International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 08, Issue – 05, May 2023, PP – 01-14

To Enhance Nighttime Vehicle Recognition with YOLO Technology Based on Optical and Thermal Images

Chih-Wei Kuan[1], Mei-Ling Yeh[2], Tien-Yin Chou[2], Wei-Zhe Chen[2]
[1](PHD student of Infrastructure Planning and Engineering, Feng Chia University, Taiwan)
[2](GIS center, Feng Chia University, Taiwan)

Abstract: This study proposes a method to enhance the recognition capability of nighttime driving vehicles by using a combination of optical and thermal imaging, along with the Yolo algorithm in AI deep learning. The proposed method is applied to the surveillance cameras at intersections at night to evaluate the traffic congestion of road sections. The Yolo algorithm is used to accurately identify the vehicles in the mixed images of optical and thermal imaging, thereby improving the recognition rate of nighttime driving vehicles. The proposed method can be used to effectively monitor the traffic situation at night and provide decision support for transportation management. Experimental results demonstrate that the proposed method can achieve high accuracy in vehicle recognition, making it a promising approach for nighttime driving safety and traffic management.
Keywords: Optical image, Thermal image, Yolo, Deep learning

I. INTRODUCTION

In recent years, with the advancement of computer hardware and software, artificial intelligence (AI) technology has flourished and various AI applications have penetrated into daily life, such as automatic detection systems for self-driving vehicles, license plate recognition in parking lots, and intersection surveillance systems, significantly improving modern traffic control systems. However, traditional optical images are easily affected by rainy weather or insufficient light at night, which can limit vehicle recognition rates. In order to improve overall traffic safety, intersection monitoring or traffic flow monitoring should provide real-time monitoring regardless of weather conditions or nighttime hours.

The physical principle of thermal imaging can distinguish features of objects with temperature changes, while traditional optical images rely on the reflection of light to identify the shape of objects. The phenomenon of light halos from vehicle headlights at night can also make optical image recognition difficult. However, thermal imaging utilizes the principle of heat energy, which is not affected by vehicle headlight lighting, making it less limited in nighttime image recognition compared to optical imaging.

Therefore, this study aims to use thermal imaging to complement the limitations of traditional optical images under low light conditions at night. The study also utilizes YOLO, a neural network method in deep learning that can simultaneously identify objects, to train and calibrate the AI model. Finally, the trained model is used to analyze and compare nighttime vehicles in both optical and thermal images.

II. RELATED WORKS

Deep learning is a type of artificial intelligence where the computer learns to recognize patterns in raw data, also known as feature learning or feature extraction, without explicit instruction from humans. The computer uses its own judgment to identify commonalities and groupings, allowing for efficient and accurate categorization of objects and recognition of target locations. This approach is particularly useful for real-time applications, such as vehicle image recognition, as it can operate with minimal human intervention.

1. Deep Learning Applied to Vehicle Identification

In recent years, deep learning has become the most promising indicator for road detection. Vehicles are an important category in traffic surveillance. Early self-driving cars relied on sensor systems that required the installation of many sensors and the data detected by these sensors to provide accurate environmental information. However, this approach had the disadvantage of requiring intensive manual parameter adjustment by professionals, making it difficult to apply in every environment. Feris (2012) and Matei (2011) used road cameras to capture images of vehicles, quickly detecting, locating, and tracking target vehicles in surveillance images. In the field of computer vision, deep learning methods have far surpassed traditional algorithms in terms of detection accuracy due to the constant increase in data volume and rapid advances in hardware and software devices.

