DMGINE: Day-Memory Guided Nighttime Image Enhancement for Dynamic Traffic Scenes

Ruizhou Liu[1,2*], Zhe Wu[2*], Zimo Liu[2], Qingfang Zheng[2†], Qingming Huang[1,2‡]

[1] School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China
[2]Department of the Networked Intelligence, Pengcheng Laboratory, Shenzhen, China
liurzh@pcl.ac.cn, wuzh02@pcl.ac.cn, liuzm@pcl.ac.cn, zhengqf01@pcl.ac.cn, qmhuang@ucas.ac.cn

Abstract

We introduce Daytime-Memory Guided Nighttime Image Enhancement (DMGNIE) framework, the first framework that turns long-running daytime surveillance videos of a single intersection into persistent "daytime memory" to guide nighttime image enhancement in traffic scenes. Our key insight is simple yet powerful: for a static scene, perfectly exposed daytime frames are, pixel-for-pixel, high-quality illumination prior for the same location under extreme low-light. Due to the complex lighting conditions in real-world traffic scenes, existing low-light image enhancement (LLIE) methods suffer from issues such as overexposure in highlight regions and noise amplification in low-light condition regions, which degrades the performance of downstream computer vision tasks. DMGNIE tackles these issues in two steps: (1) SegBMN, a semantic prior-based background modeling network, distills a clean, static daytime background from hours of video as scene prior guiding the enhancement of nighttime image; (2) a Foreground Localization-Guided Contrastive Learning module avoid the interference from the background prior with foreground objects during the guidance by maximizing the differences between foreground and background features. Finally, We conduct comprehensive experiments on real traffic surveillance datasets of two cities to evaluate the effectiveness. And the experimental results demonstrate that DMGNIE outperforms state-of-the-art baselines and achieves superior performance in challenging low-light conditions.

Introduction

In traffic scenarios with poor illumination conditions, such as rural roads and urban highways, surveillance cameras often suffer severely limited light intake due to low-light environments. This results in issues like color distortion and detail blurring in captured images, subsequently degrading the performance of computer vision tasks—including reduced target detection accuracy (Han and Lim 2024; Sun, Li, and Mu 2024) and confused semantic segmentation boundaries (Chen et al. 2023; Choe et al. 2024). Therefore, low-light image enhancement (LLIE) is crucial for ensuring the reliability of nighttime traffic monitoring systems and improving the robustness of computer vision algorithms under complex lighting conditions. So far, lots of LLIE methods are developed, such as statistical adaptive techniques, e.g., histogram equalization (Pizer et al. 1987; Poynton 2012), Retinex theory (Land 1977), and deep learning-based frameworks, e.g., LLNet (Lore, Akintayo, and Sarkar 2017), Zero-DCE (Guo et al. 2020; Li, Guo, and Loy 2021). These solutions primarily enhance low-light image quality through global illumination correction or local contrast improvement, performing well under moderate low-light conditions. However, they face significant limitations in practical applications. Traditional methods suffer from rigid parameter mappings that degrade performance under extreme dynamic ranges, while deep learning models exhibit generalization errors with rare illumination combinations due to training data constraints. As shown in Fig. 1(a), the enhanced result of a state-of-the-art LLIE method on a real nighttime traffic surveillance image exhibits overexposure in highlight regions and fails to restore fine-grained details (e.g., tree leaves) with their original colors in low-light areas.

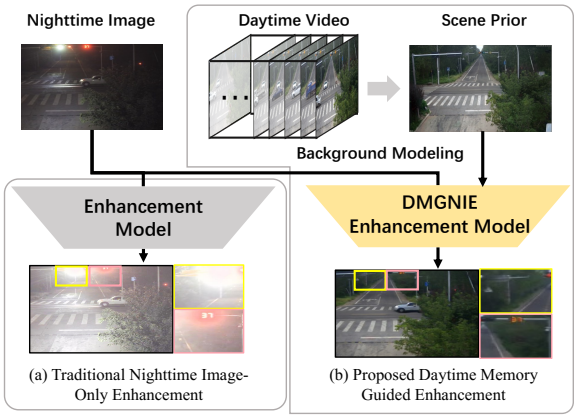


Figure 1: Illustration of our motivation. (a) Traditional LLIE methods. (b) Our proposed DMGNIE framework with daytime scene prior. As illustrated by the enhanced results, DMGNIE not only preserves foreground objects from the input image (highlighted in pink), but also effectively mitigates overexposure in bright regions (highlighted in yellow).

Such LLIE methods for video enhancement usually

*These authors contributed equally.
†Corresponding author
‡Corresponding author

treat every frame in isolation. Yet surveillance, traffic, and wildlife monitoring cameras are increasingly deployed in fixed positions and left running continuously. For these cameras, daytime recordings provide an abundant, perfectly registered, and temporally stable source of information about scene geometry, texture, and color distribution. One question is posed: *Can we treat the "daytime twin" as a privileged teacher for the night?*

To this end, we propose a novel framework, named **Daytime-Memory Guided Nighttime Image Enhancement (DMGNIE)**, the first framework that turns long-running daytime surveillance videos of a single intersection into persistent "memory" to guide high-fidelity nighttime image enhancement, as Fig. 1 (b) shows. We first construct the framework that turns long-running daytime surveillance videos of a single intersection to guide high-fidelity nighttime image enhancement. Specifically, to extract static scene information from the daytime video, a semantic prior-based background modeling module, termed SegBMN, is designed to extract clean static background image as daytime memory to guide nighttime image enhancement. Then, to avoid interference of the guidance of the background image with the information of foreground objects (e.g. vehicles, pedestrians) from the input nighttime image, we design a Foreground Localization-Guided Contrastive Learning (FLGCL) mechanism to enhance the discriminability between foreground and background features. Specifically, we firstly design a Siamese Network to obtain the location prior of foreground objects by differing the enhanced image and background image. Then, the location prior is used to guide the multi-scale contrastive leaning for the preservation of foreground objects information. Compared with traditional LLIE methods, DMGNIE can learn better illumination of daytime, and, unlike neural style transfer models, it preserves structural consistency with the original input and semantic fidelity .

