

GLARE: Low Light Image Enhancement via Generative Latent Feature based Codebook Retrieval

Han Zhou^{1,*}, Wei Dong^{1,*}, Xiaohong Liu^{2,†}, Shuaicheng Liu³,
Xiongkuo Min², Guangtao Zhai², and Jun Chen^{1,†}

¹ McMaster University, ² Shanghai Jiao Tong University,

³ University of Electronic Science and Technology of China

{zhouh115, dongw22, chenjun}@mcmaster.ca liushuaicheng@uestc.edu.cn

{xiaohongliu, minxiongkuo, zhaiguangtao}@sjtu.edu.cn

*Equal Contribution †Corresponding Authors

Abstract. Most existing Low-light Image Enhancement (LLIE) methods either directly map Low-Light (LL) to Normal-Light (NL) images or use semantic or illumination maps as guides. However, the ill-posed nature of LLIE and the difficulty of semantic retrieval from impaired inputs limit these methods, especially in extremely low-light conditions. To address this issue, we present a new LLIE network via Generative LAtent feature based codebook REtrieval (**GLARE**), in which the codebook prior is derived from undegraded NL images using a Vector Quantization (VQ) strategy. More importantly, we develop a generative Invertible Latent Normalizing Flow (I-LNF) module to align the LL feature distribution to NL latent representations, guaranteeing the correct code retrieval in the codebook. In addition, a novel Adaptive Feature Transformation (AFT) module, featuring an adjustable function for users and comprising an Adaptive Mix-up Block (AMB) along with a dual-decoder architecture, is devised to further enhance fidelity while preserving the realistic details provided by codebook prior. Extensive experiments confirm the superior performance of GLARE on various benchmark datasets and real-world data. Its effectiveness as a preprocessing tool in low-light object detection tasks further validates GLARE for high-level vision applications. Code is released at <https://github.com/LowLevelAI/GLARE>.

Keywords: Generative feature alignment · Adaptive feature transformation · Codebook Retrieval · Low-light image enhancement

1 Introduction

Low light images often suffer from various degradations, including loss of details, reduced contrast, amplified sensor noise, and color distortion, making many downstream tasks challenging, such as object detection, segmentation, or tracking [12, 15, 22, 26, 66]. Consequently, LLIE has been extensively studied recently.

Traditional techniques that leverage handcrafted priors and constraints [4, 20, 27, 29, 78] have made significant contributions to this field. However, these

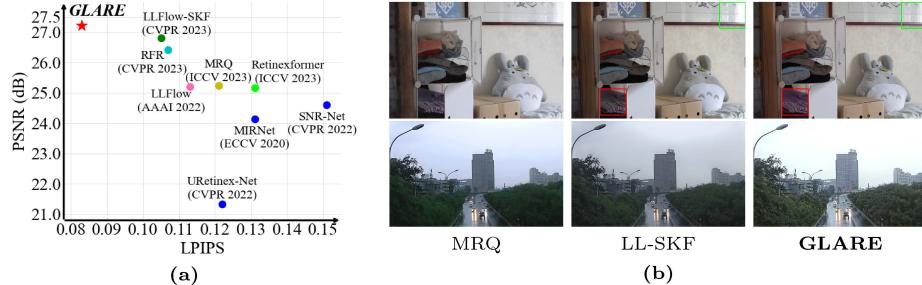


Fig. 1: (a) GLARE significantly outperforms SOTA methods on LOL [62]. (b) GLARE can generate appealing results on both LOL (upper) and real-world (lower) images.

methods still exhibit limitations in terms of adaptability across diverse illumination scenarios [56]. With the rapid advancements in deep learning, extensive approaches have been employed to learn complex mappings from LL to NL images [5, 42, 49, 70]. Although their performance surpasses that of traditional methods, once deployed in real-world scenarios with varying light conditions and significant noise, these methods tend to produce visually unsatisfactory results.

Recent methods utilize semantic priors [17, 65, 78], extracted feature [30, 51, 69, 80], and illumination maps [56] as the guidance to tackle the uncertainty and ambiguity of ill-posed LLIE problem. However, they still face challenges in extracting stable features from heavily degraded inputs, which are often overwhelmed by noise and obfuscated by low visibility. Besides, only utilizing the extracted information from degraded images to build the LL-NL transformation usually generate unsatisfactory results when testing on real-world scenarios.

To generate realistic and appealing outcomes, one possible solution is to exploit the prior knowledge of natural normal-light images. Therefore, we propose to leverage a learned Vector-Quantized (VQ) codebook prior that captures the intrinsic features of high-quality and well-lit images, to guide the learning of LL-NL mapping. The discrete codebook is learned from noise-free images via VQGAN [16]. It is worth noting that by projecting degraded images onto this confined discrete prior space, the ambiguity inherent in LL-NL transformation is substantially mitigated, thereby ensuring the quality of enhanced images. Fig. 1 illustrates the superiority of our method over current state-of-the-art (SOTA) methods, both on the benchmark dataset and real-world images.

It is important to emphasize that the superior performance of GLARE over other SOTA methods is not only attributed to the integration of the codebook prior but also to our unique designs that address two main challenges associated with leveraging the codebook prior for LL-NL mapping. First, as shown in Fig. 2b column 2, solely exploiting VQGAN and NL prior may lead to unpleasant results and the reason behind this lies in the evident discrepancy between the degraded LL features and NL features in latent space, as depicted in Fig. 2a. Since the Nearest Neighbor (NN) is commonly utilized in looking up codebook [7, 10, 21, 61], this misalignment poses challenges in accurately retrieving VQ codes for LLIE



Fig. 2: (a) T-SNE [47] visualization of distributions of NL features, LL features, and LL-NF features. Compared to LL features, NF-LL features are better aligned with NL features. (b) Visual observations on each stage of GLARE on LOL [62] dataset. Column 2-4 represent the enhanced results from LL-feat, the enhanced images from NF-LL-feat, and the final results of our **GLARE**. From column 2-4, we observe a noticeable improvement on visibility, color preservation and detail recovery, which demonstrates the effectiveness of each stage of our GLARE. [Key: NL-feat: generated by NL encoder with NL inputs, LL-feat: LL features obtained form the fine-tuned NL encoder (Stage I), NF-LL-feat: LL features generated by our generative I-LNF module (Stage II)]

task. Second, we notice that relying solely on matched codes for feature decoding [7, 10, 61] might compromise the fine details. Without integrating information from the original LL input, it could potentially introduce texture distortions.

Taking into account these issues, we further introduce two specific modules into GLARE. First, to bridge the gap between degradation features and NL representations, we introduce a generative strategy to produce LL features based on Invertible Latent conditional Normalizing Flow (I-LNF), which enables better alignment with potentially matched NL features. Specifically, given LL-NL pairs, our I-LNF transforms complicated NL features into a simple distribution with the condition of LL features via the precise log-likelihood training strategy. As shown in Fig. 2a, through this fully invertible network, our GLARE achieves a generative derivation of LL features which are closely aligned with NL representations and ensures accurate code assembly in codebook, thereby generating better enhancement results as depicted in Fig. 2b column 3.

