The 4th Workshop on UAVs in Multimedia: Capturing the World from a New Perspective

Fabian Deuser
Universität der Bundeswehr München
Germany
fabian.deuser@unibw.de

Yaxuan Li
University of Macau
China
yaxuanli.cn@gmail.com

Tingyu Wang
Hangzhou Dianzi University
China
tingyu.wang@hdu.edu.cn

Yujiao Shi
ShanghaiTech University
China
shiyj2@shanghaitech.edu.cn

Anna Bößendörfer
Universität der Bundeswehr München
Germany
anna.boessendoerfer@unibw.de

Shaofei Huang
University of Macau
China
nowherespyfly@gmail.com

Xiao Pan
Shenzhen University
China
xiaopan97@szu.edu.cn

Zhedong Zheng
University of Macau
China
zhedongzheng@um.edu.mo

Roger Zimmermann
National University of Singapore
Singapore
rogerz@comp.nus.edu.sg

Abstract

Uncrewed aerial vehicles (UAVs), also known as drones, have grown in popularity in recent years due to their ability to capture high-quality multimedia data from above. As multimedia applications such as aerial photography, cinematography, and mapping have grown in popularity, UAVs have emerged as powerful tools for gathering rich, diverse multimedia content. This workshop aims to bring together researchers, practitioners, and enthusiasts interested in UAV multimedia to explore the latest advancements, challenges, and opportunities in this exciting field. Topics covered will include aerial image and video processing, machine learning for UAV data analysis, UAV swarm technology, and UAV-based multimedia applications. In the context of the ACM Multimedia Conference, this workshop is highly relevant, as multimedia data from UAVs is becoming an important source of content for many applications. It will provide a platform for researchers to share their work and discuss potential collaborations and an opportunity for practitioners to learn about the latest developments in UAV multimedia technology. Overall, this workshop provides a unique opportunity to explore the exciting, rapidly evolving field of UAV multimedia and its potential impact on the broader multimedia community.

CCS Concepts

• **Computing methodologies** → **Vision for robotics**.

Keywords

UVA Multimedia Understanding, Drone-based Video Analysis

MM '26, Rio de Janeiro, Brazil

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:
Fabian Deuser, Yaxuan Li, Tingyu Wang, Yujiao Shi, Anna Bößendörfer, Shaofei Huang, Xiao Pan, Zhedong Zheng, and Roger Zimmermann. 2026. The 4th Workshop on UAVs in Multimedia: Capturing the World from a New Perspective. In *Proceedings of the 34th ACM International Conference on Multimedia (MM '26), November 10 - 14 2026, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Background and Motivation

Uncrewed aerial vehicles (UAVs), also known as drones, have grown in popularity in recent years due to their ability to capture high-quality multimedia data from above. This has opened up a wide range of applications, such as aerial photography [17, 38], cinematography [3, 15], mapping [1, 10, 12, 33, 42], agriculture [8, 16], geo-localization [13, 19–26] and delivery [5, 27]. UAVs have emerged as powerful tools for gathering rich, diverse multimedia content. They provide a unique vantage point with fewer obstructions [14, 28, 36, 37] and the ability to capture data from previously inaccessible or hard-to-reach locations (see Figure 1). The use of UAVs in multimedia applications has become even more significant, thanks to new technologies such as machine learning, computer vision, big data analytics, and transfer learning [4, 32, 34, 44]. These technologies have the potential to transform the way UAVs are used to capture and analyze multimedia content, opening up new possibilities for applications such as automated image and video analysis [9, 11, 14, 29, 43], real-time tracking [35, 37], predictive modeling [7, 18], natural language control [2] and citywide simulation [39].

