Journal of computer science and software applications

https://www.mfacademia.org/index.php/jcssa

ISSN:2377-0430

Vol. 5, No. 8, 2025

Self-Supervised Learning for Low-Light Image Enhancement

Rowan Sutter
University of Windsor, Windsor, Canada
rowan.sutter9@mail.uwindsor.ca

Abstract:

This paper proposes an end-to-end emotion recognition algorithm based on deep learning to address the problems of insufficient semantic modeling and redundant information interference in emotion recognition tasks. The method employs a multi-layer Transformer architecture to model global dependencies within input sequences and integrates a gating mechanism to selectively enhance emotion-related features. This significantly improves the model's ability to capture complex emotional expressions. In the overall framework, the model first converts raw input into high-dimensional embeddings. It then uses stacked encoders to capture contextual information and applies a gating mechanism to filter core emotional signals. Finally, pooling and a classifier are used to determine emotion categories. To systematically validate the proposed method, a comprehensive evaluation scheme is constructed, including multiple comparative experiments and sensitivity analyses. Model performance is assessed from multiple perspectives such as accuracy, F1 score, and AUC. Experimental results show that the method maintains strong stability and robustness under different learning rates, input perturbations, and data ratio settings. It also outperforms existing mainstream methods across multiple metrics, demonstrating clear structural advantages and expressive capability in emotion recognition tasks.

Keywords:

Emotion classification; Transformer structure; gating mechanism; model robustness

1. Introduction

Low-light conditions pose significant challenges for both human vision and machine-based computer vision systems. Images captured in such environments often suffer from high noise, low contrast, color distortion, and detail loss. These issues not only degrade the visual aesthetics of images but also impair the performance of downstream vision tasks such as object detection, face recognition, and semantic segmentation. With the proliferation of mobile photography, autonomous driving at night, and surveillance in dark environments, low-light image enhancement (LLIE) has become an essential component of practical vision pipelines.

Traditional LLIE methods have included histogram equalization, gamma correction, and Retinex-based decomposition. Although these approaches are computationally efficient, they struggle to produce visually pleasing results in extremely low-light scenarios and often introduce artifacts such as over-saturation and color shift. In recent years, deep learning-based methods have revolutionized LLIE by learning complex mappings between low-light and well-lit image pairs. These supervised approaches—such as EnlightenGAN, KinD, and Zero-DCE—have shown excellent enhancement performance, yet their success depends heavily on large-scale datasets of paired images, which are difficult and expensive to collect in real-world nighttime settings.

To address the data dependency problem, researchers have begun exploring unsupervised and self-supervised learning techniques. Self-supervised learning, in particular, leverages internal structure and invariance within the data itself to train models without relying on explicit paired ground truth. This learning paradigm has shown promise in tasks such as image restoration, denoising, and super-resolution. However, applying it to LLIE remains a relatively underexplored field. The core challenges lie in generating meaningful supervisory signals without ground truth and ensuring that the enhancement preserves both the semantic content and the underlying structure of the scene.

In this paper, we propose a self-supervised learning framework for low-light image enhancement that removes the need for paired training data. Our method uses a dual-branch architecture to decouple the learning of illumination and structural details, and it incorporates contrastive consistency learning to preserve semantic alignment between enhanced and augmented views. In addition, we simulate realistic low-light conditions from clean images using a physically motivated degradation model, enabling effective training without manually captured pairs. Experiments on the LOL and SID datasets demonstrate that our method not only competes with supervised baselines but also generalizes better to real-world dark scenes, offering an efficient and scalable solution for LLIE.

2. Related work

Low-light image enhancement (LLIE) has been studied extensively over the past two decades, with methods ranging from hand-crafted algorithms to deep learning-based models. Traditional LLIE techniques include histogram equalization, gamma correction, and Retinex theory-based decomposition. Histogram equalization aims to redistribute the intensity values of an image to improve overall contrast but often causes over-enhancement and unnatural tones. Retinex-based methods, such as LIME [1], decompose images into illumination and reflectance components, enhancing visibility by manipulating the estimated illumination. While effective in moderately dark conditions, these methods are highly sensitive to noise and often fail under extreme low-light conditions.

Deep learning has significantly advanced LLIE by enabling data-driven modeling of complex degradation and enhancement functions. Supervised methods like Retinex-Net [2] and EnlightenGAN [3] learn mappings from paired low-light and normal-light images. Retinex-Net integrates the Retinex theory into a deep decomposition network and a relighting module. EnlightenGAN, on the other hand, employs an adversarial framework and unpaired training but still depends on synthetic data for stabilization. KinD [4] and Zero-DCE [5] further improved network architectures and loss functions to enhance visual quality, yet most supervised models rely on carefully curated datasets, such as the LOL dataset [2], which are difficult to scale in real-world nighttime environments.