Second, to improve the texture details, we propose an Adaptive Feature Transformation (AFT) module equipped with learnable coefficients to effectively control the ratio of encoder features introduced to the decoder. By flexibly merging the LL information into the decoding process, our model exhibits resilience against severe image degradation and one can freely adjust these coefficients according to their preference for real-world image enhancement. Besides, the AFT module adopts a dual-decoder strategy, which includes the fixed Normal-Light Decoder (NLD) and the trainable Multi-scale Fusion Decoder (MFD). The NLD specifically processes matched codes from the codebook, facilitating the generation of realistic and natural results. Meanwhile, the MFD handles the LL features produced by our I-LNF module, enhancing the final result with more refined details and texture, as demonstrated in Fig. 2b column 4.

Contributions. The main contributions of this work are as follows: **(i)** We are the first to adopt the external NL codebook as a guidance to enhance low-light images naturally. **(ii)** We introduce **GLARE**, a novel LLIE enhancer leveraging latent normalizing flow to learn the LL feature distribution that aligns with NL features. **(iii)** A novel adaptive feature transformation module with an adjustable function for users is proposed to consolidate the fidelity while ensuring the naturalness in outputs. **(iv)** Extensive experiments indicate that our method significantly outperforms existing SOTA methods on 5 paired benchmarks and 4 real-world datasets in LLIE and our model is highly competitive while employed as a pre-processing method for high-level object detection task.

2 Related Work

2.1 Deep Learning based LLIE methods

Similar to numerous approaches in other image restoration tasks [2, 3, 13, 14, 38, 39, 54, 73], end-to-end LLIE methods [5, 19, 28, 33, 42, 67, 74] have been proposed to directly map LL images to NL ones. Most of them mainly resort to the optimization of reconstruction error between the enhanced output and ground-truth to guide the network training. However, they often fail to preserve naturalness and restore intricate details effectively. These problems have given rise to the exploration of leveraging additional information or guidance to aid the enhancement process. For instance, some methods [32, 56] achieve a simple training process for LLIE by estimating illumination maps. However, these approaches have a risk of amplifying noise and color deviations especially in real-world LL images [6].

Concurrently, other methods [17, 65, 78] argue that semantic understanding can mitigate regional degradation problems and attain pleasing visual appearance. Besides, several studies [51, 69, 80] have indicated that utilizing edge extraction can direct the generation of realistic image details and mitigate the blurry effects to an extent. Nevertheless, these methods are highly contingent upon features extracted from degraded input, which could compromise the generalization capability and introduce artifacts. In contrast to existing methods, we propose an informative codebook that encapsulates a diverse spectrum of NL feature priors. This approach demonstrates resilience against various degradations, achieving more natural and appealing image enhancement.

2.2 Vector-Quantized Codebook Learning

Vector Quantized Variational AutoEncoders (VQ-VAE) is firstly proposed in [50] to learn discrete representations. This approach effectively tackles the posterior collapse issue that is commonly encountered in VAE models. Then, VQ-VAE2 [52] explores the hierarchical VQ code for large-scale image generation. VQGAN [16] further enhances the perceptual quality by capturing a codebook of context-rich visual parts via an adversarial method. The discrete codebook has been successfully employed in image super-resolution [7], text super-resolution [34], and face restoration [21, 61, 79]. However, there remains potential

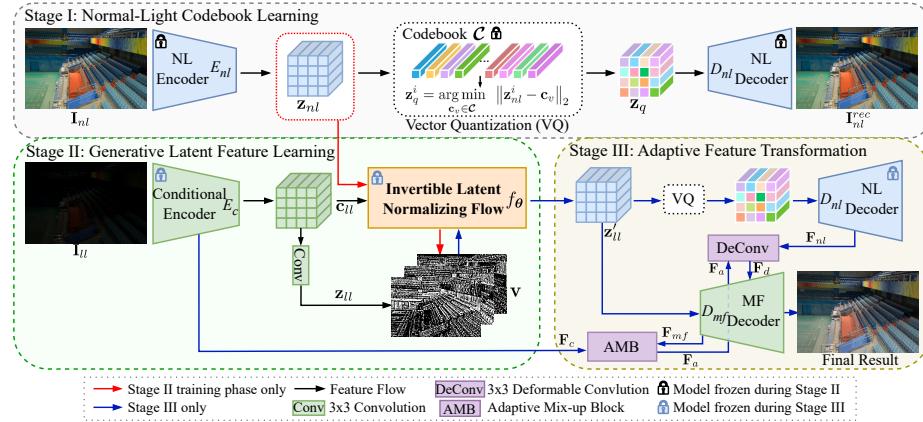


Fig. 3: The overall architecture of our proposed **GLARE**. There are three training stages in our model. Stage I aims to learn a comprehensive normal-light codebook \mathcal{C} using VQGAN. In Stage II training, given the \mathbf{c}_{ll} and \mathbf{z}_{ll} generated by the conditional encoder E_c and the convolution layer respectively, I-LNF module f_θ learns to transform the normal-light feature \mathbf{z}_{nl} to a simplified Gaussian distribution $\mathbf{v} = f_\theta(\mathbf{z}_{nl}; \mathbf{c}_{ll})$ with the mean of \mathbf{z}_{ll} . We optimize E_c and f_θ by minimizing the negative log-likelihood described in Eq. 3. In Stage III, the codebook \mathcal{C} , the NL decoder D_{nl} , the conditional encoder E_c , and the I-LNF f_θ are all fixed. Our GLARE can effectively transform a Gaussian density $p_v(\mathbf{v}) \sim \mathcal{N}(\mathbf{z}_{ll}, \Sigma)$ to the NL feature distribution $p_{\mathbf{z}_{nl}|\mathbf{c}_{ll}}(\mathbf{z}_{nl}|\mathbf{c}_{ll}, \theta)$. To further improve the enhancement performance, we propose an adaptive feature transformation strategy. By leveraging D_{mf} and AMB to flexibly incorporate LL information for decoding ($\mathbf{F}_d = DeConv(\mathbf{F}_{nl}, AMB(\mathbf{F}_c, \mathbf{F}_{mf}))$), our GLARE is capable to generate results with more refined texture and details. [Key: Σ : The unit variance]

for further improvement. One of the key research directions is how to precisely match the related correct code. Different from recent methods [72, 79] that utilize a Transformer to predict code indices in the codebook, we argue that prediction-based strategies are inherently unable to address the significant differences between LL and NL features, resulting in suboptimal performance. To this end, we propose a generative approach that produces LL features aligned with NL counterparts to successfully bridge the gap between LL and NL representations.

