International Journal of Computer Vision
https://doi.org/10.1007/s11263-024-01995-y

PIE: Physics-Inspired Low-Light Enhancement

Dong Liang[1,2] · **Zhengyan Xu**[1] · **Ling Li**[1] · **Mingqiang Wei**[1] · **Songcan Chen**[1]

Received: 1 August 2023 / Accepted: 2 January 2024

Abstract
In this paper, we propose a physics-inspired contrastive learning paradigm for low-light enhancement, called PIE. PIE primarily addresses three issues: (i) To resolve the problem of existing learning-based methods often training a LLE model with strict pixel-correspondence image pairs, we eliminate the need for pixel-correspondence paired training data and instead train with unpaired images. (ii) To address the disregard for negative samples and the inadequacy of their generation in existing methods, we incorporate physics-inspired contrastive learning for LLE and design the Bag of Curves (BoC) method to generate more reasonable negative samples that closely adhere to the underlying physical imaging principle. (iii) To overcome the reliance on semantic ground truths in existing methods, we propose an unsupervised regional segmentation module, ensuring regional brightness consistency while eliminating the dependency on semantic ground truths. Overall, the proposed PIE can effectively learn from unpaired positive/negative samples and smoothly realize non-semantic regional enhancement, which is clearly different from existing LLE efforts. Besides the novel architecture of PIE, we explore the gain of PIE on downstream tasks such as semantic segmentation and face detection. Training on readily available open data and extensive experiments demonstrate that our method surpasses the state-of-the-art LLE models over six independent cross-scenes datasets. PIE runs fast with reasonable GFLOPs in test time, making it easy to use on mobile devices. Code available

Keywords Low-light enhancement · Physics-inspired contrastive learning · Super-pixel segmentation

1 Introduction

Capturing images under low illumination remains a significant source of errors in camera imaging, further leading to image details lost, color under-saturation, low-contrast/low dynamic range, and uneven exposure. Such degeneration severely hinders downstream vision tasks, e.g., semantic segmentation (Wang et al., 2022; Cho et al., 2020; Liang et al., 2021) and object detection (Wu et al., 2022b; Al Sobbahi & Tekli, 2022; Liang et al., 2014; Geng et al., 2021), from operating smoothly in vision-based systems. Existing methods formulate low-light enhancement (LLE) as a mapping problem with three main challenges.

First, the existing learning-based methods in the low-level domain often train a model with strict pixel-correspondence image pairs via strong supervisions (Lore et al., 2017; Wei et al., 2018; Zhang et al., 2019; Xu et al., 2020; Ren et al., 2019; Ignatov et al., 2017; Zhou et al., 2023). However, high-quality pixel-correspondence image pairs are challenging to acquire in practice. For example, Ignatov et al. (2017) proposed to acquire them from a DSLR camera to refine the imaging of a mobile phone camera. It brings complicated registration and pixel-by-pixel calibration to image pairs.

Second, to release pixel-correspondence image pairs, some works (Huang et al., 2023; Shi et al., 2022) have introduced contrastive learning for LLE, which adopt the normal-

Communicated by Chongyi Li.

✉ Dong Liang
liangdong@nuaa.edu.cn

Zhengyan Xu
xuzhengyan@nuaa.edu.cn

Ling Li
liling@nuaa.edu.cn

Mingqiang Wei
mingqiang.wei@gmail.com

Songcan Chen
s.chen@nuaa.edu.cn

[1] MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Shenzhen, China

[2] Nanjing University of Aeronautics and Astronautics Shenzhen Research Institute, Nanjing, China

Published online: 25 April 2024

(a) Input (b) w/o Neg. samples (c) SCL-LLE (Liang et al, 2022) (d) PIE

Fig. 1 Impact of training data. The proposed PIE (**d**), which generates negative samples using physical laws closer to realistic imaging, produces enhanced results with better brightness, color, contrast, and naturalness under extremely dark conditions than SCL-LLE (**c**) and SCL-LLE without any negative samples (**b**). More specifically, the first sample in (**d**) has a higher dynamic range and better subjective feeling. In the second sample in (**d**), the saturation of the girl's T-shirt and the sunset surrounding is much higher and presents a better global stereoscopic atmosphere of the scene. The comparison between (**c**) and (**d**) illustrates the necessity of introducing negative samples. In contrast to (**c**), the improvement in image quality in (**d**) reflects the crucial role of negative sample quality in contrastive learning

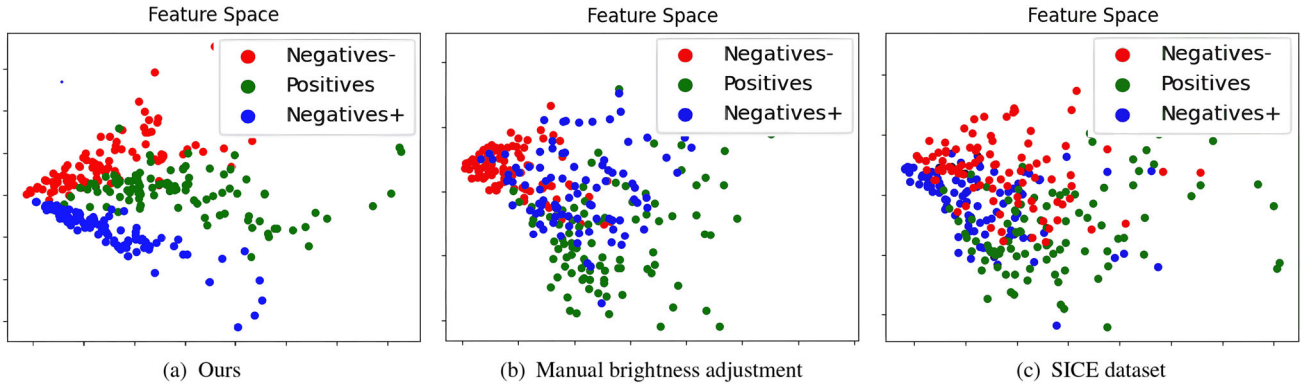
light images as positive samples and the over/underexposed images as negative samples to guide the training. As shown in Fig. 1b, the selection of negative samples in contrastive learning significantly impacts the results of LLE. The quality of negative samples and the specific contrastive learning strategy would be more significant for LLE, if we wish to release the pixel-correspondence image pairs while maintaining consistent performance with the LLE model with the strict image pairs. Therefore, another concern of this work is which negative samples to choose for what kind of contrastive learning, to provide diverse and representative negative samples, squeezing and filling the feature space, enabling the learned LLE model to provide a visual experience closer to the underlying imaging principles. Most existing methods mainly rely on directly using underexposed/overexposed images from existing low-light datasets (Huang et al., 2023) or artificially adjusting image brightness based on empirical experience (Liang et al., 2022) to obtain negative samples. However, due to limitations in the quality of the dataset itself and the constraints of human expertise, the boundaries between negative and positive samples generated by these two methods, as shown in Fig. 2b and c, may not be significant.

Third, the enhancement strategies for the background and foreground should be different. In our previous work (Liang et al., 2022), we utilize semantic information to differentiate the enhanced regions and maintain the consistency of brightness within the same semantic category. Wu et al. (2023) also employs semantic information to maintain consistent brightness for each semantic. However, the introduction of semantic segmentation destroys the universality and flexibility of the method, as semantic segmentation is a full-supervision training setting with massive pixel-level annotation.

In order to effectively learn from unpaired positive/negative samples and smoothly realize non-semantic regional enhancement with underlying imaging principles, we propose physics-inspired contrastive learning for low-light image enhancement (PIE). In contrastive learning, we design the Bag of Curves (BoC) solution by leveraging the Image Signal Processing (ISP) pipeline (i.e., the Gamma correction and Tone mapping) to destroy positive samples but follow the basic imaging rules to generate negative samples. This method generates under/overexposed samples in a way that is more closely aligned with the underlying physical imaging principles. At the same time, the design of the regional segmentation module maintains regional brightness consistency, realizes region-discriminate enhancement, and releases from semantic labels. PIE casts the image enhancement task as a multi-task joint learning problem, where LLE is converted into three constraints—contrastive learning, regional brightness consistency, and feature preservation, simultaneously ensuring the quality of global/local exposure, texture, and color. We also pay more attention to downstream tasks (i.e., semantic segmentation, and face detection) to explore if we can realize performance gain from our LLE scheme. We find that our method potentially benefits the downstream tasks under dark conditions.

The contributions are three folds:

- A physics-inspired contrastive learning approach for real-world cross-scene LLE, without any paired training images and pixel-level annotation:



(a) Ours (b) Manual brightness adjustment (c) SICE dataset

Fig. 2 Feature Visualization of generating negative samples using different methods. Compared to the method of artificially adjusting brightness (**b**) and low-light dataset (**c**), our method (**a**) exhibits a clear boundary between positive and negative samples. The samples in (**b**) are derived from positive and negative samples in SCL-LLE (Liang et al., 2022), while the images in (**c**) are sourced from the SICE (Cai et al., 2018) dataset

(1) a physics-inspired approach called "Bag of Curves" generates negative samples for contrastive learning using principles closer to the underlying physical imaging mechanism.
(2) an unsupervised regional segmentation module to maintain regional brightness consistency, realize region-discriminate enhancement, and release from semantic labels.
(3) a multi-task joint learning with three constraints—contrastive learning, regional brightness consistency, and feature preservation, simultaneously ensuring exposure, texture, and color consistency.

- PIE is compared with SOTAs via comprehensive experiments on six independent datasets in terms of visual quality, no and full-referenced image quality assessment, and human subjective survey. All results consistently endorse the superiority and efficiency of the proposed approach.
- We demonstrate that our PIE is friendly to downstream high-level vision tasks and easy to joint-learn with them.

This work is partially presented in our earlier conference version (Liang et al., 2022). We have introduced many new findings and improvements compared to the conference version. We have two new core contributions. First, we design a "Bag of Curves" solution inspired by the physical imaging principle to generate negative samples to replace the manual process. Second, we design an unsupervised regional segmentation module to maintain regional brightness consistency to replace the supervised semantic segmentation module. We also modify our contrastive learning loss for better performance, present more extensive experiments related to the aforementioned improvements compared with more recent methods, and provide more discussion with downstream tasks.

2 Related Work

2.1 Low-Light Image Enhancement

Conventional Methods LLE has been actively studied as an image-processing problem for a long. Early efforts are commonly made towards the use of handcrafted priors with empirical observations (Pizer et al., 1990; Land, 1977; Xu et al., 2014; Guo et al., 2016) to deal with the LLE problem. Histogram equalization (Pizer et al., 1990) used a cumulative distribution function to regularize the image's pixel values and evenly distribute overall intensity levels. However, this kind of operation naturally makes it easy to cause over/under-exposure. Without local adaptation, the enhancement results in intensive noise and undesirable illumination. Later methods constrained the equalization process with several kinds of priors, e.g. mean intensity preservation (Ibrahim & Kong, 2007), noise robustness, and white and black stretching (Arici et al., 2009), to improve the overall visual quality of the adjusted image. Retinex model (Land, 1977) and its multi-scale version (Jobson et al., 1997) decomposed the brightness into illumination and reflectance and then processed them separately. Wang et al. (2013) constructed a brightness filter for Retinex decomposition and tried to preserve the naturalness while enhancing details in low-light images. The reflectance component is commonly assumed to be consistent under lighting conditions; thus, light enhancement is formulated as an illumination estimation problem. The gray-scale transformation (Xu et al., 2014) is a method based on the spatial domain, which enhanced the image by modifying the distribution and dynamic range of the gray-scale value of the pixels. Guo et al. (2016) introduced a structural prior to refining the initial illumination map and finally synthesized the enhanced image according to the Retinex theory. However, these handcrafted constraints/priors are not self-adaptive to

recover image details and color. This results in washing out details, local under/over-saturation, uneven exposure, or halo artifacts.

Data-Driven Methods In the past decade, data-driven methods (Li et al., 2021a) have achieved significant advancements in the field of low-light image enhancement. Lore et al. (2017) proposed a variant stacked sparse denoising autoencoder to enhance the degraded images. RetinexNet (Wei et al., 2018) leveraged a deep architecture based on Retinex to enhance low-light images. Zhang et al. (2019) developed three subnetworks for layer decomposition, reflectance restoration, and illumination adjustment based on Retinex. RUAS (Liu et al., 2021) constructed the overall LLE network architecture by unfolding its optimization process. The above methods are trained based on image pairs with strict pixel correspondence. Zhou et al. (2022) introduced LEDNet, a powerful network designed for simultaneous low-light enhancement and deblurring tasks. Jiang et al. (2021) reported an unsupervised method using normal-light images that do not have low-light images as correspondences. Zero-DCE (Guo et al., 2020), FlexiCurve (Li et al., 2023a), CuDi (Li et al., 2022) and ReLLIE (Zhang et al., 2021a) reformulated the LLE task as an image-specific curve estimation problem with a fixed default brightness value. Fan et al. (2020) used semantic information to guide the reconstruction of the reflection of Retinex. DNF (Jin et al., 2023) is a decouple and feedback framework for the RAW-based LLIE. CLIP-LIT (Liang et al., 2023) introduced an initial prompt pair, enforcing text prompt and backlit image similarity using CLIP latent space. SCL-LLE (Liang et al., 2022) introduced semantic information to the brightness reconstruction and paid more attention to the dependency among the semantic elements via the interaction of high-level semantic knowledge and low-level signal priors. The proposed PIE maintains the brightness consistency of image regions without relying on semantic ground truths, which is clearly different from existing LLE efforts.

