Low-Light Enhancer for UAV Night Tracking Based on Zero-DCE++

Yihong Zhang, Yinjian Li, Qin Lin

School of Information Science and Technology, Donghua University, Shanghai, China
Email: 2211924@mail.dhu.edu.cn

Abstract

Unmanned aerial vehicle (UAV) target tracking tasks can currently be successfully completed in daytime situations with enough lighting, but they are unable to do so in nighttime scenes with inadequate lighting, poor contrast, and low signal-to-noise ratio. This letter presents an enhanced low-light enhancer for UAV nighttime tracking based on Zero-DCE++ due to its advantages of low processing cost and quick inference. We developed a light-weight UCBAM capable of integrating channel information and spatial features and offered a fully considered curve projection model in light of the low signal-to-noise ratio of night scenes. This method significantly improved the tracking performance of the UAV tracker in night situations when tested on the public UAVDark135 and compared to other cutting-edge low-light enhancers. By applying our work to different trackers, this search shows how broadly applicable it is.

Keywords

Low-Light Enhancement, Nighttime Tracking, Zero-DCE++, UAV Application

1. Introduction

With its compact size and adaptable motion, UAV platforms have attracted a lot of attention and applications for visual target tracking, a component of computer vision. Visual target tracking paired with UAVs opens up a larger range of applications, including maritime rescue [1], autonomous landing [2], and self-localization [3]. Trackers based on correlation filter and trackers based on Siamese have both shown advancement in a number of large-scale benchmarks [4]. These benchmarks are primarily in well-lit situations, although in real-world UAV applications, UAVs frequently operate in dimly lit areas. Modern trackers

struggle to ensure their robustness in such settings, which limits the range of conventional UAV uses.

The images captured by the UAV vision system have low brightness, low contrast, and severe noise quality issues as a result of insufficient photons and low signal-to-noise ratio in low-light settings, which prevents subsequent target trackers from gathering useful depth features. In real life, the prediction head might easily generate inaccurate results due to noise and uneven illumination. By changing the camera aperture, lengthening the exposure time, and utilizing artificial lighting, we may produce well-lit photographs with standard handheld cameras. These techniques do have some limitations, such as the fact that flash cannot be used in aerial locations where the lighting depends on the weather [5]. By employing low-light video frames as a training set, a data-driven target tracker can develop strong tracking performance in low-light conditions; however, the existing annotated dataset cannot satisfy the training needs.

In this instance, the best course of action is to place a low-light enhancer in front of the standard target tracker. Due to the real-time requirements of the tracking task, the low-light enhancer needs to be lightweight and fast inference, which is not possible with many existing low-light enhancers. Zero-DCE++ transforms the light enhancement problem into a curve projection task, improving the speed of inference. Considering the effect of noise on the enhancement task, single light curve parameter map and a set of noise curve parameter maps are used to the nonlinear curve projection. UCBAM with the addition of the convolutional block attention module (CBAM) [6] is constructed to develop a lightweight low-light enhancer and thoroughly learn the illumination and noise profiles. Testing on a benchmark shows that our approach for nighttime UAV target tracking is effective.

The contributions of this letter are summarized as follows.

This search proposes to use a light curve parameter map and a set of noise curve parameter maps to accomplish nonlinear curve projection.

A convolutional neural network (UCBAM) is constructed, which consists an encoder, a decoder and a channel space attention module to accomplish the learning of the curve parameter maps.

Testing on the open test dataset UAVDark135 shows the usefulness and broad application of our proposed low-light enhancer in the area of UAV night tracking.

2. Related Work

Low-light enhancement techniques have been rapidly improved in recent years. Chun Wei *et al.* [7] enhance the brightness of images by building a deep learning network based on a data-driven strategy and Retinex theory with BM3D noise reduction. Jiang *et al.* [8] suggest an unsupervised learning technique dubbed EnlightenGAN to get over the restriction that supervised training needs enough matched low light/normal offsets. However, the above methods require heavy calculations. A few simple techniques have been suggested; Ziteng Cui *et al.* [9]

use an adaptive transformer network for the tasks of exposure correction and dark light enhancement. By establishing a weight-sharing illumination learning method that includes a self-calibrating module, SCI [10] is able to do multi-stage cascade training while only requiring a single step of inference. Zero-DCE [11] proposes to convert the image enhancement problem into an image-specific curve estimation problem and fit this curve using a higher-order function. ChebyLighter [12] uses Chebyshev approximation to fit this curve. But these works are not directly available for UAV night tracking due to the fact that they are designed as low-level vision tasks.

The two types of target tracking algorithms are CNN-based and based on discriminant correlation filters (DCF) [13]. Siamese-based trackers stand out among CNN-based methods for their excellent balance of accuracy and efficiency. The Siamese-based tracker [14] contains a template branch and a search branch, and has the ability to find the most similar sample to the sample to be matched from multiple candidate samples. DiMP [15] enhances the traditional DCF model to learn the deep feature representation in order to increase the tracking performance, and ATOM adds an end-to-end scale estimate branch to enable end-to-end optimization on the tracking dataset. SiamRPN [16] introduces a region suggestion network to generate accurate suggestions, which achieves outstanding tracking accuracy. TransT [17] introduces a self-attentive mechanism in the feature fusion stage, which improves the tracking performance. While these general-purpose trackers exhibit exceptional performance under good illumination, their robustness is reduced significantly when under low illumination conditions.

The importance of nighttime target tracking research has been increasingly recognized, but the majority of previous studies have attempted to solve this issue by altering the data source. For example, complex sensors such as infrared cameras [18], low-light cameras [19], and event cameras [20] were used to acquire images. Some scholars have made efforts to solve the night tracking problem from the visible light perspective. DarkLighter [21] exploits the Retinex model to improve the brightness in visible images, but is not closely linked to the target tracking task. SCT [22] is trained in a task-inspired manner to be used in object tracking. Wenjing Wang *et al.* [23] propose the use of discrete integration to learn illumination-enhanced concave curves that can be used for advanced vision tasks.

3. Proposed Approach

In order to estimate single illumination curve parameter map and multiple noise curve parameter maps, as shown in Figure 1, the improved approach based on Zero-DCE++ employs a convolutional neural network created by an encoder-decoder and introducing a channel-space attention mechanism. Then, it generates images with favorable illumination conditions by applying a comprehensive curve projection model. The approach illuminates the template and searching patches for trackers.
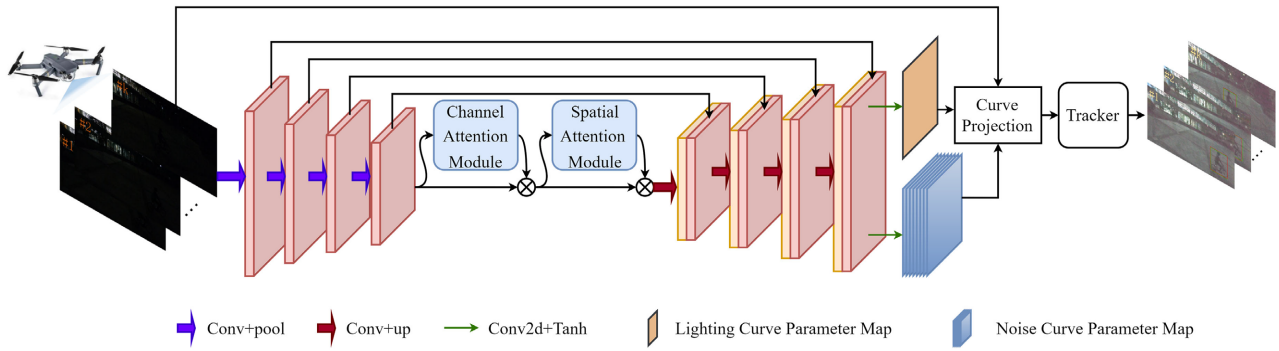
Figure 1. Overall structure diagram.

3.1. Improvement of Curve Projection

In Zero-DCE++, single light curve map is involved in the non-linear curve projection task to implement low light enhancement. The low-light image X_i is used to estimate a light curve parameter map I, and Equation (1), Equation (2) and Equation (3) are used to determine the enhanced image X_o.

$$X^0 = X_i \tag{1}$$

and

$$X^t = X^{t-1} + I \otimes X^{t-1}\left(1 - X^{t-1}\right), t = 1, \cdots, T \tag{2}$$

and

$$X_O = X^T \tag{3}$$

where, $\otimes$ denotes element-by-element multiplication and X^t signifies the intermediate outcome of the iterative procedure.

However, noise is not considered in the above curve projection model. So, a novel model is proposed which takes as input a low-illumination image X_i, a single light curve parameter map I and a set of noise curve parameter maps. The operation process can be expressed by Equation (4), Equation (5) and Equation (6).

$$X^0 = X_i \tag{4}$$

and

$$\begin{cases} \hat{X}^{t-1} = X^{t-1} - N^{t-1} \\ X^t = \hat{X}^{t-1} + I \otimes \hat{X}^{t-1} \otimes \left(1 - \hat{X}^{t-1}\right) \end{cases}, t = 1, \cdots, T \tag{5}$$

and

$$X_o = X^T \tag{6}$$

In Equation (5), the noise in X^{t-1} is removed using distinct noise curve parameter maps N^{t-1}, which improves the nonlinear curve projection for lighting correction. T is set to 8 in Equation (5) and Equation (6).

3.2. Improvement of Deep Curve Estimation Network

In Zero-DCE++, the depth curve estimation network contains only the simple

encoder and decoder, which is not enough to learn features of images. Therefore, the convolutional neural network called UCBAM is created to estimate the light curve parameter map and noise curve parameter maps. The encoder, CBAM, and decoder are the components of this network.

The encoder consists of numerous down-sampling and convolution blocks, where the down-sampling block utilizes 2D convolution to shrink the spatial dimension of the features and the convolution block uses 2D convolution to increase the number of channels of the features.

In order to execute adaptive feature optimization, CBAM, a light-weight attention module for feedforward convolutional neural networks, it calculates the attention maps for the feedforward feature maps from the channel and space dimensions separately, multiplies them by the feedforward feature maps to perform adaptive feature optimization.

The channel attention module can be expressed as [6]:

$$M_c(F_1) = \psi\left(MLP\left(AvgPool(F_1)\right) + MLP\left(MaxPool(F_1)\right)\right) \tag{7}$$

where, $AvgPool()$ represents the average pooling operation in spatial dimension, $MaxPool()$ indicates the maximum pooling operation in spatial dimension, MLP is a multilayer perceptron with an implicit layer, and $\psi()$ denotes the sigmoid operation.

Spatial attention module can be phrased as [6]:

$$M_s(F_2) = \psi\left(f^{7\times7}\left(AvgPool(F_2); MaxPool(F_2)\right)\right) \tag{8}$$

where, $AvgPool()$ indicates the average pooling operation in the channel dimension, and $MaxPool()$ defines the maximum pooling operation in the channel dimension. $f^{7\times7}$ represents the convolutional layer with the kernel size of 7×7, $\psi()$ means sigmoid operation.

The convolution block structure of the decoder is similar to that of the encoder, with the exception that the convolution blocks are intended to reduce the number of feature channels. The decoder also includes up-sampling blocks. The inverse convolution process is utilized by the up-sampling block to expand the feature space dimension.

3.3. Improvement of Loss Function

Zero-DCE++ proposes the four types of no-reference loss functions in order to satisfy the human visual perception, which are the spatial consistency loss, the exposure control loss, the color constancy loss, the illumination smoothness loss. However, the nighttime tracking task expects the target tracker to extract similar features from the light-adjusted image and the image under good illumination [22]. AlexNet [24] is often used as the backbone network for a generic target tracker. Hence, the following equation can be used to express the loss function:

$$loss = \sum_{i=3}^{5} \frac{\left\|\Phi_i(X) - \Phi_i(Y)\right\|_2}{c_i w_i h_i} \tag{9}$$

where X means the low-light image modified image, Y denotes the image taken under normal illumination, and $\Phi_i(\ )$ denotes the operation to extract the ith layer of features from the AlexNet network. c_i displays the number of channels of the feature map, w_i shows the feature map's width, and h_i indicates the feature map's height.

4. Experimental Design

4.1. Training

The LOL dataset, which has 485 pairs of low-light/normal-light photos with an image size of 600 × 400, serves as our training set, and a randomly cropped training patch of 256 × 256 is used for training. With AdamW acting as the optimizer, the training process lasts 100 periods, the first five of which serve as warm-up. The initial learning rate is set to 0.0009, the weight decay to 0.02, and the batch size to 16.

The experiments were conducted on a PC configured with NVIDIA RTX A5000 GPU, Intel(R) Xeon(R) Gold 6330 CPU.

4.2. Evaluation Metrics

This letter chooses the evaluation metrics from the single-target tracking job instead of the evaluation metrics from the standard low-light enhancement test because our study is for the performance of nighttime UAV tracking rather than a completely low-level task. The tracker is run throughout the test sequence using the OPE [25] evaluation method, providing the first frame of the grounding-truth initialized tracker, and using accuracy and success rates to gauge our performance. The distance between the center of the target tracker bounding box and the center of the matching real bounding box serves as a gauge for accuracy. The accuracy plot's vertical coordinate represents the proportion of frames whose distance is below the threshold to all frames, while the accuracy plot's horizontal coordinate represents the distance threshold. The accuracy plot's vertical coordinate at the horizontal position of 20 is utilized to determine the tracker's rating. The concatenation of the true bounding box and the tracker bounding box yields a success rate estimate. The success graph displays the percentage of frames with IOU larger than the subsequent threshold, and the trackers' success rates are graded based on the success graph's area under the curve.

5. Experimental Results and Discussion

As shown in Figure 2, Zero-DCE++, EnlightenGAN, SCI, IAT, and SCT are compared, as well as our work. These low light enhancers are used directly in the pre-processing stage of SiamRPN++, which is evaluated in UAVDark135. Clearly, the improvements in all aspects make the work gives more performance gains than Zero-DCE++ for nighttime UAV tracking. On the test dataset, the basic tracker with Zero-DCE++ only has a success rate of 0.391 and a precision of 0.49, compared to a success rate of 0.444 and a precision of 0.562 for the base tracker
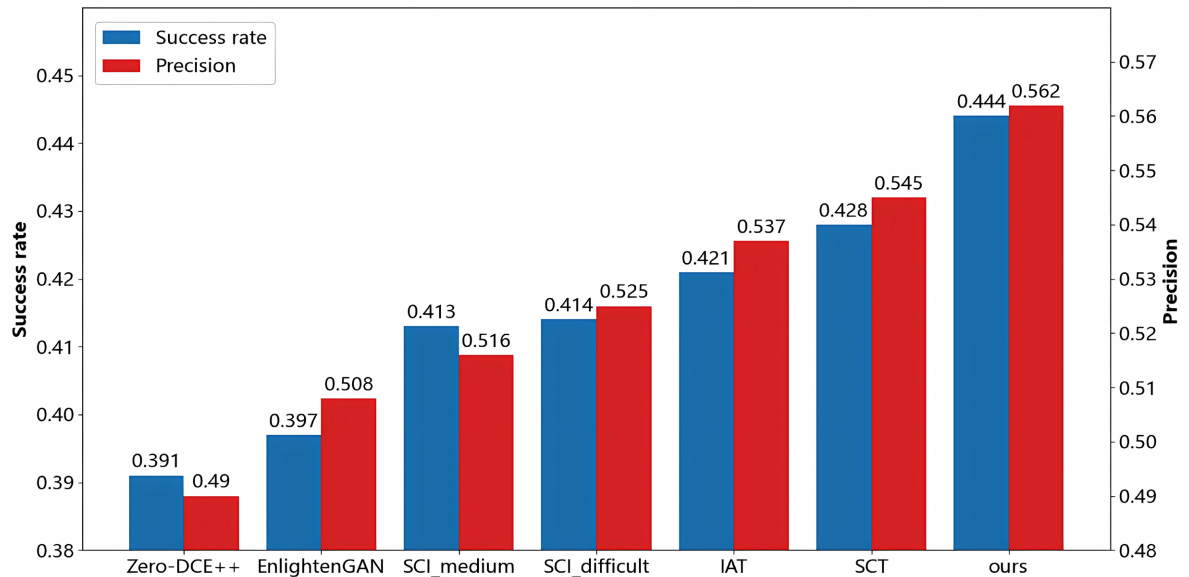
Figure 2. Comparison with different low-light enhancers.

with the work. It also has superiority over other advanced low-light enhancers in this regard. SCT, being a low light enhancer also developed for target tracking work, assists significantly for night UAV tracking task than other advanced low light enhancers. Better metrics are demonstrated in our work when compared to SCT, which it's believed that is a result of the UCBAM network's suitability for this estimate task and our curve projection model.

As shown in **Figure 3**, this work is applied to numerous single-target trackers tested on the publicly available UAVDark135, where DiMP18 and DiMP50 are based on correlation filters, and SiamAPN++, SiamRPN++ are based on twin networks. In addition, DiMP18 uses ResNet-18 as the backbone network, DiMP50 uses ResNet-50 as the backbone network, and SiamAPN++, SiamRPN++ uses AlexNet as the backbone network. **Figure 4** illustrates how this work improved performance on the tracker, for instance by assisting SiamAPN++ in achieving improvements of 27.2% and 30.9%. Our results still hold up on trackers using different networks as backbone networks even though we utilize AlexNet to extract features during training.

In **Figure 5**, some trace screenshots are provided. For visualization purposes, the first column is the original low-light image and the remaining columns are globally enhanced. During tracking, our work is enhanced for template branching and searching for patches to track. It is easy to see from the figure that our work benefits the perceptibility of the tracker at night, thus improving the performance of the tracker.

The ablation experiment is conducted on the improved part to show the necessity and superiority of the improvement. This is conducted with the noise term, the spatial attention module and the channel attention module in mind. SiamRPN++ was used as the basis tracker in the experiments, which were run on UAVDark135 to examine the effects of each absence on the enhancement of
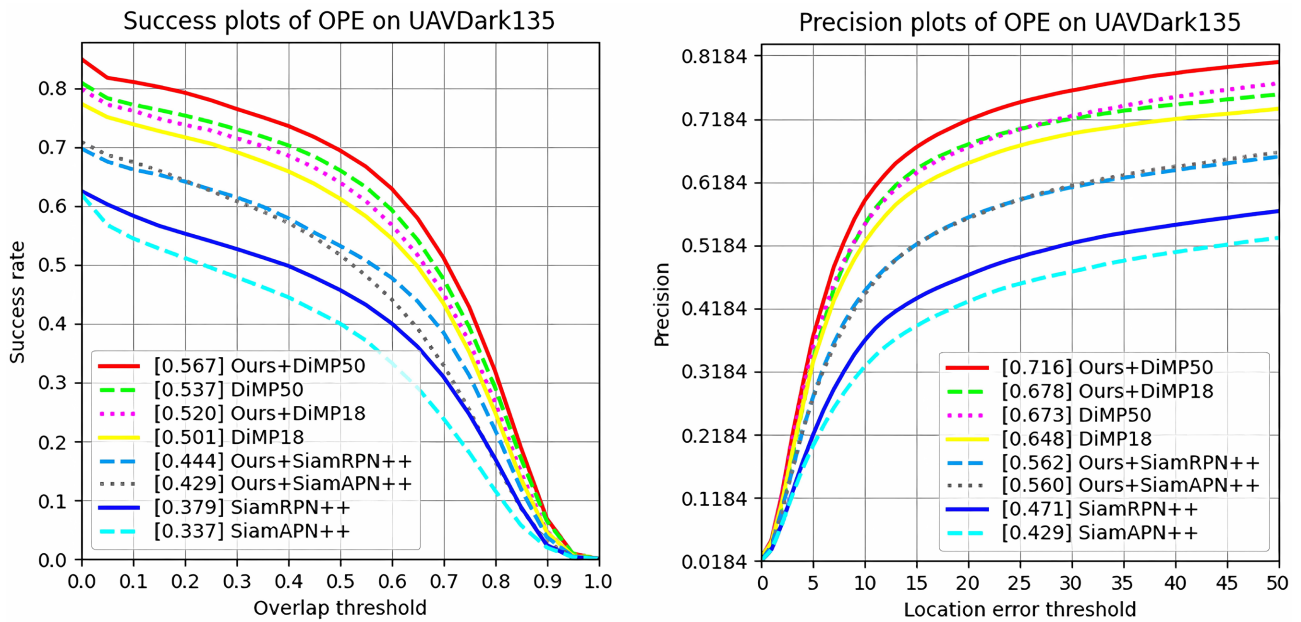
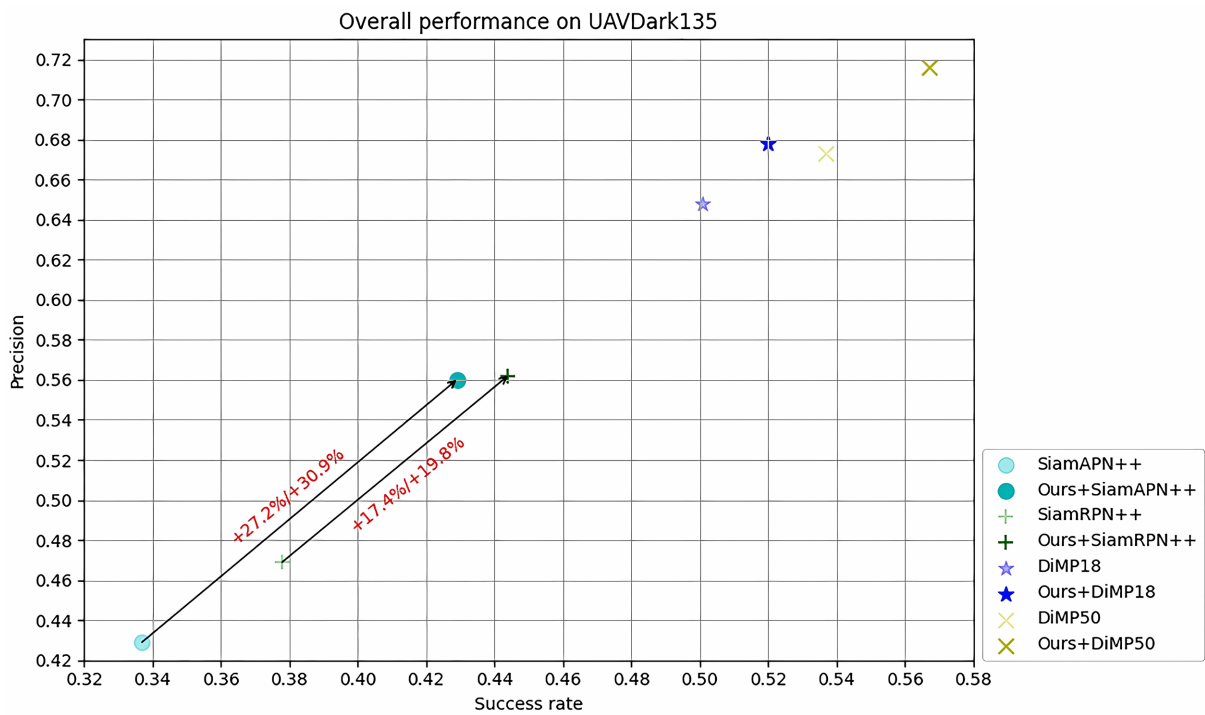**Figure 3.** Overall performance of SOTA trackers with Ours activated or not.

Figure 4. Performance improvement of SOTA trackers brought by Ours.

nighttime tracking performance. Table 1 uses the abbreviations CAM for channel attention module, SAM for spatial attention module, and NCPM for noise profile parameter maps. The performance of tracking is somewhat degraded in the absence of CAM and SAM, and when both are absent and just the noise term is present, the performance is only increased by 7.9%/9.5%, demonstrating the necessity of CBAM. In addition, the noise contour parameter map contributed
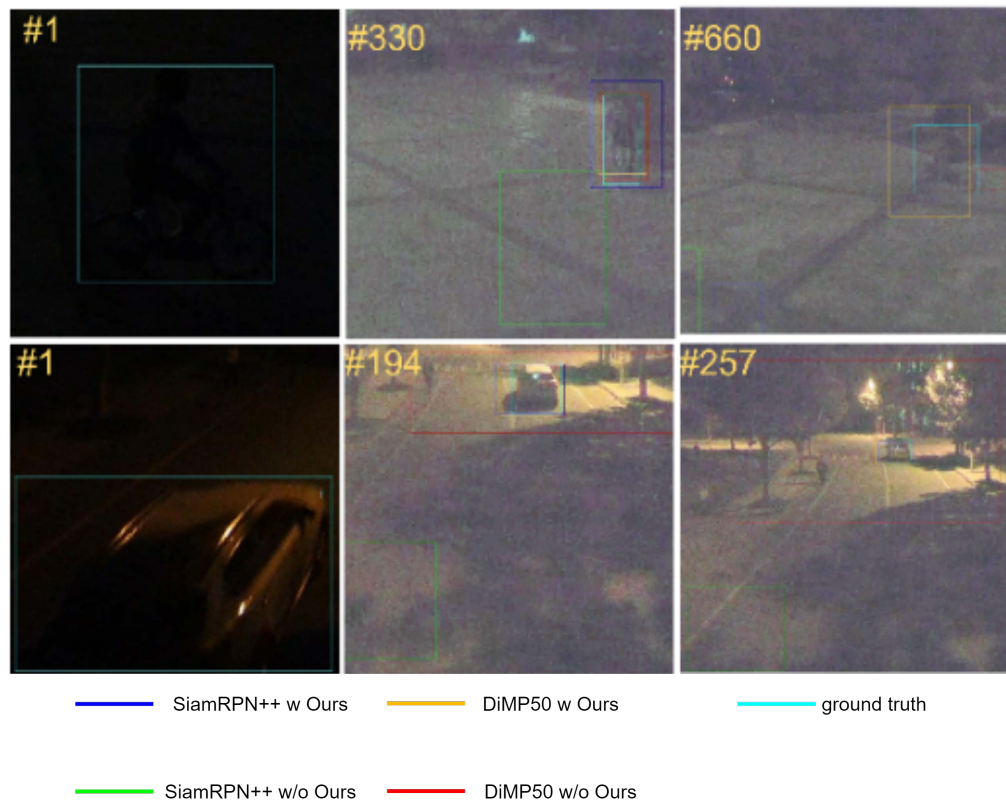
Figure 5. Overall performance of SOTA trackers with ours activated or not.

Table 1. Results of ablation experiment.

CAM	SAM	NCPM	Succ./Prec.	Δ (%)
√	√	√	0.444/0.562	16.2/18.1
	√	√	0.426/0.539	11.5/13.2
√		√	0.422/0.538	10.4/13.0
		√	0.412/0.521	7.9/9.5
√	√		0.4/0.499	4.7/4.8
			0.382/0.476	

significantly to the performance improvement, giving the tracker an increase of 11.5% in success rate and a 13.3% in precision.

6. Conclusion

This letter suggests an improved low light enhancer based on Zero-DCE++ for UAV night tracking. The low-light enhancer builds a lightweight network with encoder, decoder, and CBAM and suggests a thoroughly considered curve projection model to improve the precision and success rate of the tracker on UAVs. This is obvious that our effort will significantly contribute to the growth of UAV nighttime applications. The work also has large room for improvement, such as getting a better approach to reduce noise.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Zhang, Y., Li, S., Li, D., Zhou, W., Yang, Y., Lin, X. and Jiang, S. (2020) Parallel Three-Branch Correlation Filters for Complex Marine Environmental Object Tracking Based on a Confidence Mechanism. *Sensors*, **20**, 5210.
https://doi.org/10.3390/s20185210

[2] Yuan, B., Ma, W. and Wang, F. (2022) High Speed Safe Autonomous Landing Marker Tracking of Fixed Wing Drone Based on Deep Learning. *IEEE Access*, **10**, 80415-80436. https://doi.org/10.1109/ACCESS.2022.3195286

[3] Tang, D., Fang, Q., Shen, L. and Hu, T. (2022) Onboard Detection-Tracking-Localization. *IEEE/ASME Transactions on Mechatronics*, **25** 1555-1565.

[4] Javed, S., Danelljan, M., Khan, F.S., Khan, M.H., Felsberg, M. and Matas, J. (1983) Visual Object Tracking with Discriminative Filters and Siamese Networks: A Survey and Outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[5] Singh, A., Chougule, A., Narang, P., Chamola, V. and Yu, F.R. (2022) Low-Light Image Enhancement for UAVs with Multi-Feature Fusion Deep Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, **19**, 1-5.
https://doi.org/10.1109/LGRS.2022.3181106

[6] Woo, S., Park, J., Lee, J.-Y. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 3-19.
https://doi.org/10.1007/978-3-030-01234-2_1

[7] Wei, C., Wang, W., Yang, W. and Liu, J. (2018) Deep Retinex Decomposition for Low-Light Enhancement.

[8] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., *et al.* (2021) Enlightengan: Deep Light Enhancement without Paired Supervision. *IEEE Transactions on Image Processing*, **30**, 2340-2349. https://doi.org/10.1109/TIP.2021.3051462

[9] Cui, Z., Li, K., Gu, L., Su, S., Gao, P., Jiang, Z., *et al.* (2020) Illumination Adaptive Transformer.

[10] Ma, L., Ma, T., Liu, R., Fan, X. and Luo, Z. (2022) Toward Fast, Flexible, and Robust Low-Light Image Enhancement. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 5637-5646.
https://doi.org/10.1109/CVPR52688.2022.00555

[11] 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. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 1780-1789. https://doi.org/10.1109/CVPR42600.2020.00185

[12] Pan, J., Zhai, D., Bai, Y., Jiang, J., Zhao, D. and Liu, X. (2022) ChebyLighter: Optimal Curve Estimation for Low-Light Image Enhancement. *Proceedings of the 30th ACM International Conference on Multimedia*, Lisboa, 10-14 October 2022, 1358-1366. https://doi.org/10.1145/3503161.3548135

[13] Danelljan, M., Bhat, G., Khan, F.S. and Felsberg, M. (2019) Atom: Accurate Tracking by Overlap Maximization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 4660-4669.
https://doi.org/10.1109/CVPR.2019.00479

[14]   Bertinetto, L., Valmadre, J., Henriques, J.F., Vedaldi, A. and Torr, P.H. (2016) Fully-Convolutional Siamese Networks for Object Tracking. *Computer Vision ECCV* 2016 *Workshops*: Amsterdam, 8-10 and 15-16 October 2016, 850-865. https://doi.org/10.1007/978-3-319-48881-3_56

[15]   Bhat, G., Danelljan, M., Gool, L.V. and Timofte, R. (2019) Learning Discriminative Model Prediction for Tracking. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 6182-6191. https://doi.org/10.1109/ICCV.2019.00628

[16]   Li, B., Yan, J., Wu, W., Zhu, Z. and Hu, X. (2018) High Performance Visual Tracking with Siamese Region Proposal Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 8971-8980. https://doi.org/10.1109/CVPR.2018.00935

[17]   Wang, N., Zhou, W., Wang, J. and Li, H. (2021) Transformer Meets Tracker: Exploiting Temporal Context for Robust Visual Tracking. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 1571-1580. https://doi.org/10.1109/CVPR46437.2021.00162

[18]   Ding, M., Chen, W.-H. and Cao, Y.-F. (2022) Thermal Infrared Single-Pedestrian Tracking for Advanced Driver Assistance System. *IEEE Transactions on Intelligent Vehicles*, **1**, 814-824. https://doi.org/10.1109/TIV.2022.3140344

[19]   Ye, C. and Gu, Y. (2022) WTB-LLL: A Watercraft Tracking Benchmark Derived by Low-Light-Level Camera. *Pattern Recognition and Computer Visio*n: 5*th Chinese Conference*, *PRCV* 2022, Shenzhen, 4-7 November 2022, 707-720. https://doi.org/10.1007/978-3-031-18916-6_56

[20]   Ramesh, B., Zhang, S., Lee, Z.W., Gao, Z., Orchard, G. and Xiang, C. (2018) Long-Term Object Tracking with a Moving Event Camera. *The* 29*th BMVC*, Newcastle upon Tyne, 3-6 September 2018, 1-12.

[21]   Ye, J., Fu, C., Zheng, G., Cao, Z. and Li, B. (2021) Darklighter: Light up the Darkness for UAV Tracking. 2021 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Prague, 27 September-1 October 2021, 3079-3085. https://doi.org/10.1109/IROS51168.2021.9636680

[22]   Ye, J., Fu, C., Cao, Z., An, S., Zheng, G. and Li, B. (2017) Tracker Meets Night: A Transformer Enhancer for UAV Tracking. *IEEE Robotics and Automation Letters*, **7**, 3866-3873. https://doi.org/10.1109/LRA.2022.3146911

[23]   Wang, W., Xu, Z., Huang, H. and Liu, J. (2022) Self-Aligned Concave Curve: Illumination Enhancement for Unsupervised Adaptation. *Proceedings of the* 30*th ACM International Conference on Multimedia*, Lisboa, 10-14 October 2022, 2617-2626. https://doi.org/10.1145/3503161.3547991

[24]   Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2022) Imagenet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90. https://doi.org/10.1145/3065386

[25]   Wu, Y., Lim, J. and Yang, M.-H. (2013) Online Object Tracking: A Benchmark. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, 23-28 June 2013, 2411-2418. https://doi.org/10.1109/CVPR.2013.312