3 GLARE

Besides introducing external NL codebbok to guide the Low-Light to Normal-Light (LL-NL) mapping, the novelty of our work also lies in the distinctive Invertible Latent Normalizing Flow (I-LNF) and Adaptive Feature Transformation (AFT) modules, which are designed to maximize the potential of NL codebook prior and generate realistic results with high fidelity. The overview of our method is illustrated in Fig. 3, where the training of our method can be divided into three stages. In stage I, we pre-train a VQGAN on thousands of clear NL images to

construct a comprehensive VQ codebook (Sec. 3.1). In stage II, the I-LNF module is trained utilizing LL-NL pairs to achieve the distribution transformation between LL and NL features (Sec. 3.2). In the final stage, the AFT module, which contains the fixed NL Decoder (NLD), Adaptive Mix-up Block (AMB), and Multi-scale Fusion Decoder (MFD), is employed to enhance the fine-grained details while maintaining naturalness beneficial from the codebook (Sec. 3.3).

3.1 Stage I: Normal-Light Codebook Learning

To learn a universal and comprehensive codebook prior, we leverage a VQGAN with the structure similar to [16]. Specifically, a NL image $\mathbf{I}_{nl} \in \mathbb{R}^{3 \times W \times H}$ is first encoded and reshaped into the latent representation $\mathbf{z}_{nl} \in \mathbb{R}^{d \times N}$, where W , H , d , and $N = W/f \times H/f$ represent the image width, image height, the dimension of latent features, and the total number of latent features; f is the downsampling factor of the NL Encoder E_{nl} . Each latent vector \mathbf{z}_{nl}^i can be quantized to the corresponding code \mathbf{z}_q^i using Nearest-Neighbor (NN) matching as:

$$\mathbf{z}_q^i = \arg \min_{\mathbf{c}_v \in \mathcal{C}} \|\mathbf{z}_{nl}^i - \mathbf{c}_v\|_2, \quad (1)$$

where $\mathcal{C} \in \mathbb{R}^{d \times N_c}$ denotes the learnable codebook containing N_c discrete codes, with each element represented by \mathbf{c}_v . Then the quantized code \mathbf{z}_q is sent to NLD (denoted as D_{nl}) to generate reconstructed image \mathbf{I}_{nl}^{rec} .

To better illustrate the strengths and limitations of the codebook prior, we fine-tune the pre-trained VQGAN encoder on LL-NL pairs. Specifically, we achieve the enhanced results shown in Fig. 2b column 2 and we utilize t-SNE [47] to visualize LL features generated by fine-tuned NL encoder in Fig. 2a, which demonstrate the effectiveness of external NL prior in IILE. Besides, these visual results inspire us to design additional networks to align LL features with NL representations to further improve enhancement performance.

3.2 Stage II: Generative Latent Feature Learning

To fully exploit the potential of external codebook prior, we design additional mechanisms from the perspective of reducing the disparity between LL and NL feature distributions. Specifically, we develop an invertible latent normalizing flow to achieve the transformation between LL and NL feature distributions, thereby facilitating more accurate codes retrieval from codebook.

As shown in Fig. 3, two key components are optimized in stage II: the **Conditional Encoder** and the **Invertible LNF**. 1) The conditional encoder E_c , structurally identical to the NL encoder E_{nl} , inputs a LL image \mathbf{I}_{ll} and outputs the conditional feature \mathbf{c}_{ll} . 2) The I-LNF module in this work is realized through an invertible network, represented as f_θ . This module utilizes \mathbf{c}_{ll} as the condition to transform the complex NL feature distribution \mathbf{z}_{nl} to a latent feature, namely $\mathbf{v} = f_\theta(\mathbf{z}_{nl}; \mathbf{c}_{ll})$. Stage II focuses on obtaining a simplified distribution $p_v(\mathbf{v})$ in

the space of \mathbf{v} , such as a Gaussian distribution. Consequently, the conditional distribution $p_{\mathbf{z}_{nl}|\mathbf{c}_{ll}}(\mathbf{z}_{nl}|\mathbf{c}_{ll}, \boldsymbol{\theta})$ can be implicitly expressed as [43]:

$$p_{\mathbf{z}_{nl}|\mathbf{c}_{ll}}(\mathbf{z}_{nl}|\mathbf{c}_{ll}, \boldsymbol{\theta}) = p_{\mathbf{v}}(f_{\boldsymbol{\theta}}(\mathbf{z}_{nl}; \mathbf{c}_{ll})) | \det \frac{\partial f_{\boldsymbol{\theta}}}{\partial \mathbf{z}_{nl}}(\mathbf{z}_{nl}; \mathbf{c}_{ll})|. \quad (2)$$

Different from conventional normalizing flow applications [24, 43, 58, 65], we uniquely employ normalizing flow at the feature level rather than the image space, and our I-LNF module is designed without integrating any squeeze layers. Moreover, instead of using the standard Gaussian distribution as the prior of \mathbf{v} , we propose to use the LL feature \mathbf{z}_{ll} , generated by convolution layers based on \mathbf{c}_{ll} , as the mean value of $p_{\mathbf{v}}(\mathbf{v})$. The conditional distribution in Eq. (2) allows us to minimize the negative log-likelihood (NLL) in Eq. (3) to train the conditional encoder and I-LNF module. Besides, through the fully invertible network $f_{\boldsymbol{\theta}}$, we derive clear features \mathbf{z}'_{ll} for LL inputs by sampling $\mathbf{v} \sim p_{\mathbf{v}}(\mathbf{v})$ according to $\mathbf{z}'_{ll} = f_{\boldsymbol{\theta}}^{-1}(\mathbf{v}; \mathbf{c}_{ll})$.

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{c}_{ll}, \mathbf{z}_{nl}) = -\log p_{\mathbf{z}_{nl}|\mathbf{c}_{ll}}(\mathbf{z}_{nl}|\mathbf{c}_{ll}, \boldsymbol{\theta}). \quad (3)$$

After training, we evaluate our model on LOL dataset [62] to validate the effectiveness of the I-LNF. As shown in Fig. 2a, the LL feature distribution generated by I-LNF is closely aligned with that of the NL, facilitating accurate code assembly in the codebook. Moreover, our network achieves satisfactory enhancement results (Fig. 2b, column 3), indicating good LLIE performance after stage II. However, these results still exhibit considerable potential for improvement, especially in fidelity. For example, the color (row 1 in Fig. 2b) or structural details (row 2 in Fig. 2b) significantly diverge from the ground truth. This observation motivates and drives us to incorporate the input information into the decoding process to elevate the fidelity.