2.2 Contrastive Learning for Vision Tasks

Contrastive learning (He et al., 2020; Chen et al., 2020; Sermanet et al., 2018; Tian et al., 2020; Henaff, 2020) is from the self-supervised learning paradigm, which is characterized by using pretext tasks to mine its supervisory information from original data for downstream tasks. For a given input, contrastive learning aims to pull it together with the positives and push it apart from the negatives in feature space. Previous works have applied contrastive learning to high-level vision tasks because these tasks are inherently suited for modeling the contrast (He et al., 2020; Chen et al., 2020; Tian et al., 2020) between positive and negative samples. It has also been applied to low-level visual tasks, such as deraining (Chen et al., 2022), underwater image enhancement (Han et al., 2021), and dehazing (Wu et al., 2021). Huang et al. (2023) and Shi et al. (2022) introduced a contrastive learning module for low-light enhancement. Huang et al. (2023) employed contrastive learning to train a two-stream encoder for feature extraction. Shi et al. (2022) used contrastive learning techniques to train the SFE model for extracting structure maps. However, these methods overlooked the importance of the way to select positive and negative samples in contrastive learning. In addition, most of the existing contrastive learning methods rely heavily on a large number of negative samples and thus require either large batches or memory banks (Li et al., 2021b). In our approach, we employ only a couple of negative samples for one positive sample and introduce a random mapping strategy to avoid the risk of overfitting.

2.3 Gamma Correction and Tone Mapping

The Image Signal Processing (ISP) pipeline is used in modern digital cameras to convert raw camera sensor data into high-quality, human-readable RGB images. ISP (Karaimer & Brown, 2016) consists of several operations, including image denoising, noise reduction, white balance, color space conversion, Gamma correction, and Tone mapping.

Gamma correction (Farid, 2001; Yuan & Sun, 2012) is a standard step of the image processing pipeline to adjust the brightness of an image for display on different devices. It transformed the pixel value following a non-linear power-law function. Tone mapping (Mantiuk et al., 2008) refers to the process of converting high dynamic range (HDR) images to low dynamic range (LDR) images. HDR images in the RAW domain have higher color depth and dynamic range than LDR images, allowing them to better represent the brightness and color details in a scene, but they cannot be fully displayed on conventional RGB displays. Therefore, Tone mapping is needed for HDR images to present as much HDR image information as possible on LDR displays. Zhang et al. (2023) combined Tone mapping with GAN to adjust the brightness of images. Drago et al. (2003) and Yongqing (2013) respectively used two different Tone mapping methods based on different curves to adjust the brightness of the image. Inspired by Gamma correction and Tone mapping, we propose a physics-inspired contrastive learning method and introduce "Bag of Curves" to generate negatives for contrastive learning.

3 The Proposed PIE

3.1 Problem Formulation and Architecture

Fundamentally, low-light image enhancement can be regarded as seeking a mapping function F, such that $I_H = F(I_L)$ is the desired image, which is enhanced from the input image

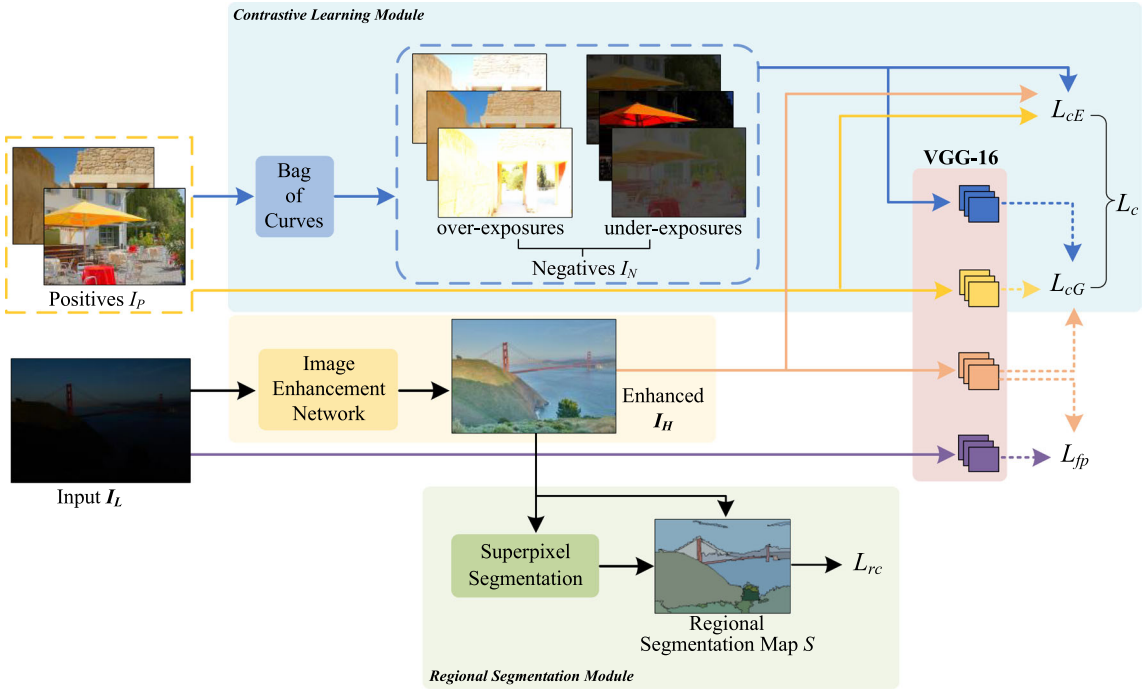


Fig. 3 The overall architecture of PIE. It includes a low-light image enhancement (LLE) network, a contrastive learning module (the blue block) boosted by Bag of Curves (BoC), a regional segmentation module (the green block), and a VGG-16 feature extractor (the red block). PIE jointly minimizes the contrastive learning loss L_c, which consists of two components, L_{cE} and L_{cG}, feature preserving loss L_{fp}, and regional brightness consistency loss L_{rc}

I_L. In our design, we introduce a prior of contrastive learning: the contrastive samples, including the negatives I_N, i.e., the under/overexposed images which are generated by our proposed Bag of Curves solution, and the positives I_P, i.e., the normal-light images. Therefore, we formulate a new mapping function as follows:

$$I_H = F(I_L, I_N, I_P) \tag{1}$$

As depicted in Fig. 3, PIE consists of a low-light image enhancement network, a contrastive learning module, and a regional segmentation module. Specifically, our approach comprises a low-light image enhancement network, which leverages a U-Net-like backbone (Guo et al., 2020) to generate a pixel correction curve that remaps each pixel. We use VGG-16 (Simonyan & Zisserman, 2014) as the feature extraction network. Specifically, for a given image I_L, it is first input to the image enhancement network. Then, the enhanced image I_H is fed into the regional segmentation module, ensuring brightness consistency within each region. For the contrastive learning module, images enhanced by image enhancement network I_H serve as the anchor for contrastive learning, images under normal lighting I_P serve as positive samples, and negative samples are over/underexposed images I_N obtained from images under normal lighting through Bag of Curves. To optimize our approach, we employ three types of losses corresponding to the framework's three key aspects: the contrastive learning loss L_c, feature preserving loss L_{fp}, and regional brightness consistency loss L_{rc}.

In PIE, we solve the challenges of manually dividing positive and negative samples in contrastive learning as on (Liang et al., 2022) and eliminate the dependency on semantic ground truths. Specifically, we first propose a "Bag of Curves" method that combines the physical imaging principle with contrastive learning to generate negative samples, which aids in compressing the feature space and enabling the model to effectively adjust the distance between the anchor and positive/negative samples. Additionally, we introduce an unsupervised regional segmentation module that maintains regional brightness consistency while removing the reliance on semantic ground truths.

3.2 Contrastive Learning Module

3.2.1 Bag of Curves

Choosing appropriate negative samples is crucial for the success of contrastive learning, as it enables the model to learn sample representations that capture the unique characteristics of the data. For an LLE task, the construction of negative

samples in contrastive learning should better follow the physical laws of imaging as closely as possible and mimic both overexposure and underexposure. Our previous work (Liang et al., 2022) has achieved non-paired contrastive learning, but the positive and negative samples used are still provided by the manual process—manually adjusting the brightness of images to generate a set of under/overexposed negatives.

We leverage Tone mapping and Gamma correction in ISP for brightness adjustment and utilize this prior knowledge to adjust the brightness of images to generate negative samples that are more consistent with physical imaging laws. We follow the following principles when choosing curves: (1) It should be able to effectively adjust the overall brightness of the image in a reasonable manner. (2) The form of the curve should be as simple as possible for ease of implementation and computation. (3) Priority is given to commonly used curves in existing methods. Taking into account these reasons, we choose Gamma, Logarithmic, and Sigmoid curves to simulate the inverse Gamma correction and Tone mapping process for adjusting the brightness of the image.

Due to the different brightness ranges captured by the human eyes and digital cameras, the brightness and color captured by cameras (when viewed on a standard monitor) look different from what the human eye perceives. When rendering high dynamic range (HDR) images, the brightness values can exceed the maximum value that a monitor can show. Therefore, we need to adjust the brightness range of the image to convert HDR images to low dynamic range (LDR) that can be appropriately displayed on a monitor. The process of adjusting the brightness of an image is commonly referred to as Tone mapping. After Tone mapping, Gamma correction is usually performed to account for humans' non-linear perception of natural brightness and to adapt to the monitor's display characteristics, ultimately outputting the corresponding brightness to the display. Tone mapping (Mantiuk et al., 2008) and Gamma correction (Farid, 2001) are commonly used on physical devices such as cameras and monitors to adjust the brightness of images, resulting in a photo effect that is more similar to human perception. Both Tone mapping and Gamma correction are intended to improve the display of images on LDR devices by transforming the range of brightness values from one distribution to another. Inspired by the above observation, and also inspired by Bag of Words (BoW) (Sivic & Zisserman, 2003) in feature engineering, we propose Bag of Curves (BoC) to generate negative samples for contrastive learning. Specifically, we leverage standard curves in Tone mapping and Gamma correction to realize reversed gone mapping and reversed Gamma correction. This way maps the brightness of $\{I_p\}$ to a certain range to generate over/underexposed images as negative samples. This range could destroy the original brightness of $\{I_p\}$ but could follow the physical imaging principle.

$$I_N \in BoC = \{C_g, C_s, C_l\} \tag{2}$$

In BoC, we select three curves, the Gamma curve C_g from Gamma correction, the Sigmoid curve C_s, and the logarithmic curve C_l in Tone mapping to generate over/underexposed images as negative samples for contrastive learning, which are directly and parallel functioned with the positives I_P. The curves representation of BoC are shown in Fig. 4.

Gamma Curve The Gamma curve C_g is defined as follows:

$$\{C_g\} = \{I_p^{\gamma}\} \tag{3}$$

A Gamma value $\gamma < 1$ is sometimes called an encoding Gamma, and the process of encoding with this compressive power-law nonlinearity is called Gamma compression; conversely, a Gamma value $\gamma > 1$ is called a decoding Gamma, and the application of the expansive power-law nonlinearity is called Gamma expansion. We set $\gamma = 0.2$ to generate overexposed images and $\gamma = 8$ to generate underexposed images. **Sigmoid curve** The Sigmoid curve C_s is a nonlinear function curve. Yongqing (2013) succeeded in expanding the local dynamic range in dark and bright areas by dodging and burning with the Sigmoid curve operator. The formula for the Sigmoid curve C_s is as follows:

$$\{C_s\} = \left\{ \frac{1}{1 + e^{10 \cdot (c - I_p)}} \right\} \tag{4}$$

where c represents the offset of the Sigmoid curve. By adjusting the values of c, the shape of the Sigmoid function can be controlled to achieve different brightness adjustment effects. For underexposure, the value of c can be set to a value smaller than the brightness of I_p, and 0.3 is selected in our setting. Similarly, for overexposure, the value of c is 0.8.

Logarithmic Curve The Logarithmic curve C_l is approaching HVS's perception of brightness (Drago et al., 2003). The formula for the Logarithmic curve C_l is as follows:

$$\{C_l\} = \{m \cdot \log_2 (1 + I_P)\} \tag{5}$$

The smaller the value of constant m, the weaker the effect of Logarithmic transformation and the lower the brightness of the image. We set $m = 10$ and $m = 40$ to generate under/overexposed negatives.

As shown in Fig. 5, the three curves can adjust the brightness of the image to guide its brightness distribution on the highlight or low-light side of the brightness histogram in quite different forms, which could be effectively negative samples for contrastive feature learning.

