Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 15, Issue. 1, January 2026, pg.56 – 65

Stereo-IDCE: Interactive Dual-Curve Estimation for Unsupervised Low-Light Stereo Image Enhancement

Jie Chu[1]; Yaqi Han[2]

[1]Hefei University of Technology, China
[2]Zhengzhou University, China
iejiechu@163.com; yqhan@gs.zzu.edu.cn

DOI: https://doi.org/10.47760/ijcsmc.2026.v15i01.005

Abstract: Enhancing low-light stereo images without supervised data faces significant challenges in effectively recovering brightness while preserving consistency between views. To tackle this, we propose Stereo-IDCE, the first unsupervised low-light stereo enhancement framework built upon interactive dual-curve estimation. Our approach simultaneously estimates pixel-wise curve parameter maps for both views and enforces photometric consistency through coupled feature interaction and specialized loss constraints. Specifically, we introduce a weight-shared IDCE-Net to extract illumination-aware features from both views, along with a cross-view feature interaction module that leverages structural and brightness cues from the opposite view to enable collaborative curve estimation. This design ensures that enhancement operations are coordinated across views rather than learned independently. For loss constraints, we design a joint curve estimation loss CDE-Loss to minimize discrepancies between the left and right curve parameter maps, and a cross-view exposure control loss to align the average brightness of corresponding local regions. These losses explicitly preserve photometric consistency without requiring ground-truth. Extensive experiments on synthetic and in-the-wild datasets demonstrate that our method outperforms existing baselines in enhancing low-light stereo images.

Keywords: **Low-light Stereo Image Enhancement, Low-light image enhancement, Unsupervised, Curve estimation, Zero shot**

I. INTRODUCTION

The challenge of acquiring high-quality visual data under low-light conditions is fundamental to numerous computer vision applications, from autonomous navigation to computational photography. Low-light image enhancement (LLIE) aims to restore visibility, suppress noise, and correct color distortion in images captured under poor illumination [1], [2]. While significant progress has been made for single images, driven by both supervised [3]–[10] and unsupervised [11]–[18] deep learning paradigms, the problem becomes markedly more complex for stereo image pairs. Enhancing low-light stereo images (LLSIE) is critical for any system reliant on

© 2026, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

3D perception, as it directly underpins the accuracy of depth estimation and scene understanding. Crucially, LLSIE introduces a unique and stringent requirement beyond simple brightness recovery: the preservation of photometric consistency between the enhanced left and right views. Failure to maintain this consistency can introduce artificial disparities, corrupt geometric inference, cause visual discomfort, and degrade the performance of downstream stereo vision tasks.

Current learning-based approaches to LLSIE are predominantly supervised, requiring meticulously aligned pairs of low-light and normal-light stereo images for training [19]–[25]. These methods typically employ dual-stream networks with shared weights and sophisticated cross-view interaction modules to facilitate information exchange. Although effective, their reliance on large-scale, perfectly paired training data is a major bottleneck. Collecting such data across diverse, real-world lighting scenarios is prohibitively expensive and often impractical, severely limiting model generalization to in-the-wild conditions. This data dependency highlights a pressing need for unsupervised LLSIE methods that can learn effective enhancement directly from unpaired or single low-light inputs.

In the realm of unsupervised LLIE, zero-reference methods like Zero-DCE++ [12] have shown remarkable promise. By formulating enhancement as the estimation of image-specific, pixel-wise high-order curves guided only by perceptual losses, they achieve strong performance without paired supervision. However, a naive application of this single-image framework independently to each view of a stereo pair is fundamentally inadequate. Such an approach ignores the intrinsic geometric and photometric correlation between views. Independent curve estimation can easily diverge for corresponding scene points, leading to inconsistencies in brightness, color, and contrast that violate photometric consistency and degrade the stereo experience.

To bridge this critical gap, we propose Stereo-IDCE (Stereo Interactive Dual-Curve Estimation), the first unsupervised framework designed explicitly for low-light stereo image enhancement. Our core insight is to perform collaborative and consistency-aware curve estimation for both views within a unified learning framework. Stereo-IDCE is built upon two synergistic components: an Interactive Dual-Curve Estimation Network (IDCE-Net) and a suite of Consistency-Driven Estimation Loss functions (CDE-Loss).

The IDCE-Net enables collaborative enhancement through a weight-shared encoder-decoder architecture constructed from efficient Nonlinear Activation-Free Blocks (NAF-Blocks) [26]. A key contribution is the new Dual-Cross Interaction Module (DCI-Module), a lightweight, asymmetric fusion unit designed for efficient cross-view feature exchange. This allows the enhancement process for one view to be informed by contextual cues from its counterpart. The network outputs pixel-wise curve parameter maps, which are applied via a Curve-Reused Light-Enhancement (CRLE) process to generate the final high-quality outputs.

To train this network without ground-truth image pairs, the CDE-Loss incorporates established non-reference losses for single-image quality [11], [12] and introduces two new stereo-specific constraints: a Joint Curve Estimation Loss that minimizes discrepancies between the left and right curve maps at the parameter estimation stage, and a Cross-view Exposure Control Loss that aligns the brightness distribution between the final enhanced views. These losses work in concert to explicitly enforce the photometric consistency essential for high-quality stereo vision.

