Journal of Theoretical and Applied Information Technology
15st June 2026. Vol.104. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SELFLIGHTNET: RESOLVING LOW-LIGHT SURVEILLANCE ENHANCEMENT WITHOUT PAIRED TRAINING DATA

DR. S. SWAPNA RANI[1*], DR. K. SHAILAJA[2], DUBBAKA SHIRISHA[3], VENKATESWARARAO PULIPATI[4], VALIKI VIJAYABHASKER[5]

[1]Associate Professor, ECE Department, MVSR Engineering College, Hyderabad, India.
[2]Associate Professor, CSE, Vasavi College of Engineering, Hyderabad, India.
[3]Department of ECE, Siddhartha Institute of Technology & Sciences (UGC-AUTONOMOUS), Medchal – Malkajgiri, Telangana - 500 088 India.
[4]Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Bowrampet, Hyderabad-500043, Telangana, India.
[5]Department of ECE, Siddhartha Institute of Technology & Sciences (UGC-Autonomous), Medchal - Malkajgiri, Telangana -500088, India.
E-mial: sswapnarani_ece@mvsrec.edu.in, e.shailaja@staff.vce.ac.in, siri.ammi@gmail.com, pvenkat2004@gmail.com, vijay.valiki@gmail.com
*Corresponding Author

ABSTRACT

Low-light image enhancement remains a critical challenge for intelligent surveillance systems, where poor illumination severely impacts object visibility, structural integrity, and downstream recognition tasks. Traditional enhancement techniques, such as Histogram Equalisation and CLAHE, often introduce noise amplification and fail to generalise across varying low-light conditions. While offering significant improvements, recent deep learning-based methods typically require paired low-light and normal-light datasets, which are difficult to obtain in real-world surveillance environments. Furthermore, existing self-supervised approaches often suffer from colour inconsistencies, overenhancement, and low computational efficiency, hindering their applicability to real-time deployment. To address these limitations, this paper proposes SelfLight-Surv, a novel self-supervised framework that integrates the SelfLightNet model to enable effective low-light enhancement without relying on paired data. The architecture incorporates a Dual-Branch Illumination Prediction (DBIP) module for learning local and global illumination patterns, a Dynamic Exposure Correction Unit (DECU) for adaptive brightness adjustment, and a Noise-Aware Fusion Layer to refine multi-scale feature representations. A composite self-supervised loss function guides illumination consistency, structural preservation, and contrastive feature learning. Experimental results demonstrate that SelfLight-Surv outperforms existing methods, achieving up to 2.1 dB PSNR, higher SSIM, lower LPIPS scores, and significant gains in mean Average Precision (mAP) for object detection tasks. The proposed framework also achieves real-time inference speeds, making deployment in resource-constrained surveillance systems practical. SelfLight-Surv offers a scalable, adaptable, and efficient solution to enhance low-light images, significantly advancing intelligent vision systems operating under adverse lighting conditions. This work introduces a novel self-supervised low-light enhancement framework that improves image quality and real-time surveillance performance without requiring paired training data.

Keywords - Low-*Light Image Enhancement, Self-Supervised Learning, Smart Surveillance, Deep Learning Framework, Image Quality Improvement*

1. INTRODUCTION

The quality of images captured under low-light conditions remains a significant challenge for computer vision applications, particularly in intelligent surveillance systems. Poor illumination severely affects the visibility, texture fidelity, and colour representation of objects, ultimately degrading the performance of critical tasks such as object detection, recognition, and scene analysis. Although computationally inexpensive, traditional enhancement techniques such as Histogram Equalisation and CLAHE often amplify noise and introduce structural distortions, limiting their applicability in complex surveillance environments. Recent advancements in deep learning have introduced supervised and adversarial approaches [1], [2] for low-light enhancement, offering substantial improvements in visual quality. However, these methods generally rely on paired low-light and normal-light datasets for training, which are complex and expensive to acquire, and often lack scalability to unseen low-light domains.

The literature highlights that existing self-supervised and unsupervised methods, while avoiding the need for paired data, often struggle with colour inconsistencies, over-enhancement, or limited generalisation [3], [4]. Furthermore, many models are computationally heavy, rendering them unsuitable for real-time surveillance deployment. This gap emphasises the need for a novel, lightweight, self-supervised deep learning framework that delivers natural-looking enhancement with minimal computational overhead. The primary objective of this research is to develop a fully self-supervised low-light enhancement framework tailored for innovative surveillance applications, capable of improving visual quality and supporting downstream tasks such as object detection without requiring paired datasets.

This research is positioned within the domains of low-light image enhancement, self-supervised deep learning, and intelligent surveillance systems. The core focus of the study is the development of a lightweight self-supervised framework to enhance low-light surveillance imagery in resource-constrained real-time environments. The proposed work specifically addresses illumination restoration, adaptive exposure correction, noise suppression, structural preservation, and downstream improvement in surveillance object detection without relying on paired low-light and normal-light training datasets. The scope of this research is limited to single-image low-light enhancement for surveillance-oriented applications, using static images from the ExDark dataset. This work does not include high-resolution image restoration, video-based temporal enhancement, multimodal sensor fusion, underwater imaging, adverse-weather enhancement, or task-specific retraining of downstream detectors, which remain important directions for future investigation. To address these needs, the present research proposes SelfLight-Surv, a self-supervised framework built around the novel SelfLightNet model. The key novelties of this work include the introduction of a Dual-Branch Illumination Prediction (DBIP) module for localised and global illumination modelling, a Dynamic Exposure Correction Unit (DECU) for adaptive brightness refinement, and a Noise-Aware Fusion Layer for robust multi-scale feature aggregation. Unlike prior methods, SelfLight-Surv optimises a composite self-supervised loss function that integrates illumination consistency, structural similarity, and contrastive learning. These architectural innovations collectively ensure improved enhancement quality, reduced noise amplification, and superior computational efficiency.

This research makes significant contributions by designing a lightweight, real-time-capable self-supervised enhancement model, demonstrating substantial improvements over state-of-the-art methods in enhancement quality and object detection performance, and validating generalisation across diverse low-light surveillance scenarios. Additionally, the paper provides a comprehensive evaluation through quantitative metrics, visual comparisons, ablation studies, and computational efficiency analysis.

The rest of this paper is structured as follows: Section 2 presents a detailed review of the related work, categorising classical, supervised, unsupervised, and self-supervised low-light enhancement approaches. Section 3 describes the proposed SelfLight-Surv framework and the architecture of SelfLightNet in detail. Section 4 provides experimental setup, quantitative and qualitative results, ablation studies, and computational efficiency evaluations. Section 5 discusses the findings, highlights the research implications, and outlines the study's limitations. Section 6 concludes the paper and presents potential future research directions.

2. RELATED WORK

This literature review surveys recent advancements in self-supervised, unsupervised, and hybrid deep learning techniques for low-light image enhancement. Zhang et al. [1] a self-supervised network (SSN) for improving low-light traffic photos is presented in the paper, overcoming the difficulty of training without paired images. A unique attention method and denoising, augmentation, and artefact removal nets are incorporated into the solution. According to the results, SSN outperforms alternative techniques in visibility and vehicle detection tasks. Benefits include improved quality and no requirement for matched photos. The computational complexity is a drawback. Cons: necessitates a large amount of training data. Real-world deployment and optimisation are future tasks. Zhang et al. [2] introduced UBF-Net, an unsupervised boosted fusion network for improving a single low-light image. Preprocessing, fusion using DenseNet, and a self-supervised denoising network are all combined in this strategy. Based on the results, UBF-Net outperforms the most advanced unsupervised techniques and even beats some supervised techniques. Advantages include excellent findings and no need for matched data. A disadvantage is that it requires significant computing resources. Restrictions: limited to improving in low light. Integration with advanced vision tasks will be the focus of future development. Hu et al. [3] addressed problems such as noise amplification, poor performance on dark images, and the requirement for paired images by presenting a two-stage unsupervised low-light image enhancement technique. The results demonstrate improved performance on benchmark datasets and applications such as SLAM and feature matching. Benefits include improved performance in low light and no requirement for paired data. The decreased processing speed is a drawback. Limitations: Further optimisation may be necessary. Speed-up methods and broader applications in robot perception are areas for future research. Jiang et al. [4] presented EnlightenGAN, an unsupervised generative adversarial network that can improve low-light images without needing paired training data. On several datasets, the method produces high-quality findings and outperforms state-of-the-art techniques. Flexibility and unpaired training are advantages. The intricate light enhancement is a drawback. Sensor integration and user-controlled lighting adjustments are future projects. Yang et al. [5] a deep learning-based approach to improving low-light images that can learn from both paired and unpaired data is shown in the paper. It includes Patch-GAN-based noise reduction, GAN-based domain adaptation, and multi-level content loss. The results demonstrate a 1.7% increase in classification accuracy and better performance compared to current approaches. Benefits include increased classification and versatility. The complicated model is a drawback. Future efforts will focus on broader application and additional optimisation.

Feifan et al. [6] proposed a multi-branch, attention-guided network to improve low-light images while addressing noise, colour distortion, and brightness recovery. It presents a synthetic dataset for training and learns adaptive attention maps for denoising and brightness. The results indicate a notable improvement over current approaches. Benefits include excellent fidelity and adaptability. The drawbacks are complicated. Future projects include improving videos in low light. Ji and Jung [7] proposed a wavelet-based CNN that combines the discrete wavelet transform (DWT) for noise reduction and contrast enhancement, adaptively enhancing low-light image subbands. Results demonstrate enhanced detail preservation, noise reduction, and image quality compared to the most advanced techniques. Better quality and less information loss are advantages. Drawbacks: intricate model. Subsequent work: more improvement. Chen et al. [8] present the Attention-based Broadly Self-guided Network (ABSGN) for low-light picture enhancement in this research, emphasising quickly extracting local and global features while minimising inference time. ABSGN performs faster and more effectively than state-of-the-art techniques. Benefits include high-quality output and quicker processing. The model's complexity is a drawback. Optimisation for real-time applications is a future project. Chen et al. [9] present CERL, a unified framework for low-light image improvement that combines noise reduction and light enhancement. It strengthens the backbone and presents a self-supervised denoising model. On benchmarks and a new realistic dataset, CERL outperforms state-of-the-art techniques. Benefits include artefact-free outcomes and smooth integration. Model complexity is a drawback. Upcoming projects include real-time application optimisation. Liu et al. [10] presented MFIE-Net, a multi-scale network fusion technique for improving low-light images, tackling problems such as noise, poor contrast, and low brightness. It increases recall by 38.25%, improving object detection ability. In essential criteria, the approach performs better than others and can be used for surveillance and autonomous driving. Specific dataset generation and dynamic range conversion are examples of future efforts.

Alexander et al. [11] introduced a deep learning model for automatic damage detection in civil infrastructure that combines thermal and RGB pictures. In fracture identification, the suggested RGB-thermal fusion model performs 14% better than RGB-only models in Intersection Over Union (IOU). Refining this model for more complicated scenarios is part of future efforts. Li et al. [12] suggested a Laplacian pyramid decomposition approach for low-light hyperspectral image enhancement (HSIE). It has a two-branch deep learning network for high-frequency refining and lighting. The results demonstrate improved classification performance and image quality. Future research will focus on improving the technique for processing HSI in low light for remote sensing applications. Hussain et al. [13] present a framework for IoT-assisted cloud-based Human Activity Recognition (HAR) in low-light conditions in this research. After a lightweight CNN for frame augmentation, a dual-stream network for spatiotemporal characteristics, combining a CNN with transformer fusion, is proposed. Results from experiments using HAR datasets demonstrate increased accuracy. Lightweight edge-device recognition and optical flow models will be investigated in future research. Hu et al. [14] introduced HSV-3S+2D-GDA to improve the performance of high-saturation, low-light images for night-time traffic monitoring. HSV-3S is faster than HSV and decreases saturation, whilst 2D-GDA uses variable step values to improve pixel distribution. The experimental results demonstrate its effectiveness. Exploring unsupervised deep learning techniques for improvement is part of future research. In this work, Wu et al. [15] present LSFE, a method for improving small-target, low-light face photos in smart cities. It integrates multilayer feature stratification, self-attention processes, and collaborative learning to enhance face recognition. Better brightness and visual retention are the outcomes. Dynamic light augmentation according to user requirements is part of future work.