In contrastive learning, it is necessary to ensure a clear boundary between negative samples and positive samples in the feature space. Negative samples should also be as clustered as possible in the feature space. Therefore, generating

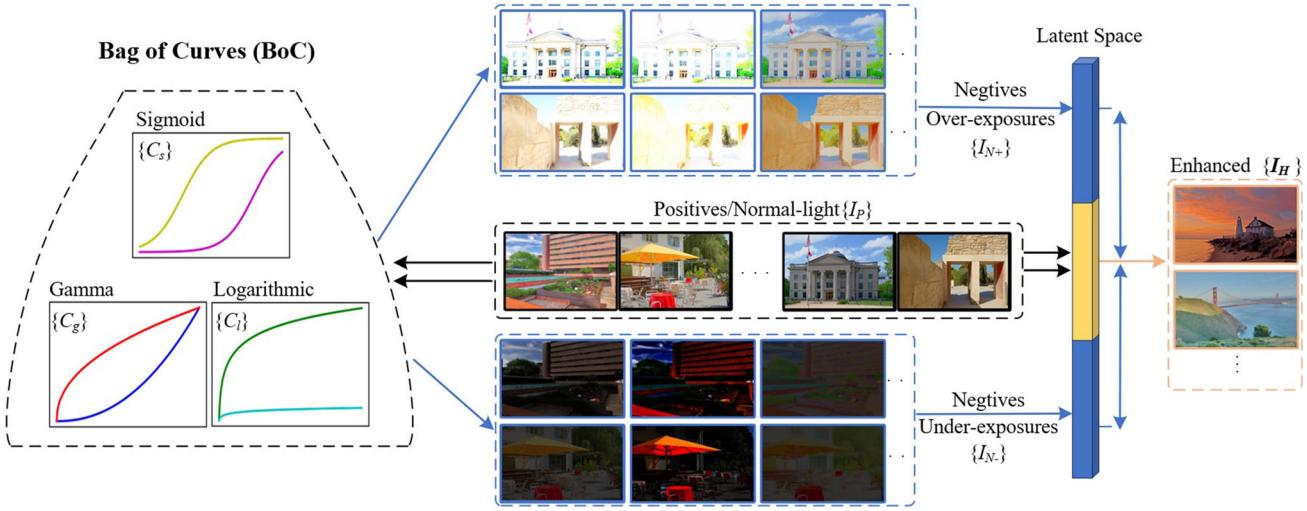


Fig. 4 Bag of Curves. There are three different groups of curves: Gamma, Sigmoid, and Logarithmic curves enable the model to learn diverse and representative characteristics of the produced negative samples

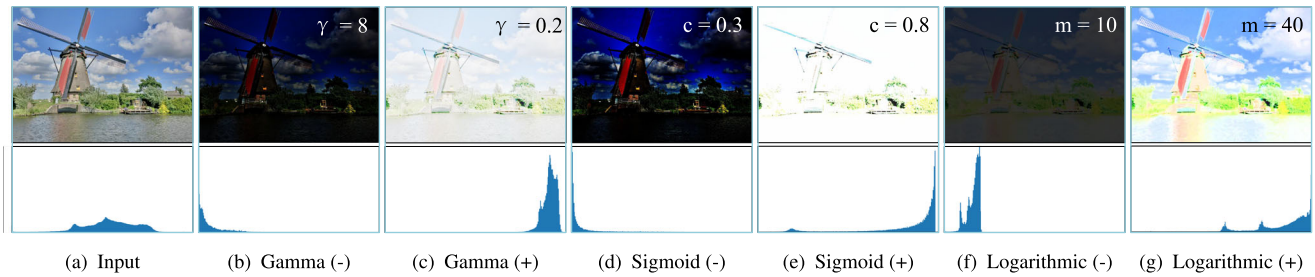


Fig. 5 BoC samples and their histogram from normal lighting sample (positives) to negatives, using Gamma, Sigmoid, and Logarithmic curves, respectively. These three curves effectively make the brightness of the normal illumination image distributed in the under (−)/overexposed (+) areas of the histograms. The (−)/(+) samples have quite different histograms and appearances but follow the physical imaging laws, making the negative samples effective for contrastive learning

representative negative samples is considered crucial. We achieve this by using fixed values to generate underexposed or overexposed negative samples that represent the entire sample category.

From Fig. 6, it can be observed that when the parameter values (e.g., γ, c, m) is fixed, the boundary between the normal illumination image (Positives) and the underexposed negative sample (Negatives-) and overexposed negative sample (Negatives+) is more distinct. This results in higher discrimination in the feature space and effective compression of the feature space. However, when the parameter value falls within a certain range, some negative samples may overlap with the normal illumination image in the feature space, causing interference in feature learning.

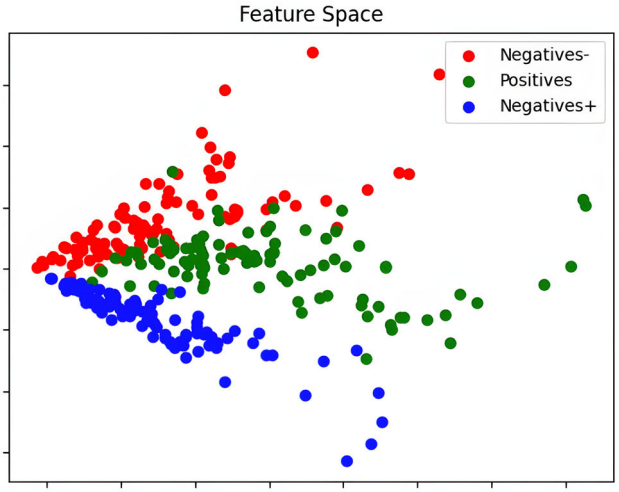
3.2.2 Contrastive Learning Detail and Its Loss

We use the images I_H enhanced by a low-light image enhancement network as anchors for contrastive learning. For negative samples, we adjust the brightness of normal-light images using BoC and transform them into over/underexposed images I_N. Positive samples are the normal-light images I_P. The positive and negative samples do not pair with each other or the anchor image, i.e., from different scenes.

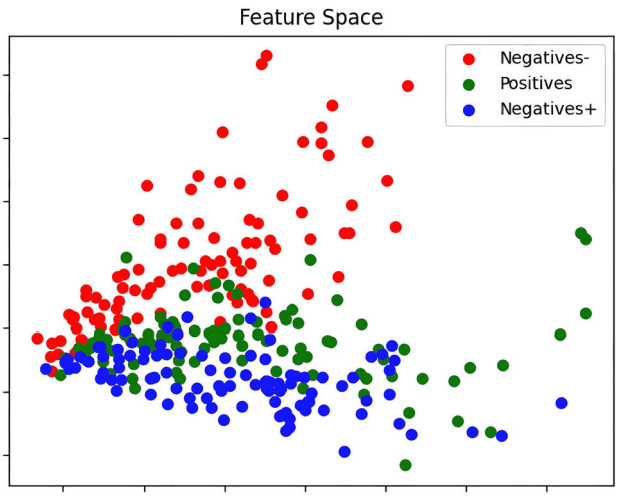
Feature Extraction Network We incorporate a pre-trained VGG-16 model to extract the feature map $f \in \mathbb{R}^{C\times H\times W}$ for the latent feature space, where G^l_{ij} is the inner product between the feature maps i and j in the layer l:

$$G^l_{ij} = \sum_k f^l_{ik} f^l_{jk} \tag{6}$$

where k represents the vector length. We then get a set of Gram matrices $\{G^1, G^2, \ldots, G^L\}$ from layers 1, ..., L in the feature extraction network. The Gram matrix G is a quantitative description of latent image features. Contrastive learning aims to learn a feature space in which samples of the same



(a) BoC with fixed parameters



(b) BoC with parameters within a certain range

Fig. 6 Visualization of features. In BoC, the three curves included in the paper have fixed parameter values in (**a**), while in (**b**), they randomly fall within a certain range. In **a**, the boundary between positive and negative samples is more pronounced, indicating a clearer separation between the two classes

category should be closer to each other while samples of different categories should be farther away.

Contrastive Loss A reasonable contrastive loss is necessary to pull the anchors into the positive samples and push away the anchors from the negative samples in the latent space. Triplet loss (Hermans et al., 2017), N-pair loss (Sohn, 2016), and InfoNCE loss (Gutmann & Hyvärinen, 2010) are commonly used loss functions in contrastive learning. The choice of which loss function to use depends on the specific task and dataset we discuss in Sect. 4.2.3. In PIE, we mix the triple loss and infoNCE loss in contrastive learning to design the contrastive learning loss.We utilize triplet loss for the Gram matrix G, aiming to:

$$d(G_{I_H}, G_{I_P}) \ll d(G_{I_H}, G_{I_N}) \tag{7}$$

where d represents the distance between features. Unlike Gram matrix G, we use infoNCE loss for the expectation value E, our goal is:

$$d(E_{I_H}, E_{I_P}) \ll d(E_{I_H}, E_{I_N}) \tag{8}$$

We wish that the distance d between features I_H and I_P is smaller than the distance between features I_H and I_N. Based on the aforementioned objectives, we designed L_{cG} and L_{cE} as two components of contrastive learning loss L_c.

For Gram matrix G:

$$L_{cG} = max\left\{d(G_{I_H}, G_{I_P}) - d(G_{I_H}, G_{I_N}) + \alpha, 0\right\} \tag{9}$$

α is a hyperparameter, and we set it to 0.3.

For the expectation E:

$$L_{cE} = -log\frac{exp(d(E_{I_H}, E_{I_P}))}{exp(d(E_{I_H}, E_{I_P})) + exp(d(E_{I_H}, E_{I_N}))} \tag{10}$$

Therefore, the contrastive loss function in PIE is expressed as follows:

$$L_c = L_{cG} + L_{cE} \tag{11}$$

The Numbers of Positive and Negative Samples In a theoretical work (Li et al., 2021b), the author argued that a 1:1 rate of positive to negative samples is sufficient for triplet loss. The author also observed significant benefits in contrastive learning of visual representations from randomness. Inspired by this work, our method involves using one underexposed and one overexposed sample as negatives for each scene. Positive and negative samples are randomly selected during each iteration of training to enhance the model's robustness. Our positive and negative samples are obtained from the SICE dataset (Cai et al., 2018), which consists of 589 scenes (360 scenes in Part1 and 229 scenes in Part2) with a total of 4413 multi-exposure images. In our method, all 360 standard images in all scenes of Part1 are used as positive samples, while negative samples are generated by applying BoC to the standard images to produce under/overexposed images.

In Sect. 4.2.3, we investigate the impact of different rates of positive and negative samples (1:1, 1:5, 5:1, and 5:5 in a batch) on low-light enhancement results. Additionally, we consider the average training time for each epoch, which involves training on all samples in the training set once.

(a) The enhanced images $\{I_H\}$

(b) Graph-based pixel-level segmentation S

Fig. 7 Demonstration of the enhanced images $\{I_H\}$ and segmentation results after Graph-based pixel-level segmentation S

3.3 Regional Segmentation Module

3.3.1 Unsupervised Super-Pixel Segmentation

In real-world scenes, it is expected that the same region of an object should have uniform brightness, while the enhancement strategies applied to the background and foreground should be different. To address this issue, Liang et al. (2022) incorporates a semantic segmentation module to prevent local overexposure or underexposure. However, the use of a semantic segmentation module resulted in the model's dependence on semantic ground truth. PIE introduces an unsupervised regional segmentation module that uses a super-pixel segmentation to maintain regional brightness consistency and enable region-discriminate enhancement while avoiding reliance on semantic labels. For this purpose, we employ a Graph-based supervised super-pixel segmentation method (Felzenszwalb & Huttenlocher, 2004) as illustrated in Fig. 7. We first use super-pixel segmentation to divide an image into super-pixel blocks. Then, we utilize the regional brightness consistency loss L_{rc} to maintain the consistency of brightness within each region.

The output is a segmentation component $S = \{c_1, c_2...\}$. During the Graph-based super-pixel segmentation process, it compares the inter-domain difference $Dif\left(c_i, c_j\right)$ between two different regions c_i and c_j, with the minimum intra-domain difference $Mint\left(c_i, c_j\right)$ between the smallest regions, c_i and c_j, within these two segmentation regions. If the difference between components is larger than the minimum internal difference, it indicates that there is a boundary between these two regions; otherwise, these two regions are merged while other regions remain unchanged. The judgment method is as follows:

$$\mathrm{D}\left(c_i, c_j\right) = \begin{cases} \text{True} & Dif\left(c_i, c_j\right) > Mint\left(c_i, c_j\right) \\ \text{False} & \text{otherwise} \end{cases} \tag{12}$$

3.3.2 Regional Brightness Consistency Loss

The application of super-pixel segmentation frees our method from dependence on semantic ground truths information. We define an average value B of the brightness level of the overall pixels in each super-pixel block $c \in S$ as follows:

$$B_c = \frac{1}{n} \sum_{i \in \theta_c} (B^i_{I_H}) \tag{13}$$

where c represents the cth super-pixel block, and we can attain multiple averages representing individual super-pixel block separately $\{B_1, B_2, \ldots\}$. n represents the number of pixels in this super-pixel block. We denote θ_c as the pixel index collection belonging to block c, $B^i_{I_H}$ as the brightness level in the enhanced image I_H at the block c. The regional brightness consistency loss L_{rc} is defined as:

$$L_{rc} = \sum_{c=1}^{C} \sum_{i \in \theta_c} (B^i_{I_H} - B_c)^2 \tag{14}$$

where C is the number of the super-pixel blocks.

