arXiv:2509.01373v1 [cs.CV] 1 Sep 2025

Unsupervised Ultra-High-Resolution UAV Low-Light Image Enhancement: A Benchmark, Metric and Framework

Wei Lu[1], **Lingyu Zhu**[2], **Si-Bao Chen**[1*]

[1]School of Computer Science and Technology, Anhui University
[2]Computer Science, City University of Hong Kong
luwei_ahu@qq.com, lingyzhu-c@my.cityu.edu.hk, sbchen@ahu.edu.cn

Abstract

Low-light conditions significantly degrade Unmanned Aerial Vehicles (UAVs) performance in critical applications. Existing Low-light Image Enhancement (LIE) methods struggle with the unique challenges of aerial imagery, including Ultra-High Resolution (UHR), lack of paired data, severe non-uniform illumination, and deployment constraints. To address these issues, we propose three key contributions. First, we present U3D, the first unsupervised UHR UAV dataset for LIE, with a unified evaluation toolkit. Second, we introduce the Edge Efficiency Index (EEI), a novel metric balancing perceptual quality with key deployment factors: speed, resolution, model complexity, and memory footprint. Third, we develop U3LIE, an efficient framework with two training-only designs—Adaptive Pre-enhancement Augmentation (APA) for input normalization and a Luminance Interval Loss ($\mathcal{L}_{int}$) for exposure control. U3LIE achieves SOTA results, processing 4K images at 23.8 FPS on a single GPU, making it ideal for real-time on-board deployment. In summary, these contributions provide a holistic solution (dataset, metric, and method) for advancing robust 24/7 UAV vision. The code and datasets are available at https://github.com/lwCVer/U3D_Toolkit.

Introduction

Unmanned Aerial Vehicles (UAVs) are increasingly deployed in mission-critical tasks such as disaster response and urban surveillance. However, their effectiveness is hindered in low-light environments, where captured imagery often suffers from noise, low contrast, and loss of detail. These degradations impair downstream vision tasks and threaten system reliability, making Low-light Image Enhancement (LIE) a critical technology for ensuring robust UAV operations 24/7. Unfortunately, existing LIE methods, mostly developed for ground-level imagery, generalize poorly to aerial views due to four fundamental challenges (Fig. 1): (1) UAVs typically operate at high altitudes, requiring **Ultra-High Resolution (UHR)** to preserve fine-scale details of small ground objects; (2) The dynamic nature of UAV flight makes it infeasible to collect aligned low/normal-light image pairs, necessitating **unsupervised learning** approaches; (3) Aerial night scenes exhibit **extreme non-uniform illumination**, with over-exposed city lights and under-exposed dark zones coexisting in a single frame; (4) **On-board deployment** demands models with low latency, memory footprint, and computational complexity for real-time inference on resource-limited platforms.



Figure 1: Examples from our U3D dataset. Left: low-light night image. Right: non-paired daytime image. The UAV low-light scenes feature: critical small-scale details, unpaired data, deep shadows, and over-exposed lights.

These combined challenges underscore the inadequacy of simply adapting ground-level LIE methods for UAV applications. This necessitates a specialized solution that jointly addresses data availability, aerial-specific visual characteristics, and deployment constraints.

To this end, we present a three-part solution designed to systematically address these challenges. **First, we address the foundational problem of data and evaluation platforms.** We construct and release the **U3D (Unsupervised UHR UAV Dataset)**, the first dataset specifically designed for unsupervised LIE on UAVs. U3D contains 1,000 real-world nocturnal aerial images at 3840×2160 resolution and 1,000 non-paired daytime images of the same size, providing a solid data foundation for the development and validation of unsupervised models. Furthermore, we develop a unified **U3D toolkit** that integrates 12 representative methods to provide a fair and reproducible benchmark for future research. **Second, to more accurately assess the overall performance of algorithms in real-world UAV applications, we propose a new evaluation metric: the Edge Efficiency Index (EEI).** Unlike traditional metrics that focus solely on visual quality, EEI combines enhancement performance with efficiency-related factors, including inference speed, image resolution, parameter count, and computational complexity. The index is designed to provide a more comprehensive measure that is closely aligned with practical deployment needs, thereby guiding algorithm design toward

*Si-Bao Chen is corresponding author.

a better balance between quality and efficiency. **Finally, to tackle the dual challenges of non-uniform illumination and on-board deployment, we propose the U3LIE framework.** U3LIE (Unsupervised UHR UAV LIE) incorporates two main innovations: **1) Adaptive Pre-enhancement Augmentation (APA):** This is a data augmentation technique used during the training phase. It preprocesses extremely dark input images through a pipeline that includes adaptive gamma correction and local contrast enhancement. The goal is to normalize the luminance distribution of the input, simplifying the learning manifold for the network, thereby mitigating over-exposure in bright areas and noise amplification in dark areas during the final enhancement. **2) Luminance Interval Loss ($\mathcal{L}_{int}$):** It is designed to address the critical trade-off between enhancing dark regions and suppressing bright regions. Unlike conventional losses that aim for a single target brightness, our $\mathcal{L}_{int}$ enforces a desirable luminance interval. It simultaneously penalizes local regions that are too dark (falling below a lower threshold) and those that are too bright (exceeding an upper threshold). This mechanism guides the network to restore details in shadows without causing over-exposure in highlights, which is suitable for high dynamic range UAV scenes.

Experimental results on the U3D dataset show that U3LIE achieves an excellent balance between enhancement performance and operational efficiency. Notably, when processing 4K (3840×2160) resolution images, U3LIE runs at 23.8 Frames Per Second (FPS), 3× faster than the competing unsupervised methods such as RUAS (7.6 FPS), while also demanding fewer parameters and FLOPs. This demonstrates its strong potential for practical on-board deployment.

Our main contributions are summarized as follows:

- We construct and release the first unsupervised UHR UAV dataset (U3D) for LIE, along with its accompanying unified evaluation U3D toolkit.
- We propose a novel Edge Efficiency Index (EEI) for the comprehensive evaluation of algorithm quality and deployment efficiency in UAV contexts.
- We design the U3LIE framework, which effectively improves the performance-efficiency balance for UAVs through the APA training strategy and a novel $\mathcal{L}_{int}$.

Related Work

Deep learning-based LIE (Wei et al. 2018) has evolved through several paradigms. Supervised methods, which rely on paired low/normal-light datasets, are often inspired by Retinex theory (Zhang, Zhang, and Guo 2019; Wu et al. 2022) or employ end-to-end architectures that learn direct mappings (Lv et al. 2018; Chen et al. 2018; Wang et al. 2018, 2022). However, acquiring aligned data for dynamic UAVs is challenging. To overcome this, unsupervised approaches leverage unpaired data, often using GANs (Zhu et al. 2017; Jiang et al. 2021; Park et al. 2020; Yang et al. 2020; Zhu et al. 2024b,a). A particularly flexible and lightweight category is zero-reference learning, pioneered by Zero-DCE (Guo et al. 2020). This approach trains models guided only by non-reference losses, yielding a family of efficient methods well-suited for on-board deployment (Liu et al. 2021; Ma et al. 2022; Li, Guo, and Loy 2022; Liang et al. 2023; Shi et al. 2024) and forming the basis of our work.

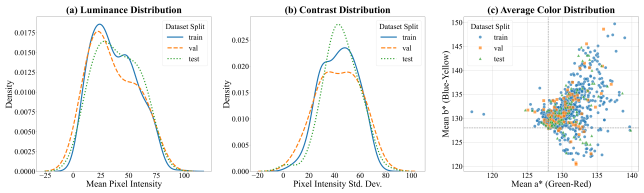


Figure 2: Statistical analysis of the U3D dataset splits. (a) Luminance distribution, (b) Contrast distribution, and (c) Average color distribution in CIELAB space. The high degree of overlap between the train, validation, and test sets demonstrates a consistent and unbiased data partitioning.

However, progress in the aerial domain is constrained by inadequate benchmarks. Existing datasets are either low-resolution (Wei et al. 2018; Chen et al. 2018), designed for supervised ground-level scenes (Wang et al. 2023), or lack the top-down perspective and UHR required for UAVs (Yu et al. 2020; Loh and Chan 2019). No benchmark is tailored for unsupervised UHR UAV enhancement. Crucially, standard evaluation metrics overlook the practical requirements of UAV applications. Metrics focused on fidelity (e.g., PSNR/SSIM (Wang et al. 2004)) or perceptual quality (e.g., NIQE (Mittal, Soundararajan, and Bovik 2012), BRISQUE (Mittal, Moorthy, and Bovik 2012), and PI (Blau and Michaeli 2018)) focus on visual results and disregard edge efficiency—a critical factor for on-board systems.

U3D Dataset and Toolkit

To address data scarcity and the lack of standardized evaluation in UAV LIE, we introduce the **U**nsupervised **U**AV **U**HR **D**ataset (**U3D**) and its evaluation **U3D toolkit**.

The U3D Dataset

U3D is an unpaired dataset designed for unsupervised learning, given the inherent difficulty in capturing perfectly aligned image pairs from dynamic UAV flights. The images were captured in real-world scenarios across diverse locations. It contains 1,000 UHR (3840×2160) low-light images and 1,000 non-paired daytime images, covering diverse urban, residential, and natural scenes. Captured with a professional gimbal-stabilized UAV at altitudes from 30-120m, the images are direct camera outputs, preserving authentic sensor noise and artifacts. To ensure fair and reproducible comparisons, U3D is partitioned into training, validation, and test sets (8:1:1). As shown in Fig. 2, the statistical properties across these splits are consistent, confirming the absence of partitioning bias and supporting fair model evaluation.

The U3D Toolkit

To facilitate standardized and reproducible research, we develop U3D toolkit, a comprehensive and modular evaluation platform. It decouples the research pipeline—including data loading, model architectures, loss functions, and training strategies—into interchangeable modules. This allows

researchers to easily integrate and test new components. To provide a robust baseline for future work, the toolkit already integrates 12 representative models, which are used for comparison in experiments. By providing this unified platform, the U3D toolkit significantly facilitates broader participation for conducting rigorous experiments in this domain.

Edge Efficiency Index (EEI)

Traditional no-reference image quality assessment (NR-IQA) metrics like NIQE (Mittal, Soundararajan, and Bovik 2012) evaluate perceptual quality but ignore computational efficiency—a critical factor for on-board UAV deployment. We introduce EEI to evaluate LIE models for resource-constrained scenarios by balancing perceptual quality with computational cost. EEI systematically balances perceptual quality with a multi-dimensional measure of computational cost. A lower score indicates a more favorable quality-efficiency trade-off. The EEI is formulated as:

$$\text{EEI} = \mathcal{Q}_{\text{PI}} \cdot \mathcal{E}_{\text{norm}}, \tag{1}$$

where $\mathcal{Q}_{\text{PI}}$ is a perceptual quality score and $\mathcal{E}_{\text{norm}}$ is a normalized multi-faceted efficiency factor.

Perceptual Quality ($\mathcal{Q}_{\text{PI}}$). For the quality component, we use the PI, defined as the average of NIQE and BRISQUE scores, to measure image naturalness and artifacts.

Normalized Efficiency ($\mathcal{E}_{\text{norm}}$). The efficiency factor $\mathcal{E}_{\text{norm}}$ aggregates three cost components: runtime, complexity, and memory. To ensure fair comparisons across different hardware, all components are normalized against a standard baseline established by running MobileNetV2 (Sandler et al. 2018) on the target device. This one-time calibration is performed at a fixed base resolution R_{base} of 3840×2160 and yields reference values for inference time (T_{ref}), FLOPs (F_{base}), parameters (P_{base}), and memory usage (M_{base}).