The main contributions of this paper are summarized as follows:

- We formulate the problem of unsupervised low-light stereo image enhancement within an interactive dual-curve estimation framework, explicitly addressing the critical issue of photometric consistency.
- We introduce an Interactive Dual-Curve Estimation Network (IDCE-Net) featuring a new, lightweight Dual-Cross Interaction Module (DCI-Module) to enable efficient and effective cross-view collaborative feature learning.
- We design a new zero-reference Consistency-Driven Estimation Loss (CDE-Loss) to enforce photometric consistency through a Joint Curve Estimation Loss and a Cross-view Exposure Control Loss, guiding the optimization of the IDCE-Net to produce coherent, high-quality enhanced stereo pairs.

Comprehensive experiments on both synthetic and real-world datasets demonstrate that our method effectively enhances low-light stereo images while superiorly maintaining inter-view consistency compared to existing state-of-the-art single-image enhancement methods and naive baselines. The remainder of this paper is organized as follows. Section II reviews related work in low-light stereo image enhancement. Section III details our proposed Stereo-IDCE. Section IV presents the experimental setup, results, and analysis. Finally, Section V concludes the paper.

II. Related Work

A. *Low-light Image Enhancement*

Low-light Image Enhancement (LLIE) aims to improve the visual quality of images captured under insufficient illumination, a critical pre-processing step for numerous computer vision tasks [1], [2]. Deep learning-based approaches to LLIE can be broadly categorized into supervised and unsupervised paradigms.

Supervised methods predominantly rely on paired low/normal-light data for training. A significant line of work is grounded in the Retinex theory, which decomposes an image into a reflectance component (encoding

© 2026, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

scene details) and an illumination component (encoding lighting). Methods like Retinex-Net [3] formalize a two-stage decompose-then-adjust pipeline, while KinD [4] further refines reflectance restoration. Beyond decomposition, DCC-Net [5] proposes an alternative framework that maintains color consistency by separating structure and color adjustments.

Unsupervised methods circumvent the need for paired data, enhancing their practicality. These can be divided into methods using unpaired image collections and those requiring only a single input image. The latter, known as zero-reference methods, are trained solely via carefully designed perceptual losses without any reference images. A seminal work, Zero-DCE [11], models enhancement as the iterative application of learnable, pixel-wise high-order curves. Its lightweight variant, Zero-DCE++ [12], reduces computational cost while maintaining performance. Another approach, SCI [13], employs a fast iterative mechanism based on Retinex principles. The clear framework and strong zero-shot generalization ability of Zero-DCE++ make it a highly extensible foundation, which we build upon in this work.

B. *Low-light Stereo Image Enhancement*

Low-light Stereo Image Enhancement (LLSIE) presents the added challenge of maintaining photometric consistency between the enhanced left and right views, which is crucial for depth perception and 3D reconstruction. Existing learning-based LLSIE methods are primarily supervised, employing shared-weight dual-network architectures with specialized modules for cross-view information exchange.

Research in this domain focuses on two key aspects: the overall network architecture and the design of the stereo interaction module. Pioneering work, DVENet [20], processes images in two stages, i.e., coarse enhancement followed by refinement via stereo interaction. Subsequent methods have introduced more sophisticated interaction mechanisms. DCI-Net [21] proposes Cross-Scale and Cross-View interaction, DGI-Net [22] employs a dynamic gated fusion mechanism for adaptive feature integration, and MSINet [19] introduces a Multi-scale Interaction Module to efficiently propagate information across views and scales.

In contrast, unsupervised methods for LLSIE remain largely unexplored. Our work introduces the first unsupervised framework for LLSIE, extending the curve estimation paradigm to enforce consistency across stereo views without paired supervision.

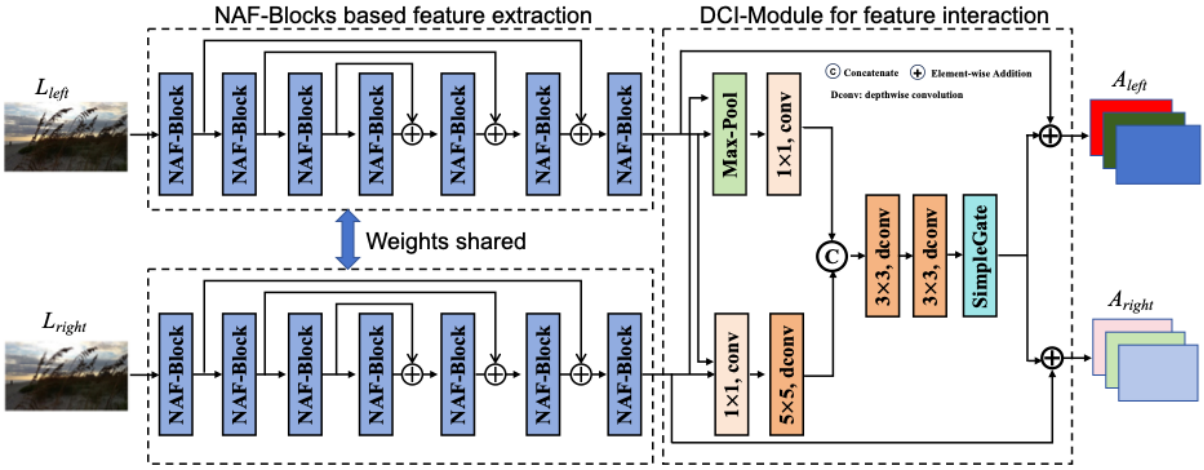

Fig. 1 **The architecture of IDCE-Net.** Our IDCE-Net takes the left and right low-light images as inputs. It leverages stacked Nonlinear Activation-Free Blocks (NAF-Blocks) for effective feature extraction and incorporates a novel Dual-Cross Interaction Module (DCI-Module) to facilitate cross-view feature interaction. The network ultimately outputs a set of pixel-wise curve parameter maps for each view, enabling view-specific yet coordinated enhancement.

