ITM Web of Conferences **54**, 02006 (2023) https://doi.org/10.1051/itmconf/20235402006
I3CS-2023

Simulated uav dataset for object detection

Avinash Kaur Sama[1,*], *Akashdeep* Sharma[2]
[1,2]University Institute of Engineering and Technology, Panjab University, Chandigarh, India

Abstract. Unmanned Aerial Vehicles (UAVs) have become increasingly popular for various applications, including object detection. Novel detector algorithms require large datasets to improve, as they are still evolving. Additionally, in countries with restrictive drone policies, simulated datasets can provide a cost-effective and efficient alternative to real-world datasets for researchers to develop and test their algorithms in a safe and controlled environment. To address this, we propose a simulated dataset for object detection through a Gazebo simulator that covers both indoor and outdoor environments. The dataset consists of 11,103 annotated frames with 27,412 annotations, of persons and cars as the objects of interest. This dataset can be used to evaluate detector proposals for object detection, providing a valuable resource for researchers in the field. The dataset is annotated using the Dark Label software, which is a popular tool for object annotation. Additionally, we assessed the dataset's performance using advanced object detection systems, with YOLOv3 achieving 86.9 mAP50-95, YOLOv3-tiny achieving 79.5 mAP50-95, YOLOv5 achieving 82.2 mAP50-95, YOLOv7 achieving 61.8 mAP50-95 and YOLOv8 achieving 87.8 mAP50-95. Overall, this simulated dataset is a valuable resource for researchers working in the field of object detection.

1 Introduction

UAVs have become crucial in various fields, such as military, agriculture, and disaster management. Object detection and localization is an essential step in many computer vision systems, including those used in UAVs, and involves identifying and localizing objects of interest in an image or video stream. Moreover, the development of object detection algorithms is a rapidly growing area of research, with new algorithms proposed regularly to enhance detection accuracy and efficiency. However, evaluating the performance of these algorithms necessitates a comprehensive and diverse dataset that accurately represents the various environments and objects encountered by UAVs in real-world situations[1].

In this study, we propose a simulated dataset utilizing a Gazebo simulation that includes both indoor and outdoor environments and encompasses two common object classes, persons, and cars. To annotate the dataset, we have used the Dark Label software and assessed its performance by employing advanced object detection techniques with various versions of You Only Look Once(YOLO) [2,3] which is a real-time object detection system.

* Corresponding author: kauravinash250@gmail.com

To create our dataset, we have used realistic 3D models of the objects of interest in the Gazebo simulation environment. Our experiments demonstrated that our dataset can serve as a valuable resource for researchers to evaluate the accuracy and efficiency of their proposed object detection algorithms in a more realistic and comprehensive environment.

We anticipate that our dataset will contribute to the development of new and improved algorithms for these essential computer vision tasks.

All YOLO versions have undergone extensive testing on diverse datasets and have demonstrated state-of-the-art performance in object detection tasks. The goal is to establish a benchmark for future researchers to compare and evaluate their object detection proposals.

1.1 Motivation

In recent years, the use of UAVs in India has gained popularity across various industries such as agriculture, surveillance, and infrastructure monitoring. However, the integration of UAVs into these applications relies heavily on the development of robust computer vision algorithms that enable them to perform tasks such as object detection, tracking, and classification. The Indian Drone Policy [5] mandates that all drones must comply with safety guidelines and be equipped with necessary safety features. Therefore, the use of simulated environments for creating datasets for UAV applications aligns with the policy's emphasis on safety and security.

This work introduces a new dataset for computer vision algorithms in UAV applications, created using a simulated environment. The motivation behind this dataset is to provide a more comprehensive and reliable resource for algorithm development and evaluation in UAV applications. Existing datasets often have limitations that hinder algorithm performance in real-world scenarios. By using a simulated environment, a wider range of scenarios can be captured at a negligible cost.

Existing datasets for computer vision algorithms in UAV applications have limitations such as a narrow range of scenarios and environments, and high costs associated with data collection using real UAVs. These limitations can hinder algorithm performance in real-world scenarios, where UAVs may encounter a wide range of obstacles, environments, and lighting conditions.

The use of simulated data can overcome these limitations and enable the development of more robust algorithms. Simulated data provides a controlled environment for algorithm development and evaluation, allowing researchers to test their algorithms under a wide range of scenarios and environments. This can ultimately lead to more reliable and effective algorithms for real-world scenarios.

Furthermore, the use of simulated data reduces safety risks and costs associated with conducting experiments using real UAVs. Novice researchers can experiment and develop algorithms without endangering human lives or damaging costly equipment.

The dataset we provide contains labeled data that is specifically designed for object detection in UAV applications and can be easily adapted for other computer vision tasks. Further, we have evaluated the performance of several state-of-the-art computer vision algorithms for object detection using our dataset. The results demonstrate the effectiveness of our dataset in improving algorithm performance in real-world scenarios.

Overall, the creation of a comprehensive dataset for computer vision algorithms in UAV applications using simulated data can significantly improve algorithm development and evaluation, leading to safer and more efficient integration of UAVs into various applications in India.

The main objectives of this paper include:

- Creation of a simulated annotated dataset for object detection in UAVs.
- Experimental evaluation of proposed dataset on different YOLO versions.

2 Related Works

Several papers are presented to discuss object detection and UAV dataset in computer vision. Over the past few years, there has been a growing interest in UAV datasets in computer vision research. A number of papers have been published on the topic, discussing various aspects of UAV datasets, such as their properties, tasks, and applications. Table 1 presents a comparison between the proposed dataset and the state-of-the-art (SOTA) datasets. The table provides an overview of the similarities and differences between the proposed dataset and the existing datasets that are considered to be the best in their respective fields.

Table 1. Comparison of Proposed Dataset with SOTA Datasets.

Dataset Name	No. of Video	Capturing Device Used	Tasks	Object of interest	Video Type	Indoor/ Outdoor	Dataset Expansi-on
KITTI [6]	5	moving car-mounted camera	3D D, T	8	Real	Outdoor	No
ImageNet Video Detection [7]	-	-	D	30	Real	-	No
UA-DETRAC [8]	100	fixed-mounted camera	T	4	Real	Outdoor	No
VisDrone [9]	263	drone-mounted camera	D, T	10	Real	Outdoor	No
UAVDT [10]	100	drone-mounted camera	D, T	2	Real	Outdoor	No
MOR-UAV[11]	30	drone-mounted camera	MOR	2	Real	Outdoor	No
Our proposed dataset	**10**	**drone-mounted camera**	**D**	**2**	**Simulated**	**Both**	**Yes**

Abbreviations: D-Detection, T-Tracking, MOR-Moving Object Recognition

One of the earliest works on UAV datasets was the KITTI dataset[6], which was released in 2013. It consists of 5 videos captured using a moving car-mounted camera. Since its release, KITTI has been widely used as a benchmark for stereo, optical flow, object detection, and other tasks. More recently, [4] introduced a simulation environment created with Unreal Engine that is intended to allow wildlife conservation researchers to safely test and enhance their algorithms.

Another notable UAV dataset is the UA-DETRAC dataset [8], which consists of 100 videos captured using a fixed-mounted camera from real-world traffic scenes. it also provides a benchmark for object detection and multiple object tracking (MOT).

In recent years, several UAV datasets have been released that are specifically designed for object detection and tracking using drone-mounted cameras. For example, the VisDrone dataset[9] consists of 263 drone-captured videos and includes tasks related to object detection and tracking in urban environments. Another is the UAVDT dataset[10] which includes drone-captured video data in real-world scenarios. Its focus on complex scenarios can provide useful insights and challenges for researchers working on object detection, single object tracking, and MOT.

Overall, these works demonstrate the importance of UAV datasets in computer vision research and highlight the need for developing more diverse and challenging datasets for evaluating different computer vision algorithms.

3 Methodology

To generate and evaluate our simulated UAV dataset, we have followed the following procedure:

3.1 Simulated Environment Creation

The Gazebo simulator, an open-source simulator, has been used to generate simulated environments. These environments include both indoor and outdoor environments, with varying levels of complexity and obstacles. The Hector quadrotor has been used in simulation, an already available robot operating system (ROS) package.

The quadrotor is a helicopter with four rotors and has six degrees of freedom (x, y, z, u, v, and w) where x, y, and z represent the position in space while u, v, and w(i.e. roll, pitch, yaw respectively) are the three Euler angles that represent the orientation in space of the quadrotor.

Table 2. Simulated Environment Specifications.

S. No.	Environment Name	Environment Type	Complexity	Total number of frames capture
1	Bookstore	Indoor	Complex	1979
2	Cafe	Indoor	Simple	1739
3	City	Outdoor	Complex	2039
4	Factory	Indoor	Simple	1348
5	Hospital	Indoor	Complex	1416
6	Neighborhood	Outdoor	Simple	1131
7	Parking Lot	Outdoor	Simple	776
8	Racetrack	Outdoor	Simple	2171
9	Small House	Indoor	Complex	2650
10	Warehouse	Indoor	Complex	2485

The video dataset captured by the quadrotor has been described in detail in Table 2, which includes information on the number of videos captured, their length, complexity, and other relevant details. In the context of the table, complexity is related to the intricate characteristics of the environments. The quadrotor's behavior has been simulated in different environments to test its ability to detect objects. The frames captured by the camera mounted on the quadrotor provide a first-person perspective of the simulation, allowing for accurate object detection evaluation.

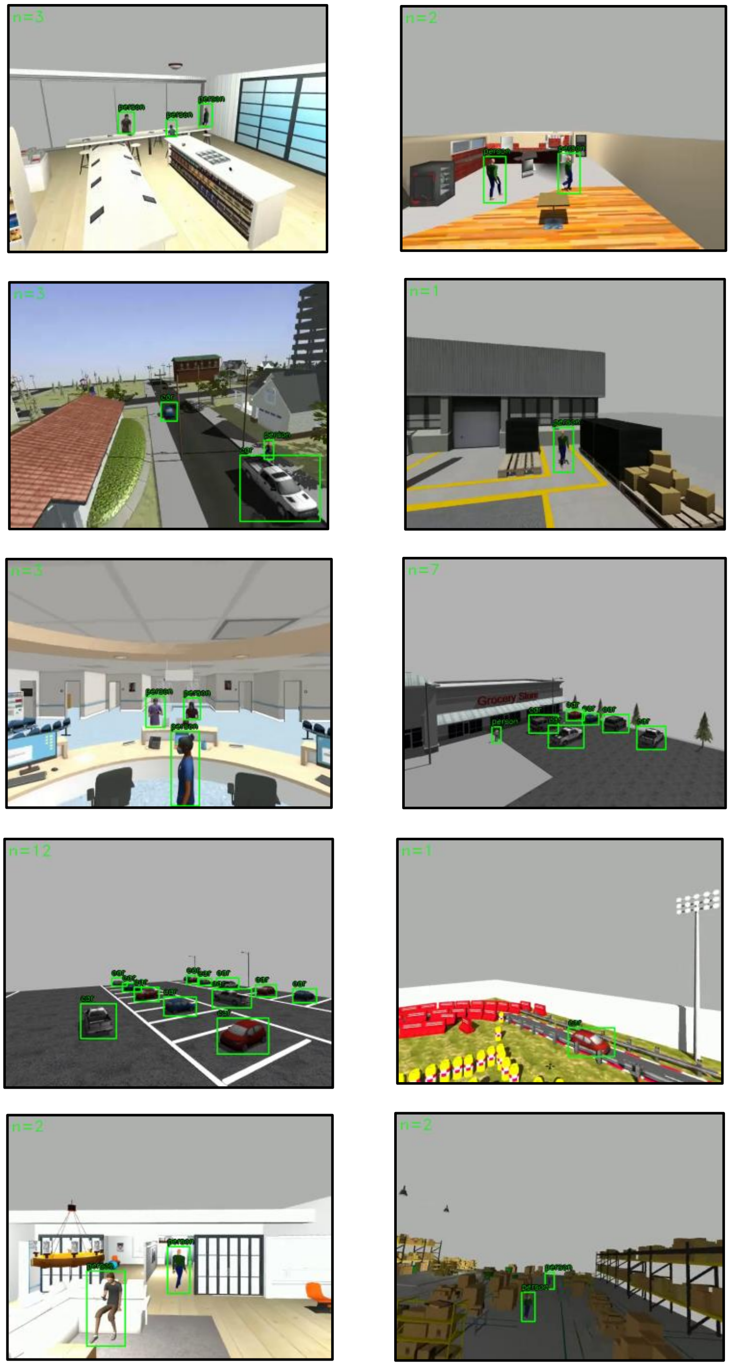

Fig. 1. Simulated Environment Datasets

Figure 1 provides a visual representation of the simulated environments. These simulated environments are intended to replicate real-world scenarios encountered by UAVs and provide a diverse and comprehensive dataset for researchers to evaluate their object detection algorithms.

3.2 Dataset Annotation

The objects of interest in our simulated dataset are persons and cars, and the dataset has 11,103 annotated frames with 27,412 annotations featuring persons and cars in different simulated scenarios. The annotations are in the YOLO (You Only Look Once) format, which provides the bounding box coordinates and class labels for each object in the frame. The Dark Label [12] annotation tool was used for the annotation process. This labeled dataset has been utilized to train and test object detection algorithms.

Additionally, the annotations in YOLO format can be easily converted into other tracking formats, such as MOT (Multiple Object Tracking), for further analysis and evaluation.

4 Evaluation and Results

Table 3. Object Detection Algorithm Performance.

YOLO Versions	YOLOv3	YOLOv3-tiny	YOLOv5	YOLOv7	YOLOv8
Precision	0.991	0.986	0.986	0.904	0.989
Recall	0.993	0.986	0.99	0.853	0.993
mAP50	0.993	0.993	0.993	0.923	0.993
mAP50:95	0.869	0.795	0.822	0.618	0.878

To test the dataset, we have trained the YOLO versions on the simulated dataset we generated and evaluated their performance. The dataset has been split into a training set, a validation set, and a test set, with 70 percent of the data used for training, 20 percent for validation, and 10 percent for testing. The training data was used to train the YOLO models, optimizing the algorithm for object detection in the simulated environments. The validation data was used to evaluate the model's performance and fine-tune its parameters to achieve the best possible results. The testing data was used to evaluate the performance of the trained models in detecting and localizing objects in new and previously unseen environments.

The YOLO models were trained on a system with a 12th gen Intel Core i7 12700*20 processor, 31.0 GiB of memory, an NVIDIA RTX A5000 graphics card, and 2TB of disk space. This hardware configuration is well-suited for deep learning applications. To run the models, we installed the TensorFlow deep learning framework and its necessary dependencies, including CUDA and cuDNN for GPU acceleration. With this setup, we were able to achieve excellent performance on our simulated dataset, demonstrating the effectiveness of YOLO models for object detection in complex environments.

We conducted experiments to evaluate the performance of different versions of YOLO for object detection in UAV applications using our simulated dataset. The models were trained for 50 epochs with a batch size of 16. The results showed that YOLOv3 and YOLOv8 outperformed YOLOv3-tiny, YOLOv5, and YOLOv7 in both indoor and outdoor environments. YOLOv3 had the highest precision, recall, and mAP50:95 scores, with a high mAP50 score, indicating good overall performance. YOLOv8 had a similar level of performance to YOLOv3 but with slightly fewer parameters and a smaller model size. The detailed results for all YOLO versions are summarized in Table 3.

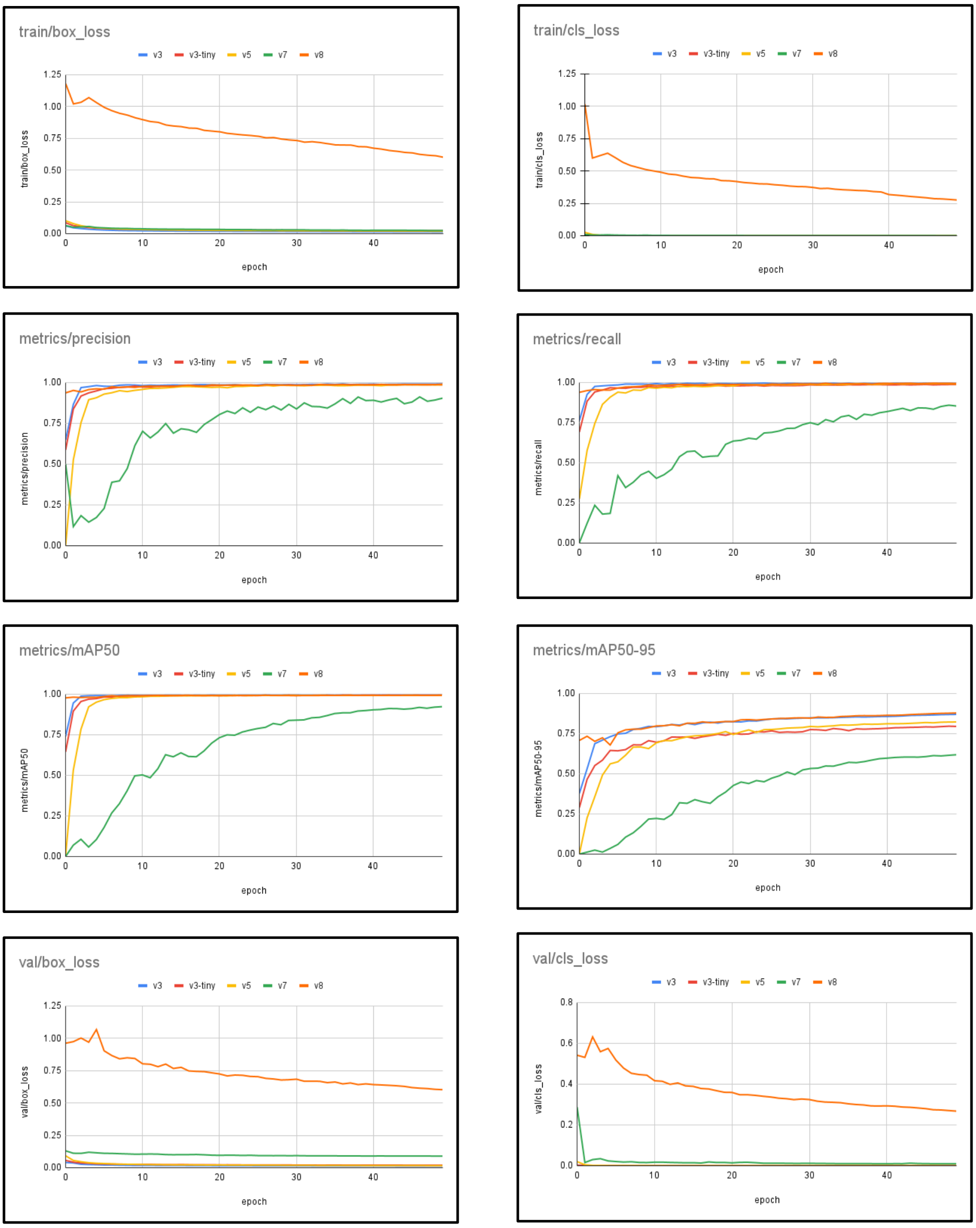

Fig. 2. Comparison of object detection performance on different YOLO versions on our simulated dataset.

The results demonstrate the effectiveness of our generated dataset for evaluating and comparing object detection algorithms, as well as the potential of using YOLOv3 and YOLOv8 for accurate and efficient object detection in UAV applications, depending on the specific requirements of the task. These versions are suitable for tasks that require high precision, recall, and mAP scores, particularly in challenging scenarios where objects may be partially occluded or have a smaller size. They are also well-suited for detecting and localizing objects in both indoor and outdoor environments.

The choice between YOLOv3 and YOLOv8 ultimately depends on the specific requirements of the task and the available resources. If resources are not a concern and you have enough computing power and time, YOLOv8 is generally preferred over YOLOv3 as it is a more advanced and accurate version of the YOLO object detection algorithm. However, if you have limited resources, such as computing power or time, and you still need good results, then YOLOv3 can be a better choice. YOLOv3 is less computationally intensive than YOLOv8, which means that it can provide reasonably good results while running faster and using fewer resources.

Figure 2 illustrates our simulated dataset's comparative analysis of object detection performance between all the YOLO versions. After examining the graphs, we also observed that YOLOv3 generally has lower values for box loss and classification loss compared to YOLOv8. This indicates that YOLOv3's object localization and classification abilities are superior to YOLOv8's. Furthermore, we noticed that YOLOv3 converges faster than YOLOv8, implying that it requires less training time to achieve optimal performance.

5 Conclusion

The study has proposed a simulated dataset through Gazebo simulation that covers both indoor and outdoor environments, consisting of 11,103 annotated frames featuring people and cars as objects of interest. This dataset provides a valuable resource for researchers to evaluate their proposed object detection algorithms in real-time UAV applications. In addition, the results of our evaluation have demonstrated that both YOLOv3 and YOLOv8 outperforms other YOLO versions in terms of object detection accuracy, particularly in challenging scenarios where objects may be partially occluded or have a smaller size. Overall, our research has demonstrated the potential of using simulated datasets and advanced object detection systems for accurate and efficient object detection in UAV applications.

References

1. Y. A. Bomantara, H. Mustafa, H. Bartholomeus, L. Kooistra, *Detection of Artificial Seed-like Objects from UAV Imagery*, Remote Sensing **15**, 1637 (2023).
2. J. Terven, D. Cordova-Esparza, *A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond*, arXiv preprint arXiv:2304.00501 (2023).
3. U. Nepal, and H. Eslamiat, *Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs*, *Sensors* **22**, 464 (2022).
4. E. Bondi, D. Dey, A. Kapoor, J. Piavis, S. Shah, F. Fang, B. Dilkina, R. Hannaford, A. Iyer, L. Joppa, M. Tambe, *Airsim-w: A simulation environment for wildlife conservation with uavs*, in proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, 1-12 (2018).
5. S. Srivastava, S. Gupta, O. Dikshit, and S. Nair, *A review of UAV regulations and policies in India*, in proceedings of UASG 2019: Unmanned Aerial System in Geomatics 1 (2020): 315-325.
6. A. Geiger, P. Lenz, C. Stiller, R. Urtasun, *Vision meets robotics: The kitti dataset*, The International Journal of Robotics Research **32**, 1231-1237 (2013).
7. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huangl, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, *Imagenet large scale visual recognition challenge*, International journal of computer vision **115**, 211-252(2015).

8. L. Wen, D. Du, Z. Cai, Z. Lei, M. Chang, H. Qi, J. Lim, M. Yang, S. Lyu, *UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking*, Computer Vision and Image Understanding **193**, 102907 (2020).
9. P. Zhu, L. Wen, D. Du, X. Bian, H. Ling, Q. Hu, H. Wu, Q. Nie, H. Cheng, C. Liu, X. Liu, W. Ma, L. Wang, A. Schumann, D. Wang, D. Ortego, E. Luna, E. Michail, E. Bochinski, F. Ni, F. Bunyak, G. Zhang, G. Seetharaman, G. Li, H. Yu, I. Kompatsiaris, J. Zhao, J. Gao, J. M. Martinez, J. C. S. Miguel, K. Palaniappan, K. Avgerinakis, L. Sommer, M. Lauer, M. Liu, N . M. Al-Shakarji, O. Acatay, P. Giannakeris, Q. Zhao, Q. Ma, Q. Huang, S. Vrochidis, T. Sikora, T. Senst, W. Song, W. Tian, W. Zhang, Y. Zhao, Y. Bai, Y. Wu, Y. Wang, Y. Li, Z. Pi, Z. Ma, *VisDrone-VDT2018: The vision meets drone video detection and tracking challenge results*, in proceedings of the European Conference on Computer Vision (ECCV) Workshops,0–0 (2018).
10. D. Du, Y. Qi, H. Yu, Y. Yang, K. Duan, G. Li, W. Zhang, Q. Huang, and Q. Tian, *The unmanned aerial vehicle benchmark: Object detection and tracking*, in proceedings of the European conference on computer vision (ECCV), 370-386 (2018)
11. M. Mandal, L.K. Kumar and, S. K. Vipparthi, *Mor-uav: A benchmark dataset and baselines for moving object recognition in uav videos*, in proceedings of the 28th ACM international conference on multimedia, 2626--2635 (2020)
12. Darkprogrammer, *https://github.com/darkpgmr/DarkLabel* (2021).