In summary, our main contributions are as follows:

- To avoid foreground ghost artifacts in background modeling in real traffic scenes, a semantic prior-based background modeling network (SegBMN) is designed to model a clean and stable static daytime background from daytime video streams.
- A Foreground Localization-Guided Contrastive Learning mechanism is designed for preservation of dynamic foreground objects in nighttime images by separating the features between foreground and background.
- We conduct extensive quantitative and qualitative experiments on three real traffic surveillance datasets, and the results demonstrate that our method consistently outperforms state-of-the-art enhancement approaches in both visual quality and restoration accuracy.

Related Works

Low-Light Image Enhancement

Learning the transformations that enhancing texture details and color fidelity in the dark regions of low-light nighttime images lies at the core of low-light image enhancement. Traditional approaches (Pizer et al. 1987; Poynton 2012; Land 1977; Jobson, Rahman, and Woodell 1997; Wang et al. 2013) rely on statistical priors. In deep learning era, some fully supervised methods (Lore, Akintayo, and Sarkar 2017; Chen et al. 2018b; Yan et al. 2025) are proposed for learning low-light enhancement on paired dataset. However, acquiring paired data from real-world is challenging. To address this, several methods adopt unsupervised learning strategies, like SCI (Ma et al. 2022) designing a self-calibration mechanism, Zero-DCE (Guo et al. 2020) and Zero-DCE++ (Li, Guo, and Loy 2021) learning adaptive gamma correction curves, UDNet (Saleh et al. 2025) exploring color uncertainty to optimize color distributions and Lighten Diffusion (Jiang et al. 2024).

In addition, some methods adopt unpaired supervised learning. EnlightenGAN (Jiang et al. 2021) pioneers the use of adversarial learning on low-loght enhancement. Cycle-Retinex (Wu et al. 2024) introduces a cycle consistency loss (Zhu et al. 2017) to ensure detail preservation. SelfEnNet (Kar et al. 2024) proposes a transformation-consistent self-supervised mechanism combined with an unpaired self-conditioning strategy. However, lack of the guidance of daytime static background image, thus these methods only achieve suboptimal enhancement performance.

Night-to-Day Domain Translation

The pioneering work of style transfer StyleGAN (Karras, Laine, and Aila 2019) has started the era of neural style transferring. Since then, lots of works are proposed (Karras et al. 2020, 2021; Wang, Zhao, and Xing 2023) and achieve impressive successes. Night-to-Day is one type of style transferring, translating a nighttime image to daytime. ToDayGAN (Anoosheh et al. 2019) first introduces the use of cycle-consistency loss to preserve content information during translation, and ForkGAN (Zheng et al. 2020) further propose a fork-shaped encoder with perceptual consistency in the latent feature space to enhance structural fidelity. AU-GAN (gi Kwak et al. 2022) incorporates uncertainty estimation to mine informative features from night images, while Fan et al. (Fan et al. 2023) leverages cross-frequency relational knowledge to simplify the Night2Day pipeline. Some studies (Jeong et al. 2021; Kim et al. 2022), impose structural constraints through manual annotations. N2D3 (Lan et al. 2024) proposes leveraging illumination-aware priors by partitioning the night image into distinct regions . However, in some practical scenarios, the objective is not to enhance generative diversity, but rather to perform faithful day-to-night translation of a fixed scene. In these cases, the preservation of essential scene content becomes crucial yet is often neglected

Methods

As discussed above, nighttime image enhancement tasks could benefit significantly from the daytime static background cues, which improve luminance distribution and structural fidelity. Firstly, we extract clean and static background image from daytime videos. There are several statistic-based background modeling methods. For example, the classical Robust Principal Component Analysis (RPCA)

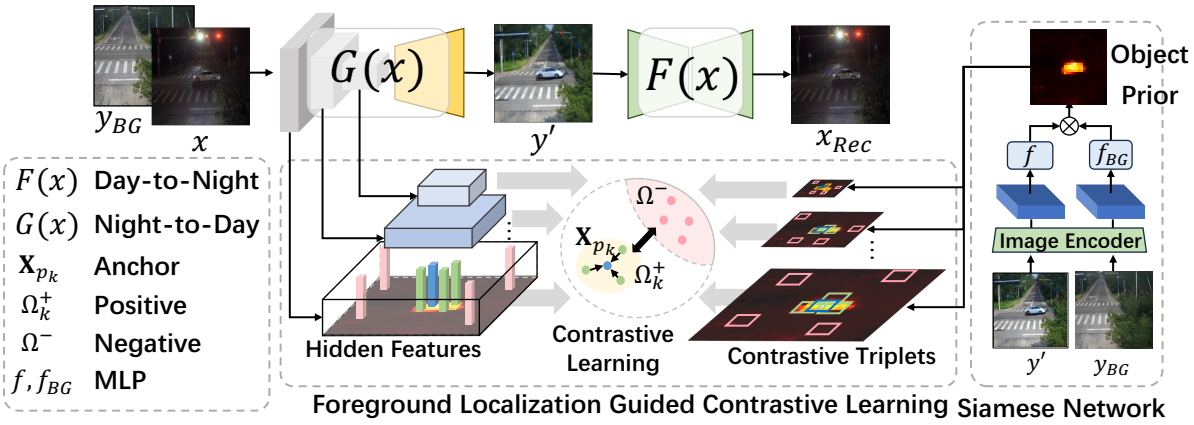


Figure 2: Overview of the DMGNIE framework. The night image x and background prior Y_{BG} are fed into network G for enhancement.

method (Candès et al. 2011) decomposes image content via iterative low-rank approximation, demonstrating robust performance in standard scenarios. However, in high-resolution complex scenes, e.g., dense traffic flows, RPCA suffers from computational inefficiency and motion artifacts due to its reliance on global optimization. To address these limitations, we propose a semantic prior-based SegBMN module, leveraging the segmentation mask to eliminate impacts of dynamic foreground objects on background. Then, we propose DMGNIE framework based on CycleGAN, as Fig. 2 shows, leveraging the daytime background guiding the enhancement of nighttime images. This design leverages cycle-consistent adversarial learning to construct an implicit day-night image mapping, thereby eliminating dependencies on strictly aligned datasets. However, the guidance of daytime background priors may interference to the feature of foreground objects, resulting in degradation of the enhanced output. To mitigate this issue, we propose a Foreground Localization-Guided Contrastive Learning mechanism. Inspired by the demonstrated advantages of contrastive learning in feature discrimination and cross-domain correlation modeling (Park et al. 2020), our method explicitly increases the discrepancy between foreground and background features. This design significantly improves the semantic consistency of foreground object representations and enhances the visual quality of nighttime scenes, particularly under challenging low-light conditions.

