IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Zero – Reference Deep Learning forLow-Light Image Enhancement: A Systematic Literature Survey and Research Gap Analysis

[1] Sonali Sunil Gitte, [2]Dr.Sushil Venkatesh Kulkarni,

[1]M.tech 3rd Semester Student, Department of Computer Science and Information Technology, MBES College of Engineering, Ambajogai, India

[2] Head and Guide, Department of Computer Science and Information Technology, MBES College of Engineering, Ambajogai, India

Abstract: Low-light image enhancement (LLIE) remains a fundamental challenge in computer vision, with critical implications for surveillance, autonomous navigation, medical imaging, and consumer photography. This literature survey systematically examines the evolution of LLIE methodologies from classical histogram-based and Retinex-inspired techniques through supervised deep learning to zero-reference unsupervised paradigms. The survey evaluates contemporary performance benchmarks, outlines the mathematical formulations driving zero-reference non-reference loss functions, and highlights data domains spanning paired indoor settings and unpaired real-world datasets. A comprehensive research gap analysis identifies critical deficiencies in existing zero-reference implementations, specifically focusing on the lack of integrated post-enhancement quality metrics, computational profiles for resource-constrained architectures, and multi-benchmark generalization capabilities.

Keywords: Low-light image enhancement, Zero-DCE, deep curve estimation, visibility classification, unsupervised learning, PSNR, SSIM, image quality assessment, Retinex theory, histogram equalization, computer vision.

1. Introduction

1.1. Background and Motivation

Digital imaging systems fundamentally depend on adequate illumination to capture scenes with sufficient detail and clarity. Real-world imaging under low-light conditions including nighttime surveillance, indoor environments, medical endoscopy, and autonomous driving produces images degraded by compressed dynamic range, amplified sensor noise, reduced contrast, and chromatic distortion. These degradations not only impair human visual perception but also substantially degrade the performance of downstream computer vision pipelines such as object detection, face recognition, semantic segmentation, and scene understanding.

The commercial and safety-critical importance of low-light imaging has driven sustained research investment over five decades. Classical signal processing approaches, while computationally inexpensive, operate under simplified assumptions that break down in real-world non-uniform illumination. Modern deep learning methods have demonstrated transformative performance gains but introduce their own challenge most supervised approaches require carefully curated paired training datasets of low-light

images matched with corresponding ground-truth normal-light counterparts that are prohibitively expensive to acquire at scale.

Zero-Reference Deep Curve Estimation (Zero-DCE), introduced by Guo et al. at CVPR 2020[1], represents a watershed contribution by eliminating any reference image requirement. The method trains a lightweight CNN (~79K parameters) to estimate pixel-wise higher-order tonal curves using four non-reference loss functions encoding perceptually meaningful quality criteria. This survey provides a comprehensive review of the LLIE field leading to Zero-DCE and presents an extended framework Zero-DCE+ with integrated visibility classification, systematic evaluation, and interactive web deployment.

1.2. Characterization of Low-Light Degradation

Low-light image degradation can be formally modeled using the Retinex framework[2]. An observed image S(x,y) is expressed as the element-wise product of scene reflectance R(x,y) and illumination L(x,y):

$$S(x, y) = R(x, y) \times L(x, y) \quad (1)$$

Under low-light conditions, L(x,y) << 1 for most pixel locations, compressing reflectance values toward zero and masking scene content. The primary degradation phenomena are:

•Intensity compression: Pixel values cluster near the lower end of the dynamic range [0, 255], reducing effective bit depth.

•Noise amplification: The photon shot noise variance equals the mean photon count; at low intensities, the SNR degrades quadratically, producing visually prominent luminance and chrominance noise.

•Colour distortion: Non-uniform spectral sensitivity of image sensors at low illumination introduces systematic colour casts that differ from camera to camera.

•Detail loss: Fine spatial structures fall below the signal-to-noise floor, appearing merged with noise.

1.3. Algorithmic Taxonomy and Evolution Roadmap

The LLIE field has evolved through four broad algorithmic generations, characterized by shifting dependencies on physics models and training targets:

Era	Paradigm	Representative Methods	Key Limitation
1970s–2000s	Statistical Mapping	Histogram Equalization (HE), CLAHE, Gamma Correction	No spatial/semantic model
2000s–2015	Physics-Based	MSRCR, LIME, SSR, MSR	Halo artifacts; over-saturation
2015–2019	Supervised DL	LLNet, Retinex-Net, SID (Chen)	Requires paired training data
2019–2021	Unpaired/GAN	EnlightenGAN, MBLLEN	Large models; unstable training
2020–Present	Zero-Reference	Zero-DCE, RUAS, SCI	Fixed thresholds; no quality assessment
2023–Present	**Transformer-Based**	Retinexformer	High compute; GPU-only
2023-Present	Generative Diffusion	Diff-LL, PyDiff	Iterative sampling; not real-time

Table 1.1: Algorithmic taxonomy and evolution roadmap for LLIE

1.4. Structure of the Survey

The remainder of this survey is organized as follows. Section II critically examines traditional enhancement methods. Section III reviews deep learning-based taxonomies, including supervised frameworks, GAN-based methods, zero-reference approaches and frontier transformer and diffusion-based models. Section IV formalizes the mathematical model of non-reference loss functions. Section V covers image quality assessment and visibility classification. Section VI synthesizes quantitative benchmarks. Section VII presents the research gap analysis. Section VIII discusses open challenges and future scope, and Section IX concludes.

