Real-Time UAV Target Tracking

Subjects: Imaging Science & Photographic Technology
Contributor: Zhuo Li

Real-time UAV tracking refers to the tracking processing that is completed within the actual time of the acquisition of the image sequence by the drone's airborne imaging device. This processing is used to acquire the motion parameters of the target in the image moment by moment, including the target's position, speed, acceleration, and motion trajectory, etc.

visual object tracking   unmanned aerial vehicle (UAV) videos

1. Visual tracking for UAV videos

In recent years, due to their many outstanding advantages in performance and cost, unmanned aerial vehicles (UAVs) have increasingly been deployed in many fields, such as security monitoring, disaster relief, agriculture, military equipment, sports and entertainments, etc. Correspondingly, a huge amount of visual data has been produced, and the demand for intelligent processing of UAV videos has increased significantly.

Due to the release of new benchmark datasets and the improved methodologies, single-target tracking has become a research hotspot, and the related work has made considerable advances. From the perspective of technical means, the current mainstream single-target trackers can be divided into two categories: trackers based on Discriminative Correlation Filter (DCF) and trackers based on deep learning. Minimum Output Sum of Squared Error (MOSSE) is one of the most representative trackers based on DCF [1]. These kind of trackers have fast tracking speed and are easy to transplant to the embedded hardware platform for real-time processing, but the tracking accuracy is relatively low. Therefore, it is difficult for them to meet the high-accuracy tracking requirements. Afterwards, the researchers proposed various improved DCF-based trackers through optimizing in many aspects, such as Circulant Structure of tracking-by-detection with Kernels (CSK) tracker [2], Kernelized Correlation Filters (KCF) tracker [3], and Spatially Regularized Discriminative Correlation Filter (SRDCF) tracker [4], etc. These trackers achieve a significant improvement in tracking accuracy, but at the same time, the tracking speed is significantly reduced.

With the rapid development of deep learning, many trackers based on Convolution Neural Networks (CNN) have emerged. Compared to the previous trackers, they can yield higher tracking accuracy [5][6][7][8][9][10][11][12][13]. However, for UAV target tracking scenarios, due to numerous challenges, such as relatively small object sizes and various orientation changes, the above trackers show degraded performance to different degrees. An accurate and efficient tracker is still needed to perform the target tracking task in UAV videos.

2. Real-time UAV Tracker by MultiRPN-DIDNet

MultiRPN-DIDNet is a real-time target tracking method based on multiple Region Proposal Networks (RPNs) and Distance-IoU Discriminative Network (DIDNet) for UAV videos. Firstly, an instance-based RPN suitable for the target tracking task is constructed under the framework of Simases Neural Network. RPN is to perform bounding box regression and classification, in which channel attention mechanism is integrated to improve the representative capability of the deep features. The RPNs built on the Block 2, Block 3 and Block 4 of ResNet50 output their own Regression (Reg) coefficients and Classification scores (Cls) respectively, which are weighted and then fused to determine the high-quality region proposals. Secondly, a DIDNet is designed to correct the candidate target's bounding box finely through the fusion of multi-layer features, which is trained with the Distance-IoU loss. Experimental results on the public datasets of UAV20L and DTB70 show that, compared with the state-of-the-art UAV trackers, the proposed MultiRPN-DIDNet can obtain better tracking performance with fewer region proposals and correction iterations. As a result, the tracking speed has reached 33.9 frames per second (FPS), which can meet the requirements of real-time tracking tasks.

As shown in Figure 1, the single object tracking method proposed in this paper consists of multiple RPNs and a DIoU discriminative network, in which ResNet-50 is used as the backbone network for feature extraction. RPN is constructed under the framework of SNN to perform bounding box regression and classification. The RPNs built on the Block 2, Block 3 and Block4 of ResNet50 output their own Reg coefficients and Cls scores respectively. They are weighted and then fused through a set of offline learning weight coefficients, obtaining the final Reg coefficients and Cls scores. The foreground with a higher Cls score is selected as the anchor, and the corresponding region proposal is determined by combining the Reg coefficients of the anchor. The convolutional features from multiple layers of ResNet50 are fused. The fused features and the information of the candidate area are input into the DIoU discriminative network, and the region proposal with the best DIoU value is finally determined as the tracking result.
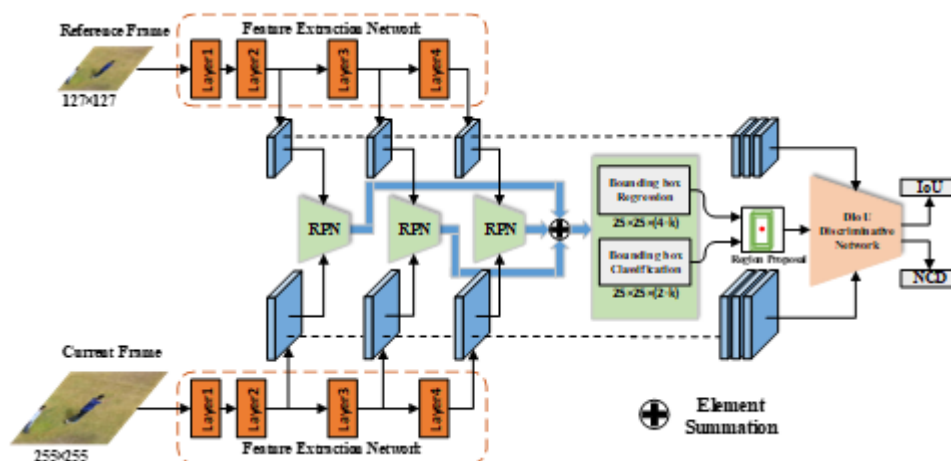

Figure 1. Target tracking framework combining multiple RPNs and DIoU discriminative network. The multiple RPNs are used to determine high-quality candidate regions, and the DIoU discriminative network performs the correction of the candidate regions, and then outputs the final result.

References

1. Bolme, D.S.; Beveridge, J.R.; Draper, B.A.; Lui, Y.M. Visual object tracking using adaptive correlation filters. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010; pp. 2544–2550.

2. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels. In Proceedings of the European Conference on Computer Vision (ECCV), Florence, Italy, 7–13 October 2012; pp. 702–715.

3. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-Speed Tracking with Kernelized Correlation Filters. IEEE Trans. Pattern Anal. Mach. Intell. 2015, 37, 583–596.

4. Danelljan, M.; Häger, G.; Khan, F.S.; Felsberg, M. Learning Spatially Regularized Correlation Filters for Visual Tracking. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 4310–4318.

5. Valmadre, J.; Bertinetto, L.; Henriques, J.; Vedaldi, A.; Torr, P.H.S. End-to-End Representation Learning for Correlation Filter Based Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5000–5008.

6. Wang, Q.; Gao, J.; Xing, J.; Zhang, M.; Hu, W. Dcfnet: Discriminant correlation filters network for visual tracking. arXiv 2017, arXiv:1704.04057.

7. Held, D.; Thrun, S.; Savarese, S. Learning to Track at 100 FPS with Deep Regression Networks. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 749–765.

8. Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H.S. Fully-Convolutional Siamese Networks for Object Tracking. In Proceedings of the European Conference on Computer Vision Workshops (ECCVW), Amsterdam, The Netherlands, 8–16 October 2016; pp. 850–865.

9. Tao, R.; Gavves, E.; Smeulders, A.W.M. Siamese Instance Search for Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1420–1429.

10. 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. 103–119.

11. Danelljan, M.; Bhat, G.; Khan, F.S.; Felsberg, M. ATOM: Accurate Tracking by Overlap Maximization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 4655–4664.

12. 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 Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 4277–4286.

13. Huang, L.; Zhao, X.; Huang, K. Bridging the gap between detection and tracking: A unified approach. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 29 October–1 November 2019; pp. 3999–4009.