International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I3P110

Volume 72 Issue 3, 99-105, March 2024

Original Article

Categorizing Video Datasets: Video Object Detection, Multiple and Single Object Tracking

Sara Bouraya[1], Abdessamad Belangour[2]

[1,2]*Laboratory of Information Technology and Modeling Hassan II University, Faculty of Sciences Ben M'sik Casablanca, Morocco.*

[1]*Corresponding Author : sarabouraya95@gmail.com*

Received: 12 August 2023 Revised: 15 November 2024 Accepted: 13 February 2024 Published: 17 March 2024

Abstract *- Video Object detection, Single Object detection, Multiple Object Detection are crucial tasks in computer vision, enabling various real-world applications. The success of these tasks algorithms heavily relies on the availability of high-quality datasets for training and evaluation. This paper presents a comprehensive categorization of datasets specifically designed for multiple object detection, single object detection, and video object detection tasks in computer vision. Object detection and tracking are fundamental problems in the field, and accurate and diverse datasets are essential for training and evaluating detection and tracking algorithms effectively. By analyzing the characteristics of datasets for multiple object detection, single object detection, and video object detection, this paper serves as a valuable resource to drive advancements in object detection, tracking algorithms and systems. Accurate and diverse datasets are pivotal in the pursuit of robust and efficient object detection, tracking solutions across various applications in computer vision.*

Keywords *- Multiple Object Tracking, Single Object Tracking, Video Object Detection, Video Dataset, VOD dataset, MOT dataset, SOT dataset.*

1. Introduction

Computer vision has revolutionized the way machines perceive and understand visual information, enabling them to interpret the world like never before. Among the diverse range of computer vision tasks, video analysis has emerged as a critical domain with numerous real-world applications. In this introduction, we explore three fundamental computer vision tasks: Video Object Detection, Multiple Object Tracking, and Single Object Tracking, each contributing to a comprehensive understanding of dynamic scenes captured in video sequences. The paper underscores the importance of thoroughly reviewing the categorization of video datasets.

Such an examination is crucial for gaining a comprehensive understanding of video object detection, single-object tracking, and multiple-object tracking. By systematically categorizing these datasets, researchers and practitioners can discern patterns, evaluate the efficacy of current methodologies, and identify areas ripe for enhancement. This analysis lays the groundwork for advancing techniques in video analysis, ultimately contributing to the development of more precise and efficient algorithms for object detection and tracking tasks.

1.1. Video Object Detection

Video object detection involves detecting and localizing objects within consecutive frames of a video. Unlike static image object detection, this task demands not only accurate object recognition but also temporal consistency to track objects seamlessly over time. Video object detection is a computer vision task that involves detecting and localizing objects of interest within video frames. It is an extension of the traditional object detection problem, which is usually performed on static images. In video object detection, the goal is to identify and track objects across consecutive frames in a video sequence.

1.2. Single Object Tracking

Single-object tracking focuses on monitoring the motion of a specific object of interest throughout a video sequence. Unlike multiple object tracking, which deals with several objects, this task concentrates on maintaining the trajectory of a single, pre-defined target. Single object tracking is widely applied in object surveillance, visual analysis of sports events, and human-computer interaction systems, where a specific object's movements and actions are of particular interest. Robust single-object tracking algorithms take into account challenges such as appearance changes, occlusions, and scale variations to ensure consistent and accurate tracking results.

1.3. Multiple Object Tracking

Multiple Object Tracking addresses the challenging problem of associating detected objects across video frames to maintain their identities and trajectories. It involves not only detecting objects individually but also understanding

Fig. 1 Categorizing Video Analysis Datasets, including Video Object Detection (VOD), Multiple Object Tracking (MOT), Single Object Tracking (SOT)

their movements and interactions as they traverse the scene. This task is particularly important in crowded scenarios and dynamic environments, where objects may occlude or interact with one another. Advanced tracking algorithms, often integrating computer vision and machine learning techniques, facilitate robust and accurate multi-object tracking in videos. Applications of multiple object tracking include surveillance, traffic monitoring, and video analytics for retail and crowd management. In summary, video object detection, multiple object tracking, and single object tracking are three pivotal computer vision tasks that collectively contribute to a deeper understanding of the dynamic world captured in video data. The advancements in deep learning, tracking algorithms, and computational resources have propelled these tasks to achieve remarkable results in various real-world applications, fostering safer, more efficient, and intelligent systems across industries. As research in computer vision continues to evolve, we can anticipate further breakthroughs, ultimately enhancing the capabilities of video analysis and its impact on society.

