Low-Light Object Tracking in UAV Videos

Subjects: Computer Science, Artificial Intelligence Contributor: Lifan Sun , Shuaibing Kong , Zhe Yang , Dan Gao , Bo Fan

Unmanned aerial vehicles (UAVs) visual object tracking under low-light conditions serves as a crucial component for applications, such as night surveillance, indoor searches, night combat, and all-weather tracking. However, the majority of the existing tracking algorithms are designed for optimal lighting conditions. In low-light environments, images captured by UAV typically exhibit reduced contrast, brightness, and a signal-to-noise ratio, which hampers the extraction of target features. Moreover, the target's appearance in low-light UAV video sequences often changes rapidly, rendering traditional fixed template tracking mechanisms inadequate, and resulting in poor tracker accuracy and robustness.

unmanned aerial vehicle

low-light tracking

1. Introduction

Visual object tracking is a fundamental task in computer vision that finds extensive applications in the unmanned aerial vehicle (UAV) domain. Recent years have witnessed the emergence of new trackers that exhibit exceptional performance in UAV tracking ^{[1][2][3]}, which is largely attributed to the fine manual annotation of large-scale datasets ^{[4][5][5][2]]}. However, the evaluation standards and tracking algorithms currently employed are primarily designed for favorable lighting conditions. In real-world scenarios, low-light conditions such as nighttime, rainy weather, and small spaces are often encountered, resulting in images with low contrast, low brightness, and low signal-to-noise ratio compared to normal lighting. These discrepancies give rise to inconsistent feature distributions between the two types of images, thereby rendering it challenging to extend trackers designed for favorable lighting conditions to low-light scenarios ^{[8][9]}, making it more challenging for UAV tracking.

Low-light UAV video sequences exhibit poor robustness and tracking drift when conventional object-tracking algorithms are employed, as illustrated in **Figure 1**. The issue of object tracking under low-light conditions can be divided into two sub-problems: enhancing low-light image features and tackling the challenge of target appearance changes in low-light video sequences. First, the low contrast, low brightness, and low signal-to-noise ratio of low-light images make feature extraction more arduous compared to normal images. Insufficient feature information hampers subsequent object-tracking tasks and constrains the performance of object-tracking algorithms. Another obstacle hindering the effectiveness of object-tracking algorithms arises from the characteristics of low-light video sequences. During tracking, the target's appearance often changes, and when it becomes occluded or deformed, its features no longer correspond to the original template features, resulting in tracking drift. Such challenges are commonplace in vision object-tracking tasks and are more pronounced under low-light conditions due to the

unstable lighting conditions, which serve as a crucial limiting factor for the performance of object-tracking algorithms.



Figure 1. Trackers performance under low-light conditions.

2. Low-Light Image Enhancement

The objective of low-light image enhancement is to improve the quality of images by making the details that are concealed in darkness visible. In recent years, this area has gained significant attention and undergone continuous development and improvement in various computer vision domains. Two main types of algorithms are used for low-light image enhancement, namely model-based methods and deep learning-based methods.

Model-based methods were developed earlier and are based on the Retinex theory ^[10]. According to this theory, low-light images can be separated into illuminance and reflectance components. The reflectance component contains the essential attributes of the image, including edge details and color information, while the illuminance component captures the general outline and brightness distribution of the objects in the image. Fu et al. ^{[11][12]} were the first to use the L2 norm to constrain illumination and proposed an image enhancement method that simultaneously estimates illuminance and reflectance components in the linear domain. This method demonstrated that the linear domain formula is more suitable than the logarithmic domain formula. Guo et al. ^[13] used relative total variation ^[14] as a constraint on illumination and developed a structure-aware smoothing model to obtain better estimates of illuminance components. However, this model has the disadvantage of overexposure. Li et al. ^[15]

regularization terms to jointly estimate a piecewise smooth illumination and a structure-displaying reflectance in the optimization problem of illumination and reflectance. They also modeled noise removal and low-light enhancement as a unified optimization goal. Additionally, Ref. ^[16] proposed a semi-decoupled decomposition model to simultaneously enhance brightness and suppress noise. Although some models use camera response characteristics (e.g., LEACRM ^[17]), their effects are often not ideal and require manual adjustment of numerous parameters when dealing with real scenes.

In recent years, deep learning-based methods have rapidly emerged with the advancement of computer technology. Li et al. [18] proposed a control-based method for optimizing UAV trajectories, which incorporates energy conversion efficiency by directly deriving the model from the voltage and current flow of the UAV's electric motor. EvoXBench ^[19] introduced an end-to-end process to address the lack of a general problem statement for NAS tasks from an optimization perspective. Zhang et al. ^[20] presented a low-complexity strategy for superresolution (SR) based on adaptive low-rank approximation (LRA), aiming to overcome the limitations of processing large-scale datasets. Jin et al. [21] developed a deep transfer learning method that leverages facial recognition techniques to achieve a computer-aided facial diagnosis, validated in both single disease and multiple diseases with healthy controls. Zheng et al. ^[22] proposed a two-stage data augmentation method for automatic modulation classification in deep learning, utilizing spectral interference in the frequency domain to enhance radio signals and aid in modulation classification. This marks the first instance where frequency domain information has been considered to enhance radio signals for modulation classification purposes. Meanwhile, deep learning-based lowlight enhancement algorithms have also made significant progress. Chen et al. [23] created a new dataset called LOL dataset by collecting low/normal light image pairs with adjusted exposure time. This dataset is the first to contain image pairs obtained from real scenes for low-light enhancement research, making a significant contribution to learning-based low-light image enhancement algorithm research. Many algorithms have been trained based on this dataset. The retinal network, designed in ^[23], generated unnatural enhancement results. KinD ^[24] improved some of the issues in the retinal network by adjusting the network architecture and introducing some training losses. DeepUPE ^[25] proposed a low-light image enhancement network that learned an image-toillumination component mapping. Yang et al. ^[26] developed a fidelity-based two-stage network that first restores signals and then further enhances the results to improve overall visual quality, trained using a semi-supervised strategy. EnGAN ^[27] used a GAN-based unsupervised training method to enhance low-light images using unpaired low/normal light data. The network was trained using carefully designed discriminators and loss functions while carefully selecting training data. SSIENet ^[28] proposed a maximum entropy-based Retinex model that could estimate illuminance and reflectance components simultaneously while being trained only with low-light images. ZeroDCE ^[29] heuristically constructed guadratic curves with learned parameters to estimate parameter mapping from low-light input and used curve projection models for iterative light enhancement of low-light images. However, these models focus on adjusting the brightness of images and do not consider the noise that inevitably occurs in real-world nighttime imaging. Liu et al. [30] introduced prior constraints based on Retinex theory to establish a lowlight image enhancement model and constructed an overall network architecture by unfolding its optimization solution process. Recently, Ma et al. [31] added self-correcting modules during training to reduce the model parameter size and improve inference speed.

However, these algorithms have limited stability, and it is difficult to achieve sustained superior performance, particularly in unknown real scenes where unclear details and inappropriate exposure are common and without good solutions for noise in images.

3. Object Tracking

In recent years, object tracking algorithms can be classified into methods based on discriminative correlation filtering ^{[32][33][34]} and methods based on Siamese networks. Achieving end-to-end training on trackers based on discriminative correlation filtering is challenging due to their complex online learning process. Moreover, limited by low-level manual features or inappropriate pre-trained classifiers, trackers based on discriminative correlation filtering become ineffective under complex conditions.

With the continuous improvement of computer performance and the establishment of large-scale datasets, tracking algorithms based on Siamese networks have become mainstream due to their superior performance. The Siamese network series of algorithms started with SINT ^[35] and SiamFC ^[36], which treat target tracking as a similarity learning problem and train Siamese networks using large amounts of image data. SiamFC introduced a correlation layer for feature fusion which significantly improved accuracy. Based on the success of SiamFC, subsequent improvements were made. CFNet [37] added a correlation filter to the template branch to make the network shallower and more efficient. DSiam ^[38] proposed a dynamic Siamese network that could be trained on labeled video sequences as a whole, fully utilizing the rich spatiotemporal information of moving objects and achieving improved accuracy with an acceptable speed loss. RASNet [39] used three attention mechanisms to weight the space and channels of SiamFC features, enhancing the network's discriminative ability by decomposing the coupling of feature extraction and discriminative analysis. SASiam ^[40] established a Siamese network containing semantic and appearance branches. During training, the two branches were separated to maintain specificity. During testing, the two branches were combined to improve accuracy. However, these methods require multi-scale testing to cope with scale changes and cannot handle proportion changes caused by changes in target appearance. To obtain more accurate target bounding boxes, B. Li et al. [41] introduced a region proposal network (RPN) ^[42] into the Siamese network framework, achieving simultaneous improvement in accuracy and speed. SiamRPN++ [43] further adopted a deeper backbone and feature aggregation architecture to exploit the potential of deep networks on Siamese networks and improve tracking accuracy. SiamMask [44] introduced a mask branch to simultaneously achieve target tracking and image segmentation. Xu et al. [45] proposed a set of criteria for estimating the target state in tracker design and designed a new Siamese network, SiamFC++, based on SiamFC. DaSiamRPN [46] introduced existing detection datasets to enrich positive sample data and difficult negative sample data to improve the generalization and discrimination ability of trackers. It also introduced a local-to-global strategy to achieve good accuracy in long-term tracking. Anchor-free methods use per-pixel regression to predict four offsets on each pixel, reducing the hyperparameters caused by the introduction of RPNs. SiamBAN [47] proposed a tracking framework, containing multiple adaptive heads, that does not require multi-scale search or predefined candidate boxes, that directly classifies objects in a unified network, and that regresses bounding boxes. SiamCAR [48] added a centrality branch to help determine the position of the target center point and further improve tracking accuracy. Recently, Transformer ^[49] was integrated into the Siamese framework to simulate global information and improve tracking performance.

Regarding target tracking algorithms under low-light conditions, a DCF framework integrated with a low-light enhancer was proposed in ^[50]. However, it is limited to hand-crafted features and lacks transferability. Ye et al. ^[51] developed a new unsupervised domain adaptation framework that uses a day-night feature discriminator to adversarially train a daytime tracking model for nighttime tracking. However, there is currently insufficient targeted research on this issue.

References

- 1. Xie, X.; Xi, J.; Yang, X.; Lu, R.; Xia, W. STFTrack: Spatio-Temporal-Focused Siamese Network for Infrared UAV Tracking. Drones 2023, 7, 296.
- 2. Memon, S.A.; Son, H.; Kim, W.G.; Khan, A.M.; Shahzad, M.; Khan, U. Tracking Multiple Unmanned Aerial Vehicles through Occlusion in Low-Altitude Airspace. Drones 2023, 7, 241.
- 3. Yeom, S. Long Distance Ground Target Tracking with Aerial Image-to-Position Conversion and Improved Track Association. Drones 2022, 6, 55.
- Fan, H.; Bai, H.; Lin, L.; Yang, F.; Chu, P.; Deng, G.; Yu, S.; Huang, M.; Liu, J.; Xu, Y.; et al. Lasot: A high-quality large-scale single object tracking benchmark. Int. J. Comput. Vis. 2021, 129, 439– 461.
- 5. Huang, L.; Zhao, X.; Huang, K. Got-10k: A large high-diversity benchmark for generic object tracking in the wild. IEEE Trans. Pattern Anal. Mach. Intell. 2019, 43, 1562–1577.
- Real, E.; Shlens, J.; Mazzocchi, S.; Pan, X.; Vanhoucke, V. Youtube-boundingboxes: A large highprecision human-annotated data set for object detection in video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5296–5305.
- Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
- 8. Ye, J.; Fu, C.; Cao, Z.; An, S.; Zheng, G.; Li, B. Tracker meets night: A transformer enhancer for UAV tracking. IEEE Robot. Autom. Lett. 2022, 7, 3866–3873.
- Ye, J.; Fu, C.; Zheng, G.; Cao, Z.; Li, B. Darklighter: Light up the darkness for uav tracking. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 3079–3085.

- 10. Rahman, Z.u.; Jobson, D.J.; Woodell, G.A. Retinex processing for automatic image enhancement. J. Electron. Imaging 2004, 13, 100–110.
- 11. Fu, X.; Liao, Y.; Zeng, D.; Huang, Y.; Zhang, X.P.; Ding, X. A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. IEEE Trans. Image Process. 2015, 24, 4965–4977.
- Fu, X.; Zeng, D.; Huang, Y.; Zhang, X.P.; Ding, X. A weighted variational model for simultaneous reflectance and illumination estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2782–2790.
- 13. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. IEEE Trans. Image Process. 2016, 26, 982–993.
- 14. Xu, L.; Yan, Q.; Xia, Y.; Jia, J. Structure extraction from texture via relative total variation. ACM Trans. Graph. (TOG) 2012, 31, 1–10.
- 15. Li, M.; Liu, J.; Yang, W.; Sun, X.; Guo, Z. Structure-revealing low-light image enhancement via robust retinex model. IEEE Trans. Image Process. 2018, 27, 2828–2841.
- 16. Hao, S.; Han, X.; Guo, Y.; Xu, X.; Wang, M. Low-light image enhancement with semi-decoupled decomposition. IEEE Trans. Multimed. 2020, 22, 3025–3038.
- 17. Ren, Y.; Ying, Z.; Li, T.H.; Li, G. LECARM: Low-light image enhancement using the camera response model. IEEE Trans. Circuits Syst. Video Technol. 2018, 29, 968–981.
- Li, B.; Li, Q.; Zeng, Y.; Rong, Y.; Zhang, R. 3D trajectory optimization for energy-efficient UAV communication: A control design perspective. IEEE Trans. Wirel. Commun. 2021, 21, 4579–4593.
- 19. Lu, Z.; Cheng, R.; Jin, Y.; Tan, K.C.; Deb, K. Neural architecture search as multiobjective optimization benchmarks: Problem formulation and performance assessment. arXiv 2023, arXiv:2208.04321.
- 20. Zhang, Y.; Luo, J.; Zhang, Y.; Huang, Y.; Cai, X.; Yang, J.; Mao, D.; Li, J.; Tuo, X.; Zhang, Y. Resolution enhancement for large-scale real beam mapping based on adaptive low-rank approximation. IEEE Trans. Geosci. Remote Sens. 2022, 60, 5116921.
- 21. Jin, B.; Cruz, L.; Gonçalves, N. Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis. IEEE Access 2020, 8, 123649–123661.
- 22. Zheng, Q.; Zhao, P.; Li, Y.; Wang, H.; Yang, Y. Spectrum interference-based two-level data augmentation method in deep learning for automatic modulation classification. Neural Comput. Appl. 2021, 33, 7723–7745.
- 23. Panareda Busto, P.; Gall, J. Open set domain adaptation. In Proceedings of the IEEE international Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 754–763.

- 24. Zhang, Y.; Guo, X.; Ma, J.; Liu, W.; Zhang, J. Beyond brightening low-light images. Int. J. Comput. Vis. 2021, 129, 1013–1037.
- 25. Wang, R.; Zhang, Q.; Fu, C.W.; Shen, X.; Zheng, W.S.; Jia, J. Underexposed photo enhancement using deep illumination estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6849–6857.
- Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; Liu, J. From fidelity to perceptual quality: A semisupervised approach for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 3063–3072.
- Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. Enlightengan: Deep light enhancement without paired supervision. IEEE Trans. Image Process. 2021, 30, 2340–2349.
- 28. Zhang, Y.; Di, X.; Zhang, B.; Wang, C. Self-supervised image enhancement network: Training with low light images only. arXiv 2020, arXiv:2002.11300.
- 29. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1780–1789.
- Liu, R.; Ma, L.; Zhang, J.; Fan, X.; Luo, Z. 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, Nashville, TN, USA, 20–25 June 2021; pp. 10561–10570.
- Ma, L.; Ma, T.; Liu, R.; Fan, X.; Luo, Z. Toward fast, flexible, and robust low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5637–5646.
- Hare, S.; Golodetz, S.; Saffari, A.; Vineet, V.; Cheng, M.M.; Hicks, S.L.; Torr, P.H. Struck: Structured output tracking with kernels. IEEE Trans. Pattern Anal. Mach. Intell. 2015, 38, 2096– 2109.
- 33. Kalal, Z.; Mikolajczyk, K.; Matas, J. Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell. 2011, 34, 1409–1422.
- 34. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. 2014, 37, 583–596.
- 35. Tao, R.; Gavves, E.; Smeulders, A.W. Siamese instance search for tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1420–1429.

- Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H. Fully-convolutional siamese networks for object tracking. In Proceedings of the Computer Vision–ECCV 2016 Workshops, Amsterdam, The Netherlands, 8–10 and 15–16 October 2016; Proceedings, Part II 14. pp. 850– 865.
- Kiani Galoogahi, H.; Fagg, A.; Lucey, S. Learning background-aware correlation filters for visual tracking. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1135–1143.
- Yang, T.; Chan, A.B. Learning dynamic memory networks for object tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 152–167.
- Wang, Q.; Teng, Z.; Xing, J.; Gao, J.; Hu, W.; Maybank, S. Learning attentions: Residual attentional siamese network for high performance online visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4854–4863.
- 40. He, A.; Luo, C.; Tian, X.; Zeng, W. A twofold siamese network for real-time object tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4834–4843.
- Li, B.; Yan, J.; Wu, W.; Zhu, Z.; Hu, X. High performance visual tracking with siamese region proposal network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8971–8980.
- 42. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. 2015, 28.
- 43. Li, B.; Wu, W.; Wang, Q.; Zhang, F.; Xing, J.; Yan, J. Siamrpn++: Evolution of siamese visual tracking with very deep networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4282–4291.
- 44. Wang, Q.; Zhang, L.; Bertinetto, L.; Hu, W.; Torr, P.H. Fast online object tracking and segmentation: A unifying approach. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1328–1338.
- 45. Xu, Y.; Wang, Z.; Li, Z.; Yuan, Y.; Yu, G. Siamfc++: Towards robust and accurate visual tracking with target estimation guidelines. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12549–12556.
- Zhu, Z.; Wang, Q.; Li, B.; Wu, W.; Yan, J.; Hu, W. Distractor-aware siamese networks for visual object tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 101–117.

- 47. Chen, Z.; Zhong, B.; Li, G.; Zhang, S.; Ji, R. Siamese box adaptive network for visual tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 6668–6677.
- Guo, D.; Wang, J.; Cui, Y.; Wang, Z.; Chen, S. SiamCAR: Siamese fully convolutional classification and regression for visual tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 6269–6277.
- 49. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. Adv. Neural Inf. Process. Syst. 2017, 30.
- 50. Li, B.; Fu, C.; Ding, F.; Ye, J.; Lin, F. ADTrack: Target-aware dual filter learning for real-time antidark UAV tracking. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 496–502.
- 51. Ye, J.; Fu, C.; Zheng, G.; Paudel, D.P.; Chen, G. Unsupervised domain adaptation for nighttime aerial tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 8896–8905.

Retrieved from https://encyclopedia.pub/entry/history/show/109067