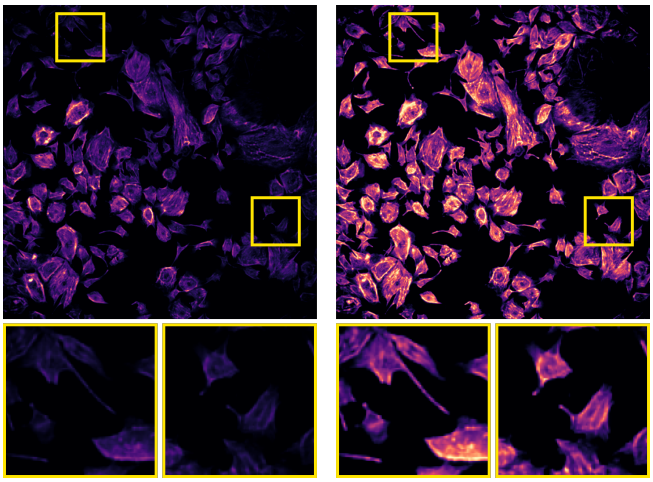


Fig. 10: Visual example of a fluorescence microscopy image (left) and the enhancement results with CoLIE (right). The restoration process rectifies illumination irregularities and effectively eliminates vignetting, which are typical characteristics of fluorescence microscopy imaging.

References

1. Badizadegan, K., Wheeler, H.E., Fujinaga, Y., Lencer, W.I.: Trafficking of cholera toxin-ganglioside gm1 complex into golgi and induction of toxicity depend on actin cytoskeleton. American Journal of Physiology-Cell Physiology **287**(5), C1453–C1462 (2004). https://doi.org/10.1152/ajpcell.00189.2004, pMID: 15294854
2. Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input / output image pairs. In: The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition (2011)
3. Cai, J., Gu, S., Zhang, L.: Learning a deep single image contrast enhancer from multi-exposure images. IEEE Transactions on Image Processing **27**(4), 2049–2062 (2018)
4. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3291–3300 (2018)
5. Chen, W., Wenjing, W., Wenhan, Y., Jiaying, L.: Deep retinex decomposition for low-light enhancement. In: British Machine Vision Conference (2018)
6. Chen, Y., Liu, S., Wang, X.: Learning continuous image representation with local implicit image function. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8628–8638 (2021)
7. Deng, B., Lewis, J.P., Jeruzalski, T., Pons-Moll, G., Hinton, G., Norouzi, M., Tagliasacchi, A.: Nasa neural articulated shape approximation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16. pp. 612–628. Springer (2020)
8. Dong, X., Xu, W., Miao, Z., Ma, L., Zhang, C., Yang, J., Jin, Z., Teoh, A.B.J., Shen, J.: Abandoning the bayer-filter to see in the dark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17431–17440 (2022)
9. Fan, C.M., Liu, T.J., Liu, K.H.: Half wavelet attention on m-net+ for low-light image enhancement. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 3878–3882. IEEE (2022)
10. Fu, X., Zeng, D., Huang, Y., Zhang, X.P., Ding, X.: A weighted variational model for simultaneous reflectance and illumination estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2782–2790 (2016)
11. 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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22252–22261 (2023)
12. Genova, K., Cole, F., Sud, A., Sarna, A., Funkhouser, T.: Local deep implicit functions for 3d shape. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4857–4866 (2020)
13. Genova, K., Cole, F., Vlasic, D., Sarna, A., Freeman, W.T., Funkhouser, T.: Learning shape templates with structured implicit functions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7154–7164 (2019)
14. Goldman, D.B.: Vignette and exposure calibration and compensation. IEEE transactions on pattern analysis and machine intelligence **32**(12), 2276–2288 (2010)
15. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1780–1789 (2020)

16. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on image processing **26**(2), 982–993 (2016)
17. Hai, J., Xuan, Z., Yang, R., Hao, Y., Zou, F., Lin, F., Han, S.: R2rnet: Low-light image enhancement via real-low to real-normal network. Journal of Visual Communication and Image Representation **90**, 103712 (2023)
18. Hao, S., Han, X., Guo, Y., Xu, X., Wang, M.: Low-light image enhancement with semi-decoupled decomposition. IEEE transactions on multimedia **22**(12), 3025–3038 (2020)
19. Huang, Y., Zha, Z.J., Fu, X., Hong, R., Li, L.: Real-world person re-identification via degradation invariance learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14084–14094 (2020)
20. Jia, M., Xu, J., Yang, R., Li, Z., Zhang, L., Wu, Y.: Three filters for the enhancement of the images acquired from fluorescence microscope and weak-light-sources and the image compression. Heliyon **9**(9) (2023)
21. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. IEEE transactions on image processing **30**, 2340–2349 (2021)
22. Kim, H., Choi, S.M., Kim, C.S., Koh, Y.J.: Representative color transform for image enhancement. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4459–4468 (2021)
23. Kimmel, R., Elad, M., Shaked, D., Keshet, R., Sobel, I.: A variational framework for retinex. International Journal of computer vision **52**, 7–23 (2003)
24. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
25. Koh, J., Lee, J., Yoon, S.: Bnudc: a two-branched deep neural network for restoring images from under-display cameras. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1950–1959 (2022)
26. Land, E.H., McCann, J.J.: Lightness and retinex theory. Josa **61**(1), 1–11 (1971)
27. Lee, J., Choi, K.P., Jin, K.H.: Learning local implicit fourier representation for image warping. In: European Conference on Computer Vision. pp. 182–200. Springer (2022)
28. Lee, J., Jin, K.H.: Local texture estimator for implicit representation function. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1929–1938 (2022)
29. Li, C., Qu, X., Gnanasambandam, A., Elgendy, O.A., Ma, J., Chan, S.H.: Photon-limited object detection using non-local feature matching and knowledge distillation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3976–3987 (2021)
30. Li, C., Guo, C.L., Zhou, M., Liang, Z., Zhou, S., Feng, R., Loy, C.C.: Embedding fourier for ultra-high-definition low-light image enhancement. In: ICLR (2023)
31. Li, C., Guo, C., Han, L., Jiang, J., Cheng, M.M., Gu, J., Loy, C.C.: Low-light image and video enhancement using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence **44**(12), 9396–9416 (2021)
32. Li, L., Qiao, H., Ye, Q., Yang, Q.: Metadata-based raw reconstruction via implicit neural functions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18196–18205 (2023)
33. 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)

34. Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10561–10570 (2021)
35. Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Li, C., Yang, J., Su, H., Zhu, J., et al.: Grounding dino: Marrying dino with grounded pre-training for open-set object detection. arXiv preprint arXiv:2303.05499 (2023)
36. Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition **61**, 650–662 (2017)
37. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward fast, flexible, and robust low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5637–5646 (2022)
38. Moran, S., Marza, P., McDonagh, S., Parisot, S., Slabaugh, G.: Deeplpf: Deep local parametric filters for image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
39. Naseer, M.M., Ranasinghe, K., Khan, S.H., Hayat, M., Shahbaz Khan, F., Yang, M.H.: Intriguing properties of vision transformers. Advances in Neural Information Processing Systems **34**, 23296–23308 (2021)
40. Ng, M.K., Wang, W.: A total variation model for retinex. SIAM Journal on Imaging Sciences **4**(1), 345–365 (2011)
41. Peng, T., Thorn, K., Schroeder, T., Wang, L., Theis, F.J., Marr, C., Navab, N.: A basic tool for background and shading correction of optical microscopy images. Nature Communications **8**(1) (Jun 2017). https://doi.org/10.1038/ncomms14836, http://dx.doi.org/10.1038/ncomms14836
42. Pizer, S.M.: Contrast-limited adaptive histogram equalization: Speed and effectiveness stephen m. pizer, r. eugene johnston, james p. ericksen, bonnie c. yankaskas, keith e. muller medical image display research group. In: Proceedings of the first conference on visualization in biomedical computing, Atlanta, Georgia. vol. 337, p. 2 (1990)
43. Ren, X., Yang, W., Cheng, W.H., Liu, J.: Lr3m: Robust low-light enhancement via low-rank regularized retinex model. IEEE Transactions on Image Processing **29**, 5862–5876 (2020)
44. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in neural information processing systems **33**, 7462–7473 (2020)
45. Smith, K., Li, Y., Piccinini, F., Csucs, G., Balazs, C., Bevilacqua, A., Horvath, P.: Cidre: an illumination-correction method for optical microscopy. Nature Methods **12**(5), 404–406 (Mar 2015). https://doi.org/10.1038/nmeth.3323, http://dx.doi.org/10.1038/nmeth.3323
46. Sun, Y., Liu, J., Xie, M., Wohlberg, B.E., Kamilov, U.S.: Coil: Coordinate-based internal learning for imaging inverse problems. IEEE Transactions on Computational Imaging **7** (11 2021). https://doi.org/10.1109/TCI.2021.3125564, https://www.osti.gov/biblio/1883143
47. Wu, H., Zheng, S., Zhang, J., Huang, K.: Fast end-to-end trainable guided filter. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1838–1847 (2018)
48. Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., Jiang, J.: 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. pp. 5901–5910 (2022)

49. Xu, X., Wang, S., Wang, Z., Zhang, X., Hu, R.: Exploring image enhancement for salient object detection in low light images. ACM transactions on multimedia computing, communications, and applications (TOMM) **17**(1s), 1–19 (2021)
50. Yang, S., Ding, M., Wu, Y., Li, Z., Zhang, J.: Implicit neural representation for cooperative low-light image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12918–12927 (2023)
51. Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3063–3072 (2020)
52. Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
53. Yang, W., Yuan, Y., Ren, W., Liu, J., Scheirer, W.J., Wang, Z., Zhang, T., Zhong, Q., Xie, D., Pu, S., et al.: Advancing image understanding in poor visibility environments: A collective benchmark study. IEEE Transactions on Image Processing **29**, 5737–5752 (2020)
54. Zhang, Y., Guo, X., Ma, J., Liu, W., Zhang, J.: Beyond brightening low-light images. International Journal of Computer Vision **129**, 1013–1037 (2021)
55. 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)
56. Zhang, Z., Zheng, H., Hong, R., Xu, M., Yan, S., Wang, M.: Deep color consistent network for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1899–1908 (2022)
57. Zhao, L., Lu, S.P., Chen, T., Yang, Z., Shamir, A.: Deep symmetric network for underexposed image enhancement with recurrent attentional learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 12075–12084 (October 2021)