Preliminaries

Given two sets of images from different domains, $X = \{x_i\}_{i=1}^{M} \in \mathcal{X}$ and $Y = \{y_i\}_{i=1}^{N} \in \mathcal{Y}$, where X denotes nighttime low-light images and Y denotes daytime well-lit images from the same scenes, with $x_i, y_i \in \mathbb{R}^{H\times W\times 3}$, our goal is to learn a mapping function $G : \mathcal{X} \xrightarrow{y_{BG}} \mathcal{Y}$, implementing *Night-to-Day* transformation, where the Y_{BG} is the static daytime background images.

Cycle-Consistency Constrain. Cycle-Consistency Constrain is proposed from CycleGAN (Zhu et al. 2017), which adopts two generators and two discriminators, one generator $G : X \to Y$ and another one $F : Y \to X$, while

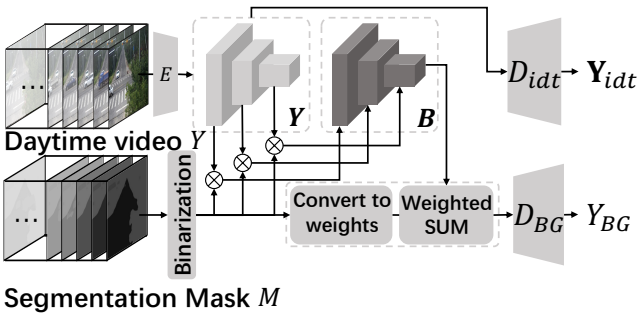


Figure 3: Demonstration of the semantic-guided SegBMN module. Inpus are daytime video frames and corresponding masks, and the Y_{BG} is the background image.

two discriminators separately discriminate the image distributions of the source and target domains, that is D_G and D_F. Here we mainly introduce the definition of cycle-consistency constrain. Mathematically, for a batch of input nighttime images $\{x_i\}_{i=1}^{M}$ and daytime images $\{y_j\}_{j=1}^{N}$, the cycle-consistency loss is defined as:

$$\begin{aligned}\mathcal{L}_{\text{cycle}} &= \mathbb{E}_{x\sim P_{\mathcal{X}}}\left[\|F(G(x)) - x\|_1\right] \\ &+ \mathbb{E}_{y\sim P_{\mathcal{Y}}}\left[\|G(F(y)) - y\|_1\right]\end{aligned} \tag{1}$$

where, $\|\cdot\|_1$ denotes the L_1-norm. $F(G(x))$ represents the forward cycle and $G(F(y))$ represents the backward cycle.

Semantic-guided for Fast Background Modeling

To address these limitations of conventional background modeling methods, such as RPCA (Candès et al. 2011), IncPCP (Chau and Rodríguez 2017), in real traffic surveillance videos, we introduce scene semantic priors provided by semantic segmentation methods (Chen et al. 2023; Choe et al. 2024) to reduce the interference of foreground objects on the background modeling process. Motivated by this, we propose a background modeling method based on Unified Network (UNet), named SegBMN, as Fig. 3 shows.

First, a pre-trained semantic segmentation model f_{seg} is adopted to identify foreground objects such as cars, pedestrians, and motorcycles from daytime video Y, like

$M = f_{\text{seg}}(Y)$. The mask M is a $\{0,1\}$-binary mask, where 0 denotes the foreground regions and 1 represents the background. The inputs of SegBMN are K daytime images $\mathbf{Y}$, and their semantic mask $\mathbf{M}$. Through a pyramid downsampling of L stages, we extract multi-scale feature maps $[\mathbf{Y}^{(1)}, \cdots, \mathbf{Y}^{(L)}] = E(\mathbf{Y})$. Then the masks $\mathbf{M}$ are interpolated and resized to match each scale, obtaining $[\mathbf{M}^{(1)}, \cdots, \mathbf{M}^{(L)}]$. Foreground features are suppressed via element-wise multiplication $\mathbf{B}^{(l)} = \mathbf{Y}^{(l)} \odot \mathbf{M}^{(l)}, l \in [1, \cdots L]$. The filtered background features are aggregated across the K images by a weighted summation:

$$Y_{BG}^{(l)} = \sum_{j=1}^{K} \frac{\mathbf{B}_j^{(l)} \odot \mathbf{M}_j^{(l)}}{\sum_{i=1}^{K} \mathbf{M}_i^{(l)}} \tag{2}$$

The multi-scale background features are decoded via an upsampling path to reconstruct the background image $Y_{BG} = D_{BG}\left(Y_{BG}^{(1)}, \cdots, Y_{BG}^{(L)}\right)$. Furthermore, to encourage the encoder $E(\cdot)$ encoding as much visual information as possible and do so accurately, we introduce another decoder branch for identical output $\mathbf{Y}_{idt} = D_{idt}\left(\mathbf{Y}^{(1)}, \cdots, \mathbf{Y}^{(L)}\right)$ and a reconstruction-guided loss that penalizes missing or distorted content. The overall training loss is defined as

$$\mathcal{L}_{BG} = \|Y_{BG} - \mathbf{Y} \odot \mathbf{M}\|_1 + \alpha \|\mathbf{Y}_{idt} - \mathbf{Y}\|_1. \tag{3}$$

which enforces both accurate background reconstruction and better encoding features.

