The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

HybridEnhancementNet: A Novel Deep Learning Architecture for Low-Light Image Enhancement through Model Fusion

Vihaan R. Mishra (author)1

1Department of Computer Science, Los Altos High School, Los Altos, United States

Abstract

Deep learning models replicate key aspects of human cognition, enabling powerful capabilities in classification, recognition, and generation. In areas like vehicle autonomy, computer vision and artificial intelligence (AI) play a critical role—especially in low-light conditions where enhanced visual input can improve safety and performance.

A promising strategy in model development is combining complementary architectures to leverage their individual strengths. This work introduces HybridEnhancementNet (HybridNet), a novel low-light image enhancement framework that fuses the convolutional enhancement and ReLU activation function of the Zero-Reference Deep Curve Estimation (Zero-DCE) model with the feature concatenation and skip-connections of the Color and Intensity Decoupling Network (CIDNet) model. The model is trained on the Low-Light (LOL) dataset, consisting of 500 image pairs captured in low-light and well-lit conditions. Experimental results demonstrate that

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

HybridEnhancementNet produces enhanced images with improved visibility while preserving detail and smooth contrast transitions. The model showed robust generalization, delivering strong performance on the 6,000-image DarkFace low-light face-detection dataset, despite having no prior exposure to it. These findings highlight the potential of hybrid deep learning architectures for real-world computer vision applications, including autonomous navigation under challenging lighting conditions. Future work should further explore model fusion strategies to optimize performance and robustness across diverse visual environments.

Keywords: Low-Light dataset, novel architecture, deep learning, computer vision, artificial intelligence

Introduction

Image processing plays a central role in modern computer vision systems, enabling applications such as autonomous navigation, surveillance, medical imaging (Siddiqi & Javaid, 2024), and consumer photography. Deep learning has changed the way computers work with images (He et al., 2016), but a fundamental challenge within this field is enhancing images captured in low-light conditions, where poor illumination leads to reduced contrast, amplified noise, color distortion, and the loss of fine structural details. These degradations make it difficult for both humans and machine-learning models to interpret scene content accurately, and traditional enhancement techniques often fail because the underlying degradations are nonlinear, spatially varying, and tightly coupled (Land & McCann, 1971; Jobson et al., 1997).

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Deep learning—particularly architectures built on convolutional neural networks (CNNs) (O'Shea & Nash, 2015)—has transformed image enhancement by enabling models to learn complex mappings directly from data. CNNs are capable of extracting multi-scale spatial features, modeling non-linear relationships, and adapting to diverse lighting conditions more effectively than handcrafted methods. Recent advances have introduced specialized architectures that estimate illumination curves (Guo et al., 2020), decouple color and intensity information (Meharaj et al., 2023), or leverage multi-branch feature fusion to recover details that would otherwise remain obscured in low-light environments.

However, no single architecture fully addresses all aspects of low-light degradation. Methods that brighten images may amplify noise, while models that preserve detail may struggle with extreme illumination deficits. This motivates the development of hybrid approaches that combine the strengths of multiple deep-learning architectures. This work introduces HybridEnhancementNet (HybridNet), a model that integrates key components from Zero-DCE and CIDNet to achieve more balanced and effective low-light image enhancement.

In this paper, we investigate the hypothesis that combining features from Zero-DCE (Guo et al., 2020) and CIDNet (Meharaj et al., 2023) will better enhance low-light images than just Zero-DCE or CIDNet.

Literature Review

Deep residual learning is an early implementation of skip-connections to minimize feature loss and has since been used in many different architectures. Specifically for low-light

image enhancement, the dominant strategy has been to implement an encoder-decoder architecture (Ronneberger et al., 2015; Hinton & Salakhutdinov, 2006). The encoder compresses the image, progressively reducing the spatial dimension while expanding the channel dimension to capture higher-level features. Then, the decoder uses the feature details extracted from the encoder to progressively reconstruct the image. Spatial dimensions are incrementally restored while channel counts decrease, resulting in an hourglass-shaped structure. This hourglass-shaped structure is effective in extracting key features and denoising.

Although image processing is a challenging problem, significant progress has been made, specifically in the area of image classification. Deep residual learning for image recognition (He et al., 2016) opened up the door for many classification models, especially in image processing. ResNet introduced skip connections (Pham, 2022), allowing for key features in an image to be maintained throughout the convolutional layers. These residual connections have been implemented in newer models such as MIRNet (Waqas Zamir et al., 2020). In low-light image enhancement tasks, most frameworks trained on datasets like the LOL Dataset use these residual connections with convolutional layers followed by ReLU activations (Agarap, 2018), allowing the network to effectively learn illumination corrections and local features. Two models that use these concepts for low-light image enhancement are Zero-DCE and CIDNet.

Zero-DCE enhances low-light images by estimating an image-specific enhancement curve that directly adjusts pixel intensities. The network predicts curve parameters from the input image and applies them iteratively to increase brightness and contrast while keeping intensities within a valid range. The architecture of Zero-DCE consists of seven convolution layers with

ReLU activation functions and skip-connections to the later feature maps that maintain the details of the original image. It then uses a deep curve estimation to adjust the brightness and intensity of each pixel value, finding the best-fitting enhancement of the image. By preserving already discernible details of the original image, the model can selectively brighten darker regions and recover key features.

CIDNet uses a different approach to low-light image enhancement. The model transforms the image into color maps and an intensity map, which are then applied to an enhancement network. It has an encoder/decoder architecture (dilating image size by ½ each layer 3 times and then scaling image size by 2 until it reverts to the original size), and uses lightweight cross-attention instead of a typical convolution layer. The idea behind attention replacing convolutions is that it can gather information from the entire image at once instead of just nearby pixels, helping the model make more accurate low-light corrections. CIDNet also makes use of the ReLU activation function and skip-connections to improve performance. The final feature maps are then passed through the convolutional layers to produce the enhanced image.

HybridEnhancementNet integrates ideas from both Zero-DCE and CIDNet into a unified architecture, designed to enhance low-light images more effectively than either model alone.

Methodology

Dataset Description

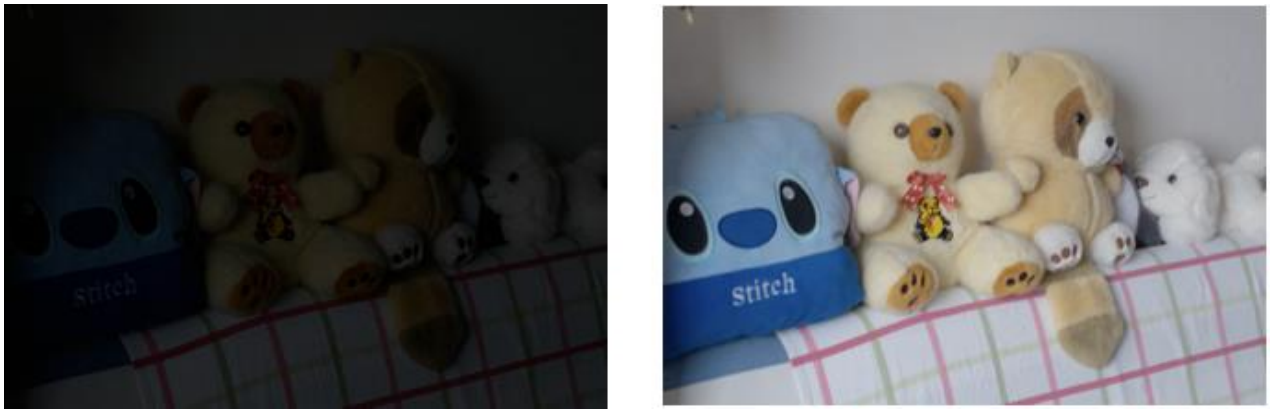
The LOL and DarkFace datasets were used to evaluate HybridEnhancementNet. The model was trained and tested on the LOL dataset, which includes 500 pairs of low-light and

ground-truth images (Wei et al., 2018). Each image in the dataset was captured indoors, featuring household items and appliances such as refrigerators and cabinets, as shown in Figure 1 below. The pictures were collected by varying the camera's ISO and exposure settings while keeping the camera position constant. DarkFace (Yang et al., 2020), a dataset containing 6,000 low-light images of humans captured in real-world environments, was also used to test the generalization capabilities of the proposed model.

Figure 1

Low-light Image to Ground Truth Comparison for LOL Dataset

Low-light vs Ground Truth



Note. Both of these images are from the LOL Dataset. The Low-light image has undergone artificial darkening, while the Ground Truth image represents the original image.

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Model Selection and Architecture

Recently, decomposing an input image into multiple feature maps and pairing this with residual learning has become standard practice in low-light image enhancement. Two prominent architectures that embody these ideas are CIDNet and Zero-DCE. Building on their complementary strengths, the most effective components were extracted from each model and integrated into a unified architecture. This fusion enables a model that is both more efficient and more effective than either approach alone.

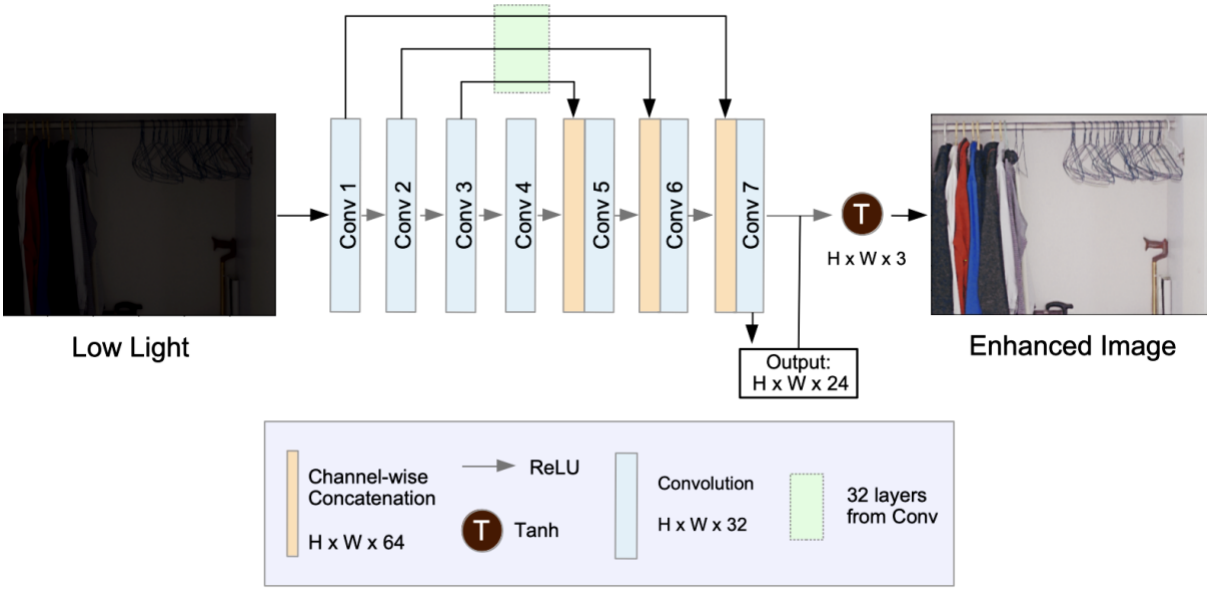
For HybridEnhancementNet, the first 4 convolution layers were taken from Zero-DCE to brighten and enhance the image. Each convolution is followed by a ReLU activation, which suppresses negative activations and preserves positive ones, enabling the network to emphasize useful features while discarding uninformative responses. HybridEnhancementNet also uses a hyperbolic tangent activation function in the output layer, similar to Zero-DCE, for controlled enhancements and smoother images. CIDNet's feature concatenation mechanism was adopted to introduce multi-scale information, allowing the model to learn both low and high-level image features. This idea from CIDNet was integrated into HybridEnhancementNet, and concatenation-based residual connections were introduced from the first 4 convolutions into the next 4 convolutions.

One of the key differences between Zero-DCE and CIDNet is the number of channels. While Zero-DCE maintains the same channels throughout the enhancement process, the fact that the residual connections in HybridEnhancementNet are concatenation-based results in channel

expansion. Additional convolution layers were introduced to handle the increased number of input channels.

Figure 2

Depiction of HybridEnhancementNet Architecture



Note. HybridEnhancementNet combines the brightness estimation capability of ZeroDCE and the multi-scale feature fusion of CIDNet.

HybridEnhancementNet combines the idea of concatenation from CIDNet's architecture with the repeated convolutional process of Zero-DCE, allowing for early features to be preserved and combined with later image representations for more enhanced image outputs, as shown above in Figure 2.

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Data Preprocessing and Augmentation

In the first step, the RGB images were transformed to 400x600x3 pixels. Then, the images were changed into PyTorch tensors and the data was normalized to increase the speed of model training convergence. An 80-20 train-test split was used with a fixed random seed for reproducibility.

Model Training

HybridEnhancementNet was trained for 200 epochs using Adam optimization (Kingma & Ba, 2014) with an initial learning rate of 1e-3, which was reduced by half every 5 epochs to ensure stable convergence. Mean-Squared Error (MSE) loss was used for training, and Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) for evaluation. PSNR measures how closely the enhanced image matches a reference image by quantifying pixel-level reconstruction accuracy. SSIM evaluates structural similarity, capturing how well the enhancement preserves edges, textures, and overall perceptual quality.

For the DarkFace dataset (Yang et al., 2020), a specialized data pipeline was implemented to automatically parse the custom annotation format, where each text file contains face coordinates and the location of the boxes around each face, with a simple structure. The final model demonstrates reliable enhancement of dark images while maintaining the structural integrity of features critical for downstream tasks like detection and recognition. Given the low memory footprint of HybridEnhancementNet, the model balances computational efficiency with

top-tier image enhancement quality, making it suitable for real-world applications where processing speed and reliability are paramount.

Experimental Results

Experimental Setup

To train all models and the proposed HybridEnhancementNet, Google Colab Pro was used with an NVIDIA A100 GPU. The implementation was developed in Python 3.12 using PyTorch as the primary deep learning framework. Image handling and preprocessing were performed using PIL (Pillow), OpenCV, and NumPy, while Matplotlib was used for visualizing intermediate outputs during training.

Performance Evaluation and Analysis

To ensure the quality of the enhanced images, Mean-Squared Error (MSE) was used. MSE is defined as $MSE = \frac{1}{n}\sum_{i=1}^{n} \quad (Y_i - \widehat{Y}_i)^2$, where n is the number of pixels, Y_i is the observed values, and $\widehat{Y}_i$ is the predicted values by HybridEnhancementNet to calculate the accuracy of the enhanced images. HybridEnhancementNet was able to enhance key parts of an image while maintaining natural dark regions and making shady areas more visible than the ground truth image. This is demonstrated in Figures 3 and 4 below.

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

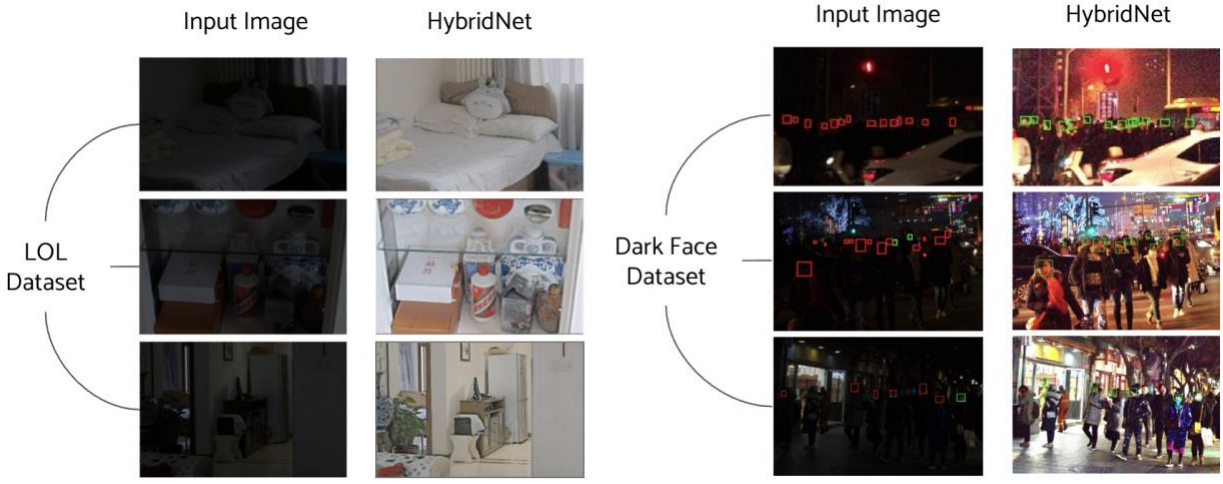
Figure 3

Three-Way Image Comparison on the LOL Dataset

Note. Comparison of Low-light Input Image, Enhanced Output Image, and Ground Truth image on the LOL Dataset.

Figure 4

Input / Output Image Comparison on LOL and DarkFace Datasets



Note. Comparison of input and corresponding enhanced output images from LOL (left) and DarkFace (right) datasets. HybridNet is the short form of HybridEnhancementNet.

To evaluate the model, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) were used to compare HybridEnhancementNet against CIDNet and Zero-DCE. PSNR is measured by the formula $PSNR = 10log_{10}(\frac{MAX^2}{MSE})$, where MAX is the maximum possible value of each pixel. A higher PSNR of 22-28 corresponds to a higher quality of image reconstruction. The formula for SSIM is $SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy}+C_2)}{(\mu_x{}^2+\mu_y{}^2+C_1)(\sigma_x{}^2+\sigma_y{}^2+C_2)}$ where μ_x and μ_y are the mean brightness value of the first image and second image, respectively; $\sigma_x{}^2$ and $\sigma_y{}^2$ are the variances of each image; σ_{xy} is the covariance between the two images, and $C_1 = K_1L$ and $C_2 = K_2L$, where $K_1 = 0.01, K_2 = 0.03$, and $L = 255$. The number 255 represents the range of possible pixel values. An SSIM of 1 means that the input and output images are identical to the reference image in terms of structure, luminance, and contrast. In Table 1 and Figure 5, it is demonstrated that HybridEnhancementNet outperformed Zero-DCE and CIDNet on the LOL Dataset.

Table 1

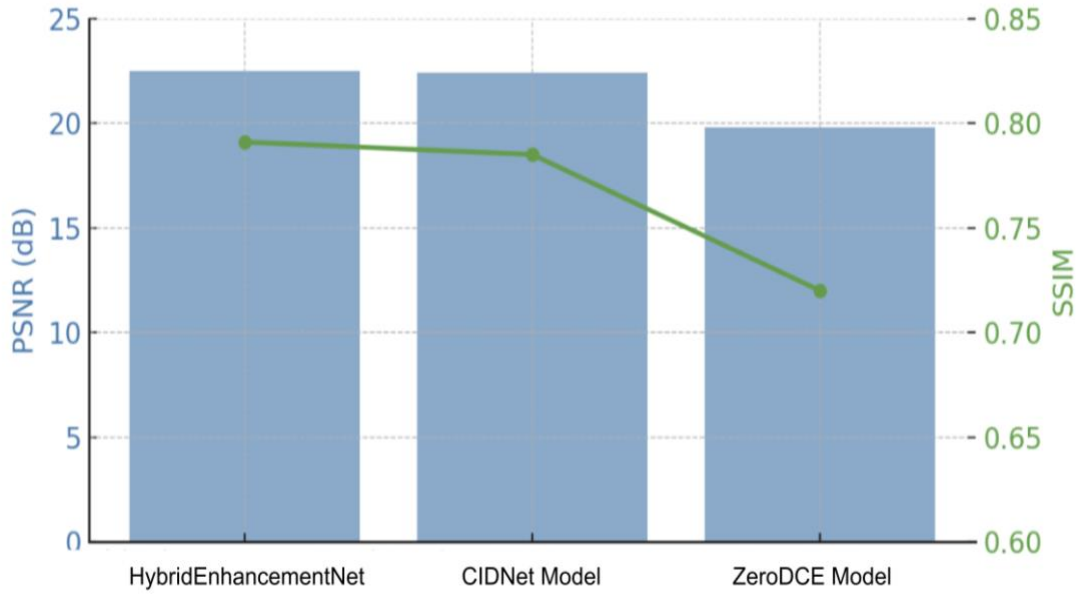
Model Evaluation Table: Results on the LOL Dataset

Model	PSNR (dB) ↑	SSIM ↑
HybridEnhancementNet	22.51	0.791
CIDNet	22.40	0.785
Zero-DCE	19.80	0.720

Note. This table compares HybridEnhancementNet's effectiveness at brightening dark images to CIDNet and Zero-DCE on the LOL Dataset.

Figure 5

Comparison of Enhancement Methods



Note. Comparison plot showing PSNR (dB) as bars and SSIM as a line for the image three enhancement methods on the LOL Dataset.

Discussion

The quantitative evaluation demonstrates the effectiveness of the proposed HybridEnhancementNet model for low-light face enhancement on the DarkFace dataset. As shown in the prior data tables, HybridEnhancementNet achieved competitive performance, outperforming both CIDNet and Zero-DCE across standard image quality metrics. In some cases,

the enhanced image can reveal details even more clearly than the ground truth. HybridEnhancementNet was able to selectively brighten only the dark parts of the image. It is important to mention that the model experienced training convergence from epochs 100 - 120. The results indicate that the proposed method improves enhancement quality and generalizes well across different visual situations, making it suitable for real-world applications that require robust face detection in challenging lighting conditions.

Conclusion

Summary of Findings

The research findings support the hypothesis: combining targeted features from strong architectures can yield a superior model. HybridEnhancementNet produced high-quality outputs with minimal loss, enhancing details beyond the ground truth, demonstrating that integrating features from models like CIDNet and Zero-DCE can deliver significant performance gains.

Implications and Applications

Research in the field of enhancing low-light images can be applied to the improvement of autonomous vehicles at night, security in surveillance cameras, medical imaging, and more. In the case of autonomous machines, HybridEnhancementNet could be combined with a fast object detection model to help navigate in dark areas. HybridEnhancementNet can improve intruder detection at night by providing clear video from a camera, ensuring the owner’s property is safe.

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Limitations and Future Work

For future research, additional datasets such as ExDark (Loh et al., 2019) could be incorporated. While training on ExDark was out of scope for this paper, this type of additional training could improve robustness and enable further generalization.

Additionally, HybridEnhancementNet could be augmented with sharpening kernels (Redmon et al., 2016)—convolutions that amplify local intensity differences to make edges and fine details more pronounced—to improve the identifiability of key features in the enhanced image. Since HybridEnhancementNet combines components of Zero-DCE and CIDNet, a natural extension would be to integrate MIRNet and its strong denoising capabilities, potentially yielding a model with improved noise suppression alongside enhanced low-light correction.

While many architectures can enhance low-light images, systematically combining their most effective components into unified models remains underexplored and is a key direction for future research.

Acknowledgement

I would like to thank my advisor, Dr. Robail Yasrab (University of Cambridge), for their valuable guidance and support throughout this project.

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

References

Agarap, Abien Fred. (2018). Deep Learning using Rectified Linear Units (ReLU). arXiv. https://arxiv.org/abs/1803.08375

Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1780–1789). https://doi.org/10.1109/CVPR42600.2020.00186

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. Science, 313(5786), 504–507. https://doi.org/10.1126/science.1127647

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv. https://arxiv.org/abs/1412.6980

Land, E. H., & McCann, J. J. (1971). Lightness and Retinex Theory. Journal of the Optical Society of America, 61(1), 1. https://doi.org/10.1364/josa.61.000001

Loh, Y. P., & Chan, C. S. (2019). Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding, 178, 30–42. https://doi.org/10.1016/j.cviu.2018.10.010

Meharaj, S., Esampelli Malathi, Chittumalla Keerthana, Poosala Harshitha, & Dasari. (2023). Low Light Image Enhancement using MIRNet. International Journal for Multidisciplinary Research, 5(3). https://doi.org/10.36948/ijfmr.2023.v05i03.4001

O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks. ArXiv:1511.08458 [Cs], 2. https://arxiv.org/abs/1511.08458

Pham, T. D. (2022). Kriging-weighted Laplacian kernels for grayscale image sharpening. IEEE Access, 10, 57094–57106. https://doi.org/10.1109/ACCESS.2022.3178000

Jobson, D. J., Rahman, Z., & Woodell, G. A. (1997). A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Transactions on Image Processing, 6(7), 965–976. https://doi.org/10.1109/83.597272

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 779–788). https://doi.org/10.1109/CVPR.2016.91

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Lecture Notes in Computer Science, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Siddiqi, R., & Javaid, S. (2024). Deep Learning for Pneumonia Detection in Chest X-ray Images: A Comprehensive Survey. Journal of Imaging, 10(8), 176–176. https://doi.org/10.3390/jimaging10080176

Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep Retinex Decomposition for Low-Light Enhancement. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.1808.04560

Yan, Q., Feng, Y., Zhang, C., Wang, P., Wu, P., Dong, W., Sun, J., & Zhang, Y. (2024). You Only Need One Color Space: An Efficient Network for Low-light Image Enhancement. ArXiv.org. https://arxiv.org/abs/2402.05809

Yang, W., Yuan, Y., Ren, W., Liu, J., Scheirer, W. J., Wang, Z., Zhang, T., Zhong, Q., Xie, D., Pu, S., Zheng, Y., Qu, Y., Xie, Y., Chen, L., Li, Z., Hong, C., Jiang, H., Yang, S., Liu, Y., & Qu, X. (2020). Advancing Image Understanding in Poor Visibility Environments: A Collective Benchmark Study. IEEE Transactions on Image Processing, 29, 5737–5752. https://doi.org/10.1109/tip.2020.2981922

Waqas Zamir, S., Arora, A., Khan, S., Hayat, M., Shahbaz Khan, F., Yang, M.-H., & Shao, L. (2020). Learning Enriched Features for Real Image Restoration and Enhancement. Retrieved February 10, 2026, from https://www.ecva.net/papers/eccv_2020/papers_ECCV/papers/123700494.pdf