A Deployable Privacy-Preserving Thermal-Based Obstacle Detection System for Indoor Navigation

Jingxiang Gao
Carnegie Mellon University
jingxia3@andrew.cmu.edu

Hend K. Gedawy
Carnegie Mellon University
hkg@andrew.cmu.edu

Eduardo Feo-Flushing
Carnegie Mellon University
efeoflus@andrew.cmu.edu

Khaled A. Harras
Carnegie Mellon University
kharras@andrew.cmu.edu

ICC 2025 - IEEE International Conference on Communications | 979-8-3315-0521-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICC52391.2025.11161365

***Abstract*—Indoor navigation and wayfinding have a wide range of applications, from enabling robots to efficiently navigate warehouses to enhancing mobility for people with visual impairments (PVI). These systems rely on a combination of localization, obstacle detection, and path-finding technologies. While real-time obstacle/object detection is essential for safe indoor navigation, current solutions are impractical, face privacy concerns, and have high barriers to adoption; particularly for PVI. In this paper, we introduce DareDevil, a highly deployable indoor navigation system that ensures safe navigation while preserving user privacy. DareDevil offers three main components: a comprehensive navigation dataset, a real-time thermal-based obstacle detection model, and a thermal colorization module that feeds thermal data to the model during inference. The thermal-based dataset supports seamless navigation in both indoor and outdoor environments. Our system provides trained models of varying sizes; making it compatible with a wide range of high-end IoT devices and eliminating the need for bulky or expensive processing units. We evaluate DareDevil's performance across different model sizes and thermal image colorization schemes. Our results show that DareDevil's models can achieve higher accuracy compared to the state-of-the-art real-time object detection model. Furthermore, the findings highlight the often-overlooked impact of thermal image colorization on thermal-based models.**

***Index Terms*—Indoor Navigation, Object Detection, Thermal Imaging, YOLO, Privacy Protection, Machine Learning**

I. Introduction

Indoor navigation involves determining and following routes within enclosed environments like shopping malls, airports, and office buildings. Unlike outdoor navigation, indoor environments present unique challenges due to the absence of satellite signals and the complexity of indoor layouts. Applications of indoor navigation range from guiding robots in warehouses to improving mobility for people with visual impairments (PVI) [1, 2, 3]. These systems depend on a combination of localization [4, 5], obstacle detection, and pathfinding technologies [6]. However, several barriers hinder their widespread adoption, particularly for PVI. Barriers include depending on external infrastructure (e.g., Bluetooth beacons, Wi-Fi networks), difficulty in maintaining accurate localization, high costs, low portability, steep learning curves, and privacy concerns; especially with camera-based solutions [7, 8].

While real-time obstacle/object detection is essential for indoor navigation, current solutions face high deployment barriers. Many solutions rely on expensive dedicated computing units and specialized sensors, which pose both financial and physical burdens to users [9, 10]. In addition, these systems often rely on RGB images [11, 12]. Despite their high performance in the detection of multiclass objects, they can collect personally identifiable information, such as details of facial features, raising privacy issues [13]. Thermal imaging offers a more reliable and privacy-preserving alternative[14], and has been widely adopted in privacy-sensitive applications; such as crowd monitoring [15, 16], smart homes [17], and autonomous vehicles [14]. To be effective, thermal images must be collected in real-time and processed by a model trained to analyze them and recognize objects. However, developing such a model requires access to indoor thermal image datasets, which are currently lacking. Popular datasets, such as KAIST [18], CVC-14 [19], LLVIP [20], are focused solely on outdoor environments, while indoor datasets like TRISTAR [21] and OdomBeyondVision [22] are tailored to tasks unrelated to object detection, such as human posture analysis and odometry. As a result, thermal-based models designed for outdoor environments [23] tend to perform poorly indoors, struggling to identify obstacles beyond humans or pets [24, 25].

This paper introduces DareDevil, a privacy-preserving deployable indoor navigation system. DareDevil provides an indoor thermal-based dataset and a real-time obstacle detection model that ensures safe navigation while maintaining user privacy. Three key factors drive its high deployment potential. First, DareDevil offers three model sizes, each capable of running on various devices, including high-end mobile and IoT devices. This eliminates the need for additional, specialized, expensive, or bulky processing units, allowing users to leverage their existing high-end mobile devices to run the obstacle detection model. Second, DareDevil provides a comprehensive dataset, enabling seamless operation in indoor and outdoor environments. This versatility is crucial for integrating the obstacle detection module with existing localization and pathfinding technologies, providing PVI with a more seamless and enhanced navigation experience. Third, the system's strong privacy protections are a significant step toward increasing the adoption of indoor navigation solutions.