3.4 Other Details

3.4.1 Feature Preservation Loss

Many low-level visual tasks (Ledig et al., 2017; Kupyn et al., 2018; Johnson et al., 2016) use the perceptual loss to make desired images and their features and ground truth perceptually consistent. We also leverage perceptual loss as our feature retention loss to preserve the image features before and after enhancement. The feature retention loss L_{fr} is defined as:

$$L_{fr} = \frac{1}{C_l W_l H_l} (f^l(I_L) - f^l(I_H))^2 \tag{15}$$

where $f^l(I_L)$ denotes the feature map $f \in \mathbb{R}^{C\times H\times W}$ of the input image I_L in the layer l, and $f^l(L_H)$ is the feature map of the enhanced image I_H in the layer l.

Since the color naturalness is one of the significant concerns of LLE, we add a color constancy term L_{cc} incorporating with the feature retention term, following the way reported in Guo et al. (2020). Based on the gray-world color constancy hypothesis (Buchsbaum, 1980), the pixel averages of the three channels tend to be the same value. L_{cc} constrains the ratio of three channels to prevent potential color deviations in the enhanced image. In addition, to avoid aggressive and sharp changes between neighboring pixels, an illumination smoothness penalty term is also embedded in L_{cc}. The formulation of L_{cc} can be expressed as:

$$\begin{aligned} L_{cc} &= \sum_{\forall(p,q)\in\xi} (J^p - J^q)^2 \\ &\quad + \lambda \frac{1}{M} \sum_{m=1}^{M} \sum_{p\in\xi} \left(\left|\nabla_x A_m^p\right| + \left|\nabla_y A_m^p\right| \right), \\ \xi &= \{R, G, B\} \end{aligned} \tag{16}$$

where J^p denotes the average intensity value of p channel in the enhanced image, (p, q) represents a pair of channels, M is the number of the iterations, and ∇_x and ∇_y denote the horizontal and vertical gradient operations, respectively. A is a parameter map with the same size as the image. Each pixel has a corresponding higher-order curve parameter generated in multiple iterations. A_m^p denotes the parameter map of channel p in mth iteration. We set λ to 200 in our experiments for the best outcome.

The feature preservation loss L_{fp} is the sum of L_{fr} and L_{cc}.

3.4.2 Efficient Training Details

In our implementation, the feature extraction network is pre-trained on ImageNet (Russakovsky et al., 2015), the CBDNet is pre-trained on BSD500 (Martin et al., 2001), Waterloo (Ma et al., 2017), MIT-Adobe FiveK (Bychkovsky et al., 2011) and RENOIR dataset (Anaya & Barbu, 2018). We train PIE end-to-end while fixing the weights of the feature extraction network. The back-propagated operation only updates the weights in the image enhancement network. Hence, most network computation is done in the image enhancement network, which efficiently learns I_H from (I_L, I_N, I_P) to recover the enhanced image with various scenes. We resize the training images to the size of 384 × 384. As for the numerical parameters, we set the maximum epoch as 10 and the batch size as 2. Our network is implemented with PyTroch on an NVIDIA 1080Ti GPU. The Adam optimizer optimizes the model with a fixed learning rate $1e^{-4}$.

3.4.3 Downstream Task-Driven LLE

We aim to explore whether PIE can benefit downstream tasks. We evaluated LLE on three tasks: semantic segmentation, and face detection.

Semantic Segmentation Our earlier work (Liang et al., 2022) with a semantic brightness consistency loss has demonstrated the effectiveness of LLE in improving downstream semantic segmentation. In this study, to validate the gain of PIE on semantic segmentation, we replace the regional segmentation module in PIE with the same semantic segmentation module used in Liang et al. (2022). Additionally, we replace the regional brightness consistency loss L_{rc} with the semantic brightness consistency loss L_{sc}. The semantic segmentation network we use here is the popular DeepLabv3+ (Chen et al., 2018), and we train our network on the training images of the Cityscapes (Cordts et al., 2016) dataset.

Face Detection For face detection, we replace the regional segmentation module in the PIE with the RetinaFace (Deng et al., 2020) trained on the WIDER FACE dataset (Yang et al., 2016) and replace the regional brightness consistency loss L_{rc} with a face detection loss L_{det}. The face detection loss L_{det} includes two components: L_{cls} and L_{box}. L_{cls} is the softmax loss for binary classification (face/not face), while L_{box} is the face box regression loss which is based ond (Girshick, 2015). The PIE network for face detection, called PIE_{det}, is fine-tuned using 4000 images from the DARK FACE dataset (Yang et al., 2020). During training, the parameters of the RetinaFace are fixed, and the RetinaFace is introduced only to calculate the detection loss to guide the optimization of the low-light enhancement model.

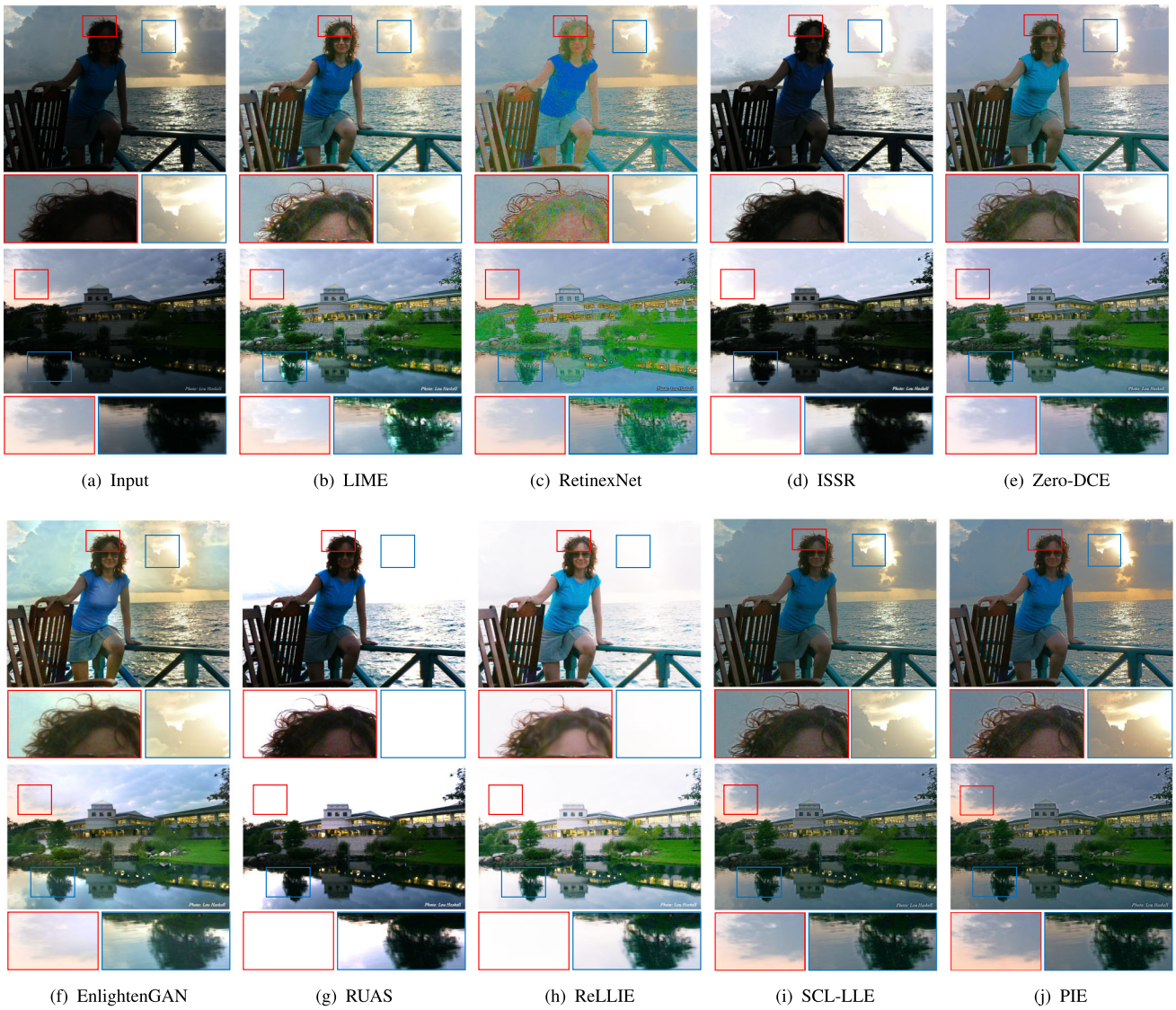
More details regarding the downstream task-driven LLE will be presented in the following experiments outlined in Sect. 4.3.

4 Experiments

4.1 Cross-Dataset Peer Comparison

For testing images, we use six publicly available low-light image datasets from other reported works, i.e., DICM (Lee et al., 2012), MEF (Ma et al., 2015), LIME (Guo et al., 2016), NPE (Wang et al., 2013), VV[1] and the Part2 of SICE (Cai et al., 2018). DICM, LIME, MEF, VV, NPE, and SICE are ad hoc test datasets, including 64, 10, 17, 24, 8, and 229 images, respectively. They are widely used in LLE testing: SCL-LLE (Liang et al., 2022), EnlightenGAN (Jiang et al., 2021), Zero-DCE (Guo et al., 2020). Images in these datasets are diverse and representative: DICM is mainly landscaped

[1] https://sites.google.com/site/vonikakis/datasets.



(a) Input (b) LIME (c) RetinexNet (d) ISSR (e) Zero-DCE

(f) EnlightenGAN (g) RUAS (h) ReLLIE (i) SCL-LLE (j) PIE

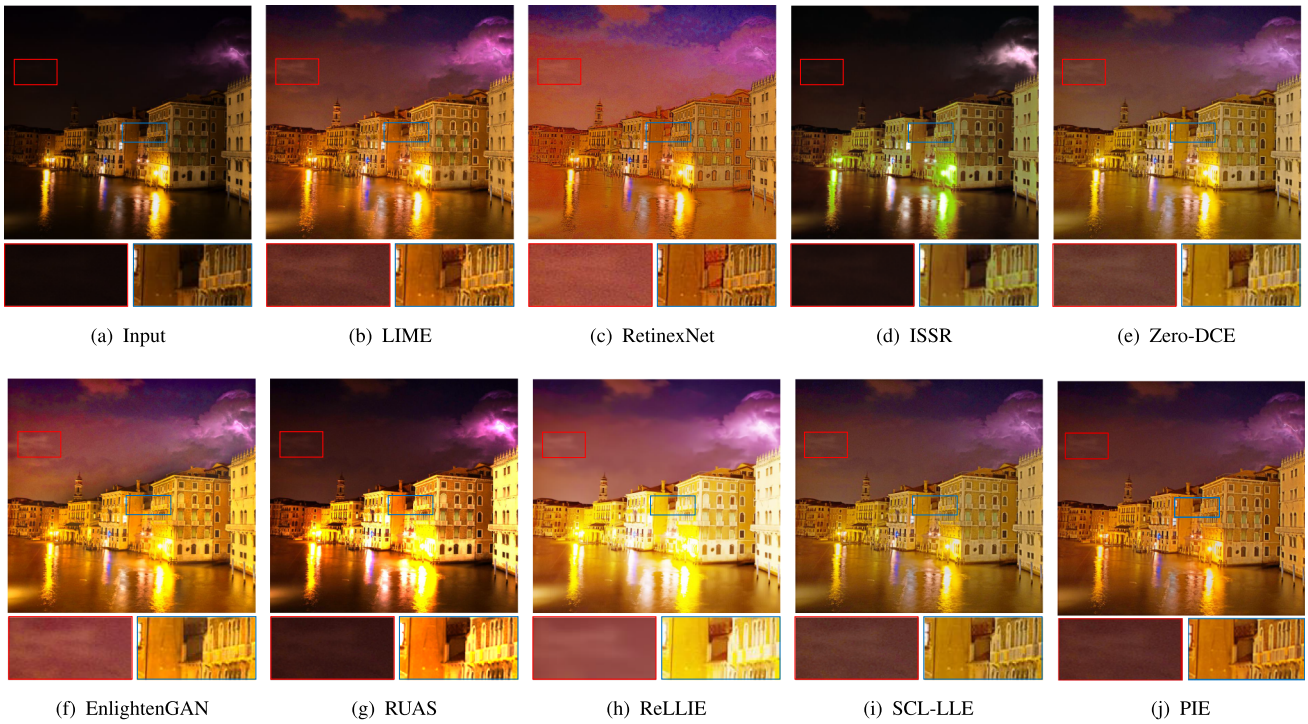
Fig. 8 Demonstration of PIE and the state-of-the-art methods over VV (the first sample) and DICM (the second sample) datasets with zoom-in regions. PIE enables the enhanced images to look more realistic and recovers better details and richer color in both foreground and background

with extreme darkness; LIME focuses on dark street landscapes; MEF focuses on dark indoor scenes and buildings; VV is mostly backlit and portraits; NPE mainly includes natural scenery in low light; SICE is a large-scale multi-exposure image dataset that contains high-resolution image sequences of multiple scenes. Note that all the images in the six datasets are independent cross-scene images without any overlapped scene of the input image and the positive/negative samples.