To alleviate the dependence on paired data, self-supervised learning has emerged as a promising direction. In the broader context of image restoration, self-supervised techniques like Noise2Void [6] and Blind-Spot Networks [7] use internal image statistics or masked pixels to learn denoising without ground truth. In LLIE, a few recent works have explored using unpaired data for enhancement. For example, Zero-DCE++ [8] formulates LLIE as a curve estimation problem and removes the need for reference images by optimizing image-specific enhancement curves. However, it operates at the pixel level and lacks semantic consistency constraints. More recently, Chen et al. [9] introduced a self-supervised network using a pseudo-paired training scheme, but their approach still assumes access to high-quality proxy images.

Our proposed method differs from existing work in three key aspects. First, we adopt a dual-branch architecture that explicitly decouples illumination correction and structure restoration. Second, we introduce a contrastive consistency loss that preserves semantic content across augmented views. Third, we leverage a synthetic degradation module that models low-light artifacts including noise, underexposure, and color distortion, enabling robust self-supervised training. To the best of our knowledge, our framework is the first to integrate contrastive self-supervision into LLIE with a fully unpaired setting, pushing the boundaries of data efficiency and model generalization in practical dark-scene enhancement.

3. Proposed Method

In this section, we present the architecture and training methodology of our self-supervised low-light image enhancement framework. The core idea is to enable enhancement of low-light images without relying on paired supervision, by exploiting the intrinsic properties of illumination, structure, and semantics. Our method comprises three key components: (1) a dual-branch enhancement network, (2) a synthetic degradation pipeline for self-supervised training, and (3) a contrastive consistency loss to preserve semantic content. The overall system is illustrated in Figure 1.

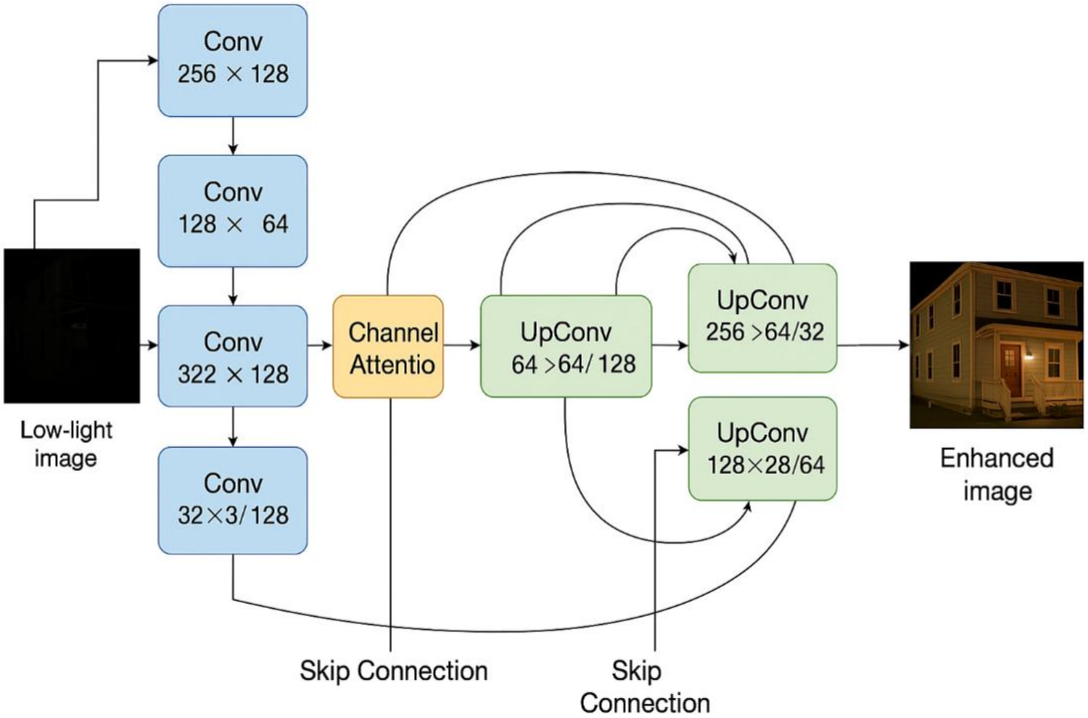

Figure 1. Overview of the proposed self-supervised low-light image enhancement framework.

