

Enhancing Nighttime UAV Tracking with Light Distribution Suppression

Liangliang Yao¹, Changhong Fu^{1,*}, Yiheng Wang¹, Haobo Zuo², Kunhan Lu¹

Abstract—Visual object tracking has boosted extensive intelligent applications for unmanned aerial vehicles (UAVs). However, the state-of-the-art (SOTA) enhancers for nighttime UAV tracking always neglect the uneven light distribution in low-light images, inevitably leading to excessive enhancement in scenarios with complex illumination. To address these issues, this work proposes a novel enhancer, *i.e.*, LDEnhancer, enhancing nighttime UAV tracking with light distribution suppression. Specifically, a novel image content refinement module is developed to decompose the light distribution information and image content information in the feature space, allowing for the targeted enhancement of the image content information. Then this work designs a new light distribution generation module to capture light distribution effectively. The features with light distribution information and image content information are fed into the different parameter estimation modules, respectively, for the parameter map prediction. Finally, leveraging two parameter maps, an innovative interweave iteration adjustment is proposed for the collaborative pixel-wise adjustment of low-light images. Additionally, a challenging nighttime UAV tracking dataset with uneven light distribution, namely NAT2024-2, is constructed to provide a comprehensive evaluation, which contains 40 challenging sequences with over 74K frames in total. Experimental results on the authoritative UAV benchmarks and the proposed NAT2024-2 demonstrate that LDEnhancer outperforms other SOTA low-light enhancers for nighttime UAV tracking. Furthermore, real-world tests on a typical UAV platform with an NVIDIA Orin NX confirm the practicality and efficiency of LDEnhancer. The code is available at <https://github.com/vision4robotics/LDEnhancer>.

I. INTRODUCTION

Visual object tracking has gained increasing attention in various applications of unmanned aerial vehicle (UAV), *e.g.*, search and rescue missions [1], obstacle avoidance [2], aerial cinematography [3], and 3D localization [4]. With the position provided in the initial frame, the UAV tracker aims to predict the target location in the following frames. For nighttime UAV tracking, the performance of the trackers [5]–[10] designed for high-visibility circumstances presents a significant drop due to the limited illumination, low contrast, and much noise in low-light images [11]. This issue is particularly pronounced in scenarios with complex illumination. Consequently, developing robust and effective nighttime UAV tracking remains a challenging problem.

Over the past years, the low-light image enhancer based on iteration enhancement [12]–[14] has been one of the mainstream solutions for nighttime UAV tracking. These methods

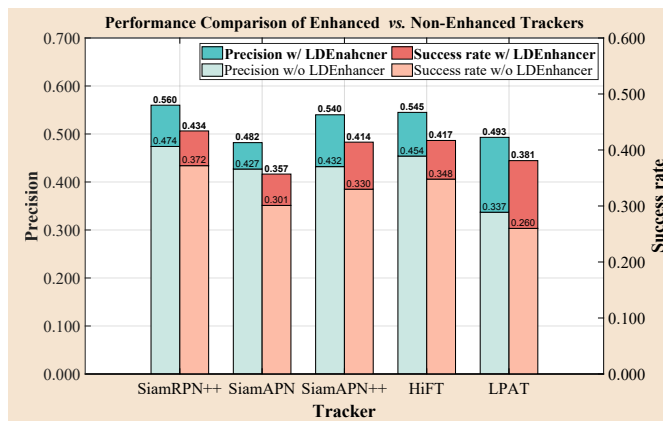


Fig. 1. Overall performance of the state-of-the-art (SOTA) trackers [5]–[8], [15] with the proposed LDEnhancer enabled or not on UAVDark135 [11]. With the support of LDEnhancer, SOTA UAV trackers, which exhibit difficulty in nighttime scenarios with uneven light distribution, have been significantly improved. SiamAPN++ [6] with LDEnhancer gains the improvement with 25.0% and 25.5% in precision and success rate, respectively. Meanwhile, LPAT [7] gets the promotion with 46.3% and 46.5% in precision and success rate, respectively.

mainly estimate an enhancement parameter map from the low-light image and adjust low-light images iteratively based on the enhancement map. By boosting the intensity of dark regions, the performance of the UAV trackers trained with daytime images has gained substantial improvement. However, when applied to low-light images captured in scenarios with complex illumination, these methods neglect the uneven light distribution and inevitably further enhance the regions that are already sufficiently bright. The over-enhancement and saturation after low-light image enhancement (LLIE) may destroy the extracted image features, leading to a reduction in image content information which is significant for nighttime UAV tracking. **Hence, how to devise a low-light enhancer capable of addressing uneven light distribution is a pressing need for nighttime UAV tracking.**

Recently, light distribution suppression [16]–[19] has garnered widespread attention in image processing. The majority of these works [16], [17] are towards light distribution suppression on haze or foggy nights, posing a challenge to directly utilize in real-world nighttime scenarios. Moreover, they are unable to effectively enhance dark regions in low-light images. Only several approaches [19] have the capability of suppressing bright regions and enhancing dark regions simultaneously. Despite the works generating some visually pleasing results in LLIE, the efficacy for enhancing nighttime UAV tracking remains notably unsatisfactory due to the misalignment of the optimization objectives with nighttime UAV tracking. **Thereby, how to effectively combine a low-light**

*Corresponding author

¹L. Yao, C. Fu, Y. Wang, and K. Lu are with the School of Mechanical Engineering, Tongji University, Shanghai 201804, China. Email: changhongfu@tongji.edu.cn

²H. Zuo is with the Department of Computer Science, University of Hong Kong, Hong Kong 999077, China.

enhancer with light distribution suppression for nighttime UAV tracking is worth exploring carefully.

In this work, a novel enhancer, *i.e.*, LDEnhancer, is proposed to enhance nighttime UAV tracking with light distribution suppression. Figure 1 shows the tracking performance of the state-of-the-art (SOTA) trackers with the proposed LDEnhancer enabled or not on UAVDark135 [11]. With LDEnhancer, the performance of UAV trackers has been greatly improved in low-light conditions. The main contributions are as follows:

- A novel low-light image enhancer, dubbed LDEnhancer, is proposed to enhance nighttime UAV tracking with light distribution suppression, where the task is treated as a collaborative pixel-wise adjustment with suppression parameter map and enhancement parameter map.
- A novel image content refinement module is designed to decompose the light distribution information and image content information in the feature space, allowing for the targeted enhancement of the image content information. A new light distribution generation module is developed to capture light distribution effectively.
- Two different parameter estimation modules are constructed to process the features with light distribution information and image content information, respectively, for the parameter map prediction.
- Comprehensive evaluation on authoritative nighttime UAV tracking benchmarks and the proposed NAT2024-2 verify the effectiveness and robustness of LDEnhancer for nighttime UAV tracking. Real-world tests demonstrate that LDEnhancer has superior advantages for the deployment of nighttime UAV tracking, especially for scenarios with complex illumination.

II. RELATED WORK

A. Low-Light Image Enhancement (LLIE)

The primary objective of LLIE [20]–[28] is to enhance the perception or interpretability of images captured in low-light environments. C. Li *et al.* [21] estimate an enhancement parameter map from the input low-light image, which is then used for image enhancement in an iteration manner. KinD++ [23], inspired by the Retinex theory, decomposes the images into two parts: illumination map and reflection map. Then this method makes adjustments to the illumination map and removes degradation on the reflection map. R. Liu *et al.* [24] combine Retinex theory and unrolled optimization, using a cooperative prior architecture search algorithm. For nighttime UAV tracking, several approaches [12]–[14] have been proposed to combine LLIE with UAV trackers for enhancing performance. Darklighter [12] designs a lightweight convolutional neural network (CNN)-based enhancement parameter map estimation network to cope with poor illumination and intricate noise. HighlightNet [13] proposes an iteration-based low-light enhancer to facilitate both the online target selection step and tracking step for nighttime UAV tracking. J. Ye *et al.* [14] construct a low-light enhancer with a spatial-channel Transformer to achieve

favorable LLIE iteratively. However, the SOTA enhancers tend to significantly overenhance bright regions in captured images under conditions of complex illumination, resulting in a suboptimal performance for nighttime UAV tracking.

B. Light Distribution Suppression

Light distribution suppression [16]–[19], a challenging problem to be resolved in image processing, has caught worldwide research attention. In night image dehazing, a few methods have been proposed to suppress light distribution. Y. Li *et al.* [16] model the glow component by convolving the light distribution with the atmosphere point spread function, and decomposing it from the image. J. Zhang *et al.* [29] refine the hazy imaging model with uneven light distribution, which is estimated by maximum reflectance prior. W. Yan *et al.* [17] defog the high and low frequency components of grayscale images and guide light distribution suppression of colored images with defogged grayscale images. A. Sharma *et al.* [18] propose a novel approach grounded in camera response function estimation and high dynamic range imaging to mitigate light distribution. For LLIE with light distribution suppression, Y. Jin *et al.* [19] propose an unsupervised night image enhancement method that combines layer decomposition and light-effects suppression to achieve better visibility, color fidelity, and naturalness in low-light conditions without manual annotations or prior knowledge. However, most works in night image dehazing are only designed for light distribution suppression, without considering enhancing dark regions simultaneously. Moreover, the misalignment of the optimization objectives between LLIE and nighttime UAV tracking poses a limitation to enhancing the tracking performance in low-light scenarios.

III. PROPOSED METHOD

In this section, the proposed enhancer, LDEnhancer, is introduced in detail. As depicted in Fig. 2, the proposed framework can be partitioned into three parts, *i.e.*, *feature extraction*, *light distribution-aware parameter inference*, and *interweave iteration adjustment*.

A. Feature Extraction

During feature extraction, this work constructs a CNN-based feature extractor to extract robust features from input images. TABLE I shows the detailed network settings of the feature extractor. As an illustrative example, an input with data dimensions of $256 \times 256 \times 3$ is considered. The network architecture comprises four lightweight CNR

TABLE I
DETAILS OF THE FEATURE EXTRACTOR.

Input	Data dimensions	Operator	Kernel	Stride	Padding	Output
I_0	$256 \times 256 \times 3$	CNR	2×2	2	0	C_1
C_1	$128 \times 128 \times 4$	CNR	2×2	2	0	C_2
C_2	$64 \times 64 \times 4$	CNR	2×2	2	0	C_3
C_3	$32 \times 32 \times 8$	CNR	2×2	2	0	F_0

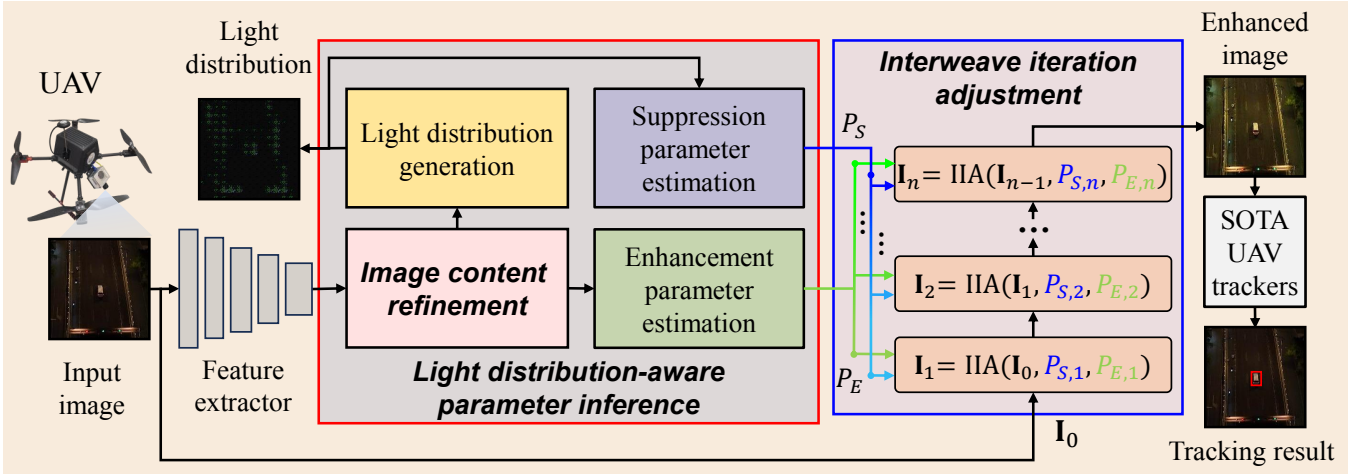


Fig. 2. Overview of our proposed LDEnhancer. The LDEnhancer is composed of three parts, from left to right: *feature extraction*, *light distribution-aware parameter inference*, and *interweave iteration enhancement*. P_S denotes the suppression parameter map. P_E represents the enhancement parameter map.

layers, where CNR denotes the sequential operations of a convolutional layer, batch normalization, and a rectified linear unit (ReLU) activation function. To extract high-level features while reducing the size of the feature maps, the feature extractor employs a convolutional layer with a 2×2 kernel and a stride size of 2 as a down-sampling operation.

B. Light Distribution-Aware Parameter Inference

In light distribution-aware parameter inference, this work designs a novel image content refinement module to process the feature generated from the feature extractor. Then the decomposed features are posted to a new light distribution generation module. The light distribution generation module captures light distribution effectively with a proposed unsupervised light loss. The resulting features containing image content information and light distribution information are then used in separate estimation modules for suppression and enhancement parameter map prediction.

Image content refinement. As shown in Fig. 3, the image content refinement module takes the features F_0 generated from the feature extractor as the input. The features F_0 are fed into the proposed decomposition layer Decomposition for generating light distribution features F_1 . Specifically, the decomposition layer is a 1×1 convolution layer. Then subtract light distribution features F_1 from the input features F_0 element by element. The results are denoted as F_2 . The whole process is defined as:

$$\begin{aligned} F_1 &= \text{Decomposition}(F_0) \\ F_2 &= F_0 - F_1 \end{aligned} \quad (1)$$

The cross-attention CA uses the feature F_2 extracted by the feature extraction network as Q (Query), while the estimated light distribution F_1 as K (Key) and V (Value). To clarify, the position encoding is denoted as P . The results of the cross-attention are fed into the deconvolution layers DeConv to adjust the dimension. The output O_1 of the image

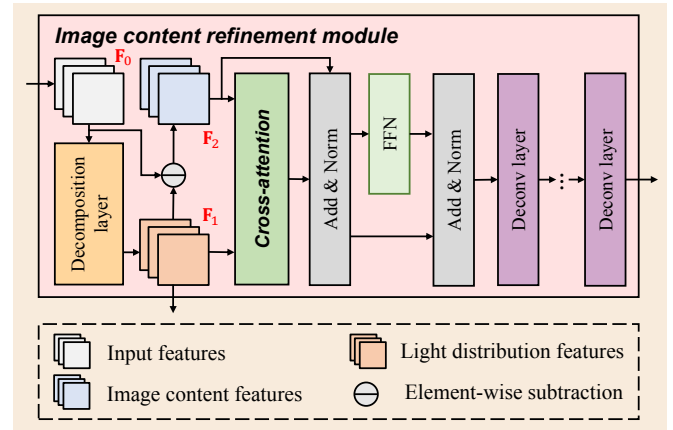


Fig. 3. Detailed workflow of the image content refinement module.

content refinement module is obtained by:

$$\begin{aligned} M_x &= \text{CA}(F_2 + P, F_1 + P, F_1) \\ M_y &= \text{Norm}(M_x + F_2) \\ O_1 &= \text{DeConv}(\text{Norm}(\text{FFN}(M_y) + M_y)) \end{aligned} \quad (2)$$

where M_x and M_y are the intermediate variable. CA represents the procedure of the multi-head attention [30]. Norm is the normalization operation. Besides, FFN stands for feed-forward network. Generally, the cross-attention CA can be calculated as follows:

$$\begin{aligned} \text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{c}}\right)\mathbf{V} \\ \text{CA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= (\text{Cat}(a^1, a^2, \dots, a^n))\mathbf{W}_c \\ a^j &= \text{Att}(\mathbf{Q}\mathbf{W}_1^j, \mathbf{K}\mathbf{W}_2^j, \mathbf{V}\mathbf{W}_3^j) \end{aligned} \quad (3)$$

where Att represents the scaled dot-product attention and \sqrt{c} is the scaling factor to avoid gradient vanishment in the softmax function. $\mathbf{W}_c \in \mathbb{R}^{C \times C}$, $\mathbf{W}_1^j \in \mathbb{R}^{C \times C_M}$, $\mathbf{W}_2^j \in \mathbb{R}^{C \times C_M}$, $\mathbf{W}_3^j \in \mathbb{R}^{C \times C_M}$ can all be regarded as fully connected layer operation, where $C_M = C/M$, and M is the number of parallel attention head.

Remark 1: Utilizing 1×1 convolution facilitates the efficient mapping of extracted features to a new feature space, thereby improving the subsequent adaptive generation of light distribution information. Under the guidance of the estimated light distribution features, the image content refinement module is capable of adaptively refining the features containing image content information by attention mechanism.

Light distribution generation. The light distribution generation module is utilized to capture light distribution based on the illumination-prior knowledge. TABLE II shows the details of the light distribution generation module. The module is composed of four lightweight DCNR layers. Specifically, DCNR represents the combination operations of the deconvolution layer, batch normalization, and ReLU.

To address the issue of ambiguity output by the light distribution generation module, it is crucial to provide an appropriate initial estimation. To this end, a light distribution label denoted as \mathbf{I}_l is calculated using illumination-prior knowledge with the relative smoothness method [31]. Since light distribution is usually smooth variations, this method extracts the light distribution from the input image by applying a second-order Laplacian filter. The unsupervised light distribution loss \mathcal{L}_l to constrain the output \mathbf{O}_2 of the light distribution generation module is defined as:

$$\mathcal{L}_l = \text{SmoothL1}(\mathbf{O}_2, \mathbf{I}_l) \quad . \quad (4)$$

Remark 2: The feature extractor and light distribution generation module are combined to form a U-Net-like network architecture, serving as the encoder and decoder, respectively. The feature extractor encodes the input image to extract robust features, while the light distribution generation module reconstructs these features to capture the light distribution.

Parameter estimation. The parameter estimation modules are constructed to predict the parameter maps. In this work, the suppression parameter estimation module and enhancement parameter estimation module share a common structure. As shown in TABLE III, the parameter estimation modules consist of five layers. In the first four layers, each

TABLE II
DETAILS OF THE LIGHT DISTRIBUTION GENERATION MODULE.

Input	Data dimensions	Operator	Kernel	Stride	Padding	Output
\mathbf{F}_1	$16 \times 16 \times 8$	DCNR	2×2	2	0	\mathbf{D}_1
\mathbf{D}_1	$32 \times 32 \times 8$	DCNR	2×2	2	0	\mathbf{D}_2
\mathbf{D}_2	$64 \times 64 \times 4$	DCNR	2×2	2	0	\mathbf{D}_3
\mathbf{D}_3	$128 \times 128 \times 4$	DCNR	2×2	2	0	\mathbf{O}_2

TABLE III
DETAILS OF THE PARAMETER ESTIMATION MODULES.

Input	Data dimensions	Operator	Kernel	Stride	Padding	Output
\mathbf{O}_2	$256 \times 256 \times 3$	CR	3×3	1	1	Conv1
Conv1	$256 \times 256 \times 16$	CR	3×3	1	1	Conv2
Conv2	$256 \times 256 \times 16$	CR	3×3	1	1	Conv3
Conv3&Conv2	$256 \times 256 \times 32$	CR	3×3	1	1	Conv4
Conv4&Conv1	$256 \times 256 \times 32$	CT	3×3	1	1	P_S

CR layer denotes convolution layers with kernel size 3×3 and stride 1, followed by the ReLU. The last CT layer represents convolution layers with kernel size 3×3 and stride 1, followed by the Tanh activation function. To clarify, take the initially captured light distribution \mathbf{O}_2 as an instance, the suppression parameter estimation module generates the suppression parameter map.

C. Interweave Iteration Adjustment

The interweave iteration adjustment (IIA) is proposed for the collaborative pixel-wise adjustment of low-light images, leveraging two parameter maps. To avoid information loss due to overexposure in the bright region, the formula is first defined with two parameter maps:

$$\mathbf{I}_n = \text{IIA}(\mathbf{I}_{n-1}, P_{S,n}, P_{E,n}) \quad . \quad (5)$$

By taking the light distribution into account, the above formula is then revised to:

$$\mathbf{I}_n = \mathbf{I}_{n-1} + (P_{E,n} - P_{S,n})\mathbf{I}_{n-1}(1 - \mathbf{I}_{n-1}) \quad , \quad (6)$$

where \mathbf{I}_n represents the result for iteration n .

Remark 3: In a collaborative way, each iteration process suppresses the bright areas and enhances the darker areas in the image simultaneously. The interweave iteration adjustment facilitates convergence toward the desired enhancement results over successive iterations.

D. Loss Functions

This work adopts five loss functions to train the framework. Specifically, this work adopts the three same loss functions as DCE++ [21], *i.e.*, spatial consistency loss \mathcal{L}_{spa} , color constancy loss \mathcal{L}_{col} , illumination smoothness loss \mathcal{L}_{tv} . For spatial consistency loss, this loss encourages spatial coherence of the enhanced image by preserving the difference of neighboring regions. Considering the impact of light distribution information on image enhancement, this work subtracts light information labels \mathbf{I}_l from the original images, using these labels \mathbf{I}_o in place of the original images. For illumination smoothness loss, the work adds this loss to each parameter map for interweave iteration adjustment, *i.e.*, $P_{E,n} - P_{S,n}$. Meanwhile, this work proposes a novel image exposure control loss \mathcal{L}_{ie} . The loss measures the distance between the average intensity value of a local region to level K . The \mathcal{L}_{ie} can be expressed as:

$$\mathcal{L}_{\text{ie}} = \left(\sum_{i=1}^T \text{SmoothL1}(\alpha E_i, K) \right) / T \quad , \quad (7)$$

where α stands for the coefficient, T represents the number of nonoverlapping local regions of size 16×16 , and E_i is the average intensity value of a local region in the enhanced image. To sum up, the overall loss function of LDEnhancer is presented as:

$$\mathcal{L}_{\text{all}} = \lambda_1 \mathcal{L}_{\text{spa}} + \lambda_2 \mathcal{L}_{\text{col}} + \lambda_3 \mathcal{L}_{\text{tv}} + \lambda_4 \mathcal{L}_{\text{ie}} + \lambda_5 \mathcal{L}_l \quad , \quad (8)$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ are the coefficients of the loss function. In this work, the value of the coefficients adopts 10, 5, 1, 10, 1, respectively.

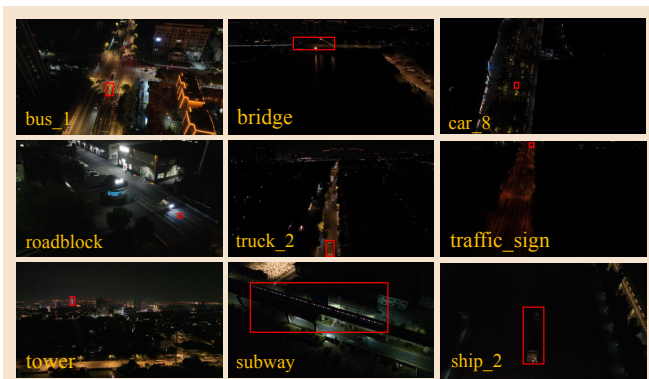


Fig. 4. First frames of sequences from NAT2024-2. The tracking object is marked with red boxes.

Remark 4: Substantial low-light images encompassing uneven light distribution are an obstacle to training an enhancer with fully supervised learning-based methods. Hence, this work proposes an unsupervised loss to train the network.

IV. NAT2024-2 BENCHMARK

For comprehensive evaluation in real nighttime scenarios, a new nighttime UAV tracking benchmark with uneven Light distribution, *i.e.*, NAT2024-2, is constructed. It consists of 40 long image sequences, which own 74K frames in total. Each sequence is over **1500** frames. The sequence categories in the NAT2024-2 dataset encompass diverse targets (*e.g.*, cars, trucks, persons, buses, subways, traffic signs, buildings, and ships) or activities (*e.g.*, cycling, running, and ball games). Employing long-term nighttime UAV tracking benchmarks can better emulate the intricate real-world scenarios, aiding in assessing the performance in complex environments.

V. EXPERIMENTS

In this section, the overall performances of the LDEnhancer in nighttime UAV tracking are presented, demonstrating the effectiveness and superiority of LDEnhancer. Real-world tests are conducted to validate the practical feasibility.

A. Implementation Details

The proposed LDEnhancer is trained for 100 epochs with PyTorch on a single NVIDIA A100 GPU. For training data, the random sampling strategy is utilized to process the real nighttime UAV tracking dataset. The images are sampled from the train set of NAT2021-*train* [32] every 50 frames. The training dataset is denoted as NAT2021-*light*. In the training process, images are first resized to 256×256 . The AdamW optimizer [33] with a fixed learning rate of 0.000001 is adopted to optimize model parameters with a weight decay of 0.0001. The batch size is set to 4.

B. Evaluation Metrics

The proposed enhancer serves for nighttime UAV tracking. Evaluation metrics in visual object tracking are applied to evaluate the enhancer. The three metrics in the one-pass evaluation (OPE) metrics [34] are precision, normalized precision, and success rate. Precision is determined by the

center location error (CLE), and the fraction of frames with a CLE lower than 20 pixels is displayed as the precision plot (PP). The success rate is measured using the intersection over union (IoU), and the percentage of frames with an IoU greater than the configured maximum threshold is shown as the success plot (SP). The area under the curve (AUC) of the success plot is used to rank the success rate of trackers. To account for variations in image resolution and bounding box scale, the normalized precision metric is selected.

C. Effectiveness and Superiority of LDEnhancer for Night-time UAV Tracking

1) *LDEnhancer with Different Trackers:* To prove the effectiveness and superiority of the proposed method, the enhancer is tested on the authoritative night tracking datasets, *i.e.*, UAVDark135, and the proposed NAT2024-2, combined with 5 typical SOTA trackers, including HiFT [8], SiamRPN++ [5], SiamAPN++ [6], SiamAPN [15], and LPAT [7]. The objects in UAVDark135 undergo various light conditions, making the uneven light distribution problem more serious. As illustrated in Fig. 5, the proposed method further strengthens the adaptability of the SOTA UAV trackers in the nighttime. The enhancer has significantly improved the performance of SiamAPN++ for nighttime UAV tracking: **25.0%** on precision, **29.8%** on normalized precision, and **25.5%** on success rate. Moreover, despite the adverse impact of complex nighttime conditions on the performance of LPAT, the implementation of LDEnhancer enables this tracker to achieve superior outcomes.

Evaluations visualization: as shown in Fig. 6, the presented screenshots visually demonstrate the effectiveness of the proposed method. The results indicate that the application of LDEnhancer significantly enhances the performance of the SOTA trackers during low-light conditions.

2) *Comparison with SOTA Low-Light Enhancers:* The comparison with other SOTA enhancement methods is shown in TABLE IV. To present the advantage of LDEnhancer for nighttime UAV tracking, it is compared with other nine SOTA low-light enhancers utilizing the same tracker SiamRPN++ on the UAVDark135 benchmark, including RUAS [24], LIME [25], EnlightenGAN [26], LLVE [27], DCE++ [21], KinD++ [23], DarkLighter [12], HighlightNet [13], and SCT [14]. The results demonstrate the impressive performance of LDEnhancer in nighttime UAV tracking, which is attributed to its light distribution suppression. Conversely, other general low-light enhancers exhibit suboptimal performance. Notably, LDEnhancer raises the SiamRPN++ by **16.7%** and **18.1%** in success rate and precision, respectively, surpassing the second-best performer, SCT.

Remark 5: Although these enhancers, *i.e.*, DarkLighter, HighlightNet, and SCT, are also nighttime UAV tracking-oriented, the inferior performance exposes the weakness of common enhancers in the face of uneven light distribution.

3) *Attribute-Based Performance:* Additional environmental changes caused by illumination variation and low resolution can exacerbate the difficulty of nighttime UAV tracking. To thoroughly assess the effectiveness of LDEnhancer,

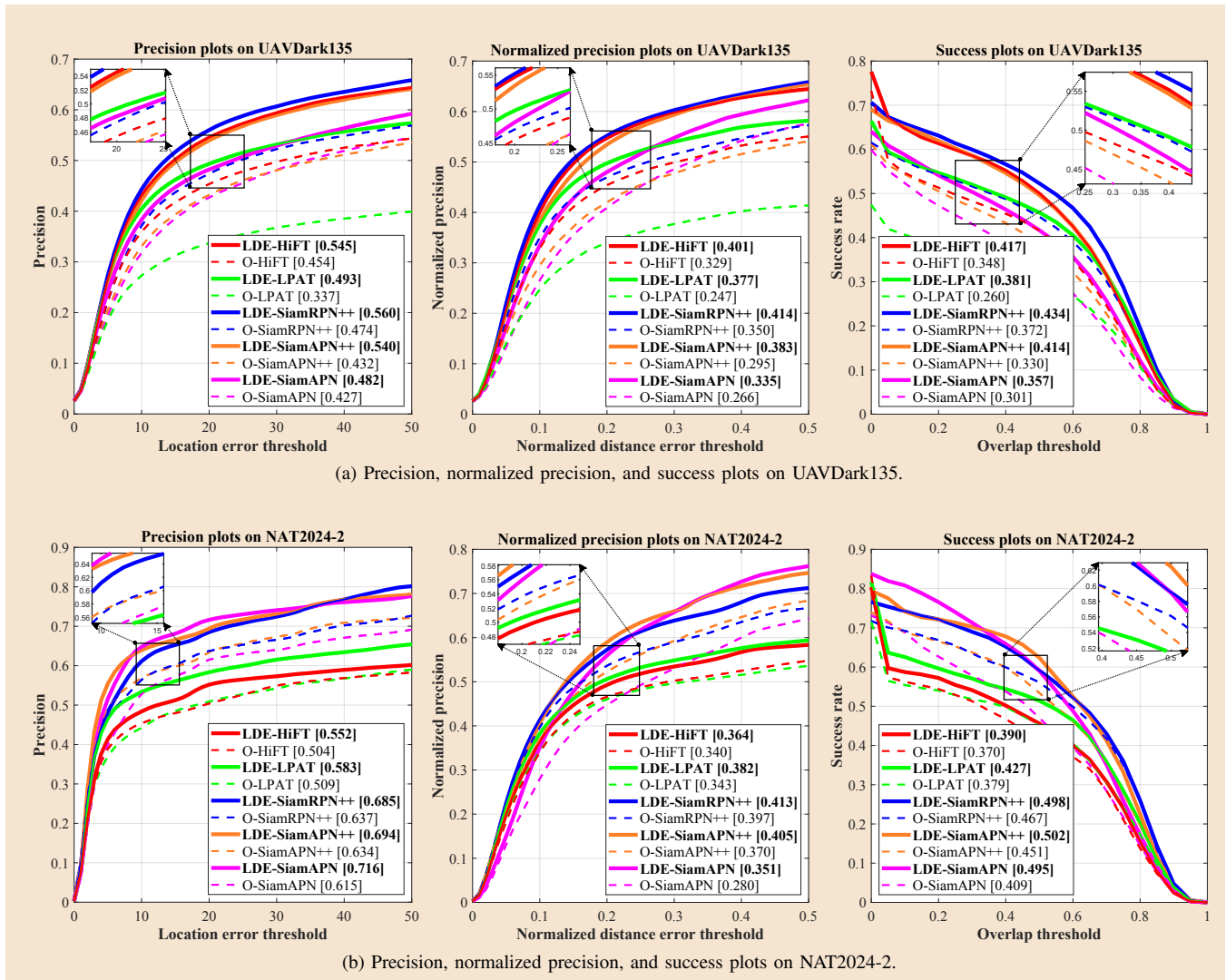


Fig. 5. Overall performance of SOTA UAV trackers with LDEhancer utilized (plots with solid lines) or not (plots with dashed lines). The same color denotes the same tracker. The results demonstrate the effectiveness and superiority of the proposed methods for nighttime UAV tracking.

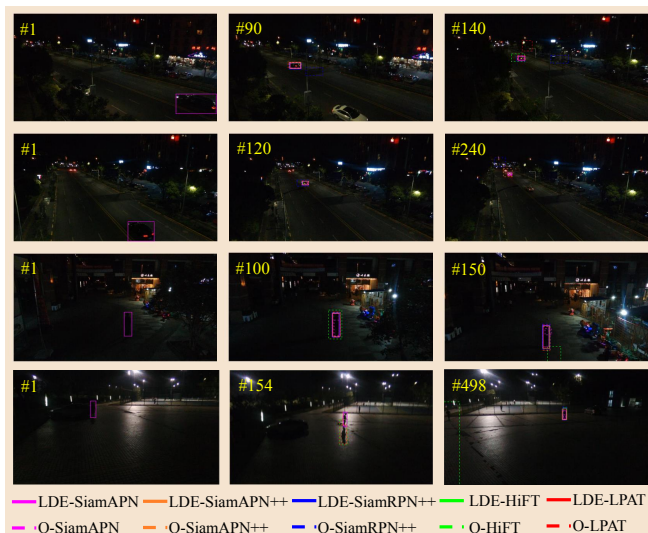


Fig. 6. Visualization of trackers with LDEhancer activated (solid boxes with the prefix ‘‘LDE’’) or not (dashed boxes with the prefix ‘‘O’’). From up to down, the sequences are *car4*, *car6*, *pedestrian5_3*, and *person8* from the UAVDark135 [11]. With LDEhancer, the trackers achieve robust tracking.

a comparison of their pertinent properties is conducted. Figure 7 proves the robustness of our framework in several challenging conditions. SiamAPN++ raises the success rate of the existing best performance by **25.7%** on UAVDark135 for illumination variation. In addition, SiamAPN++ realizes

TABLE IV
COMPARISON OF LDENHANCER WITH SOTA ENHANCERS ON UAVDARK135 [11]. THE FIRST RESULTS ARE IN **BOLD**.

UAVDark135	Succ.	Prec.	Δ_s (%)	Δ_p (%)
SiamRPN++	0.372	0.474	-	-
+RUAS [24]	0.381	0.490	2.419	3.376
+KinD++ [23]	0.386	0.485	3.763	2.321
+LIME [25]	0.393	0.501	5.645	5.696
+EnlightenGAN [26]	0.397	0.505	6.720	6.540
+LLVE [27]	0.397	0.508	6.720	7.173
+DCE++ [21]	0.402	0.512	8.065	8.017
+Darklighter [12]	0.385	0.495	3.495	4.430
+HighLightNet [13]	0.376	0.483	1.075	1.899
+SCT [14]	0.421	0.547	13.172	15.401
+LDEhancer(Ours)	0.434	0.560	16.667	18.143

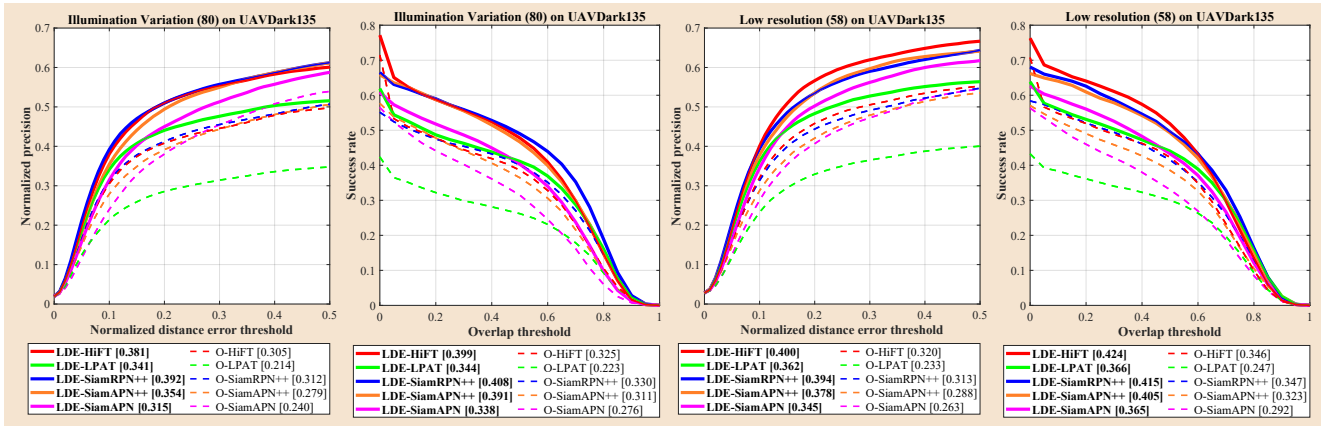


Fig. 7. Normalized precision plots and success plots of illumination variation and low-resolution attributes of SOTA UAV trackers with LDEnhancer utilized (plots with solid lines) or not (plots with dashed lines) on UAVDark135.

a success rate of **0.405** for low resolution on UAVDark135, which improves the existing best performance by **25.4%** for nighttime UAV tracking.

D. Ablation Study

Different variants of LDEnhancer are evaluated with the same UAV tracker, *i.e.*, SiamRPN++, on UAVDark135 in this subsection. Specifically, utilizing one branch structure to estimate parameter maps is denoted as the Baseline. TB stands for utilizing two branches structure to estimate different parameter maps. LS represents employing the unsupervised light distribution loss. TF denotes employing the cross-attention. As shown in TABLE V, with enhancers, the performance of UAV trackers is significantly improved.

E. Real-World Tests

To demonstrate the applicability of LDEnhancer in real-world UAV tracking, extensive real-world tests are adopted on a typical UAV platform equipped with an NVIDIA Orin NX-based edge camera. The SiamRPN++ enabled with LDEnhancer exceeding **30.3** frames per second. As shown in Fig. 8, the main challenges in the tests are illumination variation, low illumination, fast motion, and viewpoint change. In Test 1, the primary challenge is the truck moving in the bright area and the shadow area with a frequent illumination variation. In Test 2, the poor light condition exacerbates the difficulties associated with robust tracking. Test 3 presents viewpoint change and low illumination. The CLE curves show that the prediction errors are within 20 pixels, which can be regarded as reliable nighttime tracking.

TABLE V

ABLATION STUDY OF THE LDENHANCER ON UAVDARK135. Δ REFLECTS THE IMPROVEMENT COMPARED TO THE BASE TRACKER, *i.e.*, SIAMRPN++, WITH THE HIGHEST PERFORMANCE HIGHLIGHTED IN **BOLD**.

Trackers	Succ.	Δ_s (%)	Prec.	Δ_p (%)
SiamRPN++	0.372	-	0.474	-
+Baseline	0.395	+6.183	0.510	+7.595
+Baseline+TB	0.408	+9.677	0.529	+11.603
+Baseline+TB+LS	0.421	+13.172	0.541	+14.135
+Baseline+TB+LS+TF(LDEnhancer)	0.434	+16.667	0.560	+18.143

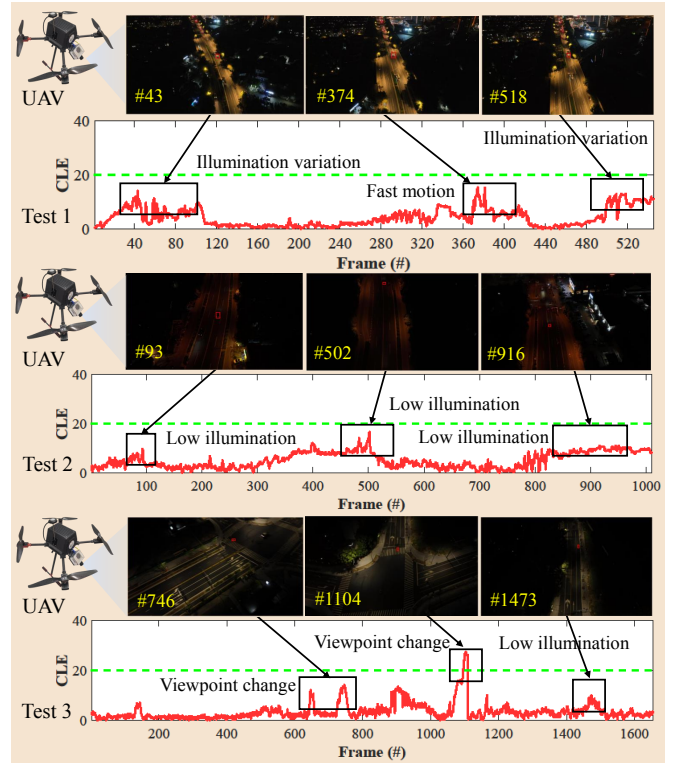


Fig. 8. Several results of real-world tests. The estimated positions are represented by the red bounding boxes. The CLE below the green dashed line is the success tracking in the real-world test. With LDEnhancer, the tracker can effectively realize robust nighttime UAV tracking.

VI. CONCLUSIONS

This work proposes a novel enhancer, LDEnhancer, enhancing nighttime UAV tracking with light distribution suppression. Specifically, a novel image content refinement module and a new light distribution generation module are proposed to capture light distribution information and image content information effectively. Two different parameter estimation modules are constructed to process the features with different information, respectively, for the parameter map prediction. Leveraging two parameter maps, an innovative interweave iteration adjustment is proposed for the collabora-

tive pixel-wise adjustment of low-light images. Experimental results prove the effectiveness as well as the efficiency of our enhancer. We believe this work would contribute to the advancement of nighttime UAV tracking.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 62173249) and the Natural Science Foundation of Shanghai (No. 20ZR1460100).

REFERENCES

- [1] L. A. Varga, B. Kiefer, M. Messmer, and A. Zell, "SeaDronesSee: A Maritime Benchmark for Detecting Humans in Open Water," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022, pp. 2260–2270.
- [2] H. Chen and P. Lu, "Computationally Efficient Obstacle Avoidance Trajectory Planner for UAVs Based On Heuristic Angular Search Method," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5693–5699.
- [3] R. Bonatti, C. Ho, W. Wang, S. Choudhury, and S. Scherer, "Towards a Robust Aerial Cinematography Platform: Localizing and Tracking Moving Targets in Unstructured Environments," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 229–236.
- [4] H. Zhang, G. Wang, Z. Lei, and J.-N. Hwang, "Eye in the Sky: Drone-Based Object Tracking and 3D Localization," in *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, 2019, pp. 899–907.
- [5] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "SiamRPN++: Evolution of Siamese Visual Tracking with very Deep Networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4282–4291.
- [6] Z. Cao, C. Fu, J. Ye, B. Li, and Y. Li, "SiamAPN++: Siamese Attentional Aggregation Network for Real-Time UAV Tracking," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 3086–3092.
- [7] C. Fu, W. Peng, S. Li, J. Ye, and Z. Cao, "Local Perception-Aware Transformer for Aerial Tracking," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 12 122–12 129.
- [8] Z. Cao, C. Fu, J. Ye, B. Li, and Y. Li, "HiFT: Hierarchical Feature Transformer for Aerial Tracking," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 15 457–15 466.
- [9] S. Gao, C. Zhou, C. Ma, X. Wang, and J. Yuan, "AiATrack: Attention in Attention for Transformer Visual Tracking," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022, pp. 146–164.
- [10] L. Yao, C. Fu, S. Li, G. Zheng, and J. Ye, "SGDViT: Saliency-Guided Dynamic Vision Transformer for UAV Tracking," in *Proceedings of the IEEE/RSJ International Conference on Robotics and Automation (ICRA)*, 2023, pp. 3353–3359.
- [11] B. Li, C. Fu, F. Ding, J. Ye, and F. Lin, "All-Day Object Tracking for Unmanned Aerial Vehicle," *IEEE Transactions on Mobile Computing*, vol. 22, no. 8, pp. 1–13, 2022.
- [12] J. Ye, C. Fu, G. Zheng, Z. Cao, and B. Li, "DarkLighter: Light up the Darkness for UAV Tracking," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 3079–3085.
- [13] C. Fu, H. Dong, J. Ye, G. Zheng, S. Li, and J. Zhao, "HighlightNet: Highlighting Low-Light Potential Features for Real-Time UAV Tracking," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 12 146–12 153.
- [14] J. Ye, C. Fu, Z. Cao, S. An, G. Zheng, and B. Li, "Tracker Meets Night: A Transformer Enhancer for UAV Tracking," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3866–3873, 2022.
- [15] C. Fu, Z. Cao, Y. Li, J. Ye, and C. Feng, "Onboard Real-Time Aerial Tracking with Efficient Siamese Anchor Proposal Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.
- [16] Y. Li, R. T. Tan, and M. S. Brown, "Nighttime Haze Removal with Glow and Multiple Light Colors," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2015, pp. 226–234.
- [17] W. Yan, R. T. Tan, and D. Dai, "Nighttime Defogging Using High-Low Frequency Decomposition and Grayscale-Color Networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020, pp. 473–488.
- [18] A. Sharma and R. T. Tan, "Nighttime Visibility Enhancement by Increasing the Dynamic Range and Suppression of Light Effects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 11 977–11 986.
- [19] Y. Jin, W. Yang, and R. T. Tan, "Unsupervised Night Image Enhancement: When Layer Decomposition Meets Light-Effects Suppression," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022, pp. 404–421.
- [20] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1780–1789.
- [21] C. Li, C. Guo, and C. C. Loy, "Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4225–4238, 2021.
- [22] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward Fast, Flexible, and Robust Low-Light Image Enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5637–5646.
- [23] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, "Beyond Brightening Low-Light Images," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1013–1037, 2021.
- [24] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "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 (CVPR)*, 2021, pp. 10 561–10 570.
- [25] X. Guo, Y. Li, and H. Ling, "LIME: Low-Light Image Enhancement via Illumination Map Estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2016.
- [26] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "EnlightenGAN: Deep Light Enhancement Without Paired Supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.
- [27] F. Zhang, Y. Li, S. You, and Y. Fu, "Learning Temporal Consistency for Low Light Video Enhancement from Single Images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 4967–4976.
- [28] X. Xu, R. Wang, and J. Lu, "Low-Light Image Enhancement via Structure Modeling and Guidance," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 9893–9903.
- [29] J. Zhang, Y. Cao, S. Fang, Y. Kang, and C. Wen Chen, "Fast Haze Removal for Nighttime Image Using Maximum Reflectance Prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 7418–7426.
- [30] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.
- [31] Y. Li and M. S. Brown, "Single Image Layer Separation Using Relative Smoothness," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2752–2759.
- [32] J. Ye, C. Fu, G. Zheng, D. P. Paudel, and G. Chen, "Unsupervised Domain Adaptation for Nighttime Aerial Tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 8886–8895.
- [33] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2019, pp. 1–11.
- [34] M. Muller, A. Bibi, S. Giancola, S. Alsubaihi, and B. Ghanem, "TrackingNet: A Large-scale Dataset and Benchmark for Object Tracking in the Wild," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 300–317.