DareDevil consists of three main components: a comprehensive navigation dataset, a real-time thermal-based obstacle detection model, and a thermal colorization module (Section II). To create a robust thermal dataset that enhances model performance, DareDevil combines an existing outdoor thermal dataset with a newly developed indoor dataset. The indoor data is extracted from collected thermal videos and undergoes several preprocessing steps. This includes video alignment, frame extraction, automated annotation, and colorization, to generate a synthesized thermal image dataset. This dataset is then used to train new models of different sizes based on the widely adopted real-time object detection framework, YOLOv8 (You Only Look Once) [26]. As YOLO is designed to process 3-channel RGB images, DareDevil features a real-time module that converts captured thermal video frames to RGB images using a selected colormap, and feeds these images to the model in real-time during inference.

We evaluate DareDevil's trained models, in terms of inference time and Mean Average Precision (mAP), using different colormaps (Section III). DareDevil demonstrates higher performance compared to the original state-of-the-art YOLOv8 RGB-based models [26]; achieving a high mAP of up to 79.3% across all model architectures and colormaps. The results also emphasize the significant impact of thermal image colorization on model performance, a factor often overlooked in previous research. In some cases, colormap selection resulted in a performance drop of up to 6.1% in mAP. Additionally, we use a DareDevil's RGB-based model, trained on RGB images, as a baseline to DareDevil's thermal-based models. The findings indicate that thermal-based models trained on colorized thermal images outperform RGB-based models. This is likely due to the thermal representation's ability to highlight key features, reduce visual noise, and provide more consistent object contrast, especially in challenging environments.

II. Methodology

This section outlines the specifications, technical challenges, and solutions implemented for DareDevil's key components. There are three key components: a comprehensive navigation dataset, real-time thermal-based object detection models, and a real-time thermal colorization module that feeds thermal data to the model during inference. Figure 1 summarizes our methodology, while further implementation details can be found in DareDevil's open-source repository [27].

A. Dataset Creation

We now describe the process of creating an indoor thermal dataset for object detection, which involves data collection and several preprocessing stages, as illustrated in Figure 1.

Training a YOLO model requires a large dataset of annotated images to enable accurate object detection and classification. For each image, YOLO needs a corresponding annotation file containing bounding box coordinates and class labels for the objects in that image. These annotations guide the model in identifying object locations and types.

1) Data Collection: To develop a comprehensive dataset that enables robust model performance in indoor environments, we augment the Teledyne FLIR ADAS Dataset [28] (FLIR) with a new dataset collected from recordings inside the Carnegie Mellon University Qatar (CMUQ) campus. Integrating these two datasets serves multiple purposes. While our primary focus is indoor navigation, training on a combined dataset enhances the model's ability to operate seamlessly in both indoor and outdoor settings. This is essential for integrating the obstacle detection module with existing localization and pathfinding technologies, providing PVI with a smoother navigation experience. Additionally, our indoor dataset has limited diversity in lighting conditions, object appearances, and spatial layouts, which could hinder model generalization. Incorporating the larger, more varied outdoor FLIR dataset introduces greater variability, improving the model's adaptability and robustness in different indoor environments.

