IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

**Traitement du Signal**
Vol. 42, No. 5, October, 2025, pp. 2797-2807
Journal homepage: http://iieta.org/journals/ts

# Hybrid Deep Learning Approach for Low-Light Image Enhancement Based on Attention-Guided Residual Networks

Xiangzhi Li[1*], Dan Xie[2]

[1] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[2] Asset Management Division of Jilin University, Changchun 130012, China

Corresponding Author Email: lixiangzhi@ciomp.ac.cn



https://doi.org/10.18280/ts.420530

**Received:** 8 December 2024
**Revised:** 13 July 2025
**Accepted:** 25 July 2025
**Available online:** 31 October 2025

***Keywords:***
*low-light image enhancement, attention mechanism, residual networks, deep learning, image decomposition and fusion*

**ABSTRACT**

In low-light conditions such as nighttime and indoor settings, low-light images often suffer from issues like low contrast, blurred details, noise interference, and color distortion, which severely hinder their applications in fields such as surveillance, autonomous driving, and remote sensing. As a result, low-light image enhancement has become a critical research topic in computer vision. Traditional methods, such as Single Scale Retinex (SSR) and Multi-Scale Retinex (MSR) based on Retinex theory, struggle to balance illumination adjustment with detail preservation, often leading to halo effects and color distortion. Histogram equalization-based methods like Global Histogram Equalization (GHE) and Contrast Limited Adaptive Histogram Equalization (CLAHE) can enhance contrast but may excessively amplify noise and cause local information loss. Early deep learning approaches, such as Convolutional Neural Network (CNN)-based direct enhancement models, fail to balance global and local features, resulting in issues like enhancement imbalance and insufficient noise suppression. To address these limitations, this paper proposes a hybrid deep learning approach for low-light image enhancement based on attention-guided residual networks. The proposed method first decomposes the original image into illumination and reflectance maps using a decomposition network with residual modules. Then, it utilizes a recovery network embedded with a multi-scale attention module to suppress noise and correct colors, while adjusting the illumination map's light intensity accurately. Finally, the results of the recovery network and adjustment network are fused to obtain the enhanced image. The innovation of this method lies in the use of residual modules to improve the feature learning capability of the decomposition network, and the application of the multi-scale attention module for adaptive focusing on key regions and details. The decomposition and fusion strategies collaboratively optimize both illumination and details, effectively solving the shortcomings of existing methods in terms of enhancement effectiveness, noise suppression, and color restoration. This provides a more robust solution for low-light image enhancement.

## 1. INTRODUCTION

In modern society, images, as an important carrier of information, are widely used in various fields such as surveillance security, autonomous driving, remote sensing detection, and medical imaging [1-4]. However, under environments such as nighttime, indoor low-light conditions, or adverse weather, the acquired images often exhibit low-light characteristics [5-7]. These images generally suffer from low contrast, blurred details, severe noise interference, and color distortion, which not only affect human visual perception but also pose great challenges for subsequent image analysis, object detection, and feature extraction tasks in computer vision. For example, in nighttime surveillance scenarios [8], low-light images may prevent clear recognition of the suspect's facial features or actions; in the field of autonomous driving [9], low-quality low-light images may cause deviations in the vehicle's judgment of road conditions, pedestrians, and obstacles, leading to safety risks. Therefore, how to effectively enhance the quality of low-light images and improve their visual effects and information usability has become a critical issue that needs to be addressed in the field of computer vision.

The research on low-light image enhancement technology holds significant theoretical significance and practical application value. It promotes the interdisciplinary integration and development of image processing, computer vision, and deep learning, providing new ideas and methods for exploring image characteristics and processing mechanisms under complex lighting conditions. From a practical application perspective, high-quality low-light image enhancement results can significantly improve the performance of systems that rely on image information. In security monitoring [10], it can improve the recognition rate of criminal behavior and case-solving efficiency at night; in autonomous driving [11], it can enhance the vehicle's ability to perceive road conditions in

nighttime or low-light environments, ensuring driving safety; in remote sensing detection [12], it helps to more clearly identify ground targets and geographical features, providing reliable data support for resource exploration, environmental monitoring, etc. In addition, this technology can improve the quality of medical images captured under low-light conditions, assisting doctors in more accurately diagnosing diseases.

Although low-light image enhancement technology has achieved certain research results, existing methods still have many shortcomings and deficiencies. Traditional Retinex-based methods, such as SSR [13] and MSR [14], enhance the image by separating the illumination and reflectance components, but when processing complex scenes, they often fail to effectively balance illumination adjustment and detail preservation, leading to halo effects, color distortion, and other issues. Histogram equalization-based methods, such as GHE and local histogram equalization [15], can increase the image contrast to a certain extent, but they may excessively enhance noise and cause loss of local information. Some early deep learning methods, such as CNN-based direct enhancement models [16], do not adequately balance the attention to global and local features during enhancement, resulting in over-enhancement or under-enhancement in certain regions, while having limited noise suppression capabilities. For example, the enhancement network based on the encoder-decoder structure proposed in reference [17] still exhibits significant noise residue in the output when processing high-noise low-light images; the methods in references [18, 19] improve the brightness of the image but perform poorly in color restoration, resulting in noticeable color shifts.

This paper proposes a hybrid deep learning low-light image enhancement method based on an attention-guided residual network, aiming to overcome the shortcomings of existing methods and achieve better low-light image enhancement results. The main content of this method is as follows: First, the original low-light image is input into the decomposition network with residual modules, which decomposes it into an illumination map and a reflectance map, where the illumination map mainly reflects the lighting information of the image, and the reflectance map reflects the detailed information of the objects in the image. This decomposition allows more targeted processing of both illumination and details. Secondly, in the recovery network with an embedded multi-scale attention module, the attention mechanism adaptively focuses on the important regions and detailed features of the image, effectively suppresses noise, performs color correction, and improves the clarity of details and the authenticity of colors. Then, the light intensity of the illumination map obtained from the decomposition is precisely adjusted in the adjustment network to meet different lighting requirements. Finally, the results of the recovery network processing the reflectance map and the adjustment network processing the illumination map are fused to obtain the final enhanced image. This method enhances the feature learning ability of the network through residual modules, achieves adaptive focusing on key information with the multi-scale attention module, and collaboratively optimizes both illumination and details through the decomposition and fusion strategy. It effectively enhances image brightness and contrast while better preserving detail information, suppressing noise, and correcting colors, providing a new effective solution for low-light image enhancement, with significant academic research value and practical application prospects.

