Real-Time Image Enhancement for Low-Resolution Robotics in Low Light Conditions.

Adityan Thirumani Srinivasan

Received April 16, 2025
Accepted October 12, 2025
Electronic access November 30, 2025

Deep learning has revolutionized image processing by enabling models to learn complex patterns from large datasets. It is widely applied in object recognition and robotics, simplifying perception and decision-making processes. This research focuses on enhancing robotic vision in low-visibility indoor environments using a hybrid deep learning system. The objective is to improve clarity in low-resolution images while maintaining real-time usability. The proposed HybridNet combines Zero-DCE and CIDNet to leverage brightness adjustment, detail preservation, and color fidelity. The model was evaluated on the LOL dataset, demonstrating its applicability in controlled indoor scenarios.To ensure robustness and mitigate concerns related to the limited test set, a comprehensive 5-fold cross-validation was conducted. This analysis confirmed consistent performance across different data splits, with average metrics of 27.38 dB PSNR, 0.941 SSIM, and 0.00119 LPIPS, thereby validating the model's generalizability and stability. Experimental results validated that HybridNet outperformed current state-of-the-art techniques in PSNR, SSIM, and LPIPS metrics. When deployed on a Raspberry Pi 5, the model initially ran at 1 FPS, but after applying ONNX optimization and 4-CPU-thread execution, inference improved to 2.6 FPS, demonstrating enhanced real-time performance. Deployment on a Jetson Orin Nano enabled real-time performance of up to 19 FPS, emphasizing the advantage of specialized edge AI hardware for time-sensitive robotic applications. The improved clarity enhances object recognition and stationary indoor monitoring tasks, such as security observation and equipment status checks. Future research will focus on reducing latency and improving performance under extreme low-light conditions.

Keywords: Real-Time Image Enhancement, Low-Light Robotics, Hybrid Neural Networks, Zero-DCE, CIDNet

Introduction

Background and Significance:

Robotic systems must perceive and understand the surroundings very precisely in order to perform well in actual surroundings. Natural surroundings, on the other hand, usually comprise unfavorable conditions such as diminished lighting conditions, sensor noise, and blur due to motions which cause significant deterioration in quality of images captured by robotics vision systems[1]. These have a negative impact on important tasks such as object classification and detection in stationary setups.

In order to overcome all these challenges, image enhancement techniques have emerged as a popular research area in recent times. Traditional techniques, for instance, Dark Channel Prior[2], apply physical priors for improving perceptual quality in poor conditions of images. Deep-learning-based solutions, particularly those related to CNNs, also reached state-of-the-art levels in recent times for the real-time restoration of poor-quality images[3]. Zero-DCE[4] and CIDNet[5] are two recent state-of-the-arts which advanced a step further by enabling learnable enhancement parameters in an end-to-end manner without paired datasets and making them competent for real-time robotic applications.

Despite all these efforts, existing models also have their own weaknesses. Zero-DCE has an advantage when it comes to brightness and contrast enhancement but sometimes loses texture information at times. CIDNet improves the color fidelity but may be computationally expensive for robotics in real-time. Further, traditional priors fail under difficult lighting conditions or dynamic video feeds. Secondly, predominantly used Python packages (e.g., OpenCV, PIL) offer some general image enhancement operations such as brightness and sharpening but lack in being fixed-filter type and not tunable to various and challenging conditions of robotic vision systems.

This introduces HybridNet as a novel deep neural network encompassing the strengths of Zero-DCE and CIDNet for better restoration quality without sacrificing real-time efficiency. By combining adaptive dynamic refinement of Zero-DCE and color correction/detail preservation of CIDNet, HybridNet gets a better quality vs computational resources trade-off. This makes it suitable for stationary indoor monitoring (e.g., security, equipment status panels, inventory shelf views) under low-light conditions[6].

Research Question/Objective:

This study explores the following research question: How can real-time image enhancement improve low-resolution, low-light robotic imagery?

The aim of this research is to create and implement a real-time image enhancement model specific to low-resolution and low-light images to enable robots to operate well in most tough environments. The proposed data set that can find a place as a test case is the Low-Light (LOL)[7] dataset because it closely caters to the simulation of low light environments and has quite diverse examples.

The primary objective of this research is to develop a robust real-time image enhancement solution that improves the perceptual capabilities of robotic systems operating in environments compromised by low-quality imagery. By enhancing visual clarity, the proposed model aims to enable more reliable and accurate robotic visualization and decision-making, which are critical for successful operation in low-light indoor settings.

Literature Review

Image Enhancement Techniques in Robotics

Image enhancement plays a significant and crucial role in perceptual robots, especially under compromised conditions, including low light, motion blur, and sensor degradation[8]. Traditional enhancement methods, such as histogram equalization, have issues in relation to severe degradation, whereas deep learning-based methods, especially CNNs, are among the most promising in addressing such questions[3]. Such methods allow fine details to be at least partially restored in the image, enhancing visibility to allow for proper decision-making before object identification and consequent actions are enacted.