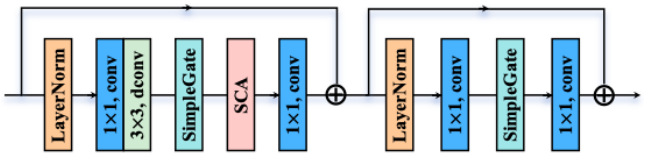

Fig. 2 **The illustration of NAF-Block.**

III. METHOD

This section introduces our proposed unsupervised framework for low-light stereo image enhancement, termed Stereo-IDCE (Stereo Interactive Dual-Curve Estimation). Our framework is built upon two core components: an Interactive Dual-Curve Estimation Network (IDCE-Net) and a set of Consistency-Driven Estimation Loss functions (CDE-Loss). The IDCE-Net takes left and right low-light images as input and collaboratively estimates pixel-wise, high-order light-enhancement curves for both views. These curves are then

© 2026, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

applied iteratively to the RGB channels of the inputs to produce the final enhanced stereo pair. The CDE-Loss functions are designed to evaluate the enhancement quality jointly and, critically, to enforce photometric consistency between the two enhanced views, thereby guiding the unsupervised training of IDCE-Net. Next, we detail the architecture of IDCE-Net and the formulation of CDE-Loss.

A. IDCE-Net

To achieve collaborative and efficient curve estimation for stereo image pairs, we develop the Interactive Dual-Curve Estimation Network (IDCE-Net). As illustrated in Fig. 1, the proposed network stacks multiple Nonlinear Activation-Free Blocks (NAF-Block) and a new Dual-Cross Interaction Module (DCI-Module). IDCE-Net takes the left and right low-light images as inputs and outputs a set of pixel-wise curve parameter maps for each view. Following the efficient design principle of Zero-DCE++ [12], we downsample the input images by a factor of 12 before feeding them into the network. This significantly reduces computational cost while maintaining sufficient representation capacity for curve estimation. The estimated parameter maps are then bilinearly upsampled back to the original input resolution for final pixel-wise enhancement.

NAF-Blocks based feature extraction. Our feature extraction backbone is built upon a U-Net architecture, as depicted in Fig. 1, where both the encoder and decoder consist of multiple stacked NAF-Blocks with symmetric skip connections. Crucially, the weights of these NAF-Blocks are completely shared for processing both the left and right low-light images, ensuring a unified feature representation. Each NAF-Block, following the efficient design principle of NAFNet [26], is architected to operate entirely without conventional nonlinear activation functions (e.g., ReLU, GELU), significantly reducing computational complexity while achieving state-of-the-art performance.

As illustrated in Fig. 2, the NAF-Block comprises two primary modules arranged sequentially: a gated depthwise-attention module and a gated pointwise-ffn module, each followed by a residual connection with learnable scaling. The Gated Depthwise-Attention module focuses on capturing spatial-contextual information. The input is first normalized by Layer Normalization. A Pointwise Convolution expands the channel dimension, followed by a 3×3 Depthwise Convolution to efficiently aggregate spatial features. The SimpleGate operation then introduces nonlinearity via channel-split and element-wise multiplication. Finally, a Simplified Channel Attention (SCA) mechanism, utilizing global average pooling and a linear layer, models channel-wise dependencies. The Gated Pointwise-FFN module performs channel-wise feature transformation. It similarly begins with Layer Normalization. A Pointwise Convolution is applied, followed by the SimpleGate for nonlinear activation, and another Pointwise Convolution projects the features back. Both modules employ a residual structure where the final output is added to a learnably scaled version of the module's input, ensuring stable gradient flow and enabling the construction of deep networks.

DCI-Module for feature interaction. The Dual-Cross Interaction Module (DCI-Module) is a lightweight, asymmetric feature fusion unit designed to facilitate efficient cross-view information exchange between the left and right feature streams within our IDCE-Net, as illustrated in Fig. 1.

The processing of a feature pair within the DCI-Module unfolds as follows. The features first enter two parallel, asymmetric processing paths. The left-view path applies a MaxPooling operation followed by a Pointwise Convolution, which down-samples and then transforms the features to a shared embedding space. Simultaneously, the right-view path first applies a Pointwise Convolution and then a larger receptive field 5×5 Depthwise Convolution, focusing on contextual feature extraction. This asymmetric design allows the module to capture complementary information at different scales from each view. The outputs from both paths are then concatenated along the channel dimension. The concatenated features proceed to a fusion processing unit, which consists of two sequential 3×3 Depthwise Convolution followed with a SimpleGate. This unit is responsible for learning and refining the cross-view interactive information from the merged features. Finally, the refined interactive features are combined with the original input features via a skip-connection. This critical step injects the learned cross-view supplementary cues directly back into each view's feature stream, producing the enhanced output features that now contain information from both perspectives.