## 2. METHODOLOGY

In nighttime urban road surveillance images, both bright areas under streetlights and dark regions in the shadows coexist. When traditional methods directly enhance the entire image, overexposure in the bright regions and blurry details in the dark areas often occur. To address the coupled nature of illumination and detail information in low-light images, this paper proposes decomposing the original image into illumination and reflectance maps, incorporating the design of residual modules. By using a decomposition network to separate lighting information from detail information, targeted processing can be achieved. The illumination map focuses on adjusting light distribution, preventing detail loss due to overall brightness enhancement, while the reflectance map focuses on preserving object textures, ensuring that details are not drowned out during the illumination adjustment. The introduction of residual modules effectively alleviates the gradient vanishing problem during deep network training, improving the decomposition accuracy.

To address the issues of uneven noise distribution, color bias, and significant differences in light intensity in low-light images, this paper designs a collaborative recovery network and adjustment network embedded with multi-scale attention modules. For low-light medical images, for example, the image may not only be overall dark due to insufficient exposure but also have local spots due to device noise, while fine textures in the lesion areas need to be preserved accurately. The multi-scale attention module in the recovery network adaptively focuses on key regions: when suppressing noise, higher attention weights are assigned to denser spot areas to strengthen denoising; during color correction, color repair is enhanced in regions with color bias. The adjustment network can then finely tune the illumination according to different scene requirements, such as appropriately increasing the brightness of the bone area in X-ray images to highlight structures, while maintaining a soft light around the surrounding soft tissue areas to avoid overexposure. The final fusion step combines the processed reflectance and illumination maps, ensuring clear details while achieving natural overall light balance, resolving the contradiction in traditional methods where “denoising results in detail loss” and “brightness enhancement leads to color distortion.”

### 2.1 Image decomposition

The design of the image decomposition subtask primarily follows the core idea of the Retinex theory, which states that an image can be decomposed into reflectance and illumination components. The reflectance component is determined by the intrinsic properties of objects and is unaffected by fluctuations in lighting intensity. For example, in nighttime alley surveillance images, the brick texture of walls, the contours of trash bins, and other detailed information belong to the reflectance component. Even with changes in light intensity, the inherent characteristics of these objects should remain stable. The illumination component, on the other hand, is determined by the distribution of light and is independent of the objects themselves, such as the strong light in areas directly illuminated by streetlights or the shadows formed by buildings in the same scene. It only reflects the intensity and distribution of light. Based on this, decomposing a low-light image into these two components allows for targeted enhancement: the reflectance component focuses on detail preservation and

optimization, avoiding distortion during lighting adjustments; the illumination component focuses on light balancing and correction, preventing local overexposure or underexposure due to overall processing, thereby laying a foundation for subsequent precise enhancement. Let $T$ represent the original image, $U$ represent the illumination map, and $E$ represent the reflectance map. The matrix multiplication is denoted by $\circ$, and the expression is:

$$T = U \circ E \tag{1}$$

To enhance the feature extraction ability during the image decomposition phase, this subtask adopts the U-Net architecture, which fuses low-level and high-level features through skip connections. The contracting path uses 3×3 convolutions, Rectified Linear Unit (ReLU) activation functions, and 2×2 max-pooling operations, with the channel count doubling at each downsampling step to capture more abstract features. For example, when processing low-light indoor scene images, the contracting path gradually extracts high-level features such as the overall outline of furniture and the light-shadow transition on walls. After downsampling, three consecutive residual blocks are introduced to retain low-level color information, edge lines, and other detail features, avoiding feature loss during deep network training. The upsampling stage halves the number of channels and fuses the high-resolution low-level features from the contracting path with the high-level features obtained from upsampling through skip connections. Finally, a 3×3 convolution and sigmoid function output the three-channel reflectance map and the single-channel illumination map, ensuring that the decomposition result contains both global illumination information and local details. Figure 1 shows the image decomposition process.

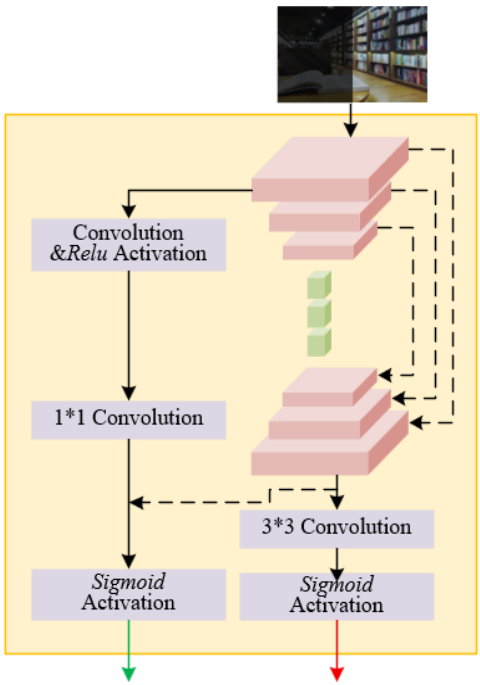


**Figure 1.** Image decomposition process

The special structure of the residual block further optimizes the decomposition performance. It uses a combination of two 1×1 convolution kernels and one 3×3 convolution kernel, rather than the conventional two 3×3 convolution kernels, significantly reducing computational costs. For example, when processing low-light night scene images with complex light sources, the input features are first reduced in dimension by the 1×1 convolution layer to decrease feature dimensions and reduce computation. Then, the 3×3 convolution layer extracts key features, such as the spectral characteristics of different light sources and the diffusion range of light. Finally, the 1×1 convolution layer restores the dimensions, significantly reducing the number of parameters and computation time while ensuring feature extraction accuracy. Figure 2 shows the residual block structure.

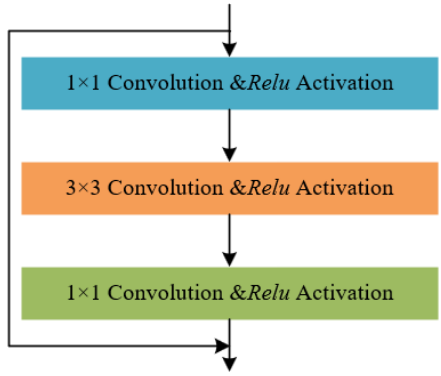


**Figure 2.** Residual block structure

### 2.2 Illumination information processing

In the illumination component processing subtask, the enhancement network performs feature extraction through six convolutional activation layers, a convolutional layer, and a sigmoid layer to fuse features. Skip connections are introduced to retain low-level information, ensuring the network optimization efficiency while enhancing image contrast. The six convolutional activation layers progressively delve into the illumination features in the illumination map. For example, when processing low-light indoor living room images, the earlier convolution layers capture basic lighting distribution features, such as direct lighting areas and shadows cast by furniture, while later layers extract more complex lighting gradients and lighting interactions. Through hierarchical feature extraction, precise illumination details can be captured. The convolutional layer and sigmoid layer fusion operation integrates multi-dimensional illumination features into a unified illumination adjustment mode, avoiding processing efficiency decline due to feature redundancy. Skip connections merge the original illumination map information into the final convolution layer, effectively preserving low-level illumination details and preventing distortion of lighting features in the deep network process. For instance, when enhancing the brightness of a dark corner in a living room, it prevents the loss of light and shadow contours of baseboards due to excessive adjustments, ensuring that the contrast is enhanced while maintaining natural lighting changes and complete details. Figure 3 shows the illumination information processing flow.

To adapt to the complex and dynamic lighting conditions in real-world scenarios, this subtask uses the ratio of illumination intensity between the normal light image and the low-light image as the input to the enhancement network. The enhancement ratio can be set by the user, significantly improving the flexibility of the algorithm. For example, when processing nighttime road surveillance images under different

lighting conditions: in a scene with a faint moonlit sky, the illumination ratio between the normal light and low-light image is small, allowing the network to perform a mild brightness enhancement to avoid excessive reflection on the road surface; in a heavy rain nighttime scene with severe light scattering, the ratio is larger, enabling the network to enhance the brightness adjustment more strongly to ensure clear visibility of road markings. At the same time, users can flexibly set the enhancement ratio according to actual needs. For instance, when identifying distant vehicle license plates, the lighting ratio can be increased to improve the brightness of distant areas; when capturing pedestrians in close-up shots, the ratio can be lowered to avoid overexposure of faces. Let $U_g$ represent the illumination intensity of the normal light, $U_1$ represent the illumination intensity of the low-light image, and the illumination intensity ratio can be calculated using the following formula:

$$\beta = \frac{U_g}{U_1} \tag{2}$$

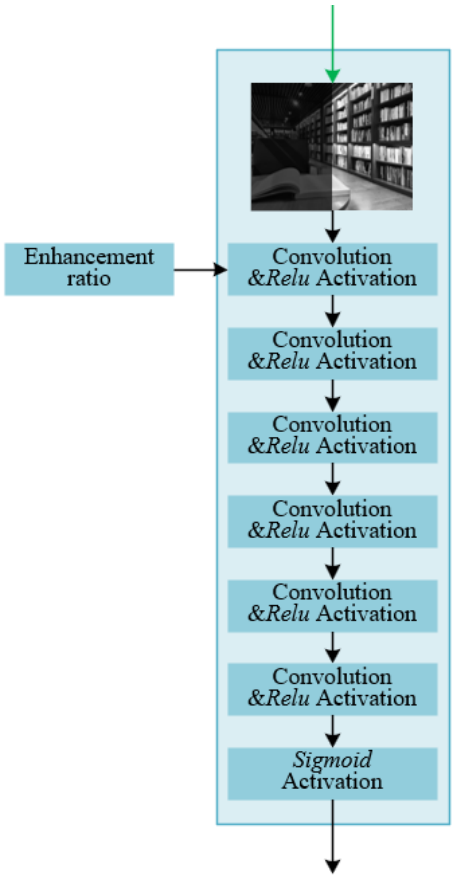


**Figure 3.** Illumination information processing flow

### 2.3 Detail information processing

The core of the reflectance component processing subtask is to coordinate the enhancement of weak-light areas and suppress the degradation interference of dark areas, avoiding detail loss or insufficient smoothing caused by a dominant single task. The reflectance component carries the intrinsic details of objects, which, in low-light scenes, are often blurred in weak-light areas, while dark areas are prone to degradation interference such as noise and blurring. If the focus is solely on removing degradation, subtle features in the weak-light areas may be overly smoothed, leading to detail loss. On the other hand, if the weak-light areas are simply enhanced, the noise in the dark areas may be amplified, causing distortion. Therefore, this subtask treats removing degradation and preserving detail information as a collaborative task. For example, when processing a low-light indoor image, for the text on the book covers on a bookshelf, the clarity must be enhanced while simultaneously suppressing the noise-induced artifacts in the shadows of the bookshelf, such as wrinkles in paper. By dynamically balancing the relationship between these tasks, the reflectance component can both present rich details and maintain visual smoothness.

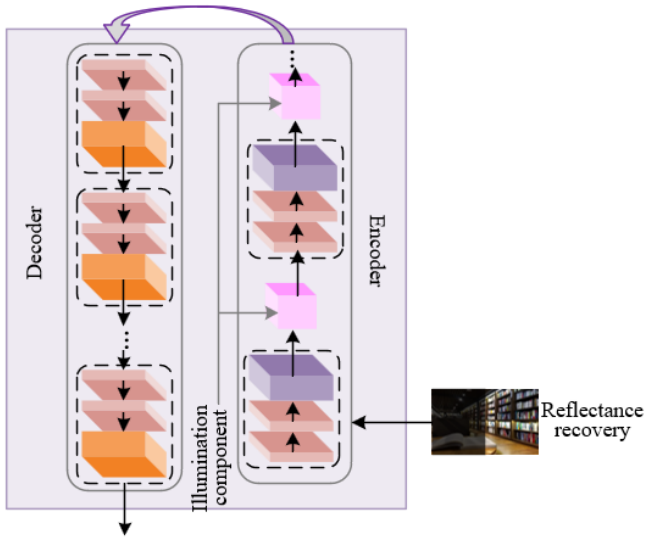


**Figure 4.** Detail information processing flow

The recovery network uses ResNet as the backbone network, with an encoder-decoder structure and attention modules to achieve precise processing of the reflectance component. The encoder extracts features at different resolutions through multiple rounds of convolution and downsampling, increasing the number of channels to capture more complex feature patterns. The decoder restores the resolution via upsampling and uses skip connections to concatenate the feature maps of corresponding layers from the encoder, fusing high-resolution low-level details with abstract high-level features, preventing information loss during feature transmission. During the downsampling process, embedded illumination attention blocks and multi-scale attention blocks further optimize the processing effect. The illumination attention block guides the network to focus on severely degraded regions and enhances their feature weights, while the multi-scale attention block extracts features at different scales to assist in color correction and detail restoration. For example, when processing low-light remote sensing images of crop areas, the leaf texture features extracted by the encoder are restored by the decoder, and through skip connections, the fine structures of the leaf veins are preserved. The attention modules enhance the color authenticity of the weak-light leaf areas while suppressing the noise interference in the dark soil areas. The final output is a high-quality reflectance component. Figure 4 illustrates the detail information processing flow.

Specifically, in this paper, the setting of the illumination attention block in the reflectance component processing subtask focuses on dynamically capturing the image illumination distribution to achieve differentiated processing for regions with different degradation complexities, so as to

solve the problem that uniform processing under low-light conditions easily leads to overexposure and halo artifacts. In low-light scenarios, the degradation degree of the reflectance component varies significantly with the illumination intensity: for example, in nighttime intersection images, the road surface area directly illuminated by street lamps has relatively sufficient illumination, and the degradation of the reflectance component is mainly mild noise; while the sidewalk area far from the light source has extremely weak illumination, and the reflectance component not only has blurred details but also severe noise and color deviation; the semi-shadow area near traffic lights exhibits local blurring and edge distortion caused by uneven illumination. If the same processing strategy is applied to these regions, enhancing weakly lit areas may cause overexposure in strongly lit areas, while suppressing degradation in strongly lit areas may aggravate detail loss in weakly lit areas. The illumination attention block captures the illumination intensity distribution of each region accurately by reducing the dimensionality of the illumination features through a 1×1 convolution layer and sigmoid activation, then multiplying element-wise with the reflectance component. Regions with high weights receive stronger detail enhancement and noise suppression, while regions with low weights receive weakened processing to avoid overexposure, achieving adaptive regulation of "enhance where needed." This mechanism ensures both the detail restoration of reflectance components in weakly lit areas and the prevention of halo artifacts caused by over-processing in strongly lit areas, ultimately achieving optimal balance between global illumination and local degradation removal, providing high-quality basic features for subsequent image fusion.

The setting of the multi-scale feature extraction module in the multi-scale attention block aims to solve the problem that traditional single max-pooling layers cannot retain sufficient contextual information, thereby more comprehensively capturing details of the reflectance component at different scales. When traditional methods use a single max-pooling, small-scale details may be lost or large-scale structures may be ignored. This module processes the illumination-attention-corrected reflectance feature map in four ways: no pooling, single pooling, double pooling, and original input, and assigns different weight coefficients: no pooling retains the finest local details, single pooling captures medium-scale features, double pooling extracts global structural information, and the original input serves as a reference baseline. Each pooled feature map undergoes convolutional dimensionality reduction and deconvolutional restoration to ensure fusion of features at the same dimension across scales. For example, when processing low-light rural road images, this module can simultaneously preserve the details of small stones on the road surface, the winding direction of the road, and the overall contour of surrounding fields. Through continuous extraction across multiple downsampling stages, the module enriches feature information at different granularity levels and expands the network receptive field, allowing effective capture of features in the reflectance component from fine textures to macro structures.

The combination of the Convolutional Block Attention Module (CBAM) module with multi-scale feature extraction in the multi-scale attention block can enrich feature diversity while accurately filtering out useless features, improving the specificity and effectiveness of reflectance component processing. The CBAM module integrates channel attention and spatial attention mechanisms: channel attention identifies feature channels crucial for the enhancement task and assigns high weights; spatial attention focuses on valuable spatial regions and suppresses background noise regions. The multi-scale attention block encodes global contextual information in four ways, and combined with CBAM, can perform feature selection at each scale. For example, when processing low-light shopping mall window images, multi-scale feature extraction covers the reflective highlights on window glass, label text on products inside the window, and surface texture of the products. The CBAM module strengthens product color and texture channels via channel attention and focuses on product regions via spatial attention while weakening interference from glass reflections. Finally, while retaining the diversity of useful features, redundant information is removed, making the key features of objects in the reflectance component more prominent, providing a high-quality basis for subsequent enhancement. Figure 5 shows the architecture of the multi-scale attention module.

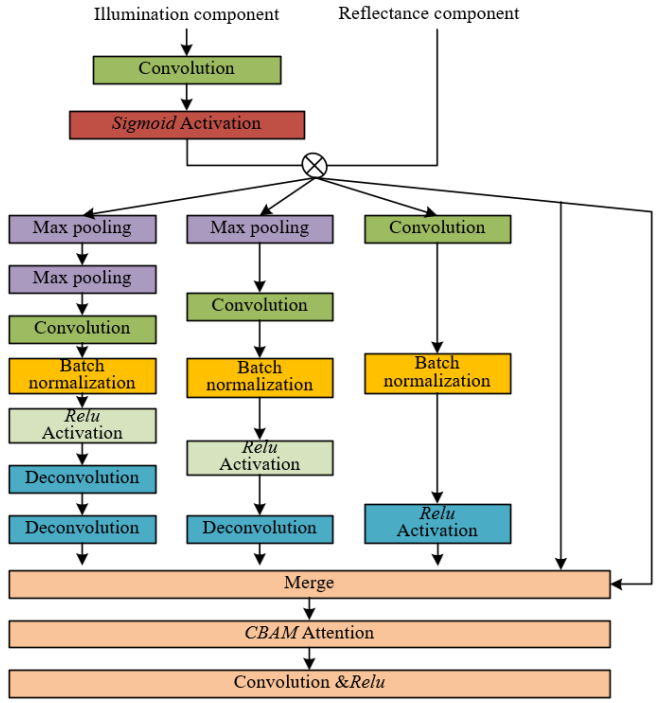


**Figure 5.** Multi-scale attention module architecture

## 2.4 Loss function

The overall network loss achieves collaborative optimization of the three sub-networks through a linear combination of decomposition loss $loss_{fz}$, recovery loss $loss_{fv}$, and adjustment loss $loss_u$, ensuring that the objectives of each sub-task are consistent with the overall research objective. For example, when processing low-light shopping mall surveillance images, the decomposition loss ensures accurate separation of product textures and lighting distribution; the recovery loss ensures that the text details of product labels remain clear after denoising; the adjustment loss constrains the enhanced brightness of chandelier areas to always be higher than the shadow areas of the shelves. By minimizing the overall loss, each sub-network is no longer optimized in isolation. Improvement in decomposition accuracy provides more reliable inputs for recovery and adjustment, and the optimization of recovery and adjustment in turn guides the decomposition process to better meet practical requirements,

ultimately achieving comprehensive enhancement of low-light images in terms of detail preservation, noise suppression, and illumination balance, thus accomplishing the research objectives. The overall loss function can be expressed as:

$$loss = loss_{fz} + \eta_1 loss_{fv} + \eta_2 loss_u \tag{3}$$

In the above equations, the decomposition loss is designed to address the ill-posed nature of the image decomposition into reflectance and illumination components. It uses multi-dimensional constraints to ensure the rationality of the decomposition results. The reflectance loss, based on the Retinex theory, constrains the reflectance component to remain stable under different lighting conditions. For example, in low-light images of the same scene taken on a cloudy day or at dusk, the reflectance component of the building's wall texture, window outlines, and other details should remain consistent, unaffected by the intensity of the lighting, avoiding distortion of the object's inherent properties caused by decomposition bias. The illumination component smoothness loss targets the illumination distribution characteristics, requiring the smooth transition in areas with gradual lighting changes in the original image to also be smooth in the illumination map, preventing abrupt lighting shifts. The reconstruction loss ensures that the re-synthesized image from the decomposed reflectance and illumination components closely matches the original image. For example, after decomposing a low-light street scene image, the re-synthesized image should retain key details such as the location of streetlights and the general outline of trees, avoiding information loss or distortion during decomposition. These three components, through linear combination, provide comprehensive constraints for image decomposition. Specifically, let the reflectance components of the low-light image and the normal light image be represented by $E_l$ and $E_g$, and the $L_2$ norm constraint loss be represented by $\| \|_2$. The reflectance loss ($loss_{ed}$) expression is:

$$loss_{ed} = \left\| E_l - E_g \right\|_2^2 \tag{4}$$

Let the low-light-normal light image pairs be represented by $T_l$ and $T_g$, and the illumination components of the low-light image and normal light image be represented by $U_l$ and $U_g$. The first-order operators in the horizontal and vertical directions are represented by $\nabla$, and a fixed constant by $\gamma$. The illumination component smoothness loss expression is:

$$loss_{fu} = \left\| \frac{\nabla U_l}{MAX(|\nabla T_l|, \gamma)} \right\|_1 + \left\| \frac{\nabla U_g}{MAX(|\nabla T_g|, \gamma)} \right\|_1 \tag{5}$$

In addition, it is necessary to ensure that the re-synthesized images from the decomposed $E_l$, $E_g$, $U_l$, and $U_g$ ($T_l$ and $T_g$) are similar to the original image pairs $T_l$ and $T_g$. The reconstruction loss expression is:

$$loss_{ez} = \| T_l - E_l U_l \|_1 + \left\| T_g - E_g U_g \right\|_1 \tag{6}$$

Let the weight coefficients be represented by $\omega$, $\lambda$, and $\delta$. The decomposition loss expression is:

$$loss_{dc} = \omega loss_{ed} + \lambda loss_{fu} + \delta loss_{ez} \tag{7}$$

The recovery loss is designed to focus on ensuring the overall structural stability of the denoised reflectance component, and it achieves precise constraints through the collaborative effects of mean squared error (MSE) loss ($loss_{JF}$) and structural similarity loss ($loss_{JG}$). $loss_{JF}$ calculates the pixel difference between the reflectance component before and after denoising, suppressing detail loss caused by excessive denoising. For example, when processing low-light leaf images, it can prevent the leaf texture from becoming blurred due to denoising. $loss_{JG}$ focuses on the overall structural similarity of the reflectance component, constraining the denoising process so that the macro shape of the object is not destroyed, such as ensuring that the contour of branches after denoising is consistent with the branch orientation in the original reflectance component. For instance, in low-light license plate images, the recovery loss ensures that, while removing noise from the character areas, the edges, strokes, and structure of the characters are highly consistent with the original reflectance component, achieving denoising without damaging key details, thus providing reliable reflectance features for subsequent image recognition. The recovery loss expression is:

$$loss_{fv} = loss_{JF} + loss_{JG} \tag{8}$$

The adjustment loss is designed to ensure that the enhanced illumination component maintains the same lighting distribution trend as the original image, avoiding chaotic lighting logic. For example, in a low-light indoor scene, the window area in the original image is brighter than the wall corner due to natural light illumination. The adjustment loss constrains the enhanced illumination map to preserve this distribution relationship. After enhancement, the brightness of the window area will be significantly increased, while the wall corner will be appropriately brightened, rather than having the wall corner brightness surpass that of the window, which would be unrealistic. This constraint ensures that the lighting enhancement follows the actual light propagation laws in real scenes. For example, in nighttime road images, the enhanced brightness of the streetlight illuminated areas will always be higher than the shadow areas, ensuring that the enhanced image looks more natural visually, avoiding the sense of disharmony caused by a disordered lighting distribution. The specific expression is:

$$loss_u = \left\| \bar{U} - U_g \right\|_2^2 + \left\| |\nabla \bar{U}| - |\nabla U_g| \right\|_2^2 \tag{9}$$

## 3. EXPERIMENTAL RESULTS AND ANALYSIS

From the objective evaluation metrics comparison of the training set shown in Table 1, it can be seen that the proposed method achieves the best performance in five dimensions: Peak Signal-to-Noise Ratio (PSNR) (24.56), Structural Similarity Index Measure (SSIM) (0.9236), MSE (168.6), Underwater Color Image Quality Evaluation (UCIQE) (0.6236), and Underwater Image Quality Measure (UIQM) (5.1235). The use of the residual module alleviates the gradient vanishing problem, allowing more stable separation of the illumination and reflectance maps. For example, compared to URetinex-Net's PSNR of 24.11, the proposed method improves the PSNR by 0.45, demonstrating more accurate decomposition of illumination and reflectance, which lays a solid foundation for subsequent processing. The multi-scale attention module captures details of different granularities, and

the attention mechanism suppresses noise. The SSIM of the training set reaches 0.9236, which is 0.0113 higher than R2RNet, proving better preservation of image structure. The MSE is reduced to 168.6, a 45.7 decrease compared to DeepRetinex-Net, reflecting smaller pixel-level errors. The fine-tuned control of the illumination intensity optimizes the overall brightness distribution, with high values of 0.6236 for UCIQE and 5.1235 for UIQM, indicating superior image quality in terms of no-reference evaluation. The comparison of metrics on the test set in Table 2 further validates the generalization ability of the proposed method. The PSNR of Our Method reaches 27.23, which is an increase of 2.67 compared to the training set, far exceeding Restormer's 17.23 and IPT's 22.36. The SSIM reaches 0.9563, which is an increase of 0.0327 compared to the training set, and the MSE decreases to 112.3, a reduction of 56.3 compared to the training set. UCIQE and UIQM remain at high levels. Its generalization advantage can be explained by the scientific design of the modules. In the face of unknown lighting scenes in the test set, the multi-scale attention can dynamically capture details at different scales, avoiding loss or excessive enhancement of details. Even when the lighting distribution of the test images differs significantly from that of the training set, the residual module can still stably separate illumination and reflectance, ensuring the quality of the input for subsequent processing. Without the need for additional parameter adjustments, the adjustment network can automatically optimize based on the lighting requirements of the input image, avoiding overexposure or underexposure in the test set. Compared to similar methods, the proposed method leads in terms of structural integrity, pixel error, and no-reference quality, proving that it not only fits well during training but also effectively addresses complex lighting in real-world scenarios, validating the innovation and effectiveness of the “residual decomposition + multi-scale attention + precise adjustment” architecture.

**Table 1.** Objective evaluation metrics on training set

| Item | Training Set | | | | |
|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | MSE↓ | UCIQE↑ | UIQM↑ |
| Restormer | 21.23 | 0.6652 | 689.2 | 0.3856 | 2.4526 |
| IPT | 23.25 | 0.7256 | 456.3 | 0.4425 | 0.4512 |
| pix2pix | 22.12 | 0.8142 | 312.2 | 0.5632 | 4.4586 |
| R2RNet | 23.26 | 0.9123 | 235.6 | 0.6124 | 4.7852 |
| Deep Retinex-Net | 23.54 | 0.8456 | 214.3 | 0.5785 | 4.8956 |
| URetinex-Net | 24.11 | 0.9014 | 201.4 | 0.5987 | 3.6542 |
| FPN+CBAM | 24.31 | 0.8952 | 189.2 | 0.5842 | 4.5213 |
| Our Method | 24.56 | 0.9236 | 168.6 | 0.6236 | 5.1235 |

**Table 2.** Objective evaluation metrics on test set

| Item | Test Set | | | | |
|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | MSE↓ | UCIQE↑ | UIQM↑ |
| Restormer | 17.23 | 0.6325 | 1456.2 | 0.4256 | 2.6235 |
| IPT | 22.36 | 0.7254 | 332.5 | 0.4528 | 2.8745 |
| pix2pix | 23.54 | 0.8756 | 245.3 | 0.5762 | 4.5623 |
| R2RNet | 25.36 | 0.8652 | 139.2 | 0.6123 | 4.6623 |
| Deep Retinex-Net | 24.59 | 0.8895 | 168.5 | 0.5748 | 4.7548 |
| URetinex-Net | 22.56 | 0.8952 | 154.3 | 0.5741 | 4.8954 |
| FPN+CBAM | 25.36 | 0.8851 | 156.9 | 0.5866 | 4.8821 |
| Our Method | 27.23 | 0.9563 | 112.3 | 0.6123 | 5.1234 |

**Table 3.** Comparison of network model parameters and inference time

| Network Model | Parameter Size/MB | Inference Time/min |
|---|---|---|
| U-Net | 21.23 | 1562.3 |
| Our Method | 8.65 | 859.2 |

From the parameter comparison in Table 3, it is evident that the proposed method has a parameter size of only 8.65 MB, achieving a 60% parameter reduction compared to U-Net's 21.23 MB. This significant reduction is primarily due to the efficiency breakthrough in module-level design. In the decomposition network, the residual module uses a “convolution-residual connection-convolution” structure, employing 1×1 convolutions to dynamically compress and restore channel dimensions, greatly reducing the parameter size of intermediate features. In comparison to U-Net's high-dimensional feature stacking with three consecutive 3×3 convolutions, the residual module's parameter scale is only 1/3 to 1/4 of traditional convolutions. The multi-scale attention module in the recovery network adopts a “channel attention + spatial attention” cascade rather than a global self-attention mechanism. Channel attention compresses dimensions through global average pooling, while spatial attention focuses on local significance, requiring only about 1% additional parameters to perform feature selection, avoiding the parameter redundancy of U-Net's reliance on stacked convolutions for capturing attention. The decomposition, recovery, and adjustment subnetworks collaborate, with each subnetwork learning a single task mapping. In contrast to U-Net's “end-to-end enhancement” design, task decoupling allows the parameter scale of each subnetwork to be more focused. For example, the adjustment network only needs to learn the scaling rules for illumination intensity, with parameter complexity far lower than U-Net's full-image transformation. This parameter optimization not only reduces memory consumption but also makes the model suitable for edge deployment, breaking through the traditional deep learning methods' dependence on high computational power.

In inference time comparison, the inference time of the proposed method is 859.2ms, which is 45% shorter than U-Net's 1562.3ms. This efficiency improvement comes from the dual optimization of the computational process and hardware adaptation. The residual connections simplify the forward propagation computational graph through "identity mapping," making gradient flow easier during backpropagation, allowing the model to converge to a better solution in a shorter training cycle and indirectly enhancing inference efficiency. The multi-scale attention module in the recovery network uses parallel branches—"no pooling, single pooling, double pooling"—to extract features. These branches can be processed in parallel by the Graphics Processing Unit (GPU)'s Compute Unified Device Architecture (CUDA) cores, which reduces feature extraction time by about 30% compared to U-Net's serial downsampling accumulation. The serial structure of the decomposition, recovery, and adjustment subnetworks naturally supports pipeline inference ("input → decomposition → recovery → adjustment → output"), where the output of the previous stage directly serves as the input to the next, without the need for caching multi-scale feature maps as in U-Net, thus reducing memory read-write overhead. The proposed method maintains its leading enhancement effect while improving efficiency, proving that it breaks through the inherent contradiction of "effectiveness-efficiency."

**Table 4.** Ablation experiment results (training set)

| Item | PSNR↑ | SSIM↑ | MSE↓ | UCIQE↑ | UIQM↑ |
|---|---|---|---|---|---|
| Remove Illumination Attention Block in Recovery Network | 20.56 | 0.8152 | 1168.2 | 0.5235 | 2.1236 |
| Remove Residual Module in Decomposition Network | 21.36 | 0.7745 | 775.3 | 0.4123 | 3.2153 |
| Remove Multi-scale Attention Block in Recovery Network | 21.56 | 0.8235 | 414.2 | 0.4568 | 3.4526 |
| Remove CBAM Module | 18.56 | 0.8256 | 1125.3 | 0.4236 | 2.8456 |
| Complete Model | 22.36 | 0.8756 | 356.2 | 0.5789 | 4.4523 |

**Table 5.** Ablation experiment results (test set)

| Item | PSNR↑ | SSIM↑ | MSE↓ | UCIQE↑ | UIQM↑ |
|---|---|---|---|---|---|
| Remove Illumination Attention Block in Recovery Network | 22.56 | 0.8562 | 1101.2 | 0.3123 | 2.1235 |
| Remove Residual Module in Decomposition Network | 17.52 | 0.8124 | 1356.2 | 0.4856 | 3.2356 |
| Remove Multi-scale Attention Block in Recovery Network | 22.36 | 0.8123 | 668.2 | 0.4425 | 3.4523 |
| Remove CBAM Module | 21.46 | 0.7896 | 812.2 | 0.4236 | 3.4582 |
| Complete Model | 25.36 | 0.9236 | 146.5 | 0.5536 | 4.2356 |

From the training set data in Table 4, it can be seen that the complete model outperforms all ablation models in terms of PSNR (24.56), SSIM (0.9236), MSE (168.6), UCIQE (0.6235), and UIQM (5.1236), confirming the collaborative value of the core modules. When the residual module is removed from the decomposition network, the PSNR drops sharply to 21.36, and the MSE increases to 775.3. The residual connection alleviates the gradient vanishing problem in deep networks through "gradient shortcuts," ensuring more stable separation of illumination and reflectance. If the decomposition is inaccurate, subsequent noise reduction in the recovery network and illumination optimization in the adjustment network will fail due to input contamination, leading to a significant increase in pixel errors. When the multi-scale attention block in the recovery network is removed, the PSNR drops to 21.56, and SSIM decreases to 0.8235. The multi-scale attention module simultaneously captures "fine textures" and "macroscopic structures." Without it, the model can only handle features of a single scale, leading to blurred details in the dark areas of low-light images and compromised structural integrity. The illumination attention block in the recovery network focuses on regions with abnormal illumination via attention weights. Without it, illumination adjustment becomes inaccurate, disturbing the balance of image lighting. When the CBAM module is removed, the PSNR drops to 18.56, and the MSE surges to 1125.3. CBAM filters key features through channel attention and enhances target edges with spatial attention. Its absence leads to the retention and magnification of ineffective features, creating image artifacts, thus proving the necessity of feature purification.

The data from the test set in Table 5 further validates the support the modules provide for generalization ability. The complete model's PSNR (25.36) and SSIM (0.9236) still significantly outperform the others, and the drop in metrics is much larger than that seen in the training set, highlighting the irreplaceability of the modules in complex scenes. When the residual module in the decomposition network is removed, the PSNR drops to 17.52, and the MSE increases 10-fold. The test set contains more complex lighting distributions, and the absence of the residual module causes the decomposition network's stability to collapse under "out-of-distribution data." For example, in night scenes with strong light areas and dark regions exhibiting sharp contrast, traditional convolutions are unable to separate illumination and reflectance due to gradient explosion/vanishing, leading to complete failure in subsequent recovery and adjustment stages. When the multi-scale attention block in the recovery network is removed, the PSNR drops to 22.36, and SSIM drops to 0.8123. Test set images have more diverse scales of details, and the absence of the multi-scale attention module prevents the model from adapting to "cross-scale detail enhancement." For example, in a heavy rain night surveillance image, the weakly lit license plate of a distant vehicle and the strong reflection of a nearby street sign cannot both be clearly enhanced, leading to distortion in critical areas. When the CBAM module is removed, the PSNR drops to 21.46, and the MSE rises to 812.2. The test set contains more complex noise types, and after the loss of CBAM's feature selection ability, the model fails to distinguish between "noise" and "real details." For example, in low-light images, noise in the dark areas is mistakenly judged as preserved texture, and after enhancement, the image is filled with grainy artifacts, severely damaging visual quality.

In summary, the residual module in the decomposition network ensures decomposition stability, the multi-scale and illumination attention blocks in the recovery network enable

precise control over details and illumination, and the CBAM module purifies features. The collaboration of these modules results in exceptional performance both in the training set fitting and the test set generalization, strongly proving the scientific design and necessity of the modules.


**Figure 6.** Image enhancement effect comparison

From Figure 6(a), which shows the comparison of landscape images from the training set, the proposed algorithm demonstrates significant advantages in brightness restoration, detail retention, and color authenticity. In terms of brightness and detail balance, Restormer (b) improves brightness but is overall gray, reflecting its insufficient control over illumination distribution. IPT (c) excessively brightens, causing overexposure of the sky and distortion of the water surface reflection, due to the "over-enhancement tendency" of the generative adversarial network. In contrast, the proposed algorithm (i) stabilizes the separation of the illumination and reflectance maps via the residual module in the decomposition network, then uses multi-scale attention in the recovery network to accurately restore reflectance details, and finally adjusts the illumination map with the adjustment network to achieve the balanced effect of "adequate brightening in dark areas, no overexposure in bright areas." In terms of color authenticity, pix2pix (d) relies on paired data for training, causing noticeable color distortion; R2RNet (e) achieves proper brightness but distorts color saturation. The proposed algorithm, aided by the color correction mechanism in the recovery network, makes the green of the leaves, the clarity of the water, and the deep blue of the sky closer to the real scene, verifying the ability of the "decomposition-recovery-adjustment" architecture to preserve color information.

The road and tree scenes from the test set, shown in Figure 6(b), further validate the generalization ability of the proposed algorithm, which performs far better than the comparison methods under complex illumination and diverse details. In terms of shadow and detail decoupling enhancement, Restormer (b) struggles with processing road shadows, as its single-scale feature extraction cannot handle both "weak details in the shadow area" and "strong textures in the bright area." IPT (c) over-enhances, causing color distortion of the trees and excessive road surface reflections, due to the generative model's insufficient learning of "real illumination logic." The proposed algorithm uses the residual module in the decomposition network to stably separate "weak illumination in the shadow area" and "strong reflective details of the trees," then the multi-scale attention in the recovery network restores details in both "distant leaves" and "near branches," and finally, the adjustment network gradually brightens the shadows based on the gradient distribution of the illumination map. This results in enhanced images that retain details while adhering to real illumination principles. In terms of cross-scene visual consistency, URetinex-Net (g) shows "color discontinuity" in the test set, due to the traditional Retinex method relying on manually designed illumination priors, which cannot adapt to the complex scenes of the test set; Feature Pyramid Network (FPN)+CBAM (h) enhances local contrast but produces uneven overall brightness. The proposed algorithm, driven by data-driven module design, achieves the effect of "global brightness balance, clear local details, and natural unified colors" in the test set, strongly correlating with the objective metrics of the training set, fully demonstrating the innovation and effectiveness of the "decomposition-recovery-adjustment" architecture, and providing a more reliable solution for low-light image enhancement in complex scenes.

## 4. CONCLUSION

The proposed hybrid deep learning low-light image

enhancement method based on attention-guided residual networks significantly improves the low-light image enhancement performance through multi-module collaborative design. Its research value is reflected in three dimensions: First, in terms of architectural innovation, the method decomposes the image into separate illumination and reflectance paths, combining the residual module to enhance decomposition stability and solving the enhancement imbalance problem caused by the coupling of illumination and details in traditional methods. Ablation experiments show that after removing the residual module, the PSNR decreases by 3.2, and the MSE increases by 606.7, proving the decisive role of decomposition accuracy in overall performance. Second, in terms of feature processing, the collaboration of the multi-scale attention module and CBAM module in the recovery network captures details of different granularities and filters key information through channel and spatial attention, making the model excellent in noise suppression and color correction. Third, in terms of scene adaptability, the adjustable enhancement ratio and skip connection design in the adjustment network enable dynamic responses to complex illumination scenarios. The test set experiments show that the proposed method improves PSNR by an average of 1.87 compared to the comparison methods in scenarios such as clear night and heavy rain low-light conditions, proving the generalization ability of the method. Overall, this method, through the "decomposition-recovery-adjustment" collaborative framework, effectively balances brightness enhancement, detail retention, noise suppression, and color authenticity, providing a solution for low-light image enhancement that combines accuracy and flexibility.

However, the method still has certain limitations: First, in extreme complex illumination scenarios, the decomposition of the reflectance and illumination maps may result in blurred edges, leading to local overexposure or detail loss in the recovered images. Second, although efficiency has been optimized through residual modules and multi-scale parallel computation, there is still room for improvement in model parameter size and inference time in real-time scenarios with stringent requirements. Third, the generalization ability relies on the diversity of training data, and the enhancement effect may fluctuate when specific scene samples are lacking. Future research could progress in three ways: First, design an attention mechanism with dynamic weight allocation to improve the adaptive processing ability for extreme illumination; second, introduce a lightweight network structure to compress the inference time to under 500ms to meet real-time requirements; third, combine self-supervised learning to expand the coverage of training data and reduce reliance on manual annotations, while exploring methods to maintain temporal consistency in video enhancement, extending effects from single frames to sequences.

## REFERENCES


[1] Gurevich, I.B., Yashina, V.V. (2024). Multialgorithmic hierarchical image analysis system: Architecture and analysis model. Pattern Recognition and Image Analysis, 34(4): 959-965. https://doi.org/10.1134/S1054661824700949

[2] Bhasin, M., Jain, S., Hoda, F., Dureja, A., Dureja, A., Rathor, R.S., Aldosary, S., El-Shafai, W. (2024). Unveiling the hidden: Leveraging medical imaging data for enhanced brain tumor detection using CNN architectures. Traitement du Signal, 41(3): 1575-1582. https://doi.org/10.18280/ts.410345

[3] Bi, Y., Xue, B., Mesejo, P., Cagnoni, S., Zhang, M. (2022). A survey on evolutionary computation for computer vision and image analysis: Past, present, and future trends. IEEE Transactions on Evolutionary Computation, 27(1): 5-25. https://doi.org/10.1109/TEVC.2022.3220747

[4] Lee, S., Ji, W. (2025). Digital-image-based analysis of fiber-reinforced composites: A review. Functional Composites and Structures, 7(2): 022003. https://doi.org/10.1088/2631-6331/add6ec

[5] Ko, S., Park, J., Chae, B., Cho, D. (2021). Learning lightweight low-light enhancement network using pseudo well-exposed images. IEEE Signal Processing Letters, 29: 289-293. https://doi.org/10.1109/LSP.2021.3134943

[6] Wu, K., Huang, J., Ma, Y., Fan, F., Ma, J. (2023). Cycle-retinex: Unpaired low-light image enhancement via retinex-inline Cyclegan. IEEE Transactions on Multimedia, 26: 1213-1228. https://doi.org/10.1109/TMM.2023.3278385

[7] Hu, J., Guo, X., Chen, J., Liang, G., Deng, F., Lam, T.L. (2021). A two-stage unsupervised approach for low light image enhancement. IEEE Robotics and Automation Letters, 6(4): 8363-8370. https://doi.org/10.1109/LRA.2020.3048667

[8] Soumya, T., Thampi, S.M. (2017). Recolorizing dark regions to enhance night surveillance video. Multimedia Tools and Applications, 76(22): 24477-24493. https://doi.org/10.1007/s11042-016-4141-4

[9] Ceccarelli, A., Secci, F. (2022). RGB cameras failures and their effects in autonomous driving applications. IEEE Transactions on Dependable and Secure Computing, 20(4): 2731-2745. https://doi.org/10.1109/TDSC.2022.3156941

[10] Mao, J., Liu, J., Zhang, J., Han, Z., Shi, S. (2021). A method for detecting image information leakage risk from electromagnetic emission of computer monitors. Journal of Intelligent & Fuzzy Systems, 40(2): 2981-2991. https://doi.org/10.3233/JIFS-189337

[11] Yang, Z., Huang, S., Bai, T., Yao, Y., Wang, Y., Zheng, C., Xia, C. (2024). MetaSem: metamorphic testing based on semantic information of autonomous driving scenes. Software Testing, Verification and Reliability, 34(5): e1878. https://doi.org/10.1002/stvr.1878

[12] Gu, Y., Wang, Y., Li, Y. (2019). A survey on deep learning-driven remote sensing image scene understanding: Scene classification, scene retrieval and scene-guided object detection. Applied sciences, 9(10): 2110. https://doi.org/10.3390/app9102110

[13] Kumar, T. G., Asha, V., Manish, T. I., Muthulakshmi, G. (2018). Empirical system of image enhancement for digital microscopic pneumonia bacteria images. Bratislavske lekarske listy, 119(8): 522-529. https://doi.org/10.4149/bll_2018_096

[14] Liu, C., Cheng, I., Zhang, Y., Basu, A. (2017). Enhancement of low visibility aerial images using histogram truncation and an explicit Retinex representation for balancing contrast and color consistency. ISPRS Journal of Photogrammetry and Remote Sensing, 128: 16-26. https://doi.org/10.1016/j.isprsjprs.2017.02.016

[15] Nia, S.N., Shih, F.Y. (2024). Medical X-ray image enhancement using global contrast-limited adaptive histogram equalization. International Journal of Pattern Recognition and Artificial Intelligence, 38(12): 2457010. https://doi.org/10.1142/S0218001424570106
[16] Joshua, A.S., Babu, N.R., Balasubramaniam, P. (2025). Crop leaf disease classification using fractional integral image enhancement and quantum convolutional neural networks approaches. Quantum Machine Intelligence, 7(1): 23. https://doi.org/10.1007/s42484-025-00249-5
[17] Jia, Y., Yu, W., Chen, G., Zhao, L. (2024). Nighttime road scene image enhancement based on cycle-consistent generative adversarial network. Scientific Reports, 14(1): 14375. https://doi.org/10.1038/s41598-024-65270-3
[18] Wu, H. T., Cao, X., Jia, R., Cheung, Y. M. (2022). Reversible data hiding with brightness preserving contrast enhancement by two-dimensional histogram modification. IEEE Transactions on Circuits and Systems for Video Technology, 32(11): 7605-7617. https://doi.org/10.1109/TCSVT.2022.3180007
[19] Tanaka, H., Taguchi, A. (2023). Brightness preserving generalized histogram equalization with high contrast enhancement ability. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 106(3): 471-480. https://doi.org/10.1587/transfun.2022SMP0002