3.3 Stage III: Adaptive Feature Transformation

To further enhance the texture details and fidelity, we propose an adaptive feature transformation module that flexibly incorporates the feature $\mathbf{F}_c = \{\mathbf{F}_c^i\}$ from the conditional encoder into the decoding process, where i denotes the resolution level. Specifically, in order to maintain the realistic output of NLD and avoid the influence of degraded LL features, we adopt a dual-decoder architecture and develop MFD inspired by [21, 63]. Dual-decoder design enables us to leverage the deformable convolution ($dconv$) to warp NLD feature (\mathbf{F}_{nl}^i) as Eq. 4 and input the warped feature (\mathbf{F}_d^i) into MFD to generate the final enhancement.

$$\mathbf{F}_d^i = dconv(\mathbf{F}_{nl}^i, \mathbf{F}_t^i), \quad (4)$$

where i and \mathbf{F}_t^i denote the resolution level and the target feature respectively. In this work, we design a novel feature fusion network that adaptively incorporates LL information into the warping operation and provides a potential adjustment choice for users when testing on real-world occasions.

Table 1: Quantitative comparisons on LOL [62], LOL-v2-real [71], LOL-v2-synthetic [71], SDSD-indoor [55], and SDSD-outdoor [55] datasets. Our GLARE achieves superior performance compared to current SOTA methods. These results are obtained either from original papers or by running their released codes. [Key: **Best**, **Second Best**, \uparrow (\downarrow): Larger (smaller) values leads to better performance].

Methods	LOL-v1			LOL-v2-real			LOL-v2-syn		SDSD-indoor		SDSD-outdoor	
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow
SID [8]	14.35	0.436	—	13.24	0.442	—	15.04	0.610	23.29	0.703	24.90	0.693
IPT [9]	16.27	0.504	—	19.80	0.813	—	18.30	0.811	26.11	0.831	27.55	0.850
UFormer [59]	16.36	0.771	—	18.82	0.771	—	19.66	0.871	23.17	0.859	23.85	0.748
Sparse [71]	17.20	0.640	—	20.06	0.816	—	22.05	0.905	23.25	0.863	25.28	0.804
RUAS [37]	18.23	0.720	—	18.37	0.723	—	16.55	0.652	23.17	0.696	23.84	0.743
SCI [46]	14.78	0.646	—	20.28	0.752	—	24.14	0.928	—	—	—	—
KinD [77]	20.87	0.802	0.207	17.54	0.669	0.375	13.29	0.578	21.95	0.672	21.97	0.654
MIRNet [74]	24.14	0.830	0.131	20.02	0.820	0.138	21.94	0.876	24.38	0.864	27.13	0.837
DRBN [70]	19.86	0.834	0.155	20.13	0.830	0.147	23.22	0.927	24.08	0.868	25.77	0.841
SNR [68]	24.61	0.842	0.151	21.48	0.849	0.157	24.14	0.928	29.44	0.894	28.66	0.866
URetinex-Net [64]	21.33	0.835	0.122	21.16	0.840	0.144	24.14	0.928	—	—	—	—
Restormer [75]	22.43	0.823	—	19.94	0.827	—	21.41	0.830	25.67	0.827	24.79	0.802
Reformer [6]	25.16	0.845	0.131	22.80	0.840	0.171	25.67	0.930	29.77	0.896	29.84	0.877
MRQ [40]	25.24	0.855	0.121	22.37	0.846	0.142	25.94	0.935	—	—	—	—
LLFlow [58]	25.19	0.870	0.113	26.53	0.892	0.135	26.23	0.943	—	—	—	—
LL-SKF [65]	26.80	0.879	0.105	28.45	0.905	0.111	29.11	0.953	—	—	—	—
GLARE (Ours)	27.35	0.883	0.083	28.98	0.905	0.097	29.84	0.958	30.10	0.896	30.85	0.884

Adaptive Mix-up Block. The MFD that aligns structurally with NLD aims to decode the generated LL feature \mathbf{z}'_H and obtain intermediate representations as $\mathbf{F}_{mf} = \{\mathbf{F}_{mf}^i\}$, where i indicates the resolution level. At each resolution level, the conditional encoder information \mathbf{F}_c^i is added to the corresponding \mathbf{F}_{mf}^i in order to bring more LL information. Different from typical feature fusion operations (*i.e.*, skip connection [40]), our approach uses an adaptive mix-up strategy:

$$\mathbf{F}_a^i = \beta \times \sigma(\theta_i) \times \mathbf{F}_c^i + (1 - \beta \times \sigma(\theta_i)) \times \mathbf{F}_{mf}^i, \quad (5)$$

where θ_i represents a learnable coefficient, σ denotes the sigmoid operator, β is used for the adjustment for real-world testing and is set to 1 when training. Unlike skip connection, these learnable parameters can be adjusted effectively during training, which contributes to enhanced performance shown in Sec. 4.5.

Flexible Adjustment. Even though β in Eq. (5) is set to 1 in the training phase, one can flexibly adjust β according to their preference when testing with real-world images. This design stems from the phenomena that many current methods usually work struggling with real-world data, which often have different illuminations with images used in the training phase.

4 Experiments

4.1 Datasets

Normal Light Datasets. To train the VQGAN in Stage I, we select images with normal lighting conditions from DIV2K [1] and Flickr2K [35] datasets to develop the NL codebook prior.

Low Light Datasets. We conduct a thorough evaluation of our method using various paired datasets, including LOL [62], LOL-v2-real [71], LOL-v2-synthetic [71], and a large-scale dataset SDSD [55]. For LOL, LOL-v2-real, and LOL-v2-synthetic, we use 485, 689, and 900 pairs for training, and 15, 100, and 100 pairs for testing. The indoor subset of SDSD dataset includes 62 training and 6 testing video pairs, while the outdoor subset contains 116 training and 10 testing pairs. Besides, we also conduct cross-dataset evaluation on several unpaired real-world datasets: MEF [45], LIME [23], DICM [31], and NPE [57].

4.2 Implementation Details

Experiment Settings. We use the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.99$) for all training stages. In Stage I, the training iteration is set to 640K with a batch size of 4, a fixed learning rate 10^{-4} , and image size of 256×256 . In Stage II, we retain the batch size, change the image size to 320×320 , and adopt a multi-stage learning rates. Then, our GLARE is trained for 60K iterations on LOL and LOL-v2, and 225K iterations on SDSD. In Stage III, the batch size is halved, the initial learning rate is set to 5×10^{-5} , and the training iterations are adjusted to 20K, 40K, and 80K for LOL, LOL-v2, and SDSD datasets respectively.

Metrics. For paired datasets, we utilize Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [60] to assess pixel-level accuracy, and use Learned Perceptual Image Patch Similarity (LPIPS) [76] for perceptual quality evaluation. As for real-world datasets, the Natural Image Quality Evaluator (NIQE) [48] is adopted.

4.3 Performance on LLIE

Quantitative Results. As reported in Tab. 1, GLARE outperforms the current SOTA methods on five benchmarks. Our GLARE excels in PSNR, outperforming LL-SKF over 0.55 dB and 0.74 dB on LOL and LOL-v2-synthetic datasets. Furthermore, it surpasses Retformer with improvement of 0.33 dB and 1.01 dB on SDSD-indoor and SDSD-outdoor datasets. Additionally, our LPIPS scores surpass the second best performance by **20.9%** and **12.6%** in Tab. 1, indicating that the enhanced results from our network are more consistent with human visual system. Tab. 2 presents the cross-dataset evaluation on unpaired real-world datasets. We first train our GLARE on LOL training split. Then, the model that performs best on LOL test data is deployed on four unpaired datasets. As compared to the current SOTA methods, GLARE outperforms them on DICM and MEF and achieves the optimal performance on average. This demonstrates not only the superiority of our method in producing high-quality visual results but also its good generalization capability.

