OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Study of Large Vehicle Safety Warning Systems Based on YOLO

Wen-Kung Tseng [1], Jin-Ren Xiao [2]

[1]Professor, Graduate Institute of Vehicle Engineering, National Changhua University of Education, Taiwan R.O.C.

[2]Postgraduate Student, Graduate Institute of Vehicle Engineering, National Changhua University of Education, Taiwan R.O.C.

Corresponding Author: andy007@cc.ncue.edu.tw

ABSTRACT: With the rapid development of technology, modern society is constantly evolving, bringing about profound changes in our daily lives. In this technological era, ensuring the safety of large vehicles has become a top priority. Whether navigating through busy city streets or cruising along highways, the interaction between large vehicles and pedestrians often poses significant risks. However, advancements in modern technology offer us ample opportunities to incorporate sophisticated solutions into large vehicles, thus reducing the likelihood of collisions with pedestrians. This study utilizes object detection cameras based on YOLO, coupled with warning systems and deep learning algorithms. These integrated systems can effectively identify motorcycles, bicycles, and pedestrians on the road, promptly alerting drivers in real-time when potential hazards are detected nearby, thereby minimizing the risk of accidents. In conclusion, the experimental results demonstrated that employing object detection cameras based on YOLO enables instantaneous detection of potential hazards, allowing for immediate warnings to be issued to drivers upon successful identification.

KEY WARDS: *Modern society, Navigating, Large vehicles, Pedestrians, Object detection, YOLO.*

Date of Submission: 05-02-2025 Date of acceptance: 16-02-2025

I. INTRODUCTION

Over the years, while traveling on the road, many pedestrians, cyclists, and motorcyclists lack awareness of the dangers posed by nearby large vehicles. According to statistics from the Road Safety Information Query Network, in 2023 alone, there were 10,037 accidents involving large vehicles, resulting in 12,119 casualties, including 317 fatalities [1]. In 2016, Redmon, Joseph, et al. published a unified, real-time object detection method called YOLO (You Only Look Once) [2]. Compared to traditional object detection methods, YOLO's main innovation lies in its ability to achieve end-to-end object detection within a single neural network, allowing simultaneous detection of multiple objects in an image at an extremely high processing speed. YOLO divides an image into a fixed-size grid and makes predictions for each grid cell, including the object's bounding box and class, thus enabling real-time detection of the entire image. This design allows YOLO to maintain high accuracy while significantly increasing detection speed, making it suitable for applications requiring efficient real-time detection.

One of the significant contributions of this study was providing a unified framework that integrates various steps of object detection into a single neural network, simplifying the process and enhancing efficiency. In addition, YOLO improves detection accuracy and generalization by using multi-scale feature maps and highly abstracted features, enabling effective detection in diverse scenarios. The paper has set the direction for the development of object detection technology in the field of computer vision. The proposed YOLO method has garnered extensive attention and application in both academic and industrial circles, becoming a crucial tool in areas such as autonomous driving, surveillance systems, and image processing. In summary, this research not only provided an efficient solution for object detection but also established important theoretical foundations and technical support for real-time target detection in practical scenarios.

In 2022, Zhang, Yu, et al. proposed real-time vehicle detection based on improved YOLOv5 [3], which aimed to enhance vehicle detection using the advanced YOLOv5 object detection framework. The research team optimized and improved YOLOv5 to enhance the accuracy and real-time performance of vehicle detection. Key improvements included optimizing the model architecture and training process. Researchers employed a more robust convolutional neural network architecture, increasing the model's depth and breadth to improve vehicle detection in complex scenarios. Additionally, they optimized training algorithms and parameter settings to enable faster convergence and achieve better performance. The research results showed that the improved YOLOv5 model has made significant advancements in real-time vehicle detection. Compared to traditional methods, this model not only achieved higher accuracy but also processed data faster while maintaining accuracy. As a result, it was suitable for various applications requiring efficient real-time detection, such as traffic monitoring and intelligent traffic management. Overall, this research contributes to enhancing the efficiency and accuracy of vehicle detection. By improving the YOLOv5 model, researchers have provided an effective and feasible solution for real-time vehicle detection, which helps elevate the level of traffic safety management and promotes sustainable urban development.