Foreground Localization-guided Contrastive Learning

Previously, we use the SegBMN to extract clean background from daytime video streams. Then, the extracted background as memory guides the enhancement transformation of model, enabling the enhanced output to better align with the distribution of daytime images. Specifically, we concatenate the background image Y_{BG} with the nighttime image X along the channel dimension and feed them into an enhancement network.

However, based on observations of the experimental results, we found that the outputs of the enhancement network often exhibits significant color deviations in foreground objects (such as vehicles). We attribute this issue primarily to the interference caused by the introduction of background priors, which hinders the model's ability to effectively learn and extract foreground features. As a result, the model struggles to capture accurate foreground semantic information, ultimately compromising the consistency and accuracy of the enhancement process in semantic.

To address the above issues, we propose a foreground localization-guided contrastive learning mechanism. Background information not only serves as a valuable prior for nighttime image enhancement but also facilitates the identification of foreground object regions based on spatial structural differences. Building upon this, we first design a Siamese network (see right part in Fig. 2) to generate a 2D probability map that estimates the spatial location of foreground regions. With the estimated foreground positions, we then construct positive and negative sample pairs between foreground and background areas. Through contrastive learning, the model is encouraged to amplify feature discrepancies between foreground and background regions while simultaneously enhancing intra-foreground feature consistency. In the following sections, we first introduce how the Siamese network is employed for foreground localization, and then elaborate on the contrastive learning strategy built upon the localization results.

Foreground Objects Location. For better locating the foreground objects in daytime images, we propose a lightweight yet effective Siamese architecture. The inputs of this network are the enhanced nighttime image $Y' = G(X, Y_{\text{BG}})$ and daytime background image Y_{BG} that belongs to the same scene. Thus, feature vectors in the background regions of both should exhibit high similarity, while those in the foreground regions of both are expected to differ. However, directly computing the difference often results in suboptimal performance as the Y' is generated, which leads some slight noises and distorts existing in Y'. To achieve more accurate localization, we aim to enhance the model's robustness to such noise. Specifically, we introduce noise to both images, producing two additional noisy counterparts. All four images are then encoded using a shared pre-trained encoder E. The resulting feature representations are subsequently projected into a common space using two learnable projectors, denoted as f and f_{BG}. Mathematically, we have $\mathbf{Y}' = (f \circ E)(Y'), \hat{\mathbf{Y}}' = (f \circ E)(Y' + z), \mathbf{Y}_{\text{BG}} = (f_{\text{BG}} \circ E)(Y_{\text{BG}}), \hat{\mathbf{Y}}_{\text{BG}} = (f_{\text{BG}} \circ E)(Y_{\text{BG}} + z)$, where $z \sim \mathcal{N}(\mu, \sigma)$ is gaussian noise, and $\circ$ is functions composition.

Then, we define an inner product operator $\otimes$ which performs the inner product of two vectors at the same position on the two input feature maps to obtain a 2D difference map. We have $D_{orig} = E(Y') \otimes E(Y_{\text{BG}}), D_{proj} = \mathbf{Y}' \otimes \mathbf{Y}_{\text{BG}}, D_Y^{self} = \mathbf{Y}' \otimes \hat{\mathbf{Y}}', D_{BG}^{self} = \mathbf{Y}_{\text{BG}} \otimes \hat{\mathbf{Y}}_{\text{BG}}$. Finally, the loss for training these two projection functions can be defined as follows

$$\begin{aligned} \mathcal{L}_{\text{FGMap}} = & KL(D_{proj} \,\|\, D_{orig}) + \beta_1 f_{\text{entropy}}(D_{proj}) \\ & - \beta_2 \left(f_{\text{entropy}}(D_Y^{self}) + f_{\text{entropy}}(D_{BG}^{self}) \right) \\ & + \beta_3 f_{TV}(D_{proj}), \end{aligned} \tag{4}$$

where the second term makes D_{proj} be more sparse, the third and fourth term for robust to noise and the final term for smooth. And we use D_{proj} for locating the foreground objects in the next stage.

Foreground-Background Contrastive Learning. To enhance semantic consistency within the foreground, a contrastive learning mechanism guided by foreground localization is introduced. Specifically, to estimate the spatial distribution of foreground regions in nighttime images, the aforementioned Siamese network is leveraged firstly to obtain a 2D foreground probability distribution map $D_X = \text{Siamese}(G(X, Y_{BG}), Y_{BG})$ with a nighttime image X and its corresponding daytime background image Y_{BG}. Then, we extract latent features of X via the encoder of the generator G as $\mathbf{X}^{(l)} = G(X, Y_{BG}; \theta_E^{(l)})$, where $\theta_E^{(l)}$ denotes the

Dataset Name	#Scenes	#Train Images			#Valid Images			#Test Images		
		#Scenes	Daytime	Nighttime	#Scenes	Daytime	Nighttime	#Scenes	Daytime	Nighttime
LQ-TLL-QD	16	14	62,307	55,865	1	21,391	5,980	1	25,073	5,980
LQ-TLL-SZ	8	6	25,019	23,942	1	2,147	2,971	1	1,905	2,971
LQ-TELL	9	7	20,981	18,806	1	1,849	2,351	1	3,175	2,351

Table 1: The statistics of the three datasets, Low Quality Traffic Low-Light in QingDao (LQ-TLL-QD), Low Quality Traffic Low-Light in ShenZhen (LQ-TLL-SZ) and Low Quality Traffic Extreme Low-Light (LQ-TELL).

encoder parameters of the l^{th} layer.