Qualitative Results. The visual quality of our GLARE against others are shown in Fig. 4, 5, and 6. Obviously, previous methods show inferior performance on noise suppression. Besides, they also tend to produce results with evident color distortion (See the enhanced results of KinD, LLFlow, Retformer,

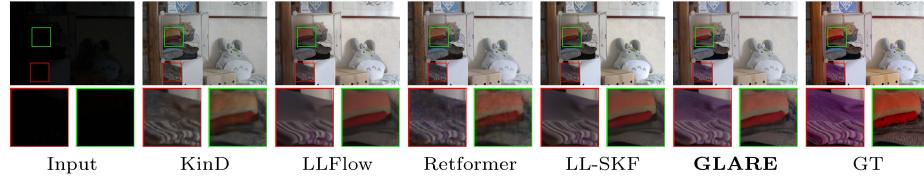


Fig. 4: Visual comparisons on LOL [62] dataset. Our method can effectively enhance visibility and generate visually appealing results.

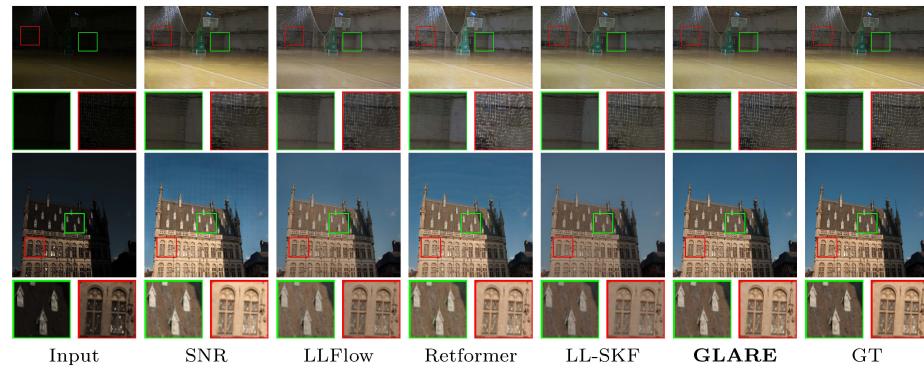


Fig. 5: Visual comparisons on LOL-v2-real [71] (top) and LOL-v2-synthetic [71] (bottom) datasets. Previous methods suffer from either severe color distortion or detail deficiency, while our GLARE performs favorably without these issues.

Table 2: Quantitative comparisons on real-world datasets. These results are obtained either from the original papers or testing with their best LOL [62] weights. [Key: **Best**, **Second Best**, ↓: Smaller value represents better quality].

Methods	MEF	LIME	DICM	NPE	Mean↓
SNR [68]	4.14	5.51	4.62	4.36	4.60
URetinex-Net [64]	3.79	3.86	4.15	4.69	4.11
LLFlow [58]	3.92	5.29	3.78	4.16	3.98
LL-SKF [65]	4.03	5.15	3.70	4.08	3.92
RFR [18]	3.92	3.81	3.75	4.13	3.81
GLARE (Ours)	3.66	4.52	3.61	4.19	3.75

LL-SKF in Fig. 4, and SNR, LLFlow in Fig. 5). Additionally, from the qualitative comparison, it can be seen that LLFlow, LL-SKF, and Retformer may induce the detail deficiency on their enhanced results (See Fig. 4 and Fig. 5), and KinD in Fig. 4 and SNR in Fig. 5 perform poorly in vision due to the introduce of unnatural artifacts. In comparison, GLARE can effectively enhance poor visibility while reliably preserving color and texture details without artifacts. The visual comparisons on unpaired real world datasets in Fig. 6 also demonstrate the strengths of our method in terms of details recovery and color maintenance.

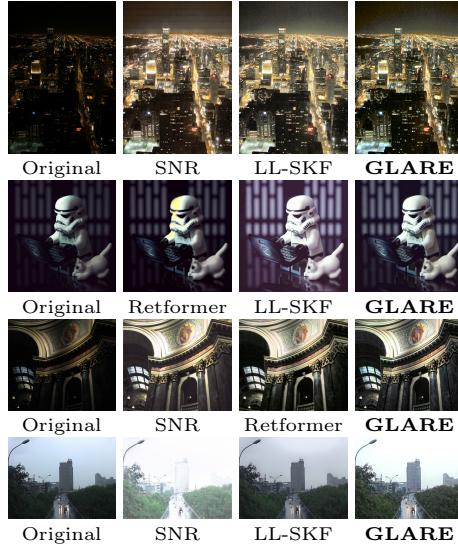


Fig. 6: Visual results of cross-dataset evaluation on unpaired real-world datasets. These four images are from DICM [31], LIME [23], MEF [45], and NPE [57] respectively. Our GLARE generates more pleasing results without noise or artifacts.



Fig. 7: Visual ablation results of AFT on LOL [62]. Our GLARE with AFT module is capable to generate results with improved edge acuity and contour definition, along with a more abundant detail texture.

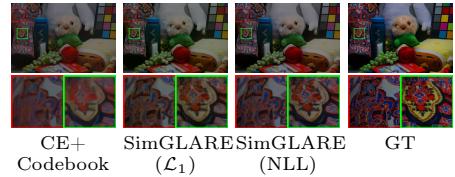


Fig. 8: Visual comparisons for ablation study of I-LNF. Without our proposed I-LNF module, the results exhibit significant detail loss and blurriness, coupled with a notably darker tone in certain areas.

4.4 Performance on Low-Light Object Detection

Implementation Details. To thoroughly evaluate our model, we also explore its potential as an effective preprocessing method in object detection task on ExDark dataset [41]. This dataset collects 7,363 low-light images, categorized into 12 classes and annotated with bounding boxes. We first employ different LLIE models trained on LOL to enhance the ExDark dataset, then carry out object detection on the enhanced images. More concretely, 5,896 images are used for training and the rest for evaluation. The adopted object detector is YOLO-v3 [53] pre-trained on COCO dataset [36].

Quantitative Results. We calculate the Average Precision (AP) and mean Average Precision (mAP) scores as our evaluation metrics. We compare our GLARE against current SOTAs in Tab. 3. As compared to KinD, MBLLEN, LLFlow, and LL-SKF, our GLARE achieves at least 0.8 improvement in terms of mAP. More importantly, our GLARE also outperforms IAT, which has demonstrated exceptional performance in low-light object detection.

