LL-DAWGNet: Degradation-Adaptive Wavelet-Gradient Network for Low-Light Remote Sensing Image Enhancement

Koyyada Dinesh Kumar[1], Sourav Phogat[1], and Sujit Kumar Sahoo[1]

School of Electrical Sciences,
Indian Institute of Technology Goa, India
{koyyada21242104,sourav2414205,sujit}@iitgoa.ac.in

Abstract. Low-light remote sensing images often suffer from poor visibility, color distortion, weak local contrast, and loss of structural details. These degradations are difficult to correct because remote sensing scenes contain both large-scale illumination variations and fine spatial structures such as roads, buildings, fields, and vegetation boundaries. In this work, we propose LL-DAWGNet, a compact degradation-adaptive wavelet-gradient network for low-light remote sensing image enhancement. The proposed method first extracts low-light prior cues, including illumination deficit, color-balance residual, saturation, intensity, and local contrast. These cues are summarized into a compact condition vector, which guides illumination-frequency adaptation, Haar wavelet subband weighting, Sobel-gradient refinement, and gated skip fusion. This allows the network to adapt its restoration behavior according to the degradation characteristics of each input image. Experiments on iSAID-dark show that LL-DAWGNet achieves 25.48 dB PSNR and 0.799 SSIM with only 0.89M parameters and 7.67G FLOPs. Results on iSAID-dark(high-pixel) further show strong structural preservation on high-resolution remote sensing images.

Keywords: Low-light image enhancement · Remote sensing · Wavelet transform · Gradient prior · Image restoration

1 Introduction

Remote sensing (RS) images support several Earth observation tasks, including urban monitoring, environmental assessment, disaster response, and surveillance. In practice, however, RS images are not always captured under favorable illumination. Images acquired at night, during low-light conditions, or under weak exposure often show poor visibility, low contrast, color distortion, weak structural details, and amplified noise. These degradations reduce visual interpretability and can also affect downstream pattern recognition tasks that depend on reliable radiometric and structural information. Low-light RS image enhancement therefore aims to recover visually clear and structurally faithful images from degraded observations.

A common way to describe low-light image formation is through the Retinex model, where an observed image is represented as the element-wise product of reflectance and illumination [8]. In practical low-light imaging, sensor noise also becomes more prominent because the useful signal is weak. We write this degradation as

$$\mathbf{I}_l = \mathbf{R}_l \odot \mathbf{L}_l + \mathbf{n}, \tag{1}$$

where $\mathbf{I}_l$ is the observed low-light image, $\mathbf{R}_l$ is the reflectance component, $\mathbf{L}_l$ is the illumination component, $\mathbf{n}$ denotes noise, and $\odot$ represents element-wise multiplication. In real RS scenes, this degradation is further coupled with color imbalance, low saturation, weak local contrast, and suppressed fine structures. Thus, low-light enhancement is not merely a brightness correction problem; it requires joint recovery of illumination, color, structure, and detail.

Compared with natural low-light images, low-light RS images introduce additional challenges. RS scenes usually contain large spatial coverage and continuous structures such as roads, buildings, fields, vegetation, and water bodies. These structures may extend across large image regions, while small objects and boundaries still require fine detail preservation. CNN-based methods are efficient, but their local receptive fields may limit broad context modeling. Transformer-based methods improve global representation, but their computational cost can be high for high-resolution imagery [2–4]. This creates a need for compact enhancement models that can handle large-scale illumination variation without losing local structure.

Recent low-light enhancement methods show that decomposition is useful for this problem. FourLLIE uses Fourier frequency information to improve lightness and detail recovery [5]. DFFN extends this idea to low-light RS enhancement by using Fourier amplitude for illumination recovery and phase information for detail refinement [1]. CIDNet introduces the HVI color space to separate chromatic and intensity information, reducing color bias and brightness artifacts in low-light enhancement [6]. In another direction, WWE-UIE shows that Haar wavelet decomposition and Sobel-gradient refinement can provide efficient frequency and structural priors [7]. These studies suggest that illumination, color, frequency, and structure should not be treated as a single entangled RGB mapping.

However, most existing designs apply their restoration operations in a fixed manner once the input representation is chosen. This is restrictive for low-light RS images because the degradation level can vary across images and regions. Some scenes need stronger illumination correction, some require color recovery, and others need careful edge restoration without amplifying noise. This motivates a degradation-adaptive framework in which the restoration process is guided by low-light cues estimated from the input image itself.

To this end, we propose **LL-DAWGNet**, a compact degradation-adaptive wavelet-gradient network for low-light RS image enhancement. The proposed method first extracts lightweight low-light prior cues, including illumination deficit, color-balance residual, saturation, intensity, and local contrast. These cues are summarized into a compact condition vector. The condition vector then guides illumination-frequency adaptation, Haar wavelet subband weight-

ing, Sobel-gradient refinement, and gated skip fusion. In this way, the network adjusts its restoration behavior according to the degradation characteristics of each input image.