Vehicle detection methods primarily consider the characteristics of vehicles, including image changes in dynamic environments, lighting and background changes in the environment, and the type, size, shape, position, and color of vehicles. The currently mainstream detection methods are mainly divided into two types: two-stage vehicle detection algorithms and one-stage vehicle detection models. The representative of two-stage object detection is faster RCNN, which generates a series of candidate boxes at the beginning and classifies and regresses them, including candidate region generation, feature extraction, classification, and position refinement. The advantages of this method are high detection accuracy, and the overall accuracy performance has been improved to 42.1% by combining the VGG16 network as the basic feature extractor (K. Simonyan, 2014). One-stage vehicle models do not have an intermediate detection process, mainly using image sampling at different positions, scales, and aspect ratios, followed by CNN feature extraction for classification and regression. Since the process has only one step, it has the advantage of being faster to operate. YOLO has proven to be effective in vehicle recognition, improving both the model's mAP detection indicator and operating speed, resulting in significant improvements in both real-time and accuracy in object detection.

However, vehicle detection at night is more challenging. In 2019, Ho Kwan Leung proposed an optimization for Faster R-CNN to address nighttime vehicle recognition. Through the collection of datasets and nighttime vehicle images, the special processing of obstructed, blurry, and small objects is necessary due to the environmental conditions at night. Traditional applications for nighttime vehicle detection rely mainly on the headlights and taillights of vehicles because there is not enough light at night. The detection accuracy of Faster R-CNN increases with the size of the vehicle image used for training. The larger the vehicle image used during training, the clearer the primary features and contours of the vehicle are displayed, providing effective training.

According to the research by Vibhanshu Singh Sindhu in 2021, vehicle detection and tracking has become an important tool in traffic image monitoring for road safety, law enforcement, and surveillance at intersections and on dashcams. The use of algorithms to process vehicle detection has become popular as traffic image data becomes more widely available. Although machine learning methods have fast recognition speeds, the variation in brightness and different speeds of vehicles in the image make the scene more complex. However, CNN has shown good results in object detection. Vuong Xuan Can in 2021 used YOLOv4 to detect and count vehicles in the traffic environment of Vietnam, where motorcycles account for over 86% of vehicles on the road. The traffic situation is complex and chaotic, and they tested five types of vehicles, including motorcycles, bicycles, cars, and buses, comparing YOLOv4, MOG, and Haar Cascade for accuracy during peak and off-peak periods. The accuracy of vehicle detection and counting depends on many factors, such as training dataset, image quality, and weather conditions. Under the same environmental conditions, YOLOv4 showed the best overall results in terms of the number of vehicles recognized and the accuracy rate. It could detect vehicles in a complex traffic environment, even when there were many vehicles overlapping in the image, with better results than the other two methods.

According to a study by Vibhanshu Singh Sindhu (2021), vehicle detection and tracking have become important tools for traffic image monitoring in various applications such as road safety, violation enforcement, and surveillance through traffic cameras and dashcams. The use of algorithms to process vehicle detection has become popular with the flow of traffic images. Although machine learning methods provide fast recognition, the variability of lighting conditions and vehicle speeds in complex traffic scenarios make the image analysis more challenging. However, Convolutional Neural Networks (CNN) have shown promising results in object detection. Vuong Xuan Can (2021) used YOLOv4 for vehicle detection and counting in Vietnam's traffic environment, where over 86% of the vehicles are motorcycles, making the traffic situation more complicated and chaotic. Five types of vehicles were tested, including motorcycles, bicycles, cars, and buses, and the YOLOv4, MOG, and Haar Cascade models were compared for their accuracy during peak and off-peak hours. The accuracy of vehicle detection and counting depends on many factors, such as training data set, image quality, and weather conditions. Under the same environment, YOLOv4 showed the best performance in terms of overall detection quantity and accuracy. It could detect vehicles in each frame of a video stream, making it effective even in scenarios with heavy traffic and overlapping vehicles.