For each latent features of $\mathbf{X}^{(l)}$ of the l^{th} encoder layer, (denoted as $\mathbf{X}$ for short), we sample K anchor patches located in the foreground region according to the probability distribution D_X as $\{\mathbf{X}_{p_k}\}_{k=1}^{K}, p_k \sim D_X$, where p_k is the 2D spatial position. Then, to enforce semantic consistency between adjacent features , for each anchor patch $\mathbf{X}_{p_k}$, we sample M positive patches within a local spatial neighborhood of radius r, constructing the positive sample set

$$\Omega_k^+ = \{\mathbf{X}_{p_k}^{(1)}, \cdots, \mathbf{X}_{p_k}^{(M)}\}. \tag{5}$$

In addition, we can derive a complementary background probability distribution based on the D_X as $\hat{D}_X = \text{Normalize}(-D_X)$. We then sample N global negative patches from the background regions, which are shared for all anchor patches

$$\Omega^- = \{\mathbf{X}_{q_1}, \cdots, \mathbf{X}_{q_N}\}, q_n \sim \hat{D}_X. \tag{6}$$

where q_n is 2D spatial position. With the above, we construct a set of contrastive triplets as follow

$$\mathcal{T} = \{\mathbf{X}_{p_k}, \Omega_k^+, \Omega^-\}. \tag{7}$$

We adopt an InfoNCE loss (van den Oord, Li, and Vinyals 2018) to minimize intra-foreground distances and maximize the inter-region discrepancy between foreground and background features. This design encourages stronger semantic coherence within the foreground and better discrimination from the background. The loss formulation is as follows:

$$\mathcal{L}_{\text{InfoNCE}} = \sum_{k=1}^{K} \left(\mathbb{E}_{\mathbf{X}^+ \sim \Omega_k^+} \left[\frac{\text{sim}(\mathbf{X}_{p_k}, \mathbf{X}^+)/\tau}{\mathbb{E}_{\mathbf{X}^- \sim \Omega^+} [\text{sim}(\mathbf{X}_{p_k}, \mathbf{X}^-)/\tau]} \right] \right) \tag{8}$$

where the similarity function $\text{sim}(\cdot, \cdot)$ is inner product.

Model Optimization

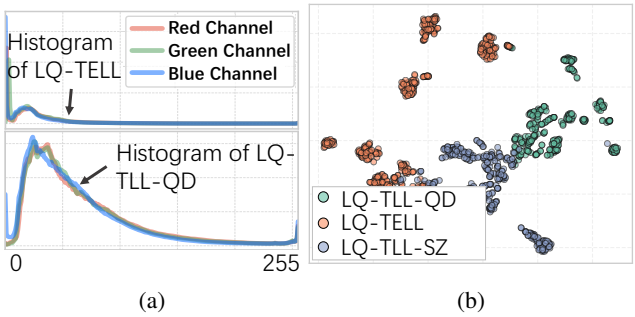
Our training process consists of two stages. In the first stage, we train the SegBMN model with Eq. 3 to get daytime background image. In the second stage, we train the generators G and F along with their corresponding discriminators, as well as the foreground localization-guided contrastive learning module. The adversarial loss is defined as follows

$$\begin{aligned} \mathcal{L}_{GAN}^{X \to Y} &= \mathbb{E}_{X \in \mathcal{X}}[\log(1 - D_G(G(X)))] + \mathbb{E}_{Y \in \mathcal{Y}}[D_G(Y)], \\ \mathcal{L}_{GAN}^{Y \to X} &= \mathbb{E}_{Y \in \mathcal{Y}}[\log(1 - D_F(F(X)))] + \mathbb{E}_{X \in \mathcal{X}}[D_F(X)]. \end{aligned} \tag{9}$$

The final loss function in the second stage training is

$$\begin{aligned} \mathcal{L} = \mathcal{L}_{GAN}^{X \to Y} + \mathcal{L}_{GAN}^{Y \to X} + \gamma_1 \mathcal{L}_{\text{cycle}} + \gamma_2 \mathcal{L}_{FGMap} \\ + \gamma_3 \mathcal{L}_{\text{InfoNCE}} \end{aligned} \tag{10}$$

where see Appendix A.1 for the definition of $\mathcal{L}_{\text{cycle}}$ loss.



(a) (b)

Figure 4: Visualization of the statistics of the LQ-TLL-QD, LQ-TLL-SZ and LQ-TELL datasets. (a) Pixel intensity distribution of LQ-TLL-QD and LQ-TELL datasets. (b) 2D visualization of image histograms from the three datasets with T-SNE (Maaten and Hinton 2008)

Datasets Description

We collect road intersections surveillance video streams deployed in two cities: Qingdao and Shenzhen. The collected data includes both daytime and nighttime images and are unpaired[1]. Based on this, we construct a unpaired nighttime image enhancement dataset *Low-Quality Traffic Low-Light (LQ-TLL)* for real road intersections, which is divided into three subsets: *LQ-TLL-QD*, *LQ-TLL-SZ* and *Extremely Low-Light Subset (LQ-TELL)*, according to brightness statistics and locations. As Fig. 4 (b) shows, the distributions of LQ-TELL are distinct from others, and Fig. 4 (a) shows the specific difference on histograms. To evaluate the generalization ability of models cross different road scene, the road intersections in training set and testing set are different, see Tab. 1. See Appendix B for more details.

Our datasets offer a challenging benchmark for nighttime image enhancement under real-world low-light conditions. Unlike LOL (Chen et al. 2018b; Yang et al. 2020), SID (Chen et al. 2018a), BDD100K (Yu et al. 2020) and Alderley (Milford and Wyeth 2012), our dataset includes extremely dark rural roads and uses static surveillance views instead of dashcams, providing a unique perspective for traffic scene understanding.

Experiments

Baselines and Metrics Descriptions

We evaluate our proposed model DMGNIE on the LQ-TLL dataset by comparing it with state-of-the-art unsupervised and supervised baseline methods. For unsupervised

[1]"unpaired" means there is no ground truth y for each x.