We compare our proposed method with 13 representative state-of-the-art methods for heterogeneous image enhancement. These included the conventional method LIME (Guo et al., 2016), the GAN-based method EnlightenGAN (Jiang et al., 2021), and four Retinex-based methods: RetinexNet (Wei et al., 2018), RUAS (Liu et al., 2021), ISSR (Fan et al., 2020), and Zero-DCE (Guo et al., 2020), which leverages the same backbone enhancement network as our proposed method. We further include two reinforcement learning based method ReLLIE (Zhang et al., 2021a) and ALL-E (Li et al., 2023b), four more recent methods Uretinex-net (Wu et al., 2022a), SCI (Ma et al., 2022), PairLLE (Fu et al., 2023), and IRN (Zhao et al., 2021), and our earlier conference version (Liang et al., 2022). We reproduce the results of these methods using recommended test settings and publicly available models.

4.1.1 Visual Quality Comparison

We first examine whether the proposed methods can achieve visually pleasing results in brightness, color, contrast, and naturalness. We observe from Figs. 8 and 9 that all the SOTAs



(a) Input (b) LIME (c) RetinexNet (d) ISSR (e) Zero-DCE

(f) EnlightenGAN (g) RUAS (h) ReLLIE (i) SCL-LLE (j) PIE

Fig. 9 Demonstration of PIE and the state-of-the-art methods with a night-view sample in LIME dataset with zoom-in regions. PIE enables enhanced images with more realistic color and details in both foreground and background and retouching with a negligible noise level on the dark background (the sky) (Color figure online)

sacrifice over/under/uneven exposure in global or local areas. Specifically, LIME (Guo et al., 2016) leads to color artifacts in strong local edges (e.g., hair and sky, and inverted reflection in the water); RetinexNet (Wei et al., 2018) and EnlightenGAN (Jiang et al., 2021) cause global color distortions with details missing; ISSR (Fan et al., 2020) and RUAS (Liu et al., 2021) generate severe global and local over/underexposure; ReLLIE (Zhang et al., 2021a) suffers from over-enhancement and over-smoothing. In contrast, PIE recovers more details and better contrast in both foreground and background, thus enabling the enhanced images to look more realistic with vivid and natural color mapping.

4.1.2 No-Referenced IQA

For testing using no-referenced image quality assessment (IQA), we adopt Natural Image Quality Evaluator (NIQE) (Mittal et al., 2013), a well-known no-reference image quality assessment for evaluating image restoration without ground truth and providing quantitative comparisons. Since some work criticizes that NIQE correlates poorly with subjective human opinion, we also adopt UNIQUE (Zhang et al., 2021b) for No-referenced IQA. Smaller NIQE and larger UNIQUE indicate more naturalistic and perceptually favored quality. The NIQE and UNIQUE results on five datasets (DICM, LIME, MEF, VV, and NPE) are reported in Table 1. Compared with other state-of-the-art methods, PIE achieves the best results for the NIQE in two of the five datasets, achieves the best results for the UNIQUE in four of the five datasets, and the average results on these five datasets are the best.

4.1.3 Full-Referenced IQA

For full-reference image quality assessment, we utilize the Peak Signal-to-Noise Ratio (PSNR, dB) and Structural Similarity (SSIM) metrics to compare the performance of various methods quantitatively. PSNR is commonly used in low-level vision tasks, and its value is always non-negative. A higher PSNR value indicates better quality. On the other hand, SSIM measures image similarity based on image brightness, contrast, and structure. Since the five datasets used in the previous test do not contain standard images, we use Part2 of the SICE dataset (Cai et al., 2018) without overlapping the training data. PIE demonstrates excellent performance on both PSNR and SSIM metrics, achieving the best performance on the SSIM metric and the second-highest PSNR score, only behind Uretinex-net (Wu et al., 2022a), as shown in Table 2.

4.1.4 Human Subjective Survey

We conduct a human subjective survey (user study) for comparisons. For each image in the five test datasets (DICM,

Table 1 NIQE ↓, UNIQUE (UN.) ↑ and User Study (U.S.) ↓ scores on DICM, LIME, MEF, VV, and NPE datasets

Methods	DICM			LIME			MEF			VV			NPE			Average		
	NIQE ↓	UN. ↑	U.S. ↓	NIQE ↓	UN. ↑	U.S. ↓	NIQE ↓	UN. ↑	U.S. ↓	NIQE ↓	UN. ↑	U.S. ↓	NIQE ↓	UN. ↑	U.S. ↓	NIQE ↓	UN. ↑	U.S. ↓
Input	4.26	0.72	3.33	4.36	0.70	4.30	4.26	0.72	4.41	3.52	**0.74**	3.38	4.32	1.17	3.92	4.13	0.75	3.67
LIME (Guo et al., 2016)	3.75	0.78	3.44	3.85	0.53	2.10	3.65	0.65	3.82	**2.54**	0.44	2.75	4.44	0.93	3.75	3.55	0.69	3.40
Retinex-Net (Wei et al., 2018)	4.47	0.75	3.59	4.60	0.52	4.00	4.41	0.97	4.06	2.70	0.36	2.88	4.60	0.81	4.13	4.13	0.69	3.75
ISSR (Fan et al., 2020)	4.14	0.59	3.13	4.17	**0.83**	3.40	4.22	0.87	4.47	3.57	0.62	3.00	4.02	0.99	3.96	4.03	0.68	3.49
Zero-DCE (Guo et al., 2020)	3.56	0.82	2.77	3.77	0.73	2.10	3.28	1.22	3.18	3.21	0.48	2.50	3.93	1.07	2.50	3.50	0.81	2.70
EnlightenGAN (Jiang et al., 2021)	3.55	0.63	2.81	**3.70**	0.49	**2.00**	**3.16**	1.03	3.29	3.25	0.58	2.12	3.95	1.07	2.85	3.47	0.69	2.72
RUAS (Liu et al., 2021)	5.21	−0.17	3.44	4.26	0.34	2.30	3.83	0.73	4.11	4.29	−0.04	3.75	5.53	0.13	4.17	4.78	0.04	3.60
ReLLIE (Zhang et al., 2021a)	4.44	0.41	2.94	5.22	0.52	3.40	5.22	1.07	4.05	3.51	0.33	2.50	5.14	0.37	3.71	4.48	0.49	3.25
Uretinex-net (Wu et al., 2022a)	3.95	0.85	2.99	4.34	0.93	2.12	3.79	1.18	3.16	3.01	0.51	2.63	4.69	0.99	2.55	3.83	0.84	2.71
SCI (Ma et al., 2022)	4.11	0.11	3.46	4.21	0.35	2.83	3.63	1.04	3.95	2.92	0.05	2.98	4.47	0.21	3.92	3.87	0.35	3.72
IRN (Zhao et al., 2021)	3.68	0.91	2.83	4.16	0.79	2.09	3.83	0.97	3.22	3.01	0.57	2.46	3.91	1.06	2.31	3.61	0.84	2.53
ALL-E (Li et al., 2023b)	3.49	0.88	2.80	3.78	0.80	2.18	3.32	1.27	2.95	3.08	0.49	2.38	3.89	1.10	2.45	3.45	0.88	2.56
PairLIE (Fu et al., 2023)	4.08	0.67	3.39	4.52	0.78	3.86	4.17	1.08	4.23	3.66	0.51	2.68	4.21	0.94	2.49	4.05	0.72	3.35
SCL-LLE (Liang et al., 2022)	3.51	0.87	**2.73**	3.78	0.76	2.20	3.31	1.25	2.47	3.16	0.49	1.63	3.88	1.08	2.08	3.46	0.85	2.46
The proposed PIE	**3.47**	**0.99**	**2.73**	3.78	**0.83**	2.16	3.22	**1.32**	**2.40**	2.98	0.58	**1.62**	**3.72**	**1.23**	**2.07**	**3.38**	**0.95**	**2.44**

The best results are highlighted in bold

Table 2 PSNR ↑ and SSIM ↑ on the Part2 of the SICE dataset

Methods	LIME	R.Net	ISSR	Z.-DCE	E.GAN	RUAS	ReLLIE	U.-net	SCI	IRN	ALL-E	PairLIE	SCL-LLE	PIE
PSNR ↑	13.67	16.98	15.01	14.78	17.82	10.62	18.17	**20.61**	12.04	13.62	8.40	15.82	17.95	<u>19.79</u>
SSIM ↑	0.62	0.66	0.65	0.62	0.66	0.44	<u>0.67</u>	0.66	0.64	0.52	0.30	0.65	**0.68**	**0.68**

The best results are highlighted in bold, and the second best results are underlined

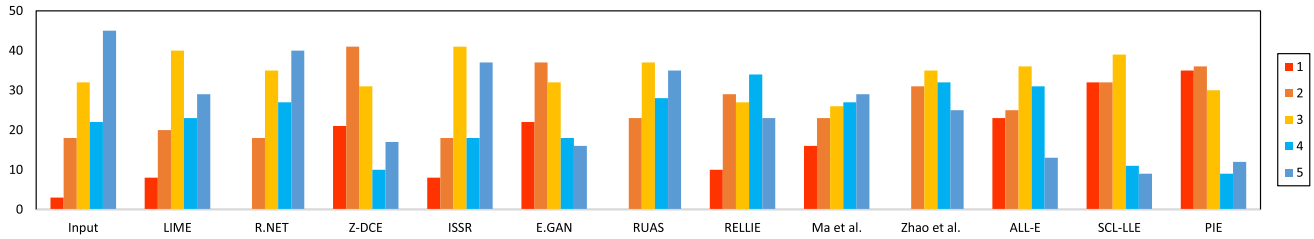
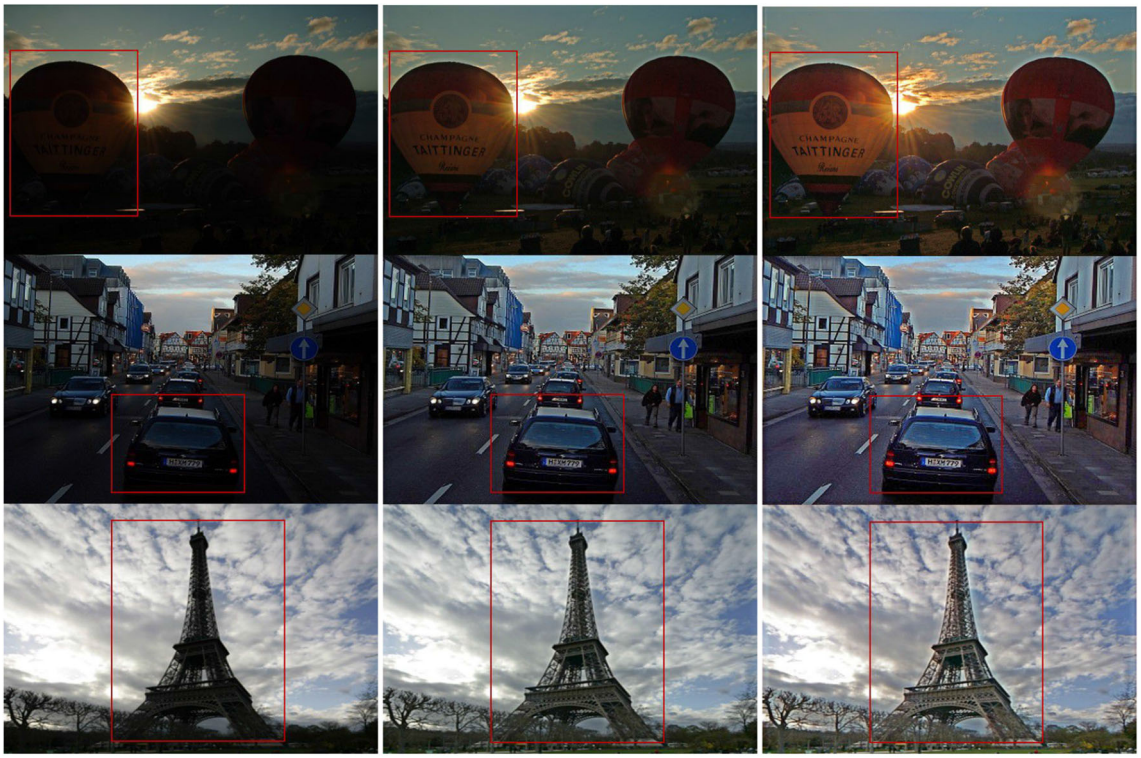


Fig. 10 The results in the human subjective survey. The color-changing from hot to cool means the quality transition from best to worst; the y-axis denotes the number of images in each ranking index (Color figure online)



(a) Input (b) w/o L_{rc} (c) w L_{rc}

Fig. 11 Ablation study on the contribution of the regional brightness consistency loss L_{rc}

LIME, MEF, VV, and NPE) enhanced by thirteen methods (LIME, Retinex-Net, Zero-DCE, ISSR, EnlightenGAN, RUAS, ReLLIE, SCL-LLE, SCI, IRN, ALL-E, and PIE), we ask 11 human subjects to rank the enhanced images. These subjects are instructed to consider:

(1) Whether or not the images contain visible noise.
(2) Whether the images have overexposed or underexposed artifacts.
(3) Whether the images show non-realistic color or texture distortion.

We assign a score to each image on a scale of 1–5, with lower values indicating better image quality.

Each image is assigned a score ranging from 1 to 5, with lower scores indicating better image quality. The final results are presented in Table 1 and Fig. 10. Among all the methods evaluated, PIE achieves the best image quality.