The main contributions of this work are as follows:

- We propose LL-DAWGNet, a compact degradation-adaptive wavelet-gradient network for low-light RS image enhancement.
- We design a low-light prior cue extractor that captures illumination, color, contrast, and structural degradation cues and summarizes them into a compact condition vector.
- We introduce degradation-conditioned frequency and gradient restoration, where the condition vector guides low/high-frequency adaptation, Haar wavelet subband weighting, and Sobel-gradient refinement.
- We use degradation-aware gated skip fusion and bottleneck strip context to improve structural preservation and broad illumination modeling in RS scenes.

2 Related Work and Positioning

Low-light image enhancement has been widely studied through model-based and learning-based approaches. Retinex-based methods model an image using illumination and reflectance components, which provides a useful physical interpretation for low-light restoration [8]. Deep learning methods further improve enhancement quality by learning degradation-aware mappings from low-light images to normal-light images. SNR-Aware LLIE uses signal-to-noise-ratio guidance to handle spatially varying degradation [2], while LLFormer studies low-light enhancement for ultra-high-definition images using an efficient transformer design [3]. RetinexFormer combines Retinex-based illumination modeling with transformer blocks to improve global context representation under low illumination [4]. These methods show the importance of illumination modeling, degradation awareness, and long-range context. However, they are mainly designed for natural images and do not explicitly adapt wavelet, gradient, and skip-fusion operations to the degradation characteristics of low-light remote sensing images.

Low-light remote sensing enhancement has received comparatively less attention. Remote sensing images differ from natural images because they contain large spatial coverage, repeated structures, and fine object boundaries. Earlier learning-based work explored CNN-based enhancement for low-light remote-sensing images, such as RSCNN [12]. DFFN directly addresses this setting using a spatial–frequency dual-domain network, where Fourier amplitude is used for brightness restoration and phase information is used for detail refinement [1]. This confirms that separating illumination-related and detail-related information is useful for low-light RS enhancement. Related frequency and representation-based methods also support this direction. FourLLIE exploits Fourier frequency information for lightness enhancement and detail recovery [5], while CIDNet introduces the HVI color space to decouple chromatic and intensity information and reduce color and brightness artifacts [6]. These works indicate that low-light

enhancement benefits from treating illumination, color, and structure as different degradation factors rather than as one entangled RGB mapping.

Wavelet and gradient priors offer another efficient way to preserve spatial structure. WWE-UIE combines white balance, Haar wavelet decomposition, and Sobel-gradient fusion in a compact restoration network [7]. Although it is designed for underwater image enhancement, it shows that wavelet and gradient operations can provide useful frequency and structural cues with low computational cost. Our method follows this general idea of using efficient priors, but adapts it to low-light remote sensing. Instead of applying wavelet and gradient operations in a fixed manner, LL-DAWGNet first extracts low-light prior cues and summarizes them into a compact condition vector. This condition vector guides illumination-frequency adaptation, Haar subband weighting, Sobel-gradient refinement, and gated skip fusion. Thus, the proposed method is positioned as a degradation-adaptive wavelet-gradient framework for low-light RS enhancement.

3 Proposed Method

We propose LL-DAWGNet, a degradation-adaptive wavelet-gradient network for low-light remote sensing image enhancement. Given a low-light image $\mathbf{I}_l \in [0,1]^{H\times W\times 3}$, the network predicts an enhanced image $\hat{\mathbf{I}}$. The main idea is to avoid treating all low-light images with a fixed restoration pipeline. Instead, the network first estimates simple low-light degradation cues from the input image and then uses these cues to guide frequency adaptation, wavelet subband weighting, gradient refinement, and skip fusion.

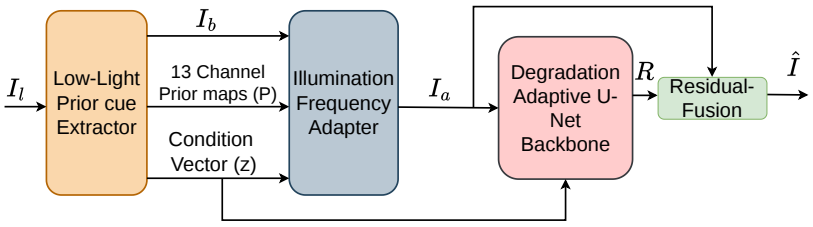


Fig. 1: Overall architecture of LL-DAWGNet. The low-light prior cue extractor generates a base image $\mathbf{I}_b$, 13-channel prior maps $\mathbf{P}$, and a degradation condition vector $\mathbf{z}$. The illumination-frequency adapter produces the prior-adapted image $\mathbf{I}_a$. The degradation-adaptive U-Net predicts a residual $\mathbf{R}$, and the final output is obtained as $\hat{\mathbf{I}} = \mathbf{I}_a + \mathbf{R}$.