Qualitative Results. The visual comparisons for low-light object detection is demonstrated in Fig. 9. It can be seen that although each LLIE method enhances the visibility to some extent, GLARE achieves the best visual performance, thus benefiting the most to the downstream detection task. Not surprisingly, the

Table 3: Quantitative low-light detection results on ExDark [41] using different LLIE method as the enhancement tool. [Key: **Best**, **Second Best**, \uparrow : The larger represents the better performance, Baseline: These scores are obtained by training the YOLO-v3 [53] detector on original ExDark [41] dataset].

Methods	Bicycle	Boat	Bottle	Bus	Car	Cat	Chair	Cup	Dog	Motor	People	Table	Mean \uparrow
Baseline [41]	80.4	76.5	77.6	89.7	84.0	71.5	69.5	76.4	78.7	76.4	81.9	52.6	76.32
MBLLEN [44]	82.2	77.5	76.3	90.3	84.1	70.9	69.4	75.9	77.7	74.7	82.0	58.2	76.59
KinD [77]	79.7	77.4	78.8	92.5	84.9	70.8	67.5	78.3	78.7	77.1	80.9	53.7	76.69
LLFlow [58]	81.6	75.5	74.3	92.5	84.5	69.7	69.0	75.8	79.0	76.5	80.9	57.9	76.44
IAT [11]	82.5	76.0	75.6	92.3	83.0	72.4	70.8	79.6	78.6	76.2	81.5	57.9	77.19
LL-SKF [65]	80.2	75.0	76.6	91.3	84.7	69.5	71.1	76.5	77.5	76.4	81.1	57.1	76.43
GLARE (Ours)	83.4	75.8	77.6	91.7	83.9	70.1	70.0	79.1	81.5	77.2	82.0	57.9	77.50

Table 4: By incorporating AFT, GLARE gains significant improvements on PSNR, SSIM, and LPIPS. Besides, our GLARE also performs better than two variants based on SimGLARE. [Key: SC: Skip Connection operation [25], Dual-d: Dual decoder architecture, **Best**, **Second Best**]

Methods	SC	AMB	Dual-d	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
(1) SimGLARE				26.51	0.855	0.109
(2) Variant 1		✓		26.60	0.867	0.093
(3) Variant 2	✓		✓	26.88	0.877	0.087
(4) GLARE	✓	✓	✓	27.35	0.883	0.083
Methods	\mathcal{L}_1	\mathcal{L}_{per}	\mathcal{L}_{ssim}	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
(5) GLARE	✓			27.02	0.870	0.099
(6) GLARE	✓	✓		27.09	0.871	0.091
(7) GLARE	✓	✓	✓	27.35	0.883	0.083

enhanced results from GLARE enables the YOLO-v3 detector to recognize more objects with higher confidence.

4.5 Ablation Study

To verify the effectiveness of each component of GLARE and justify the optimization objective utilized for training, we conduct extensive ablation experiments on LOL dataset. Specifically, we discuss the importance of AFT module, I-LNF module, and NL codebook prior in this section.

Effectiveness of Adaptive Feature Transformation. By removing the AFT module from our GLARE, we obtain a Simple LLIE model denoted as SimGLARE. Basically, SimGLARE only utilizes the information from NL codebook without feature transformation. The quantitative results of SimGLARE are shown in Tab. 4. SimGLARE is quite competitive on LLIE in terms of PSNR, SSIM, and LPIPS (compared with SOTAs in Tab. 1). However, with the proposed AFT module, our GLARE achieves further improvements on both quantitative metrics and visual results (as shown in Fig. 7). In addition, various loss functions are examined in Tab. 4, showing that our choice of losses in Stage III is reasonable.



Fig. 9: Visual comparisons and detection results of enhancement-based low-light object detection on the ExDark [41] dataset. Previous enhancement methods, when employed as preprocessing modules for object detection, encountered issues with object loss. In contrast, utilizing images enhanced by our GLARE enables YOLO-v3 to robustly detect targets with high confidence and our enhanced images exhibit better visual quality. Please zoom in for better view. [Key: Baseline: Test results of YOLO-v3 [53] detector trained on original ExDark [41] dataset].

Table 5: NLL can help achieve better enhancement result compare to \mathcal{L}_1 loss. While replacing the I-LNF module with a transformer or removing I-LNF entirely, the significant performance decline highlights the importance of the proposed I-LNF module.

Methods	PSNR↑	SSIM↑	LPIPS↓
(1) SimGLARE (\mathcal{L}_1)	25.69	0.842	0.132
(2) Codebook + Transformer (\mathcal{L}_1)	25.12	0.834	0.148
(3) Conditional Encoder + Codebook (\mathcal{L}_1)	24.53	0.825	0.161
(4) SimGLARE (NLL)	26.51	0.855	0.109

We also design two variants, named Variant 1 and Variant 2, to shed light on the importance of proposed dual-decoder architecture and AMB, respectively. Specifically, Variant 1 directly incorporates LL feature to NLD using AMB while Variant 2 adopts parallel decoder strategy but replaces AMB with skip connection operation [25]. By comparing (4) with (2) and (3) in Tab. 4, we observe that PSNR and SSIM are negatively correlated with LPIPS, which verifies the effectiveness of our AMB and dual-decoder design.

Effectiveness of Invertible Latent Normalizing Flow. To show the importance of I-LNF and the adopted NLL loss, we implement several adaptations based on SimGLARE. (1) We train SimGLARE using \mathcal{L}_1 loss to validate the effectiveness of NLL loss adopted in our work. (2) We replace the I-LNF module by leveraging a Transformer model structurally similar to [79] to directly predict the code index in the codebook. (3) We remove the I-LNF module in (1) and train the conditional encoder on LL-NL pairs. The quantitative results are reported

Table 6: The quantitative results of ablation experiments related to NL codebook prior. We observe significant decrease on metrics when the codebook is removed.

Methods	PSNR↑	SSIM↑	LPIPS↓
(1) Encoder-Decoder (\mathcal{L}_1)	22.79	0.804	0.195
(2) Conditional Encoder + Codebook (\mathcal{L}_1)	24.53	0.825	0.161
(3) Variant 3 (\mathcal{L}_1)	23.42	0.812	0.176
(4) SimGLARE (\mathcal{L}_1)	25.69	0.838	0.132

in Tab. 5. The superiority of NLL loss can be verified by comparing (4) and (1). Moreover, a comparison between the images in the second and third columns of Fig. 8 also reveals that the use of NLL loss, as opposed to \mathcal{L}_1 loss, results in sharper contours and edges. Besides, as compared to the Transformer-based code prediction, our proposed I-LNF module can help generate LL features that are better aligned with NL ones, thus ensuring more accurate code matching and achieving superior performance. More importantly, with the I-LNF module removed from SimGLARE (\mathcal{L}_1), we notice a significant decrease in PSNR (1.16 dB \downarrow) and SSIM (0.017 \downarrow), which demonstrates the effectiveness of our proposed I-LNF module.

