2024 IEEE 27th International Conference on
Intelligent Transportation Systems (ITSC)
September 24- 27, 2024. Edmonton, Canada

Comparison of Deep Object Detectors on a New Vulnerable Pedestrian Dataset

Devansh Sharma[1], Tihitina Hade[1], and Qing Tian[1,2]

Abstract— **Pedestrian safety is one primary concern in autonomous driving. The under-representation of vulnerable groups in today's pedestrian datasets points to an urgent need for a dataset of vulnerable road users. To help train well-rounded self-driving visual detectors and subsequently drive research to improve the accuracy of vulnerable pedestrian detection, we first introduce a new dataset in this paper: the Bowling Green Vulnerable Pedestrian (BGVP) dataset. The dataset includes four classes, i.e., Children without Disability, Elderly without Disability, With Disability, and Non-Vulnerable. This dataset consists of images collected from the public domain and manually-annotated bounding boxes. In addition, on the proposed dataset, we have trained and tested five classic or state-of-the-art object detection models, i.e., YOLOv4, YOLOv5, YOLOX, Faster R-CNN, and EfficientDet. Our results indicate that YOLOX and YOLOv4 perform the best on our dataset, with YOLOv4 scoring 0.7999 and YOLOX scoring 0.7779 on the mAP 0.5 metric, while YOLOX outperforms YOLOv4 by 3.8% on the mAP 0.5:0.95 metric. Overall, all five detectors do well in predicting the With Disability class and perform poorly in the Elderly without Disability class. YOLOX consistently outperforms all other detectors on the mAP 0.5:0.95 per class metric, obtaining 0.5644, 0.5242, 0.4781, and 0.6796 for the Children without Disability, Elderly without Disability, Non-vulnerable, and With Disability categories, respectively. Our dataset and codes are available at https://github.com/devvansh1997/BGVP.**

I. Introduction

The demand for pedestrian safety has surged in the era of autonomous vehicles and advanced driver assistance systems (ADAS). According to a Brookings Institution survey, 61 percent of adult users say they feel "uneasy" in autonomous vehicles [1]. A self-driving Uber car hit a pedestrian crossing the street in 2018, resulting in the death of the victim. The vehicle failed to react swiftly enough to avoid the crash because the software had incorrectly labeled the pedestrian as a "False-Positive" [2]. Since 2018, there have been 11 documented instances of Tesla Autopilot causing collisions, and approximately 37 test car crashes involving Uber, resulting in four fatalities [3]. In the US, 104,000 instances of pedestrian-related non-fatal accidents were documented in 2020 [4]. People over the age of 65 made up 20 percent of all pedestrian fatalities in 2020, while 1 in 5 children under the age of 15 who died in collisions were pedestrians [4]. These incidents underscore the critical need for well-trained object detectors that not only cater to average pedestrians but also to more vulnerable groups.

[1]Department of Computer Science, Bowling Green State University, Bowling Green, OH 43403, USA

[2]Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL 35294, USA

*Author emails: sharmad@bgsu.edu, thade@bgsu.edu, qtian@uab.edu

"Pedestrian" usually refers to a person walking along a road or in a nearby area. Visual detection of pedestrians is crucial in autonomous driving. However, deep visual detectors are at risk of bias and inaccuracies when it comes to vulnerable groups, as these road users are often underrepresented in popular pedestrian datasets today. The bias against vulnerable pedestrians stems more from the training data than from the model architecture itself. While there is a growing body of works focused on improving model architectures, few studies address the issue from a data perspective.

Most datasets overlook the vulnerability of pedestrians. Vulnerable pedestrians are less frequently encountered on the roads and differ significantly in size and motor abilities. Many existing datasets fail to acknowledge the differences and often categorize vulnerable road uses improperly. People frequently neglect the bias that can arise when using these datasets to train object detection frameworks. However, it is those vulnerable groups who need more of our attention. In light of this, we make two contributions in this work. A vulnerable pedestrian dataset will be introduced as the initial contribution, giving the community the resources they need to work on and enhance the effectiveness of object detection algorithms for vulnerable pedestrians. Our dataset, which includes underrepresented vulnerable road users, can also be utilized as a supplement to current general pedestrian datasets. The second contribution is that we examine and benchmark the performance of various cutting-edge detectors on our new dataset, including YOLOv4 [5], YOLOv5 [6], YOLOX [7], Faster R-CNN [8], and EfficientDet [9].

II. Related Work

We have observed a rise in datasets targeting specific categories of objects due to the growing use of object detection in various automation and security industries. Researchers use these datasets to optimize model performance, ensuring it performs well on both a wide range of generic object classes and specific subclasses within these classes. Pedestrians are one such category relevant to the autonomous driving industry, where the ability to detect different types of pedestrians is a critical component of effective self-driving software.

In [10], Kumar *et al.* present the P-DESTRE dataset that uses UAV video surveillance data for pedestrian detection and tracking. Richly Annotated Pedestrian (RAP) [11] is an extensive dataset that provides pedestrian data from an uncontrolled scene with various viewpoints of the same area. Aimed at object tracking, [12] provides pedestrian data comprised of pedestrians' and cyclists' trajectories. OpenPTDS [13] follows a similar approach where they introduced a dataset created from real-life experiments. They target the problem of pedestrian detection from the viewpoint of safety and security around metropolitan buildings.

The Caltech Pedestrian Dataset [14], which was released in 2009, is regarded as one of the most popular pedestrian datasets to date. It includes an annotation tool and a substantial amount of video footage with annotated frames. In addition to evaluating the models that were popular at the time on their dataset, it also provides a new standard for testing object detection algorithms on pedestrian data. In [15], a new pedestrian dataset is introduced along with a per-frame methodology for analyzing scale and occlusion. Utilizing the thermal channel offers a unique approach to pedestrian detection in low-light environments, and [16] provides a dataset of color-thermal paired images. Most existing datasets focus on ordinary pedestrians, while vulnerable pedestrians do not receive the necessary attention.

Existing research on vulnerable pedestrian detection primarily focuses on developing systems, algorithms, or models. Song *et al*. explore safety advancements and introduce an action character deduction and analysis module for video streams that concentrate on vulnerable pedestrians in [17]. Another example is the Pedestrian-Oriented Forewarning System (POFS) [18], which uses smartphone communication to detect potential accidents. The paper identifies pedestrians distracted by their mobile phones as vulnerable and proposes a system to adaptively alert them.

In summary, existing research in pedestrian detection either overlooks vulnerable pedestrians or disregards the impact of data. The lack of a dataset centered on vulnerable road users leads to biases against vulnerable groups such as children, the elderly, and individuals with disabilities. In this paper, we address this issue with a data-centric approach. We introduce a new dataset BGVP that focuses on vulnerable pedestrians or road users, rather than average pedestrians. Our focus is on the vulnerability related to a road user's age and physical disability. We hope this new benchmark dataset will inspire further research in this specific area.

III. BG Vulnerable Pedestrian (BGVP) dataset

Vulnerable pedestrians require more attention from self-driving vehicles. However, their under-representation in most of today's pedestrian datasets leads to unfair and discriminatory treatment. This bias has motivated us to create a dataset specifically focusing on vulnerable pedestrians, namely the Bowling Green Vulnerable Pedestrian (BGVP) dataset.

Most accidents are caused by human errors, which could be significantly reduced if autonomous driving becomes the norm. However, this requires object detectors on self-driving vehicles to rapidly and accurately detect not only regular pedestrians but also vulnerable ones, categorizing them by their type of vulnerability. There is a stark difference between an elderly disabled pedestrian and a young adult in their early 20s crossing the road. The vulnerability type is crucial as it can help estimate a pedestrian's actions in a traffic scenario or emergency. For instance, the actions of a 5-6 year-old child, even when accompanied by an adult, can be considered unpredictable and risky. Detecting the child quickly is essential to control speed and predict the child's future positions relative to the vehicle. This desirable behavior necessitates autonomous driving models to train on a dataset where such vulnerable groups have a significant presence.

Over the years, many pedestrian datasets have been published, but none of them is dedicated to or mention the vulnerability of pedestrians [15] [10] [19] [11] [20]. We provide a new dataset that fills this gap and meets this critical demand. It will enable testing of future object detection models against it to determine how well a particular model can learn pedestrian vulnerability.

Before diving into the dataset details, it is essential to understand the various classes in our dataset and the criteria for classifying each pedestrian. The bounding box instances in our dataset are categorized into four groups: "Children without Disability," "Elderly without Disability," "With Disability," and "Non-vulnerable."

- **Children without Disability**: This vulnerable pedestrian class encompasses any pedestrian between 1 to 16 years of age. Most children are considered vulnerable because their behavior is potentially erratic, and they may not understand traffic laws. Some early teens can also be considered a risk to a self-driving car, which is why we include them in the same category. Children without a physical disability are considered to fall into this category.
- **Elderly without Disability**: Elderly pedestrians are particularly at risk. Even though the Elderly may be well versed with traffic laws, their age may restrict many physical motions, and on average, they tend to need more time to complete the same action compared to a non-vulnerable pedestrian. The minimum age of the elderly demographic is subjective and varies across cultures. In this paper, this group is determined by the first author to be people who appear in the age group of 50 or older. People from this age group without any physical disability are deemed eligible for this class.
- **With Disability**: The most vulnerable group of all consists of pedestrians who have a physical impairment and may require some form of assistance. The pedestrian could be of any age group, but if they have a physical disability, they are categorized into this class because their age-related constraints are overshadowed by the disability. In our dataset, physical disability aids that we consider are wheelchairs, mobility walkers, scooters, crutches, and walking canes.
- **Non-Vulnerable**: This group of pedestrians comprises of people that do not fall into any of the classes mentioned above, and it is safe to say that they are less vulnerable in traffic situations than other categories. These pedestrians do not have any physical (or visible)

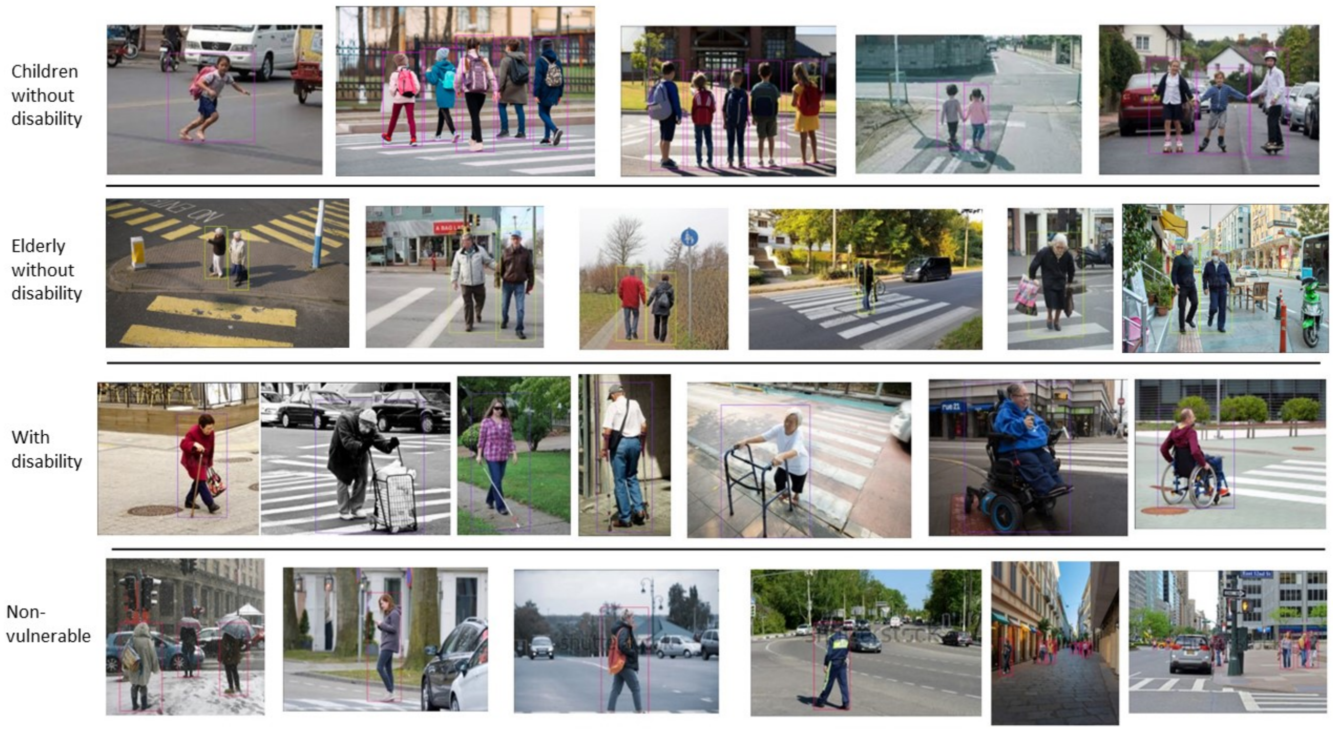


Fig. 1: Simple traffic scenes in our BGVP dataset

disability. They are supposed to act/respond quickly, and understand the situation better than vulnerable pedestrians. This is not to say that these pedestrians are not at all vulnerable; rather, their level of vulnerability in a traffic situation is lower than that of the categories mentioned above.

Fig. 1 displays a small sample of images, with each row featuring images from a single class. Fig. 2 presents a random selection of images from our database featuring complex traffic scenes, along with our annotations overlaid on them (category labels are not shown for clarity). Fig. 2 images have various class combinations, and some cover all classes. Our dataset comprises 2,000 images with a total of 5,932 bounding box annotations. On average, there are three annotations in each image and the median image size is 600 $\times$ 408 pixels. The dataset contains an assortment of images with different dimensions. The dimension of the largest image is 6000 $\times$ 4000 and the smallest is 99 $\times$ 159. All images that we have collected are from the public domain (mainly the Internet). "Children without Disability" is the most prominent vulnerable pedestrian class, with 1,646 bounding box annotations. "Elderly without Disability" has 815 bounding box instances, "With Disability" has 942 instances, and "Non-vulnerable" has 2,529 instances. Fig. 3 visualizes the distribution of bounding box categories in the dataset. Our dataset is divided into 1,405 images as the training set, 399 images as the validation set, and 196 images as the testing set. The vast majority of the images we have gathered show these vulnerable people in traffic situations, which should help provide the detection models with pertinent contextual information beneficial for training. The contextual information typically includes traffic lights, zebra crossings, vehicles, and other pedestrians (vulnerable or non-vulnerable).

Finding the right balance in annotations is crucial for helping the model learn effectively rather than causing confusion, and our annotations support this goal. We annotate all pedestrians whose categories are easily recognizable to the human eye. Over-annotation can lead to incorrect results, so in our dataset, we avoid annotating those that are too small to be identifiable in the background. This ensures we do not label them incorrectly and mislead the detectors.

IV. Models for Vulnerable Pedestrian Detection

In this paper, we experiment with a variety of state-of-the-art object detection models on our introduced benchmark. We cover both single-stage and two-stage visual detectors, i.e., YOLOv4 [5], YOLOv5 [6], YOLOX [7], EfficientDet [9], and Faster R-CNN [8].

A. Single-stage detectors

Due to their high speed and efficiency, single-stage detectors are usually preferable in time-sensitive scenarios like autonomous driving. This is also the reason why most of the detectors tested in this paper belong to this category.

A well-known one-stage detector family that has shown notable promise in the autonomous driving community is

Fig. 2: Complicated traffic scenes in our BGVP dataset

Model	Backbone	Input Size	Params	GFLOPs	mAP 0.5	mAP 0.5:0.95
EfficientDet-D0	Efficient-B0	512	3.9 M	2.5	0.7048	0.4512
Faster R-CNN	Resnet-50	640	42 M	180	0.7330	0.4860
YOLOv4	CSPDarknet-53	640	27.6 M	90.2	0.7999	0.5231
YOLOv5-s	Modified CSP v5	640	7.2 M	16.9	0.7000	0.4800
YOLOX-s	Modified CSP v5	640	8.94 M	26.8	0.7779	0.5616

TABLE I: Model Description and mAP Scores on our BGVP dataset

Model	Children w/o Disability	Elderly w/o Disability	Non-vulnerable	With Disability
EfficientDet-D0	0.4996	0.2932	0.3650	0.6469
Faster R-CNN	0.5194	0.3889	0.3970	0.6371
YOLOv4	0.5360	0.4764	0.4494	0.6304
YOLOv5-s	0.5120	0.3880	0.4040	0.6140
YOLOX-s	0.5644	0.5242	0.4781	0.6796

TABLE II: Per-class mAP 0.5:0.95 of the models on our BGVP dataset

the YOLO family [21], [22]. Redmon and his colleagues introduced the first three versions of YOLO [23]–[25], making important advancements along the way. YOLOv4 [5] uses the CSPDarknet53 backbone, the SPP additional module, and the PANet path-aggregation neck. The detector head is similar to that of YOLOv3. YOLOv4 also introduces new data augmentation methods such as Mosaic and Self-Adversarial Training(SAT). YOLOv5 was released shortly after YOLOv4 by Ultralytics via GitHub [6]. YOLOv5 is similar to YOLOv4, but one difference is that YOLOv5 leverages auto-learning bounding box anchors. YOLOv5 also has different sizes, ranging from YOLOv5-S to YOLOv5-X, with S indicating the smallest width and depth while X denoting the largest. YOLOX [7] starts with YOLOv3 as its base and makes major changes, including (1) a decoupled head for classification, regression and localization, (2) dropping the use of anchors, (3) SimOTA advanced label assignment strategy, and (4) more advanced data augmentations techniques.

EfficientDet [9], developed by Google AI, is another single-stage detector, which is highly scalable and fast. EfficientDet uses EfficientNet [26] as its backbone and a newly introduced BiFPN feature network. BiFPN enables easy and fast feature fusion and allows bi-directional information flow.

B. Two-stage Detectors

Compared to single-stage detectors, two-stage detectors are usually slow due to the extra stage of regions of interest (ROI) proposal. Faster R-CNN [8], developed by Facebook

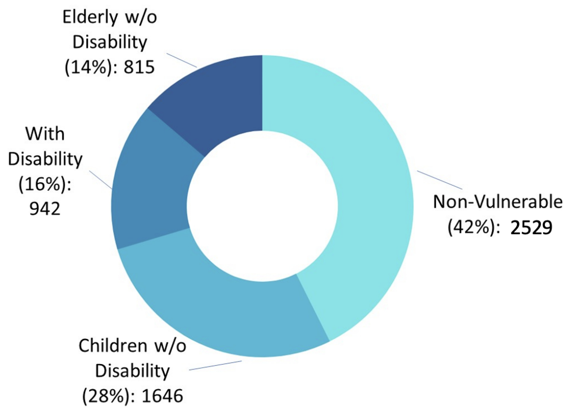


Fig. 3: Distribution of annotations among the four categories in our BGVP dataset

AI Research (FAIR), is among the most well-known two-stage detectors. It is faster than its predecessor, Fast R-CNN [27] because it replaces the slow external region proposal methods with a Region Proposal Network (RPN) that generates proposals more quickly and efficiently. It also shares convolutional features between the RPN and the detection network, streamlining the process.

V. Experiments and Discussion

In this section, we benchmark the previously mentioned models on our introduced BGVP dataset and report their results.

A. Experimental Setup

We follow the hyper-parameter settings recommended by each model's authors. All selected models were pre-trained on the MS-COCO dataset, ensuring a fair comparison of their performance on our dataset. Subsequently, we train the five models on our proposed BGVP dataset using Google Colab and the Ohio Supercomputer Center (OSC) clusters. We present our results using Mean Average Precision (mAP), specifically mAP 0.5 and mAP 0.5:0.95, as our primary metrics for evaluating model performance.

B. Experimental Results

Table I shows the details of the different models and their respective mAP scores. As we can see, YOLOX and YOLOv4 perform the best among all. YOLOv4 has the highest score on mAP 0.5, scoring 0.7999 and YOLOX is closely behind scoring 0.7779. YOLOX achieves the best mAP 0.5:0.95 score (0.5616), outperforming YOLOv4 by 3.8%. Faster R-CNN, a two-stage detector, also performs reasonably well and scores 0.7330 for mAP 0.5 and 0.4860 for the mAP 0.5:0.95 metric.

In Table II, we report the per-class scores for the mAP 0.5:0.95 metric. It is evident that all models achieve higher mAP scores for the class "With Disability". This may be attributed to the presence of regular forms, such as wheelchairs and walking canes, that constitute physical walking aids. Relatively speaking, the models do not perform well in "Elderly without Disability" and "Non-vulnerable" classes. This is perhaps due to the unavoidable ambiguity and subjectivity when determining whether a subject meets the criteria for belonging to the senior age demographic. This uncertainty can also be seen in Fig. 4, which shows the confusion matrix results of all models. EfficientDet and YOLOv5 often incorrectly predict elderly pedestrians as non-vulnerable. The YOLOX performs much better than other detectors in this regard. Performance for the class "Children without Disability" appears more consistent across all models, averaging 0.5262 (mAP 0.5:0.95), whereas for the "Elderly without Disability" class, the average is 0.4141. We also observe that all detectors struggle with images that have a darker background, often failing to detect a pedestrian at all, regardless of the category.

VI. Conclusion

In this paper, we introduced a new object detection dataset called BGVP, specifically designed for vulnerable pedestrians. We collected 2,000 images and annotated 5,932 bounding box instances from four categories, i.e., "Children without Disability", "Elderly without Disability", "With Disability", and "Non-Vulnerable". After collection and annotation, we trained and tested five state-of-the-art visual detectors and compared their results. We hope that this dataset can serve the community and motivate/facilitate more future research in this area. The dataset can also be utilized to fine-tune existing object detectors for more precise and less biased detection of vulnerable pedestrians, helping pave the way for safer and fairer autonomous driving.

Acknowledgment

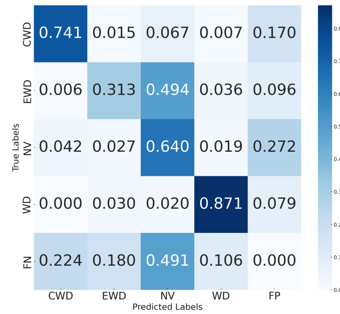
This research was enabled in part by support provided by the National Science Foundation (under Award No. 2153404 and 2412285) and the Ohio Supercomputer Center (https://www.osc.edu). Also, the authors would like to acknowledge Akinkunle A. Akinola for his help in the initial stage of the project.

References

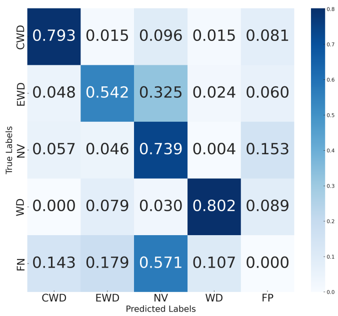
[1] A. Kopestinsky, "Self driving car statistics for 2021: Policy advice," Self driving car statistics for 2021, 29-Sep-2022. [Online]. Available: https://policyadvice.net/insurance/insights/self-driving-car-statistics/. [Accessed: 01-Dec-2022].
[2] "Fatalities caused by Autonomous Vehicles," Fatalities Caused by Autonomous Vehicles. [Online]. Available: https://qaconsultants.com/blog/fatalities-caused-by-autonomous-vehicles/. [Accessed: 01-Dec-2022].
[3] M. Tuwiner, "49 self-driving car accident rate statistics; facts," 49 Self-Driving Car Accident Rate Statistics ; Facts, 17-Jun-2022. [Online]. Available: https://annapolisclassiccars.com/self-driving-cars-statistics-facts/. [Accessed: 01-Dec-2022].
[4] "Pedestrian safety," Centers for Disease Control and Prevention, 13-May-2022. [Online]. Available: https://www.cdc.gov/transportationsafety/pedestrian_safety/index.html. [Accessed: 01-Dec-2022].
[5] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020).
[6] Ultralytics, "Ultralytics/yolov5: Yolov5 in PyTorch; ONNX; CoreML; TFLite," GitHub, 25-Jun-2020. [Online]. Available: https://github.com/ultralytics/yolov5. [Accessed: 01-Dec-2022].

(a) YOLOX

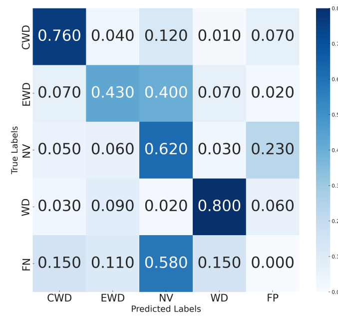
(b) YOLOv4



(c) EfficientDet



(d) Faster R-CNN



(e) YOLOv5

Fig. 4: Confusion Matrices of different models; CWD: Children without Disability, EWD: Elderly without Disability, NV: Non-Vulnerable, WD: With Disability, FP: False Positive, FN: False Negative

[7] Ge, Zheng, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. “Yolox: Exceeding yolo series in 2021.” arXiv preprint arXiv:2107.08430 (2021).

[8] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. “Faster r-cnn: Towards real-time object detection with region proposal networks.” Advances in neural information processing systems (NIPS) 28. 2015.

[9] Tan, Mingxing, Ruoming Pang, and Quoc V. Le. “Efficientdet: Scalable and efficient object detection.” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2020.

[10] Kumar, SV Aruna, Ehsan Yaghoubi, Abhijit Das, B. S. Harish, and Hugo Proença. “The p-destre: A fully annotated dataset for pedestrian detection, tracking, and short/long-term re-identification from aerial devices.” IEEE Transactions on Information Forensics and Security 16 (2020): 1696-1708.

[11] Li, Dangwei, Zhang Zhang, Xiaotang Chen, Haibin Ling, and Kaiqi Huang. “A richly annotated dataset for pedestrian attribute recognition.” arXiv preprint arXiv:1603.07054 (2016).

[12] Lambert, Jacob, Leslie Liang, Luis Yoichi Morales, Naoki Akai, Alexander Carballo, Eijiro Takeuchi, Patiphon Narksri, Shunya Seiya, and Kazuya Takeda. “Tsukuba challenge 2017 dynamic object tracks dataset for pedestrian behavior analysis.” Journal of Robotics and Mechatronics 30.4 (2018): 598-612.

[13] Song, Xiao, Jinghan Sun, Jing Liu, Kai Chen, and Hongnan Xie. “OpenPTDs dataset: Pedestrian trajectories in crowded scenarios.” Proceedings of the Asian Simulation Conference. 2018.

[14] Dollár, Piotr, Christian Wojek, Bernt Schiele, and Pietro Perona. “Pedestrian detection: A benchmark.” Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2009.

[15] Zhang, Shanshan, Rodrigo Benenson, and Bernt Schiele. “Citypersons: A diverse dataset for pedestrian detection.” Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2017.

[16] Hwang, Soonmin, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. “Multispectral pedestrian detection: Benchmark dataset and baseline.” Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2015.

[17] Song, Hyok, In Kyu Choi, Min Soo Ko, Jinwoo Bae, Sooyoung Kwak, and Jisang Yoo. “Vulnerable pedestrian detection and tracking using deep learning.” Proceedings of the IEEE International Conference on Electronics, Information, and Communication (ICEIC). 2018.

[18] Liu, Zhenyu, Lin Pu, Zhen Meng, Xinyang Yang, Konglin Zhu, and Lin Zhang. “POFS: A novel pedestrian-oriented forewarning system for vulnerable pedestrian safety.” Proceedings of the IEEE International Conference on Connected Vehicles and Expo (ICCVE). 2015.

[19] Overett, Gary, Lars Petersson, Nathan Brewer, Lars Andersson, and Niklas Pettersson. “A new pedestrian dataset for supervised learning.” Proceedings of the IEEE Intelligent Vehicles Symposium (IV). 2008.

[20] Yang, Dongfang, Linhui Li, Keith Redmill, and Ümit Özgüner. “Top-view trajectories: A pedestrian dataset of vehicle-crowd interaction from controlled experiments and crowded campus.” Proceedings of the IEEE Intelligent Vehicles Symposium (IV). 2019.

[21] Choi, Jung Im, and Qing Tian. “Adversarial Attack and Defense of YOLO Detectors in Autonomous Driving Scenarios.” Proceedings of the IEEE Intelligent Vehicles Symposium (IV). 2022.

[22] Choi, Jung Im, and Qing Tian. “Visual-saliency-guided channel pruning for deep visual detectors in autonomous driving.” Proceedings of the IEEE Intelligent Vehicles Symposium (IV). 2023.

[23] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. “You only look once: Unified, real-time object detection.” Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2016.

[24] Redmon, Joseph, and Ali Farhadi. “YOLO9000: better, faster, stronger.” Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2017.

[25] Redmon, Joseph, and Ali Farhadi. “Yolov3: An incremental improvement.” arXiv preprint arXiv:1804.02767 (2018).

[26] Tan, Mingxing, and Quoc Le. “Efficientnet: Rethinking model scaling for convolutional neural networks.” Proceedings of the International conference on machine learning (ICML). 2019.

[27] Girshick, Ross. “Fast r-cnn.” Proceedings of the IEEE international conference on computer vision (ICCV). 2015.