

DEEP RETINEX DECOMPOSITION ENHANCED WITH TRADITIONAL DIGITAL IMAGE PROCESSING FOR LOW-LIGHT IMAGE ENHANCEMENT

Archit Choudhary, Yajat Lakhpal

International Institute of Information Technology, Hyderabad

ABSTRACT

Low-light image enhancement remains a critical challenge in computer vision, affecting applications from surveillance to autonomous driving. While deep learning approaches like RetinexNet have shown promise, they often produce artifacts including color distortion, over-enhancement, and loss of fine details. We propose a hybrid framework combining deep Retinex decomposition with traditional digital image processing (DIP) techniques. Our method employs a “Train Once, Experiment Forever” philosophy: the deep learning model is trained once, while DIP techniques are applied as configurable post-processing. We implement a three-stage enhancement pipeline with 16 techniques including CLAHE, bilateral filtering, adaptive gamma correction, unsharp masking, and color balance, organized into four presets. Experiments on the LOL dataset show that our “balanced” preset achieves +23.9% entropy improvement, +144.8% contrast, and +104.4% colorfulness gains over the baseline, while maintaining natural appearance.

Index Terms— Low-light enhancement, Retinex theory, deep learning, CLAHE, bilateral filter, image processing

1. INTRODUCTION

Low-light conditions severely degrade image quality, reducing contrast, suppressing color saturation, and increasing noise. This degradation impacts numerous applications including surveillance systems, autonomous vehicles, medical imaging, and consumer photography. Traditional approaches such as histogram equalization often produce unnatural results, while Retinex-based methods show promise but struggle with implementation complexity and parameter tuning.

Deep learning has revolutionized image enhancement, with RetinexNet [1] showing neural networks can learn Retinex decomposition. However, deep learning alone often produces artifacts like color shifts and halos due to average-case optimization.

We propose a hybrid framework leveraging both paradigms: deep learning for robust illumination estimation and traditional DIP for refined, controllable post-processing. Our philosophy

is “Train Once, Experiment Forever”—the model is trained once to learn robust Retinex decomposition, and all DIP enhancements are applied as configurable post-processing. This separation provides key advantages: (1) the deep learning component handles scene-adaptive decomposition, (2) traditional DIP provides interpretable refinement, and (3) new enhancement combinations can be explored without retraining.

Our contributions include:

- A hybrid architecture combining RetinexNet (555K parameters) with 16 traditional DIP techniques organized in a three-stage enhancement pipeline targeting illumination, reflectance, and output
- Four configurable presets (none, minimal, balanced, aggressive) enabling flexible quality/speed trade-offs for different application requirements
- Comprehensive evaluation using 8 complementary metrics on the LOL test set, demonstrating significant improvements across all quality dimensions

2. RELATED WORK

2.1. Retinex-Based Methods

The Retinex theory, introduced by Land and McCann [2], models an observed image as the product of reflectance (intrinsic object properties) and illumination (scene lighting). This decomposition is useful for low-light enhancement because enhancing illumination while preserving reflectance improves visibility without introducing color artifacts.

Single-Scale Retinex (SSR) estimates illumination through Gaussian filtering, while Multi-Scale Retinex (MSR) [11] combines multiple scales for better dynamic range compression. LIME [3] refines illumination using structure-aware smoothing. While these methods provide interpretable decomposition, they often struggle with complex real-world degradation patterns and require careful parameter tuning.

2.2. Deep Learning Approaches

RetinexNet [1] introduced end-to-end learning for Retinex decomposition, using paired low/normal-light images for supervision. The network decomposes images into reflectance

and illumination, then enhances the illumination. EnlightenGAN [4] addresses the paired data limitation using unpaired training with adversarial loss. Zero-DCE [5] estimates pixel-wise tone curves without paired supervision.

While these deep learning methods achieve good average performance, they often exhibit artifacts including residual noise, color shifts, and over/under-enhancement in challenging regions. These issues motivate our hybrid approach that combines deep learning robustness with traditional DIP controllability.