Effectiveness of Codebook Prior. To investigate the importance of the NL codebook prior, based on SimGLARE (\mathcal{L}_1), we remove the codebook and the quantization process in VQGAN. The resulting model is denoted as Variant 3 and is trained using a strategy similar to that for SimGLARE (\mathcal{L}_1). Similarly, we remove the codebook in the model reported in row 3 of Tab. 5 to learn the LL-NL mapping. Quantitative results are reported in Tab. 6. The absence of a codebook prior notably impacts performance, as evidenced by an average decrease of 2.0 dB in PSNR and a 0.024 drop in SSIM. This highlights the critical importance of the codebook prior in our method.

5 Conclusion

A novel method named GLARE is proposed for LLIE. In view of the uncertainty and ambiguity caused by ill-posed nature of LLIE, we leverage the normal light codebook, which is obtained from clear and well-exposed images using VQGAN, to guide the LL-NL mapping. To better exploit the potential of codebook prior, the invertible latent normalizing flow is adopted to generate LL features aligned with NL latent representations to maximize the probability that code vectors are correctly matched in codebook. Finally, the AFT module with dual-decoder architecture is introduced to flexibly supply information into the decoding process, which further improves the fidelity of enhanced results while maintaining the perceptual quality. Extensive experiments demonstrate that our GLARE significantly outperforms the current SOTA methods on 5 paired datasets and 4 real-world datasets. The superior performance on low light object detection makes our GLARE an effective preprocessing tool in high-level vision tasks.

References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops) (2017) **8**
2. Ancuti, C.O., Ancuti, C., Vasluiianu, F.A., Timofte, R., et al.: Ntire 2023 hr nonhomogeneous dehazing challenge report. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops) (2023) **4**
3. Ancuti, C.O., Ancuti, C., Vasluiianu, F.A., Timofte, R., et al.: Ntire 2024 dense and nonhomogeneous dehazing challenge report. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops) (2024) **4**
4. Arici, T., Dikbas, S., Yucel, A.: A histogram modification framework and its application for image contrast enhancement. In: IEEE Transactions on Image Processing (TIP) (2009) **1**
5. Cai, J., Gu, S., Zhang, L.: Learning a deep single image contrast enhancer from multi-exposure images. In: IEEE Transactions on Image Processing (TIP) (2018) **2, 4**
6. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2023) **4, 8**
7. Chen, C., Shi, X., Qin, Y., Li, X., Han, X., Yang, T., Guo, S.: Real-world blind super-resolution via feature matching with implicit high resolution priors. In: Proceedings of the ACM Conference on Multimedia (MM) (2022) **2, 3, 4**
8. Chen, C., Chen, Q., Do, M.N., Koltun, V.: Seeing motion in the dark. In: Proceedings of the Conference on British Machine Vision Conference (BMVC) (2019) **8**
9. Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W.: pre-trained image processing transformer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021) **8**
10. Choi, N., Lee, S., Lee, Y., Yang, S.: Restoration of hand-drawn architectural drawings using latent space mapping with degradation generator. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2023) **2, 3**
11. Cui, Z., Li, K., Gu, L., Su, S., Gao, P., Jiang, Z., Qiao, Y., Harada, T.: You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. In: Proceedings of the Conference on British Machine Vision Conference (BMVC) (2022) **12**
12. Cui, Z., Zhu, Y., Gu, L., Qi, G.J., Li, X., Zhang, R., Zhang, Z., Harada, T.: Exploring resolution and degradation clues as self-supervised signal for low quality object detection. In: Proceedings of the European Conference on Computer Vision (ECCV) (2022) **1**
13. Dong, W., Zhou, H., Tian, Y., Sun, J., Liu, X., Zhai, G., Chen, J.: ShadowRefiner: Towards mask-free shadow removal via fast fourier transformer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops) (2024) **4**
14. Dong, W., Zhou, H., Wang, R., Liu, X., Zhai, G., Chen, J.: DehazeDCT: Towards effective non-homogeneous dehazing via deformable convolutional transformer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops) (2024) **4**

15. Dong, W., Zhou, H., Xu, D.: A new sclera segmentation and vessels extraction method for sclera recognition. In: 2018 10th International Conference on Communication Software and Networks (ICCSN) (2018) [1](#)
16. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021) [2](#), [4](#), [6](#)
17. Fan, M., Wang, W., Yang, W., Liu, J.: Integrating semantic segmentation and retinex model for low light image enhancement. In: Proceedings of the ACM Conference on Multimedia (MM) (2020) [2](#), [4](#)
18. Fu, H., Zheng, W., Meng, X., Wang, X., Wang, C., Ma, H.: You do not need additional priors or regularizers in retinex-based low-light image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2023) [10](#)
19. Fu, K., Peng, Y., Zhang, Z., Xu, Q., Liu, X., Wang, J., Zhai, G.: Attentionlut: Attention fusion-based canonical polyadic lut for real-time image enhancement. arXiv preprint arXiv:2401.01569 (2024) [4](#)
20. Fu, X., Zeng, D., Huang, Y., Xiao-Ping, Z., Xinghao, D.: A weighted variational model for simultaneous reflectance and illumination estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) [1](#)
21. Gu, Y., Wang, X., Xie, L., Dong, C., Li, G., Shan, Y., Cheng, M.M.: Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder. In: Proceedings of the European Conference on Computer Vision (ECCV) (2022) [2](#), [4](#), [7](#)
22. Guo, Q., Feng, W., Gao, R., Liu, Y., Wang, S.: Exploring the effects of blur and deblurring to visual object tracking. In: IEEE Transactions on Image Processing (TIP) (2021) [1](#)
23. Guo, X.: Lime: A method for low-light image enhancement. In: Proceedings of the ACM Conference on Multimedia (MM) (2016) [9](#), [11](#)
24. Hong, S., Park, I., Chun, S.Y.: On the robustness of normalizing flows for inverse problems in imaging. In: Proceedings of the Conference on British Machine Vision Conference (BMVC) (2023) [7](#)
25. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017) [12](#), [13](#)
26. Huang, G., Wang, X., Wu, W., Zhou, H., Wu, Y.: Real-time lane-vehicle detection and tracking system. In: Chinese Control and Decision Conference (CCDC) (2016) [1](#)
27. Ibrahim, H., Kong, N.S.P.: Brightness preserving dynamic histogram equalization for image contrast enhancement. In: Transaction on Consumer Electronics (TCE) (2007) [1](#)
28. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. In: IEEE Transactions on Image Processing (TIP) (2021) [4](#)
29. Jobsob, D., Zia-ur, R., Woodell, G.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. In: IEEE Transactions on Image Processing (TIP) (1997) [1](#)
30. Kim, E., Lee, S., Park, J., Choi, S., Seo, C., Choo, J.: Deep edge-aware interactive colorization against color-bleeding effects. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2021) [2](#)

