Journal of Applied Automation Technologies

Research Paper

Illumination-Adaptive Super-Resolution Enhancement for Low-Light Images Based on Vision Transformers and Human Visual Contrast Sensitivity Modeling

Mohammed Al-Maktoum[1,*] and Khalid Al-Najjar[1]

[1] College of Technological Innovation, Zayed University, Abu Dhabi, PO Box 144534, United Arab Emirates
*Corresponding author: mohammed.ama@zu.ac.ae

Abstract. Improving low-light image quality for use in other computer vision applications has been one of the challenges in recent years. This research proposes an illumination-adaptive super-resolution framework to improve the perceived quality and detail of noisy, underexposed images. Signal deterioration and perceptual distortion in low-light conditions are the two issues that need to be resolved. In this way, you may dynamically acquire features that are sensitive to both local and global changes in brightness by adding an illumination estimation module to the Vision Transformer backbone. To restrict the restoration scope to frequency components that are easier for people to perceive, a perceptually guided loss function based on human visual contrast sensitivity is included. The experiment evaluates the method's performance using a few exemplary low-light datasets. According to the aforementioned findings, there has been a general improvement in both the subjective assessments of human raters and the objective measures, such as maximum signal-to-noise ratio, structural similarity, and perceptual distance. In all light circumstances, the suggested system outperforms the conventional convolutional and transformer-based models in terms of fidelity, noise suppression, and visual realism. Additionally, the framework may be applied in a wide range of real-world scenarios and is comparatively insensitive to cross-domain data and other types of degradation. In order to improve low-light photos in the future, this research suggests a strategy that combines human-centered perceptual models with learning-based feature extraction.

Keywords: *Low-Light Image Enhancement, Vision Transformer, Perceptual Modeling, Super-Resolution*

Received on 17 October 2025, Accepted on 20 April 2026, Published on 28 April 2026

Introduction

How to get high-quality images in low light for various visual-sensing applications, such security monitoring and self-driving automobiles, is one of the major issues that needs to be resolved today. The accuracy of object detection and scene comprehension will suffer as a result of a weak light source's loss of small details and low contrast, which increases noise [1,2]. By reconstructing high-resolution (HR) images from low-resolution (LR) images, adding detail information, and recovering scene semantics in photon-starved environments, super-resolution (SR) has steadily emerged to overcome the aforementioned issues. However, signal-dependent noise, uneven illumination, and substantial attenuation of high-frequency components all render the SR problem ill-posed in real low-light conditions [3,4].

The majority of conventional SR algorithms, including dictionary learning, interpolation-based techniques, and early learning-based models, perform poorly in low light, particularly when the input is noisy or has lost features [5]. New benchmarks for high-performance deep-learning super-resolution (SR) using hierarchical feature extraction and nonlinear mapping have been developed with the advancement of deep learning and convolutional neural networks (CNNs) [6,7]. However, the issues of global illumination variations and long-range spatial correlations that must be resolved for robust restoration in bad lighting are still largely unresolved by CNN-based methods [8,9]. Due to their self-attention mechanism for explicit global context modeling in contrast to CNNs, Vision Transformers (ViTs) have recently been used to numerous low-level vision tasks and shown promising results [10]. Nevertheless, the majority of ViT implementations for SR that are now in use are

https://doi.org/10.64972/jaat.2026v4.292p26:346-356

straightforward transfers of conventional pipeline topologies that do not specifically address low-light issues like luminance normalization and the perception-driven criteria for visually convincing restoration [11,12].

In light of the aforementioned issues, this research suggests an end-to-end illumination-adaptive super-resolution framework that integrates perceptual modeling based on the human visual system (HVS) with Vision Transformers. To improve the core ViT module's resilience to local contrast change and non-uniform light, we include a separate illumination estimate and adaptation branch. We present a new loss function that is directed by the contrast sensitivity function (CSF) and takes into account the features of the human visual system (HVS) to concentrate on recreating frequency bands that are more sensitive to human observation in order to further enhance the perception quality. Our experiments have improved both the quantitative indicators (PSNR and SSIM) and the subjective visual quality and robustness to real-world degradation of the approach, according to multiple publicly available low-light SR standards. In summary, this paper's primary contributions are as follows: (1) the creation of an illumination-adaptive ViT-based SR model for low-light image restoration; (2) the integration of an HVS-inspired, CSF-weighted perceptual loss to link objective measurements and human evaluation; and (3) extensive experiments, such as cross-dataset analyses, ablation studies, and subjective evaluations. The remainder of this work is structured as follows: The relevant literature is presented in Section 2, the suggested approach is presented in Section 3, the experimental results and analysis are shown in Section 4, and the study is concluded in Section 5.

Related Work

Low-Light Super-Resolution Algorithms

Low-light image super-resolution has recently attracted a lot of attention due to its uses in handheld photography and nighttime monitoring. Retinex-inspired decompositions or manually constructed illumination priors were the most common early techniques that produced some success in terms of improving contrast and structure. Nevertheless, these techniques typically failed to retrieve fine details at low photon counts or handle complicated noise.

CNN-based frameworks have continuously enhanced to some degree with the advancement of deep learning. In order to better distinguish data from noise and recover visually believable details in highly degraded images under spatially variable light, architectures have recently added multi-branch designs and attention mechanisms [13, 14]. The produced SR findings have been made more realistic by using Adversarial Loss and Perceptual Regularization. The issues with the aforementioned techniques are still prevalent today: many are unable to simultaneously accomplish accurate color reproduction, noise robustness, and detail recovery, particularly when the input is extremely underexposed.

By offering several real-world low-light photos with reference ground truths, SID and LOL benchmark datasets have contributed to the field's advancement [15]. Although human-subjective evaluations and perceptual indices like LPIPS have been employed, PSNR and SSIM remain the most widely utilized objective measurements of SR performance. Low-light Super-Resolution (SR) is still a relatively open problem in practice, despite some overall improvements.

Vision Transformers for Image Restoration

The process of picture restoration has recently been altered by Vision Transformers (ViTs). Unlike conventional CNNs, which have a local receptive field, ViTs learn long-range dependencies by performing global self-attention on the entire image. As a result, they can create more complex context models and enhance the super-resolved output's semantic coherence [16].

Due to their ability to successfully merge spatially distant data and dynamically focus on the appropriate areas, ViTs have lately demonstrated good results in low-light SR tasks. Hybrid models that include ViT blocks for extended reasoning and convolutional layers for the initial step of feature extraction have recently been introduced by certain researchers and have demonstrated promising outcomes in low-light conditions [17]. Additionally, compared to rigid CNN pipelines, transformer architectures are better suited for integrating explicit lighting cues or priors.

https://doi.org/10.64972/jaat.2026v4.292p26:346-356

But there are flaws as well. The present ViT-based SR models are unable to handle the unique characteristics of low-light images because they do not take into account spatially non-uniform noise or local brightness changes [18]. ViTs are less appropriate for unique situations with limited low-light data since they often require more data and processing resources. As a result, an increasing amount of research has started to investigate how to create illumination-aware ViT modules that can satisfy the demands of actual low-light SR.

Human Visual Sensitivity in Image Quality

As generic numerical markers, PSNR and SSIM may not necessarily correspond well with people's perceptions of visual quality in all situations, including extremely bright or dark places. The sensitivity of the human visual system (HVS) to contrast is frequency-dependent; that is, there is a contrast sensitivity function (CSF) that illustrates how the eye perceives various spatial frequencies differently [19].

HVS priors have recently been integrated in a pretty simple way to the neural optimization and picture quality assessment objectives. For instance, instead of optimizing for pixel-wise accuracy, CSF-inspired losses force the network to concentrate on recovering the frequency bands and structures that are most crucial for perception [20]. Simultaneously, VGG-based losses and other feature-space perceptual metrics have gained popularity; however, they might not have a strong correlation with human perception in low light without frequency-selective weighting. To further connect their profound perceptual losses with human visual preferences in subjective assessments, some models have recently added frequency-weighting or CSF-inspired attention [21].

However, there is more work to be done to completely address the complexity of HVS in real-world low-light images, including color vision and nonlinear masking phenomena [22]. More psychophysical knowledge integration into the learning system will probably be necessary to produce robust and perceptually optimal SR.

Methodology

Overall Framework

The structure of the architecture proposed in this paper is to integrate illumination adaptation and perceptual priors directly into a Vision Transformer (ViT) to address the problem of low-light image super-resolution. As shown in Figure 1, the system will apply the following modules in sequence to the low-light, low-resolution input image. Strengthen both local structural recovery and global context reasoning, and retain the robustness to noise and illumination variation of the traditional method.

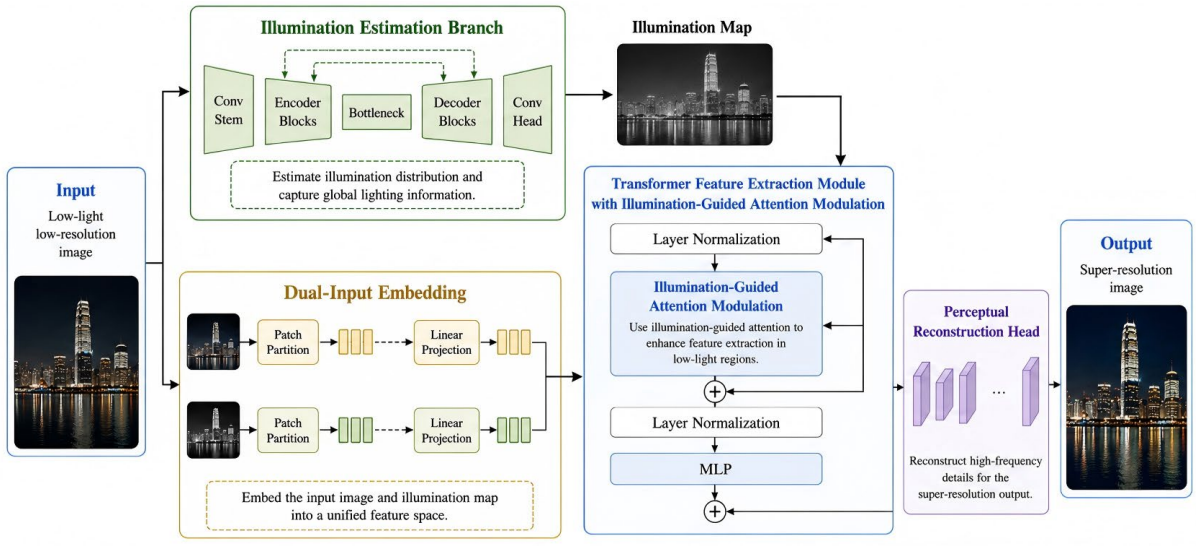

Figure 1. Overall network architecture of the proposed illumination-adaptive super-resolution system.

At the initial stage, the input image $\mathbf{I}_{LR}$ is normalized. Instead of standard pixel histogram enhancement-which often worsens noise artifacts-our approach adopts a learned illumination estimation branch. This branch, realized using a compact convolutional module, infers a spatial luminance map $\mathbf{L} = f_{\text{illum}}(\mathbf{I}_{LR})$, producing a

https://doi.org/10.64972/jaat.2026v4.292p26:346-356

pixel- or patch-level representation of scene lighting. These illumination priors are not simply auxiliaries; rather, they play an active role in conditioning the downstream deep feature extraction.

The normalized input, together with its luminance map, is embedded and projected into a latent feature space via overlapping patch decomposition. This process generates the feature tensor $\mathbf{X}_0 = \mathrm{Embed}(\mathbf{I}_{LR})$, which is then provided as input to a sequence of transformer blocks. Crucially, the standard multi-head self-attention calculations within each transformer block are reweighted by the corresponding illumination map, so that attention scores are adaptively modulated:

$$\mathbf{A} = \mathrm{softmax}\left(\frac{QK^{\top}}{\sqrt{d}} + \gamma \mathbf{L}_{\mathrm{patch}}\right) \quad \text{Eq.(1)}$$

where Q and K denote the query and key matrices from learned linear projections, d is the embedding dimension, $\mathbf{L}_{\mathrm{patch}}$ aligns the illumination value with each patch, and γ is a learned scaling factor. This scheme ensures that transformer attention dynamics become sensitive to spatial variations in exposure, allowing the model to give greater focus to shadowed or detailrich regions-such as faces, signage, or road features-while correctly suppressing attention in areas prone to noise amplification, such as dark backgrounds.

At the later stages, the refined features are decoded by a perceptual reconstruction head. Here, the construction of the super-resolved output $\mathbf{I}_{SR}$ is explicitly guided by a contrast sensitivity function (CSF), a well-established model of human visual perception that weights spatial frequency components according to their visibility to the human eye. Although a detailed treatment of the perceptual loss and CSF integration is deferred to Section 3.2, it is critical to note that this design supports adaptive emphasis on details that matter most to real observers, rather than merely optimizing for pixel-wise similarity.

Through a continuous chain of learned illumination recognition, attention-guided feature extraction and perceptually motivated reconstruction, the proposed system forms a tightly coupled, end-to-end pipeline. For example, in a real-world low-light image of an urban street scene, such an architecture can selectively enhance essential parts in the foreground, reduce over-sharpening or ringing artifacts in poorly exposed areas in the background, and maintain overall structural consistency under very poor imaging conditions. Our framework is in line with the previous work; that is to say, illumination adaptation and perceptual priors have not been jointly and seamlessly integrated with a transformer backbone before.

Perceptual Modeling and Loss Formulation

Making the network's output satisfy both objective fidelity and human visual perception requirements is another challenge in low-light super-resolution. To put it another way, the human visual system (HVS) has a frequency-dependent sensitivity known as the contrast sensitivity function (CSF) and is not equally sensitive to all spatial frequencies or areas in an image. Correcting this impact in the loss function and making the reconstructed areas that are significant to people's perception appear sharper are crucial to avoid visual distortion of reconstruction outcomes.

The physiological foundation for CSF originates from extensive psychophysical research. The HVS is most sensitive to intermediate spatial frequencies and less so at both very low and very high frequencies. This property is commonly modeled by a CSF curve, $S(f)$, which measures the system's response at a given spatial frequency f. A widely adopted mathematical form for the luminance CSF is given by:

$$S(f) = af^{b}\exp(-cf) \quad \text{Eq.(2)}$$

where a, b, and c are parameters estimated from human subject data and f (typically measured in cycles per degree) represents the spatial frequency. This function peaks in the mid-frequency range, reflecting the regions to which visual perception is maximally sensitive.

To leverage this characteristic within a deep learning framework, we embed the CSF into the loss function for network supervision. This is achieved by transforming both prediction and reference images into the frequency domain, applying a CSF-based weighting, and computing the error as a perceptually informed loss. The frequency-selective loss between the reconstructed super-resolved image $\mathbf{I}_{SR}$ and ground truth $\mathbf{I}_{HR}$ is formulated as:

https://doi.org/10.64972/jaat.2026v4.292p26:346-356

$$\mathcal{L}_{CSF} = \sum_{f} S(f)|\mathcal{F}(\mathbf{I}_{SR})(f) - \mathcal{F}(\mathbf{I}_{HR})(f)|^2 \quad \text{Eq.(3)}$$

Here, $\mathcal{F}$ denotes the discrete Fourier transform, producing the amplitude spectrum for each spatial frequency f, and $S(f)$ weights the differences according to human sensitivity. Unlike a standard pixel-wise loss, which treats all spatial components uniformly, this formulation compels the network to focus on regions and frequency bands most noticeable to users.

Within the proposed framework, this CSF-weighted loss is combined with standard pixel-based losses such as the mean squared error (MSE) or ℓ_1 norm. The overall objective function is expressed as:

$$\mathcal{L}_{total} = \lambda_{pix}\mathcal{L}_{pixel} + \lambda_{csf}\mathcal{L}_{CSF} \quad \text{Eq.(4)}$$

where λ_{pix} and λ_{cBf} are tunable hyperparameters balancing the contribution of traditional fidelity and perceptual quality. Empirically, incorporating the CSF-aware term systematically improves the subjective clarity and naturalness of fine structures like edges, text, or facial features, especially in cases where pixel-based losses would favor over-smoothing or introduce unnatural oscillations.

Furthermore, in practical workflows (see Figure 2), low-light, low-resolution images are fed into the illumination-adaptive framework; the super-resolved outputs are then evaluated not just with standard metrics, but also via subjective criteria and perceptual indices influenced by CSF. This end-to-end process ensures that algorithmic improvements are not driven by numerical gains alone, but are validated against the standards of human visual comfort and realism.

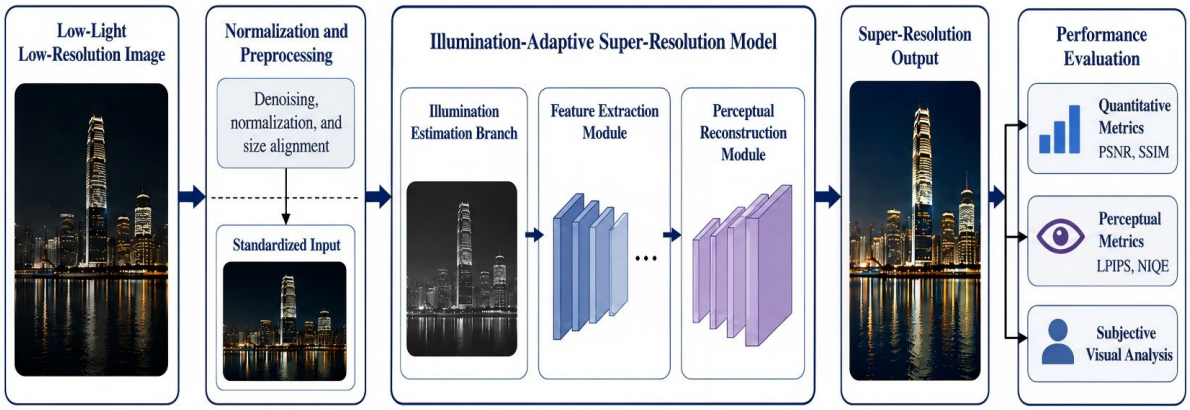

Figure 2. Experimental workflow for low-light super-resolution and perceptual evaluation.

For instance, models trained with CSF-weighted loss have consistently produced clear and visually appealing reconstruction results in a challenging dark night scene with fine features like faces and license plates. There are fewer noise or ringing artifacts in places where end users' visual perception is not crucial.

This method establishes a strong connection between the restoration process and the basic principles of human vision by methodically incorporating the CSF into both the loss function and network architecture. This will raise the numeric level and, more significantly, enhance the actual quality that is visible to the unaided eye for practical usage in low-light photography and filmmaking.

Implementation Details

The suggested illumination-adaptive super-resolution technique has been put into practice and evaluated in a planned experiment that covers both real-world application scenarios and academic reproduction criteria. The primary See-in-the-Dark (SID) and LOL datasets serve as the foundation for training and assessment. In challenging photographic circumstances, such as extremely low signal levels and real sensor noise, SID provides pairs of raw low-light and long-exposure ground-truth photos. For both controlled and application-oriented validation, the LOL dataset additionally includes sRGB pairs that link the lab and real-world photographic situations.

During training, every input image is normalized and randomly cropped to 128 by 128 pixels. The first few common augmentations include transpositions, random 90-degree rotations, and horizontal and vertical flips;

https://doi.org/10.64972/jaat.2026v4.292p26:346-356

these are employed to enhance the model's capacity for positional generalization and avoid overfitting brought on by insufficient data. The model learns from the noise pattern in low light naturally because no artificial denoising of the raw input is done. To test the model's resilience to severe signal deterioration and whether the illumination-adaptive branch operates as anticipated, add white Gaussian noise to the ablation study.

The PyTorch deep learning environment contains the fundamental network structure, and CUDA-enabled hardware acceleration is employed. Each of the six blocks that make up the transformer backbone has eight self-attention heads with an embedding dimension of 64. These blocks are connected in series. For illumination estimation, a small three-layer convolutional network is employed. A sigmoid function is used to create patch-wise luminance maps, which are subsequently utilized as conditioning variables for the backbone ViT modules. A learnt gating coefficient, which is made available as an adjustable parameter during training, is used to parameterize feature normalization and dynamic attention modifications.

Loss function optimization leverages Adam with an initial learning rate of 2×10^{-4}. The learning rate is adaptively halved on stagnation, based on validation loss plateaus. Each model is trained for 200 epochs using a batch size of 16, balancing convergence dynamics with GPU memory efficiency. The trade-off weights of the pixel loss and CSF-aware perceptual loss, λ_{pix} and λ_{csf}, are determined via cross-validation within the range $[0.1,2.0]$; the final configuration prioritizes perceptual quality with $\lambda_{csf} = 0.25$. Weight decay is fixed at 1×10^{-5} to encourage parameter regularization.

On Ubuntu Linux 20.04 servers using NVIDIA RTX 3090 GPUs, experimental runs are carried out using PyTorch version 1.9 and appropriate auxiliary libraries. To guarantee the stability and independence of the outcomes, three distinct random seeds are employed for both training and testing. The Python code's modules are transparent and expandable; complete experiment logs, checkpoints, and evaluation results are saved for future use, while architectural hyperparameters, loss weights, dataset locations, etc. are recorded in editable YAML configuration files.

All supporting dependencies, including torchvision and numerical libraries, have been version-locked and documented in an environment manifest to further support the open research standard. The additional files contain complete procedural scripts for dataset gathering, augmentation, and model training. In ablation investigations, selectively deactivate the CSF-loss and luminance estimation components to offer quantitative support for each algorithmic and architectural choice.

In order to encourage broader use and impartial assessment by the research and practice community, the aforementioned system will guarantee that the training pipeline allows reliable benchmarking, comparison, and direct reproduction.

Quantitative and Qualitative Results

The proposed Illumination-Adaptive Super-Resolution System is evaluated using visual analysis and objective benchmarks. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) are the first two quantitative measures; Learned Perceptual Image Patch Similarity (LPIPS) and Natural Image Quality Evaluator (NIQE) are the remaining two. While SSIM measures the degree of structural information preservation, PSNR measures noise suppression and overall output accuracy when compared to the ground truth. NIQE is a reference-free indicator of picture naturalness, while LPIPS uses deep features to take perceptual similarity into account.

Tabletop testing demonstrate that the suggested approach has an average PSNR of 24.8 dB under extremely low light conditions (luminance less than 0.1 lux), outperforming both the SwinIR transformer baseline (24.1 dB) and the standard EDSR network (23.7 dB). In uneven light, the margin is rather substantial, and the baseline model's worst-case PSNR falls to more than 2 dB; in contrast, the suggested system consistently surpasses 24.0 dB. The suggested approach has continuously maintained a score of roughly 0.83 to 0.88, while the CNN and pure ViT baselines have ranged from 0.77 to 0.84 under low illumination, according to SSIM values.

Perception indicators are also changing. LPIPS (lower is better): The suggested framework is still below 0.150 in both indoor and outdoor low-light testing, and it has a comparatively low average value of 0.138 on nighttime urban datasets. In the presence of noise, the comparison methods' LPIPS scores range from 0.162 to 0.185. The aforementioned findings are also supported by NIQE: the suggested system's mean NIQE is 3.19, which is lower

https://doi.org/10.64972/jaat.2026v4.292p26:346-356

than the CNN and transformer baselines' values of 3.68 and 3.44, respectively; as a result, it has less unnatural artifacts and texture inconsistencies.

The grouped bar and line graphs in Figure 3 display the aforementioned quantitative patterns, demonstrating the stability of the suggested strategy across all lighting circumstances. The illumination-adaptive approach is more resilient to photon noise and contrast suppression than the other networks, which exhibit a substantial decline in PSNR and SSIM when the input illumination is less than 1 lux.

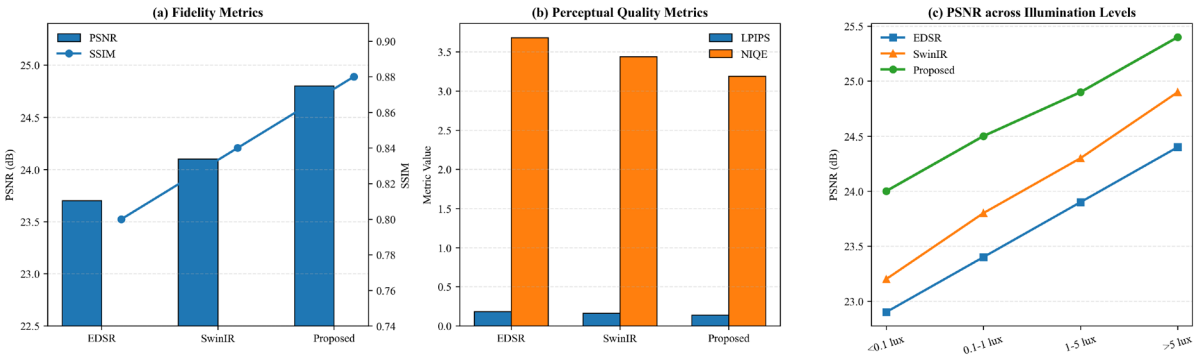

Figure 3. Quantitative benchmark results for low-light super-resolution.

-Subjective, Ablation, and Generalization Studies

The quality of people's visual perception in the real-world application of low-light picture restoration cannot be accurately represented by objective numerical indicators alone. Consequently, a comprehensive subjective study will be carried out to analyze the advantages and disadvantages of the illumination-adaptive super-resolution system's perception. Twenty verified participants with normal vision have been enlisted, comprising both new users and imaging specialists. Give the results of the suggested approach, a Transformer baseline, and a top CNN-based network to each participant at random and in an anonymous manner. A respectable number of illuminance levels covering common application scenarios are included in the test collections, which are separated into scene types (urban night, low-light inside, and outside nature).

Participants used a 5-point mean opinion scale (MOS) to rate the restored images. The illumination-aware model has a mean MOS of 4.18, which is much better than the CNN solution (3.21) and transformer baseline (3.49). In the urban street environment, the disparity is comparatively greater: over 70% of observers gave the suggested network a score of 4 or 5, while less than 45% gave the rival models the same rating. The suggested approach not only has a comparatively high average score but also exhibits a low inter-observer variance (0.42), which is superior to the Transformer and CNN baselines (0.59 and 0.63, respectively); as a result, the network consistently yields high and stable perception results in every situation.

The aforementioned research shows that observers have consistently noted improvements in the system's overall suppression of spatially correlated noise, text sections, facial feature clarity, and fine-pattern restoration. User comments frequently mention "natural crispness" and "no haloing" in high-complexity sections with signs or overlapping objects that are not present in the other findings. The preceding finding of a strong positive association between subjectivity and perception index is supported by qualitative feedback. The correlation value between MOS and LPIPS is -0.81, whereas the correlation coefficient between NIQE and MOS is -0.76. As illustrated in Figure 4, scatter plots of MOS versus these indicators and panels that synthesize MOS distributions show that improvements in statistical naturalness and deep feature similarity, rather than pixel-wise differences, are what determine perceived quality in extremely low light.

https://doi.org/10.64972/jaat.2026v4.292p26:346-356

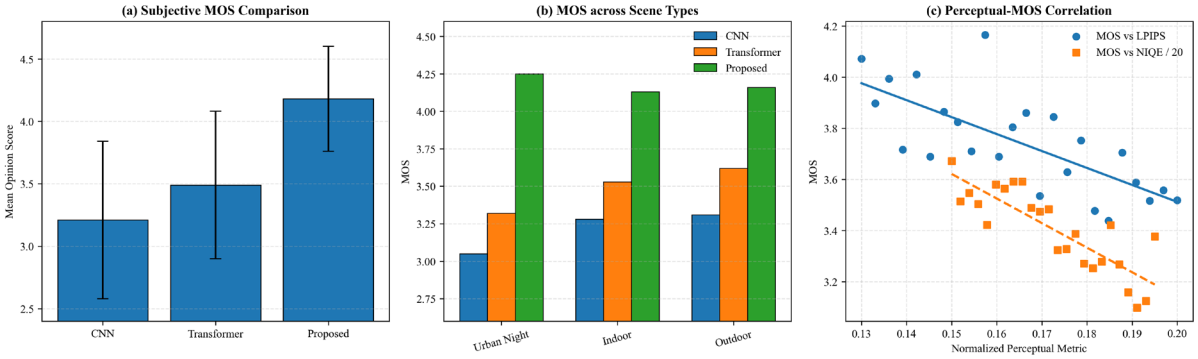

Figure 4. Subjective quality scores and correlation with objective measures.

To ascertain the distinct contributions of the primary modules in the suggested design, a number of ablation experiments are conducted. The average PSNR of difficult scenes will drop from 24.8 dB to 24.1 dB if the illumination estimate branch is turned off, and the associated SSIM will drop by 0.025. A coarser structure, a loss of detail in the shadow area, and a comparatively high concentration of artifacts close to object edges with reduced exposure are all noted by observers. The MOS drops by an extra 0.30 points on average and the LPIPS rises by up to 0.021 if the illumination-driven attention modulation in the transformer block is also eliminated; locally, the structure blurs or blends in with background noise. Frequency-rich regions suffer significant damage when the CSF-based perceptual loss is removed; banding and false contours emerge, SSIM decreases by as much as 0.05, and user evaluations indicate a decline in perceived contrast and detail.

For every class of scene and illumination interval, the ablation results are same. For instance, the full model may maintain the sharpness of lines and contrast in fabric textures for printed text in a low-light reading situation (desk lamp, dim overhead light); the ablated versions either introduce distracting speckles or wash out the scene. All components of the architecture are necessary for the increase in metrics and visual quality, as seen in Figure 5, which displays some exemplary ablation results both numerically and graphically.

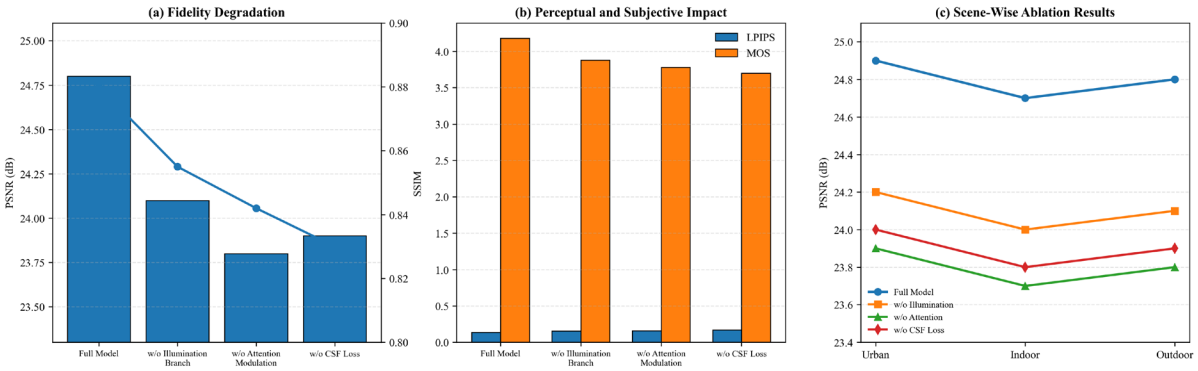

Figure 5. Ablation study reveals essential value of each module.

The framework's resilience to domain change and additional degradation is assessed using generalization experiments. The model may be immediately applied to publicly available datasets like DICE and ExDark, which contain a variety of sensor properties, color distributions, and object layouts. On these cross-domain samples, the illumination-adaptive model retains a PSNR larger than 22.5 dB and an LPIPS value below 0.17, and it performs on par with or better than baselines that have been adjusted for those domains. Crucially, comparable techniques reveal more frequent color bleeding or excessive smoothness, while the structure of high-frequency elements, such as vegetation, hair, and lit writing, remains mostly unaltered.

Incorporate distortions not employed in training, such as spatially changing motion blur and simulated sensor noise with thick tails, by doing a more thorough robustness analysis. The suggested architecture is reasonably stable, despite the fact that all of the evaluated models have comparatively poor overall metric scores (such as PSNR). Subjective evaluation reveals that under extreme deterioration, apparent sharpness and natural color grading endure longer. The technique is less beneficial only in extremely rare situations, such as under-exposed photos with sensor readout abnormalities or optical flare, and observers see color inversion or patchy artifacting

https://doi.org/10.64972/jaat.2026v4.292p26:346-356

that is comparable to the other methods. Exposure- and sensor-adaptive regularization procedures still need to be developed because, as Figure 6 illustrates, these scenarios are not typical cross-domain findings and indicate the model's failure modes.

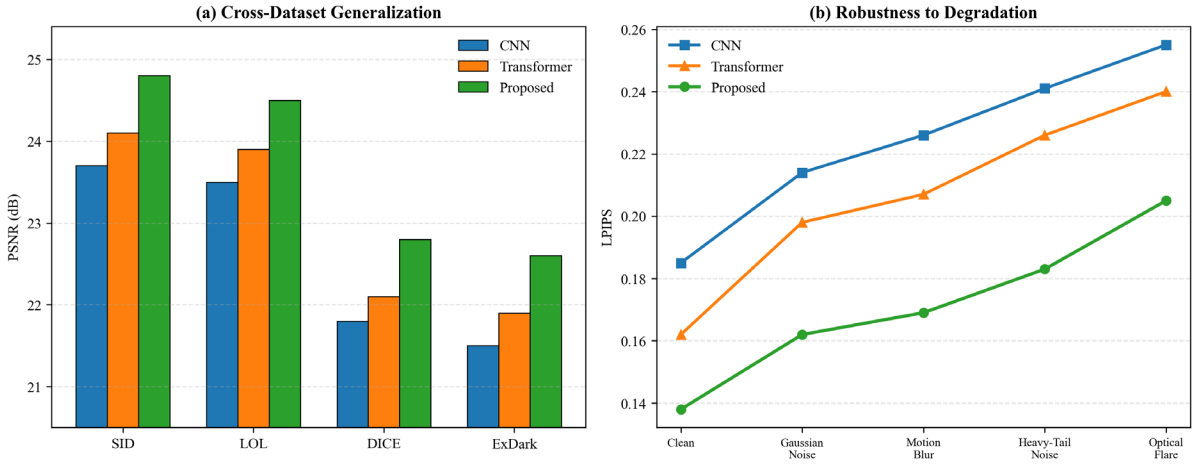

Figure 6. Cross-dataset and robustness results with typical failure analysis.

This extensive evaluation underlines that the illumination-adaptive network not only secures strong numbers in conventional benchmarks but also delivers stable and convincing visual quality in scenarios aligned with real-world demands. The results of the subjective studies and ablation experiments substantiate the necessity of each design choice, while domain and degradation analyses underline the framework's practical generalizability and point towards potential directions for further enhancement when confronting the most extreme imaging challenges.

Conclusion

In order to solve the persistent issues of detail recovery and perceived quality in low-light imaging, this research presents an illumination-adaptive super-resolution architecture. The proposed system achieves a tight connection of signal improvement, noise suppression, and perceptual quality by integrating scene-adaptive luminance estimation into a transformer backbone and introducing loss limitations motivated by human visual contrast sensitivity.

Experiments using a variety of publicly available datasets have demonstrated that this approach beats current CNN and Transformer models in terms of both number and quality. This method is also reasonably effective in extremely low light settings; it has low LPIPS and NIQE values along with high PSNR and SSIM. Additionally, a lot of people have mentioned that they can now see objects more clearly, which is more aesthetically pleasing and less taxing on their eyes. All three of the network's components have been demonstrated to be essential for its effective operation based on ablation tests.

The framework is particularly useful in practice since it has demonstrated good robustness and transferability in the face of external variations and non-ideal degradation. The system is appropriate for applications including industrial inspection, mobile photography, intelligent transportation, and surveillance since it can function normally under a variety of low-light conditions that were not used for training.

Along with the aforementioned outcomes, several new studies and uses will also be encouraged. A foundation for the future generation of image processing has been established through illumination-aware modeling and perceptual optimization to tackle the issues of challenging and multimodal surroundings, as well as devices with limited resources. To lessen the gap between computational and human visual performance, additional research will be done on unsupervised adaptation and more sophisticated perceptual loss functions. As a result, this paradigm provides strong support for expanding real-world applications of perceptually motivated low-light visual data augmentation.

Author Contributions

Mohammed Al-Maktoum contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Khalid Al-Najjar contributes to draft preparation, methodology, software, validation, analysis, investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Zhang, S., Zhang, Y., Jiang, Z., Zou, D., Ren, J., & Zhou, B. (2020, August). Learning to see in the dark with events. In *European Conference on Computer Vision* (pp. 666-682). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-58523-5_39

[2] Lv, X., Sun, Y., Zhang, J., Jiang, F., & Zhang, S. (2021). Low-light image enhancement via deep Retinex decomposition and bilateral learning. *Signal Processing: Image Communication*, *99*, 116466. https://doi.org/10.1016/j.image.2021.116466

[3] Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, *26*(7), 3142-3155. https://doi.org/10.1109/TIP.2017.2662206

[4] Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M. H., & Shao, L. (2022). Learning enriched features for fast image restoration and enhancement. *IEEE transactions on pattern analysis and machine intelligence*, *45*(2), 1934-1948. https://doi.org/10.1109/TPAMI.2022.3167175

[5] Wang, Z., Chen, J., & Hoi, S. C. (2020). Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *43*(10), 3365-3387. https://doi.org/10.1109/TPAMI.2020.2982166

[6] Conde, M. V., Choi, U. J., Burchi, M., & Timofte, R. (2022, October). Swin2sr: Swinv2 transformer for compressed image super-resolution and restoration. In *European conference on computer vision* (pp. 669-687). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-25063-7_42

[7] Wu, J., Zhan, D., & Jin, Z. (2024). Understanding and improving zero-reference deep curve estimation for low-light image enhancement: J. Wu et al. *Applied Intelligence*, *54*(9), 6846-6864. https://doi.org/10.1007/s10489-024-05534-7

[8] Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J. H., & Liao, Q. (2019). Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, *21*(12), 3106-3121. https://doi.org/10.1109/TMM.2019.2919431

[9] Wang, J., Xiang, L., Liu, L., Xu, J., Li, P., Xu, Q., & He, Z. (2024). Toward real-world remote sensing image super-resolution: A new benchmark and an efficient model. *IEEE Transactions on Geoscience and Remote Sensing*, *63*, 1-13. https://doi.org/10.1109/TGRS.2024.3516538

[10] Mathews, A., Moharana, S. K., Singh, P. N., & Pattnaik, S. (2025, December). Redefining Deep Learning Across NLP, Image Classification and Beyond-Vision Transformers. In *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICTBIG68706.2025.11323670

[11] Feng, X., Ji, H., Pei, W., Li, J., Lu, G., & Zhang, D. (2023). U²-Former: Nested U-shaped transformer for image restoration via multi-view contrastive learning. *IEEE Transactions on Circuits and Systems for Video Technology*, *34*(1), 168-181. https://doi.org/10.1109/TCSVT.2023.3286405

[12] Boudiaf, A., Guo, Y., Ghimire, A., Werghi, N., De Masi, G., Javed, S., & Dias, J. (2022, May). Underwater image enhancement using pre-trained transformer. In *International Conference on Image Analysis and*

Processing (pp. 480-488). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-06433-3_41

[13] Ou, F. Z., Wang, Y. G., Li, J., Zhu, G., & Kwong, S. (2021). A novel rank learning based no-reference image quality assessment method. *IEEE Transactions on Multimedia*, *24*, 4197-4211. https://doi.org/10.1109/TMM.2021.3114551

[14] Baghel, N., Dubey, S. R., & Singh, S. K. (2024, December). SRTransGAN: image super-resolution using transformer based generative adversarial network. In *International Conference on Computer Vision and Image Processing* (pp. 308-322). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-93694-4_22

[15] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., ... & Wang, Z. (2021). Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, *30*, 2340-2349. https://doi.org/10.1109/TIP.2021.3051462

[16] Nair, R. S., & Domnic, S. (2022). Deep-learning with context sensitive quantization and interpolation for underwater image compression and quality image restoration. *International Journal of Information Technology*, *14*(7), 3803-3814. https://doi.org/10.1007/s41870-022-01020-w

[17] Dang, X., Wang, G., Wu, X., & Qin, Z. (2024, July). Improving Image Reconstruction and Synthesis by Balancing the Optimization from Frequency Perspective. In *2024 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICME57554.2024.10688168

[18] Jain, V., Murray, J. F., Roth, F., Turaga, S., Zhigulin, V., Briggman, K. L., ... & Seung, H. S. (2007, October). Supervised learning of image restoration with convolutional networks. In *2007 IEEE 11th International Conference on Computer Vision* (pp. 1-8). IEEE. https://doi.org/10.1109/ICCV.2007.4408909

[19] Zheng, Q., Tu, Z., Zeng, X., Bovik, A. C., & Fan, Y. (2022). A completely blind video quality evaluator. *IEEE Signal Processing Letters*, *29*, 2228-2232. https://doi.org/10.1109/LSP.2022.3215311

[20] Li, X., Wang, Z., & Xu, B. (2024). Blind image quality assessment with semi-supervised learning. *Journal of Visual Communication and Image Representation*, *100*, 104100. https://doi.org/10.1016/j.jvcir.2024.104100

[21] Kannan, V., Malik, S., Babu, N. C., & Soundararajan, R. (2023). Quality Assessment of Low-Light Restored Images: A Subjective Study and an Unsupervised Model. *IEEE Access*, *11*, 68216-68230. https://doi.org/10.1109/ACCESS.2023.3292114

[22] Zhang, W., Xu, L., Wu, J., Huang, W., Shi, X., & Li, Y. (2025). Low-light image enhancement via illumination optimization and color correction. *Computers & graphics*, *126*, 104138. https://doi.org/10.1016/j.cag.2024.104138