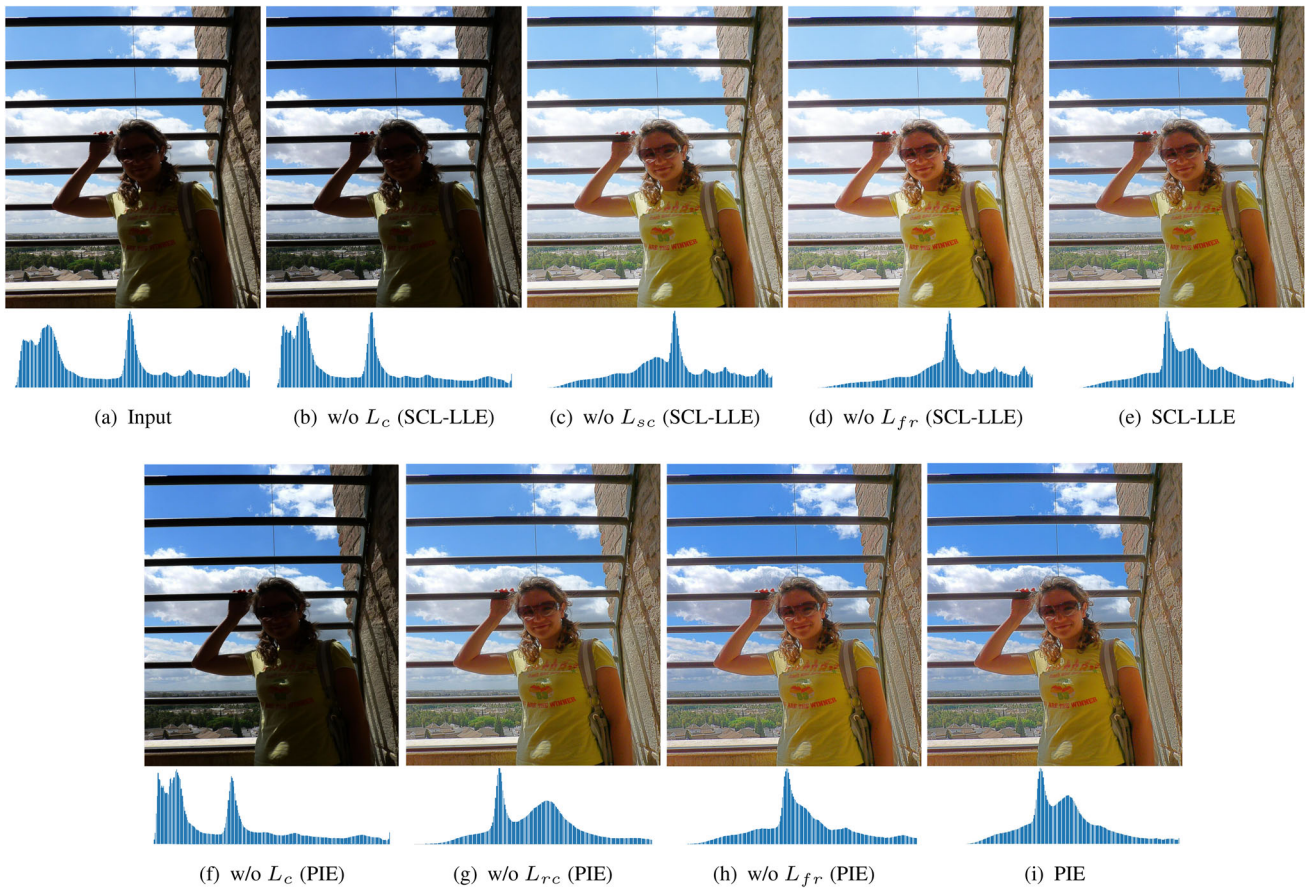


Fig. 12 Ablation study on the contribution of each loss for PIE and SCL-LLE (Liang et al., 2022)

Table 3 Ablation study. NIQE ↓ and UNIQUE (UN.) ↑ scores on the testing sets

Methods	DICM		LIME		MEF		VV		NPE		Average	
	NIQE↓	UN.↑	NIQE↓	UN.↑	NIQE↓	UN.↑	NIQE↓	UN.↑	NIQE↓	UN.↑	NIQE↓	UN.↑
Input	4.26	0.72	4.36	0.70	4.26	0.72	3.52	**0.74**	4.32	1.17	4.13	0.75
w/o L_c (SCL-LLE)	4.31	0.64	4.36	0.57	4.25	0.56	4.10	0.70	4.28	1.02	4.27	0.66
w/o L_{sc} (SCL-LLE)	3.53	0.83	3.85	0.76	3.32	1.18	3.21	0.50	3.98	1.02	3.49	0.82
w/o L_{fr} (SCL-LLE)	3.54	0.80	3.88	0.71	3.32	1.22	3.18	0.47	3.97	1.03	3.50	0.80
w/o L_c (PIE)	4.57	0.67	4.54	0.52	4.61	0.52	3.64	0.72	4.34	1.02	4.38	0.66
w/o L_{rc} (PIE)	3.54	0.94	3.85	0.76	3.32	1.32	3.01	0.54	3.87	1.02	3.45	0.90
w/o L_{fr} (PIE)	3.50	0.94	4.26	0.79	3.63	1.24	3.03	0.58	4.08	1.09	3.53	0.91
w/o Neg. samples (PIE)	3.55	0.81	3.84	0.72	3.36	1.14	3.14	0.38	3.95	1.01	3.49	0.78
w/o overexp. Neg. (PIE)	3.59	0.75	3.91	0.59	3.36	1.24	3.15	0.37	4.12	0.87	3.54	0.74
w/o underexp. Neg. (PIE)	4.58	0.57	4.52	0.48	4.69	0.46	3.58	0.65	4.36	0.86	4.38	0.58
Gamma curves only	3.52	0.96	3.79	0.81	**3.22**	1.19	3.01	0.63	3.75	1.18	3.39	**0.95**
Sigmoid curves only	3.51	0.97	**3.77**	0.82	3.31	1.19	3.10	0.61	3.80	**1.23**	3.40	0.93
Logarithmic curves only	3.52	0.97	3.86	0.79	3.32	**1.33**	3.14	0.63	3.84	1.18	3.47	0.94
SCL-LLE	3.51	0.87	**3.78**	0.76	3.31	1.25	3.16	0.49	3.88	1.08	3.46	0.85
PIE	**3.47**	**0.99**	**3.78**	**0.83**	**3.22**	**1.32**	**2.98**	0.58	**3.72**	**1.23**	**3.38**	**0.95**

The best results are highlighted in bold

Table 4 Discussion on the contrastive loss

L_{cG}			L_{cE}			$L_c = L_{cG} + L_{cE}$			
Triple	N-pair	InfoNCE	Triple	N-pair	InfoNCE	NIQE↓	UN.↑	PSNR↑	SSIM↑
✓			✓			3.44	0.93	18.44	0.67
✓				✓		3.65	0.83	15.08	0.62
✓					✓	**3.38**	**0.95**	**19.79**	**0.68**
	✓		✓			4.26	0.67	10.69	0.44
	✓			✓		5.04	0.43	8.49	0.28
	✓				✓	4.16	0.41	12.19	0.49
		✓	✓			5.13	0.43	8.37	0.27
		✓		✓		5.34	0.21	7.26	0.18
		✓			✓	4.94	0.15	9.28	0.31

The best results are highlighted in bold
The NIQE and UNIQUE scores are the average results on five datasets: DICM, LIME, MEF, VV, and NPE. The PSNR and SSIM scores are the results on the Part2 of the SICE dataset

Table 5 Comparisons of different positive and negative sample rates

Positive	Negative	DICM		LIME		MEF		VV		NPE		Average		Average training
		NIQE↓	UN.↑	NIQE↓	UN.↑	NIQE↓	UN.↑	NIQE↓	UN.↑	NIQE↓	UN.↑	NIQE↓	UN.↑	time/epoch (min.)
1	1	3.47	**0.99**	3.78	0.83	**3.22**	1.32	2.98	**0.58**	**3.72**	**1.23**	3.38	**0.95**	73.6
1	5	3.47	**0.99**	**3.74**	**0.85**	**3.22**	**1.33**	**2.96**	0.57	3.74	1.20	**3.37**	**0.95**	131.3
5	1	**3.38**	0.93	3.90	0.71	3.30	1.31	3.03	0.48	3.91	1.04	3.46	0.90	126.8
5	5	**3.38**	0.96	3.86	0.80	3.28	1.32	2.99	0.53	3.89	1.09	3.43	0.86	158.6

The best results are highlighted in bold
The baseline is PIE with the rate of 1:1

4.2 Ablation Study

In this section, we perform ablation studies to demonstrate the effectiveness of each component of PIE.

4.2.1 Contribution of Each Loss

In this section, we explore the impact of each loss on PIE and compare them with the loss in our baseline SCL-LLE (Liang et al., 2022). We consider the color consistency item L_{cc}, initially proposed and tested in Zero-DCE (Guo et al., 2020), as a baseline item without conducting an ablation study. Thus, we test the feature preservation loss L_{fp} using the first item L_{fr}.

We perform ablation studies to demonstrate the effectiveness of three different loss functions in PIE: the contrastive learning loss L_c, the feature retention loss L_{fr}, and the regional brightness consistency loss L_{rc}. Figure 12f–i shows the visualized samples with their corresponding histograms of the effects of L_c, L_{rc}, and L_{fr} functions in PIE. In Fig. 11, the enhanced result without using the regional segmentation module (b) exhibits color deviations. On the other hand, the enhanced result with the regional segmentation module (c) demonstrates that different regions can maintain their own colors, and there is a better distinction between the foreground and background (e.g., balloon, vehicles, and tower). Table 3 shows each loss's average NIQE and UNIQUE scores on five test sets. We find that the contrastive learning loss L_c significantly controls the exposure level. The results with-

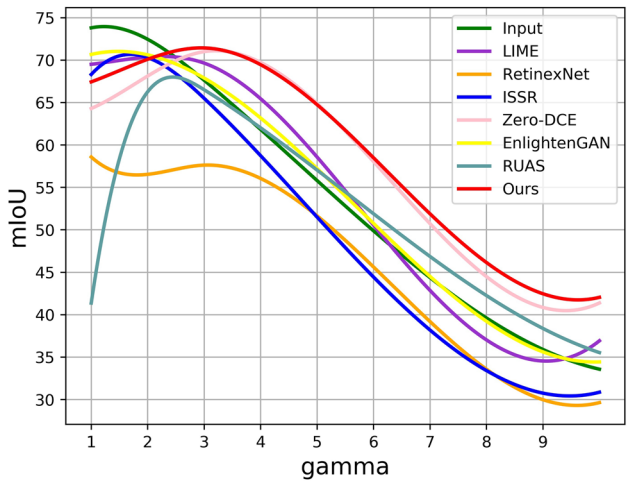
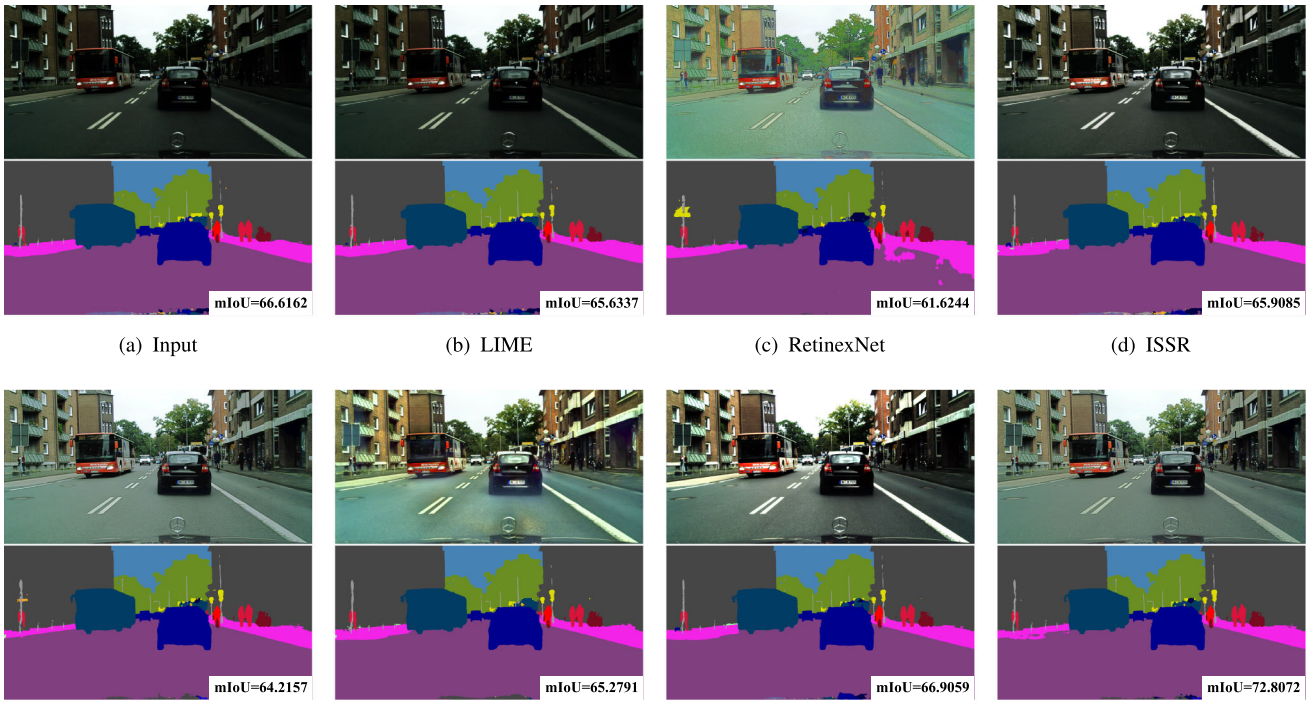


Fig. 13 The semantic segmentation results of the input low-light images after enhancement. When using the original input ($\gamma = 1$), the semantic segmentation with all the enhancement models could not surpass the initial input. When γ becomes larger, the mIoU of segmentation after using our method has been significantly better than those using the original images



(a) Input (b) LIME (c) RetinexNet (d) ISSR

(e) Zero-DCE (f) EnlightenGAN (g) RUAS (h) Ours

Fig. 14 Visual comparison of seven LLE methods for semantic segmentation

out the regional brightness consistency loss L_{rc} in PIE, and without the feature retention loss L_{fr} have relatively lower contrast (e.g., in the sky region) than the final result. Furthermore, we compare the performance of PIE with SCL-LLE (Liang et al., 2022). The contrastive learning loss L_c in PIE is more critical and effective in controlling the exposure level than in SCL-LLE. However, due to the use of an unsupervised method, the contribution of the Regional Brightness Consistency Loss L_{rc} in PIE is slightly worse than the Semantic Brightness Consistency Loss L_{sc} in SCL-LLE.