For a model under evaluation with performance metrics $T_{model}, F_{model}, P_{model}, M_{model}$ at resolution R_{model}, we define three normalized cost factors:

Time Factor ($\mathcal{E}_{\text{time}}$): $(T_{model}/T_{ref}) \cdot (R_{base}/R_{model})$, which accounts for both speed and processing resolution.

Complexity Factor ($\mathcal{E}_{\text{comp}}$): An equal-weighted average of normalized FLOPs and parameters: $0.5 \cdot (F_{model}/F_{base} + P_{model}/P_{base})$.

Resource Factor ($\mathcal{E}_{\text{rsrc}}$): Normalized memory usage: M_{model}/M_{base}.

The final efficiency factor, $\mathcal{E}_{\text{norm}}$, is a weighted sum of these factors. To robustly handle cases where profiling FLOPs/parameters fails due to out-of-memory (OOM) errors (common for large models on UHR inputs), we employ a pragmatic adaptive weighting scheme:

$$\mathcal{E}_{\text{norm}} = \begin{cases} w'_t \mathcal{E}_{\text{time}} + w'_r \mathcal{E}_{\text{rsrc}}, & \text{if } \mathcal{E}_{\text{comp}} \text{ OOM}, \\ w_t \mathcal{E}_{\text{time}} + w_c \mathcal{E}_{\text{comp}} + w_r \mathcal{E}_{\text{rsrc}}, & \text{otherwise.} \end{cases} \tag{2}$$

We set the default weights to $w_t = 0.8, w_c = 0.1, w_r = 0.1$ to prioritize inference speed (runtime), which is often the most critical bottleneck for real-time UAV applications such as obstacle avoidance or target tracking. The EEI is a flexible framework; weights can be adapted to specific constraints. For instance, on devices with limited GPU memory, one might increase w_r to penalize high memory usage. This adaptability allows the EEI to serve as a versatile tool for evaluating algorithm suitability across a spectrum of real-world, resource-constrained use cases.

The Proposed U3LIE Framework

This section presents our proposed unsupervised UHR UAV LIE framework, **U3LIE**. The pipeline includes three components: (1) Adaptive Pre-enhancement Augmentation (APA), applied only during training; (2) the enhancement network; and (3) the luminance interval loss $\mathcal{L}_{int}$, which acts as a key regularization constraint. U3LIE operates in an end-to-end manner, enabling efficient enhancement without requiring supervision or priors during inference.

Lightweight Enhancement Network

To ensure on-board efficiency, our design philosophy prioritizes computational performance without compromising visual quality. This principle guides our choice of a highly lightweight enhancement network based on the curve-based methodology of DCE++ (Li, Guo, and Loy 2022). This approach contrasts with the prevailing trend of developing larger models, which, despite their power, often remain impractical for resource-constrained UAVs.

The network learns a pixel-wise enhancement map A to parameterize an iterative, higher-order curve that adjusts the input image's illumination. The final enhanced image is generated by applying this curve. To further optimize for our specific use case, we modify the original DCE++ architecture by: **(1) removing its downsampling/upsampling stages** to process inputs at their native resolution, and **(2) reducing the network's channel depth** from 32 to 8.

Adaptive Pre-enhancement Augmentation (APA)

Directly training on UAV images with extreme dynamic range is challenging. We introduce **APA**, a **training-only, one-time pre-processing** transform, $\mathcal{T}_{APA}$, that generates a moderately-enhanced augmented training set. The pipeline is designed as a two-stage process: first, an adaptive luminance and contrast boost to improve visibility in dark regions, followed by a perceptual correction stage to address color casts and suppress highlights, common artifacts in nocturnal aerial scenes. It simplifies the learning manifold by reducing severe luminance variations and color casts, thus preventing the network from producing over-exposed results or amplifying noise. The APA pipeline is a two-stage process designed to first enhance visibility and then refine perceptual quality, as defined mathematically below.

1. Adaptive Luminance and Contrast Boosting ($\mathcal{T}_{Lum}$). After an initial denoising step $\mathcal{F}_{Denoise}$ using a bilateral filter, the core luminance enhancement is performed in the YCrCb color space for independent control. We first compute the normalized mean luminance, denoted as $\hat{\mu}_Y$, by averaging the pixel values of the input's luminance channel Y and normalizing by the maximum level L_{max} (e.g., 255). This mean value is used to derive an adaptive gamma, γ_{adapt}, which adjusts the enhancement strength based on the

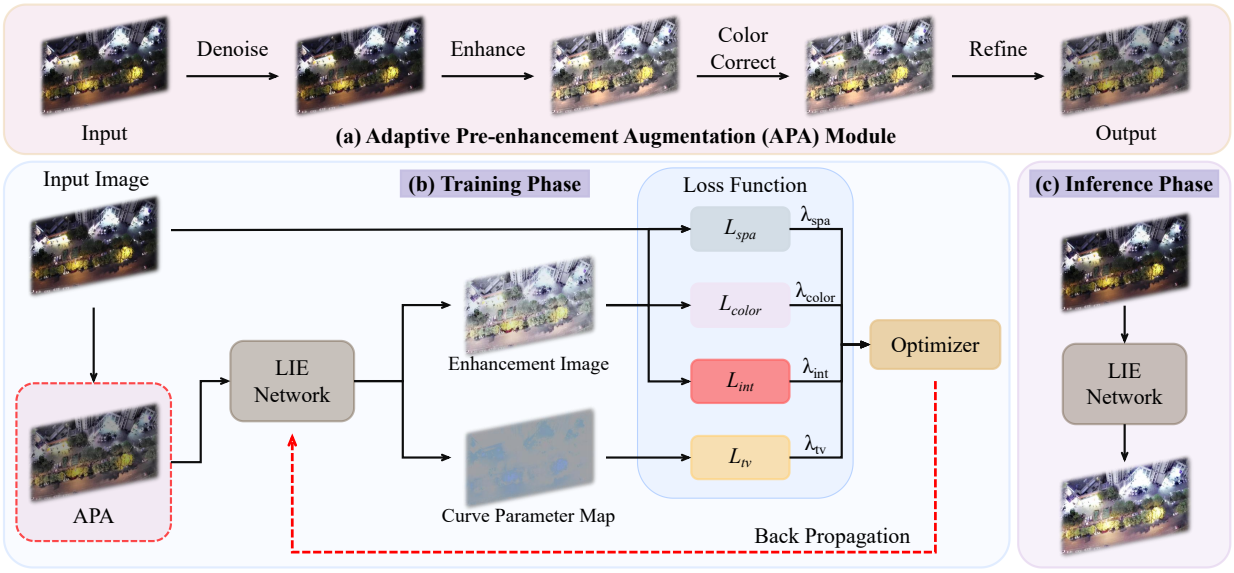


Figure 3: The U3LIE framework. During training (b), a training-only APA module (a) preprocesses input for the network, which is optimized with a loss suite including our $\mathcal{L}_{int}$. During inference (c), the input is processed directly by the trained network.

image's overall darkness:

$$\gamma_{adapt} = \text{clip}\left(\gamma_{base} - \kappa \cdot \ln(\hat{\mu}_Y + \epsilon), [\gamma_{min}, \gamma_{max}]\right). \quad (3)$$

Here, γ_{base} is a base gamma value, κ is a sensitivity hyperparameter, and ϵ is a small constant for numerical stability. The luminance channel is then enhanced via this adaptive gamma correction, followed by Contrast Limited Adaptive Histogram Equalization (CLAHE) (Pizer et al. 1987) to boost local details. The entire transformation on the luminance channel is:

$$Y'(p) = \mathcal{F}_{CLAHE}\left(L_{max} \cdot (Y(p)/L_{max})^{1/\gamma_{adapt}}\right). \quad (4)$$

The final image from this stage, I'', is obtained by merging the enhanced Y' with the original Cr and Cb channels and converting back to the RGB space.

2. Perceptual Color and Highlight Correction ($\mathcal{T}_{Color}$ and $\mathcal{T}_{Final}$). To address common color casts and preserve highlight details, we apply two subsequent correction steps. First, to correct reddish or greenish tints, the image I'' is converted to the CIEL*a*b* color space. The a* channel, which represents the green-red axis, is adjusted using a boost factor β_{red}:

$$a'(p) = (a(p) - c_{mid}) \cdot \beta_{red} + c_{mid}, \quad (5)$$

where $c_{mid} = L_{max}/2$ is the neutral midpoint of the channel. This step yields the image I'''.

Finally, to restore color vibrancy and prevent overexposure, I''' is processed in the HSV space. The saturation (S) and value (V) channels are independently scaled by factors β_{sat} and η_{supp} respectively:

$$S'(p) = \text{clip}(S(p) \cdot \beta_{sat}, [0, L_{max}]), \quad (6)$$

$$V'(p) = \text{clip}(V(p) \cdot \eta_{supp}, [0, L_{max}]). \quad (7)$$

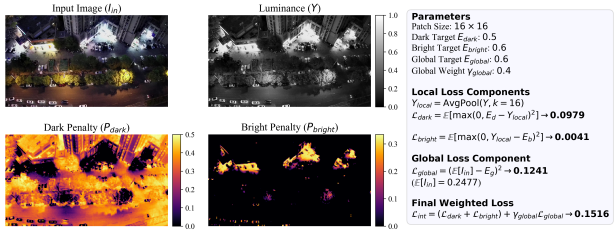


Figure 4: Analysis of $\mathcal{L}_{int}$. Left: Visualization of penalty maps for dark/bright regions. Right: Quantitative summary.

The value suppression in Eq.(7) is crucial for suppressing over-exposed regions, a common artifact in nighttime aerial scenes with artificial lighting. The resulting image I_{aug} serves as the normalized input for network training.

Luminance Interval Loss ($\mathcal{L}_{int}$)

To overcome the limitations of single-target exposure losses, we introduce the Luminance Interval Loss ($\mathcal{L}_{int}$), which enables fine-grained control over luminance. Let I_{enh} be partitioned into a set of non-overlapping patches $\{\mathcal{P}_i\}$. For each patch, we compute its mean luminance $Y_{\mathcal{P}_i} = \mathbb{E}_{p \in \mathcal{P}_i}[Y_{enh}(p)]$. The loss consists of three components:

1. Local Luminance Penalties. To enforce that patch luminances lie within a desired interval $[E_{dark}, E_{bright}]$, we define under- and over-exposure penalties:

$$\mathcal{L}_{local\_dark}(I_{enh}) = \mathbb{E}_i\left[\max(0, E_{dark} - Y_{\mathcal{P}_i})^2\right], \quad (8)$$

$$\mathcal{L}_{local\_bright}(I_{enh}) = \mathbb{E}_i\left[\max(0, Y_{\mathcal{P}_i} - E_{bright})^2\right]. \quad (9)$$

2. Global Brightness Constraint. To regulate the global image brightness, we regularize the global mean luminance

Method	Type	Venue	#P ↓	FLOPs ↓	Memory ↓	FPS ↑	NIQE ↓	BRISQUE ↓	PI ↓	EEI ↓
LLFormer	Sup	AAAI'23	OOM	OOM	5.06*	0.06	4.68	32.73	18.70	6612.54
RetinexFormer	Sup	ICCV'23	OOM	OOM	2.07*	0.33	4.44	31.63	18.04	1208.42
HVI	Sup	CVPR'25	1.97M	1029.00	15.11	0.56	4.94	35.46	20.20	742.42
MobileIE	Sup	ICCV'25	4.05K	31.95	3.11	26.95	4.65	41.34	22.99	20.50
CycleGAN	Unp	ICCV'17	11.38M	7197.0	14.03	1.38	2.97	8.55	5.76	120.74
NeRCo	Unp	ICCV'23	OOM	OOM	3.28*	0.10	4.29	16.21	10.25	2180.89
ZERO-DCE	Uns	CVPR'20	79.42K	656.92	12.89	7.76	3.89	39.43	21.66	69.29
RUAS	Uns	CVPR'21	3.44K	27.10	7.75	7.64	4.83	38.62	21.73	57.91
DCE++	Uns	PAMI'22	10.56K	83.57	17.06	7.68	4.19	37.46	20.82	56.39
SCI	Uns	CVPR'22	10.67K	1451.00	6.92	2.98	4.54	37.38	20.96	167.07
ZERO-IG	Uns	CVPR'24	123.63K	4293.00	15.78	3.09	4.78	37.50	21.14	218.31
Iformer	Uns	ICLR'25	OOM	OOM	2.09*	5.07	5.23	36.72	20.98	2620.69
U3LIE (Ours)	Uns	-	1.32K	9.91	3.23	23.80	4.30	29.92	17.11	16.03

Table 1: Quantitative comparison with SOTA methods on the U3D dataset. All models are evaluated on 4K resolution images (3840×2160). "Sup", "Unp", and "Uns" denote Supervised, Unpaired, and Unsupervised methods, respectively. For supervised (Sup) methods, we directly test using the officially provided pre-trained weights without any fine-tuning. #P: Number of parameters (K: 10^3, M: 10^6). FLOPs: Giga Floating-point operations per second (G: 10^9). Memory: GPU memory usage in Gigabytes (GB). An asterisk (*) indicates that the model was evaluated using a patch-based approach due to memory constraints. OOM denotes an Out-of-Memory error on a single NVIDIA RTX 3090 GPU. ↓: Lower is better. ↑: Higher is better.