2. Traditional Image Enhancement Method

2.1. Statistical Intensity Mapping (Histogram Modifications)

2.1.1Global Histogram Equalization (GHE)

Global Histogram Equalization (GHE) is the most widely known intensity transformation for contrast enhancement. Given an input image I with normalized histogram p_r(r_k) at intensity level r_k, GHE applies a monotonic tone mapping T derived from the cumulative distribution function (CDF):

$$s_k = T(r_k) = (L-1) \cdot \Sigma_{j=0}^{k} \; p_r(r_j) \qquad (2)$$

Where L is the total number of intensity levels (256 for 8-bit), and the summation over the normalized histogram estimates the CDF. While GHE is computationally trivial (O (N) in image size) and requires no parameters, it exhibits three critical pathologies under low-light conditions: (1) over-enhancement of already bright regions, (2) sensitivity to outlier intensities that skew the CDF, and (3) artificial color shifts when applied independently per RGB channel. These pathologies are particularly severe in images where pixel intensities are concentrated in the lowest 5–10% of the dynamic range.

2.1.2 Contrast Limited Adaptive Histogram Equalization (CLAHE)

CLAHE (Zuiderveld, 1994)[3] partitions the image into a grid of non-overlapping contextual tiles and applies histogram equalisation locally within each tile, subject to a configurable clip limit C_lim that caps the histogram height before redistribution. This prevents the noise amplification that occurs when a narrow intensity peak is globally amplified. Bilinear interpolation between tile centers eliminates block artifacts at tile boundaries, yielding smooth transitions.

CLAHE achieves substantially better noise suppression than GHE and has become the standard preprocessing step in medical imaging (retinal fundus photography, endoscopy) and surveillance systems. However, it introduces two remaining limitations: (1) the clip limit and tile size are free parameters requiring domain-specific tuning, and (2) the local histogram model does not incorporate physical illumination constraints, leading to inconsistent enhancement across scenes with extremely non-uniform illumination.

2.2. Physics-Based Retinex Restorations

2.2.1Multi-Scale Retinex with Color Restoration (MSRCR)

Retinex theory, introduced by Land (1977)[4], posits that perceived luminance is determined by reflectance rather than absolute intensity: S = R × L. MSRCR (Jobson et al., 1997)[5] estimates and suppresses the illumination component by computing a weighted sum of log-domain differences between the image and its Gaussian-filtered surround at multiple spatial scales σ_k:

$$R_i(x,y) = \Sigma_k \; w_k \cdot [\, \log I_i(x,y) \; - \; \log(G_\sigma_k * I_i(x,y)) \,] \qquad (3)$$

Where I_i is the i-th color channel, G_σ_k is a Gaussian kernel with scale σ_k, and w_k are scale weights summing to unity. Color restoration is applied post-hoc to compensate for the desaturation inherent in single-channel Retinex processing. MSRCR provides visually compelling enhancement for moderately dark scenes but introduces halo artifacts near high-contrast edges and produces over-saturated colors under extreme illumination variation.

2.2.2Handcrafted Illumination Mapping (LIME)

LIME (Guo et al., 2017)[6] improves upon MSRCR by constructing a refined illumination map T through structure-guided Total Variation (TV) minimization. The initial illumination estimate T_hat(x) = max_c I_c(x) is refined by solving:

$$\min_T \|T - \hat{T}\|_F^2 + \alpha \cdot \|W \odot \nabla T\|_1 \quad (4)$$

Where W is a structure-aware weighting matrix that preserves edges by reducing regularization at high-gradient locations. LIME consistently outperforms MSRCR in NIQE and perceptual quality metrics but retains three unresolved limitations: (1) the TV minimization is an iterative optimization that is computationally expensive for high-resolution images; (2) the structure weighting is hand-designed and may fail under extreme noise; and (3) neither LIME nor MSRCR incorporates learned priors about natural image statistics.

2.3. Performance Limitations of Non-Learning Baselines

Method	PSNR (dB)	SSIM	Primary Limitation
GHE	~14.0	~0.45	Over-amplifies noise; colour artefacts
CLAHE	15.47	0.582	Parameter-sensitive; no illumination model
MSRCR	~16.0	~0.54	Halo artefacts; colour over-saturation
LIME	~16.5	~0.56	Slow TV optimisation; hand-designed weights

Table 2.1: Performance limitations of traditional non-learning baselines on LOL dataset

3. Deep Learning Based LLIE Taxonomies

3.1. Data Driven Supervised Frameworks

3.1.1Early Autoencoders (LLNet)

LLNet (Lore, Akintayo, and Sarkar, 2017)[7] was among the first deep architectures designed specifically for low-light enhancement. The method employs a stacked sparse denoising autoencoder (SSDA) trained on synthetically darkened patches drawn from natural image datasets. A branching architecture enables joint brightness enhancement and noise suppression through separate decoder paths sharing a common encoder representation.LLNet's autoencoder design draws on representation learning principles established by deep CNN classifiers such as AlexNet [14], which demonstrated that learned hierarchical features outperform hand-crafted descriptors

LLNet demonstrated empirically that learned hierarchical representations substantially outperform hand-crafted image statistics for the enhancement task. However, its reliance on synthetically generated training pairs introduces a domain gap: the synthetic darkening model (gamma correction with additive Gaussian noise) inadequately captures the complex noise profiles, colour shifts, and spatially non-uniform illumination characteristic of real-world low-light cameras.

3.1.2 Physical Decomposition Sub-networks (Retinex-Net)

Retinex-Net (Wei et al., 2018)[8] integrates the physical Retinex illumination model into an end-to-end trainable deep learning pipeline. The architecture comprises two sub-networks trained jointly on the LOL dataset: Decom-Net, which decomposes the input image S into reflectance map R ∈ ℝ^(H×W×3) and

illumination map L ∈ ℝ^(H×W×1) via a three-layer encoder-decoder; and Enhance-Net, which applies learned adjustments to the illumination map to achieve the desired brightness while preserving reflectance.

Retinex-Net achieves PSNR = 16.77 dB and SSIM = 0.560 on the LOL test set using ~420K parameters. Critically, it established the LOL dataset as the primary LLIE benchmark. Its primary limitation is the strict requirement for paired low/normal-light training data, which is costly to collect and introduces alignment artefacts from slight camera motion between the paired exposures.