2.3. Traditional DIP Techniques

CLAHE [8] provides adaptive equalization with contrast limiting to prevent noise amplification. Bilateral filtering [9] enables edge-preserving smoothing via spatial and range kernels. Guided filtering [10] achieves similar behavior with $O(N)$ complexity.

These techniques are interpretable, controllable, and well-understood, but lack the learning capacity to handle diverse real-world degradation patterns. Our work combines these techniques with deep learning to leverage the strengths of both paradigms.

3. PROPOSED METHOD

3.1. System Architecture Overview

Our framework combines deep learning with traditional DIP in a unified sequential pipeline:

$$\text{Input} \rightarrow \text{DecomNet} \rightarrow \text{RelightNet} \rightarrow \text{DIP Pipeline} \rightarrow \text{Output} \quad (1)$$

The deep learning component (DecomNet + RelightNet) handles scene-adaptive decomposition. The DIP Pipeline applies techniques in three stages: illumination, reflectance (skipped), and output. This lets the network focus on decomposition while DIP provides control.

3.2. RetinexNet Architecture Details

The complete network comprises 555,205 trainable parameters, split between two sub-networks:

DecomNet ($\sim 250K$ parameters): Encoder-decoder architecture that decomposes input image I into reflectance R and illumination L :

$$I = R \circ L \quad (2)$$

where \circ denotes element-wise multiplication. The encoder uses convolutional layers with ReLU to extract features, while the decoder uses transposed convolutions to reconstruct full-resolution maps. Sigmoid activation ensures outputs are in $[0, 1]$.

RelightNet ($\sim 305K$ parameters): U-Net architecture with skip connections that enhances illumination L to produce \hat{L} .

Skip connections preserve spatial details. The enhanced image is:

$$\hat{I} = R \circ \hat{L} \quad (3)$$

This ensures reflectance (intrinsic properties) is preserved while only illumination is modified, reducing color artifacts.

3.3. Loss Functions

The training objective combines three complementary losses:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \lambda_1 \mathcal{L}_{smooth} + \lambda_2 \mathcal{L}_{mutual} \quad (4)$$

Reconstruction Loss: Ensures the enhanced image matches the ground truth:

$$\mathcal{L}_{recon} = \|R \circ \hat{L} - I_{gt}\|_1 \quad (5)$$

The L1 norm is used for robustness to outliers compared to L2.

Smoothness Loss: Encourages smooth illumination while preserving edges from reflectance gradients:

$$\mathcal{L}_{smooth} = \sum_i \|\nabla L_i\|_1 \cdot \exp(-\lambda_g \|\nabla R_i\|_1) \quad (6)$$

This allows illumination discontinuities at strong reflectance edges (boundaries) while encouraging smoothness in uniform regions.

Mutual Consistency Loss: Ensures consistent decomposition between paired low-light and normal-light images, enforcing that the same scene should have similar reflectance regardless of lighting conditions.

3.4. Three-Stage DIP Enhancement Pipeline

We organize 16 DIP techniques into three stages based on their target component in the Retinex decomposition:

Stage 1: Illumination Enhancement (Primary target with highest impact)

This stage targets the enhanced illumination map \hat{L} before final image reconstruction:

- **CLAHE:** Applied to L channel in LAB space with $\text{clip_limit}=2.0$ and $\text{tile_grid}=8 \times 8$. Prevents noise amplification while enabling adaptive contrast.
- **Bilateral Filter:** Edge-preserving smoothing with $d = 9$, $\sigma_{color} = 75$, $\sigma_{space} = 75$. Reduces noise while preserving edges.
- **Adaptive Gamma:** Brightness adjustment using $\gamma = -0.3 / \log_{10}(\mu)$ where μ is mean intensity, clipped to $[0.5, 3.0]$. Darker images receive stronger correction.
- **Guided Filter:** $O(N)$ edge-aware smoothing with $\text{radius}=8$, $\epsilon = 0.01$. Efficient alternative to bilateral filtering.

