

Boundless: Generating Photorealistic Synthetic Data for Object Detection in Urban Streetscapes

Mehmet Kerem Turkcan
Columbia University
New York, NY

mkt2126@columbia.edu

Javad Ghaderi
Columbia University
New York, NY

jg3465@columbia.edu

Yuyang Li
Columbia University
New York, NY

yl5339@columbia.edu

Gil Zussman
Columbia University
New York, NY

gil.zussman@columbia.edu

Chengbo Zang
Columbia University
New York, NY

cz2678@columbia.edu

Zoran Kostic
Columbia University
New York, NY

zk2172@columbia.edu

Abstract

We introduce Boundless, a photo-realistic synthetic data generation system for enabling highly accurate object detection in dense urban streetscapes. Boundless can replace massive real-world data collection and manual ground-truth object annotation (labeling) with an automated and configurable process. Boundless is based on the Unreal Engine 5 (UE5) City Sample project with improvements enabling accurate collection of 3D bounding boxes across different lighting and scene variability conditions.

We evaluate the performance of object detection models trained on the dataset generated by Boundless when used for inference on a real-world dataset acquired from medium-altitude cameras. We compare the performance of the Boundless-trained model against the CARLA-trained model and observe an improvement of 7.8 mAP. The results we achieved support the premise that synthetic data generation is a credible methodology for training/fine-tuning scalable object detection models for urban scenes.

1. Introduction

Pedestrian safety and traffic management in bustling cities can be enhanced by using video-based AI systems for real-time monitoring and interaction with street objects [42]. Autonomous vehicles will face challenges due to irregular street layouts, numerous cohabitants, and unpredictable pedestrian behavior [3, 41]. This calls for the use of infrastructure-mounted cameras and sensors to gather real-time video data at locations like traffic intersections, where object detection, tracking, trajectory prediction, and high-level reasoning can be performed on edge servers in real time [12].

To scale up video monitoring systems in large cities, deep learning (DL) models need to be trained and fine-tuned for hundreds of intersections, each with multiple cameras at varying micro-locations. Successful training of supervised DL models depends highly on the availability of ground-truth annotated (labeled) data. For traffic intersections, the annotation applies to vehicles, bicycles, pedestrians, and other moving objects, as well as immovable traffic furniture.

Real-world image collection is complicated by uncontrollable environmental conditions such as variations in lighting, weather, and the unpredictable behavior of transient objects like pedestrians and vehicles. “Manual” ground-truth annotation of street objects from arbitrary camera angles requires a large time commitment and monetary resources. Data collection in real-world scenarios additionally faces legal issues due to privacy violations. Consequently, existing urban datasets often comprise isolated scenes rather than exhaustive city maps, lacking in the ability to capture the multifaceted nature of cityscapes. These datasets offer limited perspectives, predominantly at eye-level street or high-altitude aerial views, which only partially represent the possible views that one might acquire in the urban deployment of cameras.

In this work, we investigate the use of synthetic data/image generators that can automatically create ground-truth annotations for training of object detection models, and therefore avoid the complexity and cost of manual ground-truth annotations.

We focus on the generation of synthetic image datasets using Unreal Engine 5¹ to address the shortcomings of existing object detection datasets in urban environments. We incorporate realistic variable lighting, apply post-

¹<https://www.unrealengine.com/>



(a) Fog



(b) Snow



(c) Rain



(d) Night

Figure 1. Boundless enables dynamic weather and environment changes for different situations, with changing time-of-day, weather conditions and background elements for every camera. (a-d) Different weather conditions.

processing effects to simulate weather conditions and improve render quality, and make adjustments for the level of detail of graphical assets and accurate capture of bounding boxes. The resulting simulator, which we call “Boundless”, can simulate diverse conditions, control environmental factors, and automatically generate high-quality ground-truth annotations. We investigate the suitability of medium-altitude data generated using Boundless for improving object detection performance in a setting for which available real-world training data is scarce.

2. Related Works

Object Detection in Urban Environments. A large number of datasets focus on low-altitude vehicle and pedestrian detection [13–15, 24, 30, 34, 35, 40]. Meanwhile, many other datasets focus on high-altitude aerial environments, where small object detection becomes an important challenge [5, 16, 22, 33, 37–39, 43]. However, there remains a gap for public mixed-perspective datasets that can adapt to a multitude of different deployment conditions.

Real-Time Object Detection. Many models have been proposed for object detection. For real-time applications, in recent years single-stage detectors like SSD and YOLO

models or transformer-based DETR architecture variants have achieved significant results for real-time detection [2, 4, 11, 19, 20, 26–28, 36]. In this work, we use the state-of-the-art YOLOv8x model for experiments. Rather than model development, our focus is the quality of data used for training.

Synthetic Data Generation. Realistic 3D simulators have been used extensively for various urban computer vision problems. The SYNTHIA dataset provides a collection of images from a simulated city along with pixel-level semantic annotations, to support semantic segmentation and scene understanding tasks [31]. CARLA, developed in Unreal Engine 4, is an open-source autonomous driving simulator offering extensive resources, including environments and open digital assets explicitly designed for development, testing, and validating self-driving systems [7]. CARLA-generated imagery has been used for object detection and segmentation [10, 21, 23]. Extensive work has been conducted on GTA V, using frames collected from this video game for training autonomous driving agents [29]. In addition to urban traffic simulators like CARLA, Unreal Engine itself has been considered as a rendering engine for computer vision approaches [1, 6, 25]. Recently, Matrix-City adapted the photo-realistic City Sample project as a

benchmark for training neural rendering models by designing a plugin for saving frames [17], focusing on pedestrian and vehicle-free environments and neural rendering applications. The authors in [9] used the City Sample project to detect pedestrians and hand-annotated them for this purpose.

In contrast to previous work, here we focus on enhancing the available City Sample project to enable accurate and automated data collection for training performant models for urban deep learning applications. We find that extensive customization is required to enable accurate bounding box annotation collection in Unreal Engine due to a variety of challenges. We further improve the project with live lighting changes and weather and release multiple datasets for medium-altitude object detection, a problem of interest in urban metropolises [8].

3. Boundless Simulator Design

The freely available City Sample project provides an environment for North American cityscapes, including vehicles, pedestrians, and traffic lights². To facilitate realistic data collection, we created Boundless by making technical changes to the City Sample project to enable realistic data collection for use in AI applications. We show examples of scenes created with Boundless under different weather conditions, including bounding boxes, in Figure 1 to show the capabilities of the simulator. We detail the technical improvements below.

Lighting. We allow lighting conditions to be changed dynamically before each new frame collection, thus allowing the scene to change substantially in a single capture session. We implement four new weather conditions corresponding to rain, snow, dust and heat waves. Implemented using decals projected onto the map, the rain and snow effects allow for a more realistic alternative to image-level augmentations [32]. We add particle effects for all these weather conditions. We note that the City Sample comes with a default night-time implementation; however, the stylized and overly dark night weather does not correspond to real-world night-time conditions.

Anti-Aliasing. The default temporal anti-aliasing approach in UE5 produces blurred frames. We replace the approach with MSAA and change the camera settings to output 3840x2160 resolution frames.

Level of Detail. The City Sample project includes three levels of detail for pedestrian and vehicle actors. For the medium and low levels of detail, the substituted meshes have insufficient resolution and detail quality for facilitating real-world applications. We change the available levels of detail for each vehicle and pedestrian agent, enabling

the simulator to capture distant object bounding boxes accurately from medium-altitude and street-level scenes.

Updated Bounding Boxes. Due to the design of pedestrian and vehicle actors in the City Sample project which results in large or missing collision boundaries for different 3D meshes, significant changes are needed to correctly capture bounding boxes of objects. We re-compute the level of detail, visibility, and occlusion properties of individual objects in a scene before capturing a frame to make sure annotations of all visible objects are obtained.

Export Options. Boundless exports 3D bounding boxes in the KITTI and 2D bounding boxes in the YOLO format.

4. Datasets

We give an overview of the different datasets we introduce for our experiments in Figure 2. We generated two synthetic datasets using Boundless:

- **Medium-Altitude City Sample Training Set.** The City Sample comes with default city maps. We use Boundless to collect an 8,000-frame dataset from a synthetic intersection from the City Sample project with a static camera angle. All bounding boxes are generated automatically by the simulator. Lighting conditions are changed throughout the training set in between every frame. A frame is saved from the simulation every 3 seconds in simulation time. For comparison purposes, we follow the same approach to create a corresponding dataset consisting of 22,000 frames using the CARLA simulator.

- **Medium-Altitude Digital Twin Training Set.** We create a realistic 3D digital twin of a real-world intersection within Boundless. We collect 8,700 frames from this highly accurate scene, replicating the camera angles of the medium-altitude real-world validation dataset.

In addition to our synthetically generated datasets, we use the following two real-world image datasets:

- **Medium-Altitude Real World Validation Set.** We create a real-world image dataset collected from a major North American metropolis for one intersection. This validation set consists of 3,084 frames collected on different days with a variety of weather and time-of-day conditions. The dataset contains 12,380 vehicles and 15,225 pedestrian bounding boxes.

- **VisDrone Dataset.** The VisDrone dataset [43] contains 7,019 images captured from various perspectives, including top-down and ground-level views, with varying camera angles. Although the dataset originally contains multiple object classes, we adapt it to a two-class object detection problem. We use the 561-image validation split for reporting our results.

²<https://www.unrealengine.com/marketplace/en-US/product/city-sample>

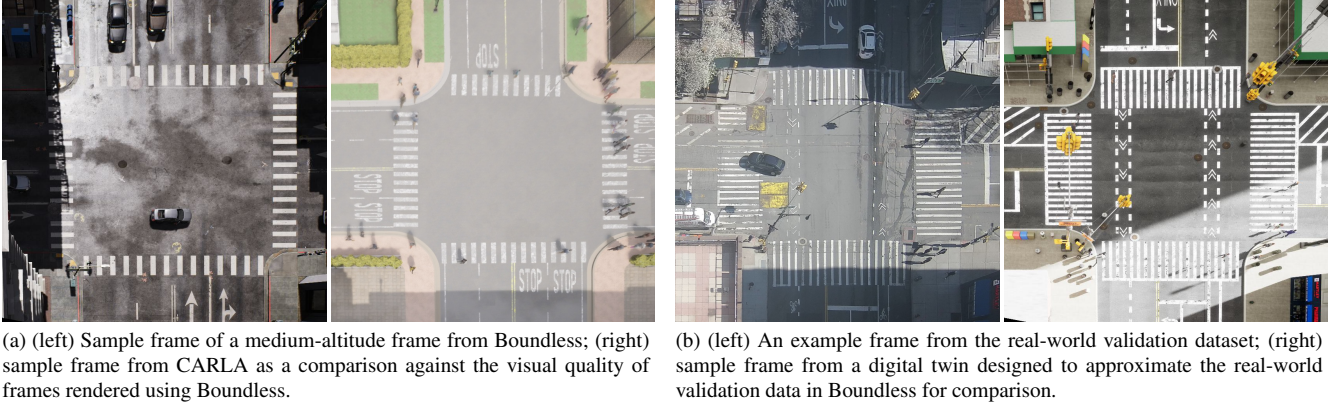


Figure 2. Examples of frames used for the medium-altitude object detection benchmark.

Table 1. Per-class average precision and mean average precision of YOLOv8x models trained on VisDrone, CARLA and Boundless datasets, evaluated on the real-world validation set.

Training Dataset	Pedestrian AP@0.5	Vehicle AP@0.5	mAP@0.5
VisDrone	24.8	86.9	55.8
CARLA	14.9	75.4	45.1
Boundless	24.0	81.8	52.9
Boundless + Digital Twin	47.5	85.7	66.6

5. Experiments

Medium-Altitude Object Detection. We used a medium-altitude object detection task to demonstrate the capabilities of Boundless, where the amount of data available is scarce compared to other types of urban data. In this task, we seek to detect pedestrians and vehicles from a static camera at $\sim 40\text{m}$ height. We wanted to explore how well a model performs on this task when trained on different datasets. To do so, we fine-tuned a COCO-pretrained YOLOv8x model [11, 18] on three different datasets: (i) VisDrone, (ii) CARLA, and (iii) Boundless. We evaluated the models on a custom dataset collected from a real-world traffic intersection. All models were trained with SGD for 10 epochs at a learning rate of $1\text{e-}3$. We show the results of this comparison in Table 1. Use of Boundless-generated data using the City Sample map yields a model that greatly outperforms the CARLA-generated data on the real-world validation set, and further exceeds VisDrone performance in this dataset.

Motivated by the superior performance obtained using Boundless, we further implemented a digital twin of the real-world intersection from the real-world validation set in the form of a map in Unreal Engine. We collected more data using Boundless from this map to see how much the results could be improved. Using a 3D model as a map, we collect an additional 8,700 frames. Repeating the experiment by adding these additional frames further improves the mAP score significantly, yielding a higher mAP than the use of VisDrone.

6. Ethical Considerations

The real-world dataset collected is deliberately installed at an altitude that makes it impossible to discern the pedestrian faces or to read car license plates. The academic institution at whose facilities the camera is installed provided an IRB waiver for sharing this “high elevation data” with the general public since the data inherently preserves privacy.

7. Conclusion

We created and described the Boundless simulation platform, and benchmarked it in a real-world medium-altitude object detection task that demonstrates challenges with the deployment of object detection models to urban streetscapes. Our best-performing model, using Boundless and incorporating a digital twin of the real-world testbed, achieved an mAP of 66.6, demonstrating the ability of the simulator to create imagery with sufficient realism to be deployed in real-world scenarios. With Boundless, we seek to support research on urban object detection for metropolises, where data collection, ground-truth annotation/labeling, and model training face technical and legal challenges. By releasing the simulator along with collected datasets, we aim to facilitate future research and applications in urban computer vision problems.

Dataset and Code Availability

Datasets and code for the experiments in this study are available at <https://github.com/zk2172-columbia/boundless>.

Acknowledgements

This work was supported in part by NSF grants CNS-1827923 and EEC-2133516, NSF grant CNS-2038984 and corresponding support from the Federal Highway Administration (FHA), NSF grant CNS-2148128 and by funds from federal agency and industry partners as specified in the Resilient & Intelligent NextG Systems (RINGS) program, and ARO grant W911NF2210031.

References

- [1] Dhruv Agarwal, Taci Kucukpinar, Joshua Fraser, Jeffrey Kerley, Andrew R Buck, Derek T Anderson, and Kannapan Palaniappan. Simulating city-scale aerial data collection using unreal engine. In *2023 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pages 1–9. IEEE, 2023. 2
- [2] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint:2004.10934*, 2020. 2
- [3] Mark Campbell, Magnus Egerstedt, Jonathan P How, and Richard M Murray. Autonomous driving in urban environments: approaches, lessons and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928):4649–4672, 2010. 1
- [4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 2
- [5] Guang Chen, Haitao Wang, Kai Chen, Zhijun Li, Zida Song, Yinlong Liu, Wenkai Chen, and Alois Knoll. A survey of the four pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal. *IEEE Transactions on systems, man, and cybernetics: systems*, 52(2):936–953, 2020. 2
- [6] Alexandru Damian, Claudiu Filip, Anamaria Nistor, Irina Petrariu, Cătălin Mariuc, and Valentin Stratan. Experimental results on synthetic data generation in unreal engine 5 for real-world object detection. In *2023 17th International Conference on Engineering of Modern Electric Systems (EMES)*, pages 1–4. IEEE, 2023. 2
- [7] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017. 2
- [8] Zhuoxu Duan, Zhengye Yang, Richard Samoilenko, Dwiref Snehal Oza, Ashvin Jagadeesan, Mingfei Sun, Hongzhe Ye, Zihao Xiong, Gil Zussman, and Zoran Kostic. Smart city traffic intersection: Impact of video quality and scene complexity on precision and inference. In *Proc. IEEE Smart City'21*, 2021. 3
- [9] Yue Hu, Gourav Datta, Kira Beerel, and Peter Beerel. Let's roll: Synthetic dataset analysis for pedestrian detection across different shutter types. *arXiv preprint arXiv:2309.08136*, 2023. 3
- [10] Jaesung Jang, Hyeongyu Lee, and Jong-Chan Kim. Carfree: Hassle-free object detection dataset generation using carla autonomous driving simulator. *Applied Sciences*, 12(1):281, 2021. 2
- [11] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLOv8, 2023. 2, 4
- [12] Zoran Kostić, Alex Angus, Zhengye Yang, Zhuoxu Duan, Ivan Seskar, Gil Zussman, and Dipankar Raychaudhuri. Smart city intersections: Intelligence nodes for future metropolises. *Computer*, 55(12):74–85, 2022. 1
- [13] Vijay Gopal Kovvali, Vassili Alexiadis, and Lin Zhang PE. Video-based vehicle trajectory data collection. Technical report, 2007. 2
- [14] Robert Krajewski, Julian Bock, Laurent Kloeker, and Lutz Eckstein. The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In *2018 21st international conference on intelligent transportation systems (ITSC)*, pages 2118–2125. IEEE, 2018.
- [15] Robert Krajewski, Tobias Moers, Julian Bock, Lennart Vater, and Lutz Eckstein. The round dataset: A drone dataset of road user trajectories at roundabouts in germany. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6. IEEE, 2020. 2
- [16] Huaqing Lai, Liangyan Chen, Weihua Liu, Zi Yan, and Sheng Ye. STC-YOLO: small object detection network for traffic signs in complex environments. *Sensors*, 23(11):5307, 2023. 2
- [17] Yixuan Li, Lihan Jiang, Linning Xu, Yuanbo Xiangli, Zhenzhi Wang, Dahua Lin, and Bo Dai. Matrixcity: A large-scale city dataset for city-scale neural rendering and beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3205–3215, 2023. 3
- [18] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common objects in context. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 4
- [19] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. SSD: Single shot multibox detector. In *Proc. ECCV*, 2016. 2
- [20] Wenyu Lv, Shangliang Xu, Yian Zhao, Guanzhong Wang, Jinman Wei, Cheng Cui, Yuning Du, Qingqing Dang, and Yi Liu. Detrs beat yolos on real-time object detection, 2023. 2
- [21] Maria Lyssenko, Christoph Gladisch, Christian Heinzemann, Matthias Woehrle, and Rudolph Triebel. Instance segmentation in carla: methodology and analysis for pedestrian-oriented synthetic data generation in crowded scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 988–996, 2021. 2
- [22] Chengqi Lyu, Wenwei Zhang, Haian Huang, Yue Zhou, Yudong Wang, Yanyi Liu, Shilong Zhang, and Kai Chen. Rtmddet: An empirical study of designing real-time object detectors. *arXiv preprint arXiv:2212.07784*, 2022. 2
- [23] DR Niranjana, BC VinayKarthik, et al. Deep learning based object detection model for autonomous driving research using carla simulator. In *2021 2nd international conference on smart electronics and communication (ICOSEC)*, pages 1251–1258. IEEE, 2021. 2

- [24] Sangmin Oh, Anthony Hoogs, Amitha Perera, Naresh Cuntoor, Chia-Chih Chen, Jong Taek Lee, Saurajit Mukherjee, JK Aggarwal, Hyungtae Lee, Larry Davis, et al. A large-scale benchmark dataset for event recognition in surveillance video. In *CVPR 2011*, pages 3153–3160. IEEE, 2011. 2
- [25] Ingeborg Rasmussen, Sigurd Kvalsvik, Per-Arne Andersen, Teodor Nilsen Aune, and Daniel Hagen. Development of a novel object detection system based on synthetic data generated from unreal game engine. *Applied Sciences*, 12(17): 8534, 2022. 2
- [26] Joseph Redmon and Ali Farhadi. YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017. 2
- [27] Joseph Redmon and Ali Farhadi. YOLOv3: An incremental improvement. *arXiv preprint:1804.02767*, 2018.
- [28] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proc. CVPR*, 2016. 2
- [29] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*, pages 102–118. Springer, 2016. 2
- [30] Alexandre Robicquet, Amir Sadeghian, Alexandre Alahi, and Silvio Savarese. Learning social etiquette: Human trajectory prediction in crowded scenes. In *European Conference on Computer Vision (ECCV)*, page 5, 2016. 2
- [31] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3234–3243, 2016. 2
- [32] Ujjwal Saxena and Rohan Giriraj. Automold. *Road augmentation library at <https://github.com/UjjwalSaxena/Automold-Road-Augmentation-Library>*, 2018. 3
- [33] Jacob Shermeyer and Adam Van Etten. The effects of super-resolution on object detection performance in satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 2
- [34] Bing Shuai, Alessandro Bergamo, Uta Buechler, Andrew Berneshawi, Alyssa Boden, and Joe Tighe. Large scale real-world multi person tracking. In *European Conference on Computer Vision*. Springer, 2022. 2
- [35] Pei Sun, Henrik Kretschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020. 2
- [36] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022. 2
- [37] Jinwang Wang, Wen Yang, Haowen Guo, Ruixiang Zhang, and Gui-Song Xia. Tiny object detection in aerial images. In *2020 25th international conference on pattern recognition (ICPR)*, pages 3791–3798. IEEE, 2021. 2
- [38] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3974–3983, 2018.
- [39] Chang Xu, Jinwang Wang, Wen Yang, and Lei Yu. Dot distance for tiny object detection in aerial images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1192–1201, 2021. 2
- [40] Dongfang Yang, Linhui Li, Keith Redmill, and Ümit Özgüner. Top-view trajectories: A pedestrian dataset of vehicle-crowd interaction from controlled experiments and crowded campus. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 899–904. IEEE, 2019. 2
- [41] Ibrar Yaqoob, Latif U Khan, SM Ahsan Kazmi, Muhammad Imran, Nadra Guizani, and Choong Seon Hong. Autonomous driving cars in smart cities: Recent advances, requirements, and challenges. *IEEE Network*, 34(1):174–181, 2019. 1
- [42] Qingyang Zhang, Hui Sun, Xiaopei Wu, and Hong Zhong. Edge video analytics for public safety: A review. *Proceedings of the IEEE*, 107(8):1675–1696, 2019. 1
- [43] Pengfei Zhu, Longyin Wen, Dawei Du, Xiao Bian, Haibin Ling, Qinghua Hu, Qinqin Nie, Hao Cheng, Chenfeng Liu, Xiaoyu Liu, et al. Visdrone-det2018: The vision meets drone object detection in image challenge results. In *Proc. ECCV Workshops*, 2018. 2, 3