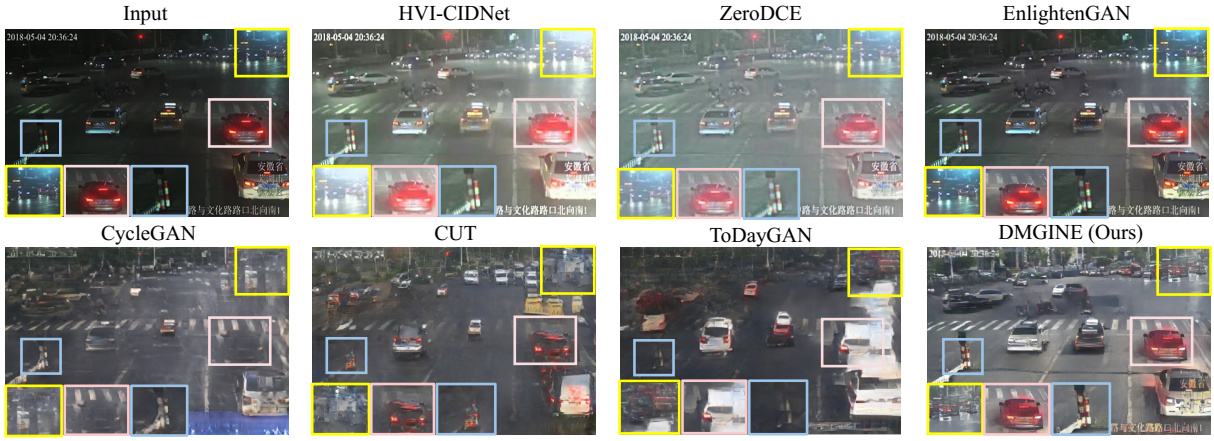


Figure 5: The qualitative comparison results on the LQ-TELL dataset.

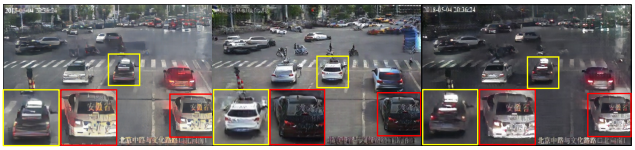
approaches, we include illumination enhancement models such as UDNet (Saleh et al. 2025), SCI (Ma et al. 2022), LightenDiffusion (Jiang et al. 2024), Zero-DCE (Guo et al. 2020), and Zero-DCE++ (Li, Guo, and Loy 2021). In the unpaired unsupervised category, we consider models including EnlightenGAN (Jiang et al. 2021), Cycle Retinex (Wu et al. 2024), SelfEnNet (Kar et al. 2024), ToDayDAN (Anoosheh et al. 2019), CUT (Park et al. 2020), and CycleGAN (Zhu et al. 2017). For supervised baselines, we evaluate HVI-CIDNet (Yan et al. 2025), URetinexNet (Wu et al. 2022), and Retinexformer (Cai et al. 2023). Since our LQ-TLL dataset is unpaired, we finetune the pretrained weights of the supervised models with our dataset to enable fair comparison.

To quantitatively evaluate the performance of both our model and the baseline methods, we adopt two widely-used no-reference perceptual quality metrics, NIQE (Mittal, Soundararajan, and Bovik 2012) and BRISQUE (Mittal, Moorthy, and Bovik 2012), to assess the perceptual quality of enhanced images. In addition, to determine if a model effectively transforms images from the night to the day, we employ the LPIPS (Zhang et al. 2018) and FID (Heusel et al. 2017) metrics. We provide more experimental results of DMGNIE in Appendix C.1 and ablation analysis of SegBMN and FLGCL in Appendix C.2 and C.3.

Comparison with Selected Baselines

As Tab. 2 shows, our method on LQ-TLL-QD, LQ-TLL-SZ and LQ-TELL datasets outperforms baselines on NIQE, FID and LPIPS metrics and achieves comparable performance on BRISQUE. As shown in Fig. 5, compared with existing LLIE baselines, our method achieves better approximation for illumination of daytime road image. Our method effectively suppresses the interference caused by overexposed regions (as highlighted in yellow box), and improve the restoration of road surfaces.

Compared with generative methods, our method also shows significant advantages in image fidelity. For road areas, our method focuses on lighting restoration while preserving fine-grained texture details. As shown in Figure 5, our method outperforms existing generative baseline methods in preserving object information. For the pink box, the vehicle is almost completely lost in CycleGAN, CUT changes the structure of the vehicle, and TodayGAN changes the color of the car from black to white. In contrast, our method preserves both the structure and color information of the vehicle. In addition, in the blue box, our method can preserve the structure and color of small objects. These comparisons highlight the fidelity and robustness of our method. The experimental settings and hyperparameters are listed in Appendix C.4.



(a) DMGINE (b) DMGINE w/o FLGCL (c) DMGINE w/o Scene Prior

Figure 6: Illustration of ablation results of DMGNIE. (a) The output of DMGNIE. (b) Result of ablating the FLGCL module. (c) Result of ablating the background prior.

Ablation Study

The Effectiveness of Foreground Localization-guided Contrastive Learning. To explore the effectiveness of FLGCL module on preserving the information of foreground objects, we ablate it from the DEGINE. As shown in Fig. 6(b), although the overall visual quality of the image shows a slight improvement, the preservation of foreground object information is notably poor. Specifically, in the yellow box, the original yellow-green taxi is incorrectly enhanced into a white car; meanwhile, in the red box, a white car is transformed into a black one. We attribute this issue to the multi-scale contrastive learning can facilitate the preservation for the foreground, preventing from interference of daytime background.