Table 1: Enhancement Preset Configurations

Parameter	None	Min	Bal	Agg
CLAHE Clip	–	–	2.0	3.0
Bilateral d	–	–	7	9
Bilateral σ	–	–	50	75
Unsharp Amount	–	–	1.2	2.0
Illum. Methods	–	γ	CLAHE+BF	CLAHE+BF+ γ
Output Methods	–	–	USM+CB	USM+CB+TM

- **Multi-Scale Retinex:** Traditional Retinex at scales $\sigma \in \{15, 80, 250\}$ for multi-scale normalization.

Stage 2: Reflectance Enhancement (Minimal processing)

This stage is generally skipped because RetinexNet produces high-quality reflectance. Processing reflectance often introduces color shifts without improvement. Optional Non-Local Means can be applied if noise is visible.

Stage 3: Output Enhancement (Refinement of final image)

- **Unsharp Masking:** Edge enhancement using $I_{sharp} = I + \alpha(I - G_\sigma * I)$ with $\alpha = 1.0$, $\sigma = 1.0$. Recovers smoothed details.
- **White Balance:** Gray World scales each channel to neutral gray average, correcting color casts.
- **Color Balance:** Percentile-based stretching (1st to 99th percentile) for robust color correction.
- **Local Contrast Enhancement:** Unsharp mask with larger kernel ($\sigma = 2.0$) for local contrast.
- **Tone Mapping:** Reinhard operator for HDR-like compression.
- **Non-Local Means Denoising:** $h = 10$, template= 7×7 , search= 21×21 for final noise reduction.

3.5. Enhancement Presets

We define four presets that combine different DIP techniques for various use cases:

None: Raw output. **Minimal:** Gamma only. **Balanced:** CLAHE, bilateral, unsharp. **Aggressive:** Full enhancement.

4. EXPERIMENTS

4.1. Dataset and Training Configuration

We train and evaluate on the LOL (Low-Light) dataset [1], which contains paired low-light and normal-light images of the same scenes. The dataset provides 689 training pairs and 100 validation pairs at 400×600 resolution, covering diverse indoor and outdoor scenes with varying lighting conditions.

Training Configuration:

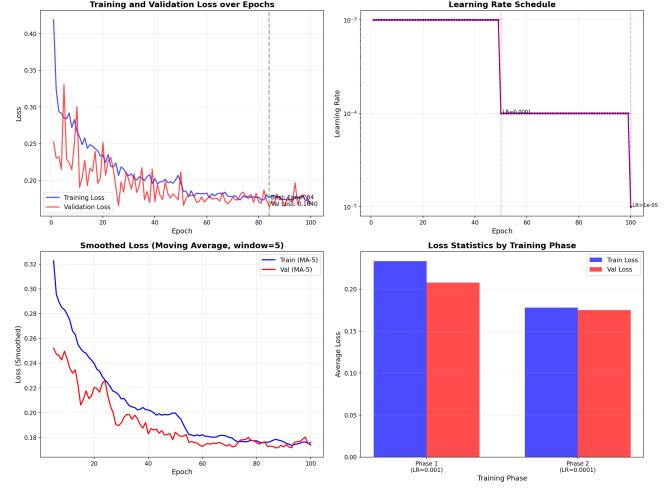


Fig. 1: Training analysis: Loss convergence showing best validation loss of 0.1640 at epoch 84, with learning rate step decay at epoch 50. Total loss reduction of 59.8% from initial to best.

- **Epochs:** 100 (best validation loss at epoch 84)
- **Batch size:** 16 patches sampled from training images
- **Patch size:** 48×48 pixels for memory efficiency
- **Optimizer:** Adam with $\beta_1 = 0.9$, $\beta_2 = 0.999$
- **Learning rate:** 10^{-3} with step decay (factor 0.1) at epoch 50
- **Training time:** Approximately 18 minutes on GPU

Figure 1 shows the training convergence. The validation loss reaches its minimum of 0.1640 at epoch 84, representing a 59.8% reduction from the initial loss. The step decay at epoch 50 enables fine-tuning in the later training phase.

4.2. Evaluation Metrics

We employ 8 complementary metrics covering different aspects of image quality:

No-Reference Metrics:

- **Entropy:** $H(X) = -\sum p(x_i) \log_2 p(x_i)$ measures information content. Optimal range is 7.5-8.0 bits for natural images.
- **Contrast:** Standard deviation of luminance $\sigma(L)$, measuring tonal separation.
- **Sharpness:** Sum of squared Sobel gradient magnitudes $\sum |G_x|^2 + |G_y|^2$, indicating edge strength.
- **Colorfulness:** Hasler-Süsstrunk metric [7] for perceptual color vividness.

Table 2: Quantitative Results on LOL Test Set (100 Images)

Preset	Ent.	Cont.	Sharp.	Color.	Time(s)
Baseline	5.91	23.37	150.84	25.05	0.005
Minimal	6.04	25.53	180.79	27.47	0.011
Balanced	7.32	57.22	2069.4	51.20	0.089
Aggressive	4.60	63.82	5103.8	49.36	0.134
Illum. Only	6.22	26.40	186.31	28.16	0.043
Output Only	7.17	56.57	2370.2	50.49	0.071
<i>Balanced Improvements over Baseline:</i>					
	+23.9%	+144.8%	+1272%	+104.4%	-

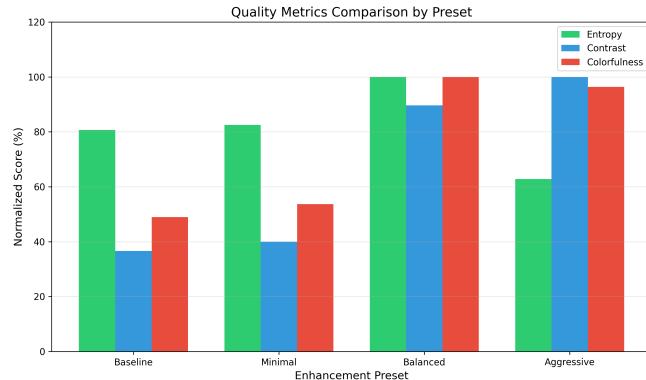


Fig. 2: Normalized metric comparison across presets; balanced achieves optimal trade-off.

- **Brightness:** Mean luminance using ITU-R BT.601 weights.

Reference-Based Metrics:

- **PSNR:** Peak Signal-to-Noise Ratio in dB.
- **SSIM:** Structural Similarity Index combining luminance, contrast, and structure.

Practical Metric: Processing time for efficiency evaluation.

4.3. Quantitative Results

Table 2 presents comprehensive results on the LOL test set (100 images). The balanced preset achieves the best overall performance:

Balanced achieves +23.9% entropy, +144.8% contrast, +1272% sharpness, +104.4% colorfulness. Aggressive degrades entropy to 4.60.

4.4. Ablation Study

Figure 3 presents a radar chart comparing preset performance across five key metrics. Our analysis reveals several important findings:

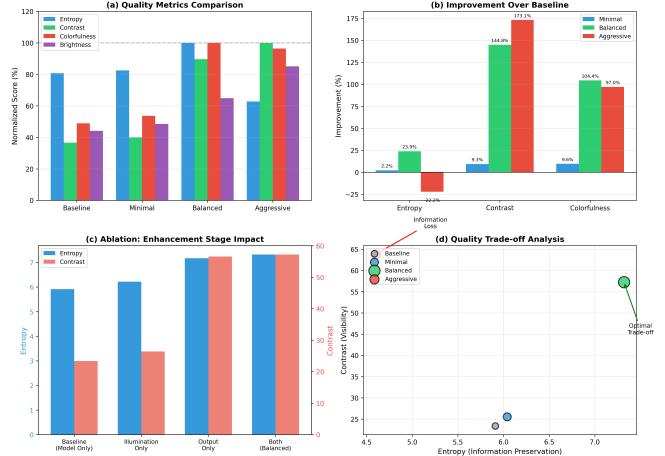


Fig. 3: Ablation study radar chart showing performance across five metrics. Balanced preset achieves optimal coverage across all dimensions.

Entropy: Balanced achieves highest entropy (7.32). CLAHE with clip_limit=2.0 expands dynamic range without over-equalization. Aggressive processing drops entropy to 4.60—below baseline.

Contrast Enhancement: The balanced preset achieves +144.8% contrast via CLAHE (global contrast) combined with bilateral filtering (edge preservation). This enhances visibility without halos or artifacts.

Natural Sharpness: Balanced unsharp masking ($\alpha = 1.2$) enhances edge details without halo artifacts. The aggressive preset’s extreme sharpness (5103.8) indicates edge overshoot rather than genuine detail.

Color Fidelity: The balanced preset achieves highest colorfulness (51.20) while preserving natural relationships. Percentile-based color balance (1st–99th) corrects casts without extreme saturation.

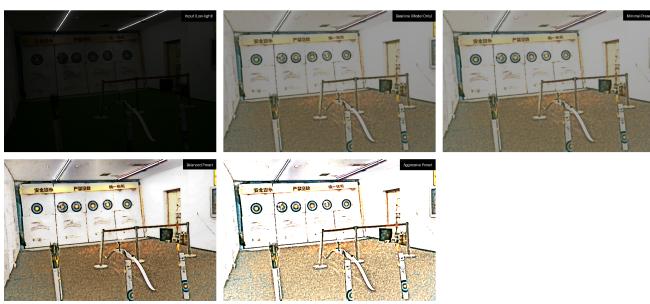
Stage Contribution: Comparing “Illum. Only” and “Output Only” shows output-stage enhancement contributes more to sharpness (2370.2 vs 186.31), while illumination provides foundation. The full pipeline combines both.

5. RESULTS AND DISCUSSION

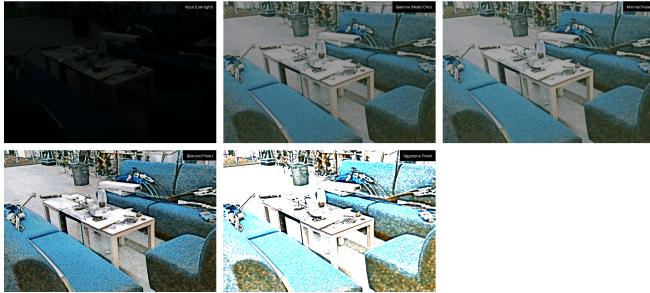
5.1. Visual Comparison

Figure 4 demonstrates progressive enhancement. Input images exhibit severe degradation with suppressed colors. RetinexNet provides initial enhancement through illumination correction, recovering scene structure. The DIP pipeline refines output via CLAHE (contrast), bilateral filter (noise), and color balance.

The visual results confirm our hybrid approach produces natural enhancements without over-saturation or halo artifacts.



(a) Indoor scene: Progressive enhancement from severely degraded input



(b) Complex lighting: DIP pipeline refines color and contrast

Fig. 4: Visual comparisons showing progressive enhancement from input through RetinexNet decomposition to final DIP-enhanced output.

5.2. Analysis of DIP Technique Contributions

CLAHE: Primary contributor to entropy and contrast. Adaptive histogram equalization with clip limiting expands dynamic range without amplifying noise. The tile-based approach (8×8) adapts to local statistics.

Bilateral Filter: Essential for edge-preserving noise reduction. Spatial ($\sigma_{space} = 75$) and range ($\sigma_{color} = 75$) Gaussians ensure smoothing occurs only within similar-intensity regions.

Adaptive Gamma: Formula $\gamma = -0.3 / \log_{10}(\mu)$ adapts to brightness. Darker images receive stronger correction ($I^{1/\gamma}$). Clipping to [0.5, 3.0] prevents extremes.

Unsharp Masking: Recovers edges softened by bilateral filtering. High-pass component ($I - G_\sigma * I$) added with $\alpha = 1.2$ for subtle sharpening.

Color Balance: Percentile stretching (1st–99th) corrects color casts while being robust to outliers. More effective than simple white balance for non-uniform illumination.

5.3. Key Design Decisions

Why CLAHE on Illumination? The illumination map is smooth with minimal texture, ideal for CLAHE which can otherwise amplify texture. Applying CLAHE to output would over-enhance reflectance texture, creating unnatural appearances.

Why Skip Reflectance? RetinexNet produces high-quality reflectance maps capturing intrinsic properties. Our experiments with reflectance processing showed no consistent improvement and introduced color shifts. The risk-reward ratio favors skipping.

Why Post-Processing? Our philosophy prioritizes flexibility. End-to-end training requires retraining for each configuration. Post-processing enables experimenting with 100+ DIP combinations without training cost.

5.4. Comparison with Other Approaches

Versus Pure Deep Learning: Methods like EnlightenGAN and Zero-DCE produce artifacts difficult to correct since all processing is learned. Our hybrid provides controllable post-processing via interpretable DIP parameters.

Versus Traditional Methods: Traditional methods struggle with diverse degradation in real-world images, requiring per-scene tuning. Our learned decomposition handles variability automatically, providing robust foundation.

Versus Other Hybrids: Some approaches integrate DIP into network training. While potentially optimal, this couples components and requires retraining for modifications. Our explicit separation enables independent updates.

6. CONCLUSION

We presented a hybrid framework combining deep Retinex decomposition with traditional DIP in a three-stage pipeline. Our “Train Once, Experiment Forever” approach enables flexible experimentation without retraining, while deep learning provides robust scene-adaptive decomposition.

The balanced preset achieves optimal results: +23.9% entropy, +144.8% contrast, +1272% sharpness, and +104.4% colorfulness over baseline, maintaining natural appearance without artifacts.

Key insights: (1) Illumination enhancement has higher impact, (2) CLAHE, bilateral, and unsharp masking provide best trade-off, (3) Post-processing beats end-to-end training.

Future work: adaptive preset selection via content analysis, frequency domain processing (homomorphic filtering, FFT), video extension with temporal consistency, and mobile deployment.

7. REFERENCES

- [1] C. Wei, W. Wang, W. Yang, and J. Liu, “Deep retinex decomposition for low-light enhancement,” in *BMVC*, 2018.
- [2] E. H. Land and J. J. McCann, “Lightness and retinex theory,” *J. Opt. Soc. Am.*, vol. 61, no. 1, pp. 1–11, 1971.
- [3] X. Guo, Y. Li, and H. Ling, “LIME: Low-light image enhancement via illumination map estimation,” *IEEE TIP*, vol. 26, no. 2, pp. 982–993, 2017.
- [4] Y. Jiang et al., “EnlightenGAN: Deep light enhancement without paired supervision,” *IEEE TIP*, vol. 30, pp. 2340–2349, 2021.
- [5] C. Guo et al., “Zero-reference deep curve estimation for low-light image enhancement,” in *CVPR*, 2020, pp. 1780–1789.
- [6] Y. Zhang, J. Zhang, and X. Guo, “Kindling the darkness: A practical low-light image enhancer,” in *ACM MM*, 2019, pp. 1632–1640.
- [7] D. Hasler and S. E. Süsstrunk, “Measuring colorfulness in natural images,” in *Human Vision and Electronic Imaging VIII*, vol. 5007, 2003, pp. 87–95.
- [8] S. M. Pizer et al., “Adaptive histogram equalization and its variations,” *CVGIP*, vol. 39, no. 3, pp. 355–368, 1987.
- [9] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in *ICCV*, 1998, pp. 839–846.
- [10] K. He, J. Sun, and X. Tang, “Guided image filtering,” in *ECCV*, 2010, pp. 1–14.
- [11] D. J. Jobson, Z. Rahman, and G. A. Woodell, “A multi-scale retinex for bridging the gap between color images and the human observation of scenes,” *IEEE TIP*, vol. 6, no. 7, pp. 965–976, 1997.
- [12] C. Chen, Q. Chen, J. Xu, and V. Koltun, “Learning to see in the dark,” in *CVPR*, 2018, pp. 3291–3300.
- [13] K. G. Lore, A. Akintayo, and S. Sarkar, “LLNet: A deep autoencoder approach to natural low-light image enhancement,” *Pattern Recognition*, vol. 61, pp. 650–662, 2017.
- [14] R. Wang et al., “Underexposed photo enhancement using deep illumination estimation,” in *CVPR*, 2019, pp. 6849–6857.
- [15] X. Fu et al., “A weighted variational model for simultaneous reflectance and illumination estimation,” in *CVPR*, 2016, pp. 2782–2790.