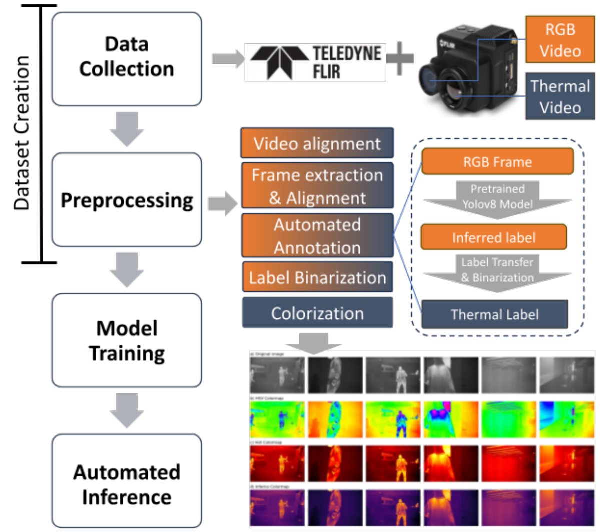

Fig. 1: DareDevil's Methodology

The FLIR dataset is a comprehensive, publicly available dataset that comprises thermal images for outdoor environments, with annotation in accordance to MSCOCO format [28]. This dataset consists of 10,742 thermal images in the training set, 3,749 images in the testing set, and 1,144 images in the validation set. This large dataset adds up to 15,635 fully annotated frames with 520,000 bounding box annotations.

FLIR dataset is augmented with our dataset that was created based on a 3-minute video collected using a FLIR DUO Pro R RGB and thermal imaging camera. Such a dual camera was critical for dataset creation since existing YOLO models work for RGB images. Hence, we used these RGB video images as the ground truth for labeling and annotating thermal video images; as explained in Section II-A2. The RGB and thermal videos capture the indoor environment of the CMUQ campus and include various objects, such as people, chairs, tables, and plants, to simulate real-world indoor navigation scenarios. The data collection aimed to create a diverse and comprehensive dataset suitable for training our obstacle detection models.

2) Preprocessing: The dual RGB and Thermal videos were used to create an annotated and labeled thermal dataset for YOLO training. RGB video images were annotated and labeled and used as a reference for thermal image annotations and labeling. However, this required multiple preprocessing steps (Fig. 1) which are listed below.

Timestamp Alignment: The collected RGB and thermal videos had different lengths, with the RGB video being one second longer due to a potential synchronization issue in the

(a) Example of frozen frame (b) Example of aligned frames

Fig. 2: Frame filtering and alignment

camera hardware. To align the timestamps, we identified a common calibration frame and manually synchronized the videos using a video editing tool. This process involved trimming the extra second from the RGB video to ensure both videos started and ended at the same time.

Video Frame Extraction and Filtering: The RGB and thermal videos were extracted into 930 frames each using OpenCV, at a frame rate of 5 frames per second to allow more frame-to-frame variation. Some of the thermal frames were frozen in place due to the camera saving time, with a blue square shown on the top right of that frame (e.g. Fig. 2a). When this happens, the subsequent frames will not match the corresponding frames in the RGB video, and will affect downstream annotation and training. After removing frozen frames, 905 valid frames are extracted from the video.

RGB-Thermal Frame Alignment: Manual alignment of RGB and thermal frames was performed to address the discrepancy in their capture parallax. We cropped and resized the frames to match the thermal image dimensions using a photo editing tool; ensuring that corresponding RGB and thermal frames were perfectly aligned (Fig. 2b).

Automated Annotation: We implemented an automatic image labeling method to avoid the time cost of manual annotation. We first used the pre-trained YOLOv8 model provided by Ultralytics [26] to annotate the RGB frames. After a label file is generated for each RGB frame, these labels are transferred to the corresponding thermal frames, marking the bounding boxes and classes of the corresponding objects. Annotations were used as a reference for the thermal images, ensuring consistency across both data modalities (Fig. 3).

Label Binarization: All labels are cast to a single class representing obstacles for three critical reasons. First, it maximizes the model's ability to detect obstacles in thermal environments agnostic to the object's class. Second, this is critical for speeding up the inference time, allowing a fast response to navigating users, which is important for PVI's safety. Third, it is sufficient for rudimentary navigation for PVI, where the main goal is to avoid colliding with objects like walls or furniture. However, to compensate for the lack of detailed object classification, the aim is to integrate this binary obstacle classification with a localization system that provides contextual information (such as the user's position and the layout of the environment) [29]. For example, when a user is navigating a shopping mall, the binary object detection quickly identifies an obstacle ahead, while the localization system knows the user is near an elevator based on the map. The system provides guidance such as "Obstacle ahead, this is likely the reception desk." The user can avoid the obstacle and move toward the next point, such as a nearby store or exit. This approach balances simplicity and effectiveness, making it a viable solution for safe indoor navigation for visually impaired users, especially in pre-mapped or well-known environments.

Fig. 3: Sample frame with RGB to thermal label transfer

Image Colorization: Thermal images often require colorization to be processed by computer vision models, which generally operate on RGB images. Thermal images are typically 1-channel (grayscale) images. They capture temperature information in a single band, where each pixel represents a temperature intensity, usually in shades of gray. For instance, black is commonly used for cooler areas and white for hotter areas. However, models like YOLOv8 (on which DareDevil's models are based) are designed to process 3-channel RGB images. Hence, 1-channel thermal images must be converted to 3-channel images by applying a colormap. However, converting high-bit thermal images (up to 14 bits per pixel) to RGB can result in significant information loss [30]. Besides, existing research in thermal-image-based object detection models has largely overlooked the impact of colorization schemes on model training outcome [31, 32, 33]. Further research is needed to understand its impact on object detection within indoor navigation systems.

To plant the seed for future research on the effects of thermal colorization in computer vision, we selected different representative colorization schemes for the DareDevil's dataset from OpenCV library. In addition to the original colorization scheme, we selected HSV, Inferno, and Hot (Fig. 4, Fig. 5). HSV has enhanced color contrast, which is prevalent in image processing. Inferno is a high-contrast scheme at low values, while Hot is a high-contrast scheme at high values.

Final Dataset Specifications: In summary, DareDevil's thermal dataset consists of 11,285 images in the training set, 4,021 in the testing set, and 1,234 in the validation set. The split is roughly 6:3:1 for training, testing, and validation data. Datasets with different colormaps are created from this data set to evaluate the impact of thermal-image colorization on model's performance; as explained in Section III.

B. Model Training

1) Model Backbone: DareDevil's models are built based on the state-of-the-art YOLO framework; specifically, YOLOv8 [26]. YOLO stands out from other object detection backbones for achieving real-time inference speed while maintaining high accuracy on diverse datasets [34, 35]. With open-source archi-

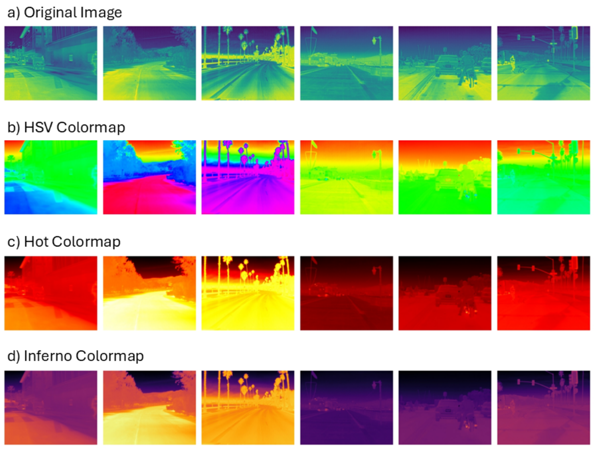

Fig. 4: Example FLIR images with applied colormaps: original, HSV, Hot, and Inferno; respectively

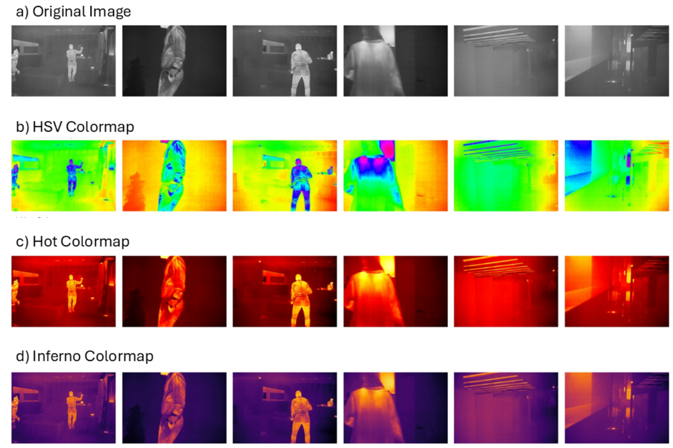

Fig. 5: Example images from collected video with applied colormaps: original, HSV, Hot, and Inferno; respectively

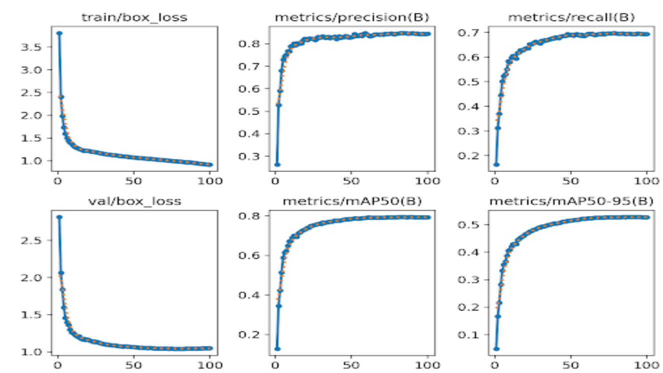

Fig. 6: Training and Validation curves for DareDevil's Medium Model using original colormap dataset.

tectures across different sizes, YOLO has been widely adopted for real-time object detection systems, in domains such as autonomous vehicles and robotics. We use YOLOv8 models (nano, small, medium) as our backbone for object detection. Nano prioritizes speed over accuracy, making it useful for real-time applications on limited hardware. Small balances speed and accuracy well, suitable for mid-range applications. Medium achieves higher accuracy but with a higher computational cost. The Nano, Small, and Medium model architectures contain 3.2, 11.2, and 25.9 million parameters, respectively. This choice of different model architecture sizes provides flexibility for running model inference on different computing devices; including high-end mobile and IoT devices. Instead of relying on additional, special, expensive, or heavy processing devices, DareDevil opens the door for users to rely on their high-end mobile devices to run the obstacle detection model.

2) Training Process: The training was performed on 2 NVIDIA A100 Tensor Core GPUs, using Python 3.12.2 and PyTorch framework. Different models were trained with our dataset using the architectures of YOLOv8 model (tiny, small, medium) from Ultralytics [26]. The training parameters were configured as follows: a learning rate of 0.01, batch size of 16, 100 epochs, and the auto optimizer, which automatically selects an appropriate optimizer based on the model's configuration. The training process involved iteratively improving the model by analyzing its performance on validation data and making necessary adjustments.

Figure 6 presents training curves for DareDevil 's Medium Model under the original colormap. Horizontal axis represents the training epoch, the blue dots are the data points, and the red dots represent the trend line; Figure 6. Other models show a similar pattern with different colormaps. Validation losses consistently decrease over 100 epochs, indicating that the models do not exhibit signs of overfitting. Plateauing observed in (Mean Average Precision) mAP50 and mAP50-95 metrics suggests that sufficient training has been performed and the models are not underfitting. The mAP is a standard metric to assess the performance of object detection models, which considers both precision and recall. mAP50 is the average precision when Intersection over Union (IoU) is 50%, where IoU is a metric that measures the overlap between the predicted bounding box and the ground truth box. mAP50-95 is the average precision across IoU thresholds from 50% to 95%, providing a more thorough measure of accuracy across various levels of object detection difficulty. Collectively, the training curves in Figure 6 indicate that the models have undergone adequate training.

C. Automated Inference

The third component DareDevil offers is a real-time thermal colorization module that feeds thermal data to the model during inference. This module extracts thermal images from the thermal camera and colorizes them to RGB images, which are then fed to the model. The model generates a label file that can be leveraged in downstream tasks during navigation, such as giving real-time warnings or navigation instructions. The model output contains information about detected objects, including their bounding boxes and confidence scores.

III. Performance Analysis

We evaluate DareDevil's trained models in terms of mAP50 and mAP50-95 using different colormaps. To investigate the

TABLE I: DareDevil's performance given different colormaps

Colormap	Backbone	mAP50	mAP50-95
Original	Nano	**0.722**	**0.452**
	Small	**0.768**	**0.498**
	Medium	**0.793**	**0.527**
HSV	Nano	0.653	0.390
	Small	0.711	0.439
	Medium	0.735	0.464
Hot	Nano	0.665	0.399
	Small	0.717	0.447
	Medium	0.744	0.475
Inferno	Nano	0.668	0.401
	Small	0.725	0.453
	Medium	0.748	0.482

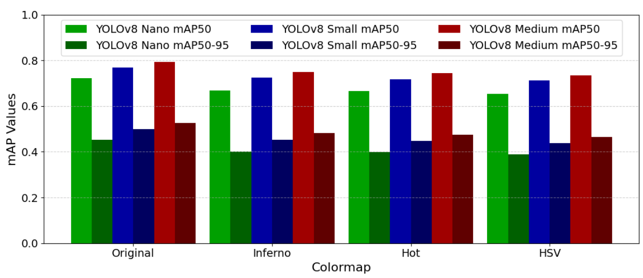

Fig. 7: Performances for trained models with different architectures under different colorizing schemes for thermal images

impact of colorization on model performance, We trained each model architecture (Nano, Small, Medium) on three more datasets constructed from the following colormappings: HSV, Hot, and Inferno (Fig. 4 and Fig. 5). As explained in Section II, these datasets are created from the original thermal images using colorization schemes from the OpenCV library.

DareDevil achieves high mAPs across all model backbones and colormaps; as shown in Table I and Fig. 7. Medium backbone models generally achieve higher accuracy than Small models, which in turn outperform the Nano ones. The mAP gap between Nano and Small is up to 6%, while the gap between Small and Medium is up to 3%; across all colormaps. This pattern is typical in YOLOv8 models, which are the base for DareDevil's trained models. Larger YOLOv8 models can learn richer and more complex data representations, handle object classification with more precision, and generalize better to diverse or challenging scenarios.

The mAPs of DareDevil's thermal models are higher than those of the original state-of-the-art YOLOv8 RGB-based models. Based on previous YOLOv8 benchmarks on other datasets (e.g. COCO), YOLOv8 achieves mAPs of 0.37, 0.44, and 0.5 for Nano, Small, and Medium architectures; respectively [26]. The higher performance of DareDevil's models is generally due to label binarization. Even though YOLOv8's mAP scores might seem lower compared to some other state-of-the-art models (like the latest versions of EfficientDet, Faster R-CNN, or Transformer-based detectors), those mAPs are generally considered "high enough" for YOLO. This is especially true given its design goals and trade-offs. It achieves a high enough mAP relative to its real-time capabilities. Hence, it is considered excellent for applications where speed, efficiency, and moderate accuracy are more important than pushing the absolute highest precision boundaries. YOLOv8 provides the best balance between real-world usability and accuracy for many use cases.

TABLE II: Original YOLOv8 models Performance

Model	Inference time (ms)	mAP50	mAP50-95
Nano	1.7	0.606	0.372
Small	2.6	0.671	0.423
Medium	5.1	0.704	0.456

The results highlight the significant impact thermal-image colorization can have on model performance, an aspect that has been largely overlooked in previous research. While much research utilizes other colormaps [23, 31, 32], the original grayscale colormap resulted in a higher accuracy across all model backbones. It achieves an average mAP increase of over 5%, compared to other colormaps. This scheme provides a more straightforward representation of thermal images, facilitating better feature extraction, clarity, and thus model performance; compared to other schemes. Inferno-based model performance is lower than that of Hot, which is in turn lower than HSV across all backbones. Hence, the most significant performance gap, 6.1% is between Original and HSV. Inferno and Hot use a more intuitive gradient that emphasizes warmer temperatures, which may highlight features more effectively for obstacle detection. However, HSV can introduce more variations in color perception, potentially making it harder for the model to differentiate between similar thermal intensities. These insights are critical for future research as it demonstrates the importance of selecting appropriate colorization techniques to maximize the efficacy of thermal-based vision systems.

To compare DareDevil's thermal-based models to a baseline, we train an RGB model for each architecture (Nano, Small, Medium); using the RGB images that correspond to the ones in our thermal dataset. These RGB images are extracted from the RGB video captured using the dual camera.

Compared to DareDevil's RGB-based models, DareDevil's thermal-based models achieve higher mAPs and comparable inference time (Table II). Generally, the thermal-based model performance on colorized thermal images is better than the RGB-based model performance on corresponding original RGB images. Nano RGB-based model is 1.8% off from the Nano-HSV model while it is 8% off from the Nano-Original model in terms of mAP50-95. The medium RGB-based model underperforms Medium-HSV and Medium-Original by 0.8% and 7.1%; respectively. Although the performance gaps are slightly larger for mAP50 compared to mAP50-95, it shows a similar performance pattern for the models. The reason for thermal-based models' outperformance is that thermal representation emphasizes key features, reduces visual noise, and provides more consistent object contrast, particularly in challenging environments. This allows the model to focus on more relevant information for obstacle detection. Regarding inference time, the performance is highly comparable to the RGB-based model, indicating that the constructed thermal-based models do not induce additional latencies. The inference time per frame for the different models is as follows: nano 0.99 ms, small 1.20 ms, and medium 1.83 ms. This low inference time is crucial for providing real-time obstacle notifications, ensuring safer navigation, particularly for PVI.

IV. Conclusion and Future Work

In this paper, we introduce DareDevil, a deployable indoor navigation system that ensures safe navigation while preserving user privacy. DareDevil offers three main components: a comprehensive navigation dataset, a real-time thermal-based obstacle detection model, and a thermal colorization module that feeds thermal data to the model during inference. The thermal-based dataset supports seamless navigation in both indoor and outdoor environments. The system provides trained models of varying sizes; making it compatible with a wide range of high-end IoT devices and eliminating the need for bulky or expensive processing units. We evaluate DareDevil across different model sizes and thermal image colorization schemes. Our results show that DareDevil's models can achieve a mean Average Precision (mAP) of up to 79.3%, which is higher than the original state-of-the-art RGB-based YOLO architectures. Our findings support the development of an accurate, real-time, and privacy-preserving thermal-based navigation solution, that addresses a significant gap in current research. In the future, we plan to expand the dataset with more objects, integrate this obstacle detection system with a localization system, and conduct extensive real-world testing and validation of the proposed system.

References

[1] Aliaa Essameldin, Mohammed Nurul Hoque, and Khaled A Harras. More than the sum of its things: Resource sharing across iots at the edge. In *Proc. IEEE/ACM Symp. Edge Comput.*, pages 206–219, 2020.

[2] Ahmed Saeed, Ahmed Abdelkader, Mouhyemen Khan, Azin Neishaboori, Khaled A Harras, and Amr Mohamed. On realistic target coverage by autonomous drones. *ACM Transactions on Sensor Netw. (TOSN)*, 15(3):1–33, 2019.

[3] Muhammad A Shah, Khaled A Harras, and Bhiksha Raj. Sherlock: A crowd-sourced system for automatic tagging of indoor floor plans. In *IEEE MASS*, pages 594–602, 2020.

[4] Omar Hashem, Khaled A Harras, and Moustafa Youssef. Deepnar: Robust time-based sub-meter indoor localization using deep learning. In *IEEE SECON*, pages 1–9, 2020.

[5] Omar Hashem, Moustafa Youssef, and Khaled A Harras. Winar: Rtt-based sub-meter indoor localization using commercial devices. In *IEEE PerCom*, pages 1–10, 2020.

[6] Ammar Karkour, Khaled A. Harras, and Eduardo Feo-Flushing. Text2Map: From navigational instructions to graph-based indoor map representations using LLMs. In *IEEE IROS*, October 2024.

[7] P. Xu, G. A. Kennedy, F.-Y. Zhao, W.-J. Zhang, and R. Van Schyndel. Wearable obstacle avoidance electronic travel aids for blind and visually impaired individuals: A systematic review. *IEEE Access*, 11:66587–66613, 2023.

[8] Mary Beatrice Dias, Ermine A Teves, George J Zimmerman, Hend K Gedawy, Sarah M Belousov, and M Bernardine Dias. Indoor navigation challenges for visually impaired people. *Indoor wayfinding and navigation*, pages 141–164, 2015.

[9] Jiaming Zhang, Kailun Yang, Angela Constantinescu, and et al. Trans4trans: Efficient transformer for transparent object and semantic scene segmentation in real-world navigation assistance. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):19173–19186, 2022.

[10] Nasrin Bayat, Jong-Hwan Kim, Renoa Choudhury, and et al. Vision transformer customized for environment detection and collision prediction to assist the visually impaired. *Journal of Imaging*, 9(8), 2023.

[11] Zongshun Wang, Ce Li, Jialin Ma, and et al. Pvi-net: Point–voxel–image fusion for semantic segmentation of point clouds in large-scale autonomous driving scenarios. *Information*, 15:148, 03 2024.

[12] Mostafa Elgendy and Cecilia Sik Lanyi. Helping people with visual impairments to avoid obstacles using deep learning. In *Proceedings of Sixth International Congress on Information and Communication Technology*, pages 909–917, Singapore, 2022. Springer Singapore.

[13] Yasha Iravantchi, Thomas Krolikowski, William Wang, Kang G. Shin, and Alanson Sample. Privacylens: On-device pii removal from rgb images using thermally-enhanced sensing. *Proceedings on Privacy Enhancing Technologies*, 2024(4):872–891, October 2024.

[14] A. N. Wilson, Khushi Anil Gupta, Balu Harshavardan Koduru, and Abhinav et.al. Kumar. Recent advances in thermal imaging and its applications using machine learning: A review. *IEEE Sensors Journal*, 23(4):3395–3407, 2023.

[15] Rita Tse, Tianchen Wang, Marcus Im, and Giovanni Pau. Privacy aware crowd-counting using thermal cameras. In *ICDIP 2020*, volume 11519.

[16] My Kieu, Andrew D. Bagdanov, Marco Bertini, and Alberto Del Bimbo. Domain adaptation for privacy-preserving pedestrian detection in thermal imagery. In *Image Analysis and Processing – ICIAP 2019*, pages 203–213, Cham, 2019. Springer International Publishing.

[17] A. Naser, A. Lotfi, and J. Zhong. Calibration of low-resolution thermal imaging for human monitoring applications. *IEEE Sens. Lett.*, 6(3):1–4, 2022.

[18] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and Inso Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. volume 20, 06 2015.

[19] Alejandro Gonzalez Alzate, Zhijie Fang, Yainuvis Socarras, and et al. Pedestrian detection at day/night time with visible and fir cameras: A comparison. *Sensors*, 16:820, 06 2016.

[20] Xinyu Jia, Chuang Zhu, Minzhen Li, and et al. Llvip: A visible-infrared paired dataset for low-light vision, 2023.

[21] Christian Stippel, Thomas Heitzinger, and Martin Kampel. A trimodal dataset: Rgb, thermal, and depth for human segmentation and temporal action detection. In Ullrich Köthe and Carsten Rother, editors, *Pattern Recognition*, pages 18–33, Cham, 2024. Springer Nature Switzerland.

[22] Peize Li, Kaiwen Cai, Muhamad Risqi U. Saputra, and et al. Odombeyondvision: An indoor multi-modal multi-platform odometry dataset beyond the visible spectrum, 2022.

[23] Rohan Ippalapally, Sri Harsha Mudumba, Meghana Adkay, and Nandi Vardhan R. Object detection using thermal imaging. In *INDICON*, pages 1–6.

[24] Zhengqiang Shao, Longbin Yan, Jie Chen, and Jingdong Chen. A dataset and a lightweight object detection network for thermal image-based home surveillance. In *APSIPA ASC*, pages 1332–1336, 2022.

[25] Fabio Martinelli, Francesco Mercaldo, and Antonella Santone. Object detection and localisation in thermal images by means of uav/drone. *Procedia Computer Science*, 225:2234–2243. KES 2023.

[26] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023.

[27] https://github.com/ericluvpython/daredevil.

[28] FLIR Systems, Inc. FLIR ADAS Dataset, 2019.

[29] Rana Sabah Naser, Meng Chun Lam, Faizan Qamar, and BB Zaidan. Smartphone-based indoor localization systems: A systematic literature review. *Electronics*, 12(8):1814, 2023.

[30] Martin Brenner, Napoleon H. Reyes, Teo Susnjak, and Andre L. C. Barczak. Rgb-d and thermal sensor fusion: A systematic literature review. *IEEE Access*, 11:82410–82442, 2023.

[31] Sayantan Nath and Cvanca Mala. Thermal image processing-based intelligent technique for object detection. *Signal, Image and Video Processing*, 16, 09 2022.

[32] Farzeen Munir, Shoaib Azam, Muhammd Aasim Rafique, and et al. Exploring thermal images for object detection in underexposure regions for autonomous driving. *Applied Soft Computing*, 121:108793, 2022.

[33] Frederik S. Leira, Håkon Hagen Helgesen, Tor Arne Johansen, and Thor I. Fossen. Object detection, recognition, and tracking from uavs using a thermal camera. *Journal of Field Robotics*, 38(2):242–267, 2021.

[34] Juan Terven, Diana-Margarita Córdova-Esparza, and Julio-Alejandro Romero-González. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction*, 5(4):1680–1716, 2023.

[35] Muhammad Hussain. Yolo-v1 to yolo-v8, the rise of yolo and its complementary nature toward digital manufacturing and industrial defect detection. *Machines*, 11(7), 2023.