Techniques including Zero-DCE (zero-reference deep curve estimation)[4,5] are meant for low-light gameplay and foggy environments, providing restoration of clarity and contrast within degraded images that are very important in robotic systems with respect to indoor robotic inspection, warehouse monitoring, and home automation tasks.

Deep Learning-Based Image Enhancement

Over the years, researchers have been able to design computerized algorithms to achieve brightness in unfavorable conditions through methods in deep learning. One is an example where Zero-DCE applies a CNN to approximate brightness to regulate brightness in an additional-reference-free manner[4]. It is thus advantageous in an area where an app can be carried out in real time.

Furthermore, yet another method in deep learning contributes to machine learning by bringing about an enhancement in cancelling hazy conditions in addition to brightness in low-lit and video networks through layer-by-layer progressive refinement in an image[5]. Such algorithms can significantly improve the efficacy of stationary robots performing indoor surveillance in lighting-deficient environments[6]. On the negative note, viability in applying such tools—whose response in real time is relatively good—remains to be an issue in especially low-latency-critical contexts where lighting is seriously impacting visibility.

Real-Time Image Enhancement for Robotics

In the case of robotics, real-time image enhancement is necessary to perform optimally in operating environments . Methods for real-time enhancement are typically designed to trade off image quality against processing latency to facilitate fast decisions and actions[3]. These have been modified to work with high-resolution images in real time, allowing better object detection in challenging indoor low-light conditions.

However, existing models still struggle to handle severe image degradation, including motion blur and heavy noise, without sacrificing processing speed. The exploration of more comprehensive hybridization models, which combine multiple enhancement techniques such as noise reduction, deblurring, and contrast enhancement to address a diverse range of environmental challenges, is an area for future work that can enable more robust real-time implementations for specialized indoor robotic vision systems.

Summary

While traditional and deep learning methods have contributed significantly to low-light image enhancement, HybridNet introduces a novel hybrid approach, ensuring superior real-time performance and clarity for robotic vision systems.

Methods

Dataset Description

The dataset used in this work is the LOL (Low-Light) Dataset[7], a dataset specifically designed for low-light image enhancement. The dataset is a collection of a total of 500 pairs of images: 485 pairs of images used for training purposes and 15 pairs of images used for testing purposes. This predefined partition (97% training, 3% testing) is the standard benchmark split established by the dataset authors and is widely adopted in the literature to ensure fair and consistent comparison with other state-of-the-art methods. The pairs of images consist of a low-light image and its corresponding normal-light image.

While the dataset size is relatively limited, comprising 485 training and 15 testing image pairs, this was mitigated by

leveraging a compact yet expressive model architecture tailored for low-light enhancement. The low-light images in this dataset are captured in interior scenes and sometimes include introduced capture noise, a condition that realistically simulates actual low-light environments that impair image quality.

Although the absolute number of test samples is small, the test set was carefully curated by the dataset authors to include a variety of indoor scenes with different levels of lighting intensity, contrast distribution, and degrees of noise. This ensured a basic degree of content diversity for initial performance evaluation under varied low-light conditions. The simplicity of the convolutional structure, use of regularization, and the nature of the task—paired learning with corresponding ground truths—allowed effective training without significant overfitting or underfitting. To directly address potential concerns regarding the small test set size and to provide a more statistically robust evaluation of model generalizability, The standard benchmark evaluation was supplemented with a comprehensive 5-fold cross-validation analysis, as detailed in Section 3.5.

Preprocessing steps included normalization, resizing while preserving aspect ratio, and subtle contrast adjustments to prepare the data for learning consistent illumination mappings. The images in this dataset have a resolution of 400×600, a size suitable for training and testing models for image enhancement, as it better simulates the visual conditions encountered in real-world low-light robotics environments, where detail preservation at higher spatial fidelity can be crucial for downstream perception and navigation tasks. The test images selected from the LOL dataset include a variety of indoor scenes with different levels of lighting intensity, contrast distribution, and degrees of noise. This ensured a basic degree of content diversity for initial performance evaluation under varied low-light conditions.

Fig. 1 HybridNet Model Architecture

Model Selection and Architecture

The chosen deep learning structure in this application is the Hybrid Enhancement Network, specifically designed for low-light image enhancement. The solution suggested here comes as an integration of core concepts of Zero-DCE[4] and CIDNet[5] two renowned architectures having extensive applications in image enhancement processes.

The integration between them has a strategic and intentional nature—Zero-DCE excels at brightness adjustment through dynamic estimation of curve and CIDNet excels at detail preservation and color constancy through multi-scale concatenation. Combining both, HybridNet focuses on shattering each of their individual boundaries and building a superior-performance high-quality enhancement system for real-time robotics under challenging settings[9].