For over 33 years, the ACM Multimedia Conference has been at the forefront of multimedia research, providing a forum for researchers and practitioners to exchange ideas, explore the latest advancements, and discuss the field's challenges. In recent years, the conference has also started to cover the topic of UAV multimedia, recognizing the growing importance of this area of research [6, 31, 36, 37, 42]. The use of UAVs for multimedia data capture is well established, and UAV data is now an important source of content for a wide range of applications. Due to the rapid

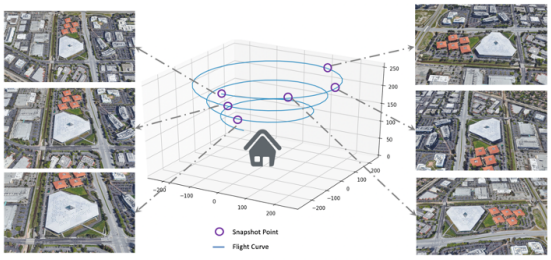

Figure 1: Different from conventional devices, UAV is a controllable aerial information capture platform with diverse viewpoints.

progress in this field and strong participation in previous events, it is crucial to maintain a dedicated space where researchers and practitioners from various disciplines can share ideas and promote UAV multimedia research.

This workshop is the 4th workshop on UAVs in Multimedia (UAVM) since 2023 [30, 40, 41]. The UAVM'23 and UAVM'24 challenges attracted approximately 50 registered participants and 20 teams from various countries. In the 3rd edition, the workshop attracted over 100 registered teams who submitted their results to the challenge. UAVM'23 accepted 13 workshop papers, UAVM'24 accepted 7 papers, and the 3rd edition accepted 13 papers.

This workshop aims to provide a platform for researchers and practitioners in UAV multimedia to discuss recent advances and emerging challenges in the field. Topics covered will include image and video processing, machine learning, swarm technology, and applications such as aerial photography, cinematography, and mapping. By bringing together experts from academia and industry, the workshop aims to promote knowledge sharing, encourage new collaborations, and advance UAV multimedia research within the broader multimedia community.

2 Target Audience & Promotion

We plan to promote the UAV workshop in order to increase audience awareness and interest, targeting researchers, academics, industry data scientists and engineers, as well as other parties interested in the latest developments and advances in the field. To achieve this, we will take several measures: **1).** Use social media platforms, such as X and Facebook, to promote the workshop topic and event. We will create an event page on Facebook and invite people to attend, as well as share updates about the workshop on X. **2).** Create a website for the UAV workshop that provides detailed information about the agenda, speakers, and registration. We will share the website link on our social media platforms to make it accessible to a wider audience. By utilizing social media, we aim to increase the visibility of our UAV workshop and attract a diverse range of attendees from various fields who are interested in learning about the latest research and opportunities in UAVs.

3 Topics and Themes

Topics covered in this workshop (but not limited to) is as follows:

- Video-based UAV Navigation
  - Satellite-guided & Ground-guided Navigation
  - Path Planning and Obstacle Avoidance
  - Visual SLAM (Simultaneous Localization and Mapping)
  - Sensor Fusion and Reinforcement Learning for Navigation
- UAV Swarm Coordination
  - Multiple Platform Collaboration
  - Multi-agent Cooperation and Communication
  - Decentralized Control and Optimization
  - Distributed Perception and Mapping
- UAV-based Object Detection and Tracking
  - Aerial-view Object Detection, Tracking and Re-identification
  - Aerial-view Action Recognition
- UAV-based Sensing and Mapping
  - 3D Mapping and Reconstruction
  - Remote Sensing and Image Analysis
  - Disaster Response and Relief
- UAV-based Delivery and Transportation
  - Package Delivery and Logistics
  - Safety and Regulations for UAV-based Transportation

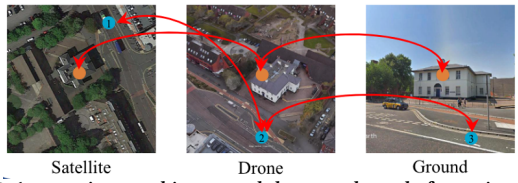

Figure 2: A cross-view matching example between three platforms, *i.e.*, satellite, drone and ground. The figure is credited by LPN [33].