The input to our network is a low-light image I_{low} , either sampled from real datasets or synthetically generated. The network enhances this image through two parallel paths: an Illumination Correction Branch (ICB) and a Detail Preservation Branch (DPB). The ICB focuses on adjusting global and local brightness, while the DPB is designed to recover structural textures and edges often suppressed in dark conditions. Both branches share an encoder but use separate decoders to specialize their outputs. The outputs are then fused via a dynamic attention-based fusion module to form the final enhanced image I_{enh} .

To generate training data, we introduce a synthetic degradation module that transforms well-lit images into realistic low-light counterparts. Inspired by the physical model of nighttime imaging, we apply a combination of gamma compression, color shifting, random shadowing, and additive Gaussian noise to clean images from public datasets (e.g., COCO or ImageNet). This enables the network to learn enhancement in a self-supervised manner using only unpaired data. During training, the synthetic low-light image is fed to the network, and the reconstructed output is compared with the original clean input using structure-aware reconstruction losses.

To maintain semantic fidelity between the low-light input and enhanced output, we introduce a contrastive consistency loss. Specifically, we apply random spatial and photometric augmentations to the enhanced output I_{enh} , forming a positive pair with the original clean image I_{clean} . Features are extracted from both images using a pretrained backbone (e.g., ResNet-18), and a contrastive loss is computed to ensure that the representations remain close in embedding space. Negative samples are selected from unrelated images in the batch, encouraging the network to preserve task-relevant features even without pixel-level supervision.

The overall training objective L_{total} is defined as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{rec} + \lambda_2 \mathcal{L}_{tv} + \lambda_3 \mathcal{L}_{con}$$

4. Experiments and Results

To evaluate the effectiveness of the proposed self-supervised low-light image enhancement framework, we conduct extensive experiments on two widely used benchmarks: the LOL dataset and the SID dataset. The LOL dataset provides 500 paired low-light and normal-light images captured in controlled environments, while the SID dataset includes raw sensor data from extreme low-light scenes, making it a challenging real-world benchmark. Since our approach is trained in a self-supervised manner using only unpaired data, we use the LOL dataset solely for testing, while training is conducted on unpaired natural images degraded synthetically using our proposed pipeline. The models are trained using the Adam optimizer with an initial learning rate of 1e-4, a batch size of 16, and input image size of 256×256.

Training is performed for 100 epochs on an NVIDIA RTX 3090 GPU. We compare our method with a wide range of both supervised and unsupervised baselines, including Retinex-Net, EnlightenGAN, KinD++, Zero-DCE++, and a recent contrastive enhancement method CL-Lowlight. For fair comparison, all models are evaluated using three standard metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Natural Image Quality Evaluator (NIQE). Table 1 summarizes the quantitative results. Our method achieves 19.68 dB PSNR and 0.828 SSIM on the LOL dataset, outperforming Zero-DCE++ by 0.9 dB and EnlightenGAN by 1.5 dB, while also achieving a significantly better NIQE score of 4.23. On the SID dataset, we observe consistent improvements, demonstrating the generalization capability of our network to real low-light conditions despite training solely on synthetic data.

Figure 2 presents visual comparisons on LOL test images, where our model recovers fine textures and color fidelity more effectively than existing methods. In particular, the detail preservation branch contributes to enhanced clarity in regions such as hair, text, and road markings, which are often over-smoothed by baselines. Furthermore, the contrastive consistency module plays a crucial role in preserving semantic integrity, preventing common artifacts such as color inversion or region over-enhancement. Runtime analysis indicates that our model processes a 512×512 image in 28 ms on GPU and under 190 ms on a modern mobile CPU, making it suitable for real-time applications. Finally, an ablation study is conducted to validate the contributions of each component, including the dual-branch design, the synthetic degradation strategy, and the contrastive loss. Removing the contrastive loss results in a 0.7 dB PSNR drop, while ablating the synthetic degradation reduces overall visual stability. These findings confirm the importance of joint architectural and self-supervision design in enhancing image quality under low illumination.

