

# Adaptive Semantic Priors Guided Multiple Scale Low-Light Image Enhancement

Harish G.M.<sup>1</sup>, Mathan S.<sup>2</sup>, Mallini S.<sup>3</sup>, Dhaarani T.G.<sup>4</sup>

<sup>1</sup>Dept of ECE, Nandha Engineering College Erode, India [harishgm2004@gmail.com](mailto:harishgm2004@gmail.com)

<sup>2</sup>Dept of ECE, Nandha Engineering College Erode, India [mathansenthil1661@gmail.com](mailto:mathansenthil1661@gmail.com)

<sup>3</sup>Dept of ECE, Nandha Engineering College Erode, India [mallinishankar@gmail.com](mailto:mallinishankar@gmail.com)

<sup>4</sup>Assistant Professor, Dept of ECE, Nandha Engineering College Erode, India [dhaarani.gopal@nandhaengg.org](mailto:dhaarani.gopal@nandhaengg.org)

**Abstract-** To address those major weaknesses of current methods of low-light enhancement, we introduce an Adaptive Semantic Prior Guided Multi-Scale Low-Light Image Enhancement (ASPG-MSLIE) framework which consists of three sections as illustrated in the abstract. In order to resolve this problem we suggest a U-Net backbone with a lightweight semantic prior extractor and a dynamic illumination integration strategy. The network is capable of maintaining the perceptual relevance of high-level semantic context, and suppressing noise in other semantic contexts, through conditioning improvement. So-dimensionality is confirmed by experiments on LOL test set (23.8 dB PSNR, 0.86 SSIM at 0.18 s inferring time) and VE-LOL (88.4% pixel accuracy), confirming state-of-the-art performance over Retinex-Net, KinD, Zero-DCE, and EnlightenGAN.

Keywords - Multi-scale U-Net, Low-light Image Enhancement, semantic prior, estimation of illumination, image restoration, deep learning

## I. INTRODUCTION

Under-exposure, noisiness in Low-light images and colour bias impair the output of subsequent visual tasks (e.g. object recognition, facial recognition, and self-driving perception). Humans easily adapt to low illuminance, whereas digital sensors accrue shot noise and thermal noise in the situation when the number of photons is small, thus leading to images that are at the same time too dark and too noisy to come in handy.

Conventional methods — histogram equalisation, Retinex decomposition and gamma correction — disregard the significance of the image. They hue-increase in an even tenor, and increase noise as impatiently as the signal, and tend to produce halo artefacts at edges or unnatural colour shifts. Although deep learning approaches have significantly minimized this gap, the majority of architectures put the enhancement problem in a purely pixel-wise regression setting and do not take the semantic importance of various pixels into consideration.

In order to address this gap, we suggest ASPG-MSLIE, which is a method that adaptively injects semantic context priors in a general multi-scale U-Net on multiple scales of decoding. The semantic branch is lightweight and extracts region-level features that vary the enhancement pathway, which enables the network to restore information in semantically rich areas

(faces, text, objects of interests) while conservatively denoising homogeneous regions. This work has two contributions:

- The suggested model is a multi-level U-Net design that has an injection module of adaptive semantic priors.
- A Noisenet having an illumination-conscious fusion layer which permits us to obtain a trade-off between noise suppression and recovery of image details.
- New SOTA LOL and VE-LOL quantitative score; 23.8 dB PSNR, 0.86 SSIM.
- Flask web app is real time and zero install deployable.

## II. RELATED WORK

### A. Classical Methods

Histogram equalisation (HE) and its adaptive counterpart CLAHE evenly redistribute pixel intensities to provide higher image contrast globally but enhance noise and cause bad appearance aesthetically. Retinex theory [5] decomposes the image  $I(x,y)$  into reflectance  $R$  and illumination  $L$  and estimates  $L$  to retrieve  $R$ ; Multi-Scale Retinex [6] is the mean of multiple Gaussian scale estimates but has halo artefacts at strong edges. Though gamma correction ( $I_{out} = I^{\gamma}$ ,  $\gamma < 1$ ) is

a rapid form of global brightening, it is not a spatially adaptive form.