$\bar{Y}_{global} = \mathbb{E}_{p\in\Omega}[Y_{enh}(p)]$ towards a target E_{global}:

$$\mathcal{L}_{global}(I_{enh}) = (\bar{Y}_{global} - E_{global})^2. \tag{10}$$

The complete $\mathcal{L}_{int}$ is a weighted sum of these components:

$$\mathcal{L}_{int}(I_{enh}) = (\mathcal{L}_{local\_dark} + \mathcal{L}_{local\_bright}) + \gamma_{\text{global}}\mathcal{L}_{global}. \tag{11}$$

Fig. 4 provides a visual breakdown of how these components operate on a sample image, demonstrating the loss's ability to identify and penalize both under- and over-exposed regions simultaneously.

Total Loss Function

The framework is trained end-to-end by minimizing a total loss function that combines our proposed $\mathcal{L}_{int}$ with a set of established non-reference losses for smoothness of the enhancement map A ($\mathcal{L}_{tv}$), spatial coherence ($\mathcal{L}_{spa}$), and color constancy ($\mathcal{L}_{col}$). The total loss is formulated as:

$$\begin{aligned}\mathcal{L}_{total} =& \lambda_{int}\mathcal{L}_{int}(I_{enh}) + \lambda_{spa}\mathcal{L}_{spa}(I_{enh}, I_{in}) \\ &+ \lambda_{tv}\mathcal{L}_{tv}(A) + \lambda_{col}\mathcal{L}_{col}(I_{enh}),\end{aligned} \tag{12}$$

where $I_{enh} = \mathcal{C}^{(N)}(I_{in}, A)$ and $A = \mathcal{F}_{net}(\mathcal{T}_{APA}(I_{in}))$. The λ terms are the corresponding weights.

Experiments

Implementation Details

Experimental Setup. All experiments were conducted on a single NVIDIA RTX 3090 GPU with 24GB of memory, using the PyTorch-based U3D toolkit on Ubuntu 20.04, and a fixed random seed of 2025 was set.

We used official implementations and default hyperparameters for all control methods. All models were evaluated on full-resolution 4K (3840×2160) images from the U3D test set. For models that encountered Out-of-Memory (OOM) errors, we employed a patch-based evaluation strategy with a 256×256 patch size and 64-pixel overlap, using Hann window blending for stitching. Validation was performed every 5 epochs, and the model with the best EEI score on the validation set was selected for the final test.

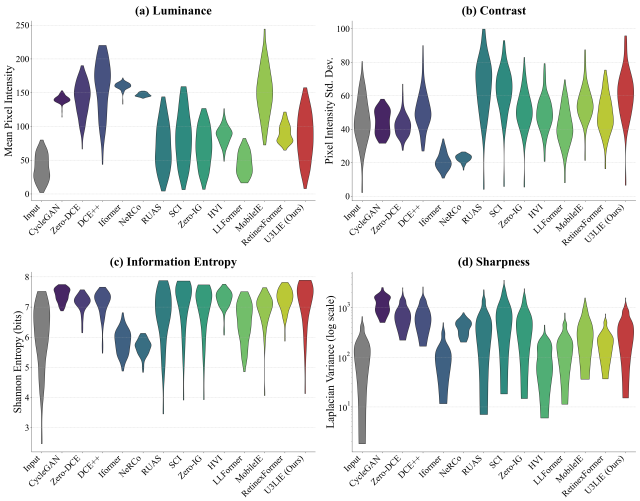


Figure 5: Quantitative analysis via distributional plots. We compare the results across four metrics: (a) Luminance, (b) Contrast, (c) Information Entropy, and (d) Sharpness.

Training Configuration of U3LIE. Our model was trained for 100 epochs using the Adam optimizer with a weight decay of $10e^{-4}$. We employed a 5-epoch linear learning rate warmup, then set the learning rate to $10e^{-5}$, halving it every 50 epochs. We trained on 2048×2048 patches with an effective batch size of 16, achieved through a batch size of 4 and 4 gradient accumulation steps. Gradient clipping was applied with a maximum norm of 0.05. For training, we used

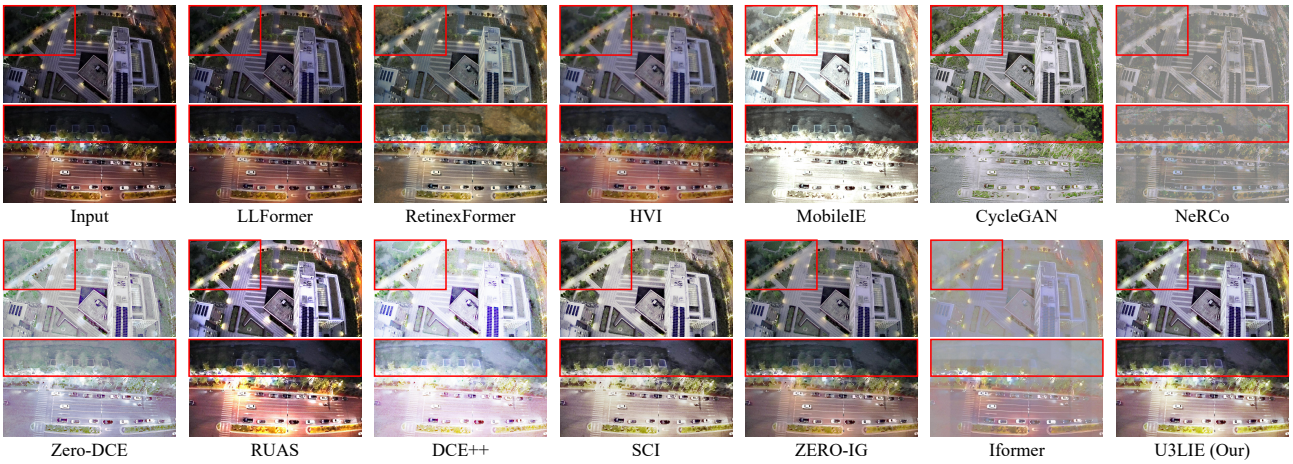


Figure 6: Visual comparison of our U3LIE with SOTA methods on the U3D dataset.

pairs of original images (I_{in}) and their pre-computed APA-processed counterparts (I_{aug}), generated once offline. The network receives I_{aug} as input, while the loss function uses both the enhanced output I_{enh} and the original input I_{in}. The loss weights in Eq.(12) were set as follows: $\lambda_{tv} = 100$, $\lambda_{spa} = 4.0$, $\lambda_{col} = 20.0$, and $\lambda_{int} = 200$. For the $\mathcal{L}_{int}$ term itself, the local patch size was 16×16, with targets for dark, bright, and global luminance set to 0.5, 0.6, and 0.6, respectively, and a global weight $\gamma_{global} = 0.4$.

Quantitative Results

We present a comprehensive quantitative comparison against SOTA methods. For unsupervised evaluation, we include lightweight curve-based methods like Zero-DCE (Guo et al. 2020), RUAS (Liu et al. 2021), DCE++ (Li, Guo, and Loy 2022), SCI (Ma et al. 2022), Zero-IG (Shi et al. 2024), and Iformer (Li, Hu, and Wang 2025). We also support unpaired methods such as CycleGAN (Zhu et al. 2017) and NeRCo (Yang et al. 2023). For comprehensive comparative analysis, we further integrate several supervised models, including LLFormer (Wang et al. 2023), RetinexFormer (Cai et al. 2023), HVI (Yan et al. 2025b), and MobileIE (Yan et al. 2025a). For supervised baselines, we directly used their officially provided pre-trained weights for inference.

Tab. 1 shows that existing methods struggle with 4K aerial imagery. Many large, Transformer-based models (e.g., LLFormer, RetinexFormer) and some unpaired/unsupervised methods (NeRCo, Iformer) fail to process full-resolution images due to Out-of-Memory (OOM) errors on a 24GB GPU. Even when a patch-based strategy is employed (*), their inference speeds (FPS) are far too low for practical real-time applications on UAVs. While lightweight unsupervised methods such as Zero-DCE, RUAS, and DCE++ achieve real-time FPS, they exhibit poor perceptual quality on the U3D dataset, with PI scores all above 20. This indicates their struggle to handle the complex, non-uniform illumination of aerial scenes. Notably, unpaired methods like CycleGAN and NeRCo achieve better PI scores (5.76 and 10.25, respectively), but their large parameter counts and slow inference speeds result in extremely high (poor) EEI scores, making them unsuitable for edge deployment.

In contrast, U3LIE demonstrates a clear advantage across both efficiency and quality. On efficiency, U3LIE is exceptionally lightweight, requiring only **1.32K** parameters and **9.91G** FLOPs. This lightweight architecture enables a remarkable **23.80 FPS** on 4K images, over **3**× **faster** than the nearest real-time competitor (Zero-DCE at 7.76 FPS). On quality, U3LIE achieves a **PI score of 17.11**, significantly outperforming other lightweight methods and comparable to heavyweight GAN-based approaches without introducing their characteristic color artifacts. This resulted in a final **EEI score of 16.03**, a new SOTA that is **3.5**× **better** than the prior best (DCE++, EEI=56.39). This outstanding result validates that our training strategy—the synergy of APA and $\mathcal{L}_{int}$—effectively unlocks the potential of a lightweight backbone for the demanding UHR-UAV context.

Distribution Analysis. Fig. 5 presents a quantitative comparison of enhancement methods across four key image quality metrics. The distributions highlight the nuanced performance and trade-offs of each approach.

Luminance (a): Many methods, including DCE++ and MobileIE, consistently push luminance towards saturation, risking over-exposure and loss of detail in bright areas. In contrast, U3LIE demonstrates superior control, centering its luminance distribution in a perceptually balanced range while maintaining high consistency across the dataset.

Contrast & Entropy (b, c): Crucially, this balanced illumination does not compromise detail restoration. U3LIE achieves top-tier performance in both contrast and information entropy. Its high and tight distributions in these metrics signify robust recovery of global structures and fine textures from dark regions, rivaling or exceeding all other methods.

Sharpness (d): The sharpness analysis reveals a trade-off. CycleGAN exhibits extreme sharpness, often at the cost of amplifying noise and introducing artifacts. U3LIE shows a more appropriate enhancement. It significantly improves clarity over the input while maintaining a controlled, moderate sharpness distribution, indicating effective detail enhancement without sacrificing image naturalness.

In summary, U3LIE excels at enhancing contrast and de-

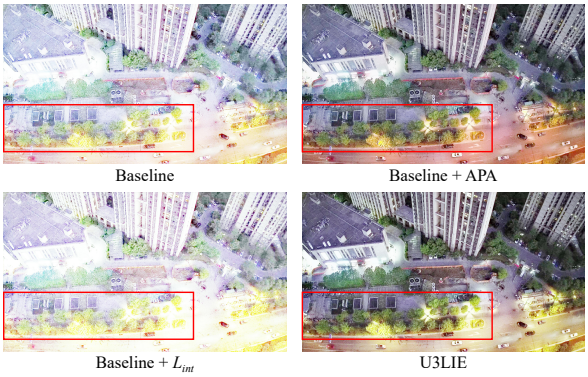


Figure 7: Ablation comparison of U3LIE on U3D test set.

APA	$\mathcal{L}_{int}$	NIQE ↓	BRISQUE ↓	PI ↓
		4.1987	38.0636	21.1311
✓		4.5081	35.9074	20.2077
	✓	4.2167	34.7033	19.4600
U3LIE		4.2954	29.9243	17.1098

Table 2: Ablation study of U3LIE on U3D test set.

tail without introducing over-exposure or noise-related artifacts, demonstrating a superior trade-off.

Qualitative Results

Fig. 6 showcases U3LIE's superior performance on complex aerial scenes with non-uniform illumination.

Supervised methods struggle with generalization to our U3D dataset. Lacking domain-specific training, they often under-enhance shadowed regions (LLFormer, Retinexformer, and HVI) or cause over-exposure and color loss in bright areas (MobileIE).

Unpaired methods such as CycleGAN and NeRCo fail to maintain color fidelity. Their domain translation process introduces severe color artifacts, resulting in color distortion.

Unsupervised methods show mixed results. Lightweight models like Zero-DCE and DCE++ tend to over-expose light sources. RUAS suffers from both color distortion and under-enhancement, while Iformer leads to inconsistent illumination. Although SCI and Zero-IG show improvements, they still fail to match the balance of enhancement and efficiency achieved by our method.

In contrast, U3LIE excels by restoring visibility in dark areas without causing color distortion or over-exposing bright regions. As shown in the red boxes, our method produces results that are both visually realistic and rich in detail, faithfully preserving the original scene structure.

Ablation Study

U3LIE framework ablation. We conducted an ablation study to verify the effectiveness of our two key components: the APA module and the luminance-aware loss $\mathcal{L}int$. As shown in Tab. 2, the baseline DCE++ achieves a PI score of 21.13. Introducing APA alone improves the score to 20.21 by guiding the model to learn from a more consistent illumination distribution during training. Applying $\mathcal{L}int$ alone brings a larger improvement (PI = 19.46), as it directly encourages better luminance balance. When both components are used together in the full U3LIE framework, the PI score further drops to 17.11, demonstrating a clear synergy between APA and $\mathcal{L}int$. Fig. 7 supports these results visually: without APA, the model tends to over-enhance bright regions, while removing $\mathcal{L}int$ leads to color distortion. These findings confirm that both APA and $\mathcal{L}_{int}$ are crucial for achieving stable and natural enhancement.

Loss Weights		Hyperparameters of $\mathcal{L}_{int}$				PI ↓
λ_{col}	λ_{int}	E_{dark}	E_{bright}	E_{global}	γ_{global}	
2	20	0.2	0.8	0.6	0.4	19.3403
2	20	0.3	0.7	0.6	0.6	19.0982
2	10	0.4	0.6	0.6	0.6	19.2415
2	20	0.4	0.6	0.6	0.6	19.0064
2	100	0.4	0.6	0.6	0.6	18.2944
2	200	0.4	0.6	0.6	0.6	16.9(CD)
50	500	0.5	0.6	0.6	0.4	17.1003
20	200	0.5	0.6	0.6	0.4	17.1098

Table 3: Ablation study of $\mathcal{L}_{int}$ and its hyperparameters. In all experiments, the loss weights λ_{tv} and λ_{spa} are fixed to 100 and 4, respectively. CD denotes color distortion.

Hyperparameter Analysis of $\mathcal{L}_{int}$. We ablate the key hyperparameters of the $\mathcal{L}_{int}$ and its associated weights, with results in Tab. 3. Our goal was to find a balance between perceptual quality (low PI) and visual fidelity.

First, we explored the target luminance intervals (E_{dark}, E_{bright}). We found that a narrow, mid-range interval like $[0.4, 0.6]$ outperforms wider or lower-biased ones (e.g., $[0.2, 0.8]$), confirming that constraining local brightness to a plausible mid-tone is effective.

Next, we analyzed the loss weights. The weight of our proposed loss, λ_{int}, is critical. Increasing λ_{int} from 20 to 100 significantly improves the PI score (from 19.01 to 18.29). However, a high λ_{int} combined with a low color loss weight ($\lambda_{col} = 2$) leads to noticeable **color distortion** (Row 6). This occurs because the strong luminance guidance overwhelms the color constancy constraint, sacrificing color accuracy to meet brightness targets. This highlights a key trade-off between luminance correction and color fidelity.

Based on these findings, our final configuration ($\lambda_{tv} = 100, \lambda_{spa} = 4, \lambda_{col} = 20, \lambda_{int} = 200$ and an interval of $[0.5, 0.6]$) strikes an optimal balance. It achieves the best PI score while maintaining natural colors.

Conclusion and Future Work

This paper addresses the challenges of unsupervised, UHR LIE for UAVs by introducing a foundational suite of tools: the **U3D** and **U3D toolkit**, the deployment-aware **EEI** metric, and the highly efficient **U3LIE** framework. U3LIE achieves a SOTA balance of visual quality and real-time performance, providing a robust baseline for developing practical solutions. For future work, we will quantify the benefit to downstream tasks and explore joint optimization to maximize performance in safety-critical UAV applications.

Appendix: Supplementary Materials for "Unsupervised Ultra-High-Resolution UAV Low-Light Image Enhancement: A Benchmark, Metric and Framework"

This supplementary document provides additional implementation and analysis details that complement the main paper. It aims to facilitate reproducibility and deepen understanding of our proposed U3D benchmark, EEI metric, and U3LIE framework.

The appendix is organized as follows:

- **Appendix A** provides a comprehensive analysis and comparison of low-light datasets, including details on the U3D data collection protocol, further justifying its necessity and novelty.
- **Appendix B** presents a detailed breakdown and robustness analysis of the EEI metric.
- **Appendix C** details the architectural design of the APA.
- **Appendix D** includes a discussion of limitations and failure cases.
- **Appendix E** provides additional visual results and qualitative comparisons.

Dataset Comparison and Analysis

The development of robust Low-Light Image Enhancement (LIE) algorithms is critically dependent on the availability of high-quality, relevant benchmark datasets. While numerous datasets have been proposed, they were largely designed for ground-level computer vision tasks and fail to address the unique constellation of challenges posed by Unmanned Aerial Vehicle (UAV) imaging. This appendix provides an in-depth comparative analysis to rigorously position our U3D dataset within the existing landscape and to articulate its foundational importance for the future of 24/7 autonomous aerial systems.

Table 4 presents a detailed, multi-faceted comparison, which is followed by a narrative analysis structured around the key gaps that U3D is designed to fill.

Dataset	Reference	Resolution	Type	Perspective	Scene Characteristics	Primary Task	Annotations	Size (Low)	Key Limitations for UAV LIE
U3D (Ours)	**This Work**	**3840×2160**	**Unpaired**	**Aerial (Top-down)**	**Urban/Suburban, Non-uniform light**	**Enhancement**	**None**	**1,000**	**Specifically designed for this task**
Supervised Ground-Level Datasets									
LOL-v1	(Wei et al. 2018)	400×600	Paired	Ground-level	Indoor/Outdoor, Mixed lighting	Enhancement	None	500	Low-res; Paired data; Ground-level view.
LOL-v2	(Wei et al. 2018)	400×600	Paired	Ground-level	Real & Synthetic scenes	Enhancement	None	789	Same limitations as LOL-v1, still low-resolution.
SID	(Chen et al. 2018)	4240×2832	Paired (RAW)	Ground-level	Indoor/Outdoor, Extreme darkness	Denoising/Enhance.	None	5,094	Focus on RAW data; Paired data; Ground-level.
UHD-LL	(Wang et al. 2023)	3840×2160	Paired	Ground-level	Street/Nightscapes, Non-uniform light	Enhancement	None	8,099	Paired data; Ground-level perspective.
FiveK	(Bychkovsky et al. 2011)	∼5K	Semi-Paired	Ground-level	General scenes, Variable light	Retouching	None	5,000	Not specific to low-light; Paired data is synthetic.
Unsupervised Ground-Level Datasets									
ExDark	(Loh and Chan 2019)	Variable (Low)	Unpaired	Ground-level	Object-centric, 12 categories	Enhancement	B-Boxes	7,363	Low/inconsistent resolution; Lacks complex scenic context.
NPE	(Wang et al. 2013)	Variable (Low)	Unpaired	Ground-level	Mixed scenes, Natural images	Enhancement	None	85	Very small size; Low resolution; Outdated.
DICM	(Lee, Lee, and Kim 2013)	512×384	Unpaired	Ground-level	General photography	Enhancement	None	69	Very low resolution; Too small for deep learning.
LIME	(Guo, Li, and Ling 2017)	Variable (Low)	Unpaired	Ground-level	Mixed scenes for qualitative eval	Enhancement	None	10	A qualitative set, not for training/testing.
Driving-Specific Datasets (Forward-Facing Perspective)									
Dark Zurich	(Sakaridis, Dai, and Van Gool 2019)	1920×1080	Unpaired	Driving	Urban/Suburban driving at dusk/night	Segmentation	Seg. Masks	8,779	Perspective mismatch; Not top-down; Moderate resolution.
BDD100K-Night	(Yu et al. 2020)	1280×720	Unpaired	Driving	Highways, City streets at night	Detection/Seg.	B-Boxes, etc.	∼10,000	Perspective mismatch; Lower resolution (720p).
NightCity	(Tan et al. 2021)	1920×1080	Unpaired	Driving	Urban driving across cities	Segmentation	Seg. Masks	4,297	Perspective mismatch; Focus on high-level tasks.
Aerial Datasets (Not for LIE)									
VisDrone	(Du et al. 2019)	∼1920×1080	N/A	Aerial (UAV)	Mostly daytime scenes	Detection/Tracking	B-Boxes	N/A	Lacks low-light focus; Not for enhancement task.
UAVDT	(Du et al. 2018)	∼1080p	N/A	Aerial (UAV)	Mostly daytime/dusk traffic scenes	Tracking	B-Boxes	N/A	Lacks sufficient night scenes for enhancement.
AU-AIR	(Bozcan and Kayacan 2020)	1920×1080	N/A	Aerial (UAV)	Daytime, Various weather	Detection	B-Boxes	N/A	No low-light data.

Table 4: A detailed and comprehensive comparison of low-light and aerial datasets. U3D is the first to uniquely combine an aerial perspective, ultra-high resolution, an unpaired data structure, and scenes with extreme non-uniform illumination, making it the essential benchmark for real-world UAV LIE.

The Critical Gaps in Existing Benchmarks

Our analysis reveals four fundamental gaps in the existing dataset landscape that have hindered the development of practical LIE models for UAVs. U3D is the first benchmark to systematically address all four.

The Resolution and Detail Gap. UAVs fly at significant altitudes, causing objects of interest on the ground (e.g., people, vehicles, debris) to occupy a very small portion of the image. Consequently, **Ultra-High Resolution (UHR)** is not a luxury but a necessity for any meaningful surveillance or inspection task. The vast majority of established LIE datasets, such as LOL (400×600), ExDark (variable low-res), and BDD100K (1280×720), lack the resolution required to retain these critical details. While UHD-LL provides 4K data, its ground-level perspective and supervised nature render it unsuitable. U3D's provision of 4K aerial imagery directly confronts this gap, enabling the development and evaluation of models that can enhance fine-scale details vital for real-world UAV missions. This UHR constraint also makes computational efficiency, measured by our EEI metric, a first-class citizen in model design.

The Paired-Data Impasse. Supervised learning, powered by datasets like LOL, SID, and UHD-LL, has been a dominant paradigm in image enhancement. However, it relies on a critical assumption: the availability of perfectly aligned low-light and ground-truth normal-light image pairs. For a dynamic, constantly moving UAV, this is a practical impossibility. Factors such as minute changes in flight path, atmospheric conditions, seasonal variations, and alterations in the ground environment between day and night make pixel-perfect alignment infeasible. Therefore, any robust, scalable UAV enhancement solution must be **unsupervised**. U3D is constructed from the ground up as an unpaired dataset, directly reflecting this real-world constraint and fostering the development of unsupervised frameworks like U3LIE that do not depend on unattainable ground-truth pairs.

The Perspective Mismatch. The world looks profoundly different from above. Datasets captured from a ground-level (e.g., LOL, ExDark) or a forward-facing driving perspective (e.g., Dark Zurich, BDD100K) exhibit different object scales, shadow patterns, and light source characteristics. A top-down aerial view, as captured in U3D, features vast, contiguous dark areas (e.g., rooftops, fields, parks) punctuated by intense, point-like light sources (streetlights, headlights). This unique **aerial perspective** creates a distinct image domain. Models trained on ground-level or driving data struggle to generalize to this domain, often failing to handle the unique spatial distribution of light and shadow. U3D provides the first domain-specific benchmark to train and validate models specifically for the challenges of aerial vision.

The Illumination Complexity Gap. Nighttime aerial scenes exhibit a form of **severe, non-uniform illumination** that is more challenging than in typical ground-level imagery. The dynamic range is extreme, with deep, information-starved shadows coexisting with saturated highlights and specular blooming artifacts in a single frame. This is not merely a few isolated bright regions in a dark image; it's about a complex interplay of varied light intensities across a massive field of view. Existing datasets often feature more diffuse or less extreme lighting conditions. The challenging scenes in U3D expose the limitations of conventional loss functions and enhancement techniques, which tend to either amplify noise in dark regions or crush details in bright ones. This complexity necessitates novel approaches, such as our Luminance Interval Loss ($\mathcal{L}_{int}$), which is specifically designed to manage this trade-off and achieve a balanced enhancement across the entire high-dynamic-range scene.

U3D: A Foundational Benchmark for Robust UAV Vision

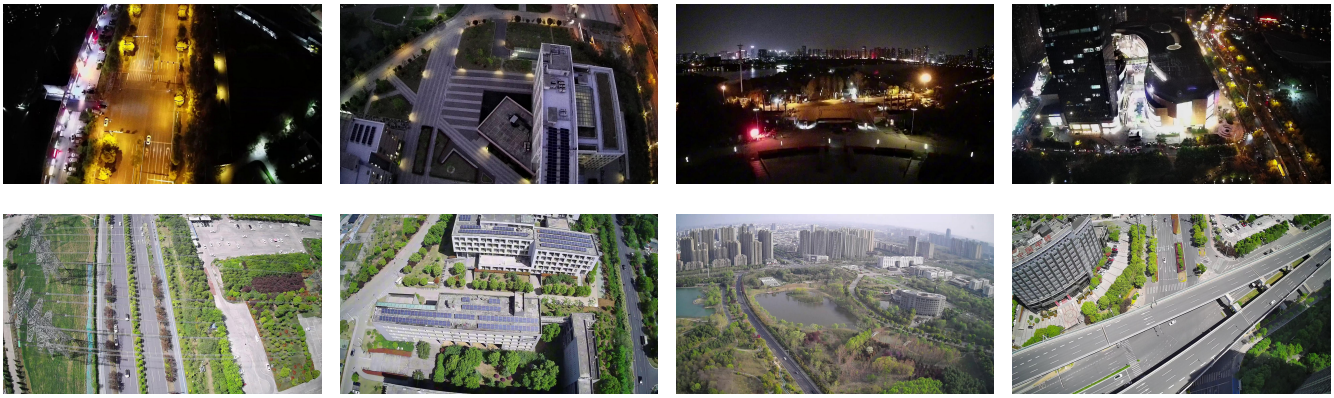


Figure 8: Visual examples from our U3D dataset. The top row showcases UHR low-light images captured by UAVs across diverse scenes. The bottom row presents non-paired daylight images from similar environments, highlighting the dataset's structure for unsupervised learning. The variety of scenes poses significant challenges in terms of non-uniform illumination and detail preservation.

To address the critical gaps in existing benchmarks, we introduce U3D, the first dataset specifically designed for unsupervised, UHR LIE on Unmanned Aerial Vehicles (UAVs).

Dataset Construction and Characteristics. The U3D benchmark comprises **1,000 UHR (3840×2160) low-light images and 1,000 non-paired daytime counterparts.** All images were captured using a professional-grade UAV equipped with a 3-axis gimbal to ensure high quality and minimize motion blur. To ensure high diversity, flights were conducted over a diverse array of real-world environments with altitudes ranging from 30m to 120m and camera angles spanning from forward-looking to nadir (top-down) views. As illustrated in Figure 8, the dataset covers a wide array of scenes, including:

- **Commercial & Business Hubs:** High-density Central Business Districts (CBDs), shopping malls, and brightly lit commercial avenues.
- **Residential Areas:** A mix of dense urban apartment complexes and sprawling suburban neighborhoods.
- **Public & Institutional Zones:** University campuses, schools, public parks, and tourist attractions.
- **Transportation Infrastructure:** Major arterial roads, city streets, and complex interchanges.

Critically, to capture authentic sensor characteristics, all images were recorded in automatic mode, allowing the on-board system to dynamically adjust ISO (typically 800-6400) and shutter speed. No in-camera noise reduction or post-processing was applied, providing direct sRGB outputs in '.jpg' format. This preserves the genuine sensor noise and compression artifacts that real-world models must learn to handle.

Significance and Impact. U3D represents a constructed benchmark that advances the state of the art in UAV-based low-light vision research. Unlike conventional datasets, our benchmark systematically addresses four key challenges in real-world aerial imaging: (1) ultra-high resolution processing, (2) unpaired data learning, (3) perspective variations, and (4) illumination heterogeneity. This provides the first comprehensive framework for developing and evaluating LIE algorithms specifically optimized for UAV platforms. Consequently, our work transitions the research paradigm from limited ground-level scenarios to practical aerial vision systems capable of reliable around-the-clock operation.

Edge Efficiency Index (EEI) Details

This section presents a rigorous, multi-faceted efficiency analysis centered on the proposed Edge Efficiency Index (EEI), a composite metric designed to provide a holistic and hardware-agnostic assessment of on-device model performance. We first define the EEI framework and its components. We then apply it to the `DCE++ light` model, evaluating its performance scaling on a single high-end GPU. Finally, we validate the robustness of the EEI metric itself through a cross-platform comparison involving desktop and edge devices.

The EEI Framework: Definition and Calculation

Our evaluation is based on the Edge Efficiency Index (EEI), defined as: $\text{EEI} = \text{PI} \times (w_T \cdot \text{TF} + w_C \cdot \text{CF} + w_R \cdot \text{RF})$, where PI is the Perceptual Index score, reflecting task-specific performance. To focus on hardware efficiency for this analysis, we normalize the PI score to a constant value of 100. The core of the metric is the Efficiency Factor, a weighted sum of three sub-metrics:

- **Time Factor (TF):** Normalized inference latency, reflecting temporal efficiency.
- **Complexity Factor (CF):** Normalized static model complexity (FLOPs and parameters), reflecting computational cost.
- **Resource Factor (RF):** Normalized peak memory footprint (VRAM/RAM), reflecting runtime resource consumption.

The weights w_T, w_C, w_R (summing to 10) represent the relative importance assigned to each factor. We propose the **8:1:1** ($w_T = 8, w_C = 1, w_R = 1$) **weighting as a pragmatic default** for evaluating on-device models. This policy allocates 80% of the importance to latency (TF), reflecting the paramount importance of user-perceived speed. The remaining 20% is divided between model complexity (CF) and memory footprint (RF), acknowledging them as critical but secondary constraints. This weighting scheme emphasizes latency (80%) to reflect its primary impact on user experience and real-time constraints, while accounting for model complexity and memory usage as secondary but important deployment factors.

Detailed Factor Calculation The following section provides the precise mathematical formulation for each factor, as implemented in our evaluation script.

Baseline Definition All factors are normalized against a hardware-specific baseline to ensure fair, cross-device comparisons. This baseline is established by running a standard MobileNetV2 model on a 4K resolution input (3840×2160 pixels) on the target hardware. The following reference values are obtained:

- Time_{ref}: The average inference time (in seconds) of MobileNetV2 at 4K.
- $\text{FLOPs}_{\text{ref}}$: The floating-point operations of MobileNetV2 for a 4K input.
- $\text{Params}_{\text{ref}}$: The number of trainable parameters in MobileNetV2.
- Mem_{ref}: The peak memory consumption (VRAM on GPU, RSS increment on CPU) of MobileNetV2 at 4K.

Time Factor (TF) The Time Factor measures the model's temporal efficiency relative to the hardware's baseline performance, normalized for resolution. It quantifies how fast the model processes each pixel compared to the baseline model. A lower TF is better.

$$\text{TF} = \frac{\text{Time}_{\text{model}}}{\text{Time}_{\text{ref}}} \times \frac{\text{Pixels}_{\text{ref}}}{\text{Pixels}_{\text{model}}}$$

where $\text{Time}_{\text{model}}$ is the measured inference time of the evaluated model, and Pixels denotes the total pixel count for the respective resolution.

Complexity Factor (CF) The Complexity Factor measures the static, hardware-independent complexity of the model. A lower CF is better. It is a weighted average of the normalized FLOPs and parameters:

$$\text{CF} = 0.5 \times \frac{\text{FLOPs}_{\text{model}}}{\text{FLOPs}_{\text{ref}}} + 0.5 \times \frac{\text{Params}_{\text{model}}}{\text{Params}_{\text{ref}}}$$

Resource Factor (RF) The Resource Factor measures the model's runtime memory consumption relative to the baseline. A lower RF is better. It is a direct ratio of the peak memory usage:

$$\text{RF} = \frac{\text{Mem}_{\text{model}}}{\text{Mem}_{\text{ref}}}$$

where $\text{Mem}_{\text{model}}$ is the peak memory usage of the model at its **specific test resolution**, while Mem_{ref} is from the baseline at a **fixed 4K resolution**. This intentionally penalizes models whose memory usage scales poorly with input size.

Algorithmic Implementation

Algorithm 1: EEI (Edge Efficiency Index) Calculation Pipeline

Require: Model M, Test Image Size S_{test}, Device D
Ensure: EEI score

function GET-HARDWARE-BASELINE(D)
  $M_{base} \leftarrow$ MobileNetV2 on device D
  $S_{base} \leftarrow (1, 3, 3840, 2160)$ ▷ Fixed 4K baseline size
  $Input_{base} \leftarrow$ Random tensor of size S_{base}
  try:
    $(FLOPs_{base}, Params_{base}) \leftarrow$ PROFILE$(M_{base}, Input_{base})$
  except Out-of-Memory (OOM):
    $(FLOPs_{base}, Params_{base}) \leftarrow$ (None, None)
  $Mem_{base} \leftarrow$ MEASURE_PEAK_MEMORY$(M_{base}, Input_{base})$
  $Time_{ref} \leftarrow$ MEASURE_AVG_INFERENCE_TIME$(M_{base}, Input_{base})$
  return ($Time_{ref}$, $FLOPs_{base}$, $Params_{base}$, Mem_{base})

function GET-MODEL-PERFORMANCE(M, S_{test}, D)
  $Input_{test} \leftarrow$ Random tensor of size S_{test}
  try:
    $(FLOPs_{model}, Params_{model}) \leftarrow$ PROFILE$(M, Input_{test})$
  except OOM:
    $(FLOPs_{model}, Params_{model}) \leftarrow$ (None, None)
  try:
    $Time_{model} \leftarrow$ MEASURE_AVG_INFERENCE_TIME$(M, Input_{test})$
    $Mem_{model} \leftarrow$ MEASURE_PEAK_MEMORY$(M, Input_{test})$
  except OOM: ▷ Fallback to patch-based processing
    $(\_, Time_{model}, Mem_{model}) \leftarrow$ PROCESS_IMAGE_ROBUSTLY$(M, Input_{test})$
  return ($Time_{model}$, $FLOPs_{model}$, $Params_{model}$, Mem_{model})

▷ — Main Calculation —
$(Time_{ref}, FLOPs_{base}, \dots) \leftarrow$ GET-HARDWARE-BASELINE(D)
$(Time_{model}, FLOPs_{model}, \dots) \leftarrow$ GET-MODEL-PERFORMANCE(M, S_{test}, D)
$PI \leftarrow$ CALCULATE-PERCEPTUAL-INDEX(M, S_{test}) ▷ Using NIQE & BRISQUE

$Pixels_{test} \leftarrow \prod S_{test}$; $Pixels_{base} \leftarrow \prod S_{base}$
$TF \leftarrow (Time_{model}/Time_{ref}) \times (Pixels_{base}/Pixels_{test})$ ▷ Time Factor
$RF \leftarrow Mem_{model}/Mem_{base}$ ▷ Resource Factor

if $FLOPs_{model}$ **and** $Params_{model}$ are available **then**
  $CF \leftarrow 0.5 \times (FLOPs_{model}/FLOPs_{base}) + 0.5 \times (Params_{model}/Params_{base})$ ▷ Complexity Factor
  $E_{norm} \leftarrow 0.8 \cdot TF + 0.1 \cdot CF + 0.1 \cdot RF$
else ▷ Fallback for OOM during profiling
  $E_{norm} \leftarrow 0.9 \cdot TF + 0.1 \cdot RF$

$EEI \leftarrow PI \times E_{norm}$
return EEI

To ensure clarity and facilitate adoption by the community, we formalize the entire EEI calculation process in Algorithm 1. The pipeline is designed to be robust and hardware-agnostic. It begins by establishing a hardware-specific baseline using a

standard MobileNetV2 model at a fixed 4K resolution (Lines 1-11). Then, it measures the performance of the model under evaluation (M) at its specific test resolution (S_{test}) (Lines 12-23).

A key feature of our implementation is its robustness to Out-of-Memory (OOM) errors, which are common when evaluating large models on UHR inputs. If profiling or full-size inference fails due to OOM, the algorithm gracefully degrades: it can fall back to a patch-based processing mode for timing measurements (Line 22) and automatically switches to a complexity-agnostic weighting scheme for the final EEI calculation (Line 35). This ensures that all models can be fairly evaluated, even those that cannot run directly on high-resolution inputs.

Experimental Analysis of `DCE++ light`

We now apply the EEI framework to evaluate the `DCE++ light` model. The analysis is two-fold: first, we examine its performance scalability and the impact of different weighting policies on a single platform; second, we assess the consistency of the EEI metric across diverse hardware.

Performance Scaling and Weighting Ablation on a Single Platform The initial evaluation was conducted on an NVIDIA GeForce RTX 3090 GPU across six resolutions, from 480p to 8K. The results, summarized in Table 5, reveal the model's efficiency profile under varying computational loads.

Resolution	Time (ms)	FPS	FLOPs (G)	Mem (GB)	TF	CF	RF	EEI Score for Weight Scenario				
								8:1:1	9:0.5:0.5	6:2:2	9:0:1	10:0:0
854×480	2.07	482.32	0.49	0.18	1.03	0.01	0.08	83.45	93.38	63.58	93.73	103.31
1280×720	4.69	213.19	1.10	0.36	1.04	0.01	0.15	84.79	94.38	65.62	95.08	103.96
1920×1080	10.48	95.45	2.48	0.72	1.03	0.02	0.30	85.83	94.51	68.45	95.92	103.20
2560×1440	18.47	54.14	4.41	1.26	1.02	0.04	0.53	87.63	94.99	72.91	97.45	102.35
3840×2160	41.82	23.91	9.91	2.76	1.03	0.09	1.16	94.96	98.98	86.92	104.34	102.99
7680×4320	167.39	5.97	39.65	10.78	1.03	0.37	4.55	131.65	117.35	160.25	138.28	103.05

Table 5: EEI evaluation of the `DCE++ light` model (PI Score = 100) on an NVIDIA GeForce RTX 3090 GPU. Time Factor (TF), Complexity Factor (CF), and Resource Factor (RF) are shown. EEI scores are presented for different weight scenarios: Default (8:1:1), 9:0.5:0.5, 6:2:2, No CF (9:0:1), and Time Only (10:0:0).

The physical metrics (Time, FLOPs, Memory) scale predictably with the input pixel count. An 81-fold increase in pixels from 480p to 8K results in a commensurate ˜81-fold increase in latency and FLOPs, confirming the model's linear computational scalability. However, peak memory usage grows substantially, reaching 10.78 GB at 8K, which may be a constraint for memory-limited devices. The normalized factors provide deeper insights. The **Time Factor (TF)** remains stable around 1.03 across all resolutions, indicating consistent per-pixel processing efficiency relative to the baseline. The **Complexity Factor (CF)** stays well below 1.0, confirming the model's lightweight design. In stark contrast, the **Resource Factor (RF)** escalates from 0.075 to 4.554, identifying memory consumption as the primary efficiency bottleneck at ultra-high resolutions, as it scales significantly faster than the baseline.

The ablation study on EEI weighting policies further illuminates this profile. This study is not intended to find a single "best" score, but rather to demonstrate how the perceived efficiency of a model changes based on application-specific priorities.

- **6:2:2.** This policy gives greater weight to complexity and resources. Consequently, `DCE++ light` achieves its lowest EEI scores under this policy for resolutions up to 4K, as the policy rewards the model's primary strengths (low CF and RF). However, this trend reverses at 8K, where the extreme RF value (4.554) is heavily penalized, resulting in the worst EEI score (160.25). This inversion powerfully illustrates that a model optimal for one set of constraints can be suboptimal for another.
- **"Latency-Priority" (9:0.5:0.5).** By further emphasizing speed, this policy highlights the model's stable but not superior TF (˜1.03). The scores are consistently higher than the Default policy at lower resolutions but become more favorable at 8K, where the impact of the high RF is minimized.
- **"Time-Only" (10:0:0).** In this boundary case, the EEI score is governed solely by TF. The stable score of ˜103 across all resolutions isolates the model's temporal performance, confirming its resolution-agnostic per-pixel efficiency if all other constraints were ignored.

Cross-Platform Consistency of the EEI Metric To validate the robustness of the EEI framework, we extended the evaluation to two additional platforms: a modern desktop NVIDIA RTX 4060 GPU and an edge-computing NVIDIA Jetson Xavier device. The results are presented in Table 6.

Resolution	GPU	Time (ms)	FPS	FLOPs (G)	Mem (GB)	TF	CF	RF	EEI Score for Weight Scenario				
									8:1:1	9:0.5:0.5	6:2:2	9:0:1	10:0:0
1920×1080	RTX 3090	10.48	95.45	2.48	0.72	1.03	0.02	0.30	85.83	94.51	68.45	95.92	103.20
	Xavier	94.54	10.58	2.48	0.69	1.08	0.02	0.29	89.10	98.27	70.75	99.61	107.45
	RTX 4060	32.20	31.06	2.48	0.70	1.09	0.02	0.30	90.69	100.04	72.00	101.40	109.39
2560×1440	RTX 3090	18.47	54.14	4.41	1.26	1.02	0.04	0.53	87.63	94.99	72.91	97.45	102.35
	Xavier	170.57	5.86	4.41	1.22	1.09	0.04	0.52	92.83	100.94	76.61	103.32	109.04
	RTX 4060	58.02	17.23	4.41	1.24	1.11	0.04	0.53	94.38	102.63	77.88	105.06	110.88
3840×2160	RTX 3090	41.82	23.91	9.91	2.76	1.03	0.09	1.16	94.96	98.98	86.92	104.34	102.99
	Xavier	391.77	2.55	9.91	2.76	1.11	0.09	1.17	101.69	106.50	92.06	111.90	111.31
	RTX 4060	130.90	7.64	9.91	2.73	1.11	0.09	1.16	101.46	106.32	91.75	111.66	111.18

Table 6: Comparative EEI evaluation of the `DCE++ light` model across different GPUs (RTX 3090, Jetson Xavier, RTX 4060) and resolutions. Time Factor (TF), Complexity Factor (CF), and Resource Factor (RF) are shown, with a fixed PI Score of 100. EEI scores are presented for various weight scenarios: Default (8:1:1), 9:0.5:0.5, 6:2:2, No CF (9:0:1), and Time Only (10:0:0).

A key finding from this cross-platform analysis is the remarkable consistency of the EEI score. Despite significant disparities in raw processing capabilities—as evidenced by the large variations in Time (ms) and FPS—the final EEI scores for any given resolution and weighting scenario remain exceptionally stable. For instance, at 1080p with the default 8:1:1 policy, the EEI scores are 85.83 (3090), 89.10 (Xavier), and 90.69 (4060). Specifically, the relative variation in EEI scores across the three platforms is consistently maintained within a 10% margin for all tested configurations. This demonstrates the robustness of the EEI framework as a hardware-agnostic metric, providing a fair and reliable assessment of a model's intrinsic efficiency, independent of the underlying computational power of the testbed.

Overall Conclusion and Implications

Our comprehensive evaluation demonstrates that `DCE++ light` is an exceptionally lightweight model with scalable temporal performance. Its primary efficiency bottleneck is memory consumption at resolutions of 4K and higher. The analysis of EEI weighting policies underscores a critical concept: **model efficiency is not an absolute measure but is context-dependent.** A model's performance profile must be evaluated against the specific priorities—be it latency, memory, or model size—of the target application. The proposed 8:1:1 default policy offers a balanced and practical standard for such evaluations.

Furthermore, the consistent EEI scores across diverse hardware validate the metric's utility for fair, hardware-agnostic comparisons. For higher-resolution deployments, future work should prioritize memory optimization to address the identified bottleneck.

Architectural Design of the APA Module

The Adaptive Pre-enhancement Augmentation (APA) module is a computational pre-processing framework designed to transform low-light aerial imagery into a domain more amenable to neural network training. Its architecture is a cascaded, multi-domain pipeline that systematically mitigates common photometric degradations (e.g., low contrast, noise) and chromatic aberrations (e.g., color casts) while preserving perceptual fidelity. The design philosophy is to create a moderately-enhanced yet naturalistic training proxy, thereby simplifying the learning task for the subsequent lightweight network. The process is decomposed into three principal stages, each addressing a specific set of visual artifacts. The mathematical and algorithmic formalization of each stage is detailed below.

Stage 1: Edge-Preserving Denoising and Luminance Decomposition

Low-light aerial images are inherently corrupted by high-gain sensor noise. The initial stage mitigates this by applying an edge-preserving bilateral filter, which smooths homogenous regions while maintaining the integrity of critical structural features. Following denoising, the image is transformed from the BGR color space to the YCrCb color space. This decomposition is crucial as it isolates the luminance component (Y channel) from the chrominance components (Cr, Cb), allowing for targeted luminance adjustments without inducing spurious color shifts.

Let the input image be denoted as $I_{in} \in \mathbb{R}^{H \times W \times 3}$. The initial processing is defined as:

$$I_{denoised} = \mathcal{F}_{Bilateral}(I_{in}, d, \sigma_{color}, \sigma_{space}), \tag{13}$$

where $\mathcal{F}_{Bilateral}$ is the bilateral filtering operator, with d as the pixel neighborhood diameter, and σ_{color} and σ_{space} as the filter's standard deviations in the color and coordinate spaces, respectively. The subsequent transformation yields the luminance and chrominance channels:

$$(Y, C_r, C_b) = \mathcal{T}_{BGR \to YCrCb}(I_{denoised}), \tag{14}$$

where $\mathcal{T}$ represents the color space transformation.

Stage 2: Adaptive Visibility and Contrast Enhancement

The core of the APA module is its adaptive, non-linear enhancement of the luminance channel. A standard gamma correction uses a fixed exponent, which is suboptimal for images with varying levels of darkness. We introduce an adaptive gamma, γ_{adapt}, whose value is dynamically computed based on the image's mean luminance. This logarithmic formulation provides a substantial boost to severely underexposed regions while applying a more conservative correction to moderately lit areas, thereby preventing over-enhancement.

The mean luminance, normalized to $[0, 1]$, is first calculated:

$$\bar{\mu}_Y = \frac{1}{255 \cdot HW} \sum_{i=1}^{H} \sum_{j=1}^{W} Y(i, j). \tag{15}$$

The adaptive gamma is then derived as:

$$\gamma_{adapt} = \text{clip}\left(\gamma_{base} - \lambda \ln(\bar{\mu}_Y + \epsilon), \gamma_{min}, \gamma_{max}\right), \tag{16}$$

where γ_{base} is the base gamma value, λ is a sensitivity parameter (0.5 in our implementation), and ϵ is a small constant for numerical stability. The result is clipped to a predefined range $[\gamma_{min}, \gamma_{max}]$ to ensure robustness.

The gamma correction is then applied to the luminance channel:

$$Y_{gamma} = 255 \cdot (Y/255)^{1/\gamma_{adapt}}. \tag{17}$$

Following this initial brightness lift, Contrast Limited Adaptive Histogram Equalization (CLAHE) is employed. Its purpose here is not global contrast stretching, but rather the enhancement of local texture and detail within large, uniform dark regions (e.g., unlit fields or rooftops), which are prevalent in aerial perspectives.

$$Y_{enh} = \mathcal{F}_{CLAHE}(Y_{gamma}). \tag{18}$$

The enhanced luminance channel Y_{enh} is then merged back with the original chrominance channels C_r and C_b.

Stage 3: Perceptual Color and Highlight Correction

The final stage refines the image's perceptual quality by addressing color casts and potential highlight clipping. Urban night scenes often exhibit a reddish or greenish tint from heterogeneous artificial lighting. To counteract this, we transition to the CIEL*a*b* color space, where the a* channel directly corresponds to the green-red axis. A linear scaling is applied to this channel to neutralize color imbalances by amplifying red tones while suppressing green ones.

First, the BGR image is converted to the CIEL*a*b* space:

$$(L^*, a^*, b^*) = \mathcal{T}_{BGR \to Lab}(\mathcal{T}_{YCrCb \to BGR}(Y_{enh}, C_r, C_b)). \tag{19}$$

The a* channel is then adjusted using a red-boost factor α_{red}:

$$a^*_{adj} = \text{clip}\left((a^* - 128.0) \cdot \alpha_{red} + 128.0, 0, 255\right). \tag{20}$$

This operation effectively expands the positive range (reds) and contracts the negative range (greens) relative to the neutral gray point (128).

After converting back to BGR, a final tuning step is performed in the HSV (Hue, Saturation, Value) space. This serves two critical functions: first, a modest saturation boost (β_{sat}) compensates for color desaturation common in low-light imagery. Second, and more importantly, a highlight suppression factor (η_{supp}) is applied to the Value (V) channel. This acts as a protective measure to prevent bright light sources (e.g., streetlights) from becoming clipped or over-exposed—a critical failure mode observed in preliminary experiments.

$$I_{tinted} = \mathcal{T}_{Lab \to BGR}(L^*, a^*_{adj}, b^*), \tag{21}$$

$$(H, S, V) = \mathcal{T}_{BGR \to HSV}(I_{tinted}), \tag{22}$$

$$S_{final} = \text{clip}(S \cdot \beta_{sat}, 0, 255), \tag{23}$$

$$V_{final} = \text{clip}(V \cdot \eta_{supp}, 0, 255). \tag{24}$$

The final output image I_{out} is obtained by merging the modified HSV channels and converting back to the RGB color space.

Algorithm 2: APA: Adaptive Pre-enhancement Augmentation Pipeline

Require: Input RGB image I_{in}, parameters $\Theta = \{d, \sigma_{color}, \sigma_{space}, \gamma_{base}, \lambda, \alpha_{red}, \beta_{sat}, \eta_{supp}\}$
Ensure: Enhanced RGB image I_{out}

$I_{bgr} \leftarrow \mathcal{T}_{RGB \to BGR}(I_{in})$
$I_{denoised} \leftarrow \mathcal{F}_{Bilateral}(I_{bgr}, d, \sigma_{color}, \sigma_{space})$ ▷ Stage 1: Denoising
$(Y, C_r, C_b) \leftarrow \mathcal{T}_{BGR \to YCrCb}(I_{denoised})$ ▷ Luminance-Chrominance separation
$\bar{\mu}_Y \leftarrow \text{mean}(Y)/255.0$ ▷ Stage 2: Adaptive Enhancement
$\gamma_{adapt} \leftarrow \text{clip}(\gamma_{base} - \lambda \ln(\bar{\mu}_Y + \epsilon), \gamma_{min}, \gamma_{max})$
$Y_{gamma} \leftarrow 255 \cdot (Y/255.0)^{1/\gamma_{adapt}}$
$Y_{enh} \leftarrow \mathcal{F}_{CLAHE}(Y_{gamma})$
$I_{enh\_bgr} \leftarrow \mathcal{T}_{YCrCb \to BGR}(Y_{enh}, C_r, C_b)$
$(L^*, a^*, b^*) \leftarrow \mathcal{T}_{BGR \to Lab}(I_{enh\_bgr})$ ▷ Stage 3: Perceptual Correction
$a^*_{adj} \leftarrow \text{clip}((a^* - 128.0) \cdot \alpha_{red} + 128.0, 0, 255)$
$I_{tinted} \leftarrow \mathcal{T}_{Lab \to BGR}(L^*, a^*_{adj}, b^*)$
$(H, S, V) \leftarrow \mathcal{T}_{BGR \to HSV}(I_{tinted})$
$S_{final} \leftarrow \text{clip}(S \cdot \beta_{sat}, 0, 255)$ ▷ Boost saturation
$V_{final} \leftarrow \text{clip}(V \cdot \eta_{supp}, 0, 255)$ ▷ Suppress highlights
$I_{final\_bgr} \leftarrow \mathcal{T}_{HSV \to BGR}(H, S_{final}, V_{final})$
$I_{out} \leftarrow \mathcal{T}_{BGR \to RGB}(I_{final\_bgr})$
return I_{out}

Algorithmic Summary

The entire APA pipeline is summarized in Algorithm 2. By systematically addressing noise, luminance, local contrast, color casts, and highlights in their respective optimal domains, the APA module transforms challenging raw inputs into a perceptually consistent and more tractable domain for the network to learn from.

Limitations and Failure Cases

While U3LIE demonstrates robust performance across a wide range of scenes in the U3D dataset, it has limitations. As shown in Figure 9, in scenarios with extreme darkness and virtually no light sources, the framework may struggle to recover meaningful information, sometimes leading to a flat, low-contrast output. This represents a fundamental limitation of any single-image enhancement method when the input signal-to-noise ratio (SNR) falls below a critical threshold, as there is insufficient information to distinguish detail from noise. Future work could explore overcoming this physical boundary by leveraging temporal information from video streams or multi-modal data fusion (e.g., with thermal imagery).

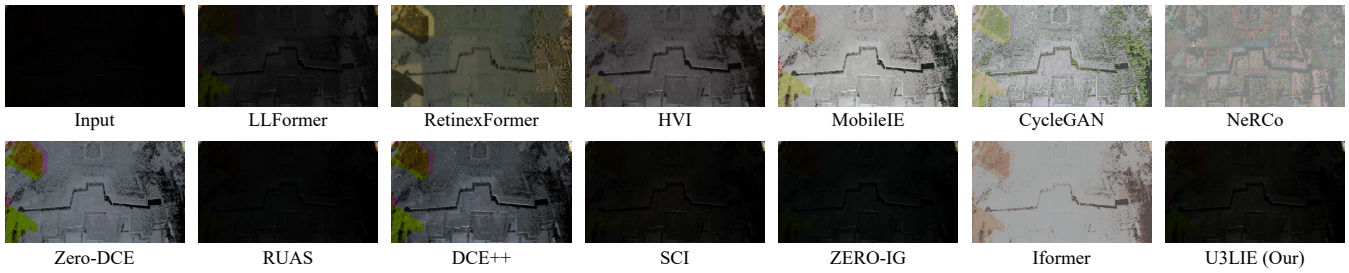


Figure 9: Illustration of a failure case for U3LIE under extreme low-light conditions.

Additional Visualizations

This section presents additional visual comparisons to highlight the effectiveness of our approach. Figure 10 showcases U3LIE's superior capability in handling challenging aerial scenes with non-uniform lighting. It adeptly restores visibility in dark regions while avoiding color artifacts and highlight clipping in bright areas. The insets (red boxes) confirm that our method generates results that are visually plausible and detailed, maintaining the scene's original structure.

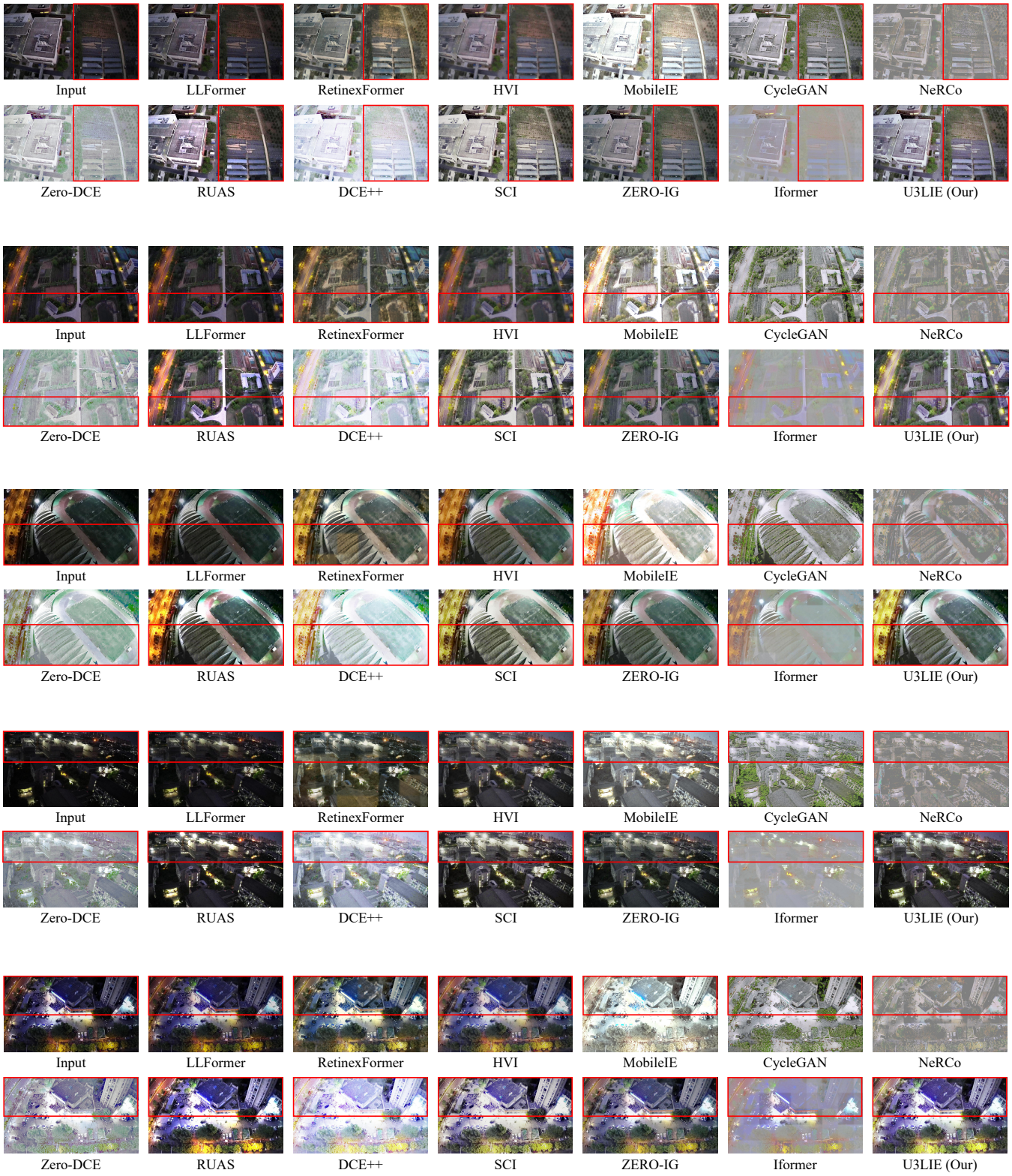


Figure 10: Visual comparison of our U3LIE with SOTA methods on the U3D dataset.

References

Blau, Y.; and Michaeli, T. 2018. The Perception-Distortion Tradeoff. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 6228–6237.

Bozcan, I.; and Kayacan, E. 2020. AU-AIR: A multi-modal unmanned aerial vehicle dataset for low altitude traffic surveillance. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 8504–8510.

Bychkovsky, V.; Paris, S.; Chan, E.; and Durand, F. 2011. Learning photographic global tonal adjustment with a database of input/output image pairs. In *2011 IEEE Conference on Computer Vision and Pattern Recognition*, 97–104.

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 *Proc. IEEE Conf. Int. Conf. Comput. Vis.*, 12470–12479.

Chen, C.; Chen, Q.; Xu, J.; and Koltun, V. 2018. Learning to See in the Dark. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 3291–3300.

Du, D.; Qi, Y.; Yu, H.; Yang, Y.; Duan, K.; Li, G.; Zhang, W.; Huang, Q.; and Tian, Q. 2018. The unmanned aerial vehicle benchmark: Object detection and tracking. In *Proceedings of the European conference on computer vision (ECCV)*, 370–386.

Du, D.; Zhu, P.; Wen, L.; Bian, X.; Lin, H.; Hu, Q.; Peng, T.; Zheng, J.; Wang, X.; Zhang, Y.; Bo, L.; Shi, H.; Zhu, R.; Kumar, A.; Li, A.; Zinollayev, A.; Askergaliyev, A.; Schumann, A.; Mao, B.; Lee, B.; Liu, C.; Chen, C.; Pan, C.; Huo, C.; Yu, D.; Cong, D.; Zeng, D.; Pailla, D. R.; Li, D.; Wang, D.; Cho, D.; Zhang, D.; Bai, F.; Jose, G.; Gao, G.; Liu, G.; Xiong, H.; Qi, H.; Wang, H.; Qiu, H.; Li, H.; Lu, H.; Kim, I.; Kim, J.; Shen, J.; Lee, J.; Ge, J.; Xu, J.; Zhou, J.; Meier, J.; Choi, J. W.; Hu, J.; Zhang, J.; Huang, J.; Huang, K.; Wang, K.; Sommer, L.; Jin, L.; Zhang, L.; Huang, L.; Sun, L.; Steinmann, L.; Jia, M.; Xu, N.; Zhang, P.; Chen, Q.; Lv, Q.; Liu, Q.; Cheng, Q.; Chennamsetty, S. S.; Chen, S.; Wei, S.; Kruthiventi, S. S. S.; Hong, S.; Kang, S.; Wu, T.; Feng, T.; Kollerathu, V. A.; Li, W.; Dai, W.; Qin, W.; Wang, W.; Wang, X.; Chen, X.; Chen, X.; Sun, X.; Zhang, X.; Zhao, X.; Zhang, X.; Zhang, X.; Chen, X.; Wei, X.; Zhang, X.; Li, Y.; Chen, Y.; Toh, Y. H.; Zhang, Y.; Zhu, Y.; Zhong, Y.; Wang, Z.; Wang, Z.; Song, Z.; and Liu, Z. 2019. VisDrone-DET2019: The Vision Meets Drone Object Detection in Image Challenge Results. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*.

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 *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 1777–1786.

Guo, X.; Li, Y.; and Ling, H. 2017. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2): 982–993.

Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; and Wang, Z. 2021. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. *IEEE Trans. Image Process.*, 30: 2340–2349.

Lee, C.-H.; Lee, C.; and Kim, C.-S. 2013. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE Transactions on Image Processing*, 22(12): 5372–5384.

Li, C.; Guo, C.; and Loy, C. C. 2022. Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(8): 4225–4238.

Li, H.; Hu, X.; and Wang, H. 2025. Interpretable Unsupervised Joint Denoising and Enhancement for Real-World low-light Scenarios. *Proc. Int. Conf. Learn. Rep.*

Liang, Z.; Li, C.; Zhou, S.; Feng, R.; and Loy, C. C. 2023. Iterative Prompt Learning for Unsupervised Backlit Image Enhancement. In *Proc. IEEE Conf. Int. Conf. Comput. Vis.*, 8060–8069.

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 *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 10556–10565.

Loh, Y. P.; and Chan, C. S. 2019. Getting to know low-light images with the Exclusively Dark dataset. *Computer Vision and Image Understanding*, 178: 30–42.

Lv, F.; Lu, F.; Wu, J.; and Lim, C. 2018. MBLLEN: Low-Light Image/Video Enhancement Using CNNs. In *Proc. Brit. Mach. Vis. Conf.*, 220.

Ma, L.; Ma, T.; Liu, R.; Fan, X.; and Luo, Z. 2022. Toward Fast, Flexible, and Robust Low-Light Image Enhancement. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 5627–5636.

Mittal, A.; Moorthy, A. K.; and Bovik, A. C. 2012. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process.*, 21(12): 4695–4708.

Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Processing Letters*, 20(3): 209–212.

Park, T.; Efros, A. A.; Zhang, R.; and Zhu, J.-Y. 2020. Contrastive Learning for Unpaired Image-to-Image Translation. In *Proc. Eur. Conf. Comput. Vis.*, 319–345.

Pizer, S. M.; Amburn, E. P.; Austin, J. D.; Cromartie, R.; Geselowitz, A.; Greer, T.; Ter Haar Romeny, B.; Zimmerman, J. B.; and Zuiderveld, K. 1987. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3): 355–368.

Sakaridis, C.; Dai, D.; and Van Gool, L. 2019. Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7374–7383.

Sandler, M.; Howard, A. G.; Zhu, M.; Zhmoginov, A.; and Chen, L.-C. 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 4510–4520.

Shi, Y.; Liu, D.; Zhang, L.; Tian, Y.; Xia, X.; and Fu, X. 2024. ZERO-IG: Zero-Shot Illumination-Guided Joint Denoising and Adaptive Enhancement for Low-Light Images. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 3015–3024.

Tan, X.; Xu, K.; Cao, Y.; Zhang, Y.; Ma, L.; and Lau, R. W. H. 2021. Night-Time Scene Parsing With a Large Real Dataset. *IEEE Transactions on Image Processing*, 30: 9085–9098.

Wang, S.; Zheng, J.; Hu, H.-M.; and Li, B. 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9): 3538–3548.

Wang, T.; Zhang, K.; Shen, T.; Luo, W.; Stenger, B.; and Lu, T. 2023. Ultra-High-Definition Low-Light Image Enhancement: A Benchmark and Transformer-Based Method. In *Proc. AAAI Conf. Artif. Intell.*, 2654–2662.

Wang, W.; Wei, C.; Yang, W.; and Liu, J. 2018. GLADNet: Low-Light Enhancement Network with Global Awareness. In *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 751–755.

Wang, Y.; Wan, R.; Yang, W.; Li, H.; Chau, L.-P.; and Kot, A. C. 2022. Low-Light Image Enhancement with Normalizing Flow. In *Proc. AAAI Conf. Artif. Intell.*, 2604–2612.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.*, 13(4): 600–612.

Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep Retinex Decomposition for Low-Light Enhancement. In *Proc. Brit. Mach. Vis. Conf.*, 155.

Wu, W.; Weng, J.; Zhang, P.; Wang, X.; Yang, W.; and Jiang, J. 2022. URetinex-Net: Retinex-based Deep Unfolding Network for Low-light Image Enhancement. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 5891–5900.

Yan, H.; Li, A.; Zhang, X.; Liu, Z.; Shi, Z.; Zhu, C.; and Zhang, L. 2025a. MobileIE: An Extremely Lightweight and Effective ConvNet for Real-Time Image Enhancement on Mobile Devices. *Proc. IEEE Conf. Int. Conf. Comput. Vis.*

Yan, Q.; Feng, Y.; Zhang, C.; Pang, G.; Shi, K.; Wu, P.; Dong, W.; Sun, J.; and Zhang, Y. 2025b. HVI: A New Color Space for Low-light Image Enhancement. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, volume abs/2502.20272, 5678–5687.

Yang, S.; Ding, M.; Wu, Y.; Li, Z.; and Zhang, J. 2023. Implicit Neural Representation for Cooperative Low-light Image Enhancement. In *Proc. IEEE Conf. Int. Conf. Comput. Vis.*, 12872–12881.

Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; and Liu, J. 2020. From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 3060–3069.

Yu, F.; Chen, H.; Wang, X.; Xian, W.; Chen, Y.; Liu, F.; Madhavan, V.; and Darrell, T. 2020. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2636–2645.

Zhang, Y.; Zhang, J.; and Guo, X. 2019. Kindling the Darkness - A Practical Low-light Image Enhancer. In *Proc. ACM Multimedia*, 1632–1640.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *Proc. IEEE Conf. Int. Conf. Comput. Vis.*, 2242–2251.

Zhu, L.; Yang, W.; Chen, B.; Zhu, H.; Meng, X.; and Wang, S. 2024a. Temporally Consistent Enhancement of Low-Light Videos via Spatial-Temporal Compatible Learning. *Int. J. Comput. Vis.*, 1–21.

Zhu, L.; Yang, W.; Chen, B.; Zhu, H.; Ni, Z.; Mao, Q.; and Wang, S. 2024b. Unrolled Decomposed Unpaired Learning for Controllable Low-Light Video Enhancement. In *Proc. Eur. Conf. Comput. Vis.*, 329–347.