3.1 Overall Pipeline

Fig. 1 shows the overall pipeline. The network has three main components: a low-light prior cue extractor($\mathcal{E}_p$), an illumination-frequency adapter($\mathcal{A}$), and a

degradation-adaptive wavelet-gradient U-Net($\mathcal{U}$). The prior cue extractor produces a base image, a set of prior maps, and a compact degradation condition vector. The adapter uses these cues to generate a prior-adapted image, and the U-Net predicts the remaining residual correction. The complete forward process is written as

$$\mathbf{I}_b, \mathbf{P}, \mathbf{z} = \mathcal{E}_p(\mathbf{I}_l), \quad \mathbf{I}_a = \mathcal{A}(\mathbf{I}_b, \mathbf{P}, \mathbf{z}), \quad \mathbf{R} = \mathcal{U}(\mathbf{I}_a, \mathbf{z}), \tag{2}$$

and the enhanced image is obtained by residual fusion:

$$\hat{\mathbf{I}} = \mathbf{I}_a + \mathbf{R}. \tag{3}$$

This design lets the restoration backbone work on a prior-adapted image rather than directly on the degraded input.

3.2 Low-Light Prior Cue Extractor

Low-light degradation is not limited to reduced brightness. It is often coupled with color imbalance, weak saturation, low local contrast, and suppressed structures. To expose these degradation factors to the network, we construct a set of lightweight prior maps from the input image, as shown in Fig. 2. First, a weak color-balanced image $\mathbf{I}_{cb}$ is obtained using a gray-world Retinex-style transform. The difference between $\mathbf{I}_{cb}$ and $\mathbf{I}_l$ gives a color-balance residual. We also compute an intensity map, an illumination-deficit map, a saturation map, and a local contrast map. These maps together describe brightness loss, color shift, and local structural visibility.

The prior maps are concatenated as

$$\mathbf{P} = [\mathbf{I}_l, \mathbf{I}_{cb}, |\mathbf{I}_{cb} - \mathbf{I}_l|, \mathbf{D}_{ill}, \mathbf{S}, \mathbf{Y}, \mathbf{C}], \quad \mathbf{P} \in \mathbb{R}^{H \times W \times 13}, \tag{4}$$

where $\mathbf{Y}$, $\mathbf{D}_{ill}$, $\mathbf{S}$, and $\mathbf{C}$ denote intensity, illumination deficit, saturation, and local contrast, respectively. The illumination-deficit map is computed from the intensity map as

$$\mathbf{D}_{ill}(x) = \frac{\max(\tau - \mathbf{Y}(x), 0)}{\tau}, \tag{5}$$

where $\tau = 0.5$ in our implementation. Global statistics are computed as a 13-dimensional vector $\mathbf{s}$: the per-channel mean and standard deviation of $\mathbf{I}_l$ (6 values), the per-channel mean of the color-balance residual $|\mathbf{I}_{cb} - \mathbf{I}_l|$ (3 values), and the spatial mean of the illumination-deficit, saturation, intensity, and local contrast maps (4 values). This vector is projected into the condition vector:

$$\mathbf{z} = \tanh\left(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{s})\right), \tag{6}$$

where $\mathbf{s}$ contains global statistics of the prior maps and $\delta(\cdot)$ denotes LeakyReLU, $\mathbf{W}_1$ and $\mathbf{W}_2$ are learnable fully connected layers that map the global prior statistics $\mathbf{s}$ to the degradation condition vector $\mathbf{z}$. The condition vector $\mathbf{z} \in \mathbb{R}^{16}$ is a single global vector per image (not spatial). In every subsequent module, a

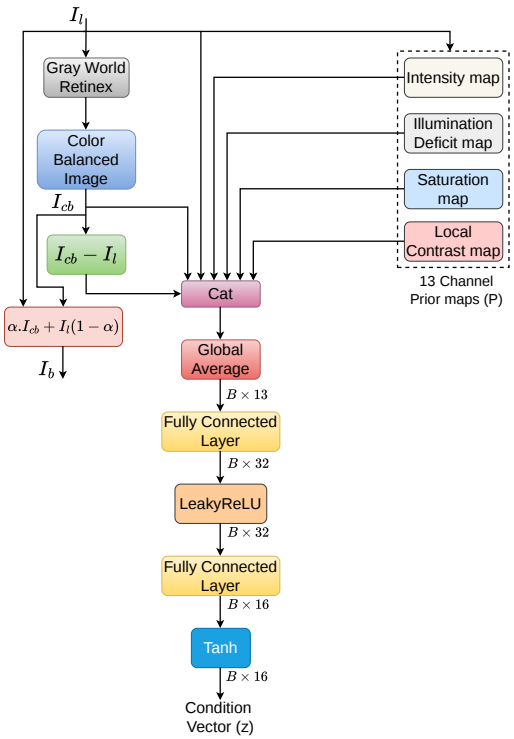


Fig. 2: Low-light prior cue extractor. The input image $\mathbf{I}_l$ is used to generate a color-balanced image $\mathbf{I}_{cb}$, illumination-deficit map, saturation map, intensity map, and local contrast map. These maps are concatenated into a 13-channel prior map $\mathbf{P}$ and summarized into the condition vector $\mathbf{z}$.