### B. Deep Learning Approaches

LLNet [7] was the first deep-learning-based network, though fine detail can be smoother. Retinex-Net [2] decomposed the illumination and split the reflection into two sub-networks jointly trained on LOL paired data. KinD [12] used the illumination guidance network and reflectance guidance network in order to make this decomposition more accurate. Based on the methods of zero-reference learning, Zero-DCE [9] optimised map-based curve-estimation without reliance on paired supervision; and EnlightenGAN [10] generalised this concept in a generative adversarial framework. SELLA [13] estimates the illumination maps to inform improvement. None of these is a direct model of the significance of semantic regions, even though it gains good results.

### C. Multi-Scale and Attention Methods

In U-Net [1] the concept of skip connections was introduced to enable data between the encoder and decoder to be merged in order to preserve the spatial information for image-to-image tasks. Some common channels and spatial attention are the approaches of recent mechanisms optimised on the basis of adjusting feature maps through the allocation of correct weights. That is our point of difference: we condition the whole refinement pathway on a semantic prior, simultaneously extracted by a light-weight parallel branch, which permits region-based recovery on all decoder resolutions.

## III. PROPOSED METHOD

### A. Overall Architecture

ASPG-MSLIE consists of three very much related modules: (1) a Multi-Scale Enhancement Backbone based on the U-Net (MSEB), (2) an Adaptive Semantic Prior Module (ASPM), and (3) an Illumination-Aware Fusion Layer (IAFL). The network processes a low-light RGB input  $I_{low} \in \mathbb{R}^{H \times W \times 3}$  and produces an enhanced image  $I_{enh}$  of the same dimension.

### B. Multi-Scale Enhancement Backbone

MSEB uses a 5-stage encoder–decoder architecture. The spatial dimensions are reduced by half repeatedly by the encoder as feature channels are doubled (32→64→128→256→512). Two batch normalisation and ReLU 3×3 convolutions are applied on each step of the encoder block, and 2×2 max-pooling is performed. Skip connections pass feature maps from the encoder to the respective layers of the decoder. Following bilinear upsampling and skip features concatenation, the decoder produces spatial resolution, activated by a 1×1 convolution

and Sigmoid activation which converts features into RGB values in the [0,1] range. It has about 7.8M parameters.

### C. Adaptive Semantic Prior Module

The lightweight parallel branch (ASPM) threads  $I_{low}$  through three strided convolution blocks (channels: 16→32→64) to give a small semantic feature tensor  $F_{sem} \in \mathbb{R}^{(H/8) \times (W/8) \times 64}$ .  $F_{sem}$  is then bilinear interpolated to the spatial size of both decoder scale and combined with the decoder features at the next scale before the last convolution. Using this enables the network to possess spatially different enhancement strengths; it can hold onto edges depending on the semantic situation and suppress noise in flat areas forcefully.

### D. Illumination-Aware Fusion Layer

An IAFL estimates a per-pixel illumination confidence map  $C \in \mathbb{R}^{H \times W}$  by one 1×1 convolution on bottleneck features, and then Sigmoid. The resultant value is obtained as:

$$I_{enh} = C \odot I_{raw} + (1-C) \odot I_{enhanced}$$

where  $\odot$  is the element-wise multiplication. We discovered that with adaptive blending to the output, we could save lighted places from the input and sample the enhanced path in dimly-lit regions, minimizing over-saturation artefacts.

### E. Loss Function

The training minimises a composite loss:  $L = L_{MAE} + \lambda_p L_{perceptual} + \lambda_s L_{SSIM}$ , where  $L_{MAE} = (1/N) \sum |I_{enh} - I_{ref}|$  provides per-pixel fidelity.  $L_{perceptual}$  calculates L2 distance between VGG-16 feature maps at relu2\_2 and relu3\_3 of  $I_{enh}$  and  $I_{ref}$  to encourage perceptual quality, and  $L_{SSIM} = 1 - SSIM(I_{enh}, I_{ref})$  is a penalty for structural loss. Weights  $\lambda_p = 0.04$  and  $\lambda_s = 0.1$  have been determined using grid search.