© 2026, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

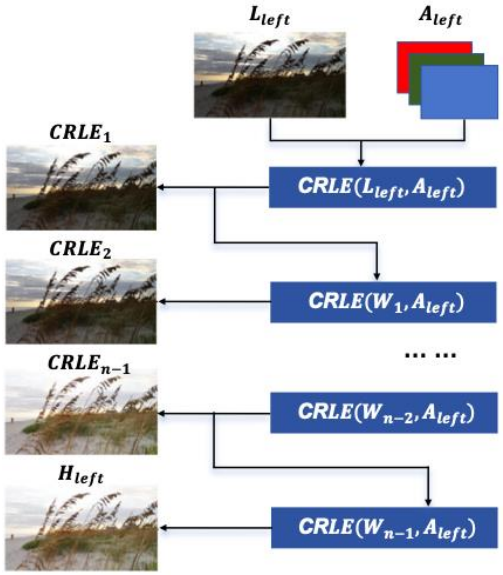

Fig. 3 **The illustration of Curve-Reused Light-Enhancement (CRLE) for left image.** CRLE operates through a sequence of iterative refinements to progressively adjust the dynamic range of an input image, where $\mathcal{A}_{left}$ is reused across all iterations.

B. Curve-Reused Light-Enhancement.

The Curve-Reused Light-Enhancement (CRLE) process forms the core of our pixel-wise enhancement strategy, building upon the iterative design principle established in Zero-DCE++ [12]. As illustrated in Fig. 3, the CRLE framework operates through a sequence of iterative refinements to progressively adjust the dynamic range of an input image. The enhancement processing flow begins with the low-light input images L_{left} and a set of curve parameter maps $\mathcal{A}_{left}$. Then, the enhancement is performed progressively, where the output of each iteration $CRLE_i$ is fed as the input to the subsequent iteration $CRLE_{i+1}$. This iterative application allows a simple curve to simulate a higher-order, more complex adjustment. Following Zero-DCE++ [12], $\mathcal{A}_{left}$ is reused across all iterations for each view. The iterative enhancement process is formulated as:

$$\begin{cases} CRLE_1 = CRLE(L_{left}, \mathcal{A}_{left}), \\ CRLE_i = CRLE(CRLE_{i-1}, \mathcal{A}_{left}), n \geq i \geq 2 \end{cases} \#(1)$$

which $CRLE_i$ represents the output of CRLE at the i-th iteration. In our implementation, the number of iterations n is set to 8, meaning the input image undergoes eight successive enhancement steps, i.e., $H_{left} = CRLE_8$, representing the fully enhanced image. $CRLE_i$ is specifically computed by:

$$CRLE_i = CRLE_{i-1} + \mathcal{A}_{left} \times CRLE_{i-1} \times (1 - CRLE_{i-1}), \#\#\#\#(2)$$

The same process is applied symmetrically to the right view, achieving transformation from low-light L_{right} to its enhanced counterpart H_{right}, ensuring a consistent enhancement pipeline across the stereo pair.

C. CDE-Loss.

The training of our IDCE-Net is guided by a set of Consistency-Driven Estimation Loss Functions (CDE-Loss), which are designed to achieve two primary objectives: 1) to ensure the high perceptual quality of each individually enhanced image, and 2) to critically enforce photometric consistency between the enhanced left and right views. The overall loss function is formulated as a weighted combination of several components, building upon the proven non-reference losses from Zero-DCE++ while introducing new constraints specific to stereo enhancement.

The base of our CDE-Loss incorporates three key loss functions from Zero-DCE++ to guide the single-image enhancement quality:

- **Spatial Consistency Loss** ($\mathcal{L}_{spa}$): This loss encourages the enhanced image to preserve the contrast differences between neighboring regions of the input image. It prevents the output from becoming spatially flat and helps maintain the texture and structural details.
- **Color Constancy Loss** ($\mathcal{L}_{col}$): Based on the Gray-World hypothesis, this loss penalizes color deviations by minimizing the differences between the average intensity values of different color channels (R, G, B), thus promoting a natural color balance in the enhanced result.
- **Illumination Smoothness Loss** ($\mathcal{L}_{tv\mathcal{A}}$): This loss is applied directly to the estimated curve parameter maps $\mathcal{A}$. It enforces smoothness by penalizing the gradients of these maps, which helps preserve the monotonicity relations between adjacent pixels and avoids abrupt, unnatural transitions in enhancement.

© 2026, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

To address the unique requirement of stereo image enhancement, we introduce two new consistency-driven losses that operate across the left and right views:

- **Joint Curve Estimation Loss** ($\mathcal{L}_{jce}$): This loss directly constrains the estimated pixel-wise curve parameter maps for the left and right views ($\mathcal{A}_{left}$ and $\mathcal{A}_{right}$) to be similar. By minimizing the discrepancy between these maps using an L1 norm, we ensure that analogous enhancement operations are applied to corresponding scene points in both views, fundamentally promoting photometric consistency from the estimation stage.
- **Cross-view Exposure Control Loss** ($\mathcal{L}_{cexp}$): This loss extends the concept of the exposure control loss to the stereo domain. It aligns the brightness distribution of corresponding local regions between the enhanced left and right views. This prevents one view from appearing noticeably brighter or darker than the other, which is a common cause of visual discomfort and degradation in stereo perception.

The final CDE-Loss is a combination of the single-view enhancement losses and the stereo consistency losses, guiding the IDCE-Net to produce high-quality, consistent stereo enhancement results without requiring any ground-truth reference images, defined by

$$\mathcal{L}_{total} = \mathcal{L}_{spa} + \mathcal{L}_{cexp} + w_{col} \times \mathcal{L}_{col} + w_{tv\mathcal{A}} \times \mathcal{L}_{tv\mathcal{A}} + w_{jce} \times \mathcal{L}_{jce}, \#(3)$$

where w_{col}, $w_{tv\mathcal{A}}$ and w_{jce} are used for balancing the scales of different losses.