linear projection of $\mathbf{z}$ (e.g. $\mathbf{W}_a\mathbf{z}$, $\mathbf{W}_b\mathbf{z}$, $\mathbf{W}_g\mathbf{z}$, $\mathbf{W}_s\mathbf{z}$) produces a per-channel vector that is broadcast spatially across all $H \times W$ locations before being combined with the corresponding feature map. Thus $\mathbf{z}$ modulates each channel uniformly in space but adaptively per image and per channel. The extractor also produces a base image by learnably mixing the input and the color-balanced prior:

$$\mathbf{I}_b = \boldsymbol{\alpha} \odot \mathbf{I}_{cb} + (1 - \boldsymbol{\alpha}) \odot \mathbf{I}_l. \tag{7}$$

3.3 Illumination-Frequency Adapter

The illumination-frequency adapter transforms the prior maps into a prior-adapted image before the main U-Net restoration stage. The motivation is that low-light degradation contains both smooth low-frequency errors, such as illumination and color shift, and high-frequency degradation, such as weak edges and local contrast loss. Therefore, the adapter separates the embedded prior feature

into low- and high-frequency components:

$$\mathbf{F} = \mathcal{E}_a(\mathbf{P}), \quad \mathbf{F}_L = \mathcal{M}_{5\times5}(\mathbf{F}), \quad \mathbf{F}_H = \mathbf{F} - \mathbf{F}_L, \tag{8}$$

where $\mathcal{E}_a(\cdot)$ is a lightweight convolutional embedding module and $\mathcal{M}_{5\times5}(\cdot)$ denotes local average pooling.

The two components are processed by separate lightweight branches and mixed using a gate predicted from the degradation condition vector:

$$\mathbf{F}_a = (1 - \mathbf{g}_a) \odot \mathcal{B}_L(\mathbf{F}_L) + \mathbf{g}_a \odot \mathcal{B}_H(\mathbf{F}_H), \quad \mathbf{g}_a = \sigma(\mathbf{W}_a\mathbf{z}). \tag{9}$$

where $B_L(\cdot)$ and $B_H(\cdot)$ denote the low-frequency and high-frequency processing branches, respectively, and $\mathbf{W}_a$ is a learnable linear projection used to predict the adaptive gate g_a from the condition vector $\mathbf{z}$. The adapted image is then generated as

$$\mathbf{I}_a = \mathbf{I}_b + \boldsymbol{\beta}_a \odot \mathcal{C}_{rgb}(\mathrm{CA}(\mathbf{F}_a)), \tag{10}$$

where $\mathrm{CA}(\cdot)$ denotes channel attention, $\mathcal{C}_{rgb}(\cdot)$ is a 3×3 convolution that maps the feature to RGB space, and $\boldsymbol{\beta}_a \in \mathbb{R}^3$ is a learnable per-channel scaling parameter (initialized to 0.1) that controls the strength of the adapter's correction independently of the degradation condition vector. In this way, the main U-Net receives an input that has already been adjusted using low-light priors.

3.4 Degradation-Adaptive Wavelet-Gradient U-Net

The restoration backbone is a compact U-Net with two encoder stages, a bottleneck, and two decoder stages, as shown in Fig. 3. Unlike a plain U-Net, each main block receives the condition vector $\mathbf{z}$. This allows the network to adapt its wavelet and gradient operations according to the estimated degradation of the input image.

The basic unit is the Low-Light Degradation-Adaptive Wavelet-Gradient Block (LLDAWGBlock). Given a feature map $\mathbf{X}$, the block first applies adaptive wavelet processing and then gradient refinement:

$$\mathcal{B}_{dawg}(\mathbf{X}, \mathbf{z}) = m\,\mathcal{S}_g(\mathcal{W}(\mathbf{X}, \mathbf{z}), \mathbf{z}) + (1 - m)\mathbf{X}, \tag{11}$$

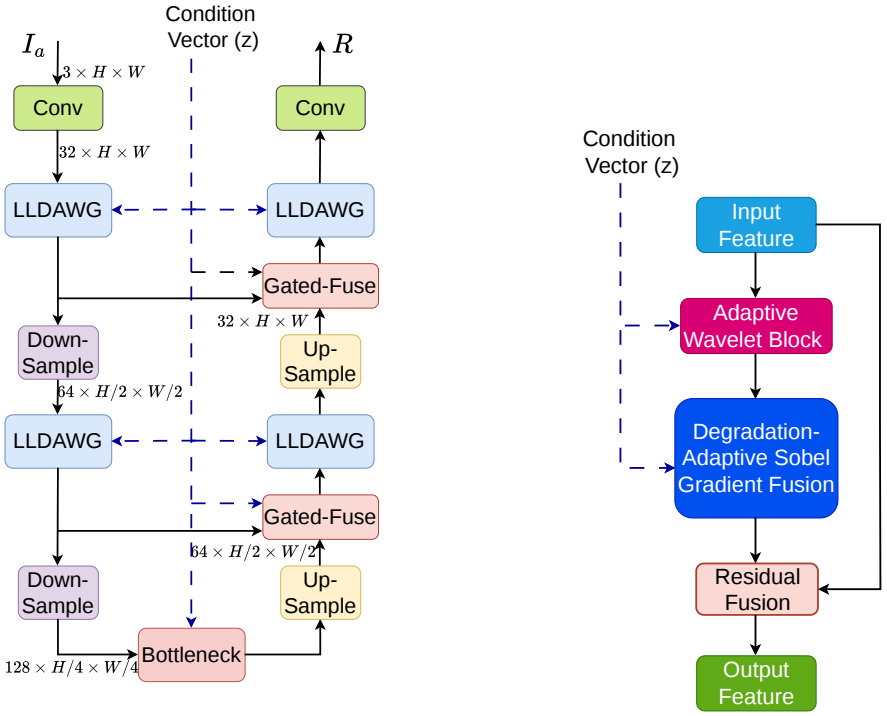
where $\mathcal{W}(\cdot)$ is the adaptive Haar wavelet block, $\mathcal{S}_g(\cdot)$ is the degradation-adaptive Sobel-gradient fusion block, and m is a learnable residual mixing weight.

In the wavelet branch, fixed Haar filters decompose the feature into four subbands:

$$\{\mathbf{X}_{LL}, \mathbf{X}_{LH}, \mathbf{X}_{HL}, \mathbf{X}_{HH}\} = \mathrm{DWT}(\mathbf{X}). \tag{12}$$

The condition vector predicts subband weights,

$$\mathbf{w} = 1 + \frac{1}{2}\tanh(\mathbf{W}_b\mathbf{z}), \tag{13}$$



(a) Degradation-adaptive U-Net backbone.

(b) LLDAWG block.

Fig. 3: Backbone and basic restoration block. The U-Net uses LLDAWG blocks in the encoder and decoder. The condition vector $\mathbf{z}$ guides adaptive wavelet processing and Sobel-gradient refinement inside each LLDAWG block, and controls gated skip fusion in the decoder.

which allows the block to emphasize low-frequency illumination information or high-frequency structural details depending on the input. The weighted subbands are fused and added back to the input feature:

$$\mathcal{W}(\mathbf{X}, \mathbf{z}) = \mathbf{X} + \boldsymbol{\lambda}_w \odot \mathrm{Up}\left(\mathcal{D}_w\left(\mathcal{C}_{1\times1}([w_{LL}\mathbf{X}_{LL}, w_{LH}\mathbf{X}_{LH}, w_{HL}\mathbf{X}_{HL}, w_{HH}\mathbf{X}_{HH}])\right)\right). \tag{14}$$

Here, $\mathcal{C}_{1\times1}(\cdot)$ is a single 1×1 convolution that fuses the four weighted subbands, $\mathcal{D}_w(\cdot)$ is a lightweight depthwise-separable refinement block, and $\mathrm{Up}(\cdot)$ denotes bilinear upsampling back to the original spatial resolution rather than an inverse discrete wavelet transform. The upsampled correction is added as a residual to the input feature $\mathbf{X}$, so the network is not required to reconstruct exact wavelet coefficients; it only needs to learn a spatially-coherent correction informed by the subband decomposition. After wavelet refinement, the block uses Sobel-gradient guidance to recover weak structures. The wavelet-refined feature $\mathcal{W}(\mathbf{X}, \mathbf{z})$ is first processed by a lightweight convolutional refinement block to

obtain a locally refined feature $\mathbf{F}_0$. The gradient magnitude is then computed as

$$\mathbf{G} = \sqrt{(\mathbf{F}_0 * \mathbf{S}_x)^2 + (\mathbf{F}_0 * \mathbf{S}_y)^2 + \epsilon}, \tag{15}$$

where $\mathbf{S}_x$ and $\mathbf{S}_y$ are Sobel kernels. The gradient magnitude is converted into a modulation mask:

$$\mathbf{M} = \sigma\left(\mathcal{C}_g(\mathbf{G})\right), \tag{16}$$

where $\mathcal{C}_g(\cdot)$ is a lightweight convolutional projection and $\sigma(\cdot)$ denotes the sigmoid function. The gradient response is injected through a gate whose strength is controlled by $\mathbf{z}$:

$$\mathbf{F}_1 = a(\mathbf{M} \odot \mathbf{F}_0) + (1-a)\mathbf{F}_0, \quad a = \sigma(a_0 + \mathbf{W}_g\mathbf{z}). \tag{17}$$

Here, a_0 is a single learnable scalar shared across the batch and channels, initialized to zero so that $a = 0.5$ before any degradation conditioning is applied; $\mathbf{W}_g\mathbf{z}$ then shifts this baseline gate per image according to the broadcast convention described in Section 3.2. This avoids applying the same sharpening strength to all images and reduces the risk of amplifying noise in severely dark regions.

3.5 Gated Skip Fusion and Bottleneck Context

A standard U-Net directly transfers encoder features to the decoder. In low-light enhancement, this can also pass dark or color-biased features forward. We therefore use degradation-aware gated skip fusion. Given a decoder feature $\mathbf{D}$ and an encoder skip feature $\mathbf{S}$, the fused feature is

$$\mathcal{G}_{skip}(\mathbf{D}, \mathbf{S}, \mathbf{z}) = \mathcal{R}_s\left(\mathbf{Q} \odot \mathbf{D} + (1-\mathbf{Q}) \odot \mathbf{S}\right), \tag{18}$$

where

$$\mathbf{Q} = \sigma\left(\mathcal{C}_{1\times1}([\mathbf{D}, \mathbf{S}]) + \mathbf{W}_s\mathbf{z}\right). \tag{19}$$

Here, $\mathcal{R}_s(\cdot)$ is a depthwise-separable 3×3 convolution applied to the gated fusion output. The gate decides how much skip information should be retained. At the bottleneck, we further use a lightweight strip-context module consisting of a depthwise 1×7 horizontal convolution followed by a depthwise 7×1 vertical convolution and a 1×1 pointwise mixing convolution, applied as a residual correction with a learnable scale. This captures long-range horizontal and vertical context at low cost compared to a dense 7×7 convolution or self-attention, which suits the elongated structures common in remote sensing scenes such as roads and field boundaries.

4 Experimental Setup

4.1 Datasets

We evaluate the proposed method on the low-light remote sensing datasets introduced in DFFN [1]. The main benchmark is **iSAID-dark**, which was introduced

in DFFN [1] and generated from the iSAID aerial image dataset [9]. Following the original protocol, iSAID-dark contains 3,755 paired training images and 66 validation images. We train the proposed model and the retrained comparison methods on the iSAID-dark training set and report full-reference results on the iSAID-dark validation set.

We also evaluate high-resolution performance on **iSAID-dark(high-pixel)**, which contains 1080p remote sensing images. This setting is useful for testing whether an enhancement model can handle large spatial structures and fine object boundaries at high resolution.

4.2 Evaluation Metrics

For both iSAID-dark and iSAID-dark(high-pixel), we report Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) [10], and Learned Perceptual Image Patch Similarity (LPIPS) [11]. PSNR measures pixel-level fidelity, SSIM evaluates structural similarity, and LPIPS measures perceptual distance between the restored image and the ground truth. Higher PSNR and SSIM indicate better restoration quality, while lower LPIPS indicates better perceptual similarity.

4.3 Implementation Details

All methods are trained or evaluated under the iSAID-dark setting. For LLFormer, FourLLIE, and CIDNet, we retrain the models on the iSAID-dark training set by following their official implementations and recommended training settings, including their original loss functions, learning-rate schedules, and number of epochs. For DFFN, we report the official iSAID-dark retrain result from the original paper, since it follows the same dataset protocol.

For the proposed method and all ablation variants, we use the same training and validation splits, patch size, and evaluation code. The proposed model is trained using randomly cropped image patches of size 256×256 with a batch size of 8 for 500 epochs. We use the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The initial learning rate is set to 8×10^{-4} and is reduced to 1×10^{-6} using cosine annealing after a 3-epoch warm-up stage.

The training objective combines mean-squared error and MS-SSIM loss:

$$\mathcal{L} = \mathcal{L}_{\text{MSE}}(\hat{\mathbf{I}}, \mathbf{I}) + 0.2 \left(1 - \text{MS-SSIM}(\hat{\mathbf{I}}, \mathbf{I})\right), \tag{20}$$

where $\hat{\mathbf{I}}$ is the enhanced image and $\mathbf{I}$ is the corresponding ground-truth image. MS-SSIM is computed with a window size of 11 and data range of 1.

5 Results and Discussion

5.1 Comparison on iSAID-dark

Table 1 reports the quantitative comparison on the iSAID-dark validation set. The proposed method achieves the best PSNR and SSIM among the compared

methods while using the fewest parameters and FLOPs. Compared with DFFN, which is specifically designed for low-light remote sensing enhancement, the proposed model improves PSNR from 25.30 dB to 25.48 dB and SSIM from 0.784 to 0.799. At the same time, it reduces the parameter count from 2.59M to 0.89M and the FLOPs from 15.96G to 7.67G.

DFFN gives the best LPIPS score, indicating stronger perceptual similarity under this metric. However, the proposed model provides a better balance between restoration accuracy and computational cost. These results suggest that degradation-adaptive wavelet-gradient processing is effective for paired low-light remote sensing enhancement.

Table 1: Quantitative comparison on the iSAID-dark validation set. The best result is shown in bold and the second-best result is underlined.

Method	**Params**	**FLOPs**	**PSNR↑**	**SSIM↑**	**LPIPS↓**
LLFormer (AAAI'23) [3]	24.55M	22.03G	23.43	0.728	0.295
FourLLIE (ACM MM'23) [5]	1.49M	16.29G	21.81	0.692	0.301
DFFN (TGRS'24) [1]	2.59M	15.96G	<u>25.30</u>	<u>0.784</u>	**0.151**
CIDNet (CVPR'25) [6]	1.98M	8.13G	24.62	0.776	<u>0.217</u>
Proposed	**0.89M**	**7.67G**	**25.48**	**0.799**	0.224

5.2 High-Resolution Evaluation

Table 2: Quantitative comparison on the iSAID-dark(high-pixel) dataset. The best result is shown in bold and the second-best result is underlined.

Method	**PSNR↑**	**SSIM↑**	**LPIPS↓**
FourLLIE (ACM MM'23) [5]	17.57	0.538	<u>0.224</u>
DFFN (TGRS'24) [1]	22.47	0.721	**0.192**
CIDNet (CVPR'25) [6]	**23.88**	<u>0.739</u>	0.289
Proposed	<u>22.61</u>	**0.752**	0.305

Table 2 reports the results on iSAID-dark(high-pixel). This setting is more challenging because the images have higher resolution and contain larger spatial structures. CIDNet obtains the highest PSNR, while the proposed method achieves the best SSIM. This indicates that the proposed model preserves structural similarity well on high-resolution remote sensing images. DFFN achieves the best LPIPS, suggesting better perceptual similarity under this metric.

Overall, the high-resolution results show that the proposed method remains competitive beyond the main validation set. Its strongest advantage is structural preservation, as reflected by the best SSIM score.

5.3 Ablation Study

We conduct an ablation study to examine the contribution of each component in the proposed model. The baseline is a lightweight U-Net that predicts a residual directly from the low-light input. B1 adds the low-light prior cue extractor and illumination-frequency adapter. B2 further adds the degradation-adaptive wavelet-gradient blocks. The final proposed model additionally introduces degradation-aware gated skip fusion and bottleneck strip context.

Table 3: Ablation study on the iSAID-dark validation set. IFA denotes the illumination-frequency adapter, WG denotes wavelet-gradient processing, GS denotes gated skip fusion, and SC denotes strip context.

Model	**Prior+IFA**	**WG**	**GS**	**SC**	**Params**	**PSNR↑**	**SSIM↑**
Baseline	×	×	×	×	0.57M	25.00	0.793
B1	✓	×	×	×	0.58M	25.09	0.795
B2	✓	✓	×	×	0.81M	25.23	0.796
Proposed	✓	✓	✓	✓	0.89M	**25.48**	**0.799**

Table 3 shows a steady improvement as the proposed components are added. The prior cue extractor and illumination-frequency adapter improve PSNR from 25.00 dB to 25.09 dB and SSIM from 0.793 to 0.795 with almost no increase in parameters. Adding adaptive wavelet-gradient blocks further improves the result to 25.23 dB and 0.796 SSIM. This confirms the value of degradation-conditioned frequency and structural refinement. The full proposed model achieves the best performance, with 25.48 dB PSNR and 0.799 SSIM using only 0.89M parameters. This improvement shows that gated skip fusion and bottleneck strip context help the network use encoder features more selectively and model broader illumination variations.

5.4 Qualitative Comparison

Fig. 4 shows visual results on two representative images selected from the 66-image iSAID-dark validation set. The examples cover different scene structures, including field boundaries, roads, vegetation regions, and aircraft regions. The proposed method improves illumination while retaining local structures. In the first example, it improves road and boundary visibility without strong color distortion. In the second example, it preserves object boundaries and surface details around the aircraft region, which is consistent with the higher PSNR/SSIM values reported below the restored images.

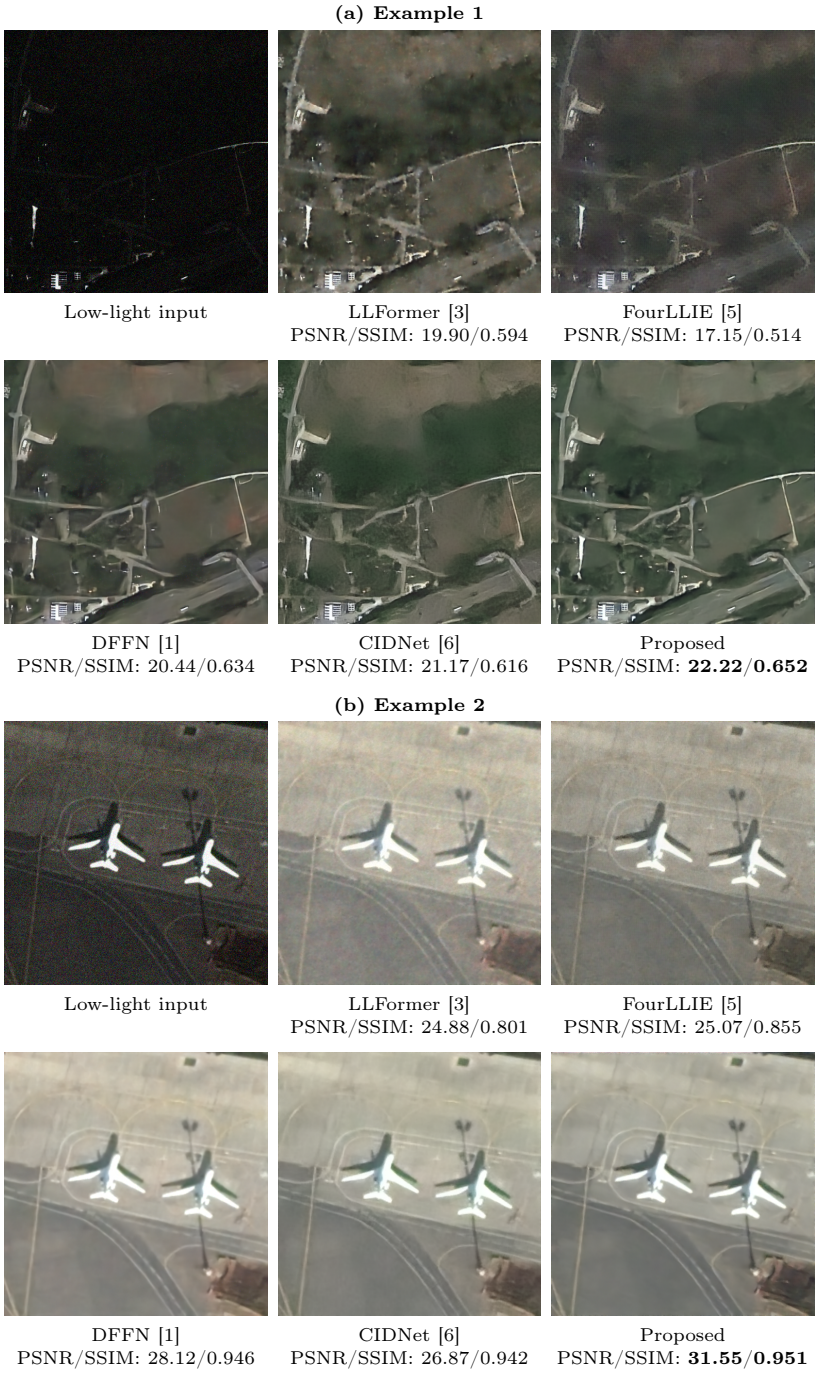


Fig. 4: Qualitative comparison on two representative images from the iSAID-dark validation set. PSNR/SSIM values are reported below the enhanced outputs.

6 Conclusion

In this paper, we proposed LL-DAWGNet, a compact degradation-adaptive wavelet-gradient network for low-light remote sensing image enhancement. The method is built on the idea that low-light degradation should not be handled using a fixed restoration process. Instead, we extract lightweight low-light prior cues and summarize them into a compact condition vector, which guides illumination-frequency adaptation, Haar wavelet subband weighting, Sobel-gradient refinement, and gated skip fusion. This allows the network to adapt its restoration behavior according to the degradation characteristics of each input image.

Experiments on iSAID-dark show that the proposed method achieves the best PSNR and SSIM among the compared methods while using fewer parameters and FLOPs. The high-resolution evaluation on iSAID-dark(high-pixel) further shows that the proposed model preserves structural similarity well in large remote sensing images. The ablation study confirms that the prior cue extractor, adaptive wavelet-gradient processing, gated skip fusion, and bottleneck strip context each contribute to the final performance. Overall, the results show that degradation-adaptive frequency and structural restoration is an effective direction for low-light remote sensing image enhancement.

References

1. Yao, Z., Fan, G., Fan, J., Gan, M., Chen, C.L.P.: Spatial–Frequency Dual-Domain Feature Fusion Network for Low-Light Remote Sensing Image Enhancement. IEEE Trans. Geosci. Remote Sens. **62**, 1–16 (2024). https://doi.org/10.1109/TGRS.2024.3434416
2. Xu, X., Wang, R., Fu, C.-W., Jia, J.: SNR-Aware Low-Light Image Enhancement. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 17714–17724 (2022)
3. Wang, T., Zhang, K., Shen, T., Luo, W., Stenger, B., Lu, T.: Ultra-High-Definition Low-Light Image Enhancement: A Benchmark and Transformer-Based Method. In: Proc. AAAI Conf. Artif. Intell. **37**(3), 2654–2662 (2023)
4. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 12470–12479 (2023). https://doi.org/10.1109/ICCV51070.2023.01149
5. Wang, C., Wu, H., Jin, Z.: FourLLIE: Boosting Low-Light Image Enhancement by Fourier Frequency Information. In: Proc. 31st ACM Int. Conf. Multimedia (ACM MM), pp. 7459–7469 (2023). https://doi.org/10.1145/3581783.3611909
6. Yan, Q., Feng, Y., Zhang, C., Pang, G., Shi, K., Wu, P., Dong, W., Sun, J., Zhang, Y.: HVI: A New Color Space for Low-light Image Enhancement. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 5678–5687 (2025)
7. Cheng, C.-H., Lee, J.-W., Lee, C.-M., Hsu, C.-C.: WWE-UIE: A Wavelet & White Balance Efficient Network for Underwater Image Enhancement. In: Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), pp. 2135–2145 (2026)
8. Wei, C., Wang, W., Yang, W., Liu, J.: Deep Retinex Decomposition for Low-Light Enhancement. In: Proc. British Machine Vision Conf. (BMVC), pp. 1–12 (2018)

9. Zamir, S.W., Arora, A., Gupta, A., Khan, S., Sun, G., Khan, F.S., Zhu, F., Shao, L., Xia, G.-S., Bai, X.: iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), pp. 28–37 (2019)
10. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
11. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 586–595 (2018)
12. Hu, L., Qin, M., Zhang, F., Du, Z., Liu, R.: RSCNN: A CNN-Based Method to Enhance Low-Light Remote-Sensing Images. Remote Sens. **13**(1), 62 (2021)