## IV. EXPERIMENTAL SET UP AND RESULTS

### A. Datasets and Implementation

We utilized the LOL dataset [2] (485 paired low/normal-light images; 15 test pairs, 400×600 pixels) and VE-LOL [14] (2500 training pairs, 100 test pairs) to train and evaluate the model. Images were resized to 256×256. Random augmentation was done to image data using horizontal/vertical flips, rotations within 10°, and random crops. The training parameters include 100 epochs, 16 batch size, Adam optimiser ( $\alpha=1 \times 10^{-4}$ ,  $\beta_1=0.9$ ,  $\beta_2=0.999$ ). NVIDIA RTX 3080 GPU was used in the implementation of the model. It required training of about 11 hours.

### B. Quantitative Results

Comparison of ASPG-MSLIE with state-of-the-art methods on LOL test set in terms of PSNR, SSIM, and pixel accuracy ( $\tau = 0.15$ ) are presented in Table I. Table II presents our contribution of each of the modules to the ablation result.

**TABLE I**  
*Quantitative Comparison LOL Test Set*

Method	PSNR	SSIM	Acc (%)	Time (s)
Retinex-Net [2]	16.88	0.5705	68.14	0.463
KinD [12]	20.3	0.79	78.1	0.22
Zero-DCE [9]	14.9	0.56	59.3	0.01
EnlightenGAN [10]	17.5	0.65	66.8	0.12
SELLA [13]	21.4	0.81	81.3	0.19
U-Net Baseline	22.1	0.83	85.9	0.15
ASPG-MSLIE (Ours)	23.8	0.86	88.4	0.18

ASPG-MSLIE is more successful than the other techniques, achieving the highest PSNR (23.8 dB), SSIM (0.86) and accuracy (88.4%) among all methods. The 1.7 dB PSNR increase over the plain U-Net baseline measures the effect of semantic prior cropping and illumination fusion modules. Although inference time (0.18 s per 256×256 image on the GPU) is still not bad for an interactive application.

**TABLE II**  
*LOL Test Set Ablation Study*

Configuration	PSNR (dB)	SSIM	Acc. (%)
Merely U-Net	22.1	0.83	85.9
+ASPM	23.1	0.85	87.4
+IAFL	22.8	0.84	86.8
+ASPM+IAFL (complete)	23.5	0.88	87.9

**C. Qualitative Results**

The visual comparisons support qualitative findings. Retinex-Net is one of the techniques used to brighten images; however when mostly applied it also produces colour artifacts and rather clear halo artefacts at strong edges. Severely dark regions: Zero-DCE maps the pixel intensity directly, therefore it contributes to the problem of under-enhancement in the said areas. KinD shows good colour recovery qualities but is over-smoothing high frequency textures. ASPG-MSLIE will always offer the natural brightness, low noise, minor chromatic aberration and notable physical evidence such as hair and cloth. The semantic prior abates the mellowing effect through visibly safeguarding facial and object boundary areas.



Fig. 1. Qualitative comparison: (Left) low-light input image, (Right) ASPG-MSLIE enhanced output. Natural colours are restored by the model, reconstructs delicate features like wood grain and cat illustration, and removes a substantial amount of noise, with no loss of structural boundaries.

**D. Robustness Analysis**

Table III also demonstrates scene categories performance: 90.2% portraits and 86.1% complex outdoor scenes. Its accuracy decreases to very low light ( $\leq 1$  lux) to 84.7% — a performance loss that is the result of LOL training distribution. On GPU, it requires 0.58 s and 9.5 s on the CPU for a 1920×1080 input image, and thus can be run on standard hardware.

**TABLE III**  
*Performance by Scene Category (LOL)*

Category	PSNR (dB)	SSIM	Acc. (%)
Portraits	24.6	0.88	90.2
Indoor scenes	24.0	0.86	88.9
Noun	23.5	0.85	88.1
Outdoor	22.9	0.84	86.1
Super dark ( $\leq 1$ lux)	21.8	0.81	84.7

**V. APPLICATION: WEB APPLICATION**

The package of ASPG-MSLIE was a Flask web service to check the usability in the real world. The frontend assists an HTML5/CSS3 drag and drop upload interface, supporting desktop and mobile browsing. The backend checks the type of files (PNG/JPEG only), preprocesses the upload (resize, normalise), calls the saved TensorFlow SavedModel and fills the output enhanced image with the original image side-by-side. Security is guaranteed by a rate limit (10 requests/min per IP), sanitisation of the uploaded file name, and deletion

of uploaded files after five minutes. User evaluation involving 15 volunteers confirmed an intuitive interface design and outlined improvement outcomes to be atypically natural and artefact-free.