31. Lee, C., Lee, C., Kim, C.S.: Contrast enhancement based on layered difference representation of 2d histograms. In: IEEE Transactions on Image Processing (TIP) (2013) [9](#), [11](#)
32. Li, C., Guo, J., Porikli, F., Pang, Y.: Lightennet: A convolutional neural network for weakly illuminated image enhancement. In: Pattern Recognition (PR) (2018) [4](#)
33. Li, W., Wu, G., Wang, W., Ren, P., Liu, X.: Fastllve: Real-time low-light video enhancement with intensity-aware lookup table. In: Proceedings of the ACM Conference on Multimedia (MM) (2023) [4](#)
34. Li, X., Zuo, W., Loy, C.C.: Learning generative structure prior for blind text image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2023) [4](#)
35. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops) (2017) [8](#)
36. Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P.: Microsoft coco: Common objects in context. In: Proceedings of the European Conference on Computer Vision (ECCV) (2014) [11](#)
37. Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021) [8](#)
38. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention based multi-scale network for image dehazing. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2019) [4](#)
39. Liu, X., Shi, Z., Wu, Z., Chen, J., Zhai, G.: Griddehazenet+: An enhanced multi-scale network with intra-task knowledge transfer for single image dehazing. IEEE Transactions on Intelligent Transportation Systems (2022) [4](#)
40. Liu, Y., Huang, T., Dong, W., Wu, F., Li, X., Shi, G.: Low-light enhancement with multi-stage residue quantization and brightness-aware attention. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2023) [8](#)
41. Loh, Y.P., Chan, C.S.: Getting to know low light images with the exclusively dark dataset. In: Computer Vision and Image Understanding (CVIU) (2019) [11](#), [12](#), [13](#)
42. Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: A deep autoencoder approach to natural low-light image enhancement. In: Pattern Recognition (PR) (2017) [2](#), [4](#)
43. Lugmayr, A., Danelljan, M., Gool, L.V., Timofte, R.: Srlflow: Learning the super-resolution space with normalizing flow. In: Proceedings of the European Conference on Computer Vision (ECCV) (2020) [7](#)
44. Lv, F., Lu, F., Wu, J., Lim, C.: Mbllen: Low-light image/video enhancement using cnns. In: Proceedings of the Conference on British Machine Vision Conference (BMVC) (2018) [12](#)
45. Ma, K., Zeng, K., Wang, Z.: Perceptual quality assessment for multi-exposure image fusion. In: IEEE Transactions on Image Processing (TIP) (2015) [9](#), [11](#)
46. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward fast, flexible, and robust low-light image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) [8](#)
47. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. In: Journal of machine learning research (2008) [3](#), [6](#)
48. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a 'completely blind' image quality analyzer. In: IEEE Signal Processing Letters (SPL) (2012) [9](#)

49. Moran, S., Marza, P., McDonagh, S., Parisot, S., Slabaugh, G.: Deeplpf: Deep local parametric filters for image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) [2](#)
50. Van den Oord, A., Vinyals, O., Kavukcuoglu, K.: Neural discrete representation learning. In: Advances in Neural Information Processing Systems (NeurIPS) (2017) [4](#)
51. Rana, D., Lal, K.J., Parihar, A.S.: Edge guided low-light image enhancement. In: Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS) (2021) [2, 4](#)
52. Razavi, A., Van den Oord, A., Vinyals, O.: Generating diverse high-fidelity images with vq-vae-2. In: Advances in Neural Information Processing Systems (NeurIPS) (2019) [4](#)
53. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018) [11, 12, 13](#)
54. Vasluiianu, F.A., Seizinger, T., Zhou, Z., Wu, Z., Chen, C., Timofte, R., et al.: NTIRE 2024 image shadow removal challenge report. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops) (2024) [4](#)
55. Wang, R., Xu, X., Fu, C.W., Lu, J., Yu, B., Jia, J.: Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2021) [8, 9](#)
56. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019) [2, 4](#)
57. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. In: IEEE Transactions on Image Processing (TIP) (2013) [9, 11](#)
58. Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.P., Kot, A.C.: Low-light image enhancement with normalizing flow. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) (2022) [7, 8, 10, 12](#)
59. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: A general u-shaped transformer for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) [8](#)
60. Wang, Z., Bovik, A.C., Sheikh, H.R., , Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. In: IEEE Transactions on Image Processing (TIP) (2004) [9](#)
61. Wang, Z., Zhang, J., Chen, R., Wang, W., Luo, P.: Restoreformer: High-quality blind face restoration from undegraded key-value pairs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) [2, 3, 4](#)
62. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: Proceedings of the Conference on British Machine Vision Conference (BMVC) (2018) [2, 3, 7, 8, 9, 10, 11](#)
63. Wu, R.Q., Duan, Z.P., Guo, C.L., Chai, Z., Li, C.: Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2023) [7](#)
64. Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., Jiang, J.: Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) [8, 10](#)

65. Wu, Y., Pan, C., Wang, G., Yang, Y., Wei, J., Li, C., Shen, H.T.: Learning semantic-aware knowledge guidance for low-light image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2023) [2](#), [4](#), [7](#), [8](#), [10](#), [12](#)
66. Xu, D., Dong, W., Zhou, H.: Sclera recognition based on efficient sclera segmentation and significant vessel matching. In: The Computer Journal (2022) [1](#)
67. Xu, K., Yang, X., Yin, B., Lau, R.H.: Learning to restore low-light images via decomposition-and-enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) [4](#)
68. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) [8](#), [10](#)
69. Xu, X., Wang, R., Lu, J.: Low-light image enhancement via structure modeling and guidance. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2023) [2](#), [4](#)
70. Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) [2](#), [8](#)
71. Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. In: IEEE Transactions on Image Processing (TIP) (2021) [8](#), [9](#), [10](#)
72. Ye, T., Chen, S., Bai, J., Shi, J., Xue, C., Jiang, J., Yin, J., Chen, E., Liu, Y.: Adverse weather removal with codebook priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2023) [5](#)
73. Yin, X., Liu, X., Liu, H.: Fmsnet: Underwater image restoration by learning from a synthesized dataset. In: International Conference on Artificial Neural Networks (ICANN) (2019) [4](#)
74. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Learning enriched features for real image restoration and enhancement. In: Proceedings of the European Conference on Computer Vision (ECCV) (2020) [4](#), [8](#)
75. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) [8](#)
76. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018) [9](#)
77. Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: A practical low-light image enhancer. In: Proceedings of the ACM Conference on Multimedia (MM) (2019) [8](#), [12](#)
78. Zheng, S., Gupta, G.: Semantic-guided zero-shot learning for low-light image/video enhancement. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV) (2022) [1](#), [2](#), [4](#)
79. Zhou, S., Chan, K.C., Li, C., Loy, C.C.: Towards robust blind face restoration with codebook lookup transformer. In: Advances in Neural Information Processing Systems (NeurIPS) (2022) [4](#), [5](#), [13](#)
80. Zhu, M., Pan, P., Chen, W., Yang, Y.: Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) (2020) [2](#), [4](#)