The losses enhance images with fine details and more naturalistic and perceptually favored quality. The corresponding histograms show that the final losses maintain a smooth mixture-of-Gaussian-like global distribution with rare over or under-saturation areas. In contrast, the undesirable unilateral over or under-saturation areas occur in the histograms of Fig. 12b–d and f.

Table 6 The average precision (AP) for face detection in low-light conditions on the DARK FACE dataset was evaluated using different IoU thresholds (0.5, 0.7, 0.9)

Methods	IoU thresholds		
	0.5	0.7	0.9
Input	0.2820	0.0693	0.0002
LIME (Guo et al., 2016)	0.4221	0.1068	**0.0004**
RetinexNet (Wei et al., 2018)	0.3874	0.1065	0.0002
ISSR (Fan et al., 2020)	0.2825	0.0674	0.0001
Zero-DCE (Guo et al., 2020)	0.4130	0.1067	0.0002
EnlightenGAN (Jiang et al., 2021)	0.3762	0.1009	0.0003
RUAS (Liu et al., 2021)	0.2782	0.0659	0.0002
ReLLIE (Zhang et al., 2021a)	0.3583	0.0958	0.0001
PIE	0.4199	0.1082	0.0003
PIE_{det}	**0.4288**	**0.1104**	0.0003

The best results are highlighted in bold

4.2.2 Contribution of the Curves in BoC

In the BoC method, we apply three types of curves—Gamma, Sigmoid, and Logarithmic to adjust the brightness of images. To investigate the contribution of each curve to our method, we separately use each type of curve to generate over/underexposed images. As shown in Table 3, all three types of curves can produce over/underexposed images as negative samples for contrastive learning, which can help the model better learn the features of the data.

4.2.3 Discussion for Contrastive Learning

Discussion for Contrastive Loss Triplet loss (Hermans et al., 2017), N-pair loss (Sohn, 2016), and InfoNCE loss (Gutmann & Hyvärinen, 2010) are commonly used in contrastive learning to pull the anchor closer to the positive sample and push

(a) RetinaFace

(b) RetinaFace with PIE$_{det}$

Fig. 15 Qualitative comparison of face detection without and with PIE. The detector is RetinaFace (Deng et al., 2020)

it away from the negative sample in the latent feature space. In PIE, we apply contrastive losses, L_{cG} and L_{cE}, respectively, to the gram matrix G and the expectation E to help the model learn features that represent positive samples and avoid features that represent negative samples. We find that applying different forms of contrastive losses to L_{cG} and L_{cE} results in different gains. Table 4 shows the results of training with different loss forms for L_{cG} and L_{cE}. We find that using triplet loss for L_{cG} and infoNCE loss for L_{cE} achieved the best results in terms of four metrics: NIQE, UNIQUE, PSNR, and SSIM. **Numbers of positive and negative samples** We conduct further experiments to explore the effect of different rates between positive and negative samples. For positive samples, we randomly select them from the positive sample dataset. We use our PIE with the rate of 1:1 as the baseline, and all experimental settings are the same as before, except for the number of samples. As the batch size increases, the GPU memory size required for training will increase, and the training time required for training an epoch (i.e., the process of using all samples in the training set to train once) will also significantly increase. Considering these factors, we use at most 5 positive or negative samples.

As shown in Table 5, adding more negative samples resulted in better performance, while adding more positive samples led to worse results. We conjecture that this is due to the different positive patterns that confuse the low-light image and hinder its capability to learn useful patterns. For negative samples, using more samples helped the model move away from the poor patterns in the over/underexposed images. However, increasing the number of negative samples also increased the training time. When we train using the rate of 1:1, the time required to train an epoch is 73.6 min. When we train using the rate of 1:5, the time required to train an epoch increases to 131.3 min. Therefore, in our experiments, we use the rate of 1:1, except for Table 5.

4.3 Gain for Downstream Tasks

4.3.1 Semantic Segmentation with PIE

Current low-light image datasets lack semantic annotation, which makes it difficult to evaluate semantic segmentation performance before and after enhancement. To address this issue, we use subsets of Frankfurt, Lindau, and Munster from the Cityscapes validation set. Additionally, we simulate low-light images with varying brightness levels using the standard positive Gamma transformation with a range of Gamma values. The trends of mean intersection-over-union (mIoU) with the brightness of the scene are shown in Fig. 13. Among all the methods, the segmentation performance with our method tends to be the best when scenes become dark. In Fig. 14, our method effectively improves semantic segmentation performance compared with LLE state-of-the-art methods. These findings motivate us to explore ways to bridge the gap between current low-light enhancement methods and downstream tasks.

4.3.2 Face Detection with PIE

We use RetinaFace (Deng et al., 2020) trained on the WIDER FACE dataset (Yang et al., 2016) as the face detector. Two thousand images in the DARK FACE dataset Yang et al. (2020) are used as test input, and different methods are used to enhance them respectively. Then, we feed the results of different low-light image enhancement methods to RetinaFace. To evaluate the accuracy of the model, we compare the average precision (AP) under different IoU thresholds (0.5, 0.7, and 0.9). A target is considered detected when the IoU is greater than 50%. Table 6 shows the AP results, with a focus on the IoU threshold of 0.5. All low-light image enhancement methods except ISSR and RUAS improve the face detection performance on the dataset. However, when we set a higher IoU threshold, the AP scores of all methods decrease. Our

Table 7 Experimental results of test-time cost comparison

Method	GFLOPs ↓	Runtime (s) ↓
LLNet (Lore et al., 2017)	4124.17	36.270
MBLLEN (Lv et al., 2018)	301.12	13.995
KinD++ (Zhang et al., 2021c)	12238.02	1.068
Zero-DCE (Guo et al., 2020)	84.99	**0.003**
EnlightenGAN (Jiang et al., 2021)	273.24	0.008
ReLLIE (Zhang et al., 2021a)	125.13	1.480
LIME (Guo et al., 2016)	(on CPU)	21.530
Retinex-Net (Wei et al., 2018)	587.470	0.120
ISSR (Fan et al., 2020)	(unavailable)	9.645
RUAS (Liu et al., 2021)	1.069	0.006
IRN (Zhao et al., 2021)	12438.282	4.216
SCI (Ma et al., 2022)	**0.580**	0.010
ALL-E (Li et al., 2023b)	113.85	1.055
SCL-LLE (Liang et al., 2022)	95.21	0.004
PIE	85.54	0.004

The best results are highlighted in bold

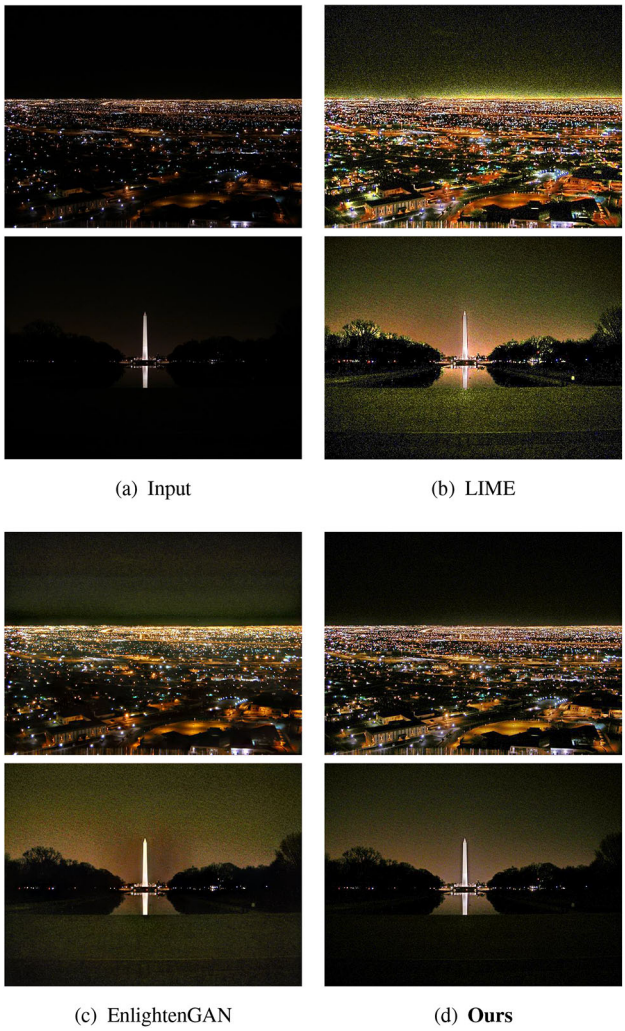
PIE method, which does not require paired training data, achieved a comparable score to the best result produced by LIME at the IoU threshold of 0.5, even without joint training with a face detection model. However, as mentioned earlier, LIME's subjective and quantitative results are not satisfactory. In contrast, our method produces better visual results. Our approach, PIE_{det}, achieves the best performance by training with a joint face detection model. Figure 15 shows the comparison between the original detection and detection with our PIE.

4.4 Test-Time Cost Comparison

We compare the test runtime and computational costs (measured in Giga floating-point operations per second, GFLOPs) of our method with state-of-the-art methods (LLNet, MBLLEN, KinD++, Zero-DCE, EnlightenGAN, ReLLIE, LIME, Retinex-Net, ISSR, RUAS, IRN, SCI, ALL-E, SCL-LLE). All indicators are recorded with full processing for 32 images of size $1200 \times 900 \times 3$ using an NVIDIA GTX1080Ti GPU. As shown in Table 7, our method is only slightly slower than Zero-DCE (our backbone LLE Network), but much faster than most other methods excluding Zero-DCE in terms of runtime. This indicates that PIE has high speed and can quickly process a large number of images. Our method achieves a moderate level in terms of computational cost (GFLOPs).

4.5 Failure Cases

In Fig. 16, we showcase two scenarios in which the Performance-Improvement Enhancement algorithm encoun-



(a) Input (b) LIME

(c) EnlightenGAN (d) **Ours**

Fig. 16 The failure cases

ters failures. These failures arise when the low-light image input contains a substantial amount of noise. Similar to other established approaches like LIME and EnlightenGAN, our method also exhibits the presence of background stripe noise in the enhanced images, which is influenced by the inherent noise in the original image. To address this issue, a module for removing stripe noise can be incorporated as a solution.

5 Conclusion

We propose physics-inspired contrastive learning for low-light image enhancement (PIE). This is achieved by introducing Bag of Curves in contrastive learning, which efficiently generates negative samples that mimic the Gamma correction and Tone mapping processes in the ISP pipeline. BoC generates under/overexposed images aligned with the underlying physical imaging principles. Additionally, the regional segmentation module is an unsupervised method that maintains regional brightness consistency and removes the dependence on semantic ground truths. Extensive experiments demonstrate that our method outperforms the state-of-the-art LLE models on six independent cross-scene datasets. Furthermore, we conduct experiments combining LLE with semantic segmentation, object detection, and image classification, demonstrating that PIE benefits downstream tasks under extremely dark conditions. The proposed method runs fast with reasonable GFLOPs in test time, making it easy to use on mobile devices.

Acknowledgements This work was partly supported by NSFC (Grant Nos. 62272229, 62076124, 62222605), the National Key R&D Program of China (2020AAA0107000), the Natural Science Foundation of Jiangsu Province (Grant Nos. BK20222012, BK20211517), and Shenzhen Science and Technology Program JCYJ20230807142001004. The authors would like to thank all the anonymous reviewers for their constructive commments.

References

Al Sobbahi, R., & Tekli, J. (2022). Comparing deep learning models for low-light natural scene image enhancement and their impact on object detection and classification: Overview, empirical evaluation, and challenges. *Signal Processing: Image Communication, 109*, 116848.

