Pedestrian and Cyclist Object Detection Using Thermal and Dash Cameras in Different Weather Conditions

Austin Miller*, Yoosuf Marikar*, Abdulla Yousif*, Hamidreza Sadreazami‡, Marzieh Amini*
*School of Information Technology, Carleton University, Ottawa, Canada
‡Bioengineering Department, McGill University, Montreal, Canada

2024 IEEE 67th International Midwest Symposium on Circuits and Systems (MWSCAS) | 979-8-3503-8717-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/MWSCAS60917.2024.10658879

***Abstract*—Ensuring the safety of cyclists and pedestrians has become imperative in our ever expanding urban centers. Despite advancements in vehicle safety technology, traditional cameras often fail in adverse weather and low-light conditions. This paper investigates the efficiency of integrating thermal cameras with dash cameras to enhance detection accuracy of vulnerable road users. We first collected and annotated datasets, comprising thermal and dash camera footage under various weather conditions. We then developed a deep learning object detection model using YOLOv8 and Roboflow. Separate models were trained for each camera, then fused to compensate for their individual limitations. It was observed that dash camera is prone to occlusions and varied lighting, whereas the thermal camera excels in low-light settings. The performance metrics for the thermal camera showed a total mAP50 of 0.92 and mAP50-95 of 0.52 for detecting both cyclists and pedestrians, reflecting a highly effective system with significant potential to improve road safety.**

***Index Terms*—Object detection, thermal camera, dash camera, YOLOv8, deep neural network**

I. Introduction

As congestion increases in urban areas, the safety of cyclists and pedestrians has become a critical concern. Despite major advancements in vehicle safety technology, the use of traditional cameras still leaves vulnerable road users susceptible to poor road conditions, adverse weather, and bad visibility [1]. Various studies have investigated the limitations of traditional vehicle safety technologies and proposed enhancements through sensor fusion techniques [1], [2]. Traditional cameras, despite their widespread use in vehicle safety systems, often struggle with detection accuracy under adverse weather and low-light conditions [3]. In [4], it was highlighted that conventional visible spectrum cameras frequently fail to detect pedestrians and cyclists at night or in foggy conditions, leading to a higher risk of accidents. This study underscores the necessity for complementary technologies that can overcome these limitations. Thermal cameras, which detect infrared radiation, provide significant advantages in low-visibility scenarios. In [5], it was demonstrated that thermal imaging effectively identifies heat signatures of pedestrians and cyclists regardless of lighting conditions. Their research showed a marked improvement in detection rates when thermal cameras were employed alongside traditional cameras.

This work was supported by Natural Sciences and Engineering Research Council (NSERC) of Canada.

In this paper, we aim to investigate the effectiveness of compensating for the downsides of typical cameras with the use of a thermal camera. The integration of a thermal camera alongside a dash camera allows for increased accuracy of detection for cyclists and pedestrians in low-light and adverse weather conditions. The incorporation of both a thermal and dash camera presents several challenges such as equipment set-up, calibration, and data fusion. Data was collected, annotated, and split for training models and testing them. To realize a accurate object detection and data fusion methods, we develop deep learning-based object detection models trained to detect pedestrians and cyclists from a moving vehicle using both thermal and dash cameras. The separately-trained models are fused utilizing distinct optical information weighted to balance an efficient model. The performance metrics are then calculated. To reinforce thermal camera reliability verification, several unique weather conditions are considered for data collection, namely, Overcast, Sunny, Night time and Sleet/Snow.

II. Data collection and Methodology

A. Data collection equipment setup

The thermal camera was attached to the front roof of the vehicle, while the dash camera was mounted to the top of the inside of the windshield. Fig. 1 shows the set-up for thermal and dash cameras. These two different locations created accuracy and consistency errors when formatting the data.

Figure 1: Thermal and dash camera set-up on the vehicle.

To account for this, the dash camera images were cropped and resized to appropriately match the size of the thermal camera images, which had very different resolutions as shown in Table I.

Table I: Camera Specifications

Camera type	Camera name	FOV (deg.)	Resolution
Thermal	FLIR ADK	75	640x512
Dash camera	RedTiger F7NP DashCam	170	3840x2160

B. Data pre-processing

Preparing the dataset involved collecting data, annotating the data and splitting it into training, validation and testing. In addition, due to the small size of the dataset and to overcome the challenges related to over-fitting and better generalizability of the models, we apply data augmentations.

More specifically, data was collected by extracting frames from video capture recordings, in thermal and visual bands. Data labeling was completed using Roboflow [6], with a manual labeling user interface. This means drawing bounding boxes around perimeter of identified subject with appropriate class. Dataset was split using a 65/21/14 ratio, corresponding to Training, validation and testing, respectively. To this end, shearing and flipping augmentations were selected due to strong vertical dependency for feature extraction. Note that a cyclist or pedestrian would never be upside down. The following data augmention is considered for both thermal and cash camera including, shearing (Vertical: 10/10 degrees, Horizontal: 10/10 degrees), and horizontal flip.

In Fig. 2, a sample of thermal camera data which is augmented using Sheering is shown.

Figure 2: Thermal cam shear-augmented sample

C. Training separate models

Models were trained using YOLOv8 [7] using the standard COCO dataset released from Microsoft for general object detection [8]. However, only cyclist and pedestrian classes were included. Fusion was completed using Roboflow to merge the datasets. The fused model was trained once more.

III. Dataset performance and accuracy

The final model performed moderately, and managed to correctly identify most pedestrians, and some cyclists. The model was weighted 50/50 due to the small size of the individual datasets. Included in this section are the dataset sizes, the performance of each separate model and the performance of the fused model.

A. Dataset Details

Augmentations were performed - as described in section II-B, to increase dataset size and ultimately enhance model performance. The datasets had identical splits as described in section II-B, which was a calibration procedure to ensure performance of each dataset were separately comparable.

Each batch of data can be defined as a group of images collected from data of unique weather conditions. Each batch was roughly 70 - 200 images in size. For an object detection model, this is a relatively small dataset. The base complete datasets sizes are listed in Tables II and III. At highest achieved performance, the thermal and dash camera datasets collectively had 528 and 506 files, respectively, and with augmentation this increased to 1268 and 1297, respectively.

Table II: Thermal dataset proportions

Training	Sunny	Overcast	Night	Sleet	Total
Pedestrian	83	19	74	52	228
Cyclist	1	6	37	-	44
Both	10	6	2	1	19
Total	10	6	2	1	19
Validation	**Sunny**	**Overcast**	**Night**	**Sleet**	**Total**
Pedestrian	26	8	23	15	72
Cyclist	1	6	37	-	44
Both	0	1	10	0	11
Total	2	4	1	0	7
Testing	**Sunny**	**Overcast**	**Night**	**Sleet**	**Total**
Pedestrian	15	0	12	8	35
Cyclist	0	0	10	0	10
Both	0	0	2	0	2
Total	15	0	24	8	47

Table III: Dash camera dataset proportions

Training	Sunny	Overcast	Night	Sleet	Total
Pedestrian	116	16	67	46	245
Cyclist	5	5	36	0	46
Both	13	10	3	0	23
Total	134	31	106	46	314
Validation	**Sunny**	**Overcast**	**Night**	**Sleet**	**Total**
Pedestrian	31	10	23	18	82
Cyclist	3	0	12	0	15
Both	10	1	0	0	11
Total	44	11	35	18	108
Testing	**Sunny**	**Overcast**	**Night**	**Sleet**	**Total**
Pedestrian	31	3	14	11	59
Cyclist	4	0	12	0	16
Both	6	1	2	0	9
Total	41	4	28	11	84

B. Performance Metrics

Mean average precision (mAP) was used to assess the accuracy of our models. mAP is calculated using the different sub metrics like confusion matrix, intersection over union (IoU), recall, and precision [9], [10], [11]. Intersection of Union (IoU): measures the degree of overlap between the predicted bounding box coordinates and the ground truth box coordinates. A higher IoU value indicates that the predicted bounding box closely matches the ground truth box. Recall: measures how often true positives occur out of all the predictions

Table IV: Performance metrics for the two cameras and different weather condition

	Thermal Camera		Dash Camera	
	mAP50	**mAP50-95**	**mAP50**	**mAP50-95**
Sunny				
Cyclist	-	-	-	-
Pedestrian	0.854	0.532	0.763	0.404
Average	0.854	0.532	0.763	0.404
Overcast				
Cyclist	0.786	0.432	0.626	0.399
Pedestrian	0.839	0.407	0.487	0.186
Average	0.813	0.42	0.556	0.293
Night				
Cyclist	0.995	0.693	0.709	0.315
Pedestrian	0.798	0.466	0.51	0.2
Average	0.897	0.58	0.609	0.258
Sleet				
Cyclist	-	-	-	-
Pedestrian	-	-	-	-
Average	0.382	0.133	0.382	0.118
Total				
Cyclist	0.962	0.585	0.871	0.554
Pedestrian	0.879	0.464	0.711	0.322
Average	**0.921**	**0.525**	**0.791**	**0.438**

- **Precision**: Indicates the probability of identifying the accurate value within all the detected targets.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

- **Recall**: Indicates the probability of correct identification in all positive samples.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

- **AP**: Summarizes the precision-recall curve as a single value by computing the average precision value for recall values over the interval [0, 1].

$$\text{AP} = \int_0^1 P(r)\, dr \tag{3}$$

- **mAP** indicates the average evaluation of AP across different categories.

C. Experimental results and Discussion

Table IV gives the performance of each dataset and their individual classes to the small size, some data could not be accurately and reliably tested. This is due to the weather not cooperating and providing a reliable time to collect data. Note that our dataset is publicly available [12]. The most efficient batch is clearly the thermal camera at night time, and the least efficient were both dash and thermal camera datasets during sleet. These results indicate how the thermal camera is more effective in low lighting environments.Figure 3 and 4 show the comparison of pedestrian and cyclist detection in sunny weather and night, respectively.

The overall mAP-50 of the thermal dataset: 0.89 was promising, however could be improved given a larger dataset. The dash cam mAP-50: 0.788 was moderately accurate and there are several reasons for these differences. The dash camera had much more variety and challenges for classification, including color, occlusion like snowflakes, rain, glare etc, which made it effectively weaker because there were more possible instances of false positives.

It should be pointed out that the thermal dataset was trained on more precise data, meaning labeled images were much more consistent and strengthened the model faster. For each dataset, and each camera used, there were unique challenges that created interference and occlusion that complicated the datasets. On sunny days, glare refracted from the windshield, interfering with dash cam, and heat would be generated from absorption, interfering with thermal results. During night time, the thermal camera would far outperform the dash camera, since it senses NIR (Near-Infrared) and not visible light. During sleet conditions, the thermal camera also outperformed the dash camera since the colder and snow occluded environment was favorable for thermal vision. The overcast environment was generally consistent in performance for both datasets since occlusion and lighting were minimized and maximized respectively.

The most significant observation related to weather occlusion was the environment definition for the thermal camera. Theoretically, thermal cameras which interpret heat signatures could visualize both pedestrians and cyclists better than dash cameras due to a variety of training data. Dash camera data includes 3 color channels, making image processing more tedious and complex. A cyclist could be wearing a backpack, have a helmet, a raincoat or a brightly coloured bicycle which could fool the model easily. The thermal camera with one channel, easily defined objects in an image and classified them, when heat signatures stood out from their environment. This means a pedestrian would need to be relatively hot and the environment relatively cool. This meant the thermal dataset excelled in all environments but sunny. The thermal camera could also be fooled if someone wore heavy clothing that protected their heat signature or if they were walking into the wind for a long period of time. We observed through experiment that a 3-channel thermal images were more effective for object detection and classification. It is to be noted that the thermal dataset had a particular tendency to classify car tires as cyclists, due to their proximity when viewed from behind without the proper context of the vehicle.

IV. Conclusion

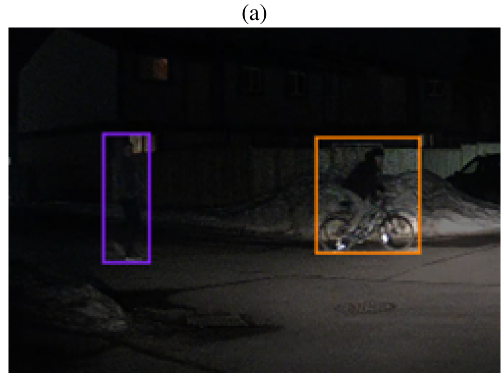
This paper has demonstrated the integration and effectiveness of advanced detection systems using dash cameras and thermal cameras powered by YOLOv8 and Roboflow, for the identification of pedestrians and cyclists in various environmental conditions. We focused on evaluating the system's performance in terms of mean Average Precision (mAP) at Intersection Over Union (IOU) of 50% (mAP50) and the more stringent 50%-95%(mAP50-95), across sunny, overcast/cloudy, night, and sleet conditions. The overall performance metrics, with a total mAP50 of 0.92 and mAP50-95 of 0.52 for detecting both cyclists and pedestrians (thermal camera), reflect a highly effective system with significant potential to improve road safety. However, the reduced effectiveness in

(a)

(b)

Figure 3: (a)-(b) Comparison of data in sunny conditions for thermal and dash camera, respectively

(a)



(b)

Figure 4: (a)-(b) Comparison of data at night for thermal and dash camera, respectively

sleet conditions and the low accuracy in detecting pedestrians in some scenarios indicate specific avenues for future improvement. In addition, the overall performance metrics, with a total mAP50 of 0.788 and mAP50-95 of 0.414 for detecting both cyclists and pedestrians (dash camera), paled in comparison in terms of performance of the thermal camera, even in favorable conditions such as sunny conditions. Future research should aim to refine detection algorithms for adverse weather conditions and develop adaptive models that can dynamically adjust to varying environmental conditions. Furthermore, expanding the dataset to include a broader range of weather conditions, pedestrian and cyclist clothing variations, movement patterns and overall more cyclist data will be crucial for improving the robustness and accuracy of detection systems.

References

[1] M. M. Narkhede and N. B. Chopade, "Review of advanced driver assistance systems and their applications for collision avoidance in urban driving scenario," in *Machine Learning and Big Data Analytics (Proceedings of International Conference on Machine Learning and Big Data Analytics (ICMLBDA) 2021)*, pp. 253–267, Springer, 2022.

[2] S. Ahmed, M. N. Huda, S. Rajbhandari, C. Saha, M. Elshaw, and S. Kanarachos, "Visual and thermal data for pedestrian and cyclist detection," in *Towards Autonomous Robotic Systems: 20th Annual Conference, TAROS 2019, London, UK, July 3–5, 2019, Proceedings, Part II 20*, pp. 223–234, Springer, 2019.

[3] H. Abbasi, M. Amini, and F. R. Yu, "Fog-aware adaptive yolo for object detection in adverse weather," in *2023 IEEE Sensors Applications Symposium (SAS)*, pp. 1–6, 2023.

[4] F. Altay and S. Velipasalar, "The use of thermal cameras for pedestrian detection," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 11489–11498, 2022.

[5] T. Fu, J. Stipancic, S. Zangenehpour, L. Miranda-Moreno, N. Saunier, *et al.*, "Automatic traffic data collection under varying lighting and temperature conditions in multimodal environments: Thermal versus visible spectrum video-based systems," *Journal of advanced transportation*, vol. 2017, 2017.

[6] S. Alexandrova, Z. Tatlock, and M. Cakmak, "Roboflow: A flow-based visual programming language for mobile manipulation tasks," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5537–5544, IEEE, 2015.

[7] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-time flying object detection with yolov8," *arXiv preprint arXiv:2305.09972*, 2023.

[8] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. 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*, pp. 740–755, Springer, 2014.

[9] H. Sadreazami, D. Mitra, M. Bolic, and S. Rajan, "Compressed domain contactless fall incident detection using UWB radar signals," in *2020 18th IEEE International New Circuits and Systems Conference (NEWCAS)*, pp. 90–93, 2020.

[10] H. Sadreazami, M. Bolic, and S. Rajan, "Residual network-based supervised learning of remotely sensed fall incidents using ultra-wideband radar," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–4, 2019.

[11] Z. Sharifisoraki, M. Amini, and S. Rajan, "Comparative analysis of mmwave radar-based object detection in autonomous vehicles," in *2024 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1–6, 2024.

[12] A. Marzieh and et.al., "Dataset of thermal and dash cameras for pedestrian and cyclist detection: https://doi.org/10.5683/sp3/jjwiev," 2024.