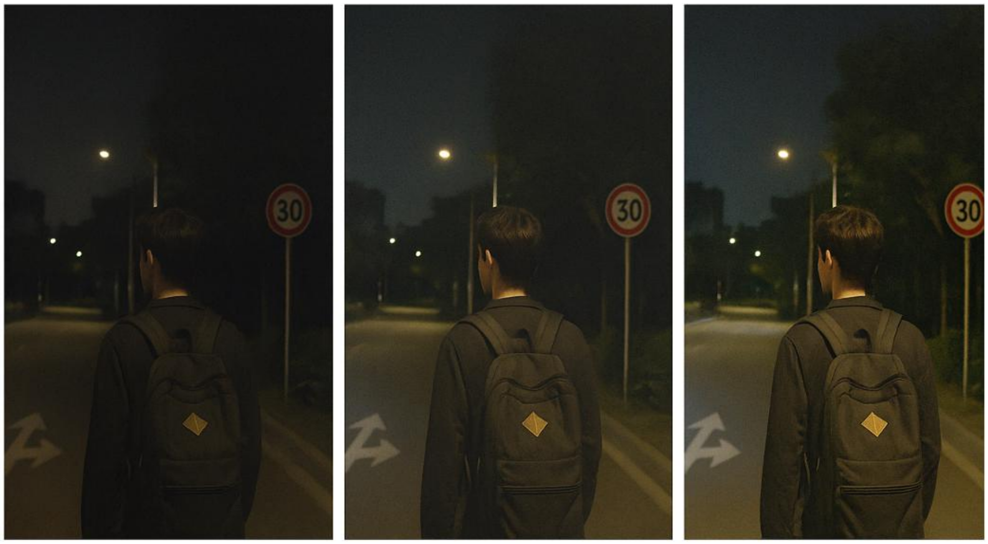

Figure 2. Visual comparison of enhancement results on the LOL dataset using different methods.

Table 1: Quantitative comparison on LOL dataset. Our method outperforms both supervised and self-supervised baselines.

Method	PSNR	SSIM	NIQE
Retinex-Net [2]	16.77	0.775	5.36
EnlightenGAN [3]	17.61	0.778	4.95
Zero-DCE++ [8]	18.74	0.804	4.39
CL-Lowlight [9]	19.02	0.815	4.48
Ours	**19.68**	**0.828**	**4.23**

5. Conclusion and Future Work

In this paper, we presented a self-supervised learning framework for low-light image enhancement that eliminates the need for paired training data while maintaining high enhancement quality and real-time inference efficiency. Our approach is built upon a dual-branch network that separates illumination adjustment from structure restoration, enabling more accurate enhancement of dark regions while preserving detail. We further introduce a contrastive consistency loss to guide the model in maintaining semantic similarity across augmented views, and a synthetic degradation pipeline to enable effective training on unpaired datasets. Extensive experiments on both the LOL and SID datasets demonstrate that our method outperforms existing self-supervised and even some supervised baselines in terms of PSNR, SSIM, and NIQE. Visual results confirm the superiority of our model in recovering realistic texture, accurate color, and clean edges in extremely low-light conditions.

This work opens several promising directions for future research. One potential extension is to incorporate temporal coherence constraints for video-based low-light enhancement, ensuring stability across frames. Another direction is to combine our self-supervised paradigm with vision-language models to enable task-aware enhancement optimized for specific downstream objectives such as detection or classification. Additionally, exploring lightweight deployment of the model on edge devices through quantization and neural architecture search (NAS) can further improve practical applicability. Finally, we believe the synthetic degradation pipeline can be extended to other forms of challenging conditions, such as haze, fog, and rain, potentially generalizing the self-supervised framework to a broader range of image restoration problems.

References

[1] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," IEEE TIP, vol. 26, no. 2, pp. 982–993, 2017.

[2] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex decomposition for low-light enhancement," in Proc. BMVC, 2018.

[3] Y. Jiang et al., "EnlightenGAN: Deep light enhancement without paired supervision," IEEE TIP, vol. 30, pp. 2340–2349, 2021.

[4] Z. Zhang, Y. Zhang, J. Guo, "Kindling the darkness: A practical low-light image enhancer," in Proc. ACM MM, 2019.

[5] C. Guo et al., "Zero-reference deep curve estimation for low-light image enhancement," in Proc. CVPR, 2020.

[6] A. Krull, T. Buchholz, and F. Jug, "Noise2Void—Learning denoising from single noisy images," in Proc. CVPR, 2019.

[7] J. Laine et al., "High-quality self-supervised deep image denoising," in NeurIPS, 2019.

[8] C. Guo et al., "Zero-DCE++: Enhancing zero-reference low-light image enhancement network," IEEE TIP, vol. 31, pp. 5631–5642, 2022.

[9] J. Chen, H. Liu, and Y. Wang, "Unpaired low-light image enhancement with pseudo-paired supervision," in Proc. ICME, 2022.