IV. Experiments and Results

A. Experimental Setting

1) Datasets: Training and Testing Datasets. During the training phase, following the data preparation protocol of iPASSR, we construct our training set from the Middlebury [27] and Flickr1024 [28] datasets, comprising 860 stereo image pairs. The images are predominantly in portrait orientation, with a typical resolution of approximately 300×500 pixels.

For evaluation, experiments are first conducted on synthetic datasets. The synthetic test set is also derived from Middlebury and Flickr1024, containing 10 and 224 image pairs, respectively. For evaluation under real-world conditions, we select 100 representative stereo pairs from the Holopix50k [29] dataset to obtain the In-The-Wilds test set, encompassing diverse and challenging low-light scenarios such as night scenes, backlit conditions, and scenes with strong localized light sources.

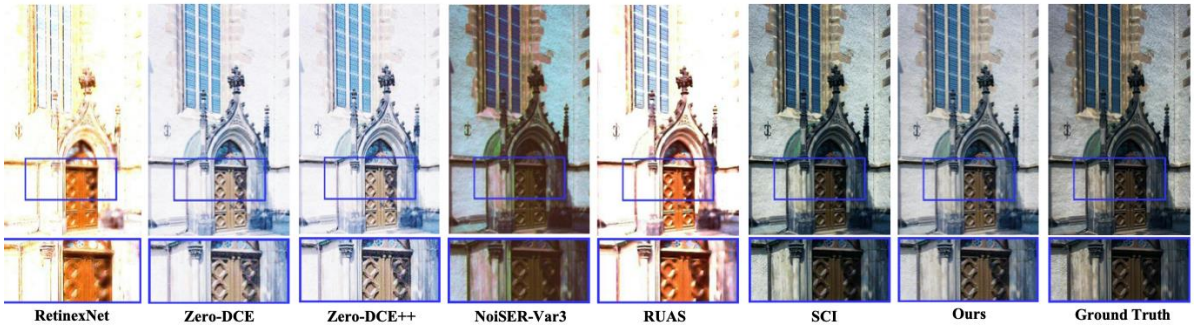

Fig. 4 **Visual results on the Flickr1024 dataset.** Our method effectively avoids overexposure and exhibits minimal color shift or deviation, as highlighted in the blue boxes.

TABLE I

THE RESULTS ARE ON FLICKR1024. OUR METHOD OUTPERFORMS ALL OTHER BASELINES. THE BEST RESULTS ARE **BOLD** AND THE SECOND BEST RESULTS ARE UNDERLINE.

Method	(left+right)/2			left		
	PSNR↑	SSIM↑	MAE↓	PSNR↑	SSIM↑	MAE↓
Retinex-Net [3]	10.58	0.424	0.130	10.56	0.424	0.130
Zero-DCE [11]	11.37	0.437	0.086	11.37	0.437	0.086
Zero-DCE++ [12]	10.49	0.413	0.115	10.49	0.413	0.111
RUAS [14]	13.91	0.434	0.075	13.89	0.433	0.076
NoiSER-Var3 [15]	13.85	0.544	0.049	14.05	0.535	0.051
NoiSER-ES [15]	14.06	0.536	0.051	14.05	0.535	0.051
NoiSER-FC [15]	14.11	0.547	0.051	14.12	0.546	0.051
SCI [13]	17.67	0.574	0.032	17.67	0.574	0.032
Ours	**18.82**	**0.601**	**0.020**	**18.84**	**0.601**	**0.019**

© 2026, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

TABLE II
THE RESULTS ARE ON MIDDLEBURY. OUR METHOD OUTPERFORMS ALL OTHER BASELINES. THE BEST RESULTS ARE **BOLD** AND THE SECOND BEST RESULTS ARE UNDERLINE.

Method	(left+right)/2			left		
	PSNR↑	SSIM↑	MAE↓	PSNR↑	SSIM↑	MAE↓
Retinex-Net [3]	10.17	0.512	0.155	10.19	0.514	0.153
Zero-DCE [11]	12.13	0.608	0.099	12.20	0.607	0.098
Zero-DCE++ [12]	9.802	0.554	0.146	9.901	0.553	0.144
RUAS [14]	14.71	0.597	0.070	14.72	0.597	0.071
NoiSER-Var3 [15]	15.39	0.694	0.031	15.53	0.698	0.030
NoiSER-ES [15]	15.39	0.729	0.031	15.54	0.731	0.030
NoiSER-FC [15]	15.39	0.694	0.031	15.53	0.698	0.030
SCI [13]	16.24	0.709	0.036	16.24	0.708	0.037
Ours	**18.92**	**0.743**	**0.031**	**18.9**	**0.741**	**0.031**

2) Metrics: We evaluate the enhanced results using standard full-reference metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Mean Absolute Error (MAE), following the evaluation protocol of prior works [19-22]. Higher PSNR/SSIM and lower MAE values indicate better enhancement quality.

3) Implementation details: For hyperparameter configuration, Stereo-IDCE follows the same learning rate and optimizer settings as Zero-DCE++: a fixed learning rate of $1e^{-4}$ and the Adam optimizer [30]. During training, we use a batch size of 1 and train the model for 50 epochs. w_{col}, $w_{tv\mathcal{A}}$ and w_{jce} in the loss function are set to 0.05, 0.2, 0.01, respectively. All experiments are implemented using the PyTorch framework and conducted on a single GeForce RTX 4090 GPU.