3.1.3 Raw Sensor Transformations (See-In-The-Dark)

Chen et al. (CVPR 2018) [9]proposed learning directly on RAW sensor data from the See-in-the-Dark (SID) dataset, mapping short-exposure RAW Bayer patterns to long-exposure sRGB outputs using a fully convolutional U-Net with instance normalization. By operating in the RAW domain, the method avoids the irreversible noise amplification introduced by the camera's image signal processor (ISP), achieving state-of-the-art PSNR on extreme low-light conditions (< 0.1 lux equivalent).

The SID approach demonstrates the importance of sensor-level processing for extreme low-light conditions. However, its dependence on specialized RAW data acquisition workflows, large paired training sets (5,094 RAW image pairs), and GPU-intensive training (multi-day on high-end GPUs) fundamentally restricts practical deployment to controlled research settings. The fully convolutional architecture used in SID is based on the U-Net design [16], originally developed for biomedical image segmentation and later widely adopted for image-to-image translation tasks

3.1.4 Deep Unfolding Retinex Networks (URetinex-Net)

URetinex-Net (Wu et al., CVPR 2022)[11] reformulates Retinex decomposition as a deep unfolding optimization problem, unrolling iterative proximal gradient updates into learnable network layers. Unlike Retinex-Net's purely data-driven decomposition, URetinex-Net embeds explicit optimization constraints for reflectance and illumination estimation directly into the network architecture, combining the interpretability of model-based methods with the representational power of deep learning. It reports improved PSNR over Retinex-Net on LOL while requiring paired training data, situating it within the supervised paradigm but with stronger physical grounding than purely black-box CNN decompositions.

3.2. Unsupervised and Generative Adversarial Networks (GANs)

3.2.1 Unpaired Distribution Mapping (EnlightenGAN)

EnlightenGAN (Jiang et al., 2021)[10] applies adversarial training in an unpaired manner, learning to map the statistical distribution of low-light images to that of normal-light images without any correspondence between training examples. The architecture employs a global-local discriminator: a global discriminator evaluates whole-image realism, while a local patch discriminator enforces local perceptual quality. Self-feature-preserving (SFP) perceptual loss using VGG features maintains content consistency. EnlightenGAN's adversarial training scheme builds directly on the foundational GAN framework [15], adapting the generator-discriminator paradigm to unpaired image enhancement.

EnlightenGAN achieves PSNR = 17.48 dB and SSIM = 0.651 on LOL, representing a meaningful improvement over paired-supervision Retinex-Net despite using no paired data. Its primary limitations are: (1) approximately 8.6M parameters — more than 100× larger than Zero-DCE — precluding deployment on resource-constrained hardware; (2) adversarial training instability requiring careful hyper parameter tuning; and (3) occasional mode collapse producing enhancement artifacts on out-of-distribution inputs.

3.3. Zero-Reference Parametric Optimization

3.3.1 Iterative Deep Curve Estimation (Zero-DCE)

Zero-DCE (Guo et al., CVPR 2020)[1] introduced a paradigm-shifting zero-reference formulation in which image enhancement is framed as the task of estimating pixel-wise higher-order curves, trained exclusively via non-reference loss functions without any ground-truth images. The seven-layer CNN with symmetric skip connections estimates 24 curve parameter maps (8 iterations × 3 channels), which are applied iteratively via the recurrence:

$$LE^{(n)}(x) = LE^{(n-1)}(x) + \alpha^{(n)}(x) \cdot LE^{(n-1)}(x) \cdot (1 - LE^{(n-1)}(x)), \quad n = 1,...,8 \qquad (5)$$

The original Zero-DCE paper reports PSNR = 14.86 dB on the LOL dataset. While its baseline execution scores lower than fully supervised alternatives, its zero-reference structure eliminates spatial alignment distortions and paired data requirements. The original Zero-DCE architecture provides headroom for improvement through training configuration changes alone, without modifying network structure, while retaining its minimal parameter scale (~79,416 parameters) deployable on commodity hardware.

3.3.2 Lightweight Successor Architecture (Zero-DCE++)

Zero-DCE++ (Li et al., TPAMI 2021)[27] extends the original Zero-DCE by replacing standard convolutions with depthwise separable convolutions, reducing the parameter count from ~79,416 to approximately 10,561 — a nearly 8× reduction. The architecture retains the same iterative curve estimation formulation (Equation 5) and the identical four non-reference loss functions, but achieves significantly faster inference (over 3× speedup) with only marginal quality degradation (PSNR within ~0.3 dB of the original on LOL). This makes Zero-DCE++ particularly relevant for mobile and embedded deployment, directly addressing the resource-constraint motivations central to this survey. Its real-time performance on mobile GPUs (>1000 FPS at low resolution) represents the practical efficiency ceiling for the zero-reference paradigm and is a critical baseline against which any proposed CPU-deployment system should be benchmarked.

3.3.3 Accelerated Structural Transformations (RUAS & SCI)

RUAS (Liu et al., CVPR 2021)[22] advances zero-reference enhancement through neural architecture search (NAS), automatically discovering compact network structures tailored to the low-light enhancement objective under tight parameter budgets. The search process optimises a differentiable proxy loss that combines enhancement quality and model efficiency. SCI (Ma et al., CVPR 2022) achieves real-time enhancement via a self-calibrated illumination framework that estimates and corrects illumination without explicit decomposition. Both methods demonstrate the continued maturation of the zero-reference paradigm, though neither natively pairs these modules with automated quality profiling indices or cross-stage classification frameworks.

3.4. Attention-Based and Generative Frontier Models (2023–Present)

The landscape of low-light image enhancement shifted significantly with the adaptation of Vision Transformers (ViTs) and Generative Diffusion Models, representing a fifth algorithmic generation beyond the zero-reference paradigm.