Both YOLOv3 and YOLOv4, even with more advanced features and higher recognition efficiency, have many applications in traffic and monitoring, as suggested by Chethan Kumar B. (2020). The YOLO network vehicle model was trained with a neural network consisting of one hidden layer with the minimum input and one output layer. The training dataset includes images and videos taken during the day and night under different lighting conditions, capturing images of car, truck, and two-wheeler categories between RGB and grayscale images. Performance parameters, such as accuracy, recall, and F1 score, were calculated from the nighttime and blurry image datasets. The results showed that using the YOLO model for vehicle detection, the accuracy of images and videos was as high as 98% and 99% (C. Kumar B, 2020). Amir Mohammad Ghoreyshi (2020) suggested that using the YOLOv3 tiny version in vehicle recognition, YOLO always maintained the same rate of detecting traffic vehicles, regardless of how the vehicles in the image change. The recognition rate is also five times faster than other models, and the difference in accuracy is not too significant, as shown in Fig. 1.

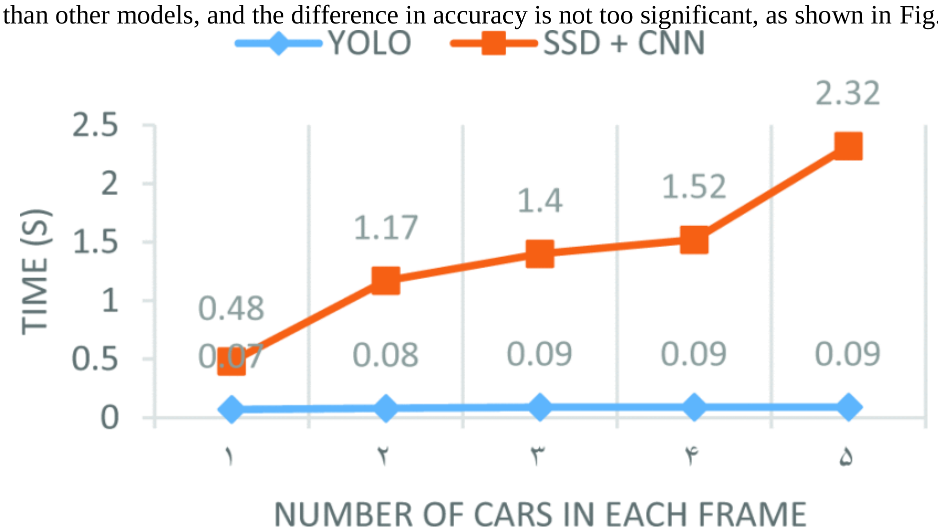

Fig. 1 Diagram of the relationship between vehicle variations and different network types(Amir Mohammad Ghoreyshi, 2020)

2. Application of thermal imaging related analysis

Infrared thermal imaging camera, also known as infrared thermography or thermal imager, is affected by changes in ambient temperature when capturing images of vehicles at night, as the image is formed by differences in radiation intensity between objects and their surrounding environment. Thus, higher ambient temperature can reduce the contrast between the detected object and its background. All objects emit infrared heat that is invisible to the naked eye, and the human eye can only receive a narrow spectrum of electromagnetic waves, known as visible light. The infrared region of the electromagnetic spectrum (wavelength between microwaves and visible light, 760 nanometers to 1 millimeter, corresponding to a frequency range of 430 THz to 300 GHz) is divided into near-infrared (NIR, IR A DIN) with a wavelength of 0.75 to 1.4 micrometers, shortwave infrared (SWIR, IR B DIN), mid-wave infrared (MWIR, IR C DIN) with a wavelength of 3 to 8 micrometers, long-wave infrared (LWIR, IR C DIN) with a wavelength of 8 to 15 micrometers, and far-infrared (FIR) with a wavelength of 50 to 1,000 micrometers, which cannot be captured by our human eyes. Thermal imaging cameras collect infrared radiation from objects in the scene and create an electronic image based on information about the temperature difference. Thermal imaging is not affected by external environmental factors such as lighting, smoke, and shadows in vehicle imaging and can measure temperature from a certain distance in a non-contact manner, converting it into a thermal image for display. One of its advantages is that it can compare the surface temperature distribution of various objects over a large area and can also measure small objects. It can operate in complete darkness, and since objects rarely have the same temperature as their surrounding objects, the thermal imager can detect them, making them look different in the thermal image. This technology is also suitable for vehicle identification on roads. This study will use thermal imaging to supplement optical imaging's shortcomings in low-light conditions at night and conduct research and analysis on vehicle identification.