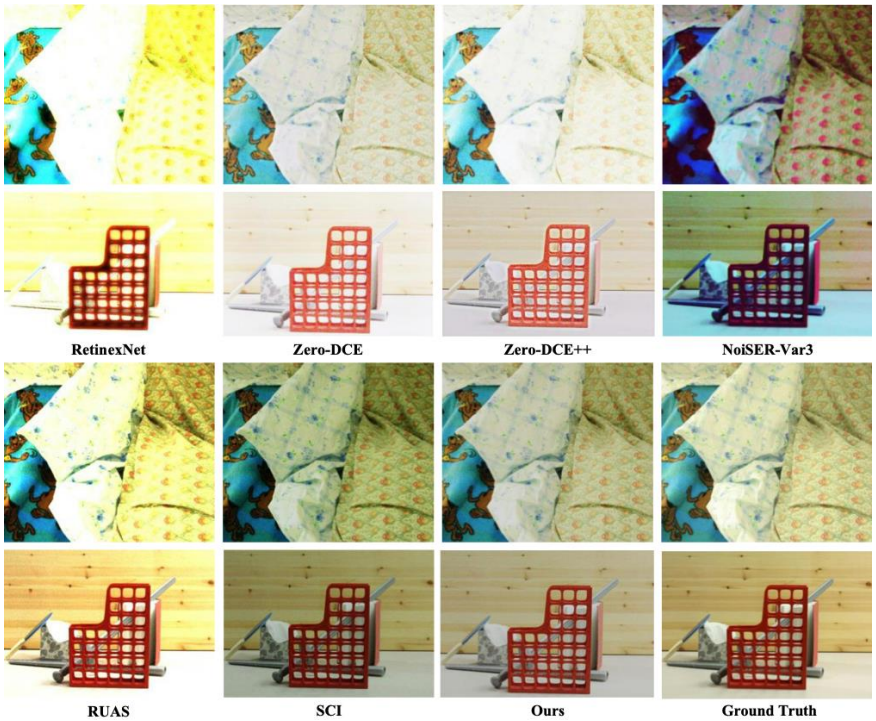

Fig. 5 **Visual results on the Middlebury datasets.** Our method produces results with only minimal brightness deviation from the ground truth, yet it achieves perceptually more natural and visually comfortable enhancements.

4) Compared Baselines: Since, to the best of our knowledge, no existing unsupervised methods are specifically designed for the LLSIE task, our comparative experiments are conducted against state-of-the-art unsupervised LLIE methods. We select five representative and influential baselines, which are briefly introduced as follows: (1) **Zero-DCE** [11] pioneers a zero-reference learning framework by formulating LLIE

© 2026, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

as a task of estimating image-specific, pixel-wise high-order curves, enabling flexible pixel-wise adjustment; (2) **Zero-DCE++** [12] is a lightweight and accelerated variant of Zero-DCE. Its primary advancement lies in a more efficient network architecture, leading to significantly reduced parameters and faster inference; (3) **RUAS** [14] is an unsupervised method based on the Retinex theory. It employs a specially designed optimization strategy that requires minimal computational resources for fast model convergence and inference. (4) **SCI** [13] achieves stable brightness enhancement by designing a self-calibrated illumination learning process with a dedicated loss for illumination smoothness. It utilizes a two-stage framework for training and inference to ensure rapid processing; (5) **NoiSER** [15] is a highly innovative work whose core contribution involves training an enhancement network using randomly sampled noise images. It innovatively introduces an autoregressive training paradigm into LLIE, grounded in the classic gray-world hypothesis [31]. For a more comprehensive comparison, we also include a classic supervised method as a reference: **Retinex-Net** [3] is a fundamental supervised method based on the Retinex theory. Its processing pipeline consists of two distinct stages: reflectance decomposition and subsequent illumination adjustment.

B. Comparison on the synthetic dataset

Results on Flickr1024. The quantitative results are listed in Table I. The PSNR of Zero-DCE++ is notably low, primarily due to severe noise amplification and overexposure issues. SCI achieves higher PSNR and SSIM scores than Zero-DCE++, while our method performs even better. The visual comparison on the Flickr1024 dataset is shown in Fig. 4. As observed, the results from Zero-DCE and Zero-DCE++ exhibit noticeable overexposure. The NoiSER method introduces visible color casts that deviate from natural visual perception. While the SCI method effectively suppresses noise, its results appear undesirably dark with a predominant bluish-black tone. In contrast, our method successfully maintains appropriate brightness, avoids overexposure and unnatural color shifts, and yields visually more pleasing tones.

Results on Middlebury. The quantitative results are shown in Table II. The performance of NoiSER shows some improvement on this dataset, indicating that its efficacy is data-dependent and leaves room for optimization. By contrast, our method consistently achieves better results. The visual results on the Middlebury dataset are presented in Fig. 5. The overexposure in Zero-DCE's result is more pronounced here, significantly washing out details. Our method produces better results in recovering vivid details from dark regions while effectively avoiding overexposure in brighter areas.