Furthermore, since DMGNIE performs contrastive learning at multiple scales, we analyze the impact of the contrastive learning with different scales on performance. We conduct ablation experiments with five scales: S1: 256, S2: (256, 128), S3: (256, 128, 64), S4: (256, 128, 64, 32) and

Methods	LQ-TLL-QD				LQ-TLL-SZ				LQ-TELL			
	BRISQUE↓	NIQE↓	FID↓	LPIPS↓	BRISQUE↓	NIQE↓	FID↓	LPIPS↓	BRISQUE↓	NIQE↓	FID↓	LPIPS↓
	Unsupervised Learning Methods											
UDNet	36.495	24.084	140.986	64.545	24.834	22.483	171.488	57.882	38.120	25.292	224.412	70.104
SCI	37.342	26.422	125.571	55.831	26.896	**21.153**	167.329	40.612	26.469	25.489	241.913	67.198
LightenDiffusion	35.258	25.132	127.872	56.637	20.967	25.745	162.934	41.692	40.605	26.399	305.927	69.123
ZeroDCE	34.603	25.489	133.391	55.037	21.141	21.826	187.717	43.281	34.910	26.308	235.222	67.863
ZeroDCE++	37.938	25.991	123.041	55.712	20.813	23.746	189.371	39.218	34.206	25.791	229.401	59.433
	Unpaired Unsupervised Learning Methods											
SelfEnNet	74.539	26.977	351.091	91.559	57.917	26.9492	451.399	83.301	48.870	26.974	347.014	66.914
EnlightenGAN	34.979	26.492	96.518	69.431	33.049	22.997	156.361	62.613	34.382	26.004	230.490	59.338
Cycle Retinex	68.282	26.866	347.735	89.518	35.915	23.471	310.931	61.802	22.232	26.676	445.280	74.264
CycleGAN	27.895	22.816	75.103	47.037	19.376	21.161	**87.753**	44.017	33.852	25.183	183.573	52.884
CUT	**26.478**	**21.826**	83.268	47.769	17.730	21.700	202.720	60.058	**21.094**	23.894	162.743	59.646
ToDayGAN	26.653	22.834	93.307	55.991	**16.443**	21.822	104.258	53.759	28.507	24.422	157.999	57.692
	Supervised Learning Methods											
HVI-CIDNet	26.889	26.055	227.544	54.44	19.727	23.801	151.430	34.887	40.425	27.087	374.456	66.651
HVI-CIDNet*	33.002	25.148	108.119	53.769	23.796	21.785	161.952	42.062	30.158	25.088	163.282	57.013
URetinexNet	32.393	25.898	117.056	56.058	20.635	22.787	159.932	42.663	32.078	25.980	175.247	58.652
URetinexNet*	29.724	25.719	179.673	56.057	20.165	24.675	171.939	44.449	33.921	26.879	276.745	63.025
Retinexformer	33.677	25.285	114.712	54.743	27.423	21.339	198.843	47.694	30.871	25.922	159.928	57.619
Retinexformer*	33.082	26.362	108.891	57.966	24.476	25.374	222.079	49.447	30.579	26.139	167.506	61.642
Ours	31.422	22.473	**43.181**	**42.961**	20.487	22.321	92.991	**31.441**	26.823	**23.612**	**55.505**	**45.244**

Table 2: Comparison with baselines. We compare our methods across LQ-TLL-QD, LQ-TLL-SZ and LQ-TELL datasets on common metrics like BRISQE, NIQE, FID and LPIPS. ↓ shows lower result is better. The methods name with "*" means the results are finetuned on LQ-TLL datasets. The best result is **bolded**, and the second best result is underlined

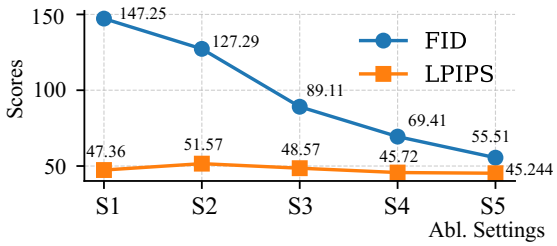


Figure 7: Impacts of performing contrastive learning of different scale on performance.

S5: (256, 128, 64, 32, 32). As Fig. 7 shows, when contrastive learning is applied only to shallow, large-scale feature maps, the results are suboptimal and even weaken effective feature encoding. As the depth of feature maps increases, contrastive learning gradually demonstrates its advantage in preserving foreground object information, leading to significant improvements in FID and LPIPS scores. We attribute this to that deeper layers capture more semantic information, and the model tends to fuse nighttime images with background prior at these stages. Without semantic constraints from contrastive learning, foreground features can be interfered by background information, ultimately degrading the enhancement quality.

The Effectiveness of Daytime Memory Guidance. Furthermore, to investigate the effectiveness of the background prior in the enhancement process, we conduct a study by removing the daytime memory guidance. As illustrated in Fig. 6 (c), the enhanced image exhibits an illumination style that shifts toward night, indicating that the background prior primarily provides the essential daytime lighting condition cues. This highlights its critical role in guiding the enhancement toward realistic and visually consistent results.

Conclusion

To address the limitations (e.g., overexposure and noise amplification) of existing LLIE methods in enhancing nighttime images in real traffic roads, we consider using daytime videos of the same scene captured by surveillance cameras as memories to guide nighttime image enhancement, and propose the DMGNIE framework. Specifically, we aim to leverage daytime scene to facilitate the restoration for nighttime image on structures and lighting conditions. Therefore, we propose a SegBMN module to extract static background information from the daytime video streams and use this information as a daytime prior to guide the enhancement model. Secondly, to preserve foreground object information of input, we propose a Foreground Localization-guided Contrastive Learning module to ensure the consistency of these objects. Finally, to evaluate the DMGNIE, we collect a large amount of surveillance data and construct the unpaired LQ-TLL dataset. Both quantitative and qualitative experimental results demonstrate the effectiveness of our approach. However, the current datasets (especially LQ-TELL) present significant challenges, such as extreme low-light conditions and severe noise. Future work will explore more robust priors and advanced architectures to address these limitations.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China: 62236008, 62441232, 62472238 and 62441232, and supported in part by Project of Peng Cheng Laboratory (PCL2025A14)

References

Anoosheh, A.; Sattler, T.; Timofte, R.; Pollefeys, M.; and Van Gool, L. 2019. Night-to-day image translation for retrieval-based localization. In *2019 International conference on robotics and automation (ICRA)*, 5958–5964. IEEE.

Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; and Zhang, Y. 2023. Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 12504–12513.

Candès, E. J.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3): 1–37.

