2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

SEEING IN THE DARK: LOW-LIGHT IMAGE ENHANCEMENT FOR MODULAR CONSTRUCTION

Yuanyang Qi, Jie Hong, Tingtian Li, and Xiao Li
CI^3 Lab, Department of Civil Engineering, The University of Hong Kong, HK SAR, China

Abstract

Modular construction uses off-site modules for on-site assembly; computer vision supports progress monitoring, inspection, and assembly. These tasks may occur in poorly lit spaces, where low light degrades image quality, while prior work has not addressed low-light enhancement for this setting. We propose RetinexRefiner, a two-stage framework that first performs illumination-based enhancement and then refines the result to suppress residual artefacts such as colour cast and noise. Trained end-to-end, it achieves improved visibility on night construction images, reaching 28.07 dB PSNR and 0.83 SSIM, and is applicable to construction sites and similar dim environments.

Introduction

Modular construction transforms the industry by fabricating volumetric units off-site and assembling them on-site (Ferdous et al., 2019; Zheng et al., 2020), with benefits in quality control, productivity, and safety (Tsz Wai et al., 2023). The on-site phase centres on module handling: monitoring progress (e.g. verifying module statuses) and ensuring safety (e.g. detecting proximity between modules and other objects) (Zheng et al., 2020; Ekanayake et al., 2021). Computer vision supports these activities through progress monitoring, inspection, and assembly (Zheng et al., 2020; Fang et al., 2020; Ekanayake et al., 2021; Qi et al., 2025). The accuracy of such vision tasks depends on image quality (Liu et al., 2023; Li et al., 2021). In practice, the work may take place in poorly lit indoor spaces, shaded areas, or at night (Chauhan and Seppänen, 2023; Ekanayake et al., 2021), where low light degrades images and can cause automated checks to miss defects or misestimate progress (Chauhan and Seppänen, 2023; Ekanayake et al., 2021; Li et al., 2021). Reliable imaging in such conditions is therefore important both for safety (e.g. detecting obstacles and workers) and for progress verification (e.g. documenting module placement and assembly state); improving low-light image quality can directly support these vision-based tasks. Prior work on modular construction has focused on module detection and progress monitoring under normal lighting, and most existing vision systems for construction assume well-lit scenes rather than dim or nocturnal conditions (Zheng et al., 2020; Fang et al., 2020; Ekanayake et al., 2021; Qi et al., 2025); low-light image enhancement for this setting has not been addressed. It is thus critical to explore low-light enhancement methods that can improve visibility for monitoring, inspection, and assembly in modular construction.

Low-light image enhancement aims to improve visibility and restore corruptions (noise, artefacts, colour distortion) hidden in the dark or introduced when brightening underexposed photos. Histogram equalisation and gamma correction directly amplify low visibility but barely consider illumination, often causing noise and colour distortion. Retinex-based methods (Land, 1977) decompose an image into reflectance and illumination; LIME (Guo et al., 2016) refines illumination with structure priors but assumes corruption-free inputs, leading to amplified noise in real low-light scenes. DUAL (Zhang et al., 2019a) performs dual illumination estimation for exposure correction. Deep learning methods have advanced the field (Li et al., 2021). RetinexNet (Wei et al., 2018) introduces a deep Retinex decomposition network and the LOL dataset; KinD (Zhang et al., 2019b) decouples illumination and reflectance for degradation removal. Both require multi-stage training. RUAS (Liu et al., 2021) combines Retinex unrolling with neural architecture search for a lightweight enhancement network. RetinexFormer (Cai et al., 2023) proposes a one-stage Transformer with illumination-guided attention blocks (IGAB) that outperform prior Retinex-based CNNs on multiple benchmarks. Other lines of work include zero-reference and physical-prior methods (Guo et al., 2020; Wang et al., 2024), diffusion-based and dual-branch pipelines (Feng et al., 2024), codebook retrieval and flow-based enhancement (Zhou et al., 2024; Wang et al., 2022), zero-shot neural implicit enhancement (Chobola et al., 2024), and learning-based pipelines for raw or extreme dark imaging (Chen et al., 2018; Jiang et al., 2025); recent work also explores new colour spaces for low-light enhancement (Yan et al., 2025). For construction and monitoring, we focus on Retinex-based paired learning so that we have an explicit illumination map and a clear reference (ground truth) for evaluation. Retinex-based methods follow the imaging relation that observed intensity is the product of reflectance and illumination; they offer interpretable decomposition, explicit illumination maps, and a natural fit for paired supervised learning. We build on RetinexFormer because it achieves strong performance

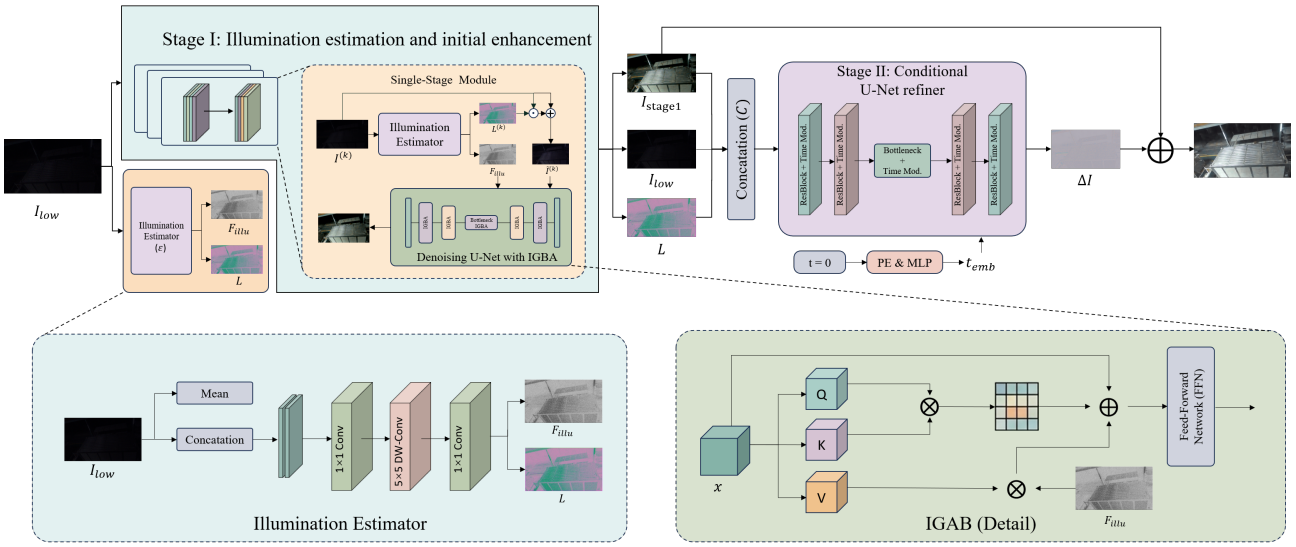

Figure 1: RetinexRefiner two-stage pipeline. Low-light input I_{low} is fed to Stage I (illumination estimation and initial enhancement), which outputs an illumination map L and an initial enhanced image I_{stage1}. Stage II (conditional U-Net refiner) takes $[I_{\text{low}}; I_{\text{stage1}}; L]$ and predicts a residual; the final output is $I_{\text{final}} = I_{\text{stage1}} + \text{Refiner}(\cdot)$. Both stages are trained end-to-end.

among Retinex-based methods and provides an explicit illumination map, which we use to condition the refiner. However, like other one-stage methods, it can leave residual artefacts in very dark regions, such as colour cast, noise, and under- or over-correction.

To address this, we propose RetinexRefiner, a two-stage pipeline that extends RetinexFormer with a refinement stage to correct such artefacts. The pipeline is trained end-to-end on paired normal-light and low-light construction images. Our main contributions can be summarised as follows:

- We address low-light image enhancement for modular construction, where on-site monitoring and assembly may occur in poorly lit or indoor conditions and prior work has not focused on this setting.
- We propose RetinexRefiner, a two-stage pipeline: stage one performs Retinex-based illumination estimation and initial enhancement; stage two uses a conditional U-Net refiner on the low-light image, stage-one output and illumination map to correct residual colour cast and noise, with both stages trained end-to-end on paired images.
- Quantitative and qualitative experiments on night construction images show improved visibility and competitive PSNR/SSIM, outperforming the stage-one baseline and comparison methods.

The paper is organised as follows. Section 2 presents the methodology (two-stage pipeline, Retinex formulation, stage I and stage II). Section 3 describes the experiments, results and ablations. Section 4 discusses limitations; Section 5 concludes and outlines future work.

Methodology

Pipeline overview

Figure 1 summarises the RetinexRefiner pipeline. Given a low-light image I_{low}, the first stage estimates illumination and yields an illumination map L and an initial enhanced image I_{stage1}. The second stage takes I_{low}, I_{stage1}, and L as input and predicts a residual. The final output is

$$I_{\text{final}} = I_{\text{stage1}} + \text{Refiner}(\cdot), \tag{1}$$

where I_{low} is the low-light input image, I_{stage1} is the stage-one enhanced image, L is the illumination map from stage one, I_{final} is the final enhanced output, and $\text{Refiner}(\cdot)$ denotes the conditional U-Net applied to the concatenation $[I_{\text{low}}; I_{\text{stage1}}; L]$. Both stages are differentiable and trained end-to-end.

Retinex-inspired formulation

We follow the Retinex assumption: an image is the element-wise product of reflectance and illumination,

$$I = R \odot L, \qquad R, L \in \mathbb{R}^{3\times H\times W}, \tag{2}$$

where I is the observed image, R is the reflectance, L is the illumination, $\odot$ denotes element-wise multiplication, and H, W are the image height and width. Stage one yields $I_{\text{stage1}} = \mathscr{F}_{\text{RF}}(I_{\text{low}})$ and $L = \mathscr{E}(I_{\text{low}})$. Stage two predicts a residual and adds it:

$$\begin{aligned} \Delta I &= \mathscr{R}([I_{\text{low}}; I_{\text{stage1}}; L], t = 0), \\ I_{\text{final}} &= I_{\text{stage1}} + \Delta I. \end{aligned} \tag{3}$$

where $[\cdot;\cdot;\cdot]$ denotes channel concatenation, $\mathscr{R}$ is the conditional U-Net refiner, ΔI is the predicted residual, t is the time step (fixed at 0 in our setup), and I_{final}, I_{stage1} are the final and stage-one enhanced images. The first stage performs illumination estimation and initial enhancement;

we extend the pipeline with the refiner for the second stage.

Stage I: Illumination estimation and initial enhancement

Stage one follows the one-stage Retinex-based framework of Cai et al. (2023), with three single-stage modules in sequence. At the k-th stage ($k = 1,2,3$), the illumination estimator forms $X_{\text{in}} = [I^{(k)}; \text{mean}_c(I^{(k)})]$ and outputs F_{illu} and $L^{(k)}$:

$$\begin{aligned} F_{\text{illu}}, L^{(k)} &= \mathscr{E}_{\text{stage}}(X_{\text{in}}) \\ &= \phi_2 \circ \phi_{\text{dw}} \circ \phi_1(X_{\text{in}}), \end{aligned} \tag{4}$$

where X_{in} is the concatenation $[I^{(k)}; \text{mean}_c(I^{(k)})]$ with mean_c the channel-wise mean, $I^{(k)}$ is the image at the k-th sub-stage ($I^{(1)} = I_{\text{low}}$), F_{illu} is the illumination feature, $L^{(k)}$ is the light-up map, $\mathscr{E}_{\text{stage}}$ is the illumination estimator, and ϕ_1, ϕ_{dw}, ϕ_2 are 1×1 conv, depth-wise conv, and 1×1 conv respectively. The input is modulated and fed to the denoising U-Net with IGAB:

$$\tilde{I}^{(k)} = I^{(k)} \odot L^{(k)} + I^{(k)}, \quad I^{(k+1)} = \mathscr{D}(\tilde{I}^{(k)}, F_{\text{illu}}) + \tilde{I}^{(k)}, \tag{5}$$

where $\tilde{I}^{(k)}$ is the lit-up image, $I^{(k+1)}$ is the output of the k-th sub-stage, and $\mathscr{D}$ is the denoising U-Net with illumination-guided attention blocks (IGAB). Each IGAB uses

$$\mathbf{A} = \text{softmax}(\mathbf{K}^\top \mathbf{Q} / \sqrt{d}), \quad \text{Attn} = \mathbf{A}(\mathbf{V} \odot F_{\text{illu}}), \tag{6}$$

and

$$\begin{aligned} \text{IGAB}(x, F_{\text{illu}}) &= W_O \text{Attn} + \text{PE}(F_{\text{illu}}) + \text{FFN}(x), \\ \text{FFN}(x) &= W_2 \, \text{GELU}(\text{DWConv}(W_1 x)), \end{aligned} \tag{7}$$

where $\mathbf{A}$ is the attention weight matrix, Attn is the attention output, $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$ are the query, key, and value from the input feature x, x is the input feature to the block, d is the head dimension, W_O, W_1, W_2 are learnable projection matrices, PE is positional encoding, FFN is the feed-forward network, GELU is the Gaussian Error Linear Unit activation, DWConv is depth-wise convolution, and $\odot$ is element-wise multiplication. Thus

$$\begin{aligned} I^{(1)} &= \mathscr{F}_1(I_{\text{low}}), \\ I^{(2)} &= \mathscr{F}_2(I^{(1)}), \\ I_{\text{stage1}} = I^{(4)} &= \mathscr{F}_3(I^{(2)}). \end{aligned} \tag{8}$$

where $\mathscr{F}_k$ denotes the k-th single-stage module ($k = 1,2,3$), and $I^{(4)}$ denotes the output after the third sub-stage (i.e. I_{stage1}). We train stage one jointly with stage two.

Stage II: Conditional U-Net refiner

Stage two takes the condition $C = [I_{\text{low}}; I_{\text{stage1}}; L] \in \mathbb{R}^{9 \times H \times W}$ and fixed $t = 0$. The time embedding is sinusoidal:

$$\mathbf{t}_{\text{emb}} = \text{MLP}(\text{PE}(t)), \tag{9}$$

where t is the time step (fixed at 0), PE is sinusoidal positional encoding, MLP is a multilayer perceptron, and $\mathbf{t}_{\text{emb}}$ is the time embedding vector. Each residual block is modulated by $\mathbf{t}_{\text{emb}}$:

$$\begin{aligned} h &= \sigma(\text{Conv}_1(x) + \text{proj}(\mathbf{t}_{\text{emb}})), \\ \text{ResBlock}(x, \mathbf{t}_{\text{emb}}) &= \sigma(\text{Conv}_2(h)) + \text{skip}(x), \end{aligned} \tag{10}$$

where x is the input to the block, h is the intermediate feature after the first conv and modulation, σ is the SiLU activation, Conv_1, Conv_2 are convolutional layers, proj projects $\mathbf{t}_{\text{emb}}$ to the feature dimension, and $\text{skip}(x)$ is the identity shortcut. The U-Net computes $x_0 = \text{Conv}_{3\times3}(C)$, then encoder ResBlocks with pooling, bottleneck, decoder ResBlocks with skip connections, and

$$\Delta I = \text{Conv}_{3\times3}(u_2), \qquad I_{\text{final}} = I_{\text{stage1}} + \Delta I, \tag{11}$$

where x_0 is the initial feature map from the 3×3 convolution on C, $C = [I_{\text{low}}; I_{\text{stage1}}; L]$ is the 9-channel condition, ΔI is the predicted residual, and u_2 is the decoder feature after the second upsampling block. This single-step refinement improves detail and reduces artefacts. Stage one already handles most of the illumination correction and denoising; the remaining errors are often local and residual (e.g. colour cast, residual noise in dark regions). The refiner, conditioned on I_{low}, I_{stage1}, and L, corrects these with a lightweight U-Net rather than enlarging the first-stage model. Both stages are trained end-to-end on paired normal/low-light images; the training objective and data generation are described in the Experiments section. Algorithm 1 summarises the inference procedure.

Experiments

Datasets

We use a paired dataset: normal-light images (reference, from a construction or site image set) and synthetically darkened low-light images. In line with common practice in supervised low-light enhancement when large-scale real low-light/normal-light pairs are scarce (Wei et al., 2018), we synthesise the low-light side from normal-light references rather than using night-time captures. Low-light images are generated with the darkening curve $y = b \cdot (a \cdot x)^\gamma$, where x is the normal-light pixel value, y is the darkened value, a, b, γ are coefficients sampled per channel (e.g. $a \in [0.38, 0.52]$, $b \in [0.22, 0.38]$, $\gamma \in [1.6, 2.0]$). Because coefficients are randomised per image and channel, some synthetic low-light inputs are intentionally very dark, covering challenging under-exposure. Optionally we add signal-dependent noise variance $\sigma_s^2 \propto L$ (with L the local illumination) and signal-independent noise variance σ_c^2, yielding $z = \text{clip}(y + n_s + n_c)$, where z is the final low-light value, and n_s, n_c are zero-mean Gaussian noise. We refer to the reference set as the normal dataset and the darkened set as the night dataset; the dataset comprises 200 image pairs. We randomly split these pairs into 160 pairs for training and 40 pairs for held-out testing (ratio 4:1). Geometric augmentation (flips, rotation) and crops of 128×128 are used during training.

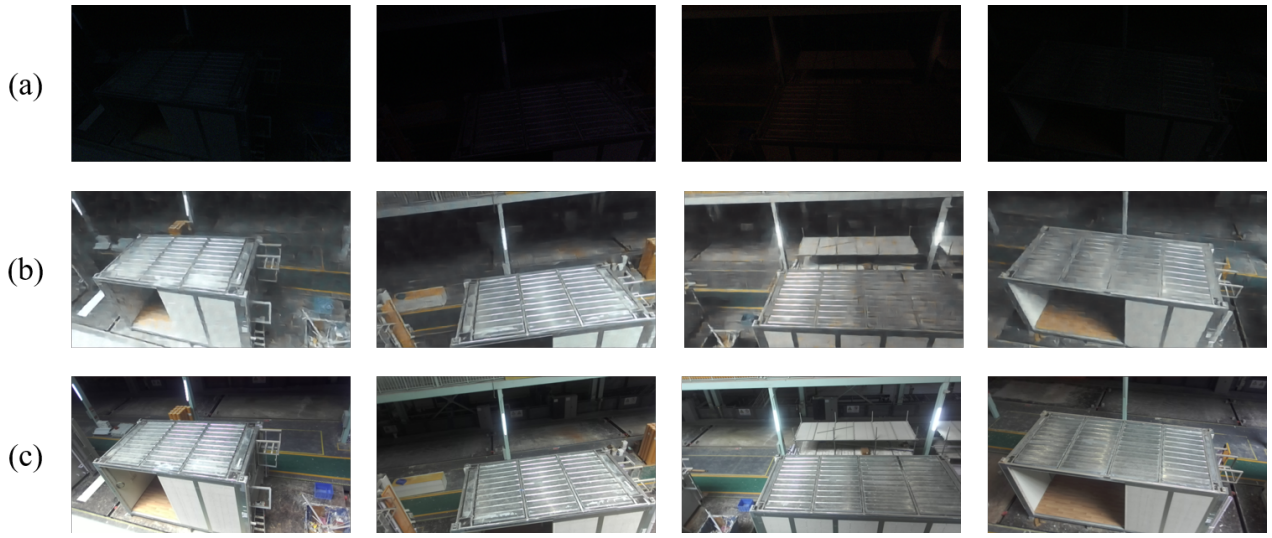

Figure 2: Multiple examples: (a) low-light input, (b) RetinexRefiner output, (c) ground truth. Each column is one sample from the construction dataset.

Implementations

We implement RetinexRefiner in PyTorch. Stage one (illumination estimation and initial enhancement) uses the RetinexFormer backbone (Cai et al., 2023) unchanged; stage two is a conditional U-Net that takes the 9-channel concatenation $[I_{\text{low}}; I_{\text{stage1}}; L]$ and predicts a residual. The refiner has an encoder-decoder structure with skip connections: it downsamples twice (average pooling), processes the bottleneck, then upsamples with bilinear interpolation and concatenates skip features before predicting the residual. The residual blocks are modulated by a sinusoidal time embedding (we fix $t = 0$ at inference). Both stages are trained end-to-end. The training objective is L1 loss between the output and the ground-truth normal-light image; we do not use a separate loss on stage one. We train for 60,000 iterations with Adam (learning rate 1×10^{-4}, betas 0.9 and 0.999), cosine-annealing restart (three periods of 5000 iterations, minimum learning rates 10^{-4} and 10^{-5}), batch size 4, and mixup (beta 1.2) with optional identity mapping. Gradient clipping is applied. Training and model settings are summarised in Table 1.

Table 1: Training and model settings for RetinexRefiner: number of iterations, batch size, crop size, optimizer (Adam with learning rate 1×10^{-4}), and loss (L1 to ground-truth normal-light image).

Setting	Value
Iterations	60k
Batch size	4
Patch size	128×128
Optimiser	Adam (1×10^{-4})
Loss	L1

Evaluation metrics

We use peak signal-to-noise ratio (PSNR, in dB) and structural similarity (SSIM) as evaluation metrics. PSNR measures the ratio between the maximum possible signal and the reconstruction error; higher is better. SSIM captures luminance, contrast, and structure similarity between the enhanced image and the reference; it lies in $[0, 1]$ and higher is better. The normal-light images serve as the reference (ground truth) for both metrics. These are standard full-reference metrics for paired low-light enhancement when references are available.

Quantitative results

We compare with traditional Retinex-based and deep learning methods: LIME (Guo et al., 2016), DUAL (Zhang et al., 2019a), RetinexNet (Wei et al., 2018), KinD (Zhang et al., 2019b), and RetinexFormer (Cai et al., 2023) (stage one only). Table 2 reports PSNR and SSIM on the held-out test set of the construction night dataset. LIME and DUAL, as unsupervised methods that do not optimise for reference matching, yield the lowest scores (PSNR $\approx$ 9.15 dB, SSIM $\approx$ 0.05). Among learning-based methods, RetinexNet reaches 21.50 dB PSNR and 0.69 SSIM; KinD attains 18.16 dB PSNR with higher SSIM (0.82). RetinexFormer (stage one) reaches 27.47 dB and 0.82 SSIM. RetinexFormer is already a strong baseline; further gains are often incremental when stacking a second stage. RetinexRefiner improves over this baseline to 28.07 dB PSNR and 0.83 SSIM (about +0.6 dB and +0.007 SSIM), showing that the conditional U-Net refiner still adds consistent gain. The refiner corrects residual colour cast, noise, and detail loss that remain after stage one. Stage one alone can leave artefacts in very dark regions; the refiner, conditioned on I_{low}, I_{stage1}, and L, both preserves well-enhanced regions and fixes under-corrected ones with modest extra parameters and one extra forward pass.

Algorithm 1 RetinexRefiner inference. Stage I (illumination estimation and initial enhancement) computes the illumination map and runs three sub-stages to obtain I_{stage1}; Stage II builds the 9-channel condition, runs the refiner U-Net, and adds the residual to obtain I_{final}.

Require: Low-light image $I_{\text{low}} \in \mathbb{R}^{3\times H\times W}$
Ensure: Enhanced image $I_{\text{final}} \in \mathbb{R}^{3\times H\times W}$

Stage I: Illumination estimation and initial enhancement
$F_{\text{illu}}, L \leftarrow \mathscr{E}(I_{\text{low}})$ ▷ Illumination map and features
$I^{(1)} \leftarrow \mathscr{F}_1(I_{\text{low}})$ ▷ First sub-stage
$I^{(2)} \leftarrow \mathscr{F}_2(I^{(1)})$ ▷ Second stage (IGAB with F_{illu})
$I_{\text{stage1}} \leftarrow \mathscr{F}_3(I^{(2)})$ ▷ Third stage; initial enhanced image
Stage II: Conditional U-Net refinement
$C \leftarrow [I_{\text{low}}; I_{\text{stage1}}; L]$ ▷ 9-channel condition
$\mathbf{t}_{\text{emb}} \leftarrow \text{MLP}(\text{PE}(t=0))$ ▷ Sinusoidal time embedding
$x_0 \leftarrow \text{Conv}_{3\times3}(C)$ ▷ Initial feature map
Encode: $x_0 \to x_1 \to x_2$ with ResBlocks + pooling, save skip features
Bottleneck: process x_2 with ResBlocks modulated by $\mathbf{t}_{\text{emb}}$ ▷ $h = \sigma(\text{Conv}_1(x) + \text{proj}(\mathbf{t}_{\text{emb}}))$, out $= \sigma(\text{Conv}_2(h)) + \text{skip}(x)$
Decode: upsample, concatenate skips, ResBlocks; obtain u_2 ▷ decoder feature after second upsampling
$\Delta I \leftarrow \text{Conv}_{3\times3}(u_2)$; $I_{\text{final}} \leftarrow I_{\text{stage1}} + \Delta I$ ▷ output equation
return I_{final}

Qualitative results

Figure 2 shows multiple examples from the construction dataset: each column is one sample, with (a) low-light input, (b) RetinexRefiner output, and (c) ground truth (normal-light references are real photographs; low-light inputs are synthetic). The enhanced results recover visibility and detail and are close to the reference. Figure 3 compares methods on the same low-light scene: the top row gives the input and results of LIME, DUAL, and KinD; the bottom row gives the ground truth and results of RetinexNet, RetinexFormer, and the proposed RetinexRefiner. LIME and DUAL exhibit strong colour cast; KinD and RetinexNet improve brightness but remain hazier or less faithful to the reference. The proposed RetinexRefiner is closest to the ground truth in brightness, colour, and detail. In construction and site imagery, text, equipment, and structural details become more visible for monitoring and inspection. These comparisons complement the PSNR/SSIM tables by illustrating perceptual differences that are not fully reflected by those scores alone.

Ablation study

We conduct ablations on the test set of the construction night dataset (Table 3). Training with darkening only yields 16.33 dB PSNR and 0.20 SSIM; adding noise during data synthesis raises performance to 28.07 dB and 0.83 SSIM, showing that noise augmentation improves robustness to real low-light conditions. For the refiner inputs, the full condition $[I_{\text{low}}; I_{\text{stage1}}; L]$ reaches 28.07 dB and 0.83 SSIM. Omitting the illumination map L gives 27.92 dB and 0.83 SSIM; omitting I_{low} gives 27.92 dB and 0.82 SSIM. Both variants are slightly worse than the full model, indicating that L helps the refiner correct residual artefacts and that I_{low} provides useful low-light context. These results validate the design of the conditional refiner.

Table 2: Quantitative comparison (PSNR / SSIM) on the test set of the construction night dataset.

Method	PSNR (dB)	SSIM
LIME (Guo et al., 2016)	9.1515	0.0499
DUAL (Zhang et al., 2019a)	9.1495	0.0522
KinD (Zhang et al., 2019b)	18.1600	0.8158
RetinexNet (Wei et al., 2018)	21.5028	0.6931
RetinexFormer (stage one) (Cai et al., 2023)	27.4707	0.8201
RetinexRefiner (ours)	**28.0747**	**0.8265**

Table 3: Ablation study (test set) on the construction night dataset.

Variant	PSNR (dB)	SSIM
Darkening only	16.3250	0.2033
Darkening + noise	**28.0747**	**0.8265**
Refiner w/o L	27.9220	0.8271
Refiner w/o I_{low}	27.9226	0.8239
Full $[I_{\text{low}}; I_{\text{stage1}}; L]$	**28.0747**	**0.8265**

Discussion

RetinexRefiner contributes to low-light enhancement for construction and built-environment imagery in two ways. First, in terms of application, prior low-light work has not targeted construction or modular assembly; we focus on night construction and similar dim environments for monitoring, inspection, assembly and related vision tasks, where RetinexRefiner can improve visibility over the compared methods. Second, RetinexRefiner provides a two-stage low-light pipeline for construction and built-environment imagery. The main novelty is a lightweight conditional refiner after a Retinex-based backbone: unlike LIME (Guo et al., 2016) and DUAL (Zhang et al., 2019a), we use end-to-end trained stages on paired images; unlike RetinexNet (Wei et al., 2018) and KinD (Zhang et al., 2019b), we train both stages jointly with a single refiner; relative to RetinexFormer (Cai et al., 2023) as stage one, the refiner conditioned on I_{low}, I_{stage1}, and L fixes residual colour cast and noise in dark regions. We position this as a practical refinement module for construction-site imagery under the paired protocol used in this study.

While this study proposes a practical design with measur-

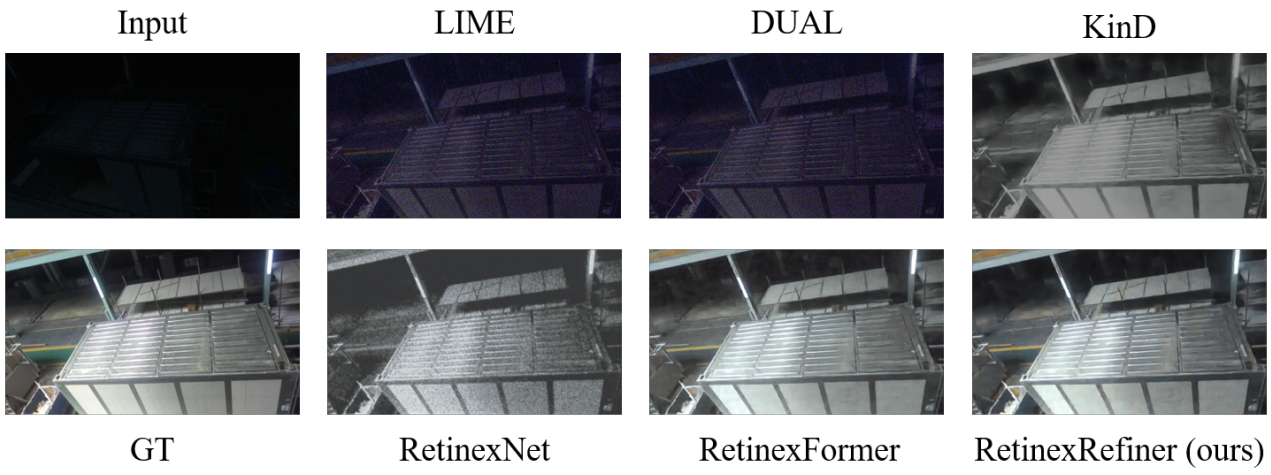

Figure 3: Comparison of methods on one low-light image (construction scene). The low-light input (top left) is a synthetic degradation of the ground-truth normal-light image (bottom left). Top row: input, LIME, DUAL, KinD. Bottom row: ground truth, RetinexNet, RetinexFormer, and the proposed RetinexRefiner.

able gains, it has the following limitations. First, the low-light image in each training/test pair is synthetically generated (darkening and optional noise) from a real normal-light construction or site photograph used as the reference; we do not use pixel-aligned real night-time field captures of the same scene paired with normal-light ground truth. The setup is therefore supervised on synthetic low-light with real references, not unpaired real night-time enhancement. Real camera response and sensor noise may differ from our synthetic low-light branch, so performance may degrade under domain shift; future work includes collecting real paired field imagery where feasible and exploring unpaired enhancement methods. Second, we report only full-reference metrics (PSNR and SSIM); they may not fully reflect perceptual issues such as colour shifts, haloing, or over-smoothing. We do not report no-reference image quality metrics or subjective user studies in this work; these are important directions for future evaluation. Third, the impact of low-light enhancement on downstream tasks such as object detection or semantic segmentation has not been measured—future work will validate this impact. Fourth, we currently compare Retinex-style and classical baselines under a paired protocol; broader cross-family comparisons will be addressed in future work.

Conclusion

In conclusion, we have presented RetinexRefiner, a two-stage Retinex-based method for low-light image enhancement in construction and built-environment contexts. Unlike multi-stage methods that train stages separately or one-stage methods that leave residual artefacts, our pipeline reuses RetinexFormer for illumination estimation and initial enhancement, then applies a lightweight conditional U-Net refiner conditioned on the low-light image, the stage-one output, and the illumination map to predict a residual that improves detail and reduces colour cast and noise; the two stages are trained end-to-end on paired images. Experiments on the construction night dataset show that RetinexRefiner reaches 28.07 dB PSNR and 0.83 SSIM, outperforming the stage-one baseline and comparison methods (LIME, RetinexNet, KinD, RetinexFormer), demonstrating its effectiveness for low-light imagery relevant to modular construction and similar dim environments. The approach is broadly applicable to construction sites, indoor sites, and other scenarios where reliable imaging in the dark is needed.

Future work will focus on: (1) exploring unpaired low-light enhancement methods and using real paired data from construction or field imagery to improve robustness under domain shift; (2) validating the impact of enhancement on downstream tasks such as object detection and semantic segmentation, and reporting task-level metrics (e.g. detection accuracy on enhanced images); (3) incorporating no-reference image quality metrics and subjective user studies to assess perceptual quality beyond PSNR and SSIM.

Acknowledgments

This research was supported by grants from the Innovation and Technology Fund, Hong Kong (Reference No. PRP-048-24LI), and by the Research Institute for Sustainable Urban Development (RISUD), The Hong Kong Polytechnic University (Project No. P0052942). We thank China State Construction Hailong Technology Company Limited for providing the data.

References

Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., and Zhang, Y. (2023). Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 12504–12513. https://doi.org/10.1109/ICCV51070.2023.01149.

Chauhan, I. and Seppänen, O. (2023). Automatic indoor

construction progress monitoring: challenges and solutions. In Proceedings of the European Conference on Computing in Construction, volume 4, pages 0–0. European Council on Computing in Construction. https://doi.org/10.35490/EC3.2023.225.

Chen, C., Chen, Q., Xu, J., and Koltun, V. (2018). Learning to see in the dark. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3291–3300. https://doi.org/10.1109/CVPR.2018.00347.

Chobola, T., Liu, Y., Zhang, H., Schnabel, J. A., and Peng, T. (2024). Fast context-based low-light image enhancement via neural implicit representations. In European Conference on Computer Vision, pages 413–430. Springer. https://doi.org/10.1007/978-3-031-73016-0_24.

Ekanayake, B., Wong, J. K.-W., Fini, A. A. F., and Smith, P. (2021). Computer vision-based interior construction progress monitoring: A literature review and future research directions. Automation in Construction, 127:103705. https://doi.org/10.1016/j.autcon.2021.103705.

Fang, W., Ding, L., Love, P. E., Luo, H., Li, H., Peña-Mora, F., Zhong, B., and Zhou, C. (2020). Computer vision applications in construction safety assurance. Automation in Construction, 110:103013. https://doi.org/10.1016/j.autcon.2019.103013.

Feng, Y., Hou, S., Lin, H., Zhu, Y., Wu, P., Dong, W., Sun, J., Yan, Q., and Zhang, Y. (2024). Diff-flight: Integrating content and detail for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6143–6152. https://doi.org/10.1109/CVPRW63382.2024.00619.

Ferdous, W., Bai, Y., Ngo, T. D., Manalo, A., and Mendis, P. (2019). New advancements, challenges and opportunities of multi-storey modular buildings–a state-of-the-art review. Engineering Structures, 183:883–893. https://doi.org/10.1016/j.engstruct.2019.01.061.

Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., and 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, pages 1780–1789. https://doi.org/10.1109/CVPR42600.2020.00185.

Guo, X., Li, Y., and Ling, H. (2016). Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 26(2):982–993. https://doi.org/10.1109/TIP.2016.2639450.

Jiang, H., Guan, B., Liu, Z., Liu, X., Yu, J., Liu, Z., Han, S., and Liu, S. (2025). Learning to see in the extremely dark. arXiv preprint arXiv:2506.21132. https://doi.org/10.48550/arXiv.2506.21132.

Land, E. H. (1977). The retinex theory of color vision. Scientific American, 237(6):108–129. https://doi.org/10.1038/scientificamerican1277-108.

Li, C., Guo, C., Han, L., Jiang, J., Cheng, M.-M., Gu, J., and Loy, C. C. (2021). Low-light image and video enhancement using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12):9396–9416. https://doi.org/10.1109/TPAMI.2021.3126387.

Liu, H., Jin, F., Zeng, H., Pu, H., and Fan, B. (2023). Image enhancement guided object detection in visually degraded scenes. IEEE Transactions on Neural Networks and Learning Systems, 35(10):14164–14177. https://doi.org/10.1109/TNNLS.2023.3274926.

Liu, R., Ma, L., Zhang, J., Fan, X., and Luo, Z. (2021). Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10561–10570. https://doi.org/10.1109/CVPR46437.2021.01042.

Qi, Y., Li, X., and Jiang, R. (2025). Vision-based segmentation, measurement, and pose estimation in modular integrated construction. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, volume 42, pages 1097–1104. IAARC Publications. https://doi.org/10.22260/ISARC2025/0142.

Tsz Wai, C., Wai Yi, P., Ibrahim Olanrewaju, O., Abdelmageed, S., Hussein, M., Tariq, S., and Zayed, T. (2023). A critical analysis of benefits and challenges of implementing modular integrated construction. International Journal of Construction Management, 23(4):656–668. https://doi.org/10.1080/15623599.2021.1907525.

Wang, W., Yang, H., Fu, J., and Liu, J. (2024). Zero-reference low-light enhancement via physical quadruple priors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 26057–26066. https://doi.org/10.1109/CVPR52733.2024.02462.

Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.-P., and Kot, A. (2022). Low-light image enhancement with normalizing flow. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 2604–2612. https://doi.org/10.1609/aaai.v36i3.20162.

Wei, C., Wang, W., Yang, W., and Liu, J. (2018). Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560. https://doi.org/10.48550/arXiv.1808.04560.

Yan, Q., Feng, Y., Zhang, C., Pang, G., Shi, K., Wu, P., Dong, W., Sun, J., and Zhang, Y. (2025). Hvi: A new color space for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5678–5687. https://doi.org/10.1109/CVPR52734.2025.00533.

Zhang, Q., Nie, Y., and Zheng, W.-S. (2019a). Dual illumination estimation for robust exposure correction. In Computer Graphics Forum, volume 38, pages 243–252. Wiley Online Library. https://doi.org/10.1111/cgf.13833.

Zhang, Y., Zhang, J., and Guo, X. (2019b). Kindling the darkness: A practical low-light image enhancer. In Proceedings of the 27th ACM International Conference on Multimedia, pages 1632–1640. https://doi.org/10.1145/3343031.3350926.

Zheng, Z., Zhang, Z., and Pan, W. (2020). Virtual prototyping-and transfer learning-enabled module detection for modular integrated construction. Automation in Construction, 120:103387. https://doi.org/10.1016/j.autcon.2020.103387.

Zhou, H., Dong, W., Liu, X., Liu, S., Min, X., Zhai, G., and Chen, J. (2024). Glare: Low light image enhancement via generative latent feature based codebook retrieval. In European Conference on Computer Vision, pages 36–54. Springer. https://doi.org/10.1007/978-3-031-73195-2_3.