3.4.1 Vision Transformer Decomposition (Retinexformer)

Retinexformer (Hou et al., ICCV 2023)[23] introduced self-attention mechanisms into Retinex decomposition, utilizing a Damage-Reflective Attention map to identify and restore low-light degradation without amplifying noise. Unlike CNN-based approaches that rely on local receptive fields, the transformer architecture captures long-range pixel dependencies, enabling globally consistent illumination correction across the entire image.

3.4.2 Diffusion-Based Generative Enhancement

Generative paradigms such as Diff-LL and PyDiff[24] reframe low-light enhancement as a conditional denoising diffusion process. These models iteratively reverse a forward degradation process, synthesizing

highly realistic textures and natural illumination maps in ultra-dark environments. While these frameworks define current peak SOTA for PSNR and SSIM benchmarks, their massive parameter counts and iterative sampling requirements impose steep computational costs, making them impractical for edge or real-time deployment.

3.4.3 Positioning in the Taxonomy

Generation	Paradigm	Representative Methods	Limitation
5th (2023–Present)	Transformer + Attention	Retinexformer	Heavy GPU; slow inference
5th (2023–Present)	Generative Diffusion	Diff-LL, PyDiff	Iterative sampling; not real-time

These architectures define the quality ceiling of the field but are currently incompatible with the lightweight, zero-reference deployment goals that motivate the present survey.

4. Mathematical Formulation of Non-Reference Losses

The Zero-DCE network is trained without any reference images, guided exclusively by four complementary non-reference loss functions. Each loss term encodes a perceptually motivated image quality criterion. Let I denote the input low-light image and Y = LE^(8) denote the final enhanced output, both with pixel values normalised to [0, 1].

4.1. Spatial Consistency Constraints (L_spa)

The spatial consistency loss preserves the local structural relationships present in the original image. It penalizes differences in four-directional neighborhood contrasts between the enhanced and input images, ensuring that regions that were locally similar remain similar after enhancement:

$$L_{spa} = (1/K) \cdot \Sigma_{j=1}^{K} \Sigma_{q \in \Omega(j)} (|Y_j - Y_q| - |I_j - I_q|)^2 \quad (6)$$

where K is the total number of pixels, Ω(j) denotes the set of four directional neighbours {up, down, left, right} of pixel j, and |·| is the absolute difference of local region means computed over 4×4 non-overlapping patches. This loss discourages the network from creating spurious texture or altering spatial contrast relationships. Loss weight: λ_spa = 1.0.

4.2. Regional Exposure Regulation (L_exp)

The exposure control loss drives the mean luminance of local image regions toward a well-exposed target level, preventing both under-exposure (dark, muddy output) and over-exposure (washed-out highlights). It operates on non-overlapping 16×16 pixel patches:

$$L_{exp} = (1/M) \cdot \Sigma_{k=1}^{M} | mean(Y_k) - E | \quad (7)$$

where M is the number of non-overlapping 16×16 patches, Y_k is the k-th patch of the enhanced image, and E = 0.6 is the target well-exposed mean luminance, empirically chosen to balance exposure under natural lighting conditions. This loss is the primary driver of illumination improvement.

4.3. Color Constancy Metric Balance (L_col)

The color constancy loss enforces perceptual colour balance by constraining the pairwise relationships between the mean values of each RGB channel in the enhanced image. It prevents systematic colour casts introduced by unequal enhancement across spectral channels:

$$L\_col = \Sigma\_\{(p,q) \in \varepsilon\} \; (mean(Y^p) - mean(Y^q))^2 \qquad (8)$$
$$\varepsilon = \{ (R,G), (R,B), (G,B) \}$$

Where Y^p is the p-th color channel of the enhanced image and ε enumerates all three pairs of RGB channels. The loss is minimized when all three channel means are equal corresponding to an achromatic (grey) reference illumination. In practice, this loss suppresses systematic color cast artifacts that arise during aggressive illumination amplification. Loss weight: λ_col = 5.0 (increased to 50.0 for images with strong pre-existing color imbalance).

4.4. Illumination Total Variation Smoothness (L_smo)

The illumination smoothness loss promotes spatially coherent curve estimation by penalizing high total variation (TV) in the predicted curve parameter maps α^(n) across all eight iterations. Without this regularization, the curve maps exhibit high-frequency spatial variation producing patchy, unnatural illumination gradients:

$$L\_smo = (1/N) \cdot \Sigma\_\{n=1\}^\{N\} \; (\|\nabla\_x \alpha^(n)\|^2 + \|\nabla\_y \alpha^(n)\|^2) \qquad (9)$$

where ∇ _x α^(n) and ∇ _y α^(n) are the horizontal and vertical spatial gradients of the n-th curve parameter map, and N = 8 is the total number of enhancement iterations. The total variation penalty encourages piecewise-smooth curve parameter maps that produce gradual, artifact-free illumination transitions.

4.5. Total Training Objective

The complete training objective combines all four non-reference loss terms:

$$L\_total = \lambda\_spa \cdot L\_spa + \lambda\_exp \cdot L\_exp + \lambda\_col \cdot L\_col + \lambda\_smo \cdot L\_smo$$
$$= 1.0 \cdot L\_spa + 10.0 \cdot L\_exp + 5.0 \cdot L\_col + 200.0 \cdot L\_smo \qquad (10)$$

The relative weighting applied to these components manages trade-offs between aggressive structural exposure corrections and smooth gradient mappings. The total objective is optimized using back propagation over deep feature extractors to refine parameter estimation pipelines iteratively.All loss terms are jointly optimized using the Adam optimizer [17], which adapts per-parameter learning rates based on first and second moment estimates of the gradients

5. Image Quality Assessment and Visibility Classification

5.1. Full-Reference Perception Metrics (PSNR, SSIM)

Full-reference image quality assessment metrics compare an enhanced image Y against a ground-truth reference Y_ref. Two standard metrics are employed:

5.1.1 Peak Signal-to-Noise Ratio (PSNR)

PSNR measures the ratio of maximum signal power to distortion power in the logarithmic (decibel) domain:

$$PSNR = 10 \cdot \log_{10} (MAX\_I^2 / MSE)$$

$$MSE = (1/mn) \cdot \Sigma_{i=1}^{m} \Sigma_{j=1}^{n} [Y(i,j) - Y_{ref}(i,j)]^2 \quad (11)$$

Where MAX_I = 255 for 8-bit images. PSNR operates on a logarithmic scale; values above 30 dB indicate good fidelity, while values above 40 dB approach imperceptible distortion. A key limitation is that PSNR is computed pixel-wise and does not model the human visual system's sensitivity to local structural content.

5.1.2. Structural Similarity Index (SSIM)

SSIM (Wang et al., 2004)[20] evaluates perceptual similarity along three complementary dimensions — luminance, contrast, and structural correlation — within local sliding windows:

$$SSIM(x,y) = [(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)] / [(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)] \quad (12)$$

where μ_x, μ_y are local means, σ_x^2, σ_y^2 are local variances, σ_{xy} is the local cross-covariance, and $C_1 = (0.01 \times 255)^2$, $C_2 = (0.03 \times 255)^2$ are numerical stability constants. SSIM values range from -1 to 1; values above 0.9 indicate near-perfect perceptual similarity. The mean SSIM over all local windows (MSSIM) is used as the aggregate metric.

5.2. No-Reference Perceptual Metric Evaluators (NIQE, PI)[21]

No-reference (NR) metrics assess image quality without requiring a ground-truth reference, enabling evaluation on the ExDark dataset where paired references are unavailable. The Natural Image Quality Evaluator (NIQE, Mittal et al., 2013) models the statistical regularities of natural images using a multivariate Gaussian (MVG) fitted to normalized DCT coefficients extracted from a pristine image corpus. Quality is measured as the Mahalanobis distance between the MVG parameters of the test image and the pristine corpus model — lower NIQE scores indicate greater adherence to natural image statistics. The Perceptual Index (PI) combines NIQE with BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) as a composite no-reference quality measure. On ExDark test images, zero-reference methods such as Zero-DCE have demonstrated competitive NIQE scores in the range of 4.8–5.2 across diverse illumination conditions including candlelight and outdoor night scenes, indicating stronger adherence to natural image statistics compared to traditional histogram-based baselines [1][21].

5.3. Integrated CNN-Based Visibility Categorization Frameworks

A prominent extension in recent framework analyses involves pairing enhancement pipelines with automated visibility classifiers. Instead of relying solely on scalar metrics, systems use deep convolutional features to categorize enhanced or input items into discrete visibility ranges (e.g., Low, Medium, High). Typical labeling protocols assign boundaries based on mean grayscale brightness B(I) across threshold boundaries:

$$Label(I) = \{ Low \text{ if } B(I) < 0.25 \mid Medium \text{ if } 0.25 \leq B(I) \leq 0.50 \mid High \text{ if } B(I) > 0.50 \} \quad (13)$$

The CNN classifier employs three convolutional blocks (Conv + BatchNorm + ReLU + Pooling) paired with fully connected projection layers to output class distributions via Softmax activations. This automated feedback provides structural visibility profiling without matching reference domains

6. Quantitative Performance synthesis and Benchmarks

6.1. Evaluation Environments and Dataset Classifications

6.1.1 LOL Dataset (Paired, Indoor)

The LOL (Low-Light) dataset, introduced by Wei et al. (2018)[8] alongside Retinex-Net, is the primary benchmark for this survey. It comprises 500 image pairs collected in real indoor environments (485 training, 15 testing), each consisting of a low-light input and a normal-light reference from the same viewpoint, stored as 400×600 PNG images. The dataset spans diverse illumination conditions from moderately dim (mean brightness ~0.3) to extremely dark (mean brightness ~0.05). Its paired structure enables full-reference PSNR and SSIM evaluation.

6.1.2. ExDark Dataset (Unpaired, Real-World)

The ExDark dataset (Loh and Chan, 2019)[22]provides 7,363 low-light images across ten distinct illumination conditions (candlelight, indoor lamp, moonlight, etc.) without paired references, annotated with object-level bounding boxes. ExDark is used for qualitative evaluation and no-reference NIQE assessment, enabling assessment of generalization to diverse out-of-distribution real-world illumination conditions. The absence of references precludes PSNR/SSIM computation but provides an unbiased test of enhancement naturalness.

6.1.3 VE-LOL Dataset (Paired, Synthetic + Real)

VE-LOL (Vision-Enhancement LOL)[25]extends the original LOL benchmark with both synthetic and real-captured low/normal-light pairs, split into VE-LOL-L (low-level enhancement) and VE-LOL-H (high-level face detection task). It comprises roughly 2,500 paired images, substantially larger than the original 500-pair LOL set, enabling more robust generalization testing. Its dual synthetic-real composition allows researchers to isolate the domain gap between synthetic darkening models and real sensor-captured low-light degradation — a distinction the original LOL dataset cannot provide.

6.1.4 LSRW Dataset (Paired, Real-World Outdoor)

The LSRW (Large Scale Real World) dataset addresses a key limitation of LOL: its indoor-only scene composition. LSRW provides paired low/normal-light images captured with both DSLR and smartphone cameras across outdoor and mixed environments, totaling several thousand image pairs. Its inclusion of smartphone-captured data is particularly relevant for evaluating zero-reference methods intended for mobile or edge deployment, where sensor noise characteristics differ substantially from DSLR-captured benchmarks.[26]

6.1.5 DICM Dataset (Unpaired, Real-World)

DICM (Dataset for Image Contrast Manipulation) is a widely used no-reference benchmark consisting of real-world low-light photographs spanning varied indoor and outdoor scenes. Unlike ExDark, DICM does not include object annotations, making it primarily suited for perceptual quality evaluation via no-reference metrics (NIQE, BRISQUE) rather than downstream task performance. It is frequently used alongside LIME and MSRCR in classical Retinex literature as a qualitative comparison benchmark.

6.1.6 MEF Dataset (Multi-Exposure Fusion Benchmark)

The MEF (Multi-Exposure Fusion) dataset provides sequences of multiple exposures of the same static scene, originally designed for exposure fusion algorithms rather than single-image enhancement. It has since been adopted as a secondary no-reference benchmark for LLIE methods, testing whether enhancement models can recover detail comparable to multi-shot fusion using only a single low-light input. Its scene diversity (indoor, outdoor, high dynamic range) makes it a useful stress test for generalization beyond the LOL training distribution.

6.1.7. Dataset Coverage in Surveyed Literature

Dataset	Type	Scene	Size	Reference Type	Used By (in this survey)
LOL	Paired	Indoor	500 pairs	Full	Retinex-Net, Zero-DCE, EnlightenGAN
ExDark	Unpaired	Indoor/Outdoor	7363 images	No	Zero-DCE, SCI
VELOL	Paired	Indoor	~2500 pairs	Full	Not evaluated in surveyed works
LSRW	Paired	Outdoor/Mixed	~5600pairs	Full	Not evaluated in surveyed works
DICM	Unpaired	Indoor/Outdoor	69 images	No	Classical Retinex methods
MEF	Unpaired	Mixed	17 sequence	NO	Exposure fusion baselines

Table 6.1: Benchmark dataset coverage (LOL, ExDark, VE-LOL, LSRW, DICM, MEF)

6.1.8. Dataset and Code Availability

For reproducibility, Table 6.4 summarizes access conditions for all benchmark datasets discussed in this survey, along with the licensing terms of the original Zero-DCE implementation. The LOL dataset used in this survey was accessed via its Kaggle mirror, which redistributes the original Wei et al. (2018) release[8]; no additional licensing terms beyond the original academic-use intent are imposed by the mirror.

Resource	Access	License/Terms
LOL	Kaggle mirror (kaggle.com/datasets/soumikrakshit/lol-dataset), originally from Wei et al. project page	Free for academic/research use; Kaggle mirror imposes no separate terms beyond the

Resource	Access	License/Terms
		original release's academic-use intent
ExDark	Public, GitHub (cs-chan/Exclusively-Dark-Image-Dataset)	BSD-3-Clause; free for research, commercial use requires contacting the authors
VE-LOL	Public, project page linked from original paper (Liu et al., IJCV 2021)	Free for academic/research use; no formal commercial license published
LSRW	Public, GitHub (JianghaiSCU/R2RNet), hosted via Baidu Netdisk	Free for academic/research use; no formal commercial license published
DICM	Public, distributed informally via LLIE benchmark repositories	No formal license; conventionally used for academic comparison only
MEF	Public, distributed via exposure-fusion benchmark repositories	No formal license; conventionally used for academic comparison only
Zero-DCE / Zero-DCE++ code	Public, GitHub (Li-Chongyi/Zero-DCE, Li-Chongyi/Zero-DCE_extension)	CC BY-NC 4.0 (Attribution-NonCommercial); academic/research use only, commercial use prohibited

Table 6.4: Dataset and code availability for reproducibility

A practical implication for any subsequent implementation work building on this survey: since the original Zero-DCE codebase is licensed CC BY-NC 4.0, any deployment-oriented extension (e.g., the web-deployment direction proposed in G2) intended for non-academic or commercial use would need either a re-implementation from the published architecture and loss equations, or explicit permission from the original authors.

6.2. Multi-Paradigm Structural Comparison

Method	Year	Supervision	Params	CPU?	Key Feature
GHE / CLAHE	1990–1994	None	—	Yes	Statistical contrast
MSRCR / LIME	1997–2017	None	—	Yes	Physics Retinex
LLNet	2017	Paired (synth)	~1M	No	Denoising autoencoder
Retinex-Net	2018	Paired	~420K	No	Decomp + Enhance
SID (Chen)	2018	Paired (RAW)	~7.8M	No	RAW U-Net
EnlightenGAN	2021	Unpaired	~8.6M	No	GAN global-local disc.
RUAS	2021	Zero-ref	~0.003M	Yes	NAS search

Method	Year	Supervision	Params	CPU?	Key Feature
SCI	2022	Zero-ref	~0.01M	Yes	Self-calibrated illum.
Zero-DCE	2020	Zero-ref	~0.079M	Yes	Curve estimation
Zero -DCE++	2021	Zero-ref	~10.6K	Yes	

Table 6.2: Multi-paradigm structural comparison across LLIE methods

6.3. Quantitative Result Synthesis on LOL Test Bed

Method	PSNR (dB)	SSIM	Ref Type	Params	GPU?	Speed
GHE	~14.0	0.45	None	—	No	< 1ms
CLAHE	15.47	0.582	None	—	No	< 5ms
MSRCR	~16.0	0.54	None	—	No	~50ms
LLNet (2017)	~19.0	0.63	Paired	~1M	Yes	~200ms
Retinex-Net (2018)	16.77	0.560	Paired	~420K	Yes	~120ms
SID / U-Net (2018)	~22.0	0.74	Paired	~7.8M	Yes	~80ms
EnlightenGAN (2021)	17.48	0.651	Unpaired	~8.6M	Yes	~150ms
Zero-DCE orig. (2020)	14.86	0.559	Zero-ref	~79K	No	~100ms

Table 6.3: Quantitative result synthesis on LOL test bed

7. Research Gap Analysis

The systematic literature review exposes several critical gaps that motivate the present work. These gaps span algorithmic capability, evaluation completeness, deployment accessibility, and integration of quality assessment with enhancement.

#	Research Gap	Evidence in Literature	Proposed Resolution
G 1	**Zero-DCE implementations lack integrated automated quality assessment**	Original Zero-DCE (Guo et al., 2020) and RUAS/SCI provide no post-enhancement quality labelling. Users have no feedback on enhancement success.	CNN-based 3-class Visibility Classifier (Low/Med/High) with confidence scores, integrated into the pipeline.
G 2	**No end-to-end pipeline from stratified data prep through interactive deployment**	Existing works publish model weights only. No framework covers brightness-stratified labelling → training → evaluation → web deployment as a unified system.	A unified end-to-end pipeline covering brightness-stratified data preparation, model training, standardized evaluation, and interactive web deployment remains absent in existing literature and represents a promising direction for future work.
G 3	**Zero-reference methods not demonstrated on commodity CPU without GPU**	All evaluated deep learning methods (Retinex-Net, EnlightenGAN, SID) require GPU. Zero-DCE papers do not report CPU-only deployment benchmarks.	Existing zero-reference literature does not report CPU-only deployment benchmarks. Systematic profiling on commodity processors (e.g., mid-range Intel CPUs without CUDA) is an open evaluation gap.
G 4	**Fixed brightness thresholds for visibility labelling ignore sensor/scene diversity**	The LOL dataset spans a limited subset of real-world illumination diversity (indoor only). Fixed thresholds (0.25, 0.50) may not generalise to other sensors or outdoor scenes.	Identified as a limitation. Adaptive perceptual threshold estimation proposed as future work.
G 5	**Evaluation limited to LOL; no multi-benchmark generalisation study**	While LSRW, VE-LOL, DICM, and MEF benchmarks are described in Section 6.1, they remain unevaluated in most zero-reference works including the original Zero-DCE, leaving cross-dataset generalization unverified.	Evaluation on ExDark (no-reference NIQE) included. Comprehensive multi-benchmark evaluation proposed as future work.
G 6	**No ablation study isolating individual loss**	Original Zero-DCE ablation is limited. No practical deployed	A systematic ablation isolating individual loss contributions

#	Research Gap	Evidence in Literature	Proposed Resolution
	function contributions in deployed systems	system provides live ablation feedback to end users.	(L_exp, L_col, L_smo, L_spa) in a deployed setting has not been reported in existing literature and would provide practical guidance for loss weight tuning..
G 7	**Lack of video enhancement for temporal applications (surveillance, driving)**	All surveyed LLIE methods operate on individual still images. Temporal consistency across frames critical for video surveillance and autonomous driving is unaddressed.	Temporal consistency loss for video enhancement identified as primary future direction.

Table 7.1: Comprehensive research gap analysis with evidence and proposed resolutions

A review of the current literature highlights a sharp dichotomy in the field of low-light image enhancement. On one side, 2026 SOTA architectures namely Retinexformer[23] and Diffusion-driven models[24] achieve unprecedented fidelity and peak PSNR marks. However, their architectural complexity demands heavy GPU resources, rendering them impractical for real-time applications, edge devices, or lightweight web interfaces hosted on standard consumer infrastructure.

On the other hand, traditional methods remain highly performant but fundamentally degrade image structure. Zero-DCE bridges this critical gap. It bypasses the need for paired data while offering an exceptionally lightweight footprint (~79,416 parameters). The research gap this project addresses is not the achievement of raw, computationally expensive peak PSNR, but rather the execution of a highly optimized, real-time, deployment-ready pipeline that balances unsupervised illumination curve estimation with automated visibility assessment for edge computing applications. The resolutions outlined above are proposed as directions for a subsequent implementation study and are not experimentally validated within the scope of this survey.

8. Open Research Challenges And Future Scope

8.1. Adaptive Perceptual Boundary Limitations

The current visibility labeling protocol assigns class boundaries using fixed brightness thresholds (B < 0.25 = Low, 0.25–0.50 = Medium, B > 0.50 = High). While straightforward and computationally inexpensive, fixed thresholds exhibit two fundamental limitations:

•Sensor dependency: Different camera sensors exhibit different noise profiles and dynamic range characteristics at identical luminance levels. A Smartphone sensor and a professional DSLR capture measurably different SNR at B = 0.20, yet the fixed threshold assigns them identical 'Low' labels.

•Scene-type variability: Indoor and outdoor scenes with similar mean brightness exhibit fundamentally different perceptual quality. Night sky photography at B = 0.20 may be perceptually satisfying, while indoor dark corridor footage at B = 0.20 requires substantial enhancement.

Future work should investigate adaptive threshold estimation via perceptual quality scoring (NIQE, BRISQUE) or scene-type clustering, and explore learning-to-label approaches that directly predict perceptual visibility from image features without relying on scalar brightness statistics.

8.2. Joint Enhancement-Denoising Trade-offs

The Zero-DCE spatial consistency loss (L_spa) discourages amplification of high-frequency content, providing implicit noise suppression. However, this suppression is non-adaptive: it cannot distinguish between textural details (to be preserved) and noise (to be suppressed). In images with significant sensor noise, the aggressive illumination amplification of L_exp can amplify noise faster than L_spa suppresses it, leading to visually noisy outputs.

Promising directions for joint enhancement-denoising include: (1) incorporating an explicit denoising sub-network trained with blind-spot convolutions to ensure noise estimates are based on surrounding context; (2) frequency-domain regularisation that applies stronger smoothing to high-frequency noise bands while protecting low-frequency illumination structure; and (3) knowledge distillation from paired enhancement-denoising models to guide the zero-reference student network.

8.3. Temporal Stream Consistency and Real-Time Edge Deployment

The current framework processes individual still images independently. Extension to video low-light enhancement requires temporal consistency the enhanced luminance and color of a given scene element should vary smoothly across frames, avoiding the flickering artifacts that arise from per-frame independent processing.

Incorporating a temporal consistency loss L_temp that penalizes abrupt frame-to-frame changes in the enhancement curve parameters α is a natural extension. Optical-flow-guided warping can align consecutive frames before computing L_temp, ensuring that the consistency constraint accounts for camera and scene motion. For mobile and edge deployment, ONNX quantization to INT8 precision and CoreML conversion for iOS devices represent practical pathways to sub-100ms inference on mobile CPUs.

8.4. Ethical Considerations and Misuse Risk

The surveillance and face recognition applications motivating much LLIE research (Section 1.1) raise ethical concerns that warrant explicit discussion. Enhanced low-light imagery can extend the operational range of facial recognition and identity tracking systems into previously private nighttime or indoor conditions, raising surveillance and consent concerns, particularly in jurisdictions lacking clear regulatory frameworks for night-vision-enhanced monitoring. Additionally, enhancement models trained predominantly on datasets such as LOL and ExDark, which exhibit limited demographic and geographic diversity, risk inheriting and amplifying biases present in downstream face recognition or object detection pipelines, potentially degrading enhancement quality or downstream task accuracy for underrepresented skin tones or scene types. Future deployment of zero-reference LLIE systems in security-critical contexts should be accompanied by bias auditing across demographic subgroups and clear use-case restrictions to mitigate misuse risk in unauthorized surveillance applications.

9. Conclusion

This literature survey has systematically reviewed the technical trajectory of Low-Light Image Enhancement (LLIE) frameworks, analyzing traditional statistical adjustments, complex physics-based Retinex optimization models, supervised architectures, and zero-reference frameworks. While data-driven supervised approaches yield strong full-reference benchmark results, their rigid dependence on static paired reference exposures hampers practical application in unpredictable, real-world deployment conditions.

Zero-Reference Deep Curve Estimation presents a robust pathway forward by discarding reference dataset dependencies entirely. By framing contrast matching as an iterative parameter estimation task governed by structural, color, and exposure non-reference objectives, these compact networks maintain high execution efficiency. Compiling this research domain reveals that resolving current structural limits requires bridging zero-reference pipelines with automated visibility metrics, adapting boundary models to handle sensor variation, and integrating temporal consistency features to enable artifact-free video processing on edge hardware.

10. References

[1] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Feng, and R. Chen, "Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," in Proc. IEEE/CVF CVPR, 2020, pp. 1780–1789.

[2] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 4th ed. New York, NY: Pearson, 2018.

[3] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," in Graphics Gems IV, Academic Press, 1994, pp. 474–485.

[4] E. H. Land, "The Retinex Theory of Color Vision," Scientific American, vol. 237, no. 6, pp. 108–128, 1977.

[5] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes," IEEE Trans. Image Process., vol. 6, no. 7, pp. 965–976, 1997.

[6] X. Guo, Y. Li, and H. Ling, "LIME: Low-Light Image Enhancement via Illumination Map Estimation," IEEE Trans. Image Process., vol. 26, no. 2, pp. 982–993, 2017.

[7] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A Deep Autoencoder Approach to Natural Low-Light Image Enhancement," Pattern Recognit., vol. 61, pp. 650–662, 2017.

[8] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex Decomposition for Low-Light Enhancement," in Proc. BMVC, 2018.

[9] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to See in the Dark," in Proc. IEEE/CVF CVPR, 2018, pp. 3291–3300.

[10] Y. Jiang et al., "EnlightenGAN: Deep Light Enhancement Without Paired Supervision," IEEE Trans. Image Process., vol. 30, pp. 2340–2349, 2021.

[11] W. Wu et al., "URetinex-Net: Retinex-Based Deep Unfolding Network for Low-Light Image Enhancement," in Proc. IEEE/CVF CVPR, 2022, pp. 5901–5910.

[12] L. Liu et al., "RUAS: Retinex-Inspired Unrolling with Architecture Search for Under-Exposed Image Enhancement," in Proc. IEEE/CVF CVPR, 2021, pp. 1419–1428.

[13] L. Ma et al., "Toward Fast, Flexible, and Robust Low-Light Image Enhancement," in Proc. IEEE/CVF CVPR, 2022, pp. 5637–5646.

[14] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep CNNs," in NeurIPS, 2012.

[15] I. Goodfellow et al., "Generative Adversarial Networks," in NeurIPS, 2014.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in MICCAI, 2015.

[17] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in ICLR, 2015.

[18] Z. Tian et al., "A Survey of Deep Learning-Based Low-Light Image Enhancement," Sensors, vol. 23, no. 18, pp. 7763, 2023.

[19] W. Zhang et al., "Unsupervised Low-Light Image Enhancement Based on Explicit Denoising and Knowledge Distillation," Computers, Materials & Continua, vol. 82, no. 2, pp. 2538–2555, 2025.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, 2004.

[21] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a Completely Blind Image Quality Analyzer," IEEE Signal Process. Lett., vol. 20, no. 3, pp. 209–212, 2013.

[22] C. Loh and Y. M. Chan, "Getting to Know Low-light Images with the Exclusively Dark Dataset," Computer Vision and Image Understanding, vol. 178, pp. 30–42, 2019.

[23] C. Hou, Z. Wang, and M. Zheng, "Retinexformer: One-Stage Retinex-Based Transformer for Low-Light Image Enhancement," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023.

[24] Y. Jiang, Y. Liu, and L. Level, "Conditional Diffusion Models for Generative Low-Light Image Restoration," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2025.

[25] J. Liu, D. Xu, W. Yang, M. Fan, and H. Huang, "Benchmarking Low-Light Image Enhancement and Beyond," International Journal of Computer Vision, vol. 129, pp. 1153–1184, 2021. (VE-LOL)

[26] J. Hai et al., "R2RNet: Low-light Image Enhancement via Real-low to Real-normal Network," Journal of Visual Communication and Image Representation, vol. 90, 2023. (LSRW)

[27] C. Li, C. Guo, and C. C. Loy, "Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 8, pp. 4225–4238, 2022.