III. RESEARCH METHOD

This study is based on YOLO v4 in the PyTorch deep learning framework, using vehicles as training samples to train an AI model for recognizing nighttime vehicle information on the road. We also conduct a simple comparison of different vehicle types to explore the difference in overall recognition accuracy between optical and thermal imaging.

1. Image analysis flow

This article will compare the recognition of driving vehicles between thermal and optical images. The image analysis process is shown in Fig. 2. We will process the vehicle images collected from the internet and train them separately with YOLOv4 and convolutional neural network to obtain the trained weights for recognition and classification of optical and thermal images. The process includes: 1. Image preprocessing (data collection, feature engineering, dividing into training and validation sets), 2. CNN model training (model learning, prediction, and evaluation), and 3. recognition of optical and thermal images.

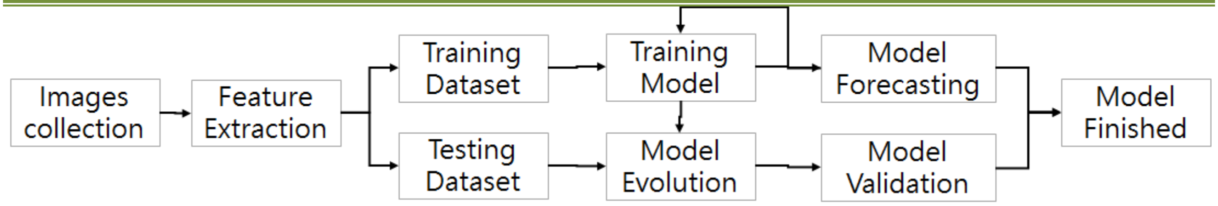

Fig. 2 Training process diagram for YOLOv4

2. Collection and analysis of thermal imaging data

The training data of this study consists mainly of self-shot vehicle photos on the road, with some using existing datasets from Yolo. The shooting locations on the road mainly include Wensin Road and Taiwan Boulevard, where the traffic volume is relatively high, as shown in Fig. 3 and Fig. 4. The advantage of selecting locations with higher traffic volume is that many types of vehicles can be labeled and trained in a single photo to improve the model's recognition rate.

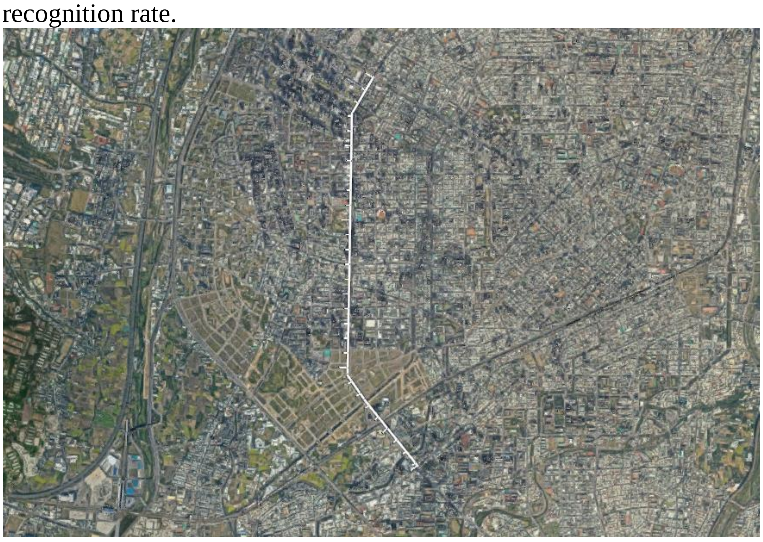

Fig. 3 Vehicle sampling range (white lines: from Section 2 of Wenxin Road to Wenxin South Road)