2. Background

In the rapidly advancing field of computer vision, the availability of high-quality and diverse video datasets has played a crucial role in driving research and development of cutting-edge algorithms. These datasets serve as valuable resources for training, evaluating, and benchmarking various computer vision tasks, particularly those involving video analysis. In the figure below (See figure 1), we explore three essential categories of video datasets: Video Object Detection datasets, Multiple Object Tracking datasets, and Single Object Tracking datasets, each catering to distinct challenges and applications in the realm of dynamic scene understanding.

These datasets facilitate fair comparisons between different algorithms, encourage the adoption of standardized evaluation metrics, and foster collaborations among researchers worldwide. As the demand for intelligent video analysis solutions continues to grow, the continuous expansion and refinement of video datasets are essential to driving progress in computer vision, paving the way for safer, more efficient, and more reliable video-based applications across various domains. Figure 1 contains three different video dataset categories; we gathered just the datasets related to our work area, which are object detection and tracking, and we categorize them as mentioned in Figure 1. For Video Object Detection datasets, we gather multiple datasets, including ImageNet VID Image, a large-scale dataset containing thousands of videos with object annotations. It covers a wide range of object categories and complex scenes and KITTI Object Detection originally designed for autonomous driving, this dataset includes videos with object annotations suitable for video object detection tasks. As well as YouTube-BoundingBoxes Contains bounding box annotations for objects in YouTube videos, making it a valuable resource for video object detection research. For Multiple Object Tracking, we arranged different datasets, including MOT challenge datasets as a benchmark dataset specifically designed for evaluating multiple object tracking algorithms. As well as VidDrone includes dronecaptured images and videos across diverse scenarios, such as urban, rural, and congested areas, enabling the evaluation of algorithms in different environmental conditions. Figure 1 contains the rest of the datasets related to this research area. Finally, for the Single Object Tracking task, we survived the most commonly used datasets for this task, including LaSot, which addresses the challenges of long-term tracking where the target object undergoes significant appearance changes, occlusions, and other complex scenarios. In addition TrackingNet dataset is a benchmark dataset designed for evaluating single object tracking algorithms.

It focuses on the task of tracking a single object across a sequence of videos. frames, encompassing a wide range of scenarios and challenges commonly encountered in real-world tracking applications.

3. Comparison

Video object detection involves identifying and localizing objects within a video sequence. To enable the development and evaluation of robust algorithms in this area, a collection of 16 datasets was compiled, each catering to different complexities and use cases. These datasets include well-known benchmarks, such as those mentioned in Table 1. Each of these datasets presents its unique challenges, such as occlusions, diverse lighting conditions, and various object categories. By combining these datasets, researchers can gain insights into the performance and limitations of their video object detection models across different scenarios. Multiple object tracking involves the simultaneous tracking of multiple objects over time. The compilation of 20 diverse datasets facilitates the exploration of algorithms that can robustly handle object interactions, occlusions, and various motion patterns. Notable datasets in this category include a variety of datasets see Table 2.These datasets encompass a wide range of scenarios, including crowded scenes, object occlusions, and varying object densities. Researchers can leverage these datasets to develop and benchmark multiple object-tracking algorithms that excel in real-world conditions. Single-object tracking involves the continuous tracking of a single object as it moves throughout a video sequence. A compilation of 25 datasets enables the exploration of algorithms that excel at maintaining the tracking accuracy and robustness of individual objects. Noteworthy datasets in this category are presented in Table 3. These datasets cover a spectrum of challenges, such as fast object motion, scale variations, and occlusions. By utilizing these datasets, researchers can assess the effectiveness of their single object tracking algorithms across different tracking scenarios.

Table 1. Video object detection datasets

Dataset	#Reference	#Frames	#Cat	#Res	Year
IMAGENET VID	[1]	2017.6K	30	1280x1080	2015
YOUTUBE-BOUNDINGBOXES	[2]	5.6M	-	-	2017
DRONECROWD	[4]	33,600	-	1920x1080	2021
VIL-100	[5]	10,000	-	-	2021
UA-DETRAC	[6]	140.0k	4	960x540	2015
MOT17DET	1	11.2K	1	1920x1080	2017
OKUTAMA-ACTION	[7]	77.4K	1	3840x2160	2017
UAVDT-DET	[8]	40.7K	3	1080x540	2018
DRONESURF	[9]	411.5k	1	1280x720	2019
VISDRONE	[10]	40.0k	10	3840x2160	2018
SEADRONESSEE	[11]	54,000	-	-	2021
CALTECH	1	250.000	-	640x480	2009
KAIST	1	-	-	-	-
KITTI-D	1	11.2k	-	1392x512	2014
CARPK	[12]	1.5k	-	-	2017
ETH	[13]	-	-	640x480	-

Table 2. Multiple object tracking datasets

Dataset	#Reference	#NbFrames	#Clips	#Resolution	Year
VISDRONE	[10]	40.0k	-	3840x2160	2018
TAO	[14]	2674.4k	-	1280x720	2020
GMOT-40	[15]	9,643	-	-	2021
BDD100K	[16]	14K	70	-	2020
DRONECROWD	[4]	33,600	122	1920×1080	2021
KITTI TRACKING	[17]	19.1	-	1392x512	2013
UA-DETRAC TRACKING	[6]	140.1k	-	960x540	2015
DUKE MTMC	[19]	2852.2k	-	1920x1080	2016
CAMPUS	[20]	929.5k	-	1417x2019	2016
UAVDT-MOT	[21]	80k	100	-	2018
SEADRONESSEE	[11]	54.105k	22	-	2021
PATHTRACK	[22]	-	-	-	2017
MVMHAT	[23]	-	-	-	2021
WILDTRACK	[24]	-	-	1920x1080	2018
DIVOTRACK	[25]	-	-	-	2023
UCSD	[26]	200	-	160×240	2010
MOT15	[27]	11.3k	-	1920x1080	2015
MOT17	1	11.2k	21	1920X1080	2017
MOT16	[28]	11.2k	-	1920X1080	2016

Table 3. Single object tracking datasets

DATASET	#Reference	#NbFrames	#Clips	Year
OXUVA	[29]	1.5M	366	2018
TNL2K	[30]	1 ,244,340	2000	2021
ALOV300+	[31]	151.6k	314	2014
NUS-PRO	[32]	135k	365	2016
UAV123	[33]	113K	123	2016
OTB2013	[34]	-	-	2013
OTB2015	[35]	-	-	2015
OTB100	[35]	59.0k	100	2015
OTB50	[35]	29 k	51	2015
DTB70	-	15.8k	70	2017
TC-128	[36]	55 k	128	2018
NFS	[37]	383k	100	2012
CDTB	[38]	102k	80	2019
TREK-150	[39]	97k	150	2022
UAVDT-SOT	[8]	37 .2k	50	2018
GOT-10K	[40]	1500.0k	10.0k	2018
TRACKINGNET	[41]	1443.1k	30.6k	2018
MDOT	[42]	259.8k	373	2020
LASOT	[43]	3870.0k	1.55k	2020
ANTI-UAV	[44]	585.9k	318	2021
VISDRONE	[10]	139.3k	167	2018
SEADRONESSEE	[11]	393,295	208	2021
VOT2016	[45]	21.5k	60	2016
VOT2017	[18]	21.3k	60	2017
VOT2019	1	20k	60	2019

4. Results and Discussion

The aggregation of 16 video object detection datasets, 20 multiple object tracking datasets, and 25 single object tracking datasets provide a comprehensive resource for advancing the fields of video analysis, object detection, and tracking.

These datasets, along with their associated reference, frame count, video count, and presentation year offer researchers a rich foundation for developing, benchmarking, and improving state-of-the-art algorithms in the exciting domains of video object detection and object tracking.

As a result, after categorizing and analyzing the various datasets related to video object detection and tracking, several noteworthy observations and insights emerge:

4.1. Diverse Application Scenarios

The datasets cover a wide range of application scenarios, from urban environments and surveillance to autonomous driving and aerial imagery. This diversity underscores the need for tracking and detection algorithms that can adapt to different real-world contexts.

4.2. Challenges in Object Detection

Some datasets focus on object detection tasks within videos, highlighting challenges such as object occlusions, scale variations, and cluttered backgrounds. This emphasizes the importance of algorithms that can accurately localize and identify objects even in complex situations.

4.3. Complex Interactions in Tracking

Multiple object tracking datasets often involve complex object interactions, crowded scenes, and occlusions. This calls for tracking algorithms capable of handling dynamic and intricate object behaviors over time.

4.4. Long-Term Tracking Challenges

Benchmarks for long-term tracking underline the difficulty of maintaining accurate tracking across extended sequences. Algorithms need to address issues such as appearance changes and target reidentification for sustained tracking performance.

4.5. Deep Learning Dominance

With the prevalence of deep learning techniques, many datasets are designed to facilitate the training and evaluation of deep neural networks. This reflects the impact of deep learning in advancing object detection and tracking capabilities.

4.6. Benchmarking and Innovation

The availability of standardized benchmark datasets encourages healthy competition and innovation in the field. Researchers can benchmark their algorithms against state-of-the-art methods, driving continuous improvements in tracking and detection accuracy.

4.7. Realism and Practicality

Many datasets focus on capturing real-world challenges, enhancing the practicality of developed algorithms for deployment in real-world scenarios. This aligns with the goal of making computer vision technologies more applicable in various industries.

4.8. Dataset Evolution

The continual updates and maintenance of datasets ensure that they remain relevant and reflective of evolving challenges. This keeps the datasets up-to-date with emerging trends and technologies.

4.9. Resource for Researchers and Practitioners

These datasets collectively provide a valuable resource for both researchers and practitioners to develop, test, and validate object detection and tracking algorithms. They serve as a foundation for advancing the field and contributing to technological progress.In summary, the analysis of categorized video object detection and tracking datasets underscores the complexity and diversity of challenges in these domains. The availability of comprehensive datasets fosters the development of more accurate, robust, and adaptable algorithms, furthering the capabilities of computer vision systems for real-world applications.

5. Conclusion

In conclusion, the fields of video object detection, multiple object tracking and single object tracking have witnessed remarkable advancements thanks to the availability of diverse and well-structured datasets. These datasets play a pivotal role in training and evaluating algorithms for video object detection as well as single and multiple object tracking tasks. As this paper highlights, practitioners and researchers have access to a comprehensive range of categorized datasets that include a plethora of challenges and scenarios.

For Single Object Tracking (SOT), datasets such as OXUVA LASOT offer a diverse collection of annotated videos. These datasets facilitate the development of algorithms capable of accurately tracking single object trajectories within various challenges and scenarios. In the realm of Multiple Object Tracking (MOT), MOT variants and other datasets emerge as a valuable resource, encompassing complex scenes with multiple objects overlapping and interacting.

These handfuls of datasets empower researchers to create robust models that can track numerous objects simultaneously. Regarding video object detection, ImageNet VID and other datasets are truly interesting resources for the computer vision era that will help researchers and practitioners acquire a plethora of different datasets that contain interesting issues. As we delve further into the realms of artificial intelligence and computer vision, the importance of high-quality datasets cannot be overstated. Datasets will help us build more reliable algorithms.

References

[1] Olga Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, pp. 211-252, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[2] Esteban Real et al., "YouTube-BoundingBoxes: A Large High-precision Human-Annotated Data Set for Object Detection in Video," *Proceeding - IEEE Conference on Computer Vision and Pattern Recognition*, *CVPR 2017*, pp. 7464-7473, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[3] Adel Ahmadyan et al., "Objectron: A Large Scale Dataset of Object-Centric Videos in the Wild with Pose Annotations," *Proceeding of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7818-7827, 2021. [Google Scholar] [Publisher Link]

[4] Longyin Wen et al., "Detection, Tracking, and Counting Meets Drones in Crowds: A Benchmark," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, pp. 7808-7817, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Yujun Zhang et al., "VIL-100: A New Dataset and A Baseline Model for Video Instance Lane Detection," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, pp. 15661-15670, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Longyin Wen et al., "UA-DETRAC: A New Benchmark and Protocol for Multi-object Detection and Tracking," *Computer Vision and Image Understanding*, vol. 193, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Mohammadamin Barekatain et al., "Okutama-Action: An Aerial View Video Dataset for Concurrent Human Action Detection," *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, pp. 2153-2160, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[8] H. Yu et al., "The Unmanned Aerial Vehicle Benchmark: Object Detection, Tracking and Baseline," *International Journal of Computer Vision*, vol. 128, pp. 1141-1159, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Isha Kalra et al., "DroneSURF: Benchmark Dataset for Drone-based Face Recognition," *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, Lille, France, pp. 1-7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Pengfei Zhu et al., "Vision Meets Drones: A Challenge," *arXiv*, pp. 1-11, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[11] Leon Amadeus Varga et al., "SeaDronesSee: A Maritime Benchmark for Detecting Humans in Open Water," *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, pp. 3686-3696, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Meng-Ru Hsieh, Yen-Liang Lin, and Winston H. Hsu, "Drone-Based Object Counting by Spatially Regularized Regional Proposal Network," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 4165-4173, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[13] Andreas Ess, Bastian Leibe, and Luc Van Gool, "Depth and Appearance for Mobile Scene Analysis," *2007 IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, Brazil, pp. 1-8, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[14] Achal Dave et al., "TAO: A Large-Scale Benchmark for Tracking Any Object," *Computer Vision-European Conference on Computer Vision ECCV 2020*, vol. 12350, pp. 436-454, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Hexin Bai et al., "GMOT-40: A Benchmark for Generic Multiple Object Tracking," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, pp. 6715-6724, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Fisher Yu et al., "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* Seattle, WA, USA, pp. 2633-2642, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] Andreas Geiger, Philip Lenz, and Raquel Urtasun, "Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[18] Matej Kristan et al., "The Visual Object Tracking VOT2017 Challenge Results," *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, Italy, pp. 1949-1972, 2017. [CrossRef] [Publisher Link]

[19] Ergys Ristani et al., "Performance Measures and a Data Set for multi-target, Multi-camera Tracking," *Computer Vision-European Conference on Computer Vision ECCV 2016 Workshop*, vol. 9914, pp. 17-35, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[20] Alexandre Robicquet et al., "Learning Social Etiquette: Human Trajectory Understanding in Crowded Scenes," *Computer Vision-European Conference on Computer Vision ECCV 2016*, vol. 9912, pp. 549-565, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[21] Dawei Du et al., "The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking," *Computer Vision-European Conference on Computer Vision ECCV 2018*, pp. 375–391, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[22] Santiago Manen et al., "PathTrack: Fast Trajectory Annotation with Path Supervision," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 290-299, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[23] Yiyang Gan et al., "Self-supervised Multi-View Multi-Human Association and Tracking," *MM '21: Proceedings of the 29th ACM International Conference on Multimedia*, pp. 282-290, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Tatjana Chavdarova et al., "WILDTRACK: A Multi-camera HD Dataset for Dense Unscripted Pedestrian Detection," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 5030-5039, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[25] Shenghao Hao et al., "DIVOTrack: A Novel Dataset and Baseline Method for Cross-View Multi-Object Tracking in DIVerse Open Scenes," a*rXiv*, pp. 1-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Vijay Mahadevan et al., "Anomaly Detection in Crowded Scenes," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, pp. 1975-1981, 2010. [CrossRef] [Publisher Link]

[27] Laura Leal-Taixé et al., "MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking," *arXiv*, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[28] Anton Milan et al., "MOT16: A Benchmark for Multi-Object Tracking," arXiv, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[29] Jack Valmadre et al., "Long-Term Tracking in the Wild: A Benchmark," *Computer Vision-European Conference on Computer Vision ECCV 2018*, vol. 11207, pp. 692–707, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[30] Xiao Wang et al., "Towards more Flexible and Accurate Object Tracking with Natural Language: Algorithms and Benchmark," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13763-13773, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[31] Arnold W.M. Smeulders et al., "Visual Tracking: An Experimental Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1442–1468, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[32] Annan Li et al., "NUS-PRO: A New Visual Tracking Challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 335–349, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[33] Matthias Mueller, Neil Smith, and Bernard Ghanem, *A Benchmark and Simulator for UAV Tracking*, European Conference on Computer Vision, vol. 9905, pp. 445-461, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[34] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang, "Online Object Tracking: A Benchmark," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, USA, pp. 2411-2418, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[35] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang, "Object Tracking Benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834-1848, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[36] Pengpeng Liang, Erik Blasch, and Haibin Ling, "Encoding Color Information for Visual Tracking: Algorithms and Benchmark," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5630-5644, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[37] Hamed Kiani Galoogahi et al., "Need for Speed: A Benchmark for Higher Frame Rate Object Tracking," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 1134-1143, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[38] Alan Lukezic et al., "CDTB: A Color and Depth Visual Object tracking Dataset and Benchmark," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), pp. 10012-10021, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[39] M. Dunnhofer et al., "Visual Object Tracking in First Person Vision," *International Journal of Computer Vision*, vol. 131, pp. 259-283, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[40] Lianghua Huang, Xin Zhao, and Kaiqi Huang, "Got-10k: A Large High-diversity Benchmark for Generic Object Tracking in the Wild," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1562-1577, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[41] Matthias Müller et al., "TrackingNet: A Large-Scale Dataset and Benchmark for Object Tracking in the Wild," *Computer Vision-European Conference on Computer Vision ECCV 2018*, vol. 11205, pp. 310-327, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[42] Pengfei Zhu et al., "Multi-Drone-Based Single Object Tracking with Agent Sharing Network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 4058-4070, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[43] Heng Fan et al., "LaSOT: A High-quality Large-Scale Single Object Tracking Benchmark," *International Journal of Computer Vision*, vol. 129, pp. 439-461, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[44] Nan Jiang et al., "Anti-UAV: A Large Multi-Modal Benchmark for UAV Tracking," a*rXiv*, pp. 1-13, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[45] Matej Kristan et al., "The Visual Object Tracking VOT2016 Challenge Results," *Computer Vision-European Conference on Computer Vision ECCV 2016*, vol. 9914, pp. 777-823, 2016. [CrossRef] [Google Scholar] [Publisher Link]