Anaya, J., & Barbu, A. (2018). Renoir: A dataset for real low-light image noise reduction. *Journal of Visual Communication and Image Representation, 51*, 144–154.

Arici, T., Dikbas, S., & Altunbasak, Y. (2009). A histogram modification framework and its application for image contrast enhancement. *IEEE Transactions on Image Processing, 18*(9), 1921–1935.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute, 310*(1), 1–26.

Bychkovsky, V., Paris, S., Chan, E., & Durand, F. (2011). Learning photographic global tonal adjustment with a database of input/output image pairs. In *IEEE conference on computer vision and pattern recognition* (pp. 97–104).

Cai, J., Gu, S., & Zhang, L. (2018). Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing, 27*(4), 2049–2062.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder–decoder with atrous separable convolution for semantic image segmentation. In *European Conference on computer vision* (pp. 801–818).

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597–1607).

Chen, X., Pan, J., Jiang, K., Li, Y., Huang, Y., Kong, C., Dai, L., & Fan, Z. (2022). Unpaired deep image deraining using dual contrastive learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2017–2026).

Cho, S. W., Baek, N. R., Koo, J. H., & Park, K. R. (2020). Modified perceptual cycle generative adversarial network-based image enhancement for improving accuracy of low light image segmentation. *IEEE Access, 9*, 6296–6324.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *IEEE conference on computer vision and pattern recognition* (pp. 3213–3223).

Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5203–5212).

Drago, F., Myszkowski, K., Annen, T., & Chiba, N. (2003). Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer graphics forum, Wiley Online Library* (pp. 419–426).

Fan, M., Wang, W., Yang, W., & Liu, J.(2020). Integrating semantic segmentation and retinex model for low-light image enhancement. In *ACM international conference on multimedia, virtual event* (pp. 2317–2325).

Farid, H. (2001). Blind inverse gamma correction. *IEEE Transactions on Image Processing, 10*(10), 1428–1433.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision, 59*, 167–181.

Fu, Z., Yang, Y., Tu, X., Huang, Y., Ding, X., & Ma, K. K. (2023). 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).

Geng, Q., Liang, D., Zhou, H., Zhang, L., Sun, H., & Liu, N. (2021). Dense face detection via high-level context mining. In *2021 16th IEEE international conference on automatic face and gesture recognition (FG 2021)* (pp. 1–8). IEEE.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440–1448).

Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., & Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. In *IEEE conference on computer vision and pattern recognition* (pp. 1780–1789).

Guo, X., Li, Y., & Ling, H. (2016). Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing, 26*(2), 982–993.

Gutmann, M., & Hyvärinen, A. (2010). Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 297–304).

Han, J., Shoeiby, M., Malthus, T., Botha, E., Anstee, J., Anwar, S., Wei, R., Petersson, L., & Armin, M. A. (2021). Single underwater image restoration by contrastive learning. In *2021 IEEE international geoscience and remote sensing symposium IGARSS* (pp. 2385–2388). IEEE.

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *IEEE*

conference on computer vision and pattern recognition (pp. 9729–9738).

Henaff, O. (2020). Data-efficient image recognition with contrastive predictive coding. In *International conference on machine learning* (pp. 4182–4192).

Hermans, A., Beyer, L., & Leibe, B. (2017). In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737

Huang, Y., Tu, X., Fu, G., Liu, T., Liu, B., Yang, M., & Feng, Z. (2023). Low-light image enhancement by learning contrastive representations in spatial and frequency domains. arXiv preprint arXiv:2303.13412

Ibrahim, H., & Kong, N. (2007). Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics, 53*(4), 1752–1758.

Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., & Van Gool, L. (2017). Dslr-quality photos on mobile devices with deep convolutional networks. In *IEEE international conference on computer vision* (pp. 3277–3285).

Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., & Wang, Z. (2021). Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing, 30*, 2340–2349.

Jin, X., Han, L. H., Li, Z., Guo, C. L., Chai, Z., & Li, C. (2023). Dnf: Decouple and feedback network for seeing in the dark. In *IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 18135–18144).

Jobson, D. J., Rahman, Zu., & Woodell, G. A. (1997). A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing, 6*(7), 965–976.

Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision* (pp. 694–711).

Karaimer, H. C., & Brown, M. S. (2016). A software platform for manipulating the camera imaging pipeline. In *Computer vision–ECCV 2016: 14th European conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I* (Vol. 14, pp. 429–444).

Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., & Matas, J., (2018). Deblurgan: Blind motion deblurring using conditional adversarial networks. In *IEEE conference on computer vision and pattern recognition* (pp. 8183–8192).

Land, E. H. (1977). The retinex theory of color vision. *Scientific American, 237*(6), 108–129.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE conference on computer vision and pattern recognition* (pp. 4681–4690).

Lee, C., Lee, C., & Kim, C. S. (2012). Contrast enhancement based on layered difference representation. In *IEEE international conference on image processing* (pp. 965–968).

Li, C., Guo, C., Feng, R., Zhou, S., & Loy, C. C. (2022). Cudi: Curve distillation for efficient and controllable exposure adjustment. arXiv preprint arXiv:2207.14273

Li, C., Guo, C., Han, L., Jiang, J., Cheng, M. M., Gu, J., & Loy, C. C. (2021a). Low-light image and video enhancement using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(12), 9396–9416.

Li, C., Guo, C., Zhou, S., Ai, Q., Feng, R., & Loy, C. C., (2023). Flexicurve: Flexible piecewise curves estimation for photo retouching. In *IEEE conference on computer vision and pattern recognition NTIRE workshop (CVPRW)-Oral* (pp. 1092–1101).

Li, L., Liang, D., Gao, Y., Huang, S. J., & Chen, S. (2023). All-e: Aesthetics-guided low-light image enhancement. arXiv preprint arXiv:2304.14610

Li, W., Yang, X., Kong, M., Wang, L., Huo, J., Gao, Y., & Luo, J. (2021b). Triplet is all you need with random mappings for unsupervised visual representation learning. arXiv preprint arXiv:2107.10419

Liang, D., Kaneko, S. I., Hashimoto, M., Iwata, K., Zhao, X., & Satoh, Y. (2014). Robust object detection in severe imaging conditions using co-occurrence background model. *International Journal of Optomechatronics, 8*(1), 14–29.

Liang, D., Kang, B., Liu, X., Gao, P., Tan, X., & Kaneko, S. I. (2021). Cross-scene foreground segmentation with supervised and unsupervised model communication. *Pattern Recognition, 117*, 107995.

Liang, D., Li, L., Wei, M., Yang, S., Zhang, L., Yang, W., Du, Y., & Zhou, H. (2022). Semantically contrastive learning for low-light image enhancement. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 1555–1563).

Liang, Z., Li, C., Zhou, S., Feng, R., & Loy, C. C. (2023). Iterative prompt learning for unsupervised backlit image enhancement. arXiv preprint arXiv:2303.17569

Liu, R., Ma, L., Zhang, J., Fan, X., & Luo, Z. (2021). Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *IEEE conference on computer vision and pattern recognition* (pp. 10561–10570).

Lore, K. G., Akintayo, A., & Sarkar, S. (2017). Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition, 61*, 650–662.

Lv, F., Lu, F., Wu, J., & Lim, C. (2018). Mbllen: Low-light image/video enhancement using cnns. In *British machine vision conference* (p 4).

Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., & Zhang, L. (2017). Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing, 26*(2), 1004–1016.

Ma, K., Zeng, K., & Wang, Z. (2015). Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing, 24*(11), 3345–3356.

Ma, L., Ma, T., Liu, R., Fan, X., & Luo, Z. (2022). 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).

Mantiuk, R., Daly, S., & Kerofsky, L. (2008). Display adaptive tone mapping. *ACM Transactions on Graphics, 27*(3), 509–518.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *IEEE international conference on computer vision* (pp. 416–423).

Mittal, A., Soundararajan, R., & Bovik, A. (2013). Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters, 20*(3), 209–212.

Pizer, S., Johnston, R., & Ericksen, J. P. (1990). Contrast-limited adaptive histogram equalization: Speed and effectiveness. *Conference on Visualization in Biomedical Computing, 337*, 337–345.

Ren, W., Liu, S., Ma, L., Xu, Q., Xu, X., Cao, X., Du, J., & Yang, M. H. (2019). Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing, 28*(9), 4364–4375.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., & Berg, A. C. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision, 115*, 211–252.

Sermanet, P., Lynch, C., Chebotar, Y., Hsu, J., Jang, E., Schaal, S., Levine, S., & Brain, G. (2018). Time-contrastive networks: Self-supervised learning from video. In *IEEE international conference on robotics and automation* (pp. 1134–1141).

Shi, Y., Wang, B., Wu, X., & Zhu, M. (2022). Unsupervised low-light image enhancement by extracting structural similarity and color consistency. *IEEE Signal Processing Letters, 29*, 997–1001.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556

Sivic & Zisserman (2003). Video google: A text retrieval approach to object matching in videos. In *Proceedings ninth IEEE international conference on computer vision* (pp. 1470–1477). IEEE.

Sohn, K. (2016). Improved deep metric learning with multi-class n-pair loss objective. In *Advances in neural information processing systems* (Vol. 29).

Tian, Y., Krishnan, D., & Isola, P. (2020). Contrastive multiview coding. In *Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, Aug 23–28, 2020, Proceedings, Part XI* (Vol. 16, pp. 776–794). Springer.

Wang, H., Chen, Y., Cai, Y., Chen, L., Li, Y., Sotelo, M. A., & Li, Z. (2022). Sfnet-n: An improved sfnet algorithm for semantic segmentation of low-light autonomous driving road scenes. *IEEE Transactions on Intelligent Transportation Systems, 23*(11), 21405–21417.

Wang, S., Zheng, J., Hu, H. M., & Li, B. (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing, 22*(9), 3538–3548.

Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560

Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., & Ma, L. (2021). Contrastive learning for compact single image dehazing. In *IEEE conference on computer vision and pattern recognition* (pp. 10551–10560).

Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., & Jiang, J. (2022a). 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).

Wu, Y., Guo, H., Chakraborty, C., Khosravi, M., Berretti, S., & Wan, S. (2022b). Edge computing driven low-light image dynamic enhancement for object detection. *IEEE Transactions on Network Science and Engineering*. https://doi.org/10.1109/TNSE.2022.3151502

Wu, Y., Pan, C., Wang, G., Yang, Y., Wei, J., Li, C., & Shen, H. T. (2023). Learning semantic-aware knowledge guidance for low-light image enhancement. In *IEEE conference on computer vision and pattern recognition (CVPR)*.

Xu, K., Yang, X., Yin, B., & Lau, R. W. (2020). Learning to restore low-light images via decomposition-and-enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2281–2290).

Xu, Q., Jiang, H., Scopigno, R., & Sbert, M. (2014). A novel approach for enhancing very dark image sequences. *Signal Processing, 103*, 309–330.

Yang, S., Luo, P., Loy, C. C., & Tang, X. (2016). Wider face: A face detection benchmark. In *IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 5525–5533).

Yang, W., Yuan, Y., Ren, W., Liu, J., Scheirer, W. J., Wang, Z., Zhang, T., Zhong, Q., Xie, D., Pu, S., & Zheng, Y. (2020). Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing, 29*, 5737–5752.

Yongqing, H. (2013). Dodging and burning inspired inverse tone mapping algorithm. *Computational Information Systems, 9*,

Yuan, L., & Sun, J. (2012). Automatic exposure correction of consumer photographs. In *Computer vision–ECCV 2012: 12th European conference on computer vision, Florence, Italy, October 7–13, 2012, Proceedings, Part IV* (Vol. 12, pp. 771–785). Springer.

Zhang, J., Wang, Y., Tohidypour, H., Pourazad, M. T., & Nasiopoulos, P. (2023). A generative adversarial network based tone mapping operator for 4k hdr images. In *2023 international conference on computing, networking and communications (ICNC)* (pp. 473–477).

Zhang, R., Guo, L., Huang, S., & Wen, B. (2021a). Rellie: Deep reinforcement learning for customized low-light image enhancement. In *Proceedings of the 29th ACM international conference on multimedia* (pp. 2429–2437).

Zhang, W., Ma, K., Zhai, G., & Yang, X. (2021b). Uncertainty-aware blind image quality assessment in the laboratory and wild. *IEEE Transactions on Image Processing, 30*, 3474–3486.

Zhang, Y., Guo, X., Ma, J., Liu, W., & Zhang, J. (2021c). Beyond brightening low-light images. *International Journal of Computer Vision, 129*, 1013–1037.

Zhang, Y., Zhang, J., & Guo, X. (2019). Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia* (pp. 1632–1640).

Zhao, L., Lu, S. P., Chen, T., Yang, Z., & Shamir, A. (2021). Deep symmetric network for underexposed image enhancement with recurrent attentional learning. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 12075–12084).

Zhou, S., Li, C., & Loy, C. C. (2022). Lednet: Joint low-light enhancement and deblurring in the dark. In *European conference on computer vision (ECCV)*.

Zhou, Y., Liang, D., Chen, S., Huang, S. J., Yang, S., & Li, C (2023). Improving lens flare removal with general-purpose pipeline and multiple light sources recovery. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 12969–12979).