Chau, G.; and Rodríguez, P. 2017. Panning and jitter invariant incremental principal component pursuit for video background modeling. In *Proceedings of the IEEE international conference on computer vision workshops*, 1844–1852.

Chen, C.; Chen, Q.; Xu, J.; and Koltun, V. 2018a. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3291–3300.

Chen, J.; Lu, J.; Zhu, X.; and Zhang, L. 2023. Generative Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7111–7120.

Chen, W.; Wenjing, W.; Wenhan, Y.; and Jiaying, L. 2018b. Deep Retinex Decomposition for Low-Light Enhancement. In *British Machine Vision Conference*.

Choe, S.-A.; Shin, A.-H.; Park, K.-H.; Choi, J.; and Park, G.-M. 2024. Open-Set Domain Adaptation for Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 23943–23953.

Fan, Z.; Wu, X.; Chen, X.; and Li, Y. 2023. Learning to See in Nighttime Driving Scenes with Inter-frequency Priors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 4218–4225.

gi Kwak, J.; Jin, Y.; Li, Y.; Yoon, D.; Kim, D.; and Ko, H. 2022. Adverse Weather Image Translation with Asymmetric and Uncertainty-aware GAN. arXiv:2112.04283.

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

Han, G.; and Lim, S.-N. 2024. Few-Shot Object Detection with Foundation Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 28608–28618.

Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Jeong, S.; Kim, Y.; Lee, E.; and Sohn, K. 2021. Memory-guided Unsupervised Image-to-image Translation . In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6554–6563. Los Alamitos, CA, USA: IEEE Computer Society.

Jiang, H.; Luo, A.; Liu, X.; Han, S.; and Liu, S. 2024. LightenDiffusion: Unsupervised Low-Light Image Enhancement with Latent-Retinex Diffusion Models. In *European Conference on Computer Vision*.

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

Jobson, D.; Rahman, Z.; and Woodell, G. 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.

Kar, A.; Dhara, S. K.; Sen, D.; and Biswas, P. K. 2024. Self-Supervision via Controlled Transformation and Unpaired Self-Conditioning for Low-Light Image Enhancement. *IEEE Transactions on Instrumentation and Measurement*, 73: 1–13.

Karras, T.; Aittala, M.; Laine, S.; Härkönen, E.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2021. Alias-free generative adversarial networks. *Advances in neural information processing systems*, 34: 852–863.

Karras, T.; Laine, S.; and Aila, T. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4401–4410.

Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8110–8119.

Kim, S.; Baek, J.; Park, J.; Kim, G.; and Kim, S. 2022. InstaFormer: Instance-Aware Image-to-Image Translation with Transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18321–18331.

Lan, G.; Yang, Y.; Wang, Z.; Wang, D.; Zhao, B.; and Li, X. 2024. Night-to-Day Translation via Illumination Degradation Disentanglement. *arXiv preprint arXiv:2411.14504*.

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

Li, C.; Guo, C. G.; and Loy, C. C. 2021. Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

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

Ma, L.; Ma, T.; Liu, R.; Fan, X.; and 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*, 5637–5646.

Maaten, L. v. d.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov): 2579–2605.

Milford, M. J.; and Wyeth, G. F. 2012. SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. In *2012 IEEE International Conference on Robotics and Automation*, 1643–1649.

Mittal, A.; Moorthy, A. K.; and Bovik, A. C. 2012. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing*, 21(12): 4695–4708.

Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3): 209–212.

Park, T.; Efros, A. A.; Zhang, R.; and Zhu, J.-Y. 2020. Contrastive Learning for Unpaired Image-to-Image Translation. In *European Conference on Computer Vision*.

Pizer, S. M.; Amburn, E. P.; Austin, J. D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J. B.; and Zuiderveld, K. 1987. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3): 355–368.

Poynton, C. 2012. *Digital Video and HD: Algorithms and Interfaces*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2 edition. ISBN 9780123919267.

Saleh, A.; Sheaves, M.; Jerry, D.; and Rahimi Azghadi, M. 2025. Adaptive deep learning framework for robust unsupervised underwater image enhancement. *Expert Systems with Applications*, 268: 126314.

Sun, Z.; Li, J.; and Mu, Y. 2024. Exploring Orthogonality in Open World Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17302–17312.

van den Oord, A.; Li, Y.; and Vinyals, O. 2018. Representation Learning with Contrastive Predictive Coding. *CoRR*, abs/1807.03748.

Wang, S.; Zheng, J.; Hu, H.-M.; and Li, B. 2013. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. *IEEE Transactions on Image Processing*, 22(9): 3538–3548.

Wang, Z.; Zhao, L.; and Xing, W. 2023. Stylediffusion: Controllable disentangled style transfer via diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*, 7677–7689.

Wu, K.; Huang, J.; Ma, Y.; Fan, F.; and Ma, J. 2024. Cycle-Retinex: Unpaired Low-Light Image Enhancement via Retinex-Inline CycleGAN. *IEEE Transactions on Multimedia*, 26: 1213–1228.

Wu, W.; Weng, J.; Zhang, P.; Wang, X.; Yang, W.; and Jiang, J. 2022. URetinex-Net: Retinex-Based Deep Unfolding Network for Low-Light Image Enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5901–5910.

Yan, Q.; Feng, Y.; Zhang, C.; Pang, G.; Shi, K.; Wu, P.; Dong, W.; Sun, J.; and Zhang, Y. 2025. Hvi: A new color space for low-light image enhancement. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 5678–5687.

Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; and Liu, J. 2020. From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3060–3069.

Yu, F.; Chen, H.; Wang, X.; Xian, W.; Chen, Y.; Liu, F.; Madhavan, V.; and Darrell, T. 2020. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2636–2645.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.

Zheng, Z.; Wu, Y.; Han, X.; and Shi, J. 2020. ForkGAN: Seeing into the Rainy Night. In *The IEEE European Conference on Computer Vision (ECCV)*.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *IEEE International Conference on Computer Vision (ICCV)*.