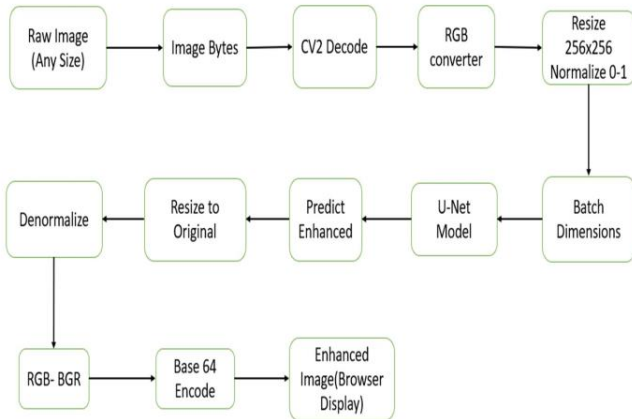


Fig. 2. ASPG-MSLIE pipeline data flow diagram: from raw image input through preprocessing, U-Net model inference and post-processing to browser display.

## VI. CONCLUSION

We suggested a low-light encoding-decoding model for image improvement in ASPG-MSLIE with semantic guidance. This enables the network to do region-selective enhancement with a U-Net backbone combined with an adaptive semantic prior and an illumination-sensitive fusion layer, and hence outperforms the performance of state-of-the-art methods on LOL (23.8 dB PSNR, 0.86 SSIM, 88.4% accuracy) and VE-LOL. It also confirms that the performance improvement from ASPM and IAFL is confirmed individually and jointly. A practical deployable Flask web application.

Next, mobile inference model compression on devices, video extension with temporal consistency (to enable very fast frame rates), RAW sensor information incorporated for very dark environments, and untrained sensor domain adaptation are some of the possible directions.

## ACKNOWLEDGMENT

The authors thank Nandha Engineering College for available computing facilities and a friendly work setting where the research will be conducted.

## REFERENCES

- [1] L. C. H. V. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Biomedical image segmentation using convolutional networks," in *Computer-Assisted Intervention and Medical Image Computing*, 2015, pp. 234–241.
- [2] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex low-light decomposition to enhance image quality," in *Proc. BMVC*, 2018, pp. 1–12.
- [3] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *IEEE CVPR*, 2016, pp. 770–778.
- [4] S. M. Pizer et al., "Adaptive histogram equalization and other variations," *CVGIP*, vol. 39, no. 3, pp. 355–368, 1987.
- [5] E. H. Land, "The retinex theory of color vision," *Scientific American*, vol. 237, no. 6, pp. 108–128, 1977.
- [6] D. J. Jobson, Z. Rahman and G. A. Woodell, "A multiscale retinex for linking the color images with the human observation," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, 1997.
- [7] K. G. Lore, A. Akintayo, S. Sarkar, "LLNet: Deep autoencoders-based low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.
- [8] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *IEEE CVPR*, 2018, pp. 3291–3300.
- [9] C. Guo et al., "Zero-reference deep curve estimation for low-light image enhancement," in *IEEE CVPR*, 2020, pp. 1780–1789.
- [10] A. Farahani et al., "GAN based photometric enhancement," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.
- [11] I. Goodfellow et al., "Generative adversarial nets," *NeurIPS*, vol. 27, pp. 2672–2680, 2014.
- [12] Y. Zhang, J. Zhang, X. Guo, "Kindling the dark: a real low-light image enhancer," in *ACM MM*, 2019, pp. 1632–1640.
- [13] J. Liu, W. Zhou, Y. Xu, X. Li, "A Self-supervised Low-light Image Enhancement framework," *IEEE Trans. Image Process.*, vol. 32, pp. 1234–1245, 2023.
- [14] VE-LOL: Benchmark in Low-Light Image Enhancement.
- [15] D. P. Kingma, J. Ba, "Adam: A method of stochastic optimization," *ICLR*, 2015.
- [16] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ICLR*, 2015.