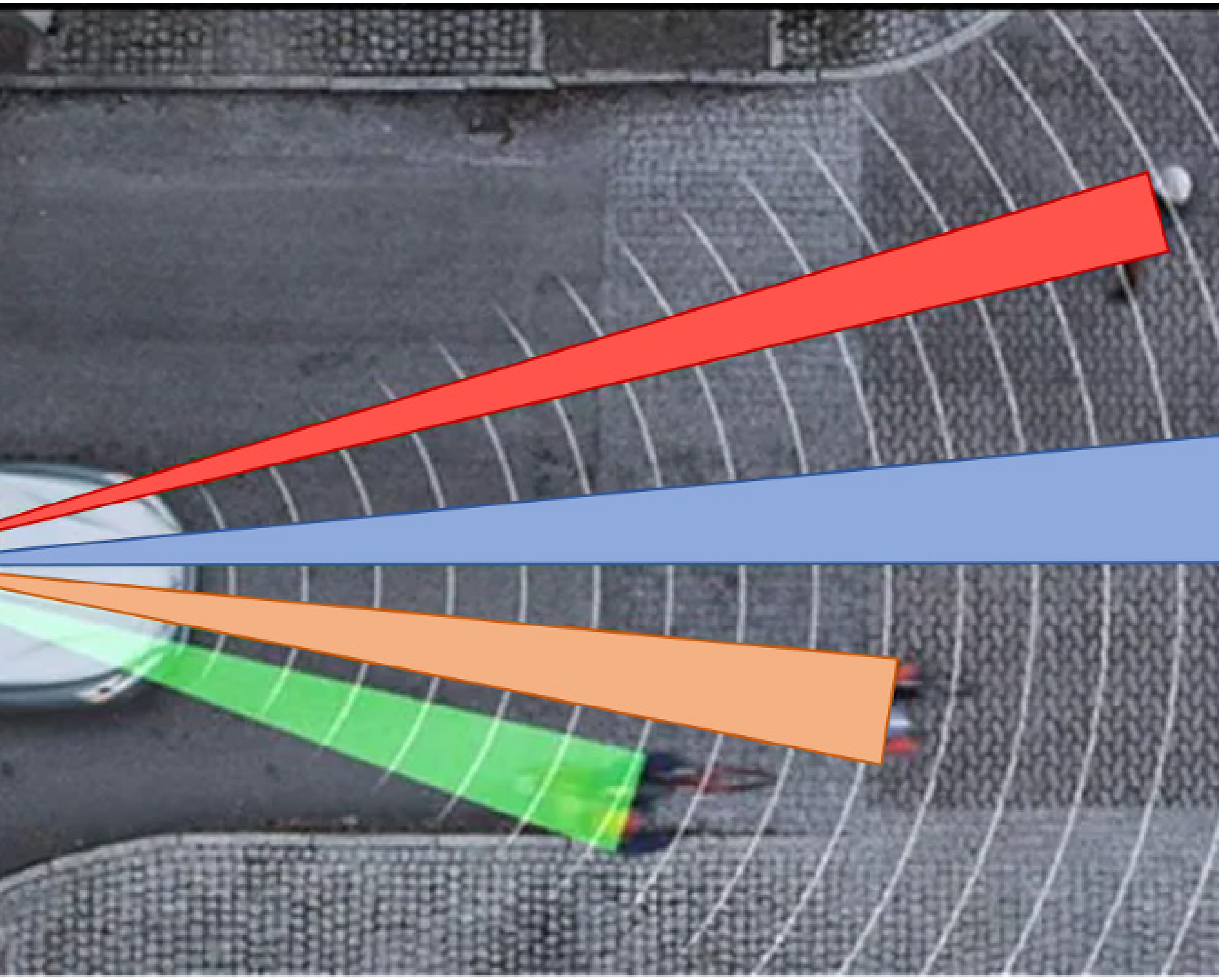
Project 2127 | November 2022

SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

AI-based Pedestrian Detection and Avoidance at Night Using an IR Camera, Radar, and a Video Camera

Hovannes Kulhandjian, PhD



CSUTC California State University Transportation Consortium

FRESNO STATE Discovery. Diversity. Distinction.

Mineta Transportation Institute

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the Mineta Consortium for Transportation Mobility (MCTM) funded by the U.S. Department of Transportation and the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 22-53

AI-based Pedestrian Detection and Avoidance at Night Using an IR Camera, Radar, and a Video Camera

Hovannes Kulhandjian

November 2022

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
22-53		

4. Title and Subtitle	5. Report Date
AI-based Pedestrian Detection and Avoidance at Night using an IR Camera, Radar, and a Video Camera	November 2022
	6. Performing Organization Code
7. Authors Hovannes Kulhandjian, PhD: 0000-0002-8983-6757	**8. Performing Organization Report** CA-MTI-2127
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219	**10. Work Unit No.** **11. Contract or Grant No.** ZSB12017-SJAUX
12. Sponsoring Agency Name and Address State of California SB1 2017/2018 Trustees of the California State University Sponsored Programs Administration 401 Golden Shore, 5th Long Beach, CA 90802	**13. Type of Report and Period Covered** **14. Sponsoring Agency Code**

15. Supplemental Notes

16. Abstract

In 2019, the United States experienced more than 6,500 pedestrian fatalities involving motor vehicles which resulted in a 67% rise in nighttime pedestrian fatalities and only a 10% rise in daytime pedestrian fatalities. In an effort to reduce fatalities, this research developed a pedestrian detection and alert system through the application of a visual camera, infrared camera, and radar sensors combined with machine learning. The research team designed the system concept to achieve a high level of accuracy in pedestrian detection and avoidance during both the day and at night to avoid potentially fatal accidents involving pedestrians crossing a street. The working prototype of pedestrian detection and collision avoidance can be installed in present-day vehicles, with the visible camera used to detect pedestrians during the day and the infrared camera to detect pedestrians primarily during the night as well as at high glare from the sun during the day. The radar sensor is also used to detect the presence of a pedestrian and calculate their range and direction of motion relative to the vehicle. Through data fusion and deep learning, the ability to quickly analyze and classify a pedestrian's presence at all times in a real-time monitoring system is achieved. The system can also be extended to cyclist and animal detection and avoidance, and could be deployed in an autonomous vehicle to assist in automatic braking systems (ABS).

17. Key Words	18. Distribution Statement
Intelligent transportation systems (ITS), Pedestrian detection, Applied machine learning, Accident prevention, Advanced driver assistance systems (ADAS)	No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	40	

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2022.2127

Mineta Transportation Institute

College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2127

ACKNOWLEDGMENTS

This study was supported by both the CSU Transportation Consortium and the Fresno State Transportation Institute. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the author and do not necessarily reflect the views of these institutes. The author would like to thank his student assistants, Jeremiah Barron, Megan Tamiyasu, and Mateo Thompson, undergraduate students who all graduated from the Electrical and Computer Engineering (ECE) department at California State University, Fresno for working on this interesting research project and implementing and testing the final proposed prototype. The author would also like to thank Editing Press for editorial services, as well as MTI staff, including Executive Director Karen Philbrick, PhD, Deputy Executive Director Hilary Nixon, PhD, Graphic Designer Alverina Eka Weinardy, and Executive Administrative Assistant Jill Carter.