In 2022, Wang et al. published YOLOv7 trainable bag-of-freebies for real-time object detectors [4]. This research introduced an object detection model named YOLOv7, which is based on trainable bag-of-freebies and achieved a new technological breakthrough in real-time object detection. The innovation of the YOLOv7 model lies in combining various free techniques such as model architecture, data augmentation, and regularization methods. Through this approach, the research team effectively enhanced the object detection model, improving its performance and generalization capabilities. In experiments, the YOLOv7 model achieved new state-of-the-art performance on multiple benchmark datasets. Compared to existing object detection models, YOLOv7 demonstrates significant improvements in accuracy, processing speed, and efficiency. Particularly in real-time object detection, YOLOv7 performs exceptionally well and is applicable in scenarios requiring efficient real-time detection, such as autonomous driving and surveillance systems. In summary, this research contributed to the advancement of real-time object detection technology. By introducing the YOLOv7 model, researchers have provided a more advanced and efficient solution for real-time object detection, which helped drive the application and development of object detection technology in practical scenarios.

In 2011, Park, Jaesik, et al. published high quality depth map upsampling for 3D-TOF cameras [5], aiming to enhance the quality of depth maps generated by 3D Time-of-Flight (TOF) cameras, with a focus on depth map upsampling techniques. The research team proposed a depth map upsampling method. This method smoothed the depth map using filters to achieve higher quality. To further enhance the accuracy and detail of the depth map, they introduced a Local Structure Similarity (LSS) filter, which optimizes the depth map in both spatial and depth domains. The research results indicated that the proposed depth map upsampling method effectively reduced distortion and improved depth accuracy while preserving the structure of the depth map. Compared to traditional methods, this approach demonstrated outstanding performance in both visual effects and quantitative assessments. Overall, this study provided an effective solution for depth map upsampling generated by 3D-TOF cameras, enhancing the quality and accuracy of depth maps to improve the three-dimensional visual effects and precision in related applications.

In 2001, JONES, Willie D, published keeping cars from crashing [6], which explored technological methods to prevent vehicle collisions. The study identified vehicle collisions as a major cause of traffic accidents and casualties. To address this issue, researchers proposed a series of technological solutions aimed at enhancing driver safety and vehicle active safety. These included adopting smart vehicle technologies to facilitate vehicle-to-vehicle communication and sensor systems for cooperative and interactive capabilities between vehicles, thereby preventing potential collision risks. The study emphasized the importance of the 2-second following distance rule. This principle suggested that drivers should maintain a distance of at least 2 seconds when driving to ensure sufficient reaction time for unexpected situations. Applying this principle effectively reduces the risk of vehicle collisions while enhancing driving safety. Smart vehicle technology is recognized as a key development direction in future vehicle safety. By integrating advanced computer vision, radar, LiDAR, and other sensor technologies, along with machine learning and artificial intelligence algorithms, smart vehicles can monitor and identify various objects and obstacles in traffic environments in real-time, automatically taking measures to prevent collisions. In summary, the study proposed strategies for preventing vehicle collisions using smart vehicle technology and the following distance principle, offering insights into future research directions and development trends. This was of significant theoretical and practical importance for improving traffic safety levels and reducing traffic accidents and casualties.

II. PROCEDURES AND METHODS

The object detection refers to the process of identifying and classifying objects in images or videos using computer vision techniques. The methods involved include feature extraction, feature description, and classification. First, features are extracted by capturing key points or regions in the image, such as edges and corners. Next, these features are described, typically converted into mathematical forms for further processing. Finally, machine learning or deep learning techniques are applied to classify these features, thereby identifying objects in the image. This technology finds widespread applications in artificial intelligence, robotics, autonomous driving, and other fields. For instance, in intelligent surveillance systems, object recognition can be used to identify people, vehicles, and other objects in surveillance footage, enabling smart monitoring and security functionalities. In autonomous driving vehicles, object recognition helps in recognizing road signs, pedestrians, obstacles, etc., facilitating intelligent navigation and safe driving.

Thus, the continuous development and application of object recognition technology are expected to bring more convenience and safety to people's lives. The principle of object detection used in this study is to use CNN (Convolutional Neural Network) [7] which is an advanced feedforward neural network specifically designed to process data with a convolutional structure. This network architecture is primarily used for the efficient processing of multidimensional data such as images, videos, and audio. The components of a CNN include multiple layers of convolutional layers, pooling layers, and fully connected layers. These layers work together to enable the model to learn features effectively from complex data and make accurate predictions, as shown in Figure 1.

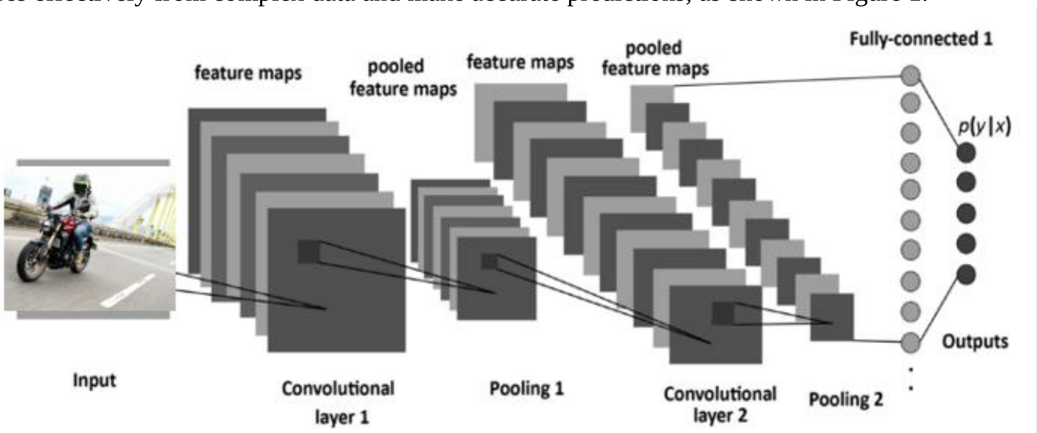

Figure 1: CNN Architecture [7]

CNN is widely used in the field of artificial intelligence, such as image recognition, object detection, and speech recognition. The key feature of CNN lies in convolutional operations, where input data is convolved with a kernel through sliding windows, as depicted in Figure 2 [8]. This process helps extract feature maps and capture various features of images. Subsequently, pooling layers sample the feature maps to reduce dimensionality and computational load while preserving important features of the image. Finally, fully connected layers convert the feature maps into one-dimensional vectors for tasks like classification or regression.

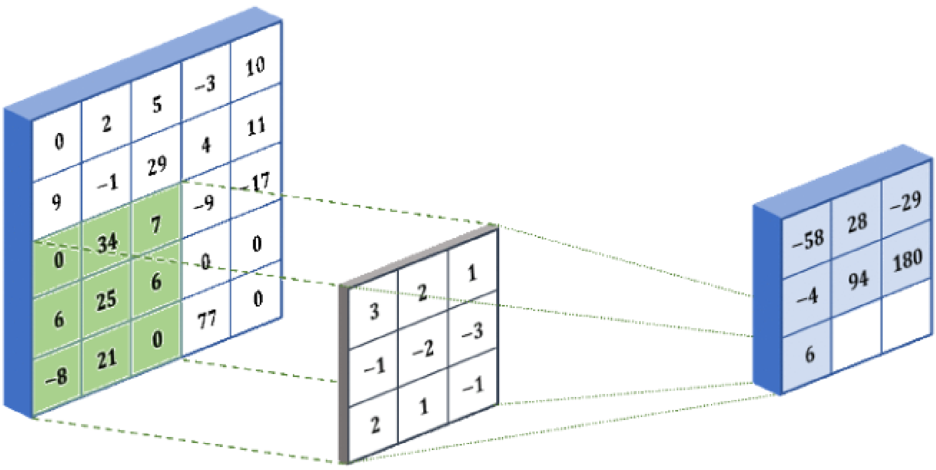

Figure II: Procedure of a 2-D CNN Error! Reference source not found.

The training process of CNN utilizes backpropagation algorithm to predict training data through multiple convolutional and fully connected layers. By computing the difference between predicted values and actual values, convolutional kernels and weight parameters are adjusted through gradient descent to maximize the model's prediction accuracy. Today, CNN has become a mainstream algorithm in fields such as image recognition and object detection, widely applied across numerous practical scenarios. After processing through convolutional layers, data enters pooling layers. The purpose of pooling is to reduce the data volume of images while preserving important information, achieved by performing a dimensionality reduction through either maximum or average pooling methods. This paper used a 2x2 max pooling as an example, as illustrated in Figure 3 [9].

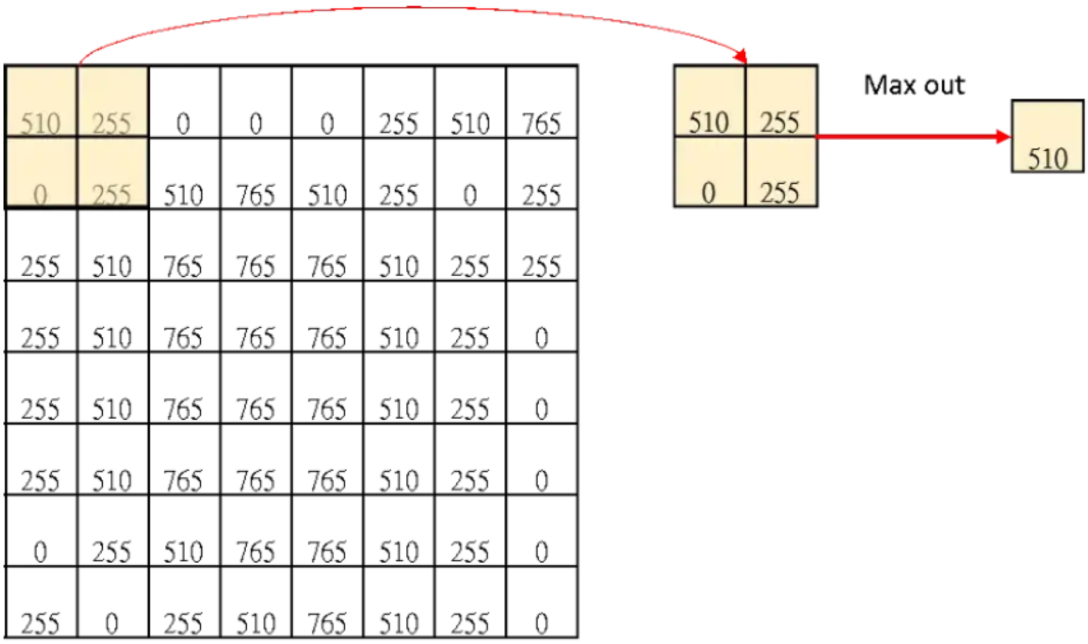

Figure II: Procedure of a 2-D CNN Error! Reference source not found.

There is a simple example online where an image is divided into smaller parts, each processed separately, and then combined together. Despite reducing the pixel count of the image, it is still possible to recognize the original image through its features as shown in Figure II **Error! Reference source not found.**.

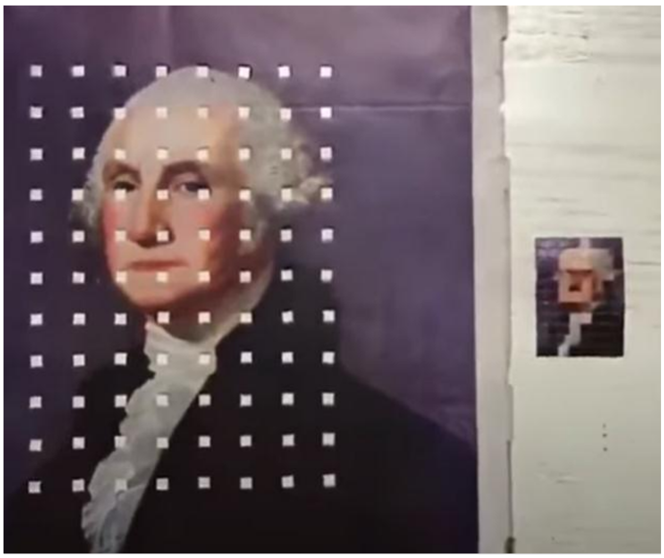

Figure II: Example of 2D CNN Error! Reference source not found.

YOLO is one of the widely used models in the field of object detection today [9]. It achieves object localization in an image through a single prediction step. This algorithm is based on Convolutional Neural Networks (CNNs). Upon inputting an image, it divides the image into an S × S grid. Within each grid cell, the algorithm predicts bounding boxes, allowing for a maximum of two bounding boxes per grid cell. Additionally, the algorithm provides confidence scores and probabilities for their associated classes or labels, totaling five values, as shown in Figure 5. After predicting the bounding boxes and confidence scores for the two grid cells, the algorithm utilizes Non-Maximum Suppression (NMS) to generate the final output results.

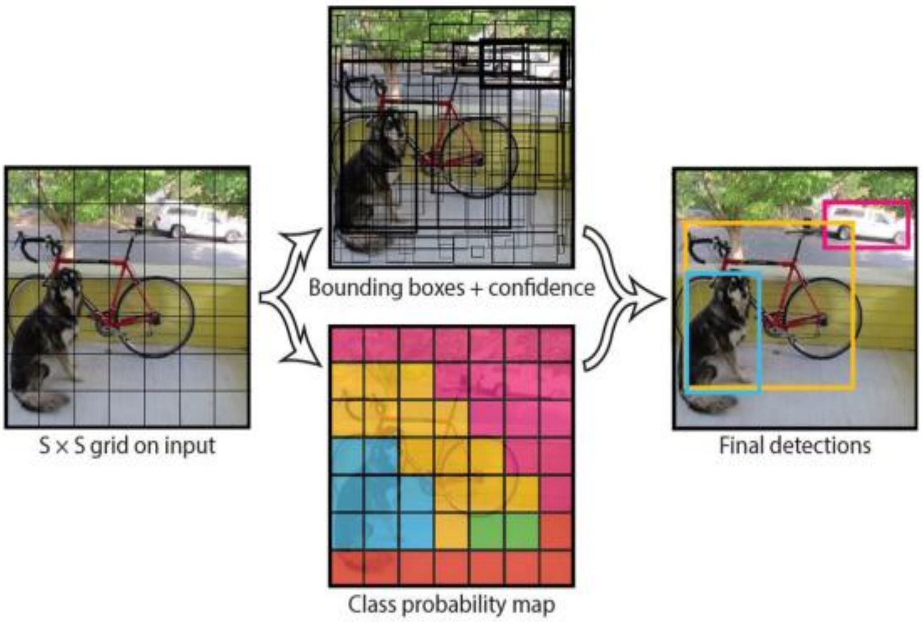

Figure 5: YOLO Object Detection Error! Reference source not found.

YOLOv5, part of the YOLO series of algorithms, was developed by Ultralytics and released in June 2020 [2], but has not yet been published in a paper. Compared to earlier versions of YOLO, YOLOv5 shows significant improvements in both accuracy and speed. It introduces new module functionalities and offers more convenience in training. YOLOv5 is built upon YOLOv4 and notably integrates PyTorch as its native framework for the first time, making model training more accessible for users. The YOLOv5 official release includes a series of models that users can utilize by adjusting parameters such as network depth and width, allowing for models of varying complexity. This flexibility and user-friendly approach enable users to adapt to diverse training requirements more easily. The models range from YOLOv5n (smallest) to YOLOv5x (largest), as depicted in Figure 6.

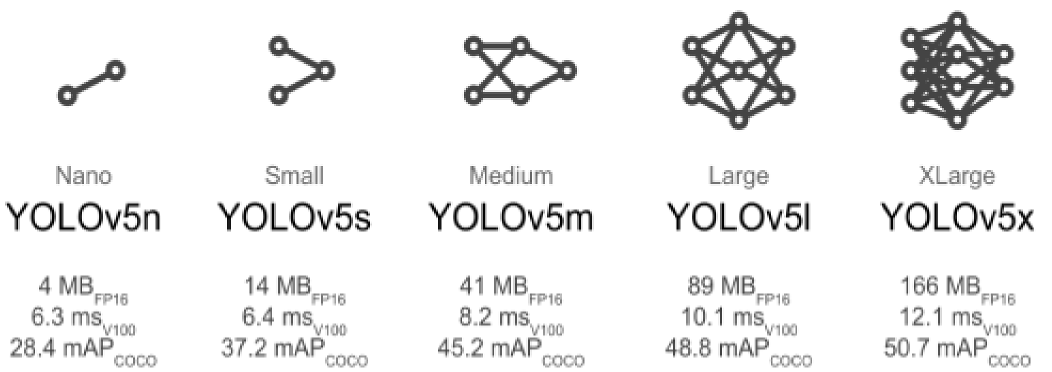

Figure 6: YOLOv5 model Error! Reference source not found.

YOLOv5 offers five different models, as shown in Figure 7. Users can choose the appropriate model based on their needs and the devices they use. This allows for achieving better training results and enables object detection on images, videos, or camera feeds.

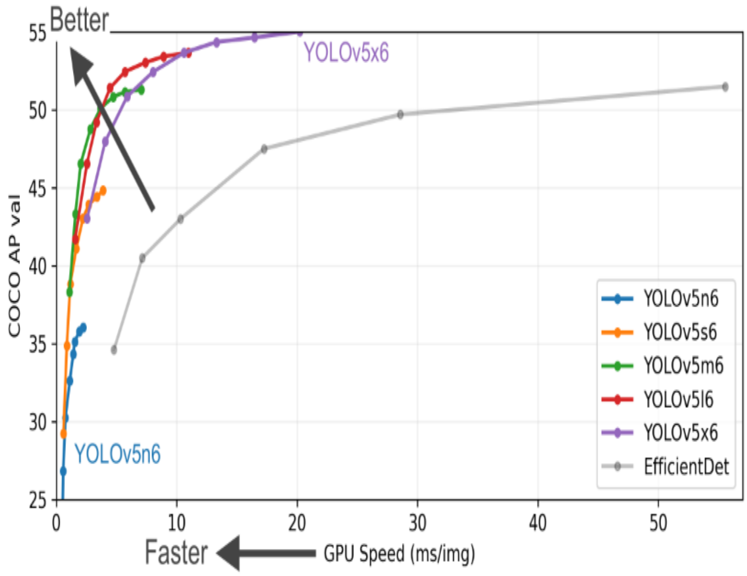

Figure 7: YOLOv5 model comparison Error! Reference source not found.

III. RESULTS AND DISCUSSIONS

This study utilized the YOLOv5 model for training. Image data on pedestrians, bicycles, and motorcycles, in various environmental conditions were taken. Additionally, each category of pedestrians, bicycles, and motorcycles consisted of 800 images, totaling 2400 images for the YOLOv5 training dataset, as depicted in Table 1.

Table 1: Datasets

project	images
pedestrian	800 images
bicycle	800 images
motorcycle	800 images
Total	2400images

After collecting image samples, it is necessary to establish a database in the YOLO format for training the YOLO object detection module. This process involved using the LabelImg software to annotate and categorize each image sample. The annotation process required accurately marking the position and category of objects in each image so that subsequent training can precisely identify different types of objects. To ensure diversity and sufficiency of training data, a large number of image samples from different scenes, angles, and lighting conditions were collected. In this study, the data labeling is based on three target features: pedestrians, bicycles, and motorcycles. The annotation software LabelImg was used to label objects and create coordinate data, marking the desired features for recognition. Such a database was essential for better training of the model, enabling it to perform with improved performance and accuracy in real-world applications as shown in Figure

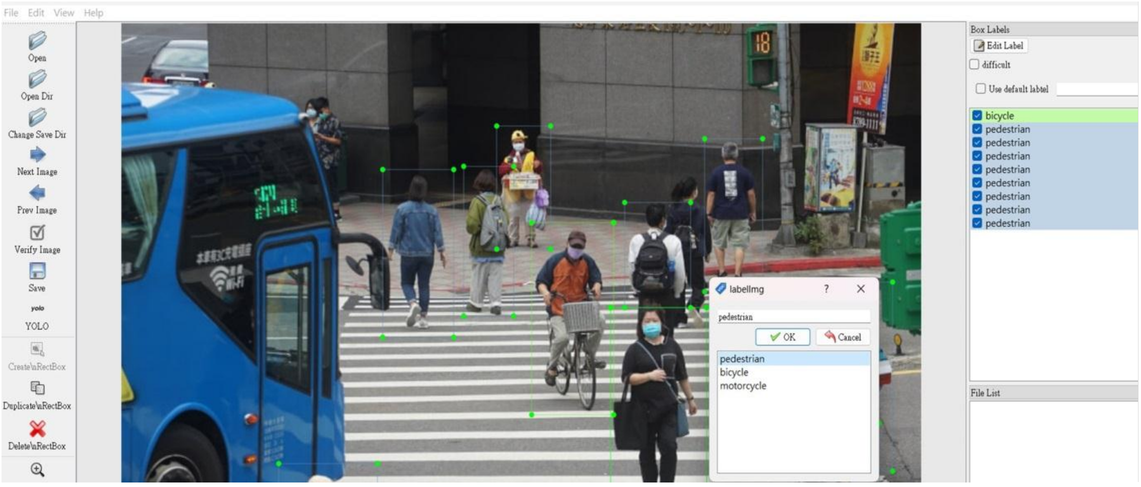

Figure 8: Images labeling

After labeling the collected images, LabelImg automatically generated YOLO format text files. Each generated text file contained five pieces of data: the first item (0-2) represented the categories for pedestrian, bicycle, and motorcycle, while the remaining four values represent the center coordinates (x, y) of the annotated target and the dimensions (width, height) of the bounding box relative to the original image size. Once all 2400 images were labeled, training of the YOLO model could commence, and training parameters were adjusted. Using the trained YOLO model, we validated its capability to observe pedestrians in real-time via a webcam. During the detection process, it accurately identified pedestrians near large vehicles and assigns corresponding class labels, as depicted in Figure 9. Pedestrians in the footage appear from different angles, facing towards or away from the camera, including adults and children of various genders and body types appearing multiple times in different locations. Despite these variations, the trained YOLOv7 model consistently recognizes pedestrians and provides correct class labels. The confidence scores for detecting distant pedestrians are around 0.80, and the recognition speed achieves 0.018 seconds, demonstrating real-time pedestrian observation capabilities.

Figure 1: Real-time pedestrian detection

Using the trained YOLO model, we verified whether the model could perform real-time bicycle detection through a webcam without being influenced by surrounding buildings and vehicles such as cars and public transportation. During the detection process, it correctly identified bicycles and assigns class labels, as shown in Figure 10. Bicycles of different sizes and in various poses were observed in the camera footage, and the trained YOLO model consistently recognized bicycles and provided accurate class names. The confidence scores for detecting moving bicycles are around 0.80, and the recognition speed reaches 0.017 seconds, demonstrating real-time bicycle observation capabilities.

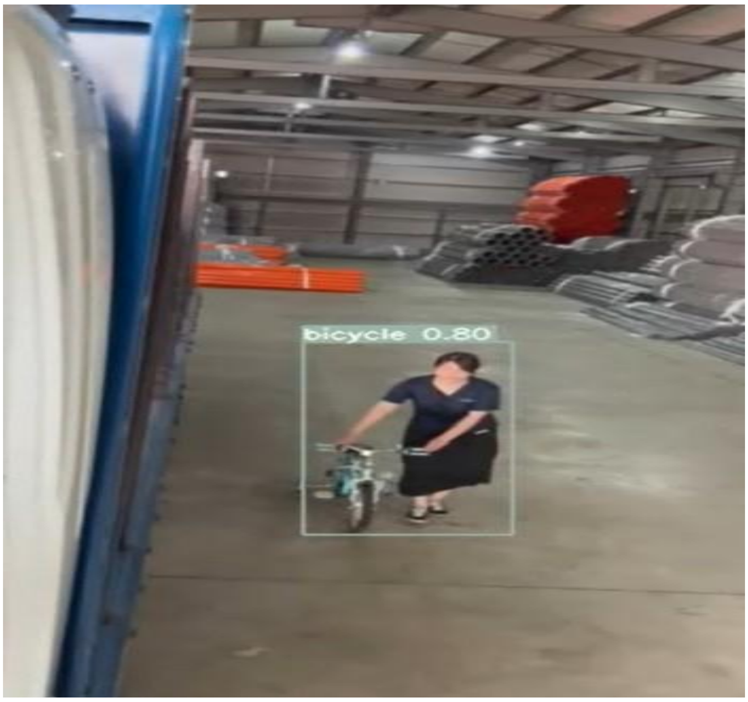

Figure 10: Real-time bicycle detection

Using the trained YOLO model, we aimed to verify whether the model could perform real-time motorcycle detection through a webcam without being affected by nearby cars and buildings, and accurately detected motorcycles during the detection process, providing class labels as shown in Figure 11: **Real-time motorcycle detection**The motorcycle in the camera moved within the frame, and the trained YOLO model could still recognize the motorcycle and provided the correct class label. It also accurately detected bicycles during the detection process and provides class labels. The motorcycles in the camera vary in style, color, angle, and distance as they moved within the frame. The confidence score was around 0.71 and achieved recognition speeds of 0.018 seconds, demonstrating real-time motorcycle detection.

Figure 11: Real-time motorcycle detection

In terms of accuracy, it can be seen that the intelligent detection system proposed in this study achieved nearly 0.8 precision. High precision and low loss rates particularly in handling small objects and complex scenes could be obtained. When the intelligent detection system detected the pedestrians, bicycles and motorcycles the system would signal the alarm system to warn the large vehicle‘s driver and the people nearby to avoid the collision.

IV. CONCLUSIONS

In this research the vehicle-mounted cameras with object detection technology were integrated into a warning system. Its purpose was to reduce accidents involving large vehicles and pedestrians, bicycles and motorcycles during driving. In the literature review, existing vehicle cameras and radars typically provided only visual display and distance information. In contrast, this study integrated YOLO object detection specifically for alerting pedestrians, cyclists, and motorcyclists. After data collection and model training, the trained YOLO model was utilized for real-time detection of targets. High accuracy and low loss factor were obtained. Due to modifications in calculation methods and model modules, the detection speed has been optimized to approximately 0.018s, allowing effective recognition of pedestrians, bicycles, and motorcycles. Overall, using YOLO as the object detection model not only provided an intuitive way to observe pedestrians, bicycles, and motorcycles, but also achieved impressive performance metrics. In summary, this paper applied YOLO object detection and warning system technology in vehicle-mounted cameras to detect pedestrians, bicycles, and motorcycles in the environment. Upon successful detection, immediate alerts were sent to the warning system to avoid the collision.

REFERENCES

[1]. Vehicle safety report , (2018). Available: https://roadsafety.tw/AccCauseOrder.
[2]. Redmon, Joseph, (2016). "You only look once: Unified, real-time object detection", Proceedings of the IEEE conference on computer vision and pattern detection.
[3]. Zhang, Yu, (2022). "Real-Time Vehicle Detection Based on Improved YOLO V5", Sustainability.
[4]. Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao, (2022). "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." arXiv preprint arXiv:2207.02696.
[5]. Park, Jaesik, (2011). "High quality depth map upsampling for 3D-TOF cameras." 2011 International Conference on Computer Vision.
[6]. JONES, Willie D, (2011). Keeping cars from crashing. IEEE spectrum, 38.9: 40-45.
[7]. LeCun, Yann, (1998). "Gradient-based learning applied to document recognition", Proceedings of the IEEE 86.11: 2278-2324.
[8]. Convolutional Neural Network (CNN), available: https://medium.com/ching- A1-convolutional-neural-network-cnn-
[9]. Convolutional Neural Network (CNN), available: https://chih-sheng-huang821.medium.com/%E5%8D%B7% B6%93%EF-convolutional-neural-network-cnn-