4 Activities and Invited Keynotes

We plan to hold a hybrid format of workshop, *i.e.*, both on-site and online. For the on-site one at least two organizers will attend in person to host the workshop. The workshop will include two major activities, the invited keynotes, and the paper presentations. We will invite keynote presentations for a half-day workshop, following by accepted workshop presentations. The speakers are the experts on the relevant community from different organizations globally. The schedule of the workshop activities are listed in Table 1.

5 Challenge and Paper Submission

In previous editions of the workshop, the challenge focused on coarse retrieval using the University-1652 dataset (see Figure 2). This year, we introduce a new dataset and task targeting fine-grained localization. Specifically, we present PairUAV [1], which shifts the focus from cross-view retrieval to terminal-range pose alignment.

5.1 Challenge Dataset

Accurate UAV navigation during the final approach phase, often referred to as "last-meter" navigation, is critical for applications such as autonomous landing, delivery, and search-and-rescue, where GNSS may be unreliable and monocular vision suffers from scale ambiguity. To address this challenge, we introduce PairUAV, a fine-grained localization dataset derived from University-1652, providing explicit supervision for relative pose estimation. Unlike existing UAV localization benchmarks, PairUAV formulates navigation as a target-driven task, where a UAV agent must align its pose with a specified target image defining the goal location and orientation.

[1] Currently under review

Table 1: Schedule of workshop activities (half-day).

Topic	Duration	Speaker	Organization
• Morning Schedule			
Opening of the workshop	5 min	Fabian Deuser	University of the Bundeswehr Munich
Keynote 1: Big Geospatial Data Management	45 min	Martin Werner	Technical University of Munich
Keynote 2: Vision-Language Models for UAV Perception	45 min	Hongxia Yang	The Hong Kong Polytechnic University
Coffee Break	15 min		
• Paper Presentations			
Paper 1 Presentation	20 min	TBD	
Paper 2 Presentation	20 min	TBD	
Paper 3 Presentation	20 min	TBD	

Table 2: Here, we show the result on PairUAV. Precision visual navigation in the final approach is quite challenging.

Methods	PairUAV			
Methods	MAE_R (m)↓	MAE_H (deg)↓	AVG↓	SR (%)↑
Baseline [45]	44.86	90.33	67.59	15.19
DreamNav	**27.04**	**24.13**	**25.59**	**24.41**

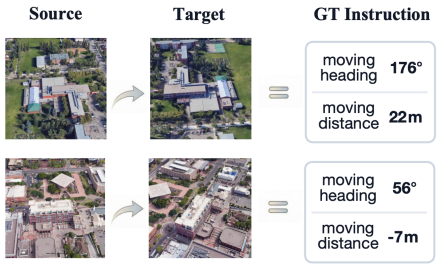

Figure 3: Samples for precision visual navigation in Our Challenge.

PairUAV is the first dataset to support target-driven UAV navigation at this scale, containing over 4.8 million image pairs across 72 scenes. The dataset will be publicly released with a leaderboard.

An example of the task is shown in Figure 3. Given a source image and a target image, participants must predict the relative translation and heading required to align the UAV with the target pose. Table 2 reports baseline results. Performance is evaluated using mean absolute range error (MAE_R, m), mean absolute heading error (MAE_H, deg), their average (AVG), and the success rate (SR, %), defined as predictions within a 10 m tolerance.

5.2 Submission Types

In this workshop, we welcome two types of submissions, all of which should relate to the topics and themes as listed in Section 3: **(1).** Challenge papers (up to 4 pages in length, plus unlimited pages for references): original solution to the Challenge data, PairUAV, in terms of effectiveness and efficiency. **(2).** Original papers (up to 4 pages in length, plus unlimited pages for references): original ideas, perspectives, research vision, and open challenges in the area of evaluation approaches for UAVs in Multimedia; Page limits include diagrams and appendices.

6 Organizer Information

Fabian Deuser (https://skyy93.github.io/) is a Doctoral researcher at the University of the Bundeswehr Munich, supervised by Prof. Norbert Oswald (UniBW Munich) and Prof. Martin Werner (TU Munich). His research focuses on retrieval with deep learning and contrastive learning. He is a two-time winner of the UAVM Workshop at ACM MM'23 & 24, and co-organized UAVM'25.

Yaxuan Li (https://yaxuanli-cn.github.io/) is currently a Research Assistant at the University of Macau, supervised by Prof. Zhedong Zheng. His research interests primarily lie in UAV navigation, embodied AI, and few-shot learning. To date, he has published over 5 papers in top-tier international journals and conferences. He has been honored with awards such as the China National Scholarship and the Academic Excellence Scholarship.

Tingyu Wang (https://scholar.google.com/citations?user=wv3H-F4AAAAJ) is an assistant professor at the School of Communication Engineering, Hangzhou Dianzi University, China. He received his Ph.D. degree from the Lab of Intelligent Information Processing, Hangzhou Dianzi University, in 2023, supervised by Prof. Chenggang Yan. He has co-organized 3 workshops at ACM MM (23, 24 & 25), focusing on scene understanding from UAVs' perspectives.

Yujiao Shi (https://yujiaoshi.github.io/) is an Assistant Professor at ShanghaiTech University. She was previously a research fellow and PhD student at the Australian National University, supervised by Prof. Hongdong Li. Her research interests include multi-modal retrieval, registration and translation, 3D vision, and self-supervised learning. She was a tutorial speaker on aerial image-based localization at CVPR 2023.

Anna Bößendörfer is a Doctoral researcher at the University of the Bundeswehr Munich, supervised by Professor Harald Görl. Her research interests include drone detection and generative AI.

Shaofei Huang (https://spyflying.github.io/) is a Postdoctoral Researcher at the University of Macau. She obtained her Ph.D. from the Institute of Information Engineering, Chinese Academy of Sciences, under the supervision of Prof. Si Liu and Prof. Jizhong Han. Her research interests include multimedia analysis and autonomous driving. She has published 20+ papers in top conferences and journals. She was awarded the China National Scholarship and Excellent Prize of President Scholarship of CAS.

Xiao Pan (https://scholar.google.com/citations?user=5Rh3yn4AAAAJ&hl=en) is currently an Assistant Professor in the College of Electronics and Information Engineering at Shenzhen University. His research focuses on 3D vision and vision generation. He received his Ph.D. in Artificial Intelligence from Zhejiang University in 2025, under the supervision of Professor Yi Yang. He has gained extensive industry experience and international exposure, having served as a research intern at Alibaba DAMO Academy and Tongyi Lab.

Zhedong Zheng (https://zdzheng.xyz) is an assistant professor with the University of Macau. He was a research fellow at School of Computing, National University of Singapore. He received the Ph.D. degree from the University of Technology Sydney, Australia, in 2021 and the B.S. degree from Fudan University, China, in 2016. He received the IEEE Circuits and Systems Society Outstanding

Young Author Award of 2021. He has organized a special session on reliable retrieval at ICME'22, two workshops at ACM MM'23 and one workshop at ACM ICMR'24. Besides, he is invited as a keynote speaker at CVPR'20, CVPR'21, a tutorial speaker at ACM MM'22. He also serves as an area chair for ACM MM'24,25 and ICASSP'25. **Roger Zimmermann** (https://www.comp.nus.edu.sg/cs/people/rogerz/) is a Full Professor at the School of Computing at the National University of Singapore (NUS). He is a Co-PI with the Grab-NUS AI Lab. From 2011 to 2021 he was a Deputy Director with the NUS Smart Systems Institute (SSI) and from 2010 to 2016 he co-directed the Centre of Social Media Innovations for Communities (COSMIC), a research institute funded by the National Research Foundation (NRF) of Singapore. Prior to joining NUS he held the positions of Research Area Director with the Integrated Media Systems Center (IMSC) and Research Assistant Professor at the University of Southern California (USC).

References

[1] Quan Chen, Tingyu Wang, Zihao Yang, Haoran Li, Rongfeng Lu, Yaoqi Sun, Bolun Zheng, and Chenggang Yan. 2024. SDPL: Shifting-Dense Partition Learning for UAV-View Geo-Localization. *IEEE Transactions on Circuits and Systems for Video Technology* 34, 11 (2024), 11810–11824. doi:10.1109/TCSVT.2024.3424196

[2] Meng Chu, Zhedong Zheng, Wei Ji, Tingyu Wang, and Tat-Seng Chua. 2024. Towards natural language-guided drones: GeoText-1652 benchmark with spatial relation matching. In *European Conference on Computer Vision*. Springer, 213–231.

[3] Yuanjie Dang, Chong Huang, Peng Chen, Ronghua Liang, Xin Yang, and Kwang-Ting Cheng. 2022. Path-Analysis-Based Reinforcement Learning Algorithm for Imitation Filming. *TMM* (2022).

[4] Fabian Deuser, Konrad Habel, and Norbert Oswald. 2023. Sample4geo: Hard negative sampling for cross-view geo-localisation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 16847–16856.

[5] Didula Dissanayaka, Thumeera R Wanasinghe, Oscar De Silva, Awantha Jayasiri, and George KI Mann. 2023. Review of Navigation Methods for UAV-Based Parcel Delivery. *TASE* (2023).

[6] Tzu-Yi Fan, Fangqi Liu, Jia-Wei Fang, Nalini Venkatasubramanian, and Cheng-Hsin Hsu. 2022. Enhancing situational awareness with adaptive firefighting drones: leveraging diverse media types and classifiers. In *ACM Multimedia*.

[7] Jianwu Fang, Lei-Lei Li, Kuan Yang, Zhedong Zheng, Jianru Xue, and Tat-Seng Chua. 2022. Cognitive Accident Prediction in Driving Scenes: A Multimodality Benchmark. *arXiv:2212.09381* (2022).

[8] Payton Goodrich, Omar Betancourt, Ana Claudia Arias, and Tarek Zohdi. 2023. Placement and drone flight path mapping of agricultural soil sensors using machine learning. *Computers and Electronics in Agriculture* (2023).

[9] Pu Jin, Lichao Mou, Gui-Song Xia, and Xiao Xiang Zhu. 2022. Anomaly Detection in Aerial Videos With Transformers. *TGRS* (2022).

[10] Haoran Li, Tingyu Wang, Quan Chen, Qiang Zhao, Shaowei Jiang, Chenggang Yan, and Bolun Zheng. 2024. Aerial-view geo-localization based on multi-layer local pattern cross-attention network. *Applied Intelligence* 54, 21 (2024), 11034–11053.

[11] Tianjiao Li, Jun Liu, Wei Zhang, Yun Ni, Wenqian Wang, and Zhiheng Li. 2021. Uav-human: A large benchmark for human behavior understanding with unmanned aerial vehicles. In *CVPR*.

[12] Jinliang Lin, Zhedong Zheng, Zhun Zhong, Zhiming Luo, Shaozi Li, Yi Yang, and Nicu Sebe. 2022. Joint Representation Learning and Keypoint Detection for Cross-view Geo-localization. *TIP* (2022).

[13] Liu Liu and Hongdong Li. 2019. Lending orientation to neural networks for cross-view geo-localization. In *CVPR*.

[14] Zhihao Liu, Yuanyuan Shang, Timing Li, Guanlin Chen, Yu Wang, Qinghua Hu, and Pengfei Zhu. 2023. Robust Multi-Drone Multi-Target Tracking to Resolve Target Occlusion: A Benchmark. *TMM* (2023).

[15] Ioannis Pitas and Ioannis Mademlis. 2022. Autonomous UAV Cinematography. In *ACM Multimedia*.

[16] Redmond Ramin Shamshiri, Ibrahim A Hameed, Siva K Balasundram, Desa Ahmad, Cornelia Weltzien, and Muhammad Yamin. 2018. Fundamental research on unmanned aerial vehicles to support precision agriculture in oil palm plantations. *Agricultural Robots-Fundamentals and Application* (2018).

[17] Mingwen Shao, Chao Wang, Wangmeng Zuo, and Deyu Meng. 2022. Efficient pyramidal GAN for versatile missing data reconstruction in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–14.

[18] Yujiao Shi, Dylan John Campbell, Xin Yu, and Hongdong Li. 2022. Geometry-guided street-view panorama synthesis from satellite imagery. *TPAMI* (2022).

[19] Yujiao Shi and Hongdong Li. 2022. Beyond Cross-view Image Retrieval: Highly Accurate Vehicle Localization Using Satellite Image. In *CVPR*.

[20] Yujiao Shi, Liu Liu, Xin Yu, and Hongdong Li. 2019. Spatial-Aware Feature Aggregation for Image based Cross-View Geo-Localization. In *NeurIPS*.

[21] Yujiao Shi, Fei Wu, Akhil Perincherry, Ankit Vora, and Hongdong Li. 2023. Boosting 3-DoF Ground-to-Satellite Camera Localization Accuracy via Geometry-Guided Cross-View Transformer. In *ICCV*.

[22] Yujiao Shi, Xin Yu, Dylan Campbell, and Hongdong Li. 2020. Where am I looking at? joint location and orientation estimation by cross-view matching. In *CVPR*.

[23] Yujiao Shi, Xin Yu, Liu Liu, Dylan Campbell, Piotr Koniusz, and Hongdong Li. 2022. Accurate 3-DoF Camera Geo-Localization via Ground-to-Satellite Image Matching. *TPAMI* (2022).

[24] Yujiao Shi, Xin Yu, Liu Liu, Tong Zhang, and Hongdong Li. 2020. Optimal feature transport for cross-view image geo-localization. In *AAAI*.

[25] Yujiao Shi, Xin Yu, Shan Wang, and Hongdong Li. 2022. CVLNet: Cross-View Feature Correspondence Learning for Video-based Camera Localization. In *ACCV*.

[26] Zhenbo Song, Xianghui Ze, Jianfeng Lu, and Yujiao Shi. 2023. Learning Dense Flow Field for Highly-accurate Cross-view Camera Localization. *arXiv:2309.15556* (2023).

[27] Francesco Betti Sorbelli, Federico Corò, Lorenzo Palazzetti, Cristina M Pinotti, and Giulio Rigoni. 2023. How the Wind Can Be Leveraged for Saving Energy in a Truck-Drone Delivery System. *TITS* (2023).

[28] Chao Wang, Mingwen Shao, Deyu Meng, and Wangmeng Zuo. 2022. Dual-pyramidal image inpainting with dynamic normalization. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 9 (2022), 5975–5988.

[29] Chao Wang, Zhedong Zheng, Ruijie Quan, Yifan Sun, and Yi Yang. 2023. Context-Aware Pretraining for Efficient Blind Image Decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18186–18195.

[30] Tingyu Wang, Yujiao Shi, Fabian Deuser, Shaofei Huang, Guosheng Hu, Si Liu, Zhedong Zheng, and Roger Zimmermann. 2025. The 3rd workshop on uavs in multimedia: Capturing the world from a new perspective. In *Proceedings of the 3rd International Workshop on UAVs in Multimedia: Capturing the World from a New Perspective*. 1–3.

[31] Tingyu Wang, Zihao Yang, Quan Chen, Yaoqi Sun, and Chenggang Yan. 2024. Rethinking Pooling for Multi-Granularity Features in Aerial-View Geo-Localization. *IEEE Signal Processing Letters* 31 (2024), 3005–3009. doi:10.1109/LSP.2024.3484330

[32] Tingyu Wang, Zhedong Zheng, Yaoqi Sun, Tat-Seng Chua, Yi Yang, and Chenggang Yan. 2024. Multiple-environment Self-adaptive Network for Aerial-view Geo-localization. *Pattern Recognition* (2024).

[33] Tingyu Wang, Zhedong Zheng, Chenggang Yan, Jiyong Zhang, Yaoqi Sun, Bolun Zheng, and Yi Yang. 2021. Each part matters: Local patterns facilitate cross-view geo-localization. *TCSVT* (2021).

[34] Tingyu Wang, Zhedong Zheng, Zunjie Zhu, Yaoqi Sun, Chenggang Yan, and Yi Yang. 2024. Learning Cross-View Geo-Localization Embeddings via Dynamic Weighted Decorrelation Regularization. *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024), 1–12. doi:10.1109/TGRS.2024.3491757

[35] Longyin Wen, Dawei Du, Pengfei Zhu, Qinghua Hu, Qilong Wang, Liefeng Bo, and Siwei Lyu. 2021. Detection, tracking, and counting meets drones in crowds: A benchmark. In *CVPR*.

[36] Cai YuanQiang, Dawei Du, Libo Zhang, Longyin Wen, Weiqiang Wang, Yanjun Wu, and Siwei Lyu. 2020. Guided attention network for object detection and counting on drones. In *ACM Multimedia*.

[37] Haotian Zhang, Gaoang Wang, Zhichao Lei, and Jenq-Neng Hwang. 2019. Eye in the sky: Drone-based object tracking and 3d localization. In *ACM Multimedia*.

[38] LiangLiang Zhao and MinLing Zhu. 2023. MS-YOLOv7: YOLOv7 Based on Multi-Scale for Object Detection on UAV Aerial Photography. *Drones* (2023).

[39] Ou Zheng, Mohamed Abdel-Aty, Lishengsa Yue, Amr Abdelraouf, Zijin Wang, and Nada Mahmoud. 2022. CitySim: A Drone-Based Vehicle Trajectory Dataset for Safety Oriented Research and Digital Twins. *arXiv:2208.11036* (2022).

[40] Zhedong Zheng, Yujiao Shi, Tingyu Wang, Chen Chen, Pengfei Zhu, and Richard Hartley. 2024. The 2nd Workshop on UAVs in Multimedia: Capturing the World from a New Perspective. In *Proceedings of the 2nd Workshop on UAVs in Multimedia: Capturing the World from a New Perspective*. 1–3. doi:10.1145/3689095.3695932

[41] Zhedong Zheng, Yujiao Shi, Tingyu Wang, Jun Liu, Jianwu Fang, Yunchao Wei, and Tat-seng Chua. 2023. UAVM'23: 2023 Workshop on UAVs in Multimedia: Capturing the World from a New Perspective. In *Proceedings of the 31st ACM International Conference on Multimedia*. 9715–9717.

[42] Zhedong Zheng, Yunchao Wei, and Yi Yang. 2020. University-1652: A multi-view multi-source benchmark for drone-based geo-localization. In *ACM Multimedia*.

[43] Zhedong Zheng and Yi Yang. 2019. Unsupervised scene adaptation with memory regularization in vivo. *IJCAI* (2019).

[44] Zhedong Zheng and Yi Yang. 2022. Adaptive boosting for domain adaptation: Toward robust predictions in scene segmentation. *IEEE Transactions on Image Processing* 31 (2022), 5371–5382.

[45] Yuke Zhu, Roozbeh Mottaghi, Eric Kolve, Joseph J. Lim, Abhinav Gupta, Li Fei-Fei, and Ali Farhadi. 2017. Target-driven visual navigation in indoor scenes using deep reinforcement learning. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 3357–3364. doi:10.1109/ICRA.2017.7989381