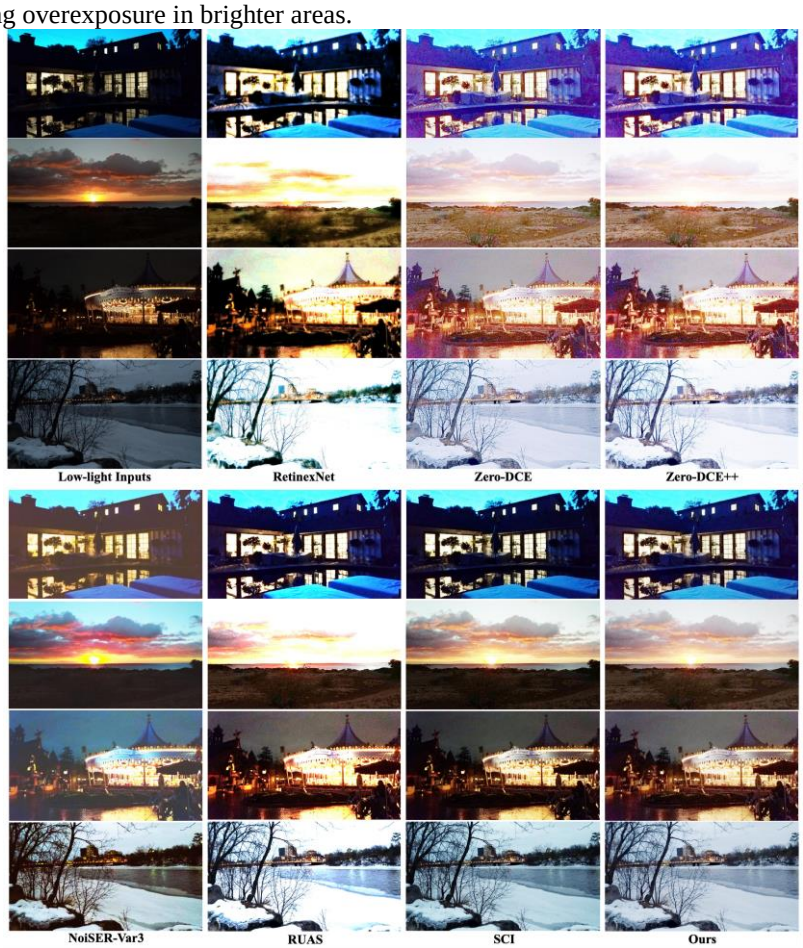

Fig. 6 **Visual results on In-the-Wild Data.** Our method successfully recovers vivid details from dark regions while effectively avoiding overexposure in brighter areas.

© 2026, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

C. *Comparison on the In-The-Wilds dataset*

Visual results on this dataset are presented in Fig. 6. The output from Zero-DCE++, while appearing subjectively brighter, demonstrates a critical flaw: it fails to preserve fine details in high-intensity regions. For instance, in the second image, cloud textures and the sun are almost completely lost, indicating that its aggressive brightening comes at the cost of compromised image content. Both NoiSER and SCI produce results that are perceptually dark and exhibit noticeable color casts on this real-world data. In contrast, our method achieves a more balanced enhancement, effectively improving visibility in dark regions while robustly avoiding overexposure and maintaining natural color fidelity across all tested scenarios.

TABLE III
THE RESULTS OF ABLATION STUDY ON FLICKR1024. BOTH DCI-MODULE AND CDE-LOSS MATTERS FOR LOW-LIGHT STEREO IMAGE ENHANCEMENT.

Method	(left+right)/2		
	PSNR↑	SSIM↑	MAE↓
Base	10.49	0.413	0.115
w/o DCI-Module	10.88	0.425	0.101
w/o CDE-Loss	18.03	0.597	0.028
Ours	**18.82**	**0.601**	**0.020**

D. *Ablation study*

We conduct an ablation study to quantitatively validate the importance of the two core components designed to enforce photometric consistency: the Dual-Cross Interaction Module (DCI-Module) in the IDCE-Net and the Consistency-Driven Estimation Loss (CDE-Loss). Three baselines are compared: 1) *Base*: A naive baseline where the left and right views are enhanced independently using the original Zero-DCE++ pipeline, with no interaction module or stereo-specific losses. 2) *w/o DCI-Module*: Our full framework but with the DCI-Module removed from the IDCE-Net, disabling explicit cross-view feature interaction. 3) *w/o CDE-Loss*: Our full network but trained solely with the original Zero-DCE++ losses, excluding our proposed Joint Curve Estimation and Cross-view Exposure Control losses.

The quantitative results are presented in Table III. Both the *w/o DCI-Module* and *w/o CDE-Loss* variants outperform the independent *Base* model, confirming that even partial integration of our stereo-aware design improves upon separate processing. However, they underperform our complete Stereo-IDCE method, which achieves the best scores. This ablation study clearly demonstrates that both the DCI-Module and the CDE-Loss are critical for optimal performance. The DCI-Module enables effective feature-level collaboration, while the CDE-Loss provides essential loss-level constraints; together, they are indispensable for ensuring high-quality photometric consistency in unsupervised low-light stereo image enhancement.

V. CONCLUSION

In this paper, we have presented the first unsupervised framework for LLSIE, named Stereo-IDCE. Our core contribution is the interactive dual curve estimation paradigm, which allows for high-dimensional, pixel-wise adjustment of both views through collaboratively estimated enhancement curves. To ensure photometric consistency, we introduce a lightweight cross-view feature interaction module that facilitates efficient information exchange between views with minimal computational overhead. Furthermore, we propose new consistency-driven loss functions, including a joint curve estimation loss and a cross-view exposure control loss, which effectively mitigate common artifacts such as overexposure and noise amplification without requiring any ground-truth supervision. Extensive experiments on both synthetic and real-world data demonstrate that our method achieves a superior performance compared to existing baselines.

REFERENCES

[1]. W. Liu, P. Zhao, X. Song, and B. Zhang, "A survey of low-light image enhancement," *Frontiers in Computing and Intelligent Systems*, vol. 1, no. 3, pp. 88–92, 2022.