Fig. 4 Vehicle sampling area (white line: from the intersection of Section 2, Taiwan Boulevard and Wenxin Road to Section 1, Taiwan Boulevard)

3. Pre-processing of training data and environment variable setting

Before training the model with images, the vehicles to be trained need to be labeled. In the learned samples during the model training, the characteristic values of the vehicles marked and the areas that are not selected by the marked bounding box are considered as background. LabelImg, an open-source software, is used for labeling. It is an image annotation tool. In this study, three labels were marked, including buses, cars, and trucks. As shown in Fig. 5, the boundary of the vehicle to be detected is marked with a bounding box and labeled with the category name. Other areas that are not marked are considered as background.

When labeling training samples of vehicles, if there are vehicles that are farther away from the screen, their features are already limited and unclear. For example, some vehicles may overlap with other vehicles. In order to avoid confusion in machine recognition during learning, the blocked vehicles are not labeled. However, there is still a need for sufficient training data sets. Vehicles with complete faces, important features such as headlights and windshields should be labeled in the images.

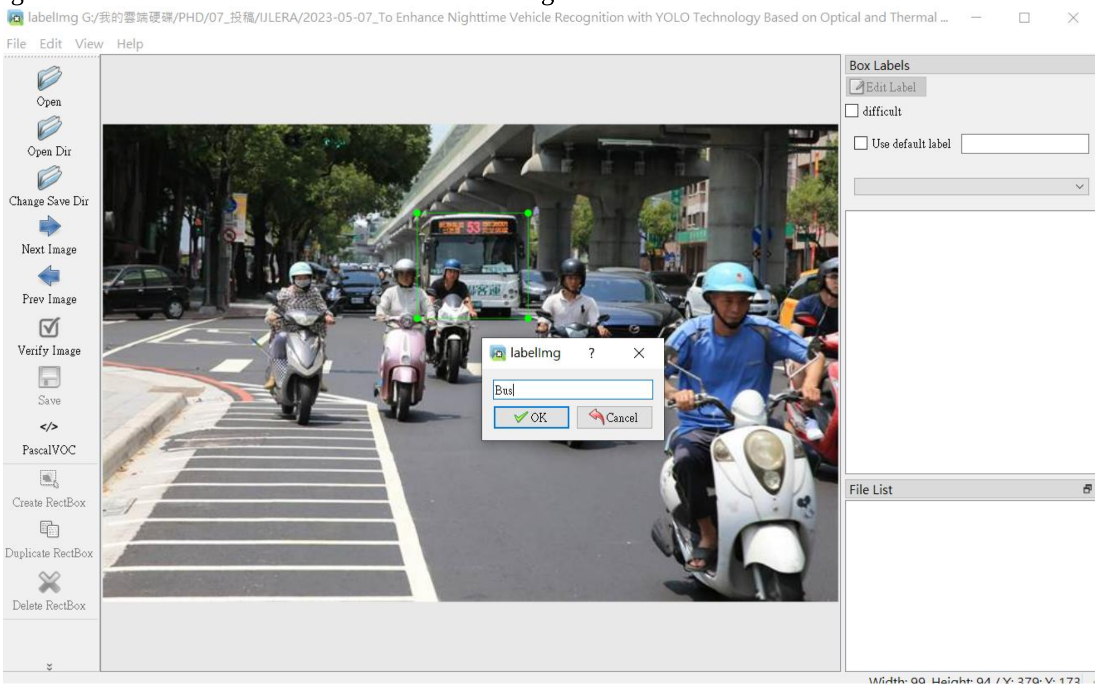

Fig. 5 Using the label Img software to label training data

After labeling a training sample using Label Img, the software will save the annotated sample as a separate file with the extension ".xml". This annotation file contains information such as the classification category of the training sample, its coordinates, as well as the X and Y dimensions of each bounding box. An example of an annotation file is shown in Fig. 6.


```
<annotation>
	<folder>images</folder>
	<filename>test01.jpg</filename>
	<path>D:\AI\09_Code\Library\masktest\images\test01.jpg</path>
	<source>
		<database>Unknown</database>
	</source>
	<size>
		<width>650</width>
		<height>540</height>
		<depth>3</depth>
	</size>
	<segmented>0</segmented>
	<object>
		<name>with_mask</name>
		<pose>Unspecified</pose>
		<truncated>0</truncated>
		<difficult>0</difficult>
		<bndbox>
			<xmin>267</xmin>
			<ymin>262</ymin>
			<xmax>431</xmax>
			<ymax>396</ymax>
		</bndbox>
	</object>
	<object>
		<name>with_mask</name>
```


Fig. 6 Content of Pascal VOC file

The required file format for our experiment is YOLO, where each line represents an object and includes the object name (object-class), the object's center x and y coordinates as a proportion of the image width and height (image_width and image_height), and the object's width and height as a proportion of the image width and height (target_width and target_height), as shown in Fig. 7. The calculation methods are as follows:

Center X = (xmin + (xmax – xmin) / 2) / width
Center Y = (ymin + (ymax – ymin) / 2) / height
W = (xmax – xmin) / width
H = (ymax – ymin) / height

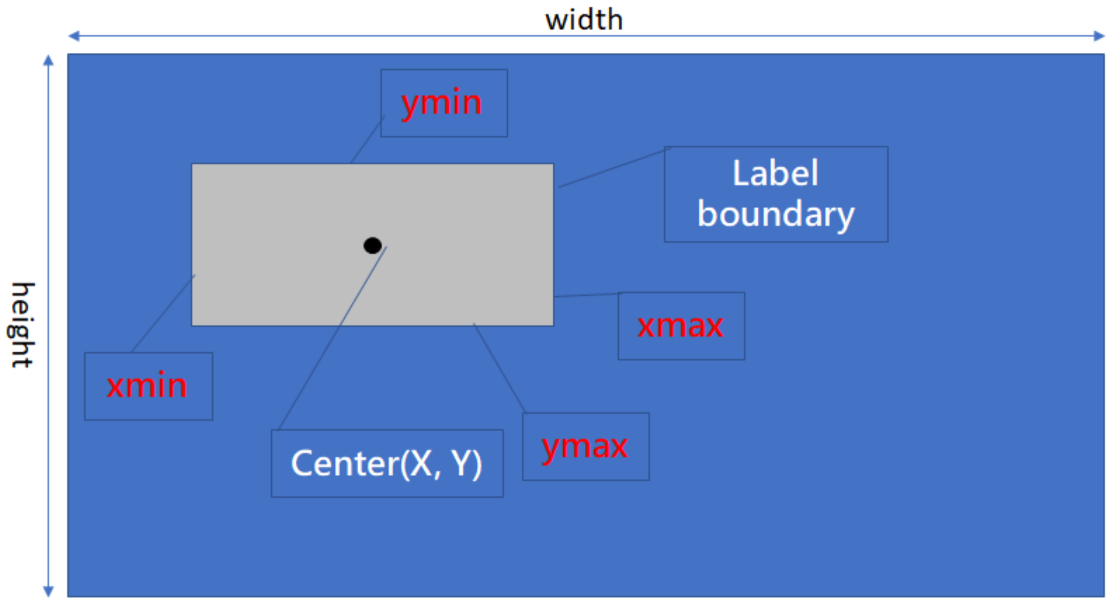

Fig. 7 YOLO labeling data format

YOLO Label File in Text Format .txt file, where each image corresponds to a txt file. To create this, a folder named YOLOv4 is created, and both the annotated images and txt files are placed in it. Five configuration files are created: names, test.txt, val.txt, yaml, and yolov4.cfg files. The classification method is shown in Fig. 8.

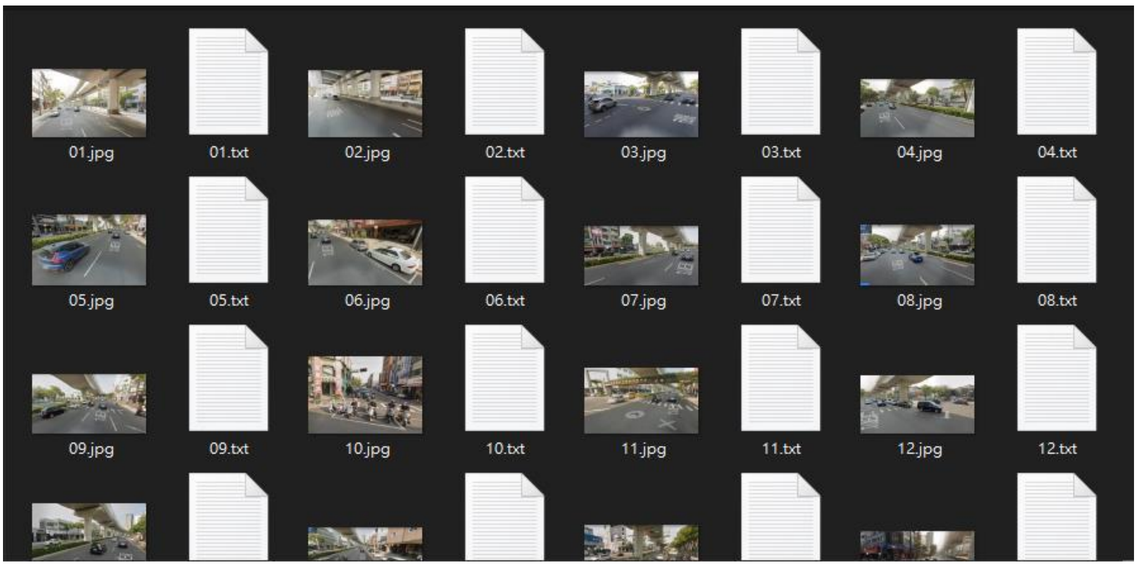

Fig. 8 Training vehicle images and their corresponding txt files of Labels are shown in the figure

The content of these files includes obj.names, which is a label list. In this study, the label name is "car." YOLO needs to read this file during training and prediction. test.txt is the training file and val.txt is the validation file. The main content of test.txt is a list of training file names, and the main content of val.txt is a list of validation file names. The train.txt file used for training this model is randomly divided into 80% for training and 20% for testing. A program was used to randomly classify 3000 images, including 1500 daytime vehicle photos and 1500 nighttime vehicle photos, resulting in 2400 samples for training and 600 samples for validation. The structure of the training data is shown in Fig. 9.

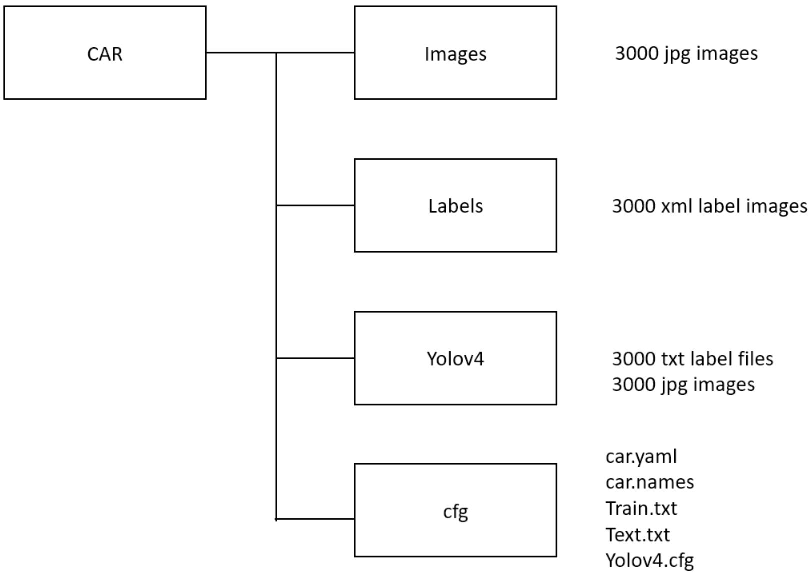

Fig. 9 Training data structure

The yolov4.cfg file contains the parameters required for training, and each module has its own parameters. To create a neural network model using PyTorch, the batch size is set to 64, which means how many samples are used for learning in each batch. In order to reduce the load on the GPU during training, the subdivisions are set to 8, which means the total batch is divided into 8 mini-batches. After all 8 mini-batches have been processed, the average gradient is obtained and updated using the gradient descent method: weight = weight - learning rate * gradient (batch). With the same number of epochs, the number of batches required for a larger batch size is reduced, so it can reduce training time. The gradient calculation for a larger batch size is more stable, and the model training curve is smoother. In fine-tuning, a larger batch size may achieve better results. The exposure is set to 1.5, which randomly changes the image brightness during training, with values between 1 and 1.5. During training, the learning rate is adjusted at 400,000 and 450,000 iterations to 0.0001 and 0.00001, respectively. With the increase of the learning rate, the model may move from underfitting to overfitting, especially on large datasets, as shown in Fig. 10.

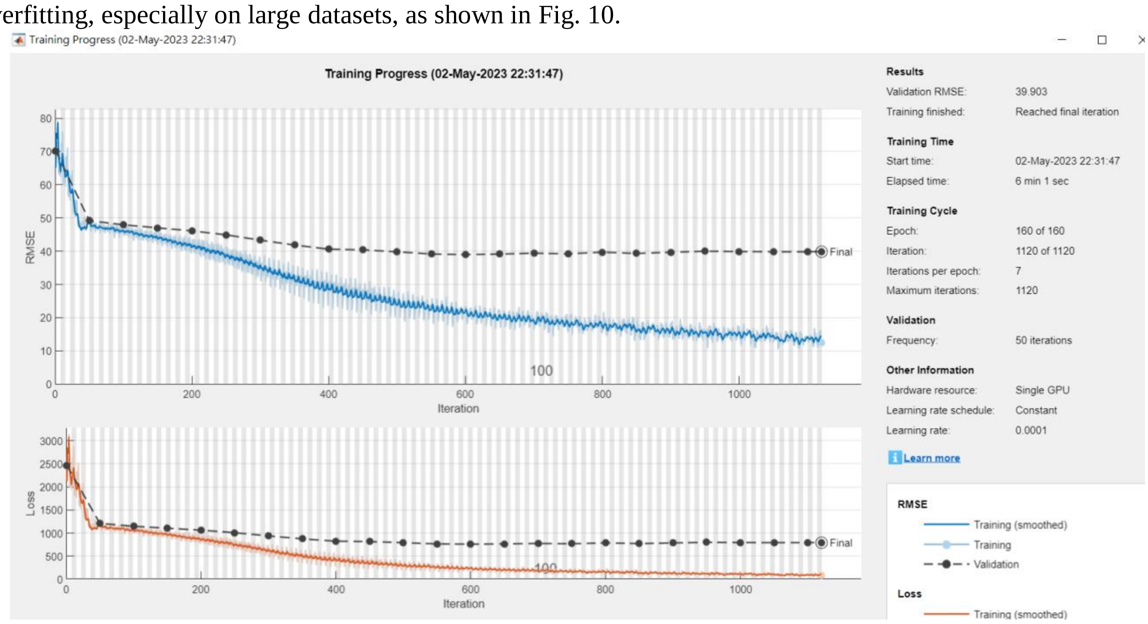

Fig. 10 Training Progress and validation RMSE

IV. