https://doi.org/10.24867/GRID-2026-p63

Artificial Intelligence for Enhancement of Low-Light Photographs

ABSTRACT

Photographs captured under low-light conditions are often characterized by reduced contrast, pronounced noise, loss of detail, and color reproduction deviations, which can significantly reduce their visual quality. Traditional image enhancement methods can partially alleviate these problems, but their effectiveness often depends on parameter selection and the characteristics of a particular scene. The development of artificial intelligence and deep learning has enabled the use of more advanced models in which image enhancement is learned directly from data. This paper reviews contemporary methods for enhancing photographs captured under low-light conditions, with particular emphasis on CNN- and Retinex-based models, GAN and zero-reference approaches, Transformer architectures, and diffusion models. Their basic principles, training strategies, datasets, and quality assessment metrics are analyzed. A comparative review of selected studies shows a shift toward models that improve perceptual quality and generalization and reduce dependence on paired training data, while computational efficiency remains an important challenge.

Iva Juretić
Ivana Jurič
Saša Petrović
Sanja Cvetojević

University of Novi Sad, Faculty of Technical Sciences, Department of Graphic Engineering and Design, Novi Sad, Serbia

KEY WORDS
Artificial intelligence, low-light image enhancement, deep learning, Retinex, Transformer, diffusion models

1. Introduction

Low-light photography remains a challenging task because only a limited amount of light reaches the camera sensor during image acquisition. As a result, low-light photographs are often affected by reduced contrast, pronounced noise, color distortion, and loss of visible detail (Guo et al., 2023). These problems become particularly severe under extremely low illumination, where the small number of captured photons leads to a low signal-to-noise ratio. Extending the exposure time can increase the amount of captured light, but it also increases the risk of motion blur caused by camera shake or movement within the scene (Chen et al., 2018).

A wide range of methods has been developed to address these degradations, from gray-level transformations, histogram equalization, and Retinex-based techniques to more recent deep learning approaches. Traditional enhancement methods can improve image brightness and contrast, but they may also amplify noise, suppress fine details, introduce halo artifacts, or produce color distortions. Their performance may also depend strongly on parameter selection and the characteristics of the input image (Zhan, Goh & Sunar, 2024). With the development of neural networks, low-light image enhancement has increasingly shifted toward data-driven methods that learn image representations directly from training data and are able to address several degradation types within a single framework.

Recent studies cover a broad range of architectures and learning strategies, including convolutional neural networks, Retinex-based deep models, generative adversarial networks, zero-reference methods, Transformer-based architectures, and diffusion models (Liu & Fan, 2025). These approaches differ not only in their enhancement mechanisms, but also in their dependence on paired training data, generalization ability, computational requirements, and evaluation strategies. In mobile photography, low-light enhancement can also be based on multi-frame or burst processing, where several consecutive images are combined to obtain a higher-quality result.

Google Night Sight is one example of this approach, using multiple captured frames for photography under very low-light conditions (Liba et al., 2019). These multi-frame methods are not considered in detail in this review, which focuses on single-image enhancement.

This paper reviews and compares recent artificial intelligence-based methods for low-light photograph enhancement, with particular emphasis on their underlying principles, datasets, evaluation procedures, advantages, and limitations. The review also considers commonly used datasets and image quality assessment metrics, while selected studies are compared to highlight the main directions in the development of this research field.

2. Characteristics and degradations of low-light photographs

Photographs captured under low-light conditions are commonly affected by reduced contrast, color reproduction errors, and image noise, all of which negatively influence their visual quality (Guo et al., 2023). Unfavorable imaging conditions may include insufficient or uneven illumination, backlighting, and other lighting situations that reduce image quality and lead to loss of visual information (Tian et al., 2023).

One of the main challenges of photography under extremely low illumination is a low signal-to-noise ratio, which is associated with the small number of photons captured by the sensor. Images acquired with short exposure times may contain substantial noise, whereas longer exposures increase the risk of motion blur caused by camera shake or movement of objects within the scene (Chen et al., 2018).

Increasing the ISO sensitivity can make an image appear brighter, but it also amplifies the noise already present in the captured signal. Subsequent image processing, such as pixel-value scaling or histogram stretching, cannot compensate for the inherently low signal-to-noise ratio (Chen et al., 2018).

Severe color distortions may also occur under extremely low-light conditions. Chen et al. (2018) describe an example in which, despite the use of a very high ISO setting, the captured image remained dark and exhibited pronounced noise and noticeable color shifts. Zheng et al. (2026) further note that low-light photographs may exhibit low brightness, poor scene visibility, high noise levels, and loss of detail. These degradations can become even more pronounced in non-uniformly illuminated scenes, where illumination varies spatially and very dark and very bright regions may occur simultaneously within the same image.

3. Low-light image enhancement methods

A variety of approaches have been developed for low-light image enhancement, ranging from traditional image processing techniques to modern methods based on artificial intelligence and deep learning. The development of this field reflects a gradual shift from predefined algorithmic procedures toward data-driven approaches (Zhan, Goh & Sunar, 2024).

3.1. Traditional enhancement methods

Traditional methods for enhancing low-light photographs include a range of mathematical and algorithmic techniques, among which gray-level transformations, histogram equalization, and Retinex-based methods are commonly used (Zhan, Goh & Sunar, 2024). These approaches can also be classified according to whether they operate in the spatial or frequency domain. Spatial-domain methods directly modify pixel values, whereas frequency-domain methods first transform the image and then apply appropriate filtering operations in the transformed domain (Liu & Fan, 2025).

Gray-level transformations are used to adjust image brightness, contrast, and dynamic range by modifying individual pixel values. This group includes linear, logarithmic, piecewise linear, and gamma transformations (Zhan, Goh & Sunar, 2024).

However, these methods may not adequately handle a wide range of illumination levels, which can result in detail loss in overexposed or underexposed regions (Zhan, Goh & Sunar, 2024).

Histogram equalization modifies the distribution of gray levels in order to increase image contrast and expand the dynamic range. Global histogram equalization is applied to the entire image, whereas adaptive histogram equalization divides the image into smaller regions that are processed independently (Zhan, Goh & Sunar, 2024). A limitation of adaptive histogram equalization is that it may excessively increase local contrast and amplify noise. Contrast Limited Adaptive Histogram Equalization (CLAHE) addresses this problem by restricting the amount of contrast enhancement, thereby reducing excessive noise amplification (Zhan, Goh & Sunar, 2024).

Retinex-based approaches rely on decomposing an image into illumination and reflectance components. These methods can improve contrast, brightness, and dynamic range while helping preserve color appearance. The development of Retinex-based enhancement has led to methods such as Single-Scale Retinex (SSR), Multi-Scale Retinex (MSR), and Multi-Scale Retinex with Color Restoration (MSRCR) (Zhan, Goh & Sunar, 2024).

Nevertheless, Retinex methods may introduce halo artifacts and color distortions, while parameter selection can be challenging because a single parameter configuration does not necessarily produce satisfactory results across different images (Zhan, Goh & Sunar, 2024).

A representative example of a traditional Retinex-based approach is LIME (Low-light Image Enhancement via Illumination Map Estimation), proposed Guo, Li and Ling (2017). Unlike methods that separately estimate both reflectance and illumination, LIME estimates only the illumination map, thereby reducing the complexity of the problem. An initial illumination map is obtained from the maximum value among the RGB channels at each pixel and is subsequently refined using a structure prior to preserve image structure.

The authors experimentally compared LIME with several traditional enhancement techniques, including HE, AHE, gamma correction, and other Retinex-based and variational methods. On the HDR dataset, LIME achieved the lowest average LOE value among the compared methods, indicating better preservation of the natural lightness order in the enhanced images (Guo, Li & Ling, 2017).

Frequency-domain methods, such as homomorphic filtering and wavelet transforms, process different frequency components of an image and may preserve image details more effectively than spatial-domain approaches. However, they are generally more sensitive to parameter selection, while both spatial- and frequency-domain techniques may amplify noise and show limited effectiveness under challenging low-light conditions (Liu & Fan, 2025).

Because traditional methods have a limited ability to model complex patterns and relationships within image data and to adapt to varying image characteristics, increasing attention has been directed toward data-driven enhancement approaches (Zhan, Goh & Sunar, 2024).

3.2. Artificial intelligence-based methods

Traditional low-light image enhancement methods often rely on physical models, predefined assumptions, and manual parameter adjustment, which can limit their adaptability to different imaging conditions (Ye, Qiu & Zhang, 2024). Learning-based methods instead use data to learn how low-light images should be enhanced.

Recent methods use different learning strategies, including supervised, unsupervised, zero-shot, and fusion-based learning (Liu & Fan, 2025). The main difference between these approaches lies in the type of training data and supervision they use. The terms zero-reference and zero-shot are not used consistently across the literature.

In this paper, zero-reference refers to methods such as Zero-DCE, which are trained without paired or unpaired reference images and use non-reference loss functions during training (Guo et al., 2020). The main distinction is that zero-reference methods may still involve model training, but without reference images, whereas zero-shot methods do not rely on prior training on an external dataset. ExCNet, for example, trains a small CNN at test time using only the input image (L. Zhang et al., 2019).

The following section presents representative CNN, Retinex-based, GAN, unsupervised and zero-reference, Transformer, and diffusion-based methods.

4. Modern AI-based approaches to low-light image enhancement

Modern artificial intelligence-based approaches to low-light image enhancement employ a variety of architectures and learning strategies. These include convolutional neural networks, Retinex-based approaches, generative adversarial networks, unsupervised and zero-reference methods, Transformer architectures, and more recent diffusion models (Liu & Fan, 2025).

4.1. Early deep learning approaches and CNN architectures

One of the early deep learning approaches to low-light image enhancement was LLNet, introduced by Lore, Akintayo and Sarkar (2017). The method is based on a deep autoencoder designed to improve contrast and reduce noise within a single framework. During training, images were artificially darkened and corrupted with Gaussian noise in order to simulate different low-light conditions.

The authors also proposed S-LLNet, a variant in which contrast enhancement and denoising are handled by two successive modules. Both models were tested on synthetically degraded images and naturally captured low-light photographs using PSNR and SSIM. On dark images with added synthetic noise, LLNet and S-LLNet performed better than the traditional methods used for comparison, with S-LLNet achieving the highest PSNR and SSIM values in several cases (Lore, Akintayo & Sarkar, 2017).

A different CNN-based strategy was introduced with MBLLEN (Multi-Branch Low-Light Enhancement Network) by Lv et al. (2018). The network follows a fully convolutional design and includes modules for feature extraction, enhancement, and fusion. Features obtained at different depths of the network are processed separately and then combined to produce the final enhanced image. This multi-branch structure allows the method to address several common low-light degradations, including insufficient brightness, reduced contrast, noise, and artifacts.

The model was trained on a synthetic dataset generated from VOC images and evaluated using several measures, including PSNR and SSIM. On synthetic low-light images without additional noise, MBLLEN reached a PSNR of 26.56 dB and an SSIM of 0.89, outperforming all methods included in the comparison (Lv et al., 2018). Similar performance gains were reported for images affected by Poisson noise, and the method was also tested on real low-light photographs.

Ren et al. (2019) approached the problem using a Deep Hybrid Network with two complementary processing streams. An encoder–decoder branch reconstructs the global scene content, while a second branch focuses on edges and structural information through a spatially varying recurrent neural network. Along with the MSE loss, perceptual and adversarial losses were used during training to improve the visual quality of the output. The contribution of the edge-processing branch was confirmed experimentally: the average PSNR and SSIM increased from 27.67 dB and 0.9156 without this branch to 28.43 dB and 0.9662 when it was included (Ren et al., 2019).

MIRNet, developed by Zamir et al. (2020), is a general image restoration and enhancement architecture rather than a model designed exclusively for low-light enhancement. Its multi-scale design uses parallel streams at different spatial resolutions, allowing information to be exchanged across scales while preserving high-resolution details. The architecture also incorporates spatial and channel attention mechanisms. For image enhancement, MIRNet was evaluated on the LOL and MIT-Adobe FiveK datasets. It achieved a PSNR of 24.14 dB and an SSIM of 0.83 on LOL, and a PSNR of 23.73 dB and an SSIM of 0.925 on MIT-Adobe FiveK, outperforming the methods included in the corresponding comparisons (Zamir et al., 2020).

4.2. Deep learning-based Retinex approaches

The development of deep learning has made it possible to combine the basic Retinex principle, in which an image is decomposed into reflectance and illumination components, with models that learn the required representations and constraints directly from data. This reduces the dependence on hand-crafted priors and careful parameter tuning that is typical of traditional Retinex methods (Wei et al., 2018). One of the early representative methods in this group is Retinex-Net, proposed by Wei et al. (2018). The model consists of two subnetworks: DecomNet, which separates the input image into reflectance and illumination components, and Enhance-Net, which adjusts the illumination component. The enhancement process includes decomposition, illumination adjustment, and reconstruction, while additional denoising is applied to the reflectance component. For training, the authors introduced the LOL (LOw-Light) dataset containing 500 pairs of images of the same scenes captured under low- and normal-light conditions. Of these, 485 pairs were used for training and 15 for evaluation. Retinex-Net was also tested on the LIME, MEF, and DICM datasets and compared with traditional methods including LIME, SRIE, NPE, and DeHz. According to the authors, the method was able to brighten dark regions without causing severe overexposure and preserved image details more effectively when enhancement and denoising were performed jointly (Wei et al., 2018).

A further development of this idea is KinD (Kindling the Darkness), introduced by Y. Zhang, J. Zhang and Guo (2019). The method also follows the Retinex formulation by separating an image into illumination and reflectance components. The illumination component is used for brightness adjustment, while the reflectance component is processed to remove degradations such as noise and color distortion. The network is trained using pairs of images captured at different exposure levels and does not require ground-truth reflectance or illumination maps (Y. Zhang, J. Zhang & Guo, 2019). A total of 450 pairs from the LOL dataset were used for training, while evaluation was carried out on the LOL, LIME, NPE, and MEF datasets using PSNR, SSIM, LOE, and NIQE. On the LOL dataset, KinD achieved a PSNR of 20.87 dB and an SSIM of 0.802, outperforming the other methods included in the comparison. It also obtained the lowest NIQE value among the compared methods on the same dataset (Y. Zhang, J. Zhang & Guo, 2019).

URetinex-Net, proposed by Wu et al. (2022), introduces a different strategy based on deep unfolding. Instead of treating Retinex decomposition as a fixed optimization procedure, the individual optimization steps are unfolded into learnable network modules that infer implicit priors from data. The architecture includes modules for initialization, iterative optimization, and illumination adjustment, with the aim of suppressing noise while preserving details in both reflectance and illumination components. The model was trained and tested on the LOL dataset, while its generalization ability was further examined on the SICE dataset without retraining or additional adaptation. On LOL, URetinex-Net achieved a PSNR of 21.33 dB and an SSIM of 0.835 and outperformed the compared methods in terms of PSNR, SSIM, and LPIPS. It also showed strong performance on the SICE dataset across most of the evaluated metrics (Wu et al., 2022).

4.3. GAN, unsupervised, and zero-reference approaches

The need for large numbers of paired low-light and normal-light images is one of the main limitations of supervised methods. This has motivated the development of approaches that can be trained without paired reference images, or even without corresponding normal-light examples.

One representative unsupervised method is EnlightenGAN, proposed by Jiang et al. (2021). The model is based on a generative adversarial network and is trained using unpaired low-light and normal-light images. Its architecture includes both global and local discriminators, together with an attention mechanism guided by illumination information from the input image.

This allows different regions of the image to be enhanced according to their local lighting conditions. The training set contained 914 low-light and 1,016 normal-light images collected from several datasets, without one-to-one pairing between them. EnlightenGAN was evaluated on the MEF, NPE, LIME, VV, and DICM datasets, where it achieved the best average NIQE score among the compared methods. In a subjective evaluation involving 23 images, it also obtained the best average rank of 1.78 (Jiang et al., 2021).

A different strategy is used by Zero-DCE (Zero-Reference Deep Curve Estimation), proposed by Guo et al. (2020). The method formulates low-light enhancement as an image-specific curve estimation problem. Its DCE-Net predicts high-order curves that adjust pixel values and the dynamic range of the input image. Zero-DCE does not require paired or unpaired reference images during training. Instead, the network is optimized using non-reference loss functions related to spatial consistency, exposure control, color constancy, and illumination smoothness (Guo et al., 2020).

A later extended version introduced Zero-DCE++, a lightweight model with approximately 10,000 parameters and substantially lower computational requirements (Li, Guo & Loy, 2022). In a subjective evaluation conducted on 199 images from five datasets, Zero-DCE received the highest average score among the compared methods, while Zero-DCE++ achieved the second-best result (Li, Guo & Loy, 2022).

SCI (Self-Calibrated Illumination), developed by Ma et al. (2022), is another lightweight unsupervised approach. The method uses a cascaded illumination learning process with shared weights and a self-calibration module. This additional calibration process is required during training, whereas inference relies only on the basic network block, which keeps the computational cost low. An unsupervised loss is used to improve the model's ability to adapt to different scenes. SCI was evaluated on the MIT and LSRW datasets and compared with traditional, supervised, and unsupervised methods.

On the MIT dataset, it achieved a PSNR of 20.45 dB and an SSIM of 0.893, while the experiments also showed that the model requires relatively low computational resources. The authors further examined its use as a preprocessing method for low-light face detection and nighttime semantic segmentation (Ma et al., 2022).

4.4.Transformer-based approaches

Transformer architectures can model relationships between spatially distant regions of an image. This is particularly useful under very low-light conditions, where local information may not be sufficient for reliable reconstruction. Recent methods therefore combine attention mechanisms with information about illumination, noise, or local image characteristics.

Xu et al. (2022) proposed SNR-Aware Low-light Image Enhancement, which combines Transformer and convolutional operations. Their contribution is adjusted according to the signal-to-noise ratio in different image regions. Transformer-based long-range processing is used more strongly in areas with very low SNR, while local convolutional operations are emphasized in regions with higher SNR.

An SNR map is also used to guide the attention mechanism and feature fusion. The model was evaluated on seven datasets, including LOL-v1, LOL-v2, SID, SMID, and SDSD. On LOL-v1, it achieved a PSNR of 24.61 dB and an SSIM of 0.842, while on LOL-v2-real it reached a PSNR of 21.48 dB and an SSIM of 0.849, outperforming the methods included in the comparison (Xu et al., 2022).

Wang et al. (2023) developed LLFormer, a Transformer model intended for the enhancement of ultra-high-resolution images. The authors also introduced the UHD-LOL dataset with 4K and 8K images in order to test how well low-light enhancement methods perform at substantially higher resolutions than those typically used in existing datasets. LLFormer uses axis-based multi-head self-attention and cross-layer feature fusion to reduce the computational cost of processing large images. The model achieved the best results among the compared methods on UHD-LOL4K and UHD-LOL8K. Additional evaluation on LOL and MIT-Adobe FiveK showed that LLFormer also performs competitively on standard datasets. On MIT-Adobe FiveK, it achieved a PSNR of 25.75 dB and an SSIM of 0.923 (Wang et al., 2023).

Cai et al. (2023) proposed Retinexformer, which combines the Retinex principle with a Transformer architecture in a one-stage framework. The model first estimates illumination information to improve image visibility and then restores degradations that are hidden in dark regions or introduced during the light-up process. Its Illumination-Guided Transformer uses illumination information to guide the modeling of non-local interactions between regions with different lighting conditions. Retinexformer was evaluated on thirteen datasets, including LOL, SID, SMID, SDSD, and MIT-Adobe FiveK. On LOL-v1, it achieved a PSNR of 25.16 dB and an SSIM of 0.845, while on MIT-Adobe FiveK it reached a PSNR of 24.94 dB, outperforming the methods included in the corresponding comparisons (Cai et al., 2023).

4.5. Diffusion models

Diffusion models represent a more recent direction in low-light image enhancement and use a generative process to reconstruct higher-quality images. Although they can produce strong perceptual results, they often require substantial computational resources and relatively long inference times. For this reason, recent work has also focused on improving their efficiency.

Jiang et al. (2023) proposed DiffLL, which combines a diffusion model with wavelet transformation. The input image is first transformed into the wavelet domain, where components representing global information and high-frequency image details are separated.

The diffusion process is then applied to a component with reduced spatial resolution, while high-frequency details are reconstructed using a dedicated High-Frequency Restoration Module (HFRM). This reduces the computational cost of the diffusion process while preserving image details.

DiffLL was evaluated on the LOL-v1, LOL-v2-real, and LSRW datasets, as well as on several real-world datasets without reference images. On LOL-v1, it achieved a PSNR of 26.34 dB and an SSIM of 0.845, while on LOL-v2-real it reached a PSNR of 28.86 dB and an SSIM of 0.876, outperforming the methods included in the comparison. The authors also report that the proposed method is at least 70 times faster than the previously compared diffusion-based image restoration approaches (Jiang et al., 2023).

A more recent method is Diff-Retinex++, introduced by Yi et al. (2025). The model combines the generative capability of diffusion models with physically motivated Retinex principles. Its architecture includes a Denoising Diffusion Model (DDM) and a Retinex-Driven Mixture of Experts (RMoE), while Retinex information is incorporated through a dedicated attention module that guides the enhancement process.

The Mixture-of-Experts design allows the model to activate different components depending on the characteristics of the input image. Diff-Retinex++ was evaluated on LOL-v1, LOL-v2-real, LOL-v2-synthetic, and LSRW, together with additional real-world scenes captured under different illumination levels. On LOL-v1, it achieved the best results among the compared methods in terms of PSNR, SSIM, FID, and MAE, while it also obtained the best overall performance on LOL-v2-real.

Experiments on extremely dark real-world scenes further showed that the model was able to preserve image details, color reproduction, and structural content (Yi et al., 2025).

5. Datasets and image quality assessment metrics

5.1. Datasets

Evaluation of low-light image enhancement methods relies on a range of datasets that include paired low-light and well-exposed images, as well as real-world datasets without reference images.

One of the most widely used datasets is LOL (LOw-Light), which contains pairs of the same scenes captured under low- and normal-light conditions. The original LOL dataset consists of 485 image pairs for training and 15 pairs for testing (Wei et al., 2018). Its later version, LOL-v2, includes real and synthetic subsets (Yang et al., 2021).

For extremely low-light conditions, the See-in-the-Dark (SID) dataset introduced by Chen et al. (2018) is particularly relevant.

SID contains 5,094 short-exposure RAW images with corresponding long-exposure reference images captured using Sony α7S II and Fujifilm X-T2 cameras.

The dataset includes both indoor and outdoor scenes recorded under very low illumination, with the reference images acquired using exposure times 100 to 300 times longer than those of the corresponding input images.

MIT-Adobe FiveK is also used in several studies considered in this review, although it was not originally developed specifically as a low-light image dataset. It is a broader photographic image enhancement dataset containing 5,000 photographs covering a wide range of scenes and lighting conditions.

Each image was independently retouched by five trained photographers (Bychkovsky et al., 2011). In MIRNet, the images produced by Expert C were used as reference images, with 4,500 images used for training and 500 for testing (Zamir et al., 2020).

To evaluate low-light enhancement methods on ultra-high-resolution images, Wang et al. (2023) introduced UHD-LOL, which consists of the UHD-LOL4K and UHD-LOL8K subsets. UHD-LOL4K contains 8,099 pairs of 4K images, while UHD-LOL8K contains 2,966 pairs of 8K images.

These datasets make it possible to assess how well enhancement models handle image resolutions substantially higher than those used in earlier low-light datasets.

The main characteristics of the datasets discussed above are summarized in Table 1.

Table 1
Summary of datasets used in the reviewed studies

Dataset	Size	Resolution	Data type	Pairing	Image format / domain
LOL-v1 (Wei et al., 2018)	500 low-/normal-light image pairs	400 × 600	Real scenes	Paired	PNG (converted from RAW)
LOL-v2-real (Yang et al., 2021)	789 pairs (689 training, 100 testing)	Not specified	Real	Paired	Not specified
LOL-v2-synthetic (Yang et al., 2021)	1,000 samples (900 training, 100 testing)	Not specified	Synthetic	Not explicitly stated	Not specified
SID (Chen et al., 2018)	5,094 short-exposure images; 424 long-exposure references	4240 × 2832 (Sony); 6000 × 4000 (Fujifilm)	Real	Paired	RAW
MIT-Adobe FiveK (Bychkovsky et al., 2011)	5,000 photographs; five retouched versions per image	Not specified	Real photographs	Input-output pairs	RAW input; retouched output versions
UHD-LOL4K (Wang et al., 2023)	8,099 low-light/normal-light image pairs	4K	Synthetic	Paired	Not specified
UHD-LOL8K (Wang et al., 2023)	2,966 low-light/normal-light image pairs	8K	Synthetic	Paired	Not specified

Note: Only characteristics explicitly reported in the cited studies are included. "Not specified" means that the cited source used for the table does not state that characteristic; no value was inferred.

5.2. Image quality assessment metrics

PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure) are among the most commonly used metrics for quantitative evaluation of enhanced images. PSNR is based on the difference between pixel values in the reference and processed images, with higher values indicating lower reconstruction error. Wang et al. (2004) noted that measures based only on pixel-wise error do not necessarily correspond well to human perception of image quality, which motivated the development of SSIM as a structure-based measure. SSIM compares luminance, contrast, and structural information between two images, with higher values indicating greater similarity to the reference image (Wang et al., 2004).

Perceptual similarity is also evaluated using metrics such as LPIPS (Learned Perceptual Image Patch Similarity). LPIPS measures the difference between images using features extracted from deep neural networks.

R. Zhang et al. (2018) showed that deep features can correspond more closely to human judgments of perceptual similarity than traditional measures such as PSNR and SSIM. For LPIPS, lower values indicate a smaller perceptual distance between the compared images.

When a suitable reference image is not available, NIQE (Natural Image Quality Evaluator) can be used. NIQE does not require a reference image or prior training on subjectively rated distorted images. Instead, image quality is estimated from the deviation between the statistical characteristics of the evaluated image and a model built from natural, undistorted images (Mittal, Soundararajan & Bovik, 2013). Lower NIQE values indicate that the image statistics are closer to those of natural images.

FID (Fréchet Inception Distance), introduced by Heusel et al. (2017), is also commonly used for generative approaches. FID compares the distributions of real and generated images by using the means and covariance matrices of their feature representations in the Inception feature space. Lower FID values indicate greater similarity between the two distributions, while stronger degradation generally results in higher FID values (Heusel et al., 2017).

Because these metrics capture different aspects of image quality, low-light image enhancement studies often use several of them together to assess both numerical fidelity and perceptual quality.

6. Comparative analysis of selected studies

The studies included in Table 2 were selected as representative examples of the main approaches discussed in this review. Selection was based on their relevance to a distinct methodological approach, the availability of quantitative evaluation on commonly used low-light datasets, and, where available, publicly accessible implementations.

Table 2 (part 1)

Comparative overview of selected studies on low-light image enhancement

Dataset	Size	Resolution	Pairing	Image format / domain
Lore, Akintayo and Sarkar (2017) LLNet / S-LLNet	Deep auto-encoder	Synthetically degraded and real low-light images; PSNR↑, SSIM↑	On dark images with added noise, LLNet and S-LLNet achieved better PSNR and SSIM results than the traditional methods used for comparison.	Joint contrast enhancement and denoising; trained on artificially darkened images with added Gaussian noise.
Wei et al. (2018) Retinex-Net	Retinex + deep learning	LOL, LIME, MEF, DICM; qualitative evaluation and comparison	The model brightened dark regions without severe overexposure and preserved details more effectively when enhancement and denoising were performed jointly.	Decom-Net + Enhance-Net; the study introduced the LOL dataset containing 500 image pairs.
Zamir et al. (2020) MIRNet	Multi-scale CNN	LOL, MIT-Adobe FiveK; PSNR↑, SSIM↑	MIRNet outperformed the compared methods on the LOL and MIT-Adobe FiveK datasets in terms of PSNR and SSIM.	Parallel streams at different resolutions and attention mechanisms; preservation of high-resolution details.
Jiang et al. (2021) EnlightenGAN	GAN, unsupervised learning	MEF, NPE, LIME, VV, DICM; NIQE↓, subjective evaluation	Achieved the best average NIQE score among the compared methods and the best average rank in the subjective evaluation.	Does not require paired low-light/normal-light images; uses global and local discriminators together with attention.
Guo et al. (2020); Li, Guo & Loy (2022) Zero-DCE / Zero-DCE++	Zero-reference	NPE, LIME, MEF, DICM, VV, etc.; subjective evaluation, PI↓, PSNR↑, SSIM↑, MAE↓	Zero-DCE received the highest average subjective score on 199 images, while Zero-DCE++ achieved the second-best result.	Training without paired or unpaired reference data; Zero-DCE++ contains approximately 10,000 parameters.
Wu et al. (2022) URetinex-Net	Retinex + deep unfolding	LOL, SICE; MAE↓, PSNR↑, SSIM↑, LPIPS↓	Outperformed most of the compared methods in terms of PSNR, SSIM, and LPIPS on LOL and showed good generalization to SICE without additional training.	Retinex optimization steps are unfolded into a learnable network; evaluated on SICE without retraining.
Xu et al. (2022) SNR-Aware	CNN + Transformer	LOL-v1, LOL-v2, SID, SMID, SDSD; PSNR↑, SSIM↑	Achieved better PSNR and SSIM results than the methods included in the comparison on LOL-v1, LOL-v2-real, and LOL-v2-synthetic.	An SNR map controls the balance between local convolutional processing and long-range Transformer processing.
Wang et al. (2023) LLFormer	Transformer	UHD-LOL4K, UHD-LOL8K, LOL, MIT-Adobe FiveK; PSNR↑, SSIM↑, LPIPS↓, MAE↓	Achieved the best results among the compared methods on UHD-LOL4K and UHD-LOL8K and also performed well on LOL and MIT-Adobe FiveK.	Designed for 4K/8K images; axis-based self-attention reduces computational complexity.
Cai et al. (2023) Retinexformer	Retinex + Transformer	LOL-v1/v2, SID, SMID, SDSD, FiveK, etc.; PSNR↑, SSIM↑	Outperformed the compared methods on several low-light datasets, with particularly strong improvements on images affected by noise and other degradations.	One-stage Retinex framework with an Illumination-Guided Transformer; evaluated on a large number of datasets.

Table 2 (part 2)
Comparative overview of selected studies on low-light image enhancement

Dataset	Size	Resolution	Pairing	Image format / domain
Jiang et al. (2023) DiffLL	Diffusion model + wavelet	LOL-v1, LOL-v2-real, LSRW and real-world unpaired datasets; PSNR↑, SSIM↑, LPIPS↓, FID↓, NIQE↓, BRISQUE↓, PI↓	Achieved the best quantitative results among the compared methods on LOL-v1, LOL-v2-real, and LSRW, together with improvements in perceptual metrics.	Diffusion is performed in a lower-resolution wavelet domain; the authors report at least 70× faster processing than the compared diffusion-based approaches.
Yi et al. (2025) Diff-Retinex++	Retinex + diffusion model	LOL-v1, LOL-v2-real/synthetic, LSRW, A-LLIE; PSNR↑, SSIM↑, LPIPS↓, FID↓, MAE↓	On LOL-v1, it achieved the best PSNR, SSIM, FID, and MAE results among the compared methods; on LOL-v2-real, it achieved the best overall performance.	Combines generative diffusion with Retinex information and a Mixture-of-Experts architecture.

Note: Numerical results from different studies should not be directly ranked when different datasets and experimental conditions are used.
Note: ↑ indicates that a higher value represents better performance, while ↓ indicates that a lower value represents better performance.

Since the reported results were obtained using different datasets, training settings, implementations, and experimental conditions, numerical values from different studies should not be treated as a single performance ranking.

A review of the selected studies shows a gradual expansion in the capabilities of low-light image enhancement methods. Early approaches such as LLNet were mainly focused on contrast enhancement and noise reduction, while Retinex-Net and later Retinex-based models introduced an explicit separation of illumination and reflectance within neural network architectures.

MIRNet highlighted the importance of multi-scale processing and attention mechanisms, whereas EnlightenGAN and Zero-DCE reduced the dependence on paired training data through unsupervised and zero-reference learning.

More recent Transformer-based approaches extend the ability to model long-range spatial dependencies and handle more complex degradation patterns. SNR-Aware uses local signal-to-noise ratio estimates to guide the enhancement process, LLFormer is specifically designed for ultra-high-resolution images, while Retinexformer combines a Transformer architecture with the Retinex principle.

Diffusion models represent an even more recent direction. DiffLL aims to reduce their computational cost through wavelet transformation, whereas Diff-Retinex++ combines generative diffusion with physically motivated Retinex information.

7. Conclusion

The reviewed studies show a gradual shift from earlier deep learning models, which mainly focused on contrast enhancement and noise reduction, toward more complex architectures that address illumination, noise, color, and structural details at the same time. Deep learning-based Retinex methods introduced an explicit separation of reflectance and illumination, while more recent models such as URetinex-Net and Retinexformer aim to overcome some of the limitations of classical Retinex approaches by learning implicit priors and modeling long-range dependencies between different image regions (Wu et al., 2022; Cai et al., 2023).

Another important direction is the reduced dependence on paired training data. Collecting low-light and normal-light images of the same scene can be difficult, especially under real-world conditions. EnlightenGAN addresses this issue by using unpaired data, while Zero-DCE does not require either paired or unpaired reference images during training (Jiang et al., 2021; Guo et al., 2020). Such approaches can make low-light enhancement more practical in cases where suitable reference datasets are difficult to obtain.

Transformer architectures improve the ability to model relationships between spatially distant image regions, but they can also increase computational complexity, particularly when processing high-resolution images. Recent models therefore use specialized attention mechanisms, hybrid CNN–Transformer structures, and architectures designed for large images (Xu et al., 2022; Wang et al., 2023; Cai et al., 2023).

A similar limitation is present in diffusion models because of their iterative processing. DiffLL shows that wavelet transformation can reduce processing time and memory requirements, although the authors note that the method is still not efficient enough for direct real-time video processing (Jiang et al., 2023).

The reviewed results also show that the quality of an enhanced image cannot be fully described by a single metric. PSNR and SSIM provide reference-based measures of reconstruction quality, while LPIPS, NIQE, and FID provide additional information related to perceptual quality. For this reason, quantitative evaluation is often supported by visual comparison and subjective assessment. At the same time, numerical results reported in different studies should be interpreted within the context of their datasets and experimental conditions rather than treated as a direct ranking of methods.

Overall, the choice of method depends on the application and available data. Supervised methods are suitable when paired low- and normal-light images are available, while unsupervised and zero-reference approaches are more practical when such data are difficult to obtain. Lightweight models are better suited to applications with limited computational resources, whereas Transformer-based models can be useful for high-resolution images. Diffusion models can provide strong enhancement results, but their higher computational cost may limit their use in real-time applications.

8. Acknowledgments

This research has been supported by the Ministry of Science, Technological Development and Innovation (Contract No. 451-03-34/2026-03/200156) and the Faculty of Technical Sciences, University of Novi Sad through project “Scientific and Artistic Research Work of Researchers in Teaching and Associate Positions at the Faculty of Technical Sciences, University of Novi Sad 2026” (No. 01-3609/1).

9. References

Bychkovsky, V., Paris, S., Chan, E. and Durand, F. (2011) 'Learning Photographic Global Tonal Adjustment with a Database of Input/Output Image Pairs'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 97–104. DOI: 10.1109/CVPR.2011.5995332

Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R. and Zhang, Y. (2023) 'Retinexformer: One-stage Retinex-based Transformer for low-light image enhancement'. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12470–12479. DOI: 10.1109/ICCV51070.2023.01149

Chen, C., Chen, Q., Xu, J. and Koltun, V. (2018) 'Learning to See in the Dark'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3291–3300.

Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S. and Cong, R. (2020) 'Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1780–1789. DOI: 10.1109/CVPR42600.2020.00185

Guo, J., Ma, J., García-Fernández, Á.F., Zhang, Y. and Liang, H. (2023) 'A survey on image enhancement for low-light images', *Heliyon*, 9(4), e14558. DOI: 10.1016/j.heliyon.2023.e14558

Guo, X., Li, Y. and Ling, H. (2017) 'LIME: Low-Light Image Enhancement via Illumination Map Estimation', *IEEE Transactions on Image Processing*, 26(2), pp. 982–993. DOI: 10.1109/TIP.2016.2639450

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. and Hochreiter, S. (2017) 'GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium'. In: *Advances in Neural Information Processing Systems*, 30, pp. 6626–6637.

Jiang, H., Luo, A., Han, S., Fan, H. and Liu, S. (2023) 'Low-Light Image Enhancement with Wavelet-based Diffusion Models', *ACM Transactions on Graphics*, 42(6), Article 238, pp. 1–14. DOI: 10.1145/3618373

Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P. and Wang, Z. (2021) 'EnlightenGAN: Deep Light Enhancement Without Paired Supervision', *IEEE Transactions on Image Processing*, 30, pp. 2340–2349. DOI: 10.1109/TIP.2021.3051462

Li, C., Guo, C. and Loy, C.C. (2022) 'Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8), pp. 4225–4238. DOI: 10.1109/TPAMI.2021.3063604

Liba, O., Murthy, K., Tsai, Y.-T., Brooks, T., Xue, T., Karnad, N., He, Q., Barron, J.T., Sharlet, D., Geiss, R., Hasinoff, S.W., Pritch, Y. and Levoy, M. (2019) 'Handheld Mobile Photography in Very Low Light', *ACM Transactions on Graphics*, 38(6), Article 164, pp. 1–16. DOI: 10.1145/3355089.3356508.

Liu, F. and Fan, L. (2025) 'A review of advancements in low-light image enhancement using deep learning', *Neurocomputing*, 652, 131052. DOI: 10.1016/j.neucom.2025.131052

Lore, K.G., Akintayo, A. and Sarkar, S. (2017) 'LLNet: A deep autoencoder approach to natural low-light image enhancement', *Pattern Recognition*, 61, pp. 650–662. DOI: 10.1016/j.patcog.2016.06.008

Lv, F., Lu, F., Wu, J. and Lim, C. (2018) 'MBLLEN: Low-Light Image/Video Enhancement Using CNNs'. In: *Proceedings of the British Machine Vision Conference (BMVC)*, paper 220.

Ma, L., Ma, T., Liu, R., Fan, X. and Luo, Z. (2022) 'Toward Fast, Flexible, and Robust Low-Light Image Enhancement'. In: *Proceedings of the IEEE/CVF Conference on*

Computer Vision and Pattern Recognition (CVPR), pp. 5637–5646. DOI: 10.1109/CVPR52688.2022.00555

Mittal, A., Soundararajan, R. and Bovik, A.C. (2013) 'Making a "Completely Blind" Image Quality Analyzer', *IEEE Signal Processing Letters*, 20(3), pp. 209–212. DOI: 10.1109/LSP.2012.2227726

Ren, W., Liu, S., Ma, L., Xu, Q., Xu, X., Cao, X., Du, J. and Yang, M.-H. (2019) 'Low-Light Image Enhancement via a Deep Hybrid Network', *IEEE Transactions on Image Processing*, 28(9), pp. 4364–4375. DOI: 10.1109/TIP.2019.2910412

Tian, Z., Qu, P., Li, J., Sun, Y., Li, G., Liang, Z. and Zhang, W. (2023) 'A Survey of Deep Learning-Based Low-Light Image Enhancement', *Sensors*, 23(18), 7763. DOI: 10.3390/s23187763

Wang, T., Zhang, K., Shen, T., Luo, W., Stenger, B. and Lu, T. (2023) 'Ultra-High-Definition Low-Light Image Enhancement: A Benchmark and Transformer-Based Method'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(3), pp. 2654–2662. DOI: 10.1609/aaai.v37i3.25364

Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) 'Image quality assessment: From error visibility to structural similarity', *IEEE Transactions on Image Processing*, 13(4), pp. 600–612. DOI: 10.1109/TIP.2003.819861

Wei, C., Wang, W., Yang, W. and Liu, J. (2018) 'Deep Retinex Decomposition for Low-Light Enhancement'. In: *Proceedings of the British Machine Vision Conference (BMVC)*, paper 155.

Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W. and Jiang, J. (2022) 'URetinex-Net: Retinex-Based Deep Unfolding Network for Low-Light Image Enhancement'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5901–5910. DOI: 10.1109/CVPR52688.2022.00581

Xu, X., Wang, R., Fu, C.-W. and Jia, J. (2022) 'SNR-Aware Low-Light Image Enhancement'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17714–17724.

Yang, W., Wang, W., Huang, H., Wang, S. and Liu, J. (2021) 'Sparse Gradient Regularized Deep Retinex Network for Robust Low-Light Image Enhancement', *IEEE Transactions on Image Processing*, 30, pp. 2072–2086. DOI: 10.1109/TIP.2021.3050850

Ye, J., Qiu, C. and Zhang, Z. (2024) 'A survey on learning-based low-light image and video enhancement', *Displays*, 81, 102614. DOI: 10.1016/j.displa.2023.102614

Yi, X., Xu, H., Zhang, H., Tang, L. and Ma, J. (2025) 'Diff-Retinex++: Retinex-Driven Reinforced Diffusion Model for Low-Light Image Enhancement', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(8), pp. 6823–6841. DOI: 10.1109/TPAMI.2025.3563612

Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H. and Shao, L. (2020) 'Learning Enriched Features for Real Image Restoration and Enhancement'. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.-M. (eds.) *Computer Vision – ECCV 2020*, Lecture Notes in Computer Science, vol. 12370, pp. 492–511. DOI: 10.1007/978-3-030-58595-2_30

Zhan, J., Goh, E.S. and Sunar, M.S. (2024) 'Low-light image enhancement: A comprehensive review on methods, datasets and evaluation metrics', *Journal of King Saud University – Computer and Information Sciences*, 36(10), 102234. DOI: 10.1016/j.jksuci.2024.102234

Zhang, L., Zhang, L., Liu, X., Shen, Y., Zhang, S. and Zhao, S. (2019) 'Zero-Shot Restoration of Back-lit Images Using Deep Internal Learning'. In: *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 1623–1631. DOI: 10.1145/3343031.3351069

Zhang, R., Isola, P., Efros, A.A., Shechtman, E. and Wang, O. (2018) 'The Unreasonable Effectiveness of Deep Features as a Perceptual Metric'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595. DOI: 10.1109/CVPR.2018.00068

Zhang, Y., Zhang, J. and Guo, X. (2019) 'Kindling the Darkness: A Practical Low-Light Image Enhancer'. In: *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 1632–1640.

Zheng, Y., Zhou, L., Jin, K., Jin, S., Liang, Z., Zhao, W. and Zhang, W. (2026) 'A comprehensive review of low-light image enhancement methods', *Signal Processing: Image Communication*, 140, 117435. DOI: 10.1016/j.image.2025.117435



© 2026 Authors. Published by the University of Novi Sad, Faculty of Technical Sciences, Department of Graphic Engineering and Design. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license 4.0 Serbia (http://creativecommons.org/licenses/by/4.0/rs/).