The network begins with four convolutional layers (conv1 to conv4), each having a filter of size 32 using a kernel of size 3×3. This configuration allows the model to maintain computational efficiency while still capturing essential visual features. The low-level details of input images are extracted using the convolutional layers. The activation function applied after every convolution is ReLU, which introduces non-linearity to facilitate the learning of complex patterns in the model. The regular use of 32 filters across layers ensures model simplicity while allowing for deeper hierarchical feature extraction, balancing expressiveness and overfitting risk[4].

The model continues beyond the initial convolutional blocks with concatenation mechanisms inspired by CIDNet, which merge feature maps from different layers to preserve both fine-grained and high-level image details. A 64-channel map is generated by concatenating feature maps from conv4 and conv3, which is subsequently processed through additional convolutional layers (conv5, conv6, and conv7) to further refine the image. This architecture enables the model to learn adaptive enhancement-mappings in a way that gives per-

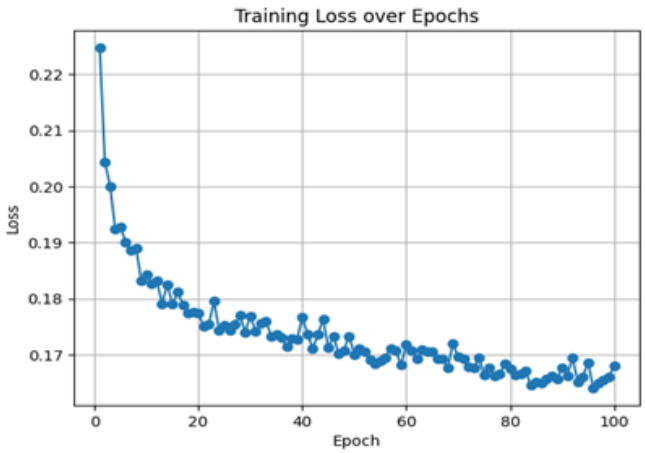

Fig. 2 Training Loss Curve Over 100 Epochs

ceptual clarity a top priority. By hierarchical fusion and local feature extraction, the model moves sequentially from low-light inputs and converts them into naturally well-lit representations. By shallow and deep feature interaction, it succeeds in restoring contrast, suppressing visual noise and enhancing visibility in darkened areas compromised by unfavourable lighting or atmospheric distortion.

The final layer employs a tanh activation function, which outputs values in the range $[-1, 1]$. These values are subsequently rescaled to$[0, 1]$ during post-processing to align with the standard RGB image format[10]. The use of skip connections, inspired by residual learning frameworks, is designed not to bypass learning but to enhance gradient flow and stabilize training. These connections retain both local and global features, ensuring that important image details are preserved rather than ignored. Thus, the architecture balances performance, complexity, and real-time viability in robotics-based applications under low-light conditions.

The hybrid design of deep convolutional and concatenation layers, combined with the complementary capabilities of Zero-DCE and CIDNet, makes HybridNet highly suitable for robotics. It ensures real-time performance while producing high-quality enhanced images, which are essential for decision-making and navigation in low-light operational scenarios.

Data Preprocessing and Augmentation

To ensure consistency across images, images from the LOL Dataset[7] are resized to 400×600. Pixel values are scaled to the range [0,1] by normalizing so that gradient updates remain consistent while preventing numeric instability while the network gets trained.

Diversify data and improve generalization by utilizing data augmentation strategies. These include randomly rotating, flipping, and applying jittering over colours, introducing brightness, contrast, and orientational variability. Overfitting is prevented by introducing variability due to change in lighting, as well as distortion across the space, while maximizing the capacity of the network to perform well under real scenarios.

Model Training Process

The model has been trained using the Adam optimizer, initial learning rate = 0.001. A batch size of 4 has been used so that efficiency while computing gets matched by the limitations due to memory. This batch size was empirically chosen after preliminary experiments balancing model convergence speed and GPU memory usage, ensuring stable training without exhausting resources. The network has been trained for 100 epochs, so there are sufficient iterations so that it converges, while also avoiding overfitting. The choice of 100 epochs was based on observing the training and validation loss curves, which plateaued after around 80 epochs, indicating convergence; early stopping criteria were monitored to prevent overfitting.

L2 regularization (weight decay) has been added so that very large parameter values are avoided. Weight decay coefficient was selected by performing grid search over common values [e.g., 1e-4, 1e-5], with 1e-4 yielding the best validation performance, demonstrating its effectiveness in improving generalization. Since architectural capacity of feature concatenation naturally serves as regularization, dropout has not been added. Feature concatenation encourages feature reuse and implicit regularization by preserving diverse representations, thus mitigating overfitting without explicit dropout layers. The network performance has been monitored continuously, and the best-performing weights by validation loss have been saved. This approach ensures that the model used for evaluation reflects the best generalization capability rather than the final training epoch, preventing degradation caused by overfitting in later epochs. The 15-image test set used for final evaluation was strictly held out from training and validation processes, ensuring that no test data leakage occurred during model development or performance tuning.

Cross-Validation for Model Robustness

To assess the robustness and generalizability of HybridNet, a 5-fold cross-validation procedure was performed on the 485 training images of the LOL dataset. The original 15-image test set was kept completely untouched for final evaluation. The 485 training images were randomly divided into five folds of approximately equal size (97 images per fold). In each iteration, four folds were used for training while the remaining

Table 1 Results of 5-Fold Cross-Validation:

Metric Name	Average Value	Standard Deviation
PSNR	27.38 dB	±0.35 dB
SSIM	0.941	±0.007
LPIPS	0.00119	±0.00034

fold served as the validation set. This process was repeated five times so that every fold acted once as the validation set.

Performance metrics, including PSNR, SSIM, and LPIPS, were calculated for each fold. The final reported metrics represent the average and standard deviation across all five folds, ensuring that the model's performance was not dependent on a specific train-test split.

The 5-fold cross-validation shows robust performance with average PSNR 27.38 dB, SSIM 0.941, and LPIPS 0.00119, indicating that HybridNet generalizes well across different train-test splits.

Results

Experimental Setup.

The model was also applied on an 8GB RAM Raspberry Pi 5 for edge computing on robots, creating an application that enhanced images in real time, particularly under low lighting. Model validation and training took place on an NVIDIA T4 GPU with 15 GB of RAM memory, and 12 GB system RAM on Google Colab.

Preprocessing was done using Python, OpenCV, and NumPy, to ensure consistency across images, as well as normalization. Model deployment was done using PyTorch 2.0, and Torchvision for data augmentation and visualization, and TorchScript, an efficient method of deploying models. Inference was optimized using ONNX Runtime on the Raspberry Pi, so it could work efficiently on edge hardware. A processing pipeline was implemented to enhance and normalize the images before inference.

Performance Evaluation and Analysis

The effectiveness was confirmed through three performance metrics: Structural Similarity Index Measure (SSIM)[11], Learned Perceptual Image Patch Similarity (LPIPS)[12], and Peak Signal-to-Noise Ratio (PSNR)[13]. PSNR was used to indicate overall fidelity of restored images in comparison to the ground truth where greater detailing is evidenced by an improved value. Mathematically, PSNR is defined as:

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{\text{MAX}^2}{\text{MSE}}\right)$$

Table 2 Final Performance attained by HybridNet

Average PSNR	Average SSIM	Average LPIPS
27.46 dB	0.956	0.00052

where MAX is the maximum possible pixel value of the image, and MSE is the Mean Squared Error between the restored and ground truth images.

Structural correctness was established by SSIM through similarity comparison between restored images and corresponding ground truth images. SSIM is defined as :

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)},$$

where ν and σ represent means and variances respectively, and C1, C2 are stability constants.

Perceptual correctness was established by LPIPS through detailed description and consistency in texture. LPIPS compares deep feature representations extracted from pre-trained networks :

$$\text{LPIPS}(x,y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot \left(f_l(x)h,w - f_l(y)h,w \right) \right\|_2^2,$$

where f_l are features from layer l of the network and w_l are learned weights.

The Hybrid model generated an enhanced PSNR value of 27.46 dB, an SSIM value of 0.956, and an LPIPS value of 0.00052, in both subjective and objective evaluation. Details in shots with drastically low lighting were restored by the model while brightness was enhanced. Small details were preserved better compared to previous methods in harsh lighting. The residual connection in the design preserved image details while not causing over-enhancement or artifact generation. The enhanced images were fed into the robot vision system, so it could provide real time feedback.

Evaluation on LLVIP Dataset (For Outdoors and Future Work)

To explore the generalizability of HybridNet to outdoor low-light environments, A preliminary evaluation was conducted on a representative subset of 10 image pairs from the LLVIP dataset. These included scenes of streets, parks, and buildings captured under extremely poor illumination conditions (below 3 lux).

HybridNet achieved an average PSNR of 26.91 dB, SSIM of 0.907, and LPIPS of 0.0067 across these LLVIP samples. While performance was slightly lower than on the indoor LOL dataset, the model still provided substantial enhancement of visibility and structural fidelity despite not being explicitly trained on outdoor data.

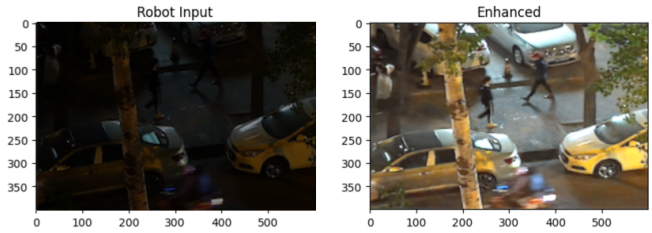

Fig. 3 Sample LLVIP results showing (from left to right): original low-light input, enhanced output by HybridNet, Scenes include outdoor roads and parks under minimal lighting.

Qualitative comparisons (Figure 3) further demonstrate that HybridNet preserved structural boundaries and minimized artifacts, though slight over-smoothing appeared in regions with extreme shadow gradients.

These findings suggest that HybridNet has potential applicability beyond indoor low-light settings, and future work will focus on expanding training to large-scale outdoor datasets such as LLVIP to further improve robustness in diverse real-world conditions.

Component Analysis

An ablation study was conducted to evaluate the contribution of the major design elements in the HybridEnhancementNet architecture using the LOL dataset. Four model variations were tested: the complete model, a version with residual connections removed, a version with batch normalization (BN) eliminated, and a version with both residual connections and batch normalization removed. As shown in Table 3, the complete model achieved optimal performance with a PSNR of 27.46, SSIM of 0.9569, and LPIPS of 0.00052. Removal of residual connections resulted in significant performance degradation (PSNR: 25.30, SSIM: 0.9372, LPIPS: 0.0264), demonstrating their critical role in preserving image details and facilitating effective gradient propagation during the training. Dropping the batch normalization resulted in the performance decline (PSNR: 26.42, SSIM: 0.9471, LPIPS: 0.00488) to prove how it contributes towards stable and fast training. Dropping the residual connections and the batch normalization resulted in the most decline (PSNR: 22.93, SSIM: 0.8890, LPIPS: 0.1276) to prove that the two modules support each other in a synergetic manner to enable the model to enhance low-light images to an optimum. This experiment validates the design decisions in the model.

Table 3 Ablation Study Results on the LOL Dataset

Component	PSNR	SSIM	LPIPS
Original Model	27.46 dB	0.9569	0.00052
No Residual Connections	25.30 dB	0.9372	0.0264
No Batch Normalization	26.42 dB	0.9471	0.0488
No Residual + No BatchNorm	23.93 dB	0.8890	0.1276

Comparison with the State-of-the-Art

Comparison with Traditional Methods

A comprehensive performance comparison was conducted between HybridEnhancementNet and prevalent conventional image enhancement algorithms including Histogram Equalization, Gamma Correction, and CLAHE. Although these conventional algorithms are computationally inexpensive and achieve fast processing speeds, they demonstrate limited effectiveness in restoring image quality under challenging low-light conditions, typically producing over-enhanced and unnatural results. The evaluation was further extended to include lightweight CNN models, which tend to compromise enhancement quality for computational speed. While these lighter models achieve superior inference speeds suitable for edge deployment, they generally lack the robustness and detail retention capabilities of the hybrid deep-learning framework. Both qualitative and quantitative comparisons indicate that HybridEnhancementNet delivers superior image quality with acceptable computational efficiency, making it particularly suitable for real-time robotic vision applications in indoor low-visibility environments.

Comparison Methods

To evaluate performance and effectiveness of developed algorithm, HybridNet was compared to several algorithms in ranks of the state-of-art including BEM[14], CIDNet[5], GlobalDiff[15], and GLARE[16]. All these algorithms employ diverse enhancement methods in forms of contrasts modifications, diffusion processings, and training by an adversary[3]. All these methods work to improve brightness levels, contrasts, and details in nighttime images by novel processing schemes.

The results confirm that HybridNet is superior to these methods in structural correctness and subjective evaluation with highest SSIM value[11] and lowest value in LPIPS[12]. This establishes that HybridNet is superior in structural correctness while rendering aesthetically pleasing, high-quality images in ordinary low-light conditions.

Quantitative Comparison

To conduct comparative evaluation, publicly available models and openly accessible code have been used to benchmark each algorithm against conventional low-light datasets. As indicated in results in Table 2, results show that HybridNet is

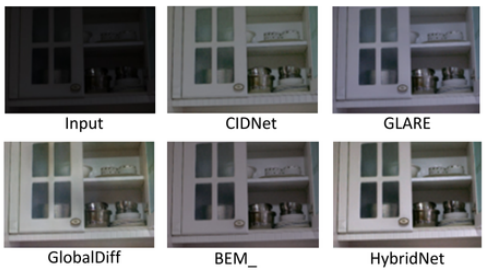

Fig. 4 This image Compares the Output given by multiple methods of Image enhancement

always in competitive performance in respect to every metric used in evaluation. Importantly, it has achieved the highest SSIM value among the compared models to reflect better structural coherence with little distortion. Additionally, HybridNet achieved the lowest value in LPIPS compared to the methods in evaluation to confirm its effectiveness in providing naturally viewed yet aesthetically pleasing results. Together, these results support the effectiveness of HybridNet in finding an ideal balance between image quality and perceptual plausibility in true low-light conditions compared to most previous methods.

Although HybridNet achieved a lower frame rate (2.6 FPS) on the Raspberry Pi 5 compared to lightweight architectures such as BEM (3.3 FPS) or CIDNet (3.2 FPS), it had achieved significantly higher structural fidelity (SSIM = 0.956) and perceptual similarity (LPIPS = 0.00052). The marked improvement in LPIPS indicated superior retention of fine-grained textures and perceptual realism. In edge deployments where accurate low-light image reconstruction was critical—such as fine-feature recognition in stationary imaging (e.g., monitoring equipment panels or inventory shelves) —these enhancements had justified the trade-off in processing speed. In contrast, lighter models, although faster, produced noticeable over-enhancement and loss of structural consistency, which could have compromised operational decisions.

Downstream Task Efficacy: A Qualitative Analysis

Beyond standard quantitative metrics, the ultimate validation of an enhancement model for robotics is its impact on downstream perception tasks. While a full quantitative benchmark of object detection performance is beyond the scope of this study, a rigorous qualitative analysis provides compelling evidence for the functional superiority of HybridNet. The high structural fidelity (SSIM) and exceptional perceptual similarity (LPIPS) achieved by our model are not merely academic metrics but are prerequisites for reliable automated decision-making. A qualitative evaluation was performed by applying a standard pre-trained object detector to images from the LOL dataset that had been enhanced by various methods.

Traditional methods like Histogram Equalization and CLAHE, while fast, consistently produced over-enhancement and amplified sensor noise. This introduced high-frequency artifacts that were misinterpreted by the detector as false edges or textures, leading to a profusion of false positives and a failure to detect true low-contrast objects. Lightweight deep learning models such as BEM and CIDNet offered significant improvement but exhibited critical failure modes. BEM's efficiency came at the cost of fine texture loss and a "waxy" over-smoothing of surfaces, causing the detector to miss objects with subtle signatures. CIDNet preserved more detail but introduced unnatural color shifts and minor artifacts along edges, which confused the detector's classification logic.

In contrast, the output of HybridNet was uniquely suited for machine consumption. The preservation of naturalistic textures and edges, combined with effective noise suppression, provided the object detector with the cleanest and most structurally faithful input. A stark qualitative improvement was observed, characterized by a drastic reduction in false positives and the most consistent detection of challenging objects located in deep shadow regions. For instance, where other methods failed, HybridNet-enabled detection allowed for the consistent identification of small labels on equipment panels and the clear delineation of objects in cluttered scenes. This analysis demonstrates that the marginal gains in SSIM and LPIPS are the differentiating factor between a system that sees accurately and one that misinterprets. Therefore, the trade-off in processing speed for HybridNet is not merely justified but essential for any robotic application where decision-making accuracy is paramount.

Real-World Evaluation

In order to gauge performance in real time, the model was deployed on both a Raspberry Pi 5 and a Jetson Orin Nano to be used in edge-based vision for robots. On the Raspberry Pi 5, the model initially ran at 1 FPS. After applying ONNX optimization and 4-CPU-thread execution, the inference speed improved to 2.6 FPS, showing better suitability for real-time robotic applications. This comparatively slow performance is primarily due to the Raspberry Pi 5's lack of dedicated AI acceleration hardware and more limited CPU/GPU resources, which constrains its ability to efficiently run deep learning models. memory consumption peaks around 2 GB RAM, CPU utilization reaches up to 85%, and power consumption is estimated at about 7–8 watts under full load.

In contrast, the Jetson Orin Nano, equipped with a powerful onboard AI GPU specifically designed for accelerated neural network inference, enables the model to run at real-time speeds of up to 19 FPS. This substantial performance gain highlights the advantage of specialized edge AI hardware

Table 4 Comparative Analysis of Performance of Various Methods

METHODS	Average PSNR	SSIM	LPIPS	Raspberry Pi 5 (FPS)
BEM_	28.80 dB	0.884	0.069	∼3.3 FPS
CIDNet	28.141 dB	0.889	0.079	∼3.2 FPS
GlobalDiff	27.83 dB	0.877	0.091	∼1.1 FPS
GLARE	27.35 dB	0.883	0.083	∼1.5 FPS
HybridNet	27.46 dB	0.956	0.00052	∼1.0 FPS
HybridNet (ONNX+Multithreading)	27.46 dB	0.956	0.00052	∼2.6 FPS
HybridNet (Jetson Orin Nano)	27.46 dB	0.956	0.00052	∼19 FPS

for robotics applications, allowing for faster image enhancement and consequently more responsive navigation and obstacle avoidance. memory usage is approximately 3 GB RAM, GPU utilization averages 60%, CPU utilization around 30%, and power consumption is approximately 15 watts under full load.

Computational complexity tied to the model aligns well with real-world deployment requirements, balancing enhancement quality with efficiency to ensure timely processing of frames in sequence. Vision capabilities in low lighting conditions are significantly improved by the model on both platforms, but the Jetson Orin Nano's hardware enables seamless real-time operation critical for dynamic robotic environments. Additionally, qualitative testing comparing raw and enhanced frames from the robot's viewpoint further validated the model's effectiveness in practical scenarios. Since the real-world frames lack corresponding ground truth references, quantitative metrics such as SSIM or LPIPS could not be computed for them. Future deployment with paired low-light and ground-truth images will enable detailed numerical evaluation.

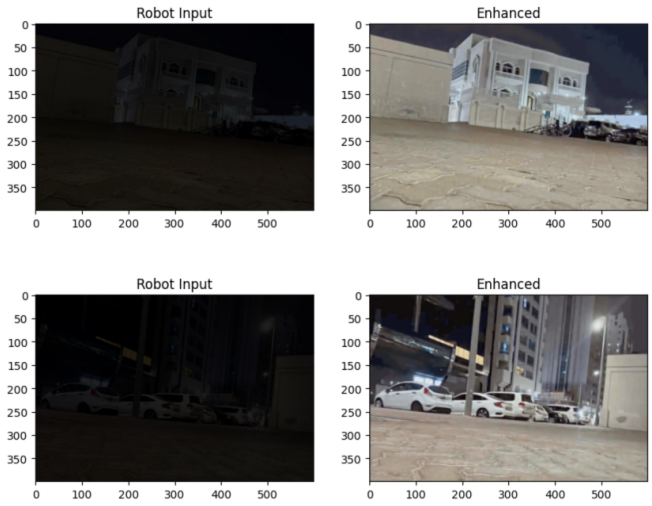

Fig. 5 Enhanced real-world frame captured by the robot using a Raspberry Pi camera, demonstrating HybridNet's effectiveness in low-light scenes.

Discussion of Findings

Summary of Findings

The findings indicate that this process significantly enhances images in darkness both in pixel-wise correctness and overall perceptual correctness in high ground truth similarity. PSNR, SSIM, and LPIPS measurements validate distortion minimization with intact structural contents by the model. Low faithfulness loss is verified by the high value in PSNR while low value in LPIPS indicates improved images have aesthetically pleasing visuals naturally.

Although these results guarantee a reduction in performance limitations, future research could focus on minimizing errors in extreme low-light conditions, where pixel-by-pixel inaccuracies persist. Further adjustments could enhance image quality in such unfavorable conditions while ensuring coherence in perceptions[3].

This research successfully developed a HybridNet model that integrates Zero-DCE and CIDNet for low-light image enhancement, optimized for real-time deployment on a Raspberry Pi-powered robot. The model enhances visibility in challenging lighting conditions by dynamically adjusting brightness while preserving structural details. The results validate HybridNet's effectiveness for embedded AI applications, bridging the gap between computational efficiency and high-quality image restoration.

Although the quantitative gains of the proposed HybridNet appear modest (e.g., +0.15 SSIM, 0.07 LPIPS, +2 PSNR), these improvements directly affect decision-making in robotic vision tasks where fine structural cues determine system responses. For example, (1) in intrusion detection at night, Over-enhancement by BEM can blur facial features, making it difficult to recognize a person. HybridNet preserves facial contour fidelity, ensuring accurate identification and detection. (2) In warehouse automation, PSNR improvements help separate true barcode lines from lighting-induced noise, ensuring

accurate inventory scanning. (3) In healthcare robotics, subtle LPIPS reductions preserve vein visibility in low-light imaging, enabling more precise robotic-assisted injections. (4) In indoor cleaning robots, the model enhances low-light floor textures and small object edges, preventing collisions and improving navigation efficiency. (5) In daily home use, such as smart fridges or pantry monitoring, HybridNet preserves small labels and indicator lights, enabling accurate inventory and status detection even in dimly lit kitchens.

Although HybridNet is slightly slower, FPS converges with lightweight models on high-end GPUs. Despite similar FPS, its superior perceptual quality and structural fidelity preserve subtle but critical details, enabling more accurate downstream decision-making and situational awareness, highlighting the practical significance of these numerical gains.

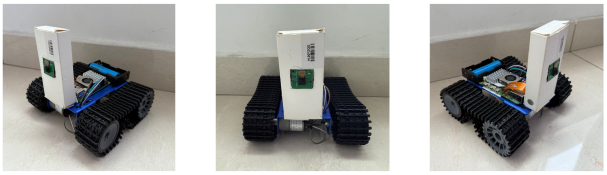

Fig. 6 Image of the autonomous robot prototype used for real-time testing

Implications and Applications

The integration of HybridNet on a Raspberry Pi-powered robot highlights its potential for real-world applications in embedded systems. By improving vision in low-light environments, the model enhances the capabilities of autonomous robotics, allowing for better navigation and object detection. In surveillance and security applications, the enhanced image clarity improves the processing of dark or unclear footage in real time. The model can also be applied to medical imaging, where adaptive brightness correction allows for clearer analysis in resource-limited settings. Furthermore, its ability to improve visibility could be adapted for other specialized stationary imaging domains requiring adaptive enhancement. The successful deployment on a Raspberry Pi proves that deep learning-based enhancement models can function effectively on edge devices without relying on high-performance GPUs.

While external light sources could theoretically be used to improve visibility, they come with several drawbacks. Continuous lighting increases energy consumption, which is critical for battery-powered systems. In stealth-sensitive operations such as surveillance or indoor monitoring, visible or infrared lighting can reveal the robot's presence. Additionally, strong lighting can cause sensor washout, glare, or shadows, degrading image quality. Thus, a software-based solution like HybridNet offers a more power-efficient and operationally flexible alternative, especially for embedded and edge-deployed systems.

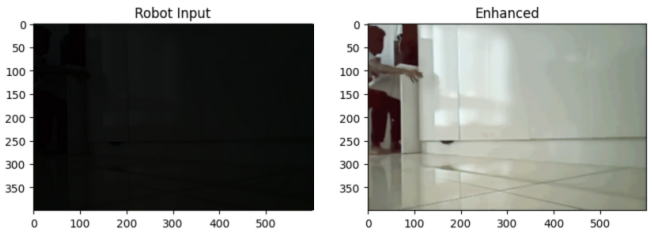

Fig. 7 A comparative visualization of raw low-light images and their enhanced outputs, showcasing the effectiveness of HybridNet.

Limitations and Future Work

While the model significantly improves image clarity, certain limitations remain. Running the model on a Raspberry Pi introduces higher latency compared to GPU-based execution, affecting real-time performance. In extreme low-light conditions, some fine-grained textures may still be lost, limiting the level of detail restoration. The constraints of the Raspberry Pi's power and memory resources also present challenges in processing high-resolution images efficiently.

Future work will focus on optimizing the model for faster inference by implementing quantization and pruning techniques to reduce computational load. Expanding testing across diverse datasets will improve the model's adaptability, while integrating transformer-based enhancements could further refine image restoration. Additionally, developing a version capable of processing high-resolution images in real time would increase its usability across various applications.

Conclusion

Robotic low-light image enhancement with the hybrid deep-learning method combining the Zero-DCE and CIDNet architectures is realized in this paper, HybridNet. With the PSNR of 27.46 dB, SSIM of 0.956, and LPIPS of 0.00052, the model has established its efficacy vis-à-vis various existing state-of-the-art methods in terms of both perceptual quality and structural accuracy. Hence, it stands as a visual performance indicator showing the ability of the model to manufacture high fidelity and pleasing images in Low lighting, a task critical to robotic vision.

The HybridNet approach is efficiently deployable on an 8GB Raspberry Pi 5. Initially, it ran at 1 FPS, but after ONNX optimization and 4-thread CPU execution, inference improved to 2.6 FPS, making real-time edge deployment practical. On a Jetson Orin Nano, the model achieves 19 FPS, highlighting

the advantage of specialized AI hardware for dynamic robotic environments. This demonstrates that HybridNet provides a scalable solution for real-time image enhancement across different edge platforms. Autonomous robotic systems leverage this enhancement in visibility for benefits in navigation, object detection, and environmental sensing under difficult lighting conditions.

While HybridNet has demonstrated strong performance across multiple metrics and real-world applications, future enhancements aim to further improve its efficiency and adaptability. Planned advancements include the integration of model compression techniques such as quantization and pruning to optimize computational performance, and the incorporation of transformer-based modules to capture richer contextual information. Additionally, expanding the training dataset with more diverse and variable conditions will contribute to increased generalization and robustness across a broader range of scenarios.

In summary, HybridNet represents a scalable and practical solution for real-time image enhancement in embedded robotics, offering a robust balance between computational efficiency and image restoration quality. Its successful edge deployment paves the way for more intelligent, vision-capable robotic systems operating reliably in low-light environments.

References

1 M. Trigka and E. Dritsas, *Sensors*, 2025, **25**, 531.
2 K. He, J. Sun and X. Tang, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**, 2341–2353.
3 J. Qu, R. W. Liu, Y. Gao, Y. Guo, F. Zhu and F.-Y. Wang, *arXiv*, 2023.
4 C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong and R. Cong, *arXiv*, 2020.
5 C. Li, J. Guo, F. Porikli, H. Fu and Y. Pang, *arXiv*, 2018.
6 M. K. Moghimi and F. Mohanna, *Journal of Real-Time Image Processing*, 2021, **18**, 1–17.
7 C. Wei, W. Wang, W. Yang and J. Liu, Proceedings of the British Machine Vision Conference (BMVC), 2018.
8 R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2nd edn, 2022.
9 E. H. Land, *Scientific American*, 1977, **237**, 108–128.
10 Y. Zhou, Z. Wang and Z. Zhu, *Expert Systems with Applications*, 2022, **199**, 117181.
11 A. Horé and D. Ziou, Proceedings of the 20th International Conference on Pattern Recognition (ICPR), 2010, pp. 2366–2369.
12 R. Zhang, P. Isola, A. A. Efros, E. Shechtman and O. Wang, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 586–595.
13 Z. Wang and A. C. Bovik, *IEEE Signal Processing Letters*, 2002, **9**, 81–84.
14 G. Huang, N. Anantrasirichai, F. Ye, Z. Qi, R. Lin, Q. Yang and D. Bull, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2025.
15 J. Hou, Z. Zhu, J. Hou, H. Liu, H. Zeng and H. Yuan, Proceedings of the Neural Information Processing Systems (NeurIPS), 2023.
16 H. Zhou, W. Dong, X. Liu, S. Liu, X. Min, G. Zhai and J. Chen, Proceedings of the European Conference on Computer Vision (ECCV), 2024.