CONTENTS

LIST OF FIGURES

Executive Summary

Pedestrian fatalities have surged in the United States over the past decade. During this 10-year period, pedestrian fatalities have increased by 46%, from 4,302 deaths in 2010 to an estimated 6,301 in 2019. The number of pedestrian fatalities at night grew by 54%, while daytime pedestrian fatalities climbed by just 16% during that same decade (NHTSA 2021). Of these fatal accidents, about 75% of them occurred after dark (Feese 2020). In addition, an American Automobile Association (AAA) research study that tested pedestrian detection in current vehicles found that the evaluated pedestrian detection systems that consisted of radar (radio detection and ranging), image sensors (camera), LIDAR (light detection and ranging), and ultrasonic sonar were ineffective during nighttime conditions (Edmonds 2019). The goal of this project is to reduce the number of nighttime pedestrian fatalities by combining data acquired from three separate sensors in real-time and using machine learning algorithms to detect pedestrians at night to alert the driver of the possibility of a collision with a pedestrian. The proposed project also has applications in autonomous vehicles where a signal can be developed to engage the automatic braking system if necessary. This project focuses on AI-based pedestrian detection and avoidance at night using an infrared (IR) camera, an RGB (red, green, and blue) video camera, and micro-Doppler RADAR. Specifically, this project will use machine learning with deep learning algorithms to detect humans at night and data fuse the information to alert the driver of a possible accident with a pedestrian. One possible solution is to use a video camera, a radar system, or a LIDAR system in a vehicle. More recently, the advancement of thermal IR cameras has shown a potential possible solution. The research on pedestrian detection and avoidance is still in its infancy. Several methods have been explored to detect a pedestrian and avoid an accident. The main contribution of this research work lies in utilizing three different sensors (i.e., a thermal infrared camera, radar sensor, and a visible camera) combined with advanced machine learning (ML) for pedestrian detection and avoidance. Therefore, we believe that this research could lead to new artificial intelligence-based application tools for drivers that can save lives. We will explore state-of-the-art ML techniques combined with data fusion to achieve this objective. The goal of this research work is to maximize pedestrian detection, especially at night, by effectively data fusing the information gathered from a thermal camera, a radar sensor, and a video camera along with the use of advanced machine learning algorithms to detect and avoid pedestrian collision in real-time. Using this multi-dimensional data, the system can make intelligent decisions during different conditions of the road, both during the day or at night. The proposed system can be embedded into a smart vehicle system that provides real-time pedestrian detection and alerting mechanisms by vibrating the driver's wheel and displaying a message on a monitor/dashboard to warn the driver of an incoming pedestrian. The developed system can be used both during the day and at night using a combination of a thermal infrared camera, a radar system, and a video camera. The system can also be installed in an autonomous vehicle.

1. Introduction

On average, a pedestrian is killed every 88 minutes in traffic crashes in the United States. That is more than 16 people a day, almost 115 people a week. The traffic fatalities for the first nine months of 2020 shows that an estimated 28,190 people died in motor vehicle traffic crashes, which is a 4.5% increase compared to 2019 (NHTSA 2021). In 2019, vehicle accidents in the United States killed more than 6,500 pedestrians, the highest annual total ever recorded, and sent more than 100,000 to hospitals with injuries (Feese 2020). Additionally, 75% percent of pedestrian fatalities occurred in the dark as compared to daylight (21%), dusk (2%), and dawn (2%).

A recent study published by U.S. News & World Report found that roughly 430 Californian pedestrians were killed in the first six months of 2018 (Galvin 2020). Comparing traffic data from the first half of 2021 to the first half of 2020 reveals that there is a 40% increase in the number of pedestrian deaths due to car crashes (Romero 2022). Fresno is not exempt from California's problems with pedestrian safety. According to a 2018 police report, 64% of fatal crashes in Fresno involved both a pedestrian and a vehicle (Valera 2018).

A research study conducted by the Volpe National Transportation Systems Center suggests that automatic emergency braking systems with pedestrian detection functionality could reduce up to 5,000 annual vehicle-pedestrian crashes and 810 fatal vehicle-pedestrian crashes (Yanagisawa et al. 2017). Pedestrian detection systems with automatic braking functionality have the potential to prevent or reduce the severity of collisions that result in property damage, personal injury, and/or death.

One possible solution is to use a video camera, a radar system, or a LIDAR system in a vehicle. More recently, the advancement of thermal infrared (IR) cameras has shown a potential possible solution. The research on pedestrian detection and avoidance is still in its infancy. Several methods have been explored to detect a pedestrian and avoid an accident (Tubaishat et al.; El-Faouzi et al.; Lim et al.; Cao et al.; Datondji et al.,). To the best of our knowledge, no prior research work has explored or experimented with using data fusion from multiple sensors (i.e., a thermal infrared camera, radar sensor, and a visible camera) combined with advanced machine learning (ML) for pedestrian detection and avoidance. Therefore, we believe that this research could lead to new artificial intelligence-based application tools for drivers that could potentially save lives. In this study, we will explore state-of-the-art ML techniques combined with data fusion to maximize pedestrian detection, especially at night. By effectively data fusing the information gathered from a thermal camera, a radar sensor, and a video camera along with the use of advanced machine learning algorithms, we can detect and avoid pedestrian collisions in real-time. Using this multi-dimensional data could allow for intelligent decisions during different conditions of the road, be it during the day or at night. The proposed system can be integrated into a smart car system to provide real-time pedestrian detection and alerting mechanisms. These mechanisms vibrate the driver's steering wheel and display a message on a monitor or dashboard to warn the driver to avoid hitting the pedestrian. The designed system combines a thermal infrared camera, a radar system, and a video camera to be used both during the day and at night. Additionally, an autonomous car might use it.

2. System Overview

2.1 System Design

This system is designed to inform drivers in the event a pedestrian crosses the street and is in danger. The system utilizes machine learning combined with a video camera, an IR camera, and a radar sensor to make an accurate assessment of road conditions, and determine if a pedestrian is in proximity to the vehicle. The video camera captures RGB (red, green, and blue) images. The system installed in a vehicle starts with the video camera, IR camera, and the radar sensor. The three sensors simultaneously scan the road conditions in front of the vehicle and gather three different types of data. The video camera, IR camera, and radar sensor all work hand in hand, and are responsible for capturing sample images and radar data in front of the vehicle. The two cameras on the vehicle are strategically placed so that they capture a wide view in front of the vehicle. Once a pedestrian is detected in the captured images, the images are automatically cropped by running a computer script, and are formatted to serve as the input to their respective deep convolutional neural network models, which is a machine learning algorithm that will be used to train them. The radar sensor is mainly used to detect the pedestrian's distance from the vehicle, their direction of motion, and their speed. After having being trained with the labeled trained data using the RGB camera, IR camera images, and the micro-Doppler signal, the deep convolutional neural networks algorithm will make predictions corresponding to the driver's behaviors captured in the sample images and the radar signals. The predictions made are then used to detect the presence and distance of a pedestrian. The deep convolutional neural network algorithm runs on a Raspberry Pi paired with a Coral USB Accelerator to reduce inference times and real-time processing time. The detections are made continuously over a sample period of 60 seconds. Once each sample period has elapsed, the cumulative detections of pedestrians in front of the vehicle using the three different sensors are performed. The three different prediction models captured by the sensors are displayed to the driver through the touch screen display in real-time. If the system detects the presence of a pedestrian in front of the vehicle that might be in danger, a signal is sent to a vibration motor connected to the steering wheel of the vehicle. The vibration of the steering wheel alerts the driver and deters them from crashing into the pedestrian.

Figure 1 shows an overview of the pedestrian detection hardware design. The RGB camera, FLIR IR camera, and Doppler radar first capture images and radar signals in front of the vehicle. Then, the pedestrian detection algorithm is ran and once it detects a pedestrian, it displays an alert message to the driver and vibrates the steering wheel to warn them.

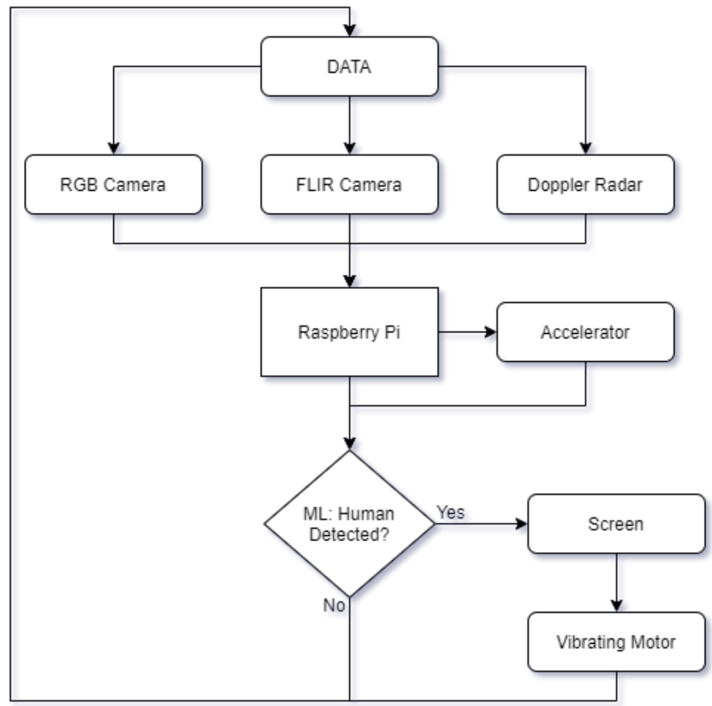


Figure 1. Block Diagram of the Pedestrian Detection Hardware Design

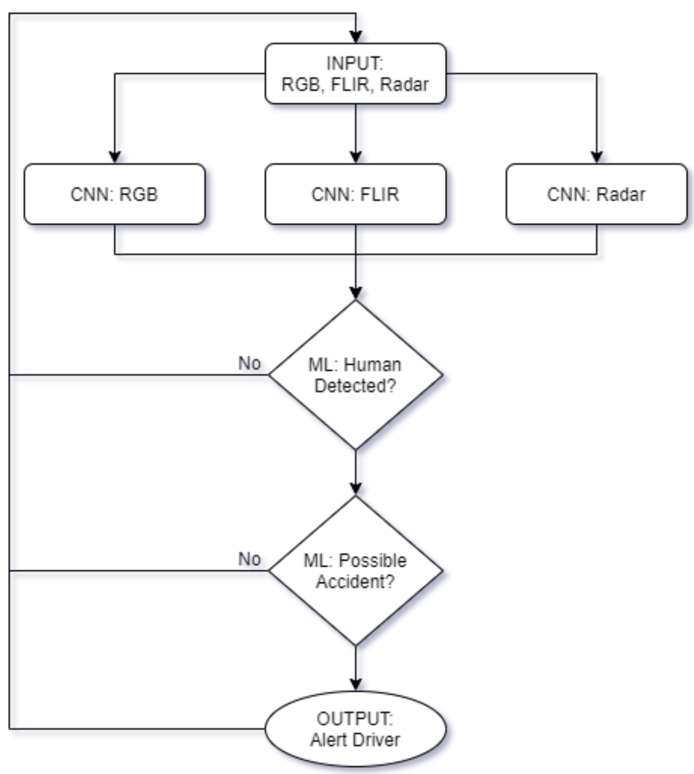


Figure 2. Block Diagram of the Pedestrian Detection Software Design

Figure 2 shows an overview of the pedestrian detection software flow diagram. Initially, the sensor information, RGB camera images, and the FLIR camera images are fed into the convolutional neural network (CNN) ML architecture to train the different models. Specifically, TensorFlow Lite, an open-source deep learning framework for on-device inference, is used to train the deep convolutional neural network. Once the models are trained, they can be transferred on to a computer on a chip machine and detect a pedestrian in real time. By running the pre-trained ML models as the real time data is gathered by the three sensors, they are continuously fed into the ML algorithms to determine if a pedestrian is detected. If a pedestrian is detected, the system will alert the driver. If not, then it will keep gathering more data in the next time slot and continue to test the new gathered images through the pre-trained ML algorithms.

The main equipment used is a video camera, a FLIR IR camera, a radar sensor, a microcomputer, a vehicle, and an alert system. The alert system consists of some type of notification to the driver utilizing a vibration motor attached to the steering wheel. The machine-learning model will initially be trained using MATLAB, which will then be transferred to Python to be used with Tensorflow. The sensors will be mounted on the vehicle with some inside the vehicle, and others on the exterior.

2.2 Deep Convolutional Neural Network Design

Deep learning is a branch of machine learning in artificial intelligence whose algorithms use artificial neural networks (ANN) that replicate the functionality of a brain. ANN is made up of layers of artificial nodes that carry raw input data through each layer to the final output layer. These neural networks are powerful in decision making and are able to learn from unstructured data. This project created a deep convolutional neural network (DCNN) model, as DCNNs are most commonly used for image classification. The architecture of a DCNN algorithm implemented in this project is shown in Figures 3 and 4 for the RGB image input and IR camera image input, respectively.

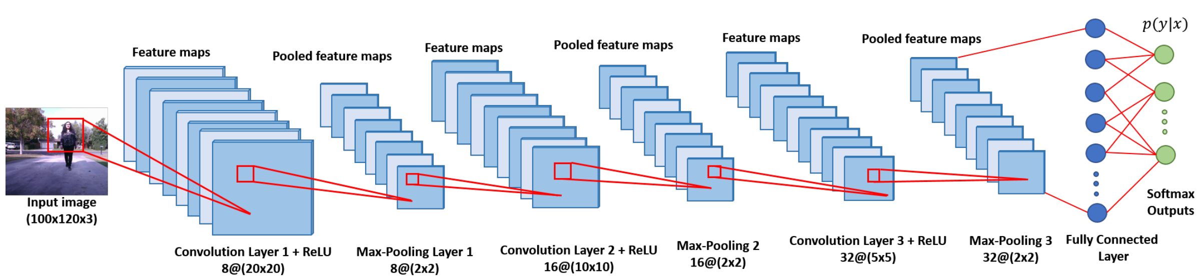


Figure 3. Convolutional Neural Network Architecture for the RGB Image Input

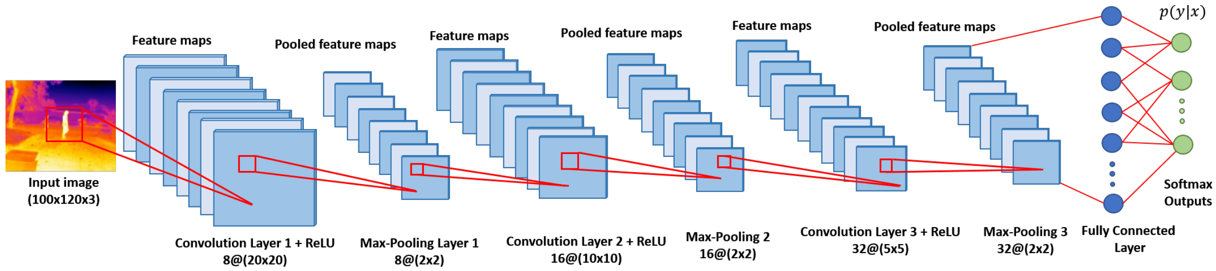


Figure 4. Convolutional Neural Network Architecture for the IR Camera Image Input

First, the captured RGB and IR camera images are manually cropped to 100 × 120 pixel RGB images. The input images undergo feature extraction network by first being processed by the convolution layer consisting of eight convolution filters of size 20 × 20 pixels. The output from the convolution layer goes through the rectified linear unit (ReLU) function followed by the pooling layer, which employs a max pooling process of 2 × 2 matrices. This process is repeated several times to create the output and train the machine with inherent features of the image. The output of the pooling layer is fed into a second convolution layer consisting of 16 convolution filters sized 10 × 10 pixels. Similarly, after passing the output through the ReLU function, it undergoes the pooling layer with a max pooling size of 2 × 2 matrices. Finally, it is passed through a third round of convolution layer consisting of 32 convolution filters of size 5 × 5 after which it is processed by the ReLU function and the pooling layer with max pooling size of 2 × 2 matrices. The max pooling concept is demonstrated in Figure 5.

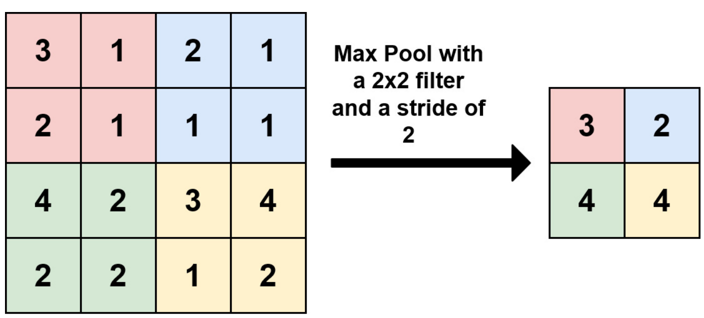


Figure 5. Max Pooling Principle

The stride is the sliding window operation, which is used in the convolution layer and in the max pooling operation, in which case the stride is 2 (Kulhandjian et al., ACM WUWNet, 2019). Supposing n × n convolution is performed, then the stride represents the movement by S elements with every step. If the stride is defined as 1, then the convolution layer will move with a sliding window of one pixel and move every third pixel by skipping the second pixel. Max pooling, shown in Figure 4, is a down sampling process where the maximum value from each view is selected (Kulhandjian et al., IEEE GLOBECOM, 2019). Since the video and spectrogram images contain sharp edges, max pooling, instead of average pooling, is used to extracts the most important features, such as edges. The classifier network consists of a fully connected layer comprised of 100 hidden nodes, which produce a softmax output that, in turn, is used for classifying the driver's status. The output layer of the DCNN represents the probability distribution containing the probabilities that each class is assigned in accordance to the input images. Using a maximum ratio combining the three gathered sensor data, once the algorithm detects a pedestrian, it will send an alert to the driver. If no pedestrian is detected it will continue to capture new images and radar data and pass it to the algorithm to perform further pedestrian detection.

In order to obtain reliable results regarding the system's capabilities of capturing rapid changes in biometric data, the deep convolutional neural network model was converted into a model capable of running on the Raspberry Pi coupled with the Coral USB Accelerator. The model conversion process is shown in Figure 6, in which the model must be compiled to run on the Edge Tensor Processing Unit (TPU), which is an artificial intelligence (AI) accelerator application-specific integrated circuit (ASIC) developed by Google specifically for neural network machine learning, particularly using Google's own TensorFlow software.

First, the TensorFlow model is converted into a TensorFlow Lite model. TensorFlow Lite models are scaled down in size, in terms of data representation, and they produce faster inference times. TensorFlow Lite allows for models to run efficiently on low-power devices with limited memory resources. The converted parameters of the Tensorflow Lite model are represented using 32-bit floating point numbers. However, the Edge TPU requires 8-bit fixed point numbers. The 8-bit fixed point numbers are obtained in the next step of the conversion process, which is post-training quantization. The method of post-training quantization used in this project was called full integer quantization. Full integer quantization scales the model down again by four times and speeds up inferencing times by a factor of three. The last step in the model conversion process is compiling the quantized TensorFlow Lite model using the Edge TPU compiler. In this project, multiple models were compiled to further increase the model performance.

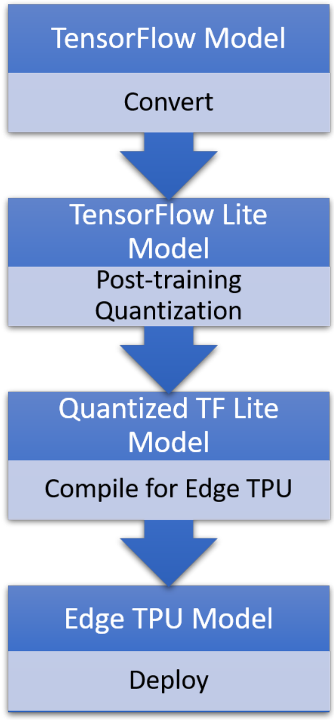


Figure 6. Machine Learning Model Conversion Process

3. Pedestrian Detection Using a Video Camera

3.1 RGB Camera

An ELP RGB camera, the USB4K02AF-KL100W, was used to gather data. An RGB camera is one that takes images that replicate standard human vision using red, blue, and green colors. The USB4K02AF-KL100W camera has a 4K resolution and is one of the better cameras to capture very clear images on the market. An array of LED lights around the camera automatically adjusts the brightness.

Figure 7. RGB Camera ELP USB4K02AF-KL100W

3.2 Data Collection using the RGB Camera

The RGB camera shown in Figure 7 was used to gather images for the dataset to configure a custom machine learning model and to detect a person while the custom machine learning model ran. Data collection with this camera was performed exclusively in the daylight since the nighttime data was captured by the FLIR IR camera, discussed in Section IV. Data was gathered at different distances; several sample images are shown in Figure 8. Bounding boxes were used on each image with a pedestrian to indicate their exact location to help the DCNN algorithm better detect and classify a pedestrian in an image. The number of images gathered and used in the final model was 1,200, 600 of which had pedestrians and 600 of which did not. In addition, we used 800 additional images from the FLIR dataset (Teledyne FLIR Thermal Dataset) (a sample image along with the bounding box is shown in Figure 9), of which 50% had pedestrians and 50% did not. We used a total of 2,000 images to train and experiment our ML model.

a) Sample Image 1 b) Sample Image 1 with a Bounding Box

c) Sample Image 2 d) Sample Image 3

Figure 8. RGB Camera Sample Images with a Bounding Box Image on the Top Right

Figure 9. Sample RGB Images from the FLIR Dataset with the Bounding Box on the Top Right

The data gathered was used to create a custom machine-learning model using a pre-trained network called Squeezenet. Analysis of the trained model indicated that the model detects the images with a high accuracy, as shown with the loss data and accuracy data in Figure 10. The loss graph determines the difference between the prediction and the actual value. This difference should be minimized as much as possible in a very accurate model. The accuracy graph determines how accurate the model is by comparing it to the test images used to validate the model. The accuracy data was gathered in MATLAB using the Squeezenet pre-trained network. Standard camera images from the FLIR dataset (Teledyne FLIR Thermal Dataset) were used to differentiate the person images when gathering the accuracy graphs in MATLAB. The training and creation of the custom model was done on a separate computer, and then transferred over to the Raspberry Pi 4 as a TFLite model. The Raspberry Pi 4 was used to test the functionality of the custom model in real-time.

3.3 Performance Results

After curating the dataset, the deep convolutional neural network was trained. A validation accuracy of 99.6% was achieved, as shown in Figure 10. The validation accuracy begins to plateau after nine epochs of training, indicating that the model has reached its peak performance. One epoch is when the entire training data set is passed forward and backward through the neural network once. Since the training data set is often limited, in practice, multiple epochs are utilized to allow the learning algorithm to run until the error from the model is sufficiently minimized.

The validation accuracy represents the accuracy the model can practically achieve when new samples are input into the model. The loss metric of the attained model describes how well the model responds to training after each iteration, or epoch. The loss optimizes the model so that the next prediction can be more accurate. Ideally, the loss will continue to decrease as training continues.

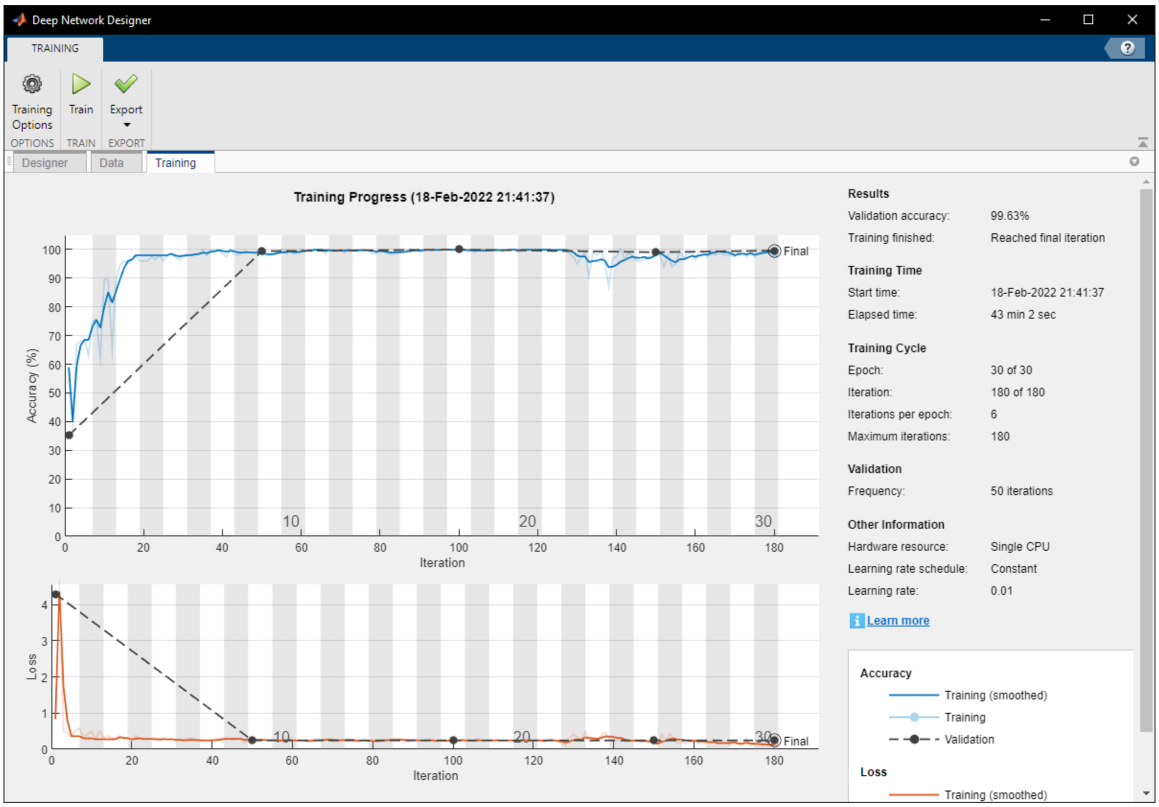


Figure 10. Accuracy and Loss Training Results using the RGB Camera

3.4 Limitations

Given the hardware used during the experimentations, there are some natural limitations. Due to the camera's resolution limitations, the farther away a pedestrian is the more the accuracy of detecting the pedestrian degrades. In order to improve this accuracy, a higher resolution camera with autofocus functionalities can be used to increase the captured image resolution at far distances.

4. Pedestrian Detection Using an IR camera

4.1 Infrared

Infrared is part of the electromagnetic spectrum with frequencies between 300 GHz–430 THz, which corresponds to wavelengths of 700 nm–1 mm. The part of the spectrum used by the camera in this project is the long-wave infrared of 8 μm–14 μm (FLIR 2018).

4.2 Data Collection using Infrared Camera

Data was gathered using the FLIR Lepton 3.5 camera that is smaller than a dime and is shown in Figure 11. Several examples of the images captured using the FLIR camera are shown in Figure 12. The gathered images were then labelled using LabelImg labelling software (Tzutalin 2015), a graphical image annotation tool written in Python that uses Qt for its graphical user interface to identify persons in the training and testing sets, as shown in Figure 13. All images were collected from the camera—no datasets were used in this study. Most images were taken at night (~85%), the rest were taken during the day.

After the images were collected, the DCNN model was trained using the pre-trained network SSD MobileNet V2 FPNLite. This network was chosen due to its relatively high speed and that the image size of this network is closest to the native resolution of the FLIR camera, so image distortion is kept to a minimum. Other networks were tried, the best results were obtained using MobileNet, a pre-trained machine learning algorithm, for the IR dataset.

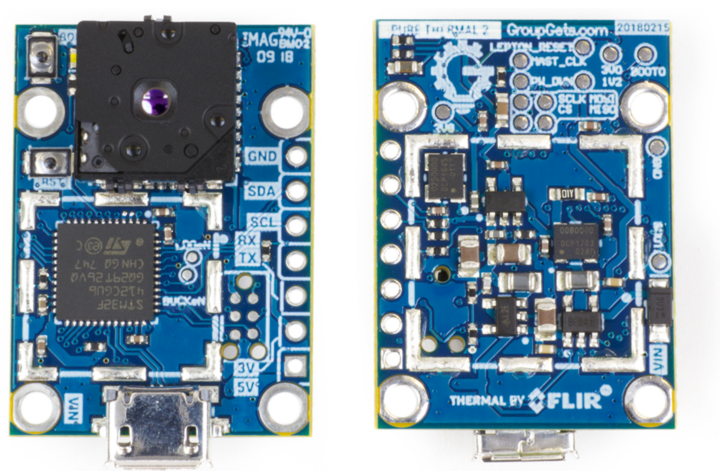


Figure 11. IR Camera: FLIR Lepton 3.5 with Pure Thermal 2 Breakout Board

Figure 12. IR Camera Sample Images Taken at Night

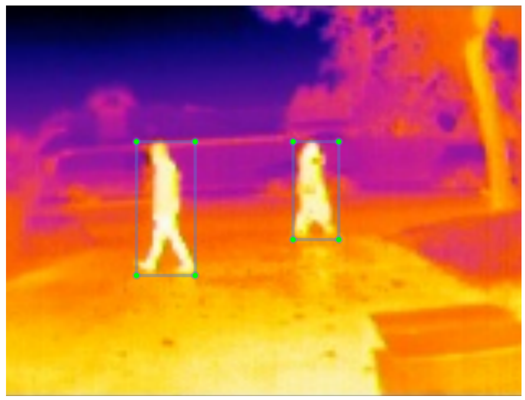


Figure 13. FLIR Camera: Object Detection (Image Labeler)

4.3. Performance Results

After curating the dataset, the deep convolutional neural network was trained. A validation accuracy of 97.26% was achieved, as shown in Figure 14. Validation accuracy began to plateau after four epochs of training. In the test data, the model successfully detected all images with humans within a 15-meter range of the vehicle. The accuracy quickly declined outside of this range when pedestrians were beyond 25 meters from the IR camera, and the algorithm began to degrade.

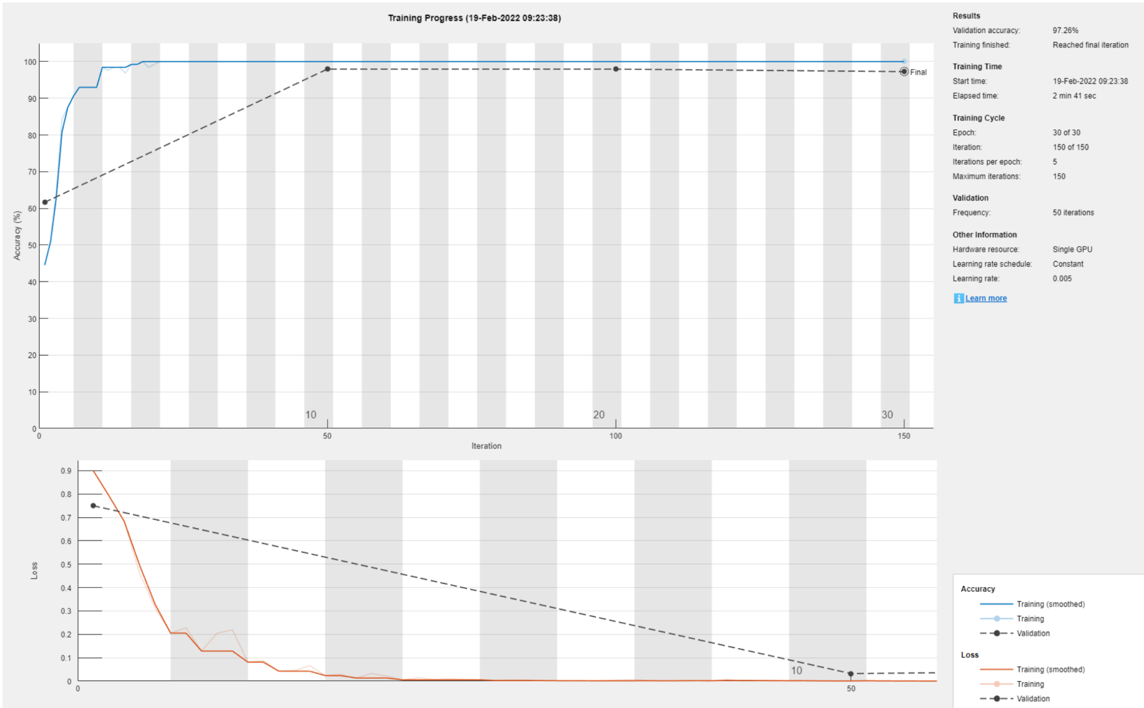


Figure 14. FLIR Camera DCNN Training and Validity Accuracy and Loss Plots

4.4 Limitations

The FLIR camera is limited by its relatively small resolution of 160 x 120 pixels. This low resolution results in poor image quality making it difficult for the training network to capture human features from the images for recognition. This is especially evident when pedestrians are farther away as they occupy fewer pixels, and generally appear as bright rectangles. This is believed to be the primary contributor to the model's poor performance when pedestrians are farther away. To address this in the future, a higher-resolution IR camera can be used, which would better detect pedestrians at farther distances.

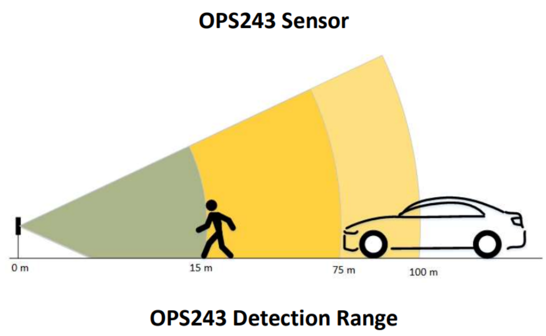
5. Pedestrian Detection Using Micro-Doppler Radar

5.1 Micro-Doppler Radar Setup

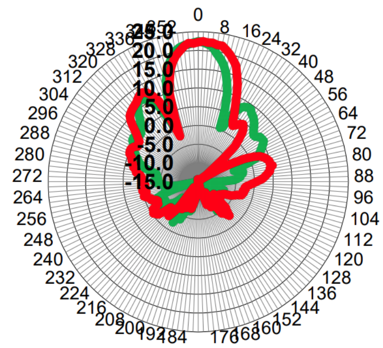
To capture radar data, OmniPreSense's OPS243-C sensor, a complete short-range radar (SRR) solution providing motion detection, speed, direction, and range reporting, was used and is shown in Figure 15. All radar signal processing is done on board of the sensor, and a simple application programming interface (API) reports the processed data. Flexible control over the reporting format, sample rate, and module power level is provided. The sensor is used to gather a pedestrian's speed, direction of motion, and distance from the vehicle. This single-board radar sensor can detect objects up to 60 meters away. The sensor is ideal for security, traffic monitoring, drone collision avoidance, robotics, and an Internet of Things (IoT) sensor applications. The captured signals from the micro-Doppler radar were further converted to spectrograms, which is a visual representation of the frequency variations of the received signal with respect to time.

Figure 15. OmiPreSence OPS24-C FMCW Doppler Sensor

Several features of the sensor include a 1 m–100 m detection range, speed reporting up to 348 mph, range reporting up to 60 m, and a narrow 20-degree beam width (-3 dB), as shown in the Figure 16b power pattern. The sensor operating frequency is 24 GHz–24.25 GHz on the industrial, scientific and medical (ISM) band. The OPS243 sensor outputs data over USB, UART, RS-232, or WiFi interfaces for simple connection to any embedded processor (Arduino, Raspberry Pi, PC) or directly to the cloud via WiFi.



a) Detection Range.



b) Power Pattern.

Figure 16. OmiPreSence OPS24-C FMCW Doppler Range and Power Pattern

Figures 16 show the radar sensor detection range capabilities on the left and the power pattern shown on the right. As we can see from the power pattern the radar sensor has a narrow 20-degree beam width at -3 dB power.

The radar sensor gathers data in the form of spectrograms, which can be used as an input to the DCNN algorithm to predict the driver's status. The vertical axis in the spectrogram plot shown in Figure 17 represents Doppler frequency variations in hertz while the horizontal axis denotes time in seconds. The redder parts of the image show that a lot of motion activity occurs compared to the blueish color code, indicating minimum variations or motion occurring in front of the radar sensor.

Figure 17 represents no pedestrian being detected (left) and a pedestrian being detected at 20 m (right), while Figure 18 represents a pedestrian being detected at 10 m (left) and at 5 m (right). As we can see, Figure 18's right side has more Doppler frequency variations since the pedestrian is at a closer proximity to the vehicle, but when the pedestrian is farther away at 20 m, less perturbations occur, as indicated in Figure 17 (right).

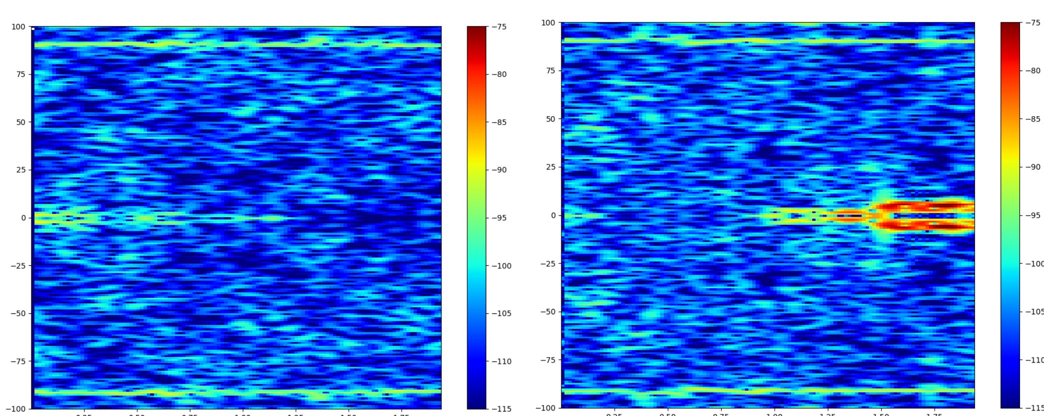


Figure 17. No Pedestrian Detected (Left), Pedestrian Detected at 20 m (Right)

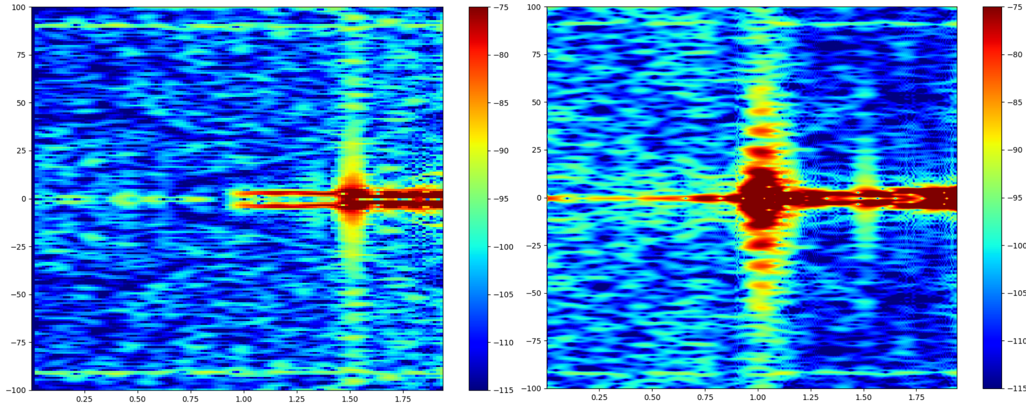


Figure 18. Pedestrian Detected at 10 m (Left) and 5 m (Right)

5.2 Experimental Results using the Radar Sensor

The OPS243-C FMCW Micro Doppler Radar is capable of not only measuring the speed of moving objects but also the distance from the Doppler. The FMCW Radar uses two antennae: one a transmit antenna that sends an FMCW signal; and another antenna that receives the echoed signal. The OPS243-C FMCW Doppler was configured through its onboard application programming interface (API). Data was collected and complied through a Python script that was executed on a Raspberry Pi 4. This data was then used to detect whether a human is in the proximity of the vehicle. Within a proximity of about 25 m, the cameras can be used to determine the possibility of an accident. Due to the limitations of the Micro Doppler found in testing, the OPS243-C can only pick up a human signature within tens of meters. This hindered our plan to use the Doppler. The goal of the Doppler was to generate spectrograms that would be fed into a convolutional neural network used for machine learning to read the frequency patterns of a car approaching a human. After finding the human detected range of tens of meters, the Raspberry Pi could only produce measurable spectrograms for an average of 2 seconds. This does not give the detection system enough time to respond. Instead, when capturing linear data and using the radar as a range finder, range depths were able to be taken instantly with no time delay. This range finder was used to detect an object within the detected human range of about 10 to 15 meters of the vehicle. All other objects found outside this range were ignored. If an object was detected within that range, the radar would set a flag high and allow the machine learning algorithms created for the RGB and FLIR cameras to be implemented. For example, for the system to determine whether a possible accident may occur with a pedestrian, the system operates in the following way: (1) All sensors gather the data and write it to a file to be read; (2) when the radar detects an object, it will set a flag high; (3) once this flag is set high, the two cameras' detection scores will be calculated with their corresponding weights of 75% and 25% for the FLIR and RGB cameras, respectively; (4) these detected scores are weighed and evaluated; (5) if the combined detection weights between the FLIR and RGB cameras are above 50%, then the system will output a signal to the vibrating motor to alert the driver of the possibility of an accident with a pedestrian.

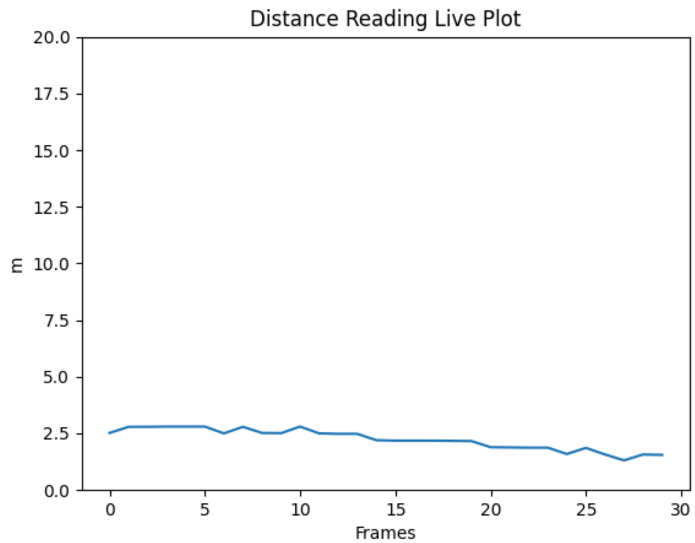


Figure 19. Pedestrian Range Detection at Different Frame Instances using the Micro-Doppler Sensor

Figure 19 shows pedestrian signature being detected within the human-detected range. The range measurements were used as additional information fed into the algorithm to provide information on how far the pedestrian is from the vehicle. These distance plot readings were used in the real experimental deployment instead of the spectrograms due to the ability of readings to take place in real time.

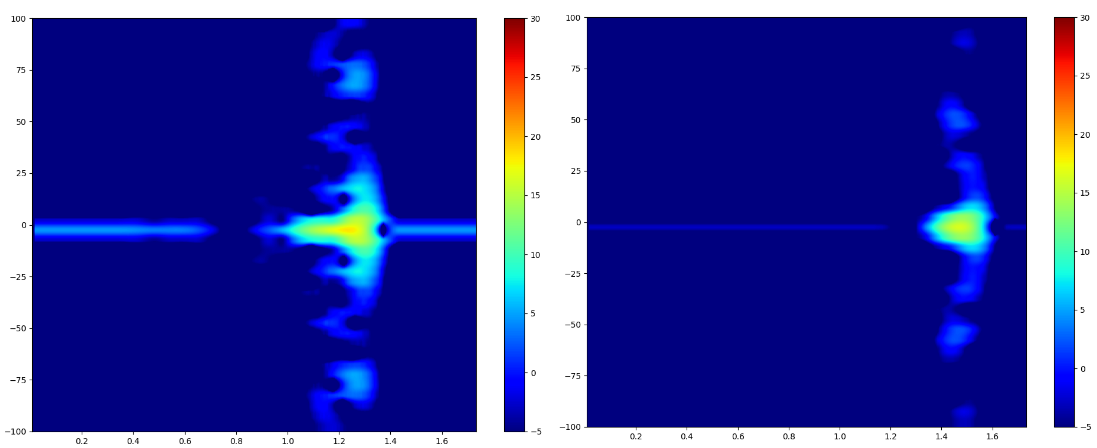


Figure 20. Spectrograms: Represents a Vehicle Approaching a Pedestrian at Low Speed (1–5 mph) (Left), a Vehicle Approaching Another Vehicle at Medium Speed (5–10 mph) (Right)

The spectrogram in Figure 20 (left) represents a vehicle approaching a human at low speeds (1–5 mph). One can see the concavity of the spectrogram pointing towards the left.

This phenomenon was recognized when the vehicle approached a human. The spectrogram on the (right) represents a vehicle approaching a human at medium speeds (5–10 mph) where the concavity of the spectrogram points toward the left. The same phenomenon was recognized when the vehicle approached humans at low speeds. This spectrogram captures less information due to the amount of time it can capture in the short amount of distance (0.5 m ~7.5 m).

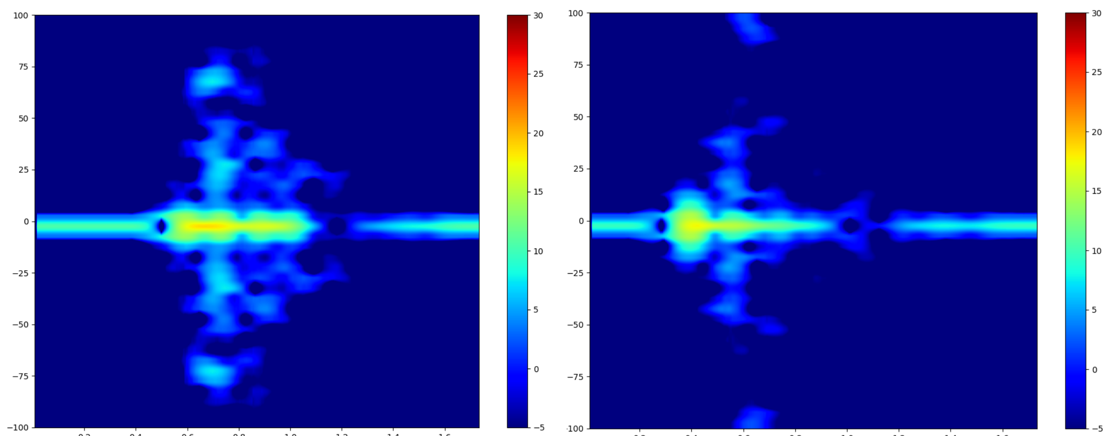


Figure 21. Spectrograms: Represents a Vehicle Approaching Another Vehicle (Left) and a Vehicle Traveling Away from Target Vehicle at Rest (Right)

The spectrogram in Figure 21 (left) represents a vehicle approaching another vehicle (10–15 mph). Here, the concavity of the spectrogram points towards the left, the same phenomenon recognized when the vehicle approached humans at low speeds. This spectrogram was able to capture more information compared to the human spectrograms because the micro-Doppler was able to capture vehicle movement within ~20 m. While the spectrogram on the (right) represents a vehicle traveling away from the target vehicle from rest (0–15 mph). You can see the concavity of the spectrogram pointing towards the right. This concavity was found when objects traveled away from the radar sensor. More information was captured in this spectrogram compared to the human spectrograms since the radar sensor was able to capture vehicle movement within ~20 m.

6. Prototype Experimentation

In this section, we evaluate the performance of the proposed pedestrian detection scheme, which was developed to be tested in a vehicle.

6.1 System Setup in a Vehicle

The experimental setup for the in-vehicle test is as follows. The Google Coral USB Accelerator used in the project adds an Edge TPU coprocessor to the system, enabling high-speed machine learning inferencing on a wide range of systems, simply by connecting it to a USB port. This on-device ML processing reduces latency, increases data privacy, and removes the need for a constant internet connection. The Raspberry Pi 4 Model B is used for processing real-time data with the machine-learning model incorporated into the Raspberry Pi. Some of the key features include a high-performance 64-bit quad-core processor, dual-display support at resolutions up to 4K via a pair of micro-HDMI ports, hardware video decoding at up to 4Kp60, up to 4GB of RAM (random access memory), dual-band 2.4/5.0 GHz wireless LAN, Bluetooth 5.0, Gigabit Ethernet, and USB 3.0.

Figure 22. Touchscreen Display for User Interface

Figure 22 shows the touch screen display attached to the car's dashboard, which provides real-time information to the driver on road status and pedestrian detection using the three sensors. It shows if a pedestrian is detected or not, and the distance measurements using the radar sensor. The system also has a warning message: if it detects a pedestrian, it blinks in red and activates the motor on the steering wheel to alert the driver.

6.2 Testbed Experimentation in a Vehicle

The pedestrian detection system was deployed in a vehicle, as shown in Figures 23 and 24. The motor was attached to the steering wheel to alert the driver as soon as it detects a pedestrian on the road that might be in danger.

Figure 23. System Integration in a Vehicle

Figure 24. System Setup on a Vehicle

A live demonstration was performed in a vehicle on a street to detect the presence of a pedestrian, which is shown in Figure 25. Overall, the developed system performed well. Over 97% accuracy was achieved in detecting a pedestrian during both the day and night.

Figure 25. Live Demonstration of Pedestrian Detection at Night: System Setup on a Vehicle

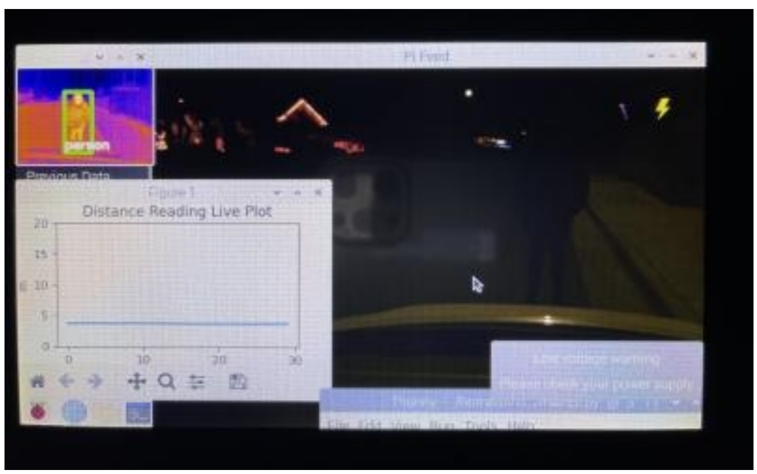


Figure 26. Live Demonstration of Pedestrian Detection at Night: Detection Scheme in Action Inside the Vehicle

7. Conclusion

In this research project, a pedestrian detection and avoidance scheme-using multi-sensor data collection along with machine learning is implemented and experimented with, which can be used for intelligent transportation systems (ITSs). The system is composed of a video camera, an infrared camera, and a micro-Doppler radar for data collection and training. A deep convolutional neural network (DCNN) model is used to train both the RGB images as well as the IR camera images gathered. The number of images gathered and used in the final model for the RGB is 1,200, 600 of which are with pedestrians and the rest with no pedestrians. Similarly, 1,000 images (500 with pedestrians and 500 without) were taken by the IR camera, 85% of which were taken at night. After curating the two datasets, two different DCNNs were trained. A validation accuracy (success rate of detecting pedestrians) of 99.6% was achieved using the RGB camera, while the IR camera that was mainly trained on images taken at night provided a 97.3% validation accuracy. The radar sensor was used to detect the range of the pedestrian and the direction of travel. Experimentations were also carried out in a vehicle. Whenever the multi-sensor detection scheme detected the presence of a pedestrian that might be in danger, the system triggered a signal to the vibrating motor on the wheel as well as displayed a warning message on the passenger's touchscreen computer. The system can be used during the day or at night.

Abbreviations and Acronyms

NHTSA	National Highway Traffic Safety Administration
ADAS	Advanced Driver Assistance Systems
AAA	American Automobile Association
LIDAR	Light Detection and Ranging
ECE	Electrical and Computer Engineering
DCNN	Deep Convolution Neural Network
IR	Infrared
RGB	Red Green Blue
Caltrans	California Department of Transportation
ITS	Intelligent Transportation System
ML	Machine Learning
CNN	Convolutional Neural Network
ANN	Artificial Neural Networks
SRR	Short-Range Radar (SRR)
API	Application Programming Interface
IoT	Internet of Things
ReLU	Rectified Linear Unit
TPU	Tensor Processing Unit
AI	Artificial Intelligence
ASIC	Accelerator Application-Specific Integrated Circuit
ADC	Analog-To-Digital Converter
DAC	Digital-To-Analog Converter

Bibliography

Cao Z, Bıyık E, Wang WZ, Raventos A, Gaidon A, Rosman G, Sadigh D, "Reinforcement learning based control of imitative policies for near-accident driving". arXiv preprint arXiv:2007.00178. 2020 Jul 1.

Datondji SR, Dupuis Y, Subirats P, Vasseur P, "A survey of vision-based traffic monitoring of road intersections". IEEE transactions on intelligent transportation systems. 2016 Apr. 22;17(10):2681-98.

Dennis Valera, Your Central Valley News, 64 percent of Fresno's deadly collisions are vehicle vs. pedestrian ones, police say, Oct 21, 2018. [Online]. Available: https://www.yourcentralvalley.com/news/64-percent-of-fresnos-deadly-collisions-are-vehicle-vs-pedestrian-ones-police-say/

El-Faouzi NE, Leung H, Kurian A, "Data fusion in intelligent transportation systems: Progress and challenges–A survey", Information Fusion. 2011 Jan 1;12(1):4-10.

Ellen Edmonds, AAA Warns Pedestrian Detection Systems Don't Work When Needed Most, Nov. 2019. [Online]. Available: https://newsroom.aaa.com/2019/10/aaa-warns-pedestrian-detection-systems-dont-work-when-needed-most/

FLIR. (2018) Flir lepton engineering datasheet. [Online]. Available: https://lepton.flir.com/wp-content/uploads/2019/02/ flir-lepton-engineering-datasheet-203.pdf

Gaby Galvin, The 10 Worst States for Pedestrian Traffic Deaths, Feb. 2022. [Online]. Available: https://www.usnews.com/news/healthiest-communities/slideshows/10-worst-states-for-pedestrian-traffic-deaths?slide=3

H. Kulhandjian, "A Visible Light Communications Framework for Intelligent Transportation Systems," in California State Transportation Consortium, Project No. 1911, pp. 1-36, Aug., 2020.

H. Kulhandjian, N. Ramachandran, M. Kulhandjian, C. D'Amours, "Human Activity Classification in Underwater using sonar and Deep Learning," in Proc. of ACM Conference on Underwater Networks and Systems (WUWNet), Atlanta, GA, USA, October, 2019.

H. Kulhandjian, P. Sharma, M. Kulhandjian, C. D'Amours, "Sign Language Gesture Recognition using Doppler Radar and Deep Learning," in Proc. of IEEE GLOBECOM Workshop on Machine Learning for Wireless Communications, Waikoloa, HI, USA, December, 2019.

Lim HS, Lee JE, Park HM, Lee S, "Stationary Target Identification in a Traffic Monitoring Radar System". Applied Sciences. 2020 Jan;10(17):5838.

M. Yanagisawa, E. Swanson, P. Azeredo and W. Najm, "Estimation of potential safety benefits for pedestrian crash avoidance systems," National Highway Traffic Safety Administration, 2017.

National Highway Traffic Safety Administration (NHTSA). United States. Early Estimate of Motor Vehicle Traffic Fatalities for the First 9 Months (Jan–Sep) of 2021. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/Publication/813240 Joe Feese, Highway Safety Association, New Projection: 2019 Pedestrian Fatalities Highest Since 1988. Feb. 2020. [Online]. Available: https://www.ghsa.org/resources/news-releases/pedestrians20

Simon Romero, Pedestrian Deaths Spike in U.S. as Reckless Driving Surges, Feb. 14, 2022. [Online]. Available: https://www.nytimes.com/2022/02/14/us/pedestrian-deaths-pandemic.html

Teledyne FLIR Thermal Dataset for Algorithm Training. [Online]. Available: https://www.flir.com/oem/adas/adas-dataset-form/

Tubaishat M, Zhuang P, Qi Q, Shang Y, "Wireless sensor networks in intelligent transportation systems", Wireless communications and mobile computing. 2009 Mar. 9(3):287-302.

Tzutalin. (2015) LabelImg. git code. [Online]. Available: https://github.com/tzutalin/labelImg

About the Author

Hovannes Kulhandjian, PhD

Dr. Kulhandjian is an Associate Professor in the Department of Electrical and Computer Engineering at California State University, Fresno (Fresno State), which he joined in Fall 2015 after having been an Associate Research Engineer in the Department of Electrical and Computer Engineering at Northeastern University. He received his BS degree in Electronics Engineering with high honors from the American University in Cairo (AUC) in 2008, and his MS and PhD degrees in Electrical Engineering from the State University of New York at Buffalo in 2010 and 2014, respectively. His current research interests are in digital signal processing, wireless communications, and networking, with applications to underwater and visible light communications, as well as networking geared towards Intelligent Transportation Systems (ITS).

Dr. Kulhandjian has received numerous awards and research grants while at Fresno State including: four research grants from Fresno State Transportation Institute (FSTI); the Claude C. Laval III Award for Commercialization of Research, Innovation and Creativity 2021; and the Claude C. Laval Award for Innovative Technology and Research 2020. In April 2021, he received a grant as a PI from the Department of Defense (DOD) Research and Education Program for Historically Black Colleges and Universities and Minority-Serving Institutions (HBCU/MI) Equipment/Instrumentation to establish a secure Communications and Embedded Systems Laboratory at California State University, Fresno.

Dr. Kulhandjian is an active member of the Association for Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE) including the IEEE Vehicular Technology Society (VTS). He is a Senior Member of IEEE. He is serving as a Guest Editor for the Special Issue "Advances in Intelligent Transportation Systems (ITS)". He has served as a Guest Editor for IEEE Access Special Section Journal, Session Co-Chair for IEEE UComms'20 Conference, Session Chair for ACM WUWNet'19 Conference, and Publicity Co-Chair for the IEEE BlackSeaCom Conference. He also serves as a member of the Technical Program Committee (TCP) for ACM and IEEE Conferences such as GLOBECOM 2022, WTS 2022, WD 2021, WD 2018, ICC 2018, WUWNet 2020, and WiMob2019. He is a recipient of the Outstanding Reviewer Award from ELSEVIER Ad Hoc Networks and ELSEVIER Computer Networks.

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors

Karen Philbrick, PhD
Executive Director

Hilary Nixon, PhD
Deputy Executive Director

Asha Weinstein Agrawal, PhD
Education Director
National Transportation Finance
Center Director

Brian Michael Jenkins
National Transportation Security
Center Director