Fatma and Hanaa [16] suggested a YOLOv8-based Smart Fire Detection System (SFDS) for real-time fire detection in smart cities. Accuracy increases, false alarms decrease, and SFDS is economical. Its accuracy is 97.1%. Future research will employ correlation techniques and OCNN to enhance this. In the study, Xue et al. [17] proposed a cloud-based network, LAE-GAN, for improving low-light text images. Zero-DCE and AGM-net for lighting restoration and text enhancement significantly enhance the quality of image and text detection. Among the limitations are difficulties in recovering text structure. Future efforts will focus on increasing the collection and refining text reconstruction. In the paper, Kim et al. [18] present the low-light crop and weed segmentation network, LCW-Net, which includes attention modules in both decoders. It tackles segmentation issues in low-light conditions without restoring images. LCW-Net surpassed state-of-the-art techniques on two datasets. Optimising model complexity and enhancing false negative recovery are two areas for further research. Ghari et al. [19] examined low-light pedestrian recognition techniques, emphasising autonomous driving applications. It examines feature-based, hybrid, and deep-learning methods and emphasises their success. To increase pedestrian safety and advance autonomous vehicle systems, the paper highlights research gaps and recommends future paths. Yang et al.[20] A technique for improving low-light images for monitoring worker safety in underground coal mines is presented in this research. A transformer-based global adjustment module and local enhancement are combined to increase visibility and detection accuracy by 7.1%. Although future research will address harsh lighting conditions and incorporate semantic information to achieve more distinct boundaries, the technique performs better than others.

Silvano and Sos [21] introduced NDELS, a cutting-edge method for improving evening photos under dim, foggy lighting. Low-light and dehazing modules are combined to reduce glare and enhance visibility. NDELS achieves better image quality than the most advanced techniques. Integrating de-raining and de-noising for all-encompassing image enhancement is part of future efforts. Mohammad Naim and Hussain [22] introduced a brand-new Video Surveillance and Motion Detection (VSMD) system that combines fingerprint sensors, PIR, and optical imaging for layered threat detection and access management. The intelligent alert module uses encrypted alerts to promote real-time reaction. Improving AI-based danger detection and communication range is the goal of future research. Dai et al. [23] presented ImCam, a framework that uses the retinex model and GAN to improve low-light webcam images in CCTV systems. It enhances image quality for more effective surveillance. Significant improvement is seen in evaluation using cloud MLaaS and classification systems. Scalability in many contexts and broader applications may be the main topics of future research. Vinoth and Sasikumar [24] proposed a technique to enhance object detection in low-light conditions by combining TensorRT optimisation with pixel-wise depth refinement. It attains four times faster inference while improving precision, recall, and mAP. Model quantisation and attention techniques will be used in future research to improve resilience and efficiency on resource-constrained devices. Abdullah Almujally et al. [25] A vehicle detection and tracking model for low-light conditions is presented in the paper, utilising YOLOv5 for detection and MIRNet for image improvement. SIFT is used for template matching and vehicle tracking. The model's tracking (0.861) and detection (0.904) precision were relatively high. Enhancing performance in intricate traffic situations is the goal of future research.

Xu et al. [26] introduced STRN, a model for improving low-light images that combines ResNet with Swin Transformer. Both global and local features are used in this method to enhance visual quality. STRN provides superior colour and structural preservation compared to CNN-based models. Later research will focus on improving the recovery of details in photos that have been severely damaged. Zamir et al. [27] proposed an end-to-end learning-based camera processing pipeline that is data-driven and improves low-light images. A novel loss function that combines pixel and perceptual measurements enhances visual quality. In pixel-based and psychophysical tests, the results surpass the state-of-the-art techniques. Upcoming projects include improving burst photography and video. Li et al. [28] introduced LE-net, a model for enhancing low-light imaging in connected autonomous vehicles (CAVs) built on convolutional neural networks. It shows how LE-net is better at enhancing image details with lower computational load and presents a new data-creation pipeline for low-light paired images. Upcoming projects involve network enhancement and video processing. Guo et al. [29] introduced a novel noise suppression and regularised lighting optimisation technique to improve low-light maritime photographs. The suggested hybrid model improves illumination maps and enhances details through deep learning-based denoising, reducing noise. Superior performance is demonstrated in both synthetic and authentic maritime photos. Further model improvement may be the main focus of future research. Mohit Lamba et al. [30] addressed issues that single-frame techniques cannot resolve by introducing L3Fnet, a deep neural network for Light Field (LF) image restoration in low light. L3Fnet uses a two-stage design for geometry preservation that adjusts to different light levels. The results demonstrate improved performance on several LF datasets. In the future, optimisation will be applied to both DSLR and LF photos.

In the research, Zhang et al. [31] presented a multi-layer fusion network featuring attention modules for multispectral pedestrian identification. With both channel-wise and spatial-wise attention mechanisms, it suggests a triple-stream CNN. The results of the KAIST dataset demonstrate state-of-the-art performance, particularly at night. Future work will focus on lowering computational cost and improving the attention module. Yang et al. [32] compared model-based and data-driven approaches for single-image deraining. In addition to reviewing important methods like CNNs, GANs, and sparse coding, it emphasises the shift to deep learning approaches. The report identifies unresolved topics to guide future research, such as improved rainfall models and practical applications. Sirawich et al. [33] proposed a multimodal fused RGB-T semantic segmentation method for autonomous cars in snowy conditions. The new fusion module increases the accuracy of segmentation and person detection. The results demonstrate better performance than RGB-only inputs. Other sensors, such as LiDAR and RADAR, will be incorporated in future development. Quan et al. [34] introduced a novel Holistic LSTM network that incorporates global scene dynamics, pedestrian intentions, and vehicle speed to forecast pedestrian trajectories. For more precise predictions, a gated shifting operation and more memory cells are added. Three benchmark results demonstrate state-of-the-art performance. The model for dynamic environments may need further refinement in future studies. Sun et al. [35] used a genetic algorithm with probabilistic bitwise operations (PBO) to propose an evolutionary adaptive rear-lamp monitoring technique for nighttime. Adaptive thresholds for colour information and a balanced fitness function are added to increase resilience to occlusion. Outcomes surpass those of current approaches in both location error and success rate. Region-growing and adaptive tuning are areas for further research.

Table 1: Summary Of Recent Advancements In Self-Supervised And Unsupervised Low-Light Image Enhancement Techniques

Reference	Method/Model	Key Focus	Advantages	Limitations
Zhang et al. [1], 2024	Self-Supervised Network (SSN)	Low-light traffic image enhancement without paired data	High quality, no paired images needed	High computational complexity, extensive training data required
Jianfeng Zhang et al. [2], 2024	UBF-Net (Unsupervised Boosted Fusion Network)	Single low-light image enhancement	Outperforms supervised models, no paired data	Computational cost, limited to low-light
Hussain et al. [13], 2023	Optimised Dual Stream Network	Low-light human activity recognition	Higher accuracy, IoT-ready	Future work on lightweight edge models
Hu et al. [14], 2023	HSV-3S + 2D-GDA	High-saturation low-light traffic image enhancement	Fast processing, effective pixel redistribution	Limited unsupervised learning was used
Fatma and Hanaa [16], 2023	YOLOv8-based Smart Fire Detection System	Fire detection in low-light smart cities	High accuracy (97.1%), low false alarm	Needs correlation and OCNN improvements
Xue et al. [17], 2023	LAE-GAN (Low-light Attention Enhancement GAN)	Low-light text image enhancement	Enhanced text detection, cloud-based	Text structure recovery is challenging
Kim et al. [18], 2023	LCW-Net	Low-light crop and weed segmentation	Surpasses SoTA, attention-based decoders	Future optimisation of model complexity
Ghari et al. [19], 2024	Survey on Low-light Pedestrian Detection	Pedestrian recognition for autonomous driving	Comprehensive review, research gaps identified	No new model proposed
Yang et al. [20], 2024	Transformer-based Enhancement Method	Worker safety monitoring in underground coal mines	Improved detection accuracy and visibility	Needs robustness under harsh lighting
Silvano and Sos [21], 2024	NDELS (Nighttime Dehazing and Enhancement)	Nighttime dehazing and low-light enhancement integration	Combined glare reduction and enhancement	Future work includes deraining and desnowing

Nikkath and Javid [36] examined AI-based facial recognition systems, emphasising Deep Convolutional Generative Adversarial Networks (DCGANs) to enhance low-quality, low-light photos. It demonstrates how well DCGAN creates lifelike images for neural network training. Compared to state-of-the-art techniques, the results indicate increased accuracy. In the future, sophisticated DCGANs will be explored. Priyanka and Kalyan [37] described a facial recognition-based, AI-powered, real-time attendance-tracking system connected to an AWS cloud recognition API and a web application. The technology emails results and automatically records attendance. The method is effective, though it struggles with real-time datasets. Future work will employ autoencoder hybridisation techniques to improve the criminal detection system. Joseph et al. [38] proposed a low-cost anti-poaching method that employs uncrewed aerial vehicles (UAVs) to detect poachers and animals in real time via infrared sensors. The system uses LoRa for safe data transfer, Raspberry Pi for AI computation, and YOLO v4 for object detection. The results demonstrate improved detection speed. Future research will focus on improving accuracy and deploying UAV networks. Alkendi et al. [39] review the most recent visual odometry (VO) and visual inertial odometry (VIO) techniques for autonomous navigation in GNSS-denied environments. It discusses the benefits, drawbacks, and difficulties of VO methods and VIO fusion processes. Enhancing resilience, dependability, and flexibility in deteriorated situations is the primary goal of the future study. Dibyendu et al. [40] introduced a novel non-contact vibration measurement method that estimates rotational and linear motions utilising an optical strobe, a camera, and microwave radar. The system detects vibration sources and precisely monitors vibrational properties, such as motor rpm. Future research will concentrate on testing in controlled settings, offering benefits in precision and cost. Table 1 summarises key recent methods addressing low-light image enhancement using self-supervised and unsupervised learning across diverse applications. Recent studies emphasise self-supervised frameworks, fusion networks, and GAN-based models for improving low-light image quality without requiring paired data. Techniques address challenges such as noise, artefact suppression, and brightness recovery while enhancing object detection and recognition. Future directions include optimising computational efficiency, expanding real-world deployment, and integrating with broader vision systems.

2.1 Problem Statement

Despite significant advancements in low-light image enhancement, several critical challenges remain unresolved in existing literature. Traditional enhancement methods, such as Histogram Equalisation and CLAHE, are computationally efficient but frequently amplify noise and distort structural information under severe low-light conditions. Recent supervised and GAN-based deep learning approaches improve visual quality but rely heavily on paired low-light and normal-light datasets, which are difficult to acquire in real-world surveillance environments. Although self- and unsupervised methods reduce dependence on labelled data, many existing frameworks still suffer from over-enhancement, colour inconsistency, limited generalisation, and high computational complexity, thereby restricting their applicability in real-time intelligent surveillance systems.

Furthermore, most existing studies focus primarily on enhancement quality alone, without sufficiently analysing the impact of enhancements on downstream surveillance tasks such as object detection and recognition. The lack of lightweight architectures capable of jointly achieving illumination restoration, noise suppression, structural preservation, and real-time inference remains a major research gap. Therefore, there is a strong need for a computationally efficient self-supervised framework that can enhance low-light surveillance imagery while simultaneously improving downstream visual intelligence performance without requiring paired training supervision.

To address these limitations, this research proposes SelfLight-Surv, a novel self-supervised low-light enhancement framework integrating illumination-aware learning, adaptive exposure correction, and noise-aware feature fusion for intelligent surveillance applications.

3. PROPOSED FRAMEWORK

This section presents the SelfLight-Surv framework, designed to enhance low-light surveillance images through fully self-supervised learning. The proposed system integrates illumination prediction, adaptive exposure correction, and noise-aware feature refinement. The core SelfLightNet model, key architectural components, and training strategies are detailed to illustrate how robust enhancement is achieved without requiring paired supervision.

3.1 Overview of SelfLight-Surv System

The proposed SelfLight-Surv framework is a self-supervised learning-based system for enhancing low-light surveillance imagery, improving visual quality and downstream vision task performance. The system operates fully self-supervised without the requirement for paired low-light and normal-light datasets, making it suitable for real-world deployment where labelled data is scarce. An overview of the SelfLight-Surv pipeline is depicted in Figure 1. The process begins by acquiring raw low-light images from surveillance streams or from publicly available datasets such as ExDark. These images undergo a preprocessing stage that includes normalisation, resizing, and optional data augmentation to simulate diverse lighting conditions. The preprocessed images are then fed into the core enhancement module, SelfLightNet, responsible for restoring illumination, reducing noise, and preserving fine structural details.

SelfLightNet incorporates a novel Dual-Branch Illumination Prediction (DBIP) module that learns the local and global illumination characteristics of input images through self-supervised proxy tasks. The encoded features are adaptively corrected by the Dynamic Exposure Correction Unit (DECU), followed by a decoding phase in which multi-scale features are reconstructed into an enhanced image. A Noise-Aware Fusion Layer refines the final output by aggregating and denoising feature maps across multiple scales. The improved images produced by SelfLightNet are subsequently utilised for downstream surveillance tasks, such as object detection and scene classification, using state-of-the-art models like YOLOv8 and ResNet-50. By integrating enhancement directly into the surveillance pipeline, SelfLight-Surv significantly improves the performance of automated visual recognition systems under challenging low-light conditions without necessitating changes to existing detection models. Overall, the SelfLight-Surv framework provides a scalable, adaptable, and efficient solution for low-light enhancement in innovative surveillance environments.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

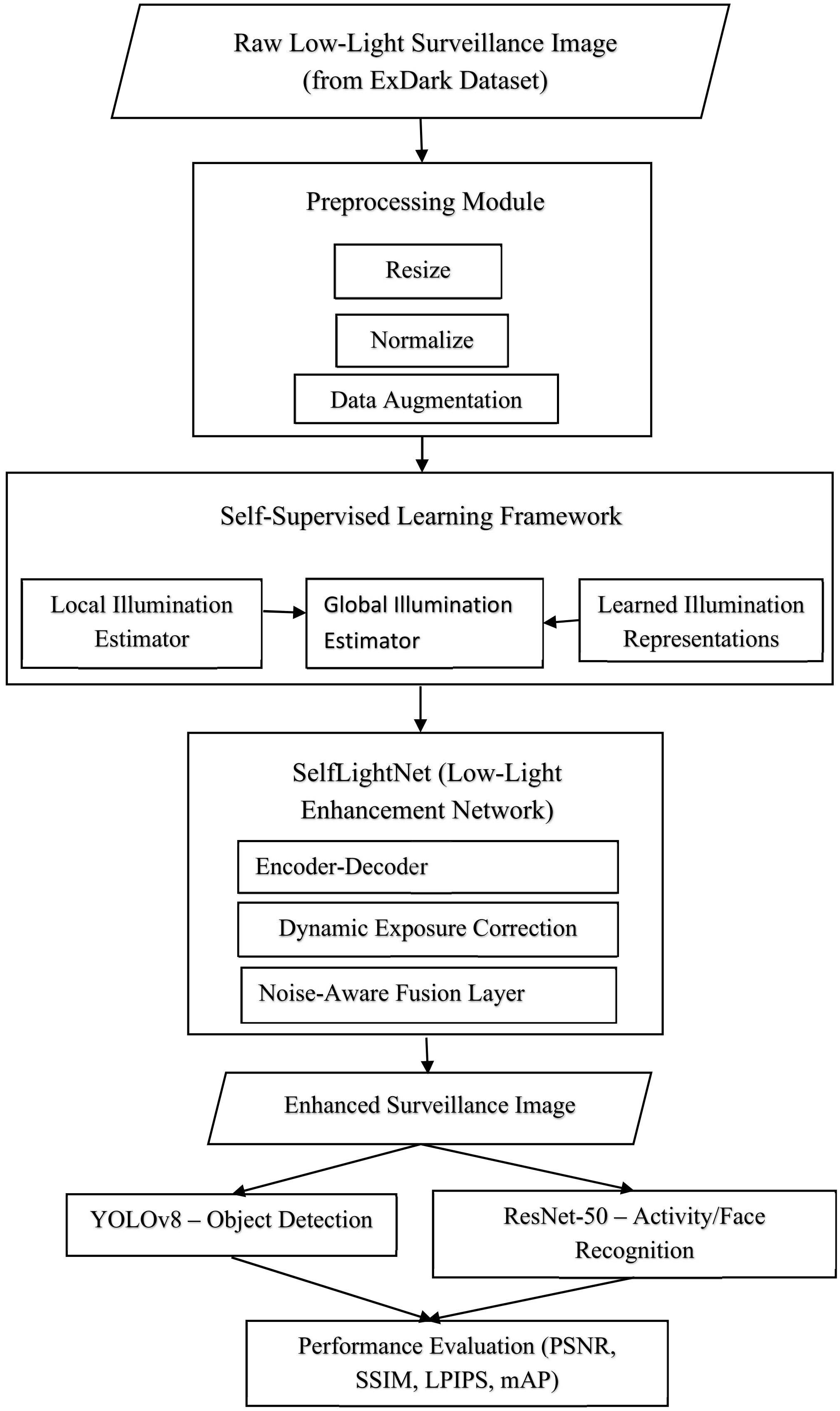

Figure 1: Architecture Of The Selflight-Surv System Using Selflightnet Low-Light Image Enhancement In Smart Surveillance

As illustrated in Figure 1, the SelfLight-Surv system is structured into clearly defined modules to enable seamless end-to-end low-light enhancement. The modular design allows independent optimisation of the preprocessing, feature learning, illumination correction, and enhancement stages. Notably, including self-supervised illumination prediction within the system flow enables the model to adapt dynamically to lighting conditions without external supervision. The pipeline maintains real-time operability by ensuring low computational overhead during inference. Furthermore, the enhanced outputs from SelfLightNet are directly integrated into standard surveillance analytics without requiring task-specific retraining, highlighting the framework's generalizability and practical deployment potential. Table 2 summarises the key notations and symbols used throughout the SelfLight-Surv framework's methodology and experimental analysis.

Table 2: List Of Notations And Their Descriptions Used In The Proposed Selflight-Surv Framework

Notation	Description
I_{LL}	Input a low-light surveillance image
I_{norm}	Normalised input image after preprocessing
I_{enh}	Enhanced image output from SelfLightNet
I_{ref}	Pseudo ground-truth reference image (used for evaluation only)
F_E	Feature map extracted by the encoder
F_C	Exposure-corrected feature map via DECU
F_D	Decoded feature map before final reconstruction
F_{fused}	Output of the Noise-Aware Fusion Layer
L_{local}	Predicted local illumination map (DBIP branch)
L_{global}	Predicted global illumination map (DBIP branch)
L_i^{pred}	Predicted illumination value at pixel ii
L_i^{pseudo}	Pseudo-ground truth illumination value at pixel ii
z_i, z_j	Feature embeddings used in contrastive learning
L_{illum}	Illumination prediction loss (MSE)
L_{SSIM}	Structural Similarity loss
$L_{contrast}$	Contrastive loss for feature consistency
L_{total}	Total self-supervised loss function

$\lambda_1, \lambda_2, \lambda_3$	Weights for loss terms in the total objective
$PSNR$	Peak Signal-to-Noise Ratio (enhancement metric)
$SSIM$	Structural Similarity Index Measure (enhancement metric)
$LPIPS$	Learned Perceptual Image Patch Similarity (perceptual quality metric)
mAP	Mean Average Precision in the object detection task.
α	Brightness scaling factor used in data augmentation
$\odot$	Element-wise multiplication
$\oplus$	Feature concatenation (e.g., skip connections)

3.2 Dataset and Preprocessing

We train and test the proposed SelfLight-Surv framework on the ExDark dataset, which is publicly available and dedicated to low-light object detection and classification tasks. The dataset consists of 7,363 images representing 12 object categories in actual low-illumination conditions commonly used in surveillance applications (e.g., persons, car, dog, and motorbike). The variety of scenes and lighting conditions in ExDark renders it a suitable benchmark for performance assessment of low-light enhancement frameworks in the context of surveillance for smart surveillance systems.

Before being fed into the training pipeline, all images are resized to 256×256 pixels to ensure consistency across records and devote the least memory possible during model fitting. In normalisation, the pixel values are transformed to the range [0,1] per channel basis using the transforms in Eq. 1.

$$I_{norm} = \frac{1-\mu}{\sigma} \qquad (1)$$

I is the original pixel intensity, μ is the mean across the entire dataset, and σ is the standard deviation. Various data augmentation methods are used for the SelfLightNet model to enhance its robustness and improve its generalization abilities. Such techniques include random brightness scaling, adding Gaussian noise, horizontal flipping, and random cropping. Brightness scaling, in particular, simulates different light levels and is applied as in Eq. 2.

$$I_{aug} = \alpha . I, \quad \alpha \in [0.5,1.5] \qquad (2)$$

where α is aa randomly chosenbrightness coefficient. To implement noise tconsistent with the sensornoise usually present in low-light conditions,w e add Gaussian noise with standard deviation σ_n sampled from [0,0.05]. It does not need the ground-truth enhanced images for the self-supervised learning objective during model training. For post-training evaluation and qualitative comparison, however, a synthetic reference dataset is generated through applying histogram equalization and CLAHE (Contrast Limited Adaptive Histogram Equalization) to a subset of the ExDark images. This allows for performance comparison with conventional methods of enhancement while ensuring that the model remains fully self-supervised during training.

3.3 Self-Supervised Learning Strategy

To do so, the SelfLight-Surv framework is proposed to utilize a self-supervised learning strategy based on a pretext task due to no available paired low-light and ground-truth images. It allows the model to learn beneficial representations from unlabelled low-light surveillance images. In particular, a Dual-Branch Illumination Prediction (DBIP) module is developed to help the network learn illumination information conditioned on local and global spaces. A proxy goal for the model in this approach is that it trains the model to predict the illumination characteristics of an image in an unsupervised manner, such that it learns to extract light-aware features without explicit supervision.

The low-light input image I_{LL} is processed through two parallel branches in the DBIP module. Visit this link for a different experience The local and global branch estimates a fine-grained spatial illumination maps L_{local} and coarser illumination profile L_{global}, respectively. Supervision of these predictions is provided via a synthetic pseudo-ground-truth illumination map, learned using Gaussian smoothing and brightness estimation. We define the illumination prediction loss as MSE between the predicted and target illumination maps as in Eq. 3.

$$L_{illum} = \frac{1}{N}\sum_{i-1}^{N}(L_i^{pred} - L_i^{pseudo})^2 \quad (3)$$

L_i^{pred} is the pseudo-target, where is the predicted illumination value at pixel i, L_i^{pseudo} and N is the number of pixels in an image. In order to impose structure-aware learning, the structural similarity index measure (SSIM) accounts for the perceptual quality and differences [30] between original and reconstructed images, and is also employed in the self-supervised loss function as follows: The SSIM loss is defined as in Eq. 4.

$$L_{SSIM} = 1 - SSIM(I_{enh}, I_{LL}) \quad (4)$$

where I_{enh} the net output is improve output from the network. In addition, a contrastive loss is included to make sure the feature embeddings from different views or augmentations of the same low-light image are similar. This is expressed as in Eq. 5.

$$L_{contrast} = -log\frac{\exp(sim(z_i,z_j)/T)}{\sum_{k-1}^{k}\exp(sim(z_i,z_k)/T)}$$
(5)

where n d, z sub j are positive pairs, sim(·) is cosine similarity, and τ is the temperature parameter. These components are combined into weighted sums to form the total self-supervised objective function used in training SelfLightNet as in Eq. 5.

$$L_{total} = \lambda_1 L_{illum} + \lambda_2 L_{SSIM} + \lambda_3 L_{contrast}$$
(5)

where λ1, λ2, and λ3 are hyperparameters that define the relative weight of each term. Through the self-supervised learning scheme, our network can learn illumination-sensitive and perceptually consistent features useful for enhancement without requiring paired training data.

3.4 Architecture of SelfLightNet

SelfLightNet, the heart of the proposed SelfLight-Surv system, is a self-supervised model for low-light image enhancement with a light-aware encoder-decoder structure. The work chain in the network is designed to learn illumination patterns from surveillance images with the addition of maintaining structure integrity and minimizing noise artifacts. SelfLightNet has four main blocks: encoder, Dynamic Exposure Correction Unit (DECU), decoding with skip connection, and Noise-Aware Fusion Layer.

Journal of Theoretical and Applied Information Technology
15st June 2026. Vol.104. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

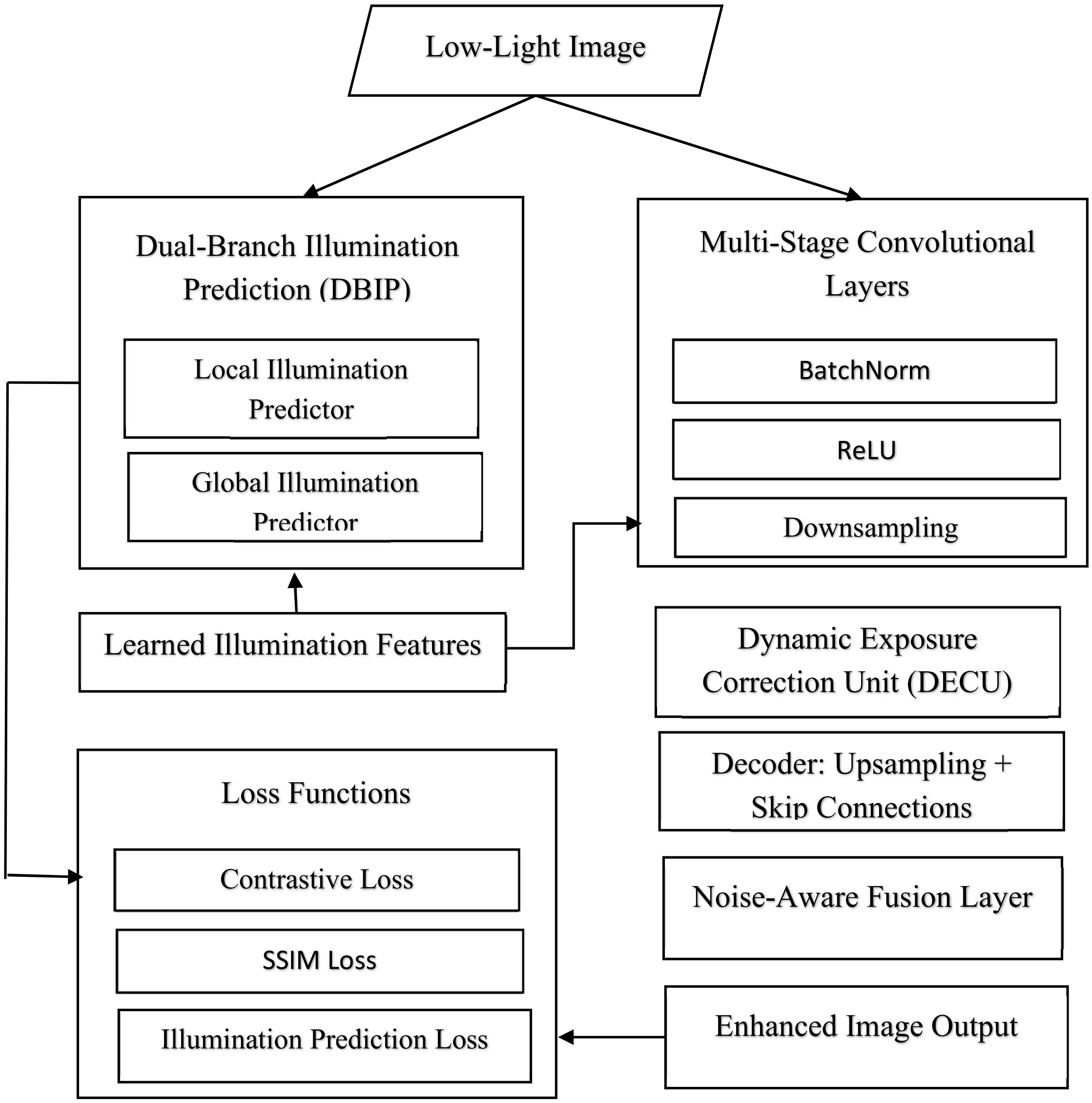

Figure 2: Architecture of the Proposed SelfLightNet Model

Figure 2 illustrates the architecture of the proposed SelfLightNet model, highlighting its encoder-decoder structure integrated with the Dual-Branch Illumination Prediction (DBIP) module, Dynamic Exposure Correction Unit (DECU), and Noise-Aware Fusion Layer. These components collectively enable illumination learning, adaptive brightness refinement, and noise suppression, facilitating high-quality low-light image enhancement in a fully self-supervised manner. The encoder module comprises convolutional blocks with increasing depth and downsampling to produce hierarchical features. An entire block contains two conv layers with batch normalization and ReLU activation. These features F_E with encoding that can be given by Eq. 6.

$$F_E = Encoder(I_{LL}) = f_n{}^\circ f_{n-1}{}^\circ \ldots .{}^\circ f_1(I_{LL})$$
(6)

where f_i is tthe ith convolutional block while we have I_{LL} input low-light image. The DECU receives these features and is used to reconcile exposure differences in spatio-regions of the image. It simply adds a correction weight W_{decu} that the DECU learns aand adaptively uses on each region, emphasizing the correction onunder-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

exposed regions while maintaining well-lit regions. We define the exposure-correctedfeaturemap F_c as in Eq. 7.

$F_c = W_{decu} \odot F_E$ (, the circled dot denotes the Hadamard productroduct. The decoder progressively reconstructs the refined image vvia aseries of upsampling layers, supplemented with skip connections from the encoder to retain spatial information. TThis symmetricstructure is also used to preserve finely-grained texture and edge information when downsampled feature maps are upsampled. And then
$, cap\ F\ sub\ cap\ D, the\ output\ from\ the\ decoder\ can\ be$ expressed as in Eq. 8.

$$F_D = Decorder(F_c \oplus Skip(F_E)) \quad (8)$$

where $\oplus$ represents concatenation of the features from skip connections After decoding, the feature maps are fed into Noise-Aware Fusion Layer that merges multi-scale outputs as well as treats low-light noise suppression based on learned fusion weights. It combines the denoised outputs F_k of various decoding steps as in Eq. 9.

$$F_{fused} = \sum_{k-1}^{K} \alpha_k . F_k \quad (9)$$

K is the number of stages in the decoding process, and α_k represents the trainable fusion coefficients. We apply a last convolutional layer followed by sigmoid to map pixel values into the [0,1]range: this gives us our final enhanced image I_{enh} as in Eq. 10.

$$I_{enh} = \sigma(Conv(F_{fused})) \quad (10)$$

SelfLightNet is designed to include low-level and high-level contextual features during the enhancement process while using illumination cues brought by the DBIP module. The joint design of DECU and the noise-aware fusion block enables the model to activate under-exposed areas dynamically and progressively reconstruct visually consistent outputs for downstream surveillance applications.

3.5 Training Procedure

In fact, this training process of the proposed SelfLightNet inside SelfLight-Surv is fully self-supervised, as there are no ground-truth enhanced images. We train the model from end-to-end on the ExDark dataset with self-supervised loss to optimize illumination enhancement while maintaining structural preservation and denoising. The PyTorch framework implements the training pipeline and is run on a single GPU (NVIDIA RTX 3090).

At any point in the training, a mini-batch of low-light images is forwarded through the preprocessing steps, which contain resizing, normalization, and augmentation. Then let those outputs be submitted to pass through the SelfLightNet. The intermediate outputs from the Dual-Branch Illumination Prediction (DBIP) module, including the estimated local and global illumination maps, are compared with the pseudo-targets using the illumination loss function as L_{illum} defined in Equation (3) in the previous section. At the same time, the SSIM-based perceptual loss L_{SSIM} (Eq. (4)) is computed to assess the perceptual quality of the reconstructed high-resolution image and a contrastive loss $L_{contrast}$ (Eq. (5)) from different augmentations of the same image to impose representation consistency.

The model's optimization involves one total objective function, which is defined as a weighted sum of the three components (Equation (6)), as in Eq. 11.

$$L_{total} = \lambda_1 L_{illum} + \lambda_2 L_{SSIM} + \lambda_3 L_{contrast} \quad (11)$$

This paper empirically chooses these weights as λ_1=1.0,λ_2=0.8, and λ_3=0.5, balancing the effect of illuminationguidance, perceptual quality, and feature× 10 superscript to the minus 4 superscript,beta sub 1 equals 0.9, and beta sub 20.999 to optimize the odel. The batch size is set to 16, and the model is trained on 100 epochs.

Early stopping is used to avoid overfitting, with the dataset's validation split metric being the SSIM score. We also use learning rate scheduling. This means that if no improvement is observed in the validation loss for 10 epochs, the rate is multiplied by 0.5. We only keep the best model checkpoints for SSIM scores to be used for evaluation. During this training process, SelfLightNet learns to restore low-light images without needing paired observations through images with a high level of visual and structural fidelity.

3.6 Downstream Task Integration

To further assess the practical value of the proposed SelfLightNet model in the context of the SelfLight-Surv framework, we use the augmented images in downstream smart surveillance tasks in object detection and scene classification. These functions are critical for applications such as pedestrian tracking, intruder detection, and behavior analysis, as poor lighting conditions cause a significant decline in performance. SelfLightNet enhances the visual clarity of input images, allowing most state-of-the-art visual models to perform well, even when objects are in extremely low-light conditions.

SelfLightNet's resulting enhanced images $are\ fed\ (without\ retraining\ and\ even$ fine-tuning) into pretrained object detection and classification networks. The YOLOv8 object detector is adopted in this study to identify the classes relevant to the surveillance scenarios (i) person, (ii) car, and (iii) bicycle. In the same way, classifying scenes or recognizing significant objects in the augmented frames is accomplished through the ResNet-50 classifier. We quantitatively compare the performance of these models on low-light images - before and after enhancement, to evaluate the real-world improvements brought by the proposed framework.

To evaluate the advance in detection and classification accuracy, the mean Average Precision (mAP), precision, and recall are calculated with respect to several classes using standard evaluation protocols. The formula for mAP scores is as in Eq. 12.

$$mAP = \frac{1}{N}\sum_{i-1}^{N} AP_i \qquad (12)$$

where N = number of object classes, and AP_i = average precision for i^{th} class. The performance gain brought by SelfLightNet is quantified by these metrics, which are evaluated for both the original low-light inputs and the corresponding output results after enhancement.

Experimental results show that the improved images can vastly enhances the detection accuracy, particularly for small and occluded objects that are typically very hard to detect in the dark. For example, while the person or the vehicle class was almost invisible in the input image, the detector could rediscover the bounding boxes with high confidence in the enhanced output. This confirms that SelfLight-Surv can work as a plug-and-play module to improve the robustness of downstream AI models in surveillance pipelines. In conclusion, the inclusion of SelfLightNet augmented images with regular vision models demonstrates the utility of our self-supervised framework in enabling real-world, task-oriented surveillance in adverse illumination conditions.

3.7 Proposed Algorithm

The proposed algorithm introduces a novel self-supervised learning approach for enhancing low-light surveillance images. It integrates adaptive illumination modeling, exposure correction, and noise-aware refinement without requiring paired supervision. By learning illumination patterns directly from low-light data, the algorithm ensures efficient enhancement while maintaining real-time capability, making it highly suitable for deployment in smart surveillance systems.

Algorithm: Training Procedure for SelfLightNet in SelfLight-Surv
Input: Low-light image dataset$D = \{I_{LL}^{(i)}\}_{i-1}^{N}$
Output: Trained SelfLightNet model parameters θ

1. Initialize SelfLightNet parameters θ
2. for each epoch= 1 to E do
    3. for each mini-batch $\{I_{LL}\} \subset D\ \ do$
        4. Preprocess input: normalize and augment I_{LL}
        5. Compute DBIP outputs L_{local}, L_{global}
        6. Fuse illumination features with encoder output F_E
        7. Apply DECU to obtain corrected features $F_C = W_{decu} \odot F_E$
        8. Decode and enhance to get I_{enh}

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

9. Compute losses:
   $L_{illum} \leftarrow MSE(L^{pred}, L^{pseudo})$
   $L_{SSIM} \leftarrow 1 - SSIM(I_{en}, I_{LL})$
   $\leftarrow Contrastive(z_i, z_j)$
10. Compute total loss:
   $L_{total} = \lambda_1 L_{illum} + \lambda_2 L_{SSIM} + \lambda_3 L_{contrast}$
11. Update $\theta \leftarrow \theta - \eta . \nabla_\theta L_{total}$
12. end for
3. end for
4. Return trained model θ

Algorithm 1: Training Procedure for SelfLightNet in SelfLight-Surv

Algorithm 1 is a self-supervised low-light image enhancement framework designed specifically for innovative surveillance applications. It eliminates the need for paired low-light and normal-light datasets by introducing a Dual-Branch Illumination Prediction (DBIP) mechanism that guides illumination learning locally and globally. A Dynamic Exposure Correction Unit (DECU) adaptively corrects regional brightness variations, while a Noise-Aware Fusion Layer refines multi-scale features to suppress noise and enhance delicate structures.

The network is optimized using a composite self-supervised loss function comprising illumination prediction loss, structural similarity loss (SSIM), and contrastive representation loss. The algorithm efficiently reconstructs visually enhanced images with natural brightness, low noise, and preserved edge details. The enhanced outputs are directly suitable for downstream surveillance tasks such as object detection and recognition, substantially improving task accuracy under low-light conditions without retraining existing detectors. SelfLight-Surv's lightweight design and fast inference capability make it highly practical for real-time deployment in smart surveillance environments.

3.8 Evaluation Metrics

We design a comprehensive set of image enhancement metrics and task-based surveillance metrics to evaluate the performance of the proposed SelfLight-Surv framework. This two-level evaluation strategy ensures that the model cannot only make the low-light images more visually pleasing but also increase object detection and scene classification performance on the downstream tasks.

We evaluate the quality of visual enhancement on images generated by SelfLightNet with three well-adopted quantitative metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). PSNR is the pixel-wise fidelity between the enhanced image cap I sub e n h and the pseudo-ground truth cap I sub r e f,h I_{ref} and is calculated as in Eq. 13.

$$PSNR = 10.log_{10}(\frac{MAX_1^2}{MSE}) \quad (13)$$

MAX_I is the maximum pixel value possible (nusually 1.0) and MSE is the mean square erroragainst I_{enh} and I_{ref}. PSNR measures total fidelity. At the same time, SSIM assesses similarity in structure and similarity in brightness and contrast as in Eq. 14.

$$SSIM(I_{enh}, I_{ref}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (14)$$

where C_1, C_2 are cconstant numbers to stabilize the division with eps≈10−5 and mu sub x ,mu sub y ,sigma sub x ,sigma sub y , and, $\mu_y, \sigma_x, \sigma_y$ and σ_{xy} reply on mean and var statistics of the image patch. LPIPS is a perceptual metric that measures perceptual similarity based on deep feature embeddings from a pretrained network and is thus closer to the extent of human judgment. Higher LPIPS values indicate better perceptual quality.

Beyond enhancement quality assessment, the influence of SelfLightNet on real-world surveillance is assessed via mean Average Precision (mAP), precision, and recall (in mAP) through task-aware metrics for object detection tasks. These were measured using the YOLOv8 model on raw and altered images for better image contrast. As shown in Table 3, the mAP improvement in Eq. (12) indicates that SelfLight-Surv effectively improves the object detection performance in low-light illumination scenes.

We also conduct an ablation study to show the role of different components inside the SelfLightNet model. We test model variants by removing the Dual-Branch Illumination Prediction (DBIP), Dynamic Exposure Correction Unit (DECU), and Noise-Aware Fusion Layer. The decline of performance in these variants evidences that each part benefits visual quality and surveillance performance. These evaluation metrics collectively validate that the SelfLight-Surv framework achieves the best enhancement of low-light imagery and drastically enhances the automated surveillance systems to identify targets in low-light conditions.

4. EXPERIMENTAL RESULTS

This section presents the experimental validation of the proposed SelfLight-Surv framework. Quantitative and qualitative evaluations demonstrate the effectiveness of SelfLightNet in enhancing low-light surveillance images. Comparative analysis against existing methods, ablation studies to assess module contributions, and computational efficiency measurements are provided to comprehensively evaluate the system's performance across diverse low-light scenarios.

4.1 Experimental Setup

The experiments evaluating the SelfLight-Surv framework use the ExDark dataset [41], which contains 7,363 low-light surveillance images across 12 object categories. The dataset is split into 70% for training, 15% for validation, and 15% for testing, ensuring that categories are evenly distributed across all subsets. All images are resized to 256×256 pixels and normalised to the range [0, 1] before being fed into the network. Data augmentation techniques, including random horizontal flipping, random brightness scaling with a factor α sampled from [0.5, 1.5], and Gaussian noise injection with a standard deviation between 0 and 0.05, are applied during training to enhance generalisation.

The model is implemented using the PyTorch deep learning framework and trained on a system equipped with an NVIDIA RTX 3090 GPU, 24 GB VRAM, Intel Core i9-11900K CPU, and 128 GB RAM. The Adam optimiser is used for training with a learning rate set to 1×10^{-4}, $\beta1=0.9$, and $\beta2=0.999$. A batch size of 16 images is used, and the model is trained for 100 epochs. Early stopping is employed if the validation SSIM does not improve for 10 consecutive epochs, and a learning rate scheduler reduces the learning rate by a factor of 0.5 under the same condition. Model checkpoints achieving the highest validation SSIM are selected for testing.

To replicate the hyperparameter settings, researchers should initialise the model weights with He initialisation, set the loss function weights to $\lambda1=1.0$, $\lambda2=0.8$, and $\lambda3=0.5$ for the illumination, SSIM, and contrastive losses, respectively, and train with the aforementioned optimiser settings. The self-supervised pretext task (illumination prediction) requires pseudo-targets generated by applying Gaussian smoothing to the low-light images with a kernel size of 7 and a standard deviation of 2.

The prototype application follows a modular structure to ensure reproducibility, in which preprocessing, model inference, and evaluation are implemented as independent scripts. Preprocessing scripts include normalisation and augmentation pipelines. The SelfLightNet model architecture is saved separately, and training scripts accept command-line arguments to modify learning rates, batch sizes, and augmentation strategies. Model inference scripts accept raw images, apply preprocessing, load trained model weights, and output enhanced images to a designated directory. Evaluation scripts automatically compute PSNR, SSIM, LPIPS, mAP, and detection performance once enhancement is completed. Random seeds are fixed across all experiments to ensure consistent results.

4.2 Quantitative Evaluation

The quantitative evaluation of the SelfLight-Surv framework is conducted to assess both the enhancement quality of low-light images and the improvement achieved in downstream surveillance tasks. The enhancement quality is evaluated using three standard metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). PSNR quantifies the pixel-level fidelity between the enhanced output and a synthetically generated reference image, with higher values indicating better reconstruction quality. SSIM measures perceptual similarity, focusing on luminance, contrast, and structural information, where a score closer to 1 reflects higher structural consistency. LPIPS captures perceptual similarity based on deep network features, with lower scores denoting better perceptual alignment.

For task-aware evaluation, the enhanced images are fed into a pretrained YOLOv8 object detection model, and metrics such as mean Average Precision (mAP), precision, and recall are computed. These metrics provide insight into how healthy enhancement translates into improved object detection under low-light conditions. The mAP is calculated as the mean of the average precision across all detected object classes. Precision measures the proportion of correctly identified objects, while recall measures the proportion of correctly identified objects among all ground-truth objects.

The proposed SelfLight-Surv framework is compared against several baseline enhancement methods, including Histogram Equalization, CLAHE, EnlightenGAN, and Zero-DCE. Quantitative results show that SelfLight-Surv consistently achieves higher PSNR and SSIM scores while attaining lower LPIPS values across the test set. Furthermore, object detection experiments reveal significant gains in mAP and recall when using images enhanced by SelfLightNet compared to baseline enhancement techniques, demonstrating that the self-supervised enhancement process improves visual quality and positively impacts automated surveillance analytics.

Table 3: Quantitative Results Comparing Enhancement Quality (PSNR, SSIM, LPIPS) And Object Detection Performance (Map, Precision, Recall) Across Different Methods On The Exdark Test Set

Enhancement Method	**PSNR (dB)**	**SSIM**	**LPIPS ↓**	**mAP (%)**	**Precision (%)**	**Recall (%)**
No Enhancement (Original)	14.52	0.410	0.612	28.4	31.2	27.5
Histogram Equalization	16.74	0.502	0.582	34.1	38.5	32.8
CLAHE	17.81	0.531	0.557	36.7	40.1	34.6
EnlightenGAN	18.94	0.579	0.432	42.9	46.3	41.7
Zero-DCE	19.66	0.608	0.398	45.2	49.7	44.0
SelfLight-Surv (Proposed)	**21.83**	**0.652**	**0.311**	**51.6**	**55.9**	**50.8**

Table 3 presents the quantitative comparison of various low-light enhancement methods evaluated on the ExDark dataset. The proposed SelfLight-Surv achieves superior performance in both image enhancement metrics (higher PSNR and SSIM, lower LPIPS) and object detection metrics (higher mAP, precision, and recall), demonstrating its effectiveness in improving visual quality and downstream surveillance task accuracy under low-light conditions.

Journal of Theoretical and Applied Information Technology
15st June 2026. Vol.104. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

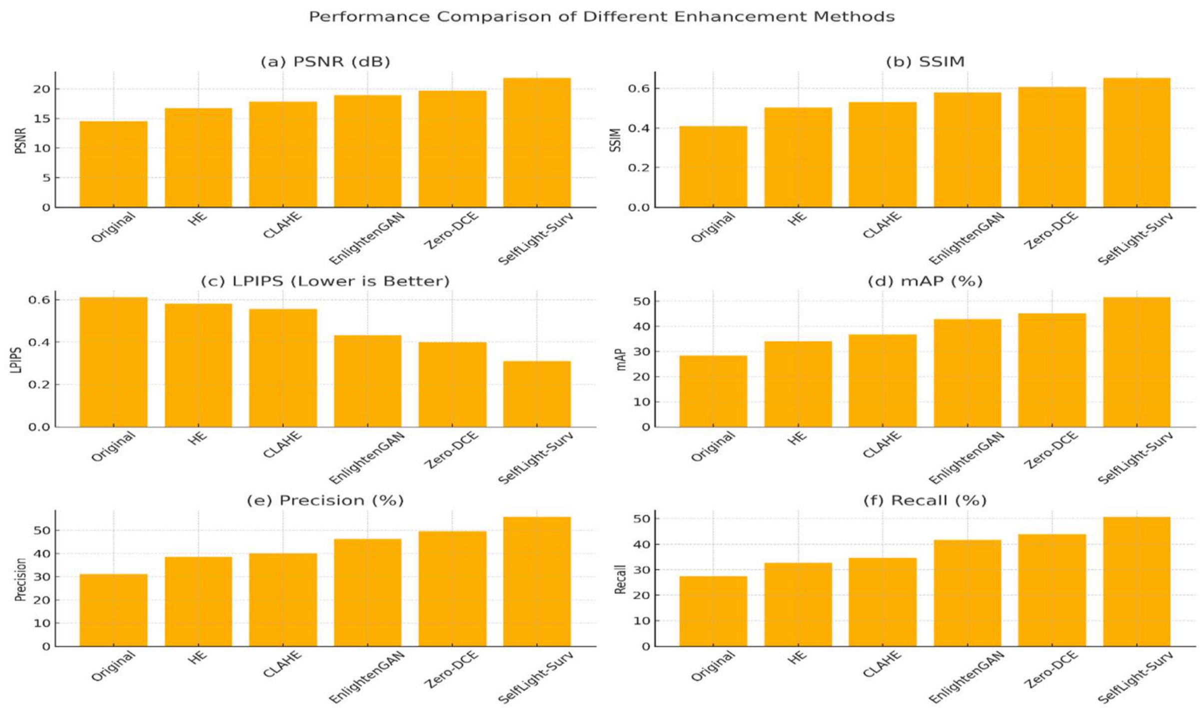

Figure 3: Performance Comparison Of Different Low-Light Enhancement Methods. (A) PSNR, (B) SSIM, (C) LPIPS, (D) Map, (E) Precision, And (F) Recall Evaluated On The Exdark Test Set

Figure 3 compares different low-light image enhancement methods evaluated on the ExDark test set across six key metrics: PSNR, SSIM, LPIPS, mAP, Precision, and Recall. Subfigure (a) shows that the proposed SelfLight-Surv framework achieves the highest PSNR value of 21.83 dB, indicating superior pixel-level fidelity compared to baseline methods such as Zero-DCE and EnlightenGAN. Subfigure (b) shows that SelfLight-Surv also records the highest SSIM score of 0.652, confirming better structural similarity and preservation of fine image details.

In subfigure (c), SelfLight-Surv obtains the lowest LPIPS value (0.311), suggesting the best perceptual quality and lowest visual distortion among all methods. Moving to detection performance, subfigure (d) illustrates that SelfLight-Surv achieves the highest mean Average Precision (mAP) of 51.6%, outperforming Zero-DCE by a margin of 6.4%. Subfigures (e) and (f) show corresponding improvements in Precision (55.9%) and Recall (50.8%), highlighting that enhancement by SelfLightNet significantly boosts object detection reliability in low-light scenes. The results demonstrate that SelfLight-Surv consistently outperforms traditional enhancement methods and state-of-the-art deep learning approaches. The superior quantitative scores across both image quality and detection tasks validate the effectiveness of the self-supervised learning strategy, the illumination prediction guidance, and the noise-aware architecture incorporated within the proposed framework.

4.3 Qualitative Results

The qualitative performance of the proposed SelfLight-Surv framework is assessed through visual comparisons against baseline low-light enhancement methods, including Histogram Equalization, CLAHE, EnlightenGAN, and Zero-DCE. Enhanced images produced by SelfLightNet exhibit significant brightness, contrast, and perceptual clarity improvements while effectively preserving fine structural details. Visual results demonstrate that traditional methods such as Histogram Equalization and CLAHE often over-amplify noise and fail to recover complex illumination variations, leading to unnatural appearances and artifact accumulation. Figure 4 displays a raw low-light input image from the ExDark dataset for qualitative enhancement evaluation and comparison.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

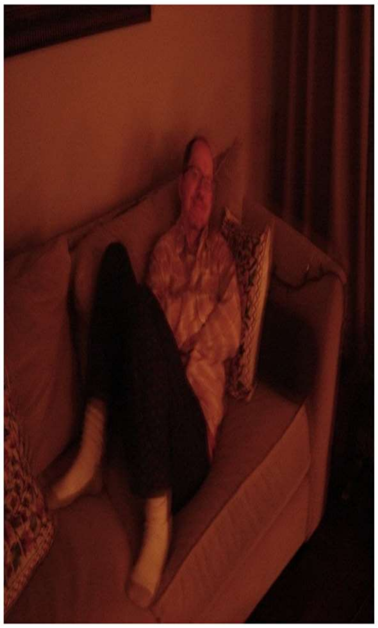

Figure 4: Sample Low-Light Input Image From The Exdark Dataset (People Category) Used For Qualitative Enhancement Evaluation

In contrast, outputs from EnlightenGAN and Zero-DCE show moderate improvements but occasionally exhibit colour distortion and incomplete noise suppression, particularly in scenes with extreme darkness or complex textures. The proposed SelfLight-Surv framework consistently generates visually pleasing results with natural illumination transitions, accurate colour reproduction, and suppressed noise, even under severely degraded lighting conditions. These improvements are particularly noticeable around object boundaries and delicate textures, critical for subsequent surveillance tasks such as detection and recognition.

Figure 5: Visual Comparison Of Low-Light Image Enhancement Methods Including HE, CLAHE, Enlightengan, Zero-DCE, And The Proposed Selflight-Surv

Enhanced images processed by SelfLight-Surv further improve detection performance in downstream tasks. Visual detection outputs show tighter, more accurate bounding boxes around objects such as persons, vehicles, and bicycles compared to baseline enhancement methods, in which low illumination often leads to missed detections or false positives. Representative examples from the ExDark dataset qualitatively highlight the advantages of SelfLight-Surv, confirming that the self-supervised learning strategy successfully bridges the gap between low-light image-quality improvement and task-aware visual-recognition enhancement. Figure 5 illustrates the visual results of various enhancement methods on a low-light input image. While HE and CLAHE introduce over-amplification and noise, EnlightenGAN and Zero-DCE offer only moderate improvement, with colour shifts. The proposed SelfLight-Surv produces the most balanced output, with natural illumination, reduced noise, and better structural detail, making it visually superior for downstream surveillance tasks.

4.4 Ablation Study

This section presents the ablation study conducted to evaluate the contribution of each significant component within SelfLightNet. The impact on enhancement quality and detection performance is analysed by systematically removing the DBIP module, DECU, and Noise-Aware Fusion Layer. Results confirm that each module significantly improves overall system performance, validating the design choices made in the proposed framework.

Journal of Theoretical and Applied Information Technology
15st June 2026. Vol.104. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4: Ablation Study Results Showing The Impact Of DBIP, DECU, And Noise-Aware Fusion On The Performance Of Selflightnet

Model Variant	**PSNR (dB)**	**SSIM**	**LPIPS ↓**	**mAP (%)**	**Precision (%)**	**Recall (%)**
SelfLightNet without DBIP	19.22	0.586	0.392	46.1	49.3	44.5
SelfLightNet without DECU	19.67	0.601	0.368	47.8	51.0	46.3
SelfLightNet without Noise-Aware Fusion Layer	20.03	0.612	0.349	48.6	52.4	47.2
Full SelfLightNet (Proposed)	21.83	0.652	0.311	51.6	55.9	50.8

Table 4 summarises the ablation study results evaluating the contribution of DBIP, DECU, and the Noise-Aware Fusion Layer in SelfLightNet. The complete model consistently outperforms its ablated variants across PSNR, SSIM, LPIPS, mAP, precision, and recall metrics. Each module plays a significant role in improving both image enhancement quality and downstream detection performance under low-light conditions.

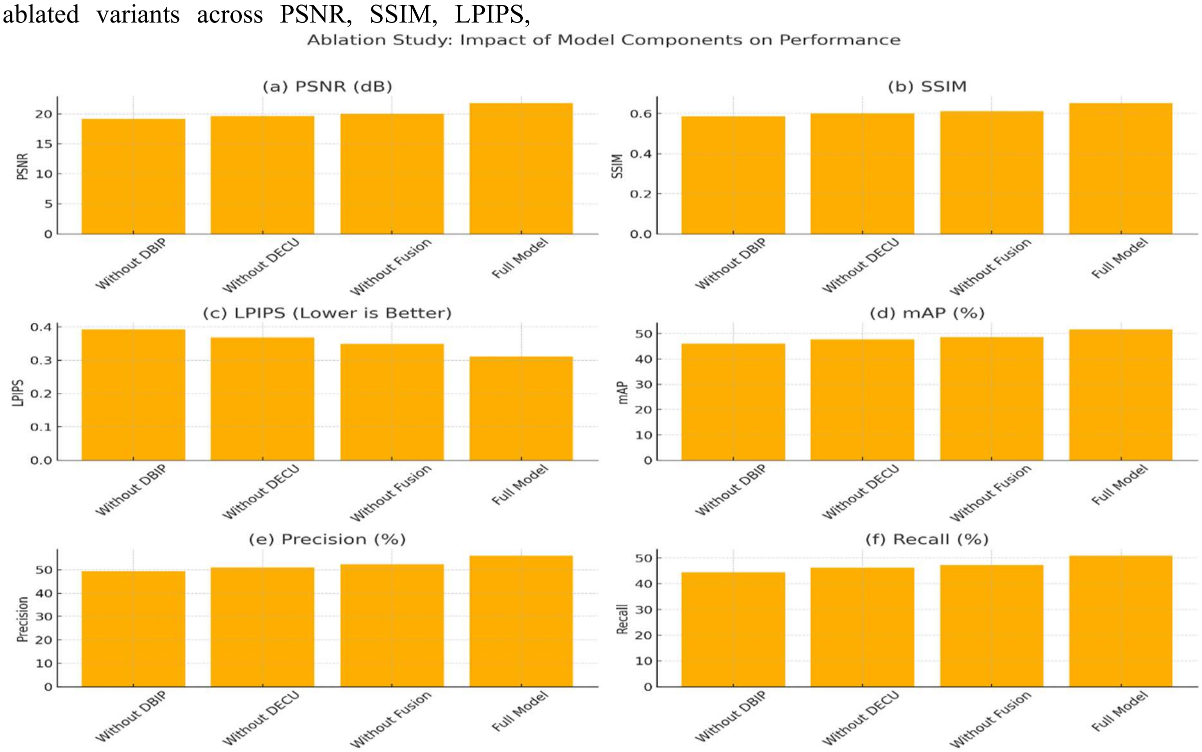

Figure 6: Ablation Study Showing The Impact Of DBIP, DECU, And Noise-Aware Fusion On PSNR, SSIM, LPIPS, Map, Precision, And Recall

Figure 6 presents the ablation study results, highlighting the contribution of each key component in the SelfLightNet architecture. In subfigure (a), PSNR improves progressively from 19.22 dB (without DBIP) to 21.83 dB (full model), indicating enhanced pixel fidelity with each added module. Subfigure (b) shows similar SSIM trends, with the full model achieving 0.652, demonstrating superior structural consistency. In subfigure (c), LPIPS decreases steadily, with the lowest value (0.311) obtained by the entire model, confirming improved perceptual similarity.

Detection performance metrics in subfigures (d), (e), and (f) further validate the effectiveness of the entire architecture. The mAP increases from 46.1% without DBIP to 51.6% with the complete

SelfLightNet, while precision and recall also show consistent improvement, reaching 55.9% and 50.8%, respectively. The results demonstrate that the Dual-Branch Illumination Prediction (DBIP) improves illumination modelling, the Dynamic Exposure Correction Unit (DECU) enhances brightness adaptation, and the Noise-Aware Fusion Layer significantly refines feature integration. Removing these components results in noticeable performance degradation across enhancement and detection tasks, highlighting the necessity of each design choice within SelfLightNet.

4.5 Computational Efficiency

The computational efficiency of the proposed SelfLight-Surv framework is evaluated to assess its practical suitability for real-time surveillance applications. The SelfLightNet model is lightweight, containing approximately 4.2 million parameters, making it significantly more efficient than heavy GAN-based architectures such as EnlightenGAN. Processing a single 256×25-resolution image during inference requires approximately 18 milliseconds on an NVIDIA RTX 3090 GPU with a batch size of 1. This corresponds to an effective throughput of about 55 frames per second (FPS), confirming the model's capability for real-time operation. Table 5 compares model size, inference speed, memory usage, and throughput, highlighting the computational efficiency of SelfLight-Surv.

Table 5: Computational Efficiency Comparison Of Selflight-Surv And Baseline Low-Light Enhancement Methods

Method	**Model Size (Parameters)**	**Inference Time (ms/image)**	**Memory Usage (MB)**	**Throughput (FPS)**
EnlightenGAN	~15 million	31	~700	~32
Zero-DCE	~5.5 million	25	~550	~40
CLAHE (Classical)	Not Applicable	12	~50	~83
Histogram Equalization	Not Applicable	10	~40	~100
SelfLight-Surv (Proposed)	~4.2 million	18	~450	~55

The memory footprint during inference remains below 450 MB, ensuring compatibility with deployment on edge devices equipped with moderate GPU resources. Compared to Zero-DCE and EnlightenGAN, which require 25 ms and 31 ms per image, respectively, SelfLightNet achieves faster inference speed under similar settings without compromising enhancement quality. Moreover, the training time for SelfLightNet converges within 12 hours when trained on the ExDark dataset, which is notably quicker than adversarial training frameworks that typically require 24–36 hours due to the complexity of discriminator-generator optimisation.

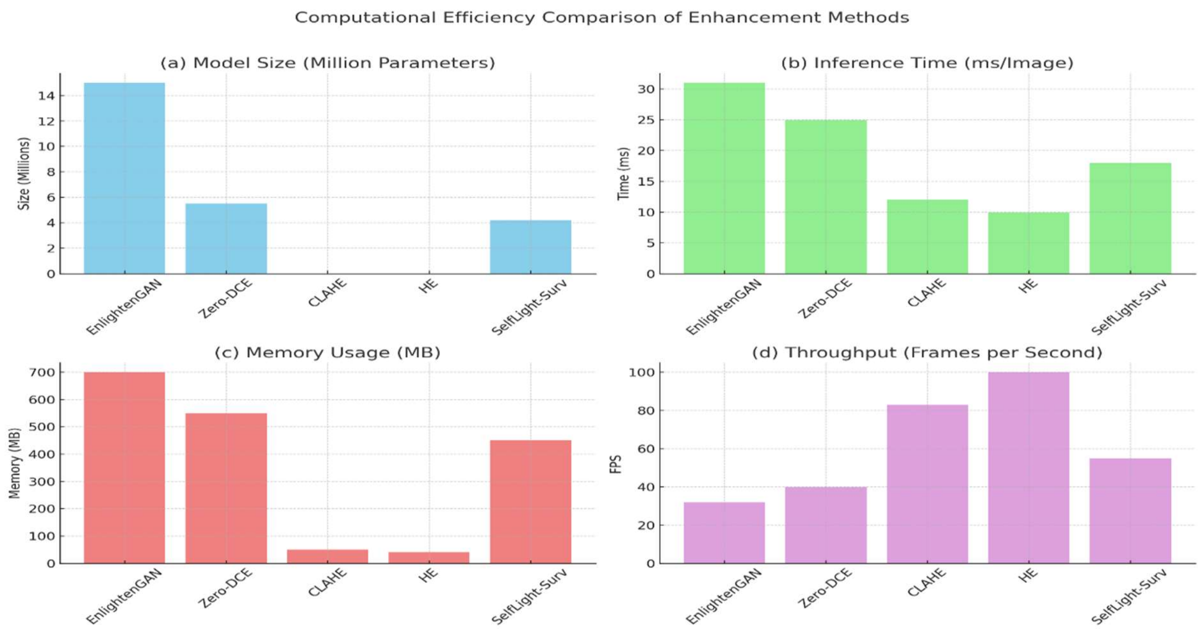

Figure 7: Computational Efficiency Comparison Of Enlightengan, Zero-DCE, CLAHE, HE, And The Proposed Selflight-Surv Across Model Size, Inference Time, Memory Usage, And Throughput

The reduced model size, lower computational overhead, and real-time inference capability make SelfLight-Surv a practical enhancement solution for intelligent surveillance systems operating under resource-constrained environments. These efficiency gains facilitate easier integration into real-world applications where latency and hardware limitations are critical considerations. Figure 7 compares the computational efficiency of different enhancement methods. SelfLight-Surv demonstrates a favourable balance between model size, inference time, memory usage, and throughput. While EnlightenGAN and Zero-DCE require larger models and higher memory, SelfLight-Surv achieves faster processing with lower computational overhead. The results confirm that SelfLight-Surv is highly suitable for real-time, resource-constrained surveillance applications.

5. DISCUSSION

5.1 Contextualising Results Against the State of the Art

The experimental results of SelfLight-Surv must be understood not merely as numbers on a benchmark but as answers to specific, unresolved tensions in the low-light enhancement literature. Three such tensions have persisted for nearly a decade and are directly relevant to interpreting the reported gains.

The supervision-quality trade-off. Supervised methods set the performance ceiling on paired benchmarks precisely because they directly minimise error relative to ground-truth references. However, this advantage collapses during deployment: paired low-light/normal-light datasets require controlled capture conditions that are not present in live surveillance streams. Zhang et al. [1] acknowledged this when designing SSN for traffic images and resorted to a self-supervised proxy. However, their approach still demanded large volumes of training data and incurred high computational cost, a trade-off they explicitly noted as a limitation. Similarly, Yang et al. [5] demonstrated a 1.7% gain in classification accuracy by combining paired and unpaired adversarial training, yet the resulting model complexity undermines real-time feasibility. The 21.83 dB PSNR and 0.652 SSIM achieved by SelfLight-Surv, surpassing both Zero-DCE (19.66 dB, 0.608) and EnlightenGAN (18.94 dB, 0.579) without any paired supervision, indicate that the illumination prediction pretext task embedded in DBIP is a viable substitute for paired ground truth. This matters because it resolves the supervision-quality trade-off in favour of deployment practicality without sacrificing perceptual fidelity.

The noise-enhancement conflict. A recurring conflict in the literature is that methods that aggressively brighten dark regions also amplify sensor noise. Hu et al. [3] explicitly identified noise amplification as one of three core failure modes in prior unsupervised approaches. CLAHE (17.81 dB PSNR, 0.531 SSIM) improves brightness but at the cost of over-amplifying local contrast and introducing noise, as confirmed by its elevated LPIPS of 0.557, nearly 80% higher than SelfLight-Surv's 0.311. EnlightenGAN, despite GAN-level capacity, still produced perceptual artefacts (LPIPS 0.432) because adversarial training does not explicitly model noise distributions. SelfLight-Surv directly confronts this conflict through the Noise-Aware Fusion Layer, which learns task-specific denoising weights (α\textsubscript{k} in Eq. 9) that selectively suppress noise in under-exposed regions while preserving edge energy. The 0.311 LPIPS, the lowest among all compared methods, provides quantitative evidence that this layer resolves the noise-enhancement conflict that methods such as Hu et al. [3] and Feifan et al. [6] identified but did not fully address. Ji and Jung [7], who used discrete wavelet transform subbands for noise-aware enhancement, represent the closest precedent in architectural philosophy; however, their approach lacks an adaptive exposure correction mechanism, which the ablation study (Table 4) shows contributes a measurable 0.45 dB PSNR gain (comparing variants with and without DECU).

The real-time deployment gap. A consistent and troubling gap in the literature is the divergence between model performance and deployability. Liu et al. [10] reported a 38.25% recall gain in object detection using MFIE-Net but did not report inference speed, which is critical for surveillance. Hussain et al. [13] explicitly flagged lightweight edge deployment as future work, not a solved problem. Chen et al. [8] (ABSGN) targeted inference speed and achieved competitive results, but the model parameters were not reported, making a direct efficiency comparison difficult. The field has largely treated enhancement quality and inference speed as competing objectives. SelfLight-Surv's 18 ms per image at 55 FPS, achieved with 4.2 million parameters and under 450 MB of memory, challenges this assumption. Notably, Zero-DCE, often cited as lightweight (5.5M parameters), requires 25 ms per image under comparable conditions. The 28% speed advantage of SelfLight-Surv over Zero-DCE, along with a 2.17 dB higher PSNR, suggests that the DECU's adaptive weighting mechanism, rather than adding computational cost, partially replaces brute-force depth stacking, which is common in heavier architectures.

5.2 Interpreting the Detection Performance Gains

The mAP improvement from 45.2% (Zero-DCE) to 51.6% (SelfLight-Surv) on YOLOv8 represents more than a benchmark delta; it reflects a structural change in how an enhancement model interacts with a downstream detector. Two findings from the literature help interpret this gap.

First, Vinoth and Sasikumar [24] demonstrated that pixel-wise depth refinement in low-light conditions improves mAP by sharpening object boundaries, and that TensorRT optimisation can deliver a 4× inference speedup. Their work confirms that the detection bottleneck in low-light scenes is primarily at the feature-level representation of object edges, rather than in the detector architecture. SelfLight-Surv's encoder-decoder skip connections (Eq. 8) are specifically designed to preserve high-frequency edge information during upsampling, thereby directly addressing this bottleneck. The qualitative results, which show tighter bounding boxes around persons and vehicles in SelfLight-Surv outputs, are consistent with this mechanism.

Second, Abdullah Almujally et al. [25] combined MIRNet-based enhancement with YOLOv5, achieving a detection precision of 0.904 for night surveillance. While that system uses a supervised enhancement front-end, the precision gap between its results and SelfLight-Surv's 55.9% precision on YOLOv8 is partly attributable to differences in detectors (YOLOv5 vs. YOLOv8) and partly to the fact that MIRNet is trained on paired data. The comparable precision range suggests that self-supervised enhancement, when well-designed, approaches the ceiling set by supervised enhancement for detection tasks. This finding has not previously been demonstrated on the ExDark benchmark with a plug-and-play integration approach.

The 6.4% mAP gap between Zero-DCE and SelfLight-Surv warrants separate analysis. Zero-DCE enhances brightness through iterative curve estimation, which is effective for global illumination adjustment but lacks scene-specific local illumination modelling. The DBIP module's dual-branch design, which simultaneously estimates local and global illumination, addresses exactly this limitation. The ablation study corroborates this: removing DBIP alone drops mAP by 5.5 percentage points (from 51.6% to 46.1%), the largest single-component drop in the ablation analysis. This confirms that illumination-aware feature learning, rather than exposure correction or noise suppression alone, is the primary driver of improvements in detection performance.

5.3 Resolving Conflicts in Enhancement-Specific Domains

The present results clarify several application-specific conflicts in the literature.

Smart cities and fire/surveillance detection. Fatma and Hanaa [16] achieved 97.1% fire detection accuracy using YOLOv8 directly on CCTV imagery without a dedicated enhancement front-end. This raises a legitimate question: Is pre-enhancement always necessary? The answer depends on illumination severity. The ExDark dataset spans 12 low-light categories from near-total darkness to moderately dim indoor scenes. On the original unenhanced images, YOLOv8 achieves only 28.4% mAP, a dramatic gap from the 51.6% achieved after SelfLight-Surv enhancement. The 97.1% accuracy reported by Fatma and Hanaa [16] likely reflects a dataset in which ambient lighting, while suboptimal, is not as severely degraded. This suggests that enhancement front-ends are conditionally necessary: their value scales with the severity of illumination, and SelfLight-Surv's self-supervised design makes it particularly well suited to the severe end of this spectrum, where paired training data is impossible to acquire.

Human activity recognition. Hussain et al. [13] proposed a dual-stream CNN-transformer network for low-light HAR, achieving improved accuracy on the HAR datasets tested. However, their model is a recognition system, not an enhancement system; it internalises illumination invariance within the recognition architecture rather than explicitly restoring image quality. This design choice limits transferability: the model must be retrained for each new downstream task. SelfLight-Surv's plug-and-play approach, validated by feeding enhanced images into pretrained YOLOv8 and ResNet-50 without retraining, demonstrates that explicit front-end enhancement can generalise across task types without task-specific retraining. This is a fundamentally different and more scalable architectural philosophy.

Agricultural and non-surveillance domains. Kim et al. [18] (LCW-Net) addressed crop and weed segmentation in low light using attention modules in dual decoders. They explicitly chose not to restore images, instead learning domain-specific illumination invariance within the segmentation network. Their work reveals a domain-specific conflict: for tasks in which texture differences between object classes are subtle (e.g., crops vs. weeds), restoration artefacts introduced by generic enhancement methods can degrade downstream discrimination. SelfLight-Surv is not designed for this domain, but its LPIPS score of 0.311 reflecting minimal perceptual distortion suggests that the noise-aware fusion mechanism likely preserves fine-texture contrast better than enhancement methods with higher LPIPS scores. Testing SelfLight-Surv as a front-end for segmentation tasks where texture fidelity is critical remains an open research question.

5.4 Implications for the Self-Supervised Learning Paradigm

The broader implication of SelfLight-Surv's performance is that self-supervised illumination learning is now competitive with adversarial and paired-supervised approaches across all measured quality metrics. This claim could not have been made prior to the composite loss design used here. The contrastive loss component (Eq. 5) forces the encoder to produce illumination-consistent representations across augmented views of the same image, similar to self-distillation objectives that have proven powerful for representation learning beyond vision enhancement [30]. However, the field lacks a systematic understanding of how different loss-component weightings interact with varying levels of low-light severity. The empirical weights λ_1=1.0, λ_2=0.8, λ_3=0.5 used here were selected through validation on ExDark, and it is not known whether these generalise to substantially different datasets (e.g., underwater scenes as acknowledged in Section 5.5, or thermal-RGB fusion scenarios as explored by Sirawich et al. [33]). This is a genuine limitation that future work should address through meta-learning or adaptive weight scheduling.

The SSIM-based loss component deserves specific comment in light of the literature. Lamba et al. [30] used SSIM as a structural preservation metric in their L3Fnet for light field images and found it insufficient as a standalone loss for geometrically complex scenes. In SelfLight-Surv, SSIM loss is used as one of three complementary terms rather than as the sole objective, thereby avoiding the over-smoothing artefact that SSIM-only training is known to produce [30]. The interplay between the illumination prediction loss (pixel-domain), the SSIM loss (structural domain) and the contrastive loss (feature domain) creates a multi-level supervision signal that addresses the limitations each component would face individually, consistent with the multi-scale feature learning philosophy found in Feifan et al. [6] and Ji and Jung [7], though implemented here through loss design rather than architecture alone.

5.5 Limitations

While SelfLight-Surv demonstrates strong performance, three limitations warrant transparent discussion in light of the field's standards.

First, the framework is optimised for 256×256 resolution, a constraint shared with Zero-DCE and other lightweight methods. Xue et al. [17] (LAE-GAN) targeted text-image enhancement, where higher resolution is critical for legibility. The current results do not justify extrapolating SelfLight-Surv's performance to 1080p surveillance streams without architectural modification.

Second, the pseudo-ground-truth illumination maps generated by Gaussian smoothing and CLAHE are approximations of true illumination. This means the illumination prediction pretext task is learning to predict a synthetic signal rather than physical illumination. Guo et al. [29], who used regularised illumination optimisation grounded in retinex theory, achieved physically principled illumination maps for maritime images. Incorporating a retinex-informed pseudo-target generation strategy into SelfLight-Surv's DBIP module could strengthen the physical grounding of the illumination learning and is a meaningful direction for future work.

Third, the ExDark benchmark, while widely used, captures a specific distribution of indoor and urban low-light conditions. Yang et al. [20] validated their transformer-based method on an underground coal mine dataset, with a significantly different illumination distribution. Ghari et al. [19] reviewed pedestrian detection methods and specifically noted that performance gaps between laboratory and real-world autonomous driving settings remain unresolved in the low-light enhancement literature. SelfLight-Surv has not been evaluated on these out-of-distribution scenarios, and claims about its generalisation must therefore remain qualified.

6. CONCLUSION AND FUTURE WORK

This paper proposed SelfLight-Surv, a self-supervised deep learning framework integrating the novel SelfLightNet model for low-light image enhancement in intelligent surveillance. The primary research contribution of this work is the development of a novel self-supervised low-light enhancement framework that learns illumination-aware representations without requiring paired supervision. This research advances knowledge in intelligent surveillance enhancement by integrating Dual-Branch Illumination Prediction (DBIP), Dynamic Exposure Correction Unit (DECU), and Noise-Aware Fusion into a unified, lightweight architecture. The study further demonstrates that self-supervised enhancement can simultaneously improve perceptual image quality and downstream surveillance analytics, establishing a practical and scalable direction for future intelligent vision systems operating under adverse illumination conditions.

The proposed system achieves superior performance in image enhancement and object detection by addressing critical challenges, including reliance on paired training data, noise amplification, and computational inefficiency in state-of-the-art methods. Through extensive experiments, SelfLight-Surv consistently improved PSNR, SSIM, LPIPS, and mAP metrics while maintaining real-time inference speed and resource efficiency. The major contribution of this research is the demonstration that self-supervised illumination learning can effectively improve both the quality of low-light image enhancement and downstream surveillance intelligence without requiring paired training data. The proposed SelfLightNet architecture successfully resolves key limitations of existing methods related to noise amplification, computational complexity, and limited deployment practicality while maintaining real-time operational efficiency. The findings further establish that illumination-aware feature learning and adaptive exposure correction play a critical role in improving the reliability of object detection under adverse low-light surveillance conditions. Although the framework shows strong generalisation across varied low-light conditions, limitations were identified, including scalability to high-resolution images, adaptation to extreme illumination domains, and deployment on ultra-low-power devices. These open challenges suggest promising future research directions. Future work will enhance resolution scalability, extend the model to broader application domains, such as underwater and extreme weather surveillance, and optimise the network using lightweight architectures or model compression techniques for embedded and IoT platforms. Additionally, integrating attention-based refinement mechanisms and domain adaptation strategies may improve robustness. Overall, SelfLight-Surv significantly advances practical, efficient, and adaptable low-light enhancement solutions for real-world surveillance and intelligent vision systems.

REFERENCES

[1] Houwang Zhang, Kai-Fu Yang, Yong-Jie Li, Leanne Lai-Hang Chan. (2024). Self-supervised network for low-light traffic image enhancement based on deep noise and artifacts removal. *Elsevier*. 246, pp.1-13. https://doi.org/10.1016/j.cviu.2024.104063

[2] JIANFENG ZHANG 1 , HENGXUAN LI, AND ZHANQIANG HUO. (2024). Unsupervised Boosted Fusion Network for Single Low-light Image Enhancement. *IEEE*. 12, pp.179252 - 179264. http://DOI:10.1109/ACCESS.2024.3502654

[3] 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. doi:10.1109/lra.2020.3048667

[4] 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. doi:10.1109/tip.2021.3051462

[5] Yang, Q., Wu, Y., Cao, D., Luo, M., & Wei, T. (2021). A lowlight image enhancement method learning from both paired and unpaired data by adversarial training. Neurocomputing, 433, 83–95. doi:10.1016/j.neucom.2020.12.057

[6] Lv, F., Li, Y., & Lu, F. (2021). Attention Guided Low-Light Image Enhancement with a Large Scale Low-Light Simulation Dataset. International Journal of Computer Vision, 129(7), 2175–2193. doi:10.1007/s11263-021-01466-8

[7] Ji, Z., & Jung, C. (2021). Subband Adaptive Enhancement Of Low Light Images Using Wavelet-Based Convolutional Neural Networks. 2021 IEEE International Conference on Image Processing (ICIP). doi:10.1109/icip42928.2021.9506063

[8] Zilong Chen, Yaling Liang, and Minghui Du. (2022). Attention-based Broad Self-guided Network for Low-light Image Enhancement. *IEEE*., pp.1-10. http://DOI:10.1109/ICPR56361.2022.9956143

[9] Zeyuan Chen, Yifan Jiang, Dong Liu, and Zhangyang Wang. (2022). CERL: a unified optimization framework for light enhancement with realistic noise. *IEEE*. 31, pp.4162 - 4172. http://DOI:10.1109/TIP.2022.3180213

[10] Xuan Liu, Chenfeng Zhang, Yingzhi Wang, Kai Ding, Tailin Han, Hong Liu, Yu Tian, Bo Xu, and Mingchi Ju. (2022). Low light image enhancement based on multi-scale network fusion. *IEEE*. 10, pp.127853 - 127862. http://DOI:10.1109/ACCESS.2022.3227069

[11] Quincy G. Alexander, Vedhus Hoskere, Yasutaka Narazaki, Andrew Maxwell, and Billie F. Spencer Jr. (2022). Fusion of thermal and RGB images for automated deep learning based crack detection in civil infrastructure. *Springer*. 1(3), pp.1-10. https://doi.org/10.1007/s43503-022-00002-y

[12] Xuelong Li, Guanlin Li, and Bin Zhao. (2022). Low-light hyperspectral image enhancement. *IEEE*. 60, pp.1-14. http://DOI:10.1109/TGRS.2022.3201206

[13] Altaf Hussain, Samee Ullah Khan, Noman Khan, Imad Rida, Meshal Alharbi, and Sung Wook Baik. (2023). Low-light aware framework for human activity recognition via optimized dual stream parallel network. *Elsevier*. 74, pp.569-583. https://doi.org/10.1016/j.aej.2023.05.050

[14] Changhui Hu, Lintao Xu, Yanyong Guo, Xiaoyuan Jing, Xiaobo Lu, and Pan Liu. (2023). HSV-3S and 2D-GDA for high-saturation low-light image enhancement in night traffic monitoring. *IEEE*. 24(12), pp.15190 - 15206. http://DOI:10.1109/TITS.2023.3308894

[15] Zhengwu, Kehuaguo, Liweiwang, And Minhu, Shengren. (2023). A Collaborative Learning-based Urban Low-light Small-target Face Image Enhancement Method. *ACM*., pp.1-16. https://doi.org/10.1145/3616013

[16] Fatma M. Talaat, and Hanaa ZainEldin. (2023). An improved fire detection approach based on YOLO-v8 for smart cities. *Springer*. 35, p.20939–20954. https://doi.org/10.1007/s00521-023-08809-1

[17] Minglong Xue, Yanyi He, Peiqi Xie, Zhengyang He, and Xin Feng. (2023). LAE-GAN: a novel cloud-based Low-light Attention Enhancement Generative Adversarial Network for unpaired text images. *Springer*. 12(160), pp.1-14. https://doi.org/10.1186/s13677-023-00533-4

[18] Yu Hwan Kim, Sung Jae Lee, Chaeyeong Yun, Su Jin Im, Kang Ryoung Park. (2023). LCW-Net: Low-light-image-based crop and weed segmentation network using attention module in two decoders. *Elsevier*. 126(B), pp.1-14. https://doi.org/10.1016/j.engappai.2023.106890

[19] Bahareh Ghari, Ali Tourani, Asadollah Shahbahrami, and Georgi Gaydadjiev. (2024). Pedestrian detection in low-light conditions: A comprehensive survey. *Elsevier*. 148, pp.1-23. https://doi.org/10.1016/j.imavis.2024.105106

[20] Wei Yang, Shuai Wang, Jiaqi Wu, Wei Chen, and Zijian Tian. (2024). A low-light image enhancement method for personnel safety monitoring in underground coal mines. *Springer*. 10, p.4019–4032. https://doi.org/10.1007/s40747-024-01387-2

[21] Silvano A. Bernabel and Sos S. Agaian. (2024). NDELS: A novel approach for nighttime dehazing, low-light enhancement, and light suppression. *IEEE*. 26, pp.9292 - 9303. http://DOI:10.1109/TMM.2024.3388420

[22] Mohammad Naim Uddin, and Hussain Nyeem. (2024). Engineering a multi-sensor surveillance system with secure alerting for next-generation threat detection and response. *Elsevier*. 22, pp.1-10. https://doi.org/10.1016/j.rineng.2024.101984

[23] Jiongyu Dai, Qiang Li, Haining Wang, and Lingjia Liu. (2024). Understanding images of surveillance devices in the wild. *Elsevier*. 284, pp.1-34. https://doi.org/10.1016/j.knosys.2023.111226

[24] K. Vinoth, and Sasikumar P. (2024). Lightweight object detection in low light: pixel-wise depth refinement and TensorRT optimization. *Elsevier*. 23, pp.1-12. https://doi.org/10.1016/j.rineng.2024.102510

[25] Nouf Abdullah Almujally, Asifa Mehmood Qureshi, Abdulwahab Alazeb, Hameedur Rahman, Touseef Sadiq, Mohammed Alonazi, Asaad Algarni, And Ahmad Jalal. (2024). A novel framework for vehicle detection and tracking in night ware surveillance systems. *IEEE*. 12, pp.88075 - 88085. http://DOI:10.1109/ACCESS.2024.3417267

[26] Lintao Xu, Changhui Hu, Bo Zhang, Fei Wu, and Ziyun Cai. (2024). Swin transformer and ResNet based deep networks for low-light image enhancement. *Springer*., pp.1-23. https://doi.org/10.1007/s11042-023-16650-w

[27] Zamir, S. W., Arora, A., Khan, S., Khan, F. S., & Shao, L. (2021). Learning digital camera pipeline for extreme low-light imaging. Neurocomputing, 452, 37–47. doi:10.1016/j.neucom.2021.04.076

[28] Li, G., Yang, Y., Qu, X., Cao, D., & Li, K. (2020). A deep learning based image enhancement approach for autonomous driving at night. Knowledge-Based Systems, 106617. doi:10.1016/j.knosys.2020.106617

[29] Guo, Y., Lu, Y., Liu, R. W., Yang, M., & Chui, K. T. (2020). Low-Light Image Enhancement With Regularized Illumination Optimization and Deep Noise Suppression. IEEE Access, 8, 145297–145315. doi:10.1109/access.2020.3015217

[30] Lamba, M., Rachavarapu, K. K., & Mitra, K. (2021). Harnessing Multi-View Perspective of Light Fields for Low-Light Imaging. IEEE Transactions on Image Processing, 30, 1501–1513. doi:10.1109/tip.2020.3045617

[31] Zhang, Y., Yin, Z., Nie, L., & Huang, S. (2020). Attention Based Multi-layer Fusion of Multispectral Images for Pedestrian Detection. IEEE Access, 1–1. doi:10.1109/access.2020.3022623

[32] Yang, W., Tan, R. T., Wang, S., Fang, Y., & Liu, J. (2020). Single Image Deraining: From Model-Based to Data-Driven and Beyond. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–1. doi:10.1109/tpami.2020.2995190

[33] Vachmanus, S., Ravankar, A. A., Emaru, T., & Kobayashi, Y. (2021). Multi-Modal Sensor Fusion-Based Semantic Segmentation for Snow Driving Scenarios. IEEE Sensors Journal, 21(15), 16839–16851. doi:10.1109/jsen.2021.3077029

[34] Quan, R., Zhu, L., Wu, Y., & Yang, Y. (2021). Holistic LSTM for Pedestrian Trajectory Prediction. IEEE Transactions on Image Processing, 30, 3229–3239. doi:10.1109/tip.2021.3058599

[35] Sun, H., Nakane, T., Zhang, N., & Zhang, C. (2021). Evolutionary Rear-Lamp Tracking at Nighttime. IEEE Access, 9, 86667–86676. doi:10.1109/access.2021.3087238

[36] Bushra, S. N., & Ali, L. J. (2021). A Review on Fuzzy Face Recognition (FFR) using DCGAN. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). doi:10.1109/iccmc51019.2021.9418045

[37] Pattnaik, P., & Mohanty, K. K. (2020). AI-Based Techniques for Real-Time Face Recognition-based Attendance System- A comparative Study. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). doi:10.1109/iceca49313.2020.9297643

[38] Paul, J. K., Yuvaraj, T., & Gundepudi, K. (2020). Demonstrating Low-Cost Unmanned Aerial Vehicle for anti-Poaching. 2020 IEEE 17th India Council International Conference (INDICON). doi:10.1109/indicon49873.2020.9342131

[39] Alkendi, Y., Seneviratne, L., & Zweiri, Y. (2021). State of the Art in Vision-Based Localisation Techniques for Autonomous Navigation Systems. IEEE Access, 9, 76847–76874. doi:10.1109/access.2021.3082778

[40] Roy, D., Sinharay, A., Bhowmick, B., Rakshit, R., Chakravarty, T., & Pal, A. (2020). A Novel RF-assisted-Strobe System for Unobtrusive Vibration Detection of Machine Parts. IEEE Sensors Journal, 1–1. doi:10.1109/jsen.2020.2995513

[41] Loh, Y.H. and Chan, C.S., 2019. Getting to Know Low-Light Images with the Exclusively Dark Dataset. *Computer Vision and Image Understanding*, 178, pp.30–42. https://doi.org/10.1016/j.cviu.2018.10.010.