[2]. Q. Zhao, G. Li, B. He, and R. Shen, "Deep learning for low-light vision: A comprehensive survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, pp. 15685–15705, 2025.

[3]. C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint* arXiv:1808.04560, 2018.

[4]. Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. on Multimedia*, 2019, pp. 1632–1640.

© 2026, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

[5]. Z. Zhang, H. Zheng, R. Hong, M. Xu, S. Yan, and M. Wang, "Deep color consistent network for low-light image enhancement," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2022, pp. 1899–1908.

[6]. W. Wu, J. Pang, S. Gao, P. Zhang, Z. Chen, and G. Qiu, "Low-light image enhancement through learning a simplified inverse rendering model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 10, pp. 9608–9622, 2025.

[7]. M. Liu, Y. Cui, W. Ren, J. Zhou, and A. C. Knoll, "LIEDNet: A lightweight network for low-light enhancement and deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, pp. 6602–6615, 2025.

[8]. H. Xie, K. Qian *et al.*, "Research on low-light image enhancement algorithm based on attention mechanism," *Journal of Advanced Computing Systems*, vol. 5, no. 5, pp. 1–14, 2025.

[9]. T. Zhang, P. Liu, Y. Lu, M. Cai, Z. Zhang, Z. Zhang, and Q. Zhou, "CWNet: Causal wavelet network for low-light image enhancement," in *Proc. IEEE/CVF Int. Conf. on Computer Vision*, 2025, pp. 8789–8799.

[10]. L. Ma, T. Ma, C. Xu, J. Liu, X. Fan, Z. Luo, and R. Liu, "Learning with self-calibrator for fast and robust low-light image enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, pp. 9095–9112, 2025.

[11]. C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2020, pp. 1780–1789.

[12]. C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4225–4238, 2021.

[13]. L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2022, pp. 5637–5646.

[14]. R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2021, pp. 10561–10570.

[15]. Z. Zhang, S. Zhao, X. Jin, M. Xu, Y. Yang, and S. Yan, "NoiSER: Noise is all you need for low-light image enhancement," *CoRR*, 2022.

[16]. X. Wang, Y. Wang, J. Zhou, J. Liu, Y. Gao, Y. Wang, and J. Zheng, "An unsupervised learning method based on U-Net++ for low-light image enhancement," *Signal, Image and Video Processing*, vol. 19, no. 4, p. 282, 2025.

[17]. C. Da, Y. Qian, J. Zeng, X. Wei, and F. Zhang, "MCATD: Multi-scale contextual attention transformer diffusion for unsupervised low-light image enhancement," *IEEE Access*, vol. 13, pp. 97 348–97 363, 2025.

[18]. X. Xu and J. Du, "Unsupervised low-light image enhancement based on curve estimation and illumination perception," *Signal, Image and Video Processing*, vol. 19, no. 8, p. 614, 2025.

[19]. Z. Ji, H. Zheng, Z. Zhang, Q. Ye, Y. Zhao, and M. Xu, "Multi-scale interaction network for low-light stereo image enhancement," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 3626–3634, 2023.

[20]. J. Huang, X. Fu, Z. Xiao, F. Zhao, and Z. Xiong, "Low-light stereo image enhancement," *IEEE Transactions on Multimedia*, pp. 1–1, 2022.

[21]. H. Zheng, Z. Zhang, J. Fan, R. Hong, Y. Yang, and S. Yan, "Decoupled cross-scale cross-view interaction for stereo image enhancement in the dark," in *Proc. 31st ACM Int. Conf. on Multimedia*, 2023, pp. 1475–1484.

[22]. B. Li, H. Zheng, Z. Zhang, Y. Zhao, Z. Zhao, and H. Zhang, "Dynamic grouped interaction network for low-light stereo image enhancement," in *Proc. 31st ACM Int. Conf. on Multimedia*, 2023, pp. 2468–2476.

[23]. S. Du, X. Qin, M. Zhao, J. Hu, L. Wang, and P. Li, "Non-uniform low-light stereo image enhancement based on transformer," in *Proc. 2nd Asia Symp. on Image and Graphics*, 2024, pp. 164–170.

[24]. S. Du, S. Yan, Z. Shi, Z. You, and L. Sun, "Wavelet-based decoupling framework for low-light stereo image enhancement," *arXiv preprint* arXiv:2507.12188, 2025.

[25]. M. Zhao, X. Qin, S. Du, X. Bai, J. Lyu, and Y. Liu, "Low-light stereo image enhancement and de-noising in the low-frequency information enhanced image space," *Expert Systems with Applications*, vol. 265, p. 125803, 2025.

[26]. L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *European Conf. on Computer Vision*. Springer, 2022, pp. 17–33.

[27]. D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," in *German Conf. on Pattern Recognition*. Springer, 2014, pp. 31–42.

[28]. L. Wang, Y. Wang, Z. Liang, Z. Lin, J. Yang, W. An, and Y. Guo, "Learning parallax attention for stereo image super-resolution," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2019, pp. 12 250–12 259.

[29]. Y. Hua, P. Kohli, P. Uplavikar, A. Ravi, S. Gunaseelan, J. Orozco, and E. Li, "Holopix50k: A large-scale in-the-wild stereo image dataset," *arXiv preprint* arXiv:2003.11172, 2020.

[30]. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conf. on Learning Representations*, 2015.

[31]. G. Buchsbaum, "A spatial processor model for object colour perception," *Journal of the Franklin Institute*, vol. 310, no. 1, pp. 1–26, 1980.

© 2026, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden