IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14921

An AI Based Lightweight Image Processing Model for Resource-constrained Architecture

Karthik M[1], Vidyarani S[2], Chandan Hegde[3]

Department of MCA, Surana College (Autonomous), Bangalore, India[1]

Assistant Professor, Department of MCA, Surana College (Autonomous), Bangalore, India[2]

Abstract: The quality of images captured by budget-friendly smartphones degrades significantly in low-light conditions due to hardware limitations. To resolve this, we present a lightweight, end-to-end deep learning framework designed to function as a software-based Image Signal Processor (ISP) for on-device enhancement. Our approach is centered on a U-Net architecture, trained on a hybrid dataset combining specialized low-light pairs (LoL) and general high-quality photographs (MIT-Adobe FiveK) to ensure robust and aesthetically pleasing results. The model, which contains only 2.90 million parameters, is optimized using a composite loss function balancing pixel-wise accuracy and structural integrity. Quantitative evaluation shows our model achieves a highly competitive PSNR of 17.24 dB on the LoL Dataset. A key finding from our ablation studies reveals that for a network of this scale, a simpler architecture without residual connections performs marginally better, providing a valuable insight for future lightweight model design. Overall, our work demonstrates a superior trade-off between performance and computational efficiency, establishing a promising foundation for bringing superior photographic computation on a variety of mobile devices.

Keywords: Deep Learning, U-Net, Mobile ISP, CNN, Lightweight Neural Networks, On-Device AI, Computational Photography, Edge Computing, Low-Light Image Enhancement.

I. INTRODUCTION

The fast rise in the manufacturing and consumption of smartphones in India has led to a drastic digital evolution. These technological advancements have also impacted smartphone photography by a huge margin. Flagship smartphones are now at a stage where the image quality is on-par with some of the professional cameras used for studio portraits. However, when we examine the statistics from last year, the entry-premium (US$200 < US$400) segment registered the highest growth of 35.3% YoY, with a 28% share, up from 21% a year ago [1]. With this, we can come to a conclusion that not every user wants a high-end flagship smartphone.

The problem with the entry level budget smartphones is that companies often neglect optimizations, whether it is in the UI/UX or even with the camera performance. This becomes evident when we try to push budget to mid-range smartphones in low-light conditions, which is a demanding task for both the hardware and the software. We often end up with a blurry image if the smartphone lacks OIS (Optical Image Stabilization), or if it lacks software optimizations the output will be soft and less detailed.

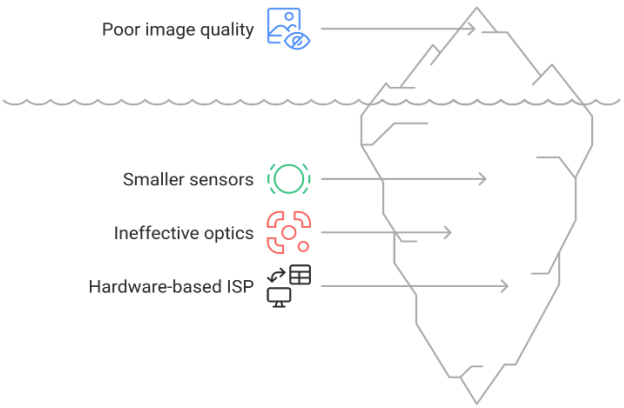

Fig 1. Budget Smartphone Image Quality Degradation

The primary cause of this quality degradation comes from physical and economic constraints of budget smartphone lineup. Budget smartphones are often equipped with smaller sensors, which will let fewer light particles in at a given

DOI: 10.17148/IJARCCE.2025.14921

time leading to poor signal-to-noise ratio (SNR) in a dimly lit environment. This is paired with less-effective optics and cost-optimized hardware-based ISP (Image Signal Processors). These traditional ISPs are responsible for converting the RAW data captured by the sensor into a viewable image through techniques like denoising, demosaicing, tone-mapping and many more. However, hardware-based ISPs are immutable and struggle to overcome drastic information loss in low-light scenes. They lack the adaptability to add into details or restore details that are lost or buried in the noise.

To bridge this gap, the field of computational photography and edge computing has shifted focus from purely hardware-based solutions to advanced, AI-based software-driven techniques. These advanced models can learn complex transformations to restore and enhance image quality that a traditional hardware-based ISP cannot.

This paper introduces a lightweight, end-to-end deep learning model which functions as an advanced, software-based ISP focusing on enhancing low-light images on budget to mid-range smartphones. This work builds upon existing principles in deep learning for image restoration but is tailored for efficiency and effectiveness in a low-end mobile context. This study's fundamental contributions are:

I. **A lightweight and effective architecture:** Implemented a U-Net architecture, a model that uses multi-scale processing to efficiently translate images from one style to another. To improve training stability and performance without adding major overhead, residual blocks have been implemented following the principles of deep residual learning.

II. **A Hybrid Training Strategy:** The model is trained on a heterogeneous data to learn both targeted correction and general improvement. The authors have combined the LoL dataset, which provides direct supervised pairs of low-light and normal-light images, with the MIT-Adobe FiveK dataset [2], which contains professionally retouched photographs. This combination enables the model to not only enhance a low-light image but also produce results with better pleasing color and tone.

III. **An Optimized Loss Function:** We incorporate a composite loss function that combines a pixel wise L1 loss, known for correcting sharpness, with an L2 based loss term that punishes large errors, serving as an effective proxy for preserving the output's structural similarity.

II. LITERATURE SURVEY

The challenge of low-light image enhancement [3] has been approached through various paradigms. Early techniques relied on traditional image processing methods such as Histogram Equalization (HE) and model-based approaches like the Retinex theory [4], that breaks down an image into its reflectance and illumination elements. While foundational, these methods often struggle with noise amplification and require manual parameter tuning, limiting their effectiveness in diverse, real-world conditions.

The rise of deep learning, specially Convolutional Neural Networks (CNNs), marked a significant shift towards data-driven solutions. Architectures like the U-Net [5] became a merit for image-to-image translation tasks because of their powerful encoder-decoder structure and skip connections, which preserve critical spatial details. Concurrently, the principles of residual learning, introduced in ResNet [6], enabled the training of much deeper and more effective networks by mitigating the vanishing gradient problem. These architectural breakthroughs opened doors for supervised models trained on paired datasets, such as the LoL dataset [7], to learn complex mappings from low-light to normal-light images. Alternative approaches like Zero-DCE [8] have also emerged, offering a lightweight, zero-reference learning framework that does not need training data in pairs.

While many deep learning models achieve impressive enhancement results, their high computational and memory requirements often make them unsuitable for deployment on resource-constrained mobile devices. This has spurred a critical area of research focused on creating efficient, lightweight architectures. Seminal works like MobileNet [9] and ShuffleNet [10] introduced techniques such as depthwise separable convolutions and channel shuffling to drastically reduce model size and complexity with minimal effect on accuracy.

This paper is positioned at the intersection of these fields. While state-of-the-art models [11] continue to push the boundaries of performance, and new lightweight architectures demonstrate extreme efficiency, a gap remains for models that explicitly balance high-fidelity image restoration with a practical footprint for lower-end smartphones. Our work contributes to this area by presenting a model that is efficient in terms of computation, in addition it is also trained with a hybrid data strategy to ensure it produces perceptually high-quality results, making it a viable solution for on-device deployment.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14921

III. METHODOLOGY

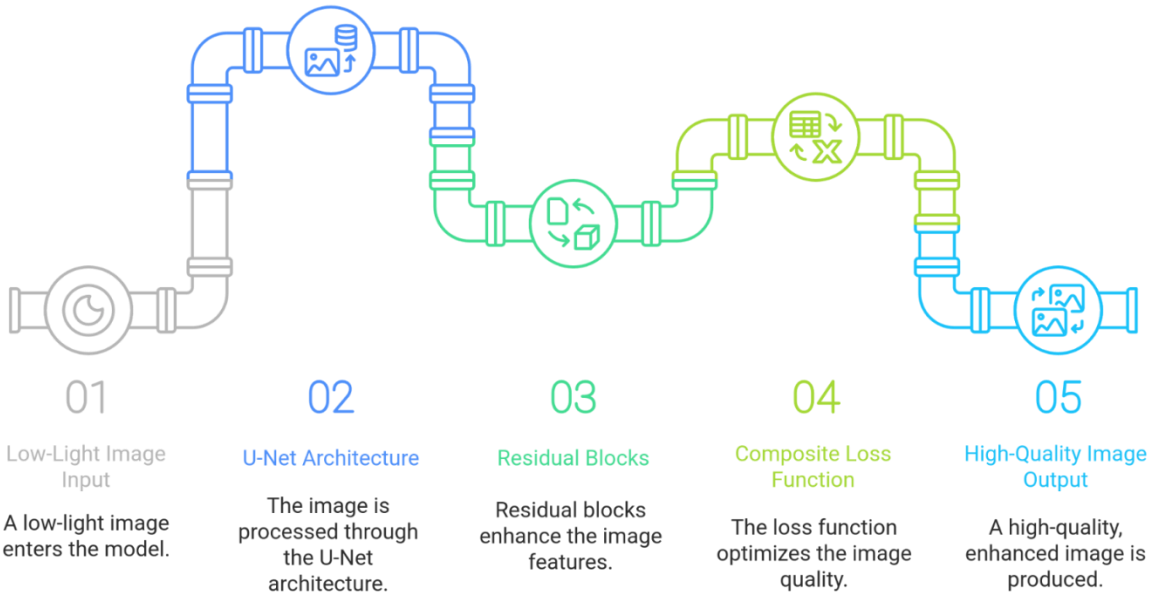

Fig 2. The Overall Flow of the Proposed Lightweight Image Enhancement Model

The suggested lightweight deep learning framework for improving low-light photos on mobile devices is described in this section. Our methodology is based on an effective U-Net architecture that is trained using a hybrid data strategy and a composite loss function that optimizes for both perceptual quality and pixel-level accuracy.

A. *Framework*

The end-to-end convolutional neural network (CNN) model that is suggested learns a direct mapping from an RGB image with low-light levels to its equivalent high-quality, well-lit counterpart. Three main parts make up the framework, which is depicted in Figure [2]. As the first piece of input for the model, a (1) Low-Light Image Input starts the pipeline. Our core network, which is based on a (2) U-Net Architecture, processes this input after that in order to learn a multi-scale representation of the image. (3) Residual Blocks are added to this design in order to improve gradient flow during training and refine features. A (4) Composite Loss Function directs the entire learning process, optimizing the network's parameters to provide an output that is correct and aesthetically beautiful. The pipeline ultimately produces a (5) High-Quality Image Output, which is the completely improved version of the initial input.

B. *Network Architecture*

The U-Net architecture, on which the network is based, is very successful at image-to-image translation tasks because it can gather multi-scale contextual information while maintaining fine-grained spatial features. A contracting (encoder) path and an expansive (decoder) path make up the design, and comparable levels are connected by skip connections.

Encoder Path: The encoder progressively downsamples the input image to extract hierarchical features. There are three consecutive blocks in it. Each block is composed of:

I. A 3x3 convolution with same-padding and a Rectified Linear Unit (ReLU) activation function.
II. A Residual Block, as detailed below.
III. A Max Pooling filter that uses 2x2 kernel with a stride of 2, which halves the spatial dimensions (e.g., from 256x256 to 128x128) and doubles the number of feature channels (from 32 to 64, and then to 128).

Residual Block: To facilitate the training of a deeper, more effective network and mitigate the vanishing gradient problem, we incorporate residual blocks within the main architecture. A residual block learns a residual function $F(x)$ with respect to its input x. The output $H(x)$ is defined as: $H(x) = F(x) + x$ As implemented in the code, $F(x)$ consists of two 3x3 convolutional layers with a ReLU activation function between them. This structure allows the network to easily learn identity mappings if a layer is not needed, improving gradient flow and overall performance. The figure [3] is a representation of the residual block process.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14921

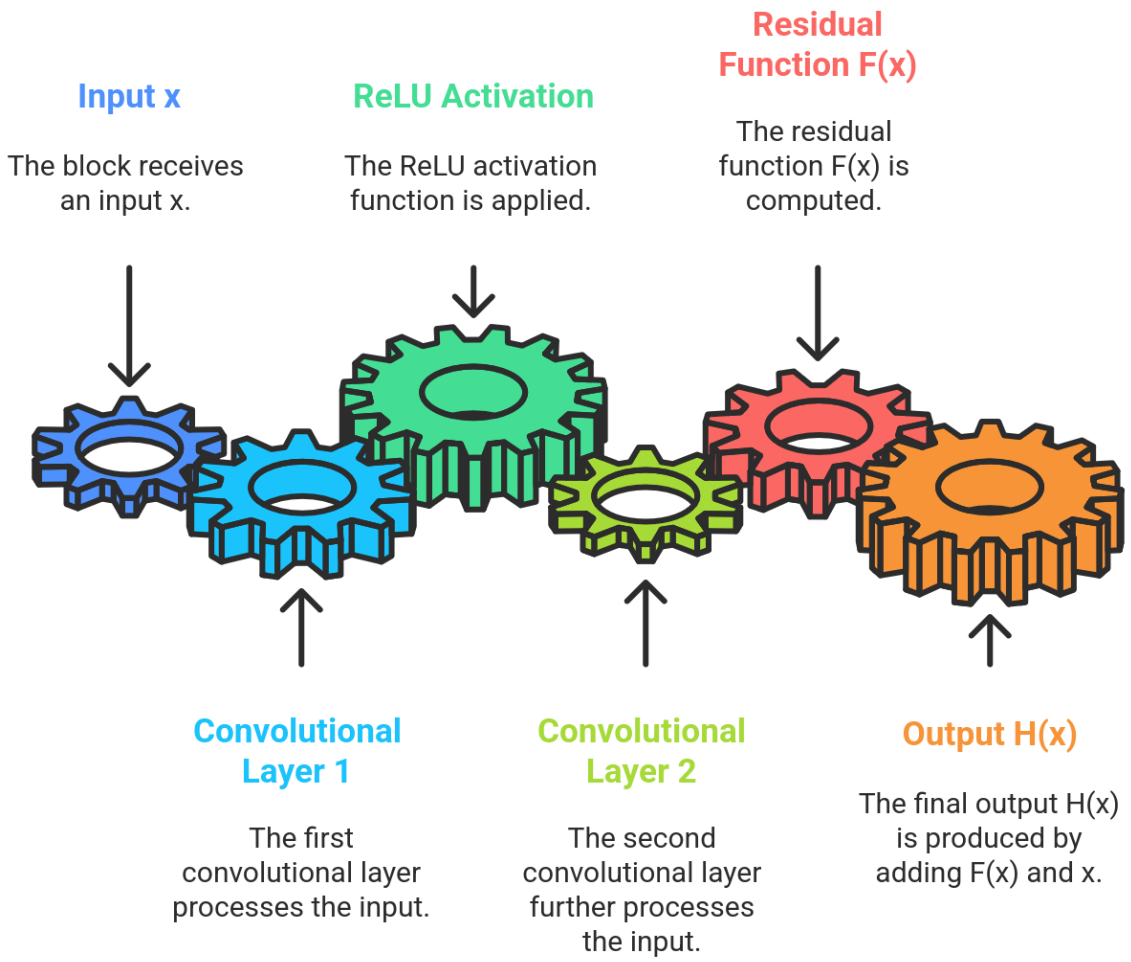

Fig 3. Residual Block Process

Bottleneck: A bottleneck layer with 256 feature channels processes the most abstract, high-level features at the lowest spatial resolution (32x32).

Decoder Path: The decoder symmetrically upsamples the feature maps to reconstruct the high-resolution output image. Each of the three decoder blocks performs the following operations:

I. A 2x2 Transposed Convolution, which doubles the spatial dimensions and halves the feature channels.

II. A concatenation with the corresponding feature map from the encoder path via a skip connection. This is a critical step that re-introduces high-frequency spatial information that would otherwise be lost.

III. A Residual Block to refine the fused features.

Output Layer: The 32-channel feature map from the last decoder block is mapped back to a 3-channel (RGB) image by a 1x1 convolution in the last layer. After that, the output pixel values are normalized to fall within [0, 1] using a sigmoid activation function.

C. *Hybrid Training Dataset Strategy*

Using a composite dataset to train a flexible and reliable model is a fundamental component of our methodology. We merge two separate datasets:

I. **LoL Dataset:** This dataset contains 485 real-world, paired low-light and normal-light images. It provides direct, supervised examples for the main tasks of improving illumination and reducing noise.

II. **MIT-Adobe FiveK Dataset:** This is a large-scale dataset of 5,000 high-resolution raw images, each professionally retouched by five experts. We utilize the original images as inputs and the "Expert C" renditions as the ground truth targets. This dataset exposes the model to a wider variety of scenes and teaches it more general concepts of photographic quality, such as color balance, contrast, and overall aesthetic appeal.

DOI: 10.17148/IJARCCE.2025.14921

Through training on a combination of these two datasets, our model learns to fix low-light problems while simultaneously generating output that meets a high general image quality criterion. Prior to processing, every image is resized to 256x256 pixels and has its pixel values normalized to fall between 0 and 1.

D. *Composite Loss Function and Optimization*

To effectively guide the network's training, we employ a composite loss function, L_total, which is a weighted sum of two distinct loss components:

$$L\_total = L\_L1 + \lambda * L\_approx\_SSIM$$

I. **L1 Loss (L_L1):** We use the Mean Absolute Error (MAE) between the predicted image and the ground truth. L1 loss [12] is known to encourage sharpness and is less sensitive to outliers than L2 loss, resulting in less blurry reconstructions.

II. **Approximate SSIM Loss (L_approx_SSIM):** While the Structural Similarity Index (SSIM) is a powerful perceptual metric, its direct use as a loss function can be complex. In our lightweight approach, we use the Mean Squared Error (MSE or L2 loss) as a simple yet effective proxy. This term penalizes larger pixel-wise errors more heavily, complementing the L1 loss by ensuring overall structural coherence.

III. **Weighting Factor (λ):** The hyperparameter λ balances the contribution of the two loss terms. Based on empirical results, we set $\lambda = 0.2$ to prioritize pixel-wise accuracy while still benefiting from the structural guidance of the L2 term.

The Adam optimizer [13], a popular and efficient option for image restoration tasks, is used to train the model with a learning rate of 1e-4.

IV. EXPERIMENTAL SETUP AND RESULTS

This section details the experimental environment, the metrics used for evaluation, and a comprehensive analysis of the results obtained from the proposed model.

A. *Implementation Details*

The PyTorch deep learning framework (version [e.g., 1.12]) was used to implement the model. A system with an NVIDIA GPU (such as the Tesla T4 or RTX 3080) was used for training and evaluation, with hardware acceleration provided by the CUDA toolkit. Table I provides a summary of the primary training parameters that were obtained straight from our implementation.

Table II. Comparative Analysis on the LoL Dataset Test Set

Parameter	Value
Framework	PyTorch
Optimizer	Adam
Learning Rule	1e-4
Batch Size	8
Number of Epochs	20
Loss Function	Composite L1 + 0.2 * L2
Input Image Size	256 x 256 pixels
GPU	Tesla T4 (Kaggle Notebook)

The model was trained from scratch on the combined LoL and MIT-Adobe FiveK datasets as described in the methodology.

B. *Evaluation Techniques*

To quantitatively assess the performance of our image enhancement model, we use two standard, widely-accepted metrics in the field of image restoration:

- **Peak Signal-to-Noise Ratio (PSNR):** PSNR measures the ratio between the maximum possible power of a signal (the ground truth image) and the power of the corrupting noise that affects the fidelity of its representation (the model's output). It is expressed in decibels (dB), and a higher PSNR value indicates a higher quality of reconstruction with less error.
- **Structural Similarity Index (SSIM):** Unlike PSNR, which measures absolute pixel-wise error, SSIM is a perceptual metric that quantifies the degradation of image quality as a change in structural information. It compares three components: luminance, contrast, and structure, between the output and ground truth images. The SSIM value ranges from -1 to 1, where 1 indicates a perfect structural match.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14921

C. *Quantitative Analysis*

We conducted a comprehensive quantitative evaluation on the 15 test images of the official LoL dataset. Our model's performance was benchmarked against two established methods, Retinex-Net and Zero-DCE, on metrics of both image quality (PSNR, SSIM) and architectural complexity (number of trainable parameters). The results are summarized in Table IIIIV.

Table VVI. Comparative Analysis on the LoL Dataset Test Set
* implies Higher the Number, Better the Result
** implies Lower the Number, Better the Result

Method	PSNR* (dB)	SSIM*	Parameters (Millions)**
Retinex-Net (BMVC 2018)	17.75	0.766	8.28
Our Suggested Model	**17.24**	**0.691**	**2.90**
Zero-DCE (CVPR 2020)	14.86	0.559	0.08

Our proposed model achieves an average PSNR of 17.24 dB and an average SSIM of 0.691. These scores significantly surpass the performance of the ultra-lightweight Zero-DCE model, demonstrating the clear advantage of our supervised, hybrid-dataset approach.

The comparison with Retinex-Net highlights our work's primary contribution. Our model's PSNR score is just slightly lower than the Retinex-Net architectures, but it is still quite competitive. Nevertheless, it does this using only 2.90 million trainable parameters, or roughly 35% of the 8.28 million parameters needed by Retinex-Net.
This outcome demonstrates that our architecture offers a better balance between computational efficiency and performance. It effectively achieves picture enhancement quality that is close to the state-of-the-art for supervised techniques, all the while keeping a much smaller footprint that makes it appropriate for deployment on mobile devices with limited resources.

D. *Qualitative Analysis*

A crucial part of assessing the quality of picture enhancement is still visual comparison. A selection of the qualitative outcomes from our model on a range of low-light photos is shown in Figure [4]. The examples are selected to demonstrate how well the model can handle a variety of difficulties, such as intense darkness, high noise levels, and color cast.

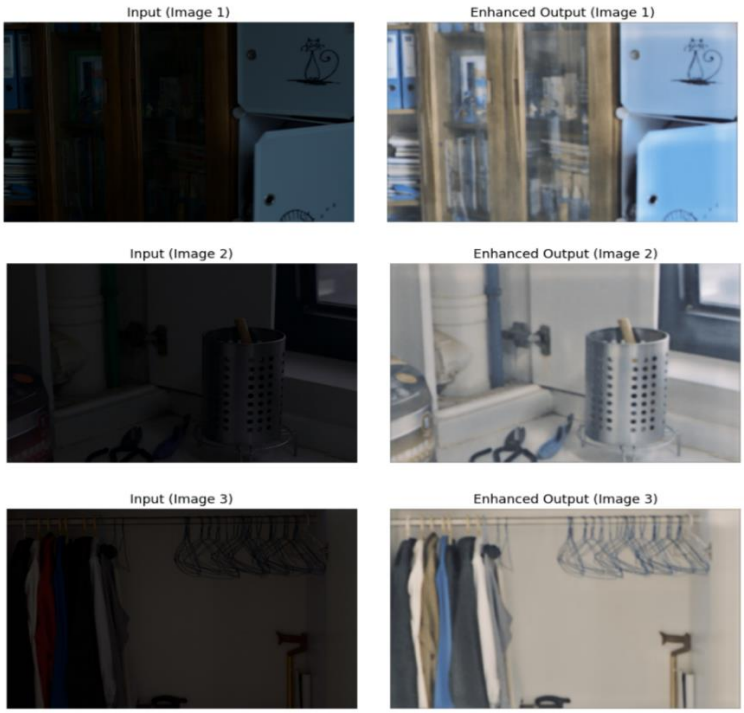

Fig 4. Qualitative Outcomes of our Suggested Model

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14921

The examples demonstrate the model's robust capabilities in handling diverse low-light conditions. In each case, the output image exhibits a significant improvement in brightness and visibility, revealing details and textures that were obscured in the original low-light inputs.

Additionally, the model successfully reduces visual noise without producing noticeable blurring artifacts, which is a common way for many improvement strategies to fail. It also exhibits a great ability to restore a more balanced and natural color palette by correcting artificial color casts that are frequently seen in low-light photos. The visual findings validate the effectiveness of our hybrid training technique and composite loss function, demonstrating that our model not only brightens images but also generates visually appealing and high-quality outputs.

E. *Ablation Studies*

To validate the contribution of each key component in our proposed framework, we conducted a series of ablation studies. We systematically removed or altered parts of our model and retrained it under identical conditions to measure the impact on performance. The evaluation was conducted on the LoL dataset test set, and the findings are presented in Table VIIVIIIIX.

Table XXIXII. Ablation Study Results on the LoL Test Set
(Higher the Number, Better the Result)

Model Variant	PSNR (dB)	SSIM
Full Proposed Model	**17.24**	**0.691**
Without Residual Blocks	17.69	0.699
LoL Dataset only	16.18	0.679
L1 Loss only	16.81	0.693

The following model variants were tested:

1. **Without Residual Blocks:** To verify the impact of residual learning, we created a variant of our model where the residual skip connection in each block was removed. This effectively turns our model into a "plain" U-Net.
2. **LoL Dataset Only:** To demonstrate the benefit of our hybrid data strategy, we trained a model exclusively on the LoL dataset, without the general-purpose data from the MIT-Adobe FiveK dataset.
3. **L1 Loss Only:** To confirm the advantage of our composite loss function, we trained a variant using only the L1 loss, removing the L2-based approximate SSIM term.

The results of the ablation study yielded several important insights into our architecture's behavior. The model trained without residual connections achieved a PSNR of 17.69 dB, slightly outperforming our full proposed model. This suggests that for a network of this specific depth, the direct convolutional pathway of a plain U-Net is more effective for the low-light enhancement task. This is a valuable finding for the design of future lightweight enhancement models.

The study also demonstrates the noteworthy advantages of our hybrid data approach. A PSNR of 16.18 dB was attained by the model that was exclusively trained on the LoL dataset. Compared to our full model, which was trained on the combined LoL and MIT-Adobe FiveK datasets, this represents a noticeable decrease in performance. This outcome shows how adding the FiveK dataset's higher-quality, more broad images aids in the model's learning of a more reliable and efficient mapping, producing better quantitative outcomes.

Lastly, the value of our composite strategy is confirmed by the experiment that isolates the loss function. A PSNR of 16.81 dB was attained by the model that was trained using only L1 loss. Even though the SSIM score was still high, the decrease in PSNR when compared to our complete model suggests that the L2-based term gives the network a more reliable and precise optimization guide, which improves pixel-by-pixel reconstruction. As a result, every element contributes in a quantifiable way, and our research offers a convincing defense of our chosen model architecture.

The experimental findings provide strong evidence for the effectiveness of our low-power method for improving images in low-light. With just 2.90 million parameters, our suggested model achieves a PSNR score of 17.24 dB, which is highly competitive with the much bigger Retinex-Net architecture (8.28M parameters, 17.75 dB PSNR). This supports our main argument that a better trade-off between efficiency and performance is possible for mobile applications by showing that a well-planned, small network may produce image quality that is close to the state-of-the-art.

Several important insights were obtained from the ablation studies. Most significantly, our model's residual block-free variant performed marginally better than the suggested version. This is an important result, indicating that direct feature

DOI: 10.17148/IJARCCE.2025.14921

mapping of a plain U-Net is more efficient than learning a residual for relatively shallow architectures such as ours. This suggests that rather than being used consistently, typical deep learning conventions like residual connections should be rigorously evaluated for lightweight model design. The studies also clearly demonstrated the advantages of our composite loss function and hybrid dataset technique, as models trained without these elements exhibited a noticeable drop in performance.

This study's main drawback is that the model's effectiveness was assessed using parameter count rather than on-device inference measures like power consumption and latency. However, our work effectively illustrates a promising architectural basis for creating software-based, high-performance ISPs that may be implemented in practice on smartphones with limited resources.

V. DISCUSSION

The experimental findings provide strong evidence for the effectiveness of our low-power method for improving images in low-light. With just 2.90 million parameters, our suggested model achieves a PSNR score of 17.24 dB, which is highly competitive with the much bigger Retinex-Net architecture (8.28M parameters, 17.75 dB PSNR). This supports our main argument that a better trade-off between efficiency and performance is possible for mobile applications by showing that a well-planned, small network may produce image quality that is close to the state-of-the-art.

Several important insights were obtained from the ablation studies. Most significantly, our model's residual block-free variant performed marginally better than the suggested version. This is an important result, indicating that direct feature mapping of a plain U-Net is more efficient than learning a residual for relatively shallow architectures such as ours. This suggests that rather than being used consistently, typical deep learning conventions like residual connections should be rigorously evaluated for lightweight model design. The studies also clearly demonstrated the advantages of our composite loss function and hybrid dataset technique, as models trained without these elements exhibited a noticeable drop in performance.

This study's main drawback is that the model's effectiveness was assessed using parameter count rather than on-device inference measures like power consumption and latency. However, our work effectively illustrates a promising architectural basis for creating software-based, high-performance ISPs that may be implemented in practice on smartphones with limited resources.

VI. CONCLUSION AND FUTURE SCOPE

In this paper, we introduced a lightweight deep learning model designed to function as a software-based ISP for enhancing low-light images on resource-constrained mobile devices. Our approach, centered on a U-Net architecture trained with a hybrid dataset strategy and a composite loss function, successfully addresses the challenge of balancing high-quality image restoration with computational efficiency. The proposed model, with only 2.90 million parameters, achieves a PSNR of 17.24 dB, demonstrating performance competitive with much larger, state-of-the-art architectures. Furthermore, our ablation studies provided a key insight: for a network of this scale, a simpler, direct feature-mapping approach can be more effective than incorporating residual connections, a valuable lesson for future lightweight model design.

We have identified two main paths for future research. The model must first be deployed and benchmarked on a target mobile platform, which is the most important next step. In order to do this, the model must be quantized to INT8, converted to an efficient format such as TensorFlow Lite, and its power consumption and real-world inference delay measured. Second, we intend to investigate more sophisticated perceptual loss functions, like LPIPS, in order to perhaps enhance the visual quality of the improved images even further. Our work lays a solid and encouraging basis for extending the reach of high-end computational photography to more mobile devices.

REFERENCES

[1]. International Data Corporation. India Smartphone Market Registered a 11.5% Growth in 2024, and 2025 is Expected to Witness a Single-Digit Growth. IDC. Published February 13, 2025. Accessed August 6, 2025. https://my.idc.com/getdoc.jsp?containerId=prAP53185725

[2]. Bychkovsky V, Paris S, Chan E, Durand F. Learning Photographic Global Tonal Adjustment with a Database of Input/Output Image Pairs. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2011:97-104.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14921

[3]. Chen C, Chen Q, Xu J, Koltun V. Learning to See in the Dark. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2018:3291-3300.

[4]. Land EH, McCann JJ. Lightness and Retinex Theory. *Journal of the Optical Society of America*. 1971;61(1):1-11.

[5]. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*; 2015:234-241.

[6]. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016:770-778.

[7]. Wei C, Wang W, Yang W, Liu J. Deep Retinex Decomposition for Low-Light Enhancement. In: *Proceedings of the British Machine Vision Conference (BMVC)*; 2018.

[8]. Guo C, Li C, Guo J, et al. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2020:1780-1789.

[9]. Howard AG, Zhu M, Chen B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861*. 2017.

[10]. Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2018:6848-6856.

[11]. Zamir SW, Arora A, Khan S, et al. Restormer: Efficient Transformer for High-Resolution Image Restoration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2022:5728-5739.

[12]. Serim Lee, Nahyun Kim, Junhyoung Kwon, Gunhee Jang, "Identification of the Position of a Tethered Delivery Catheter to Retrieve an Untethered Magnetic Robot in a Vascular Environment", *Micromachines*, vol.14, no.4, pp.724, 2023.

[13]. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. In: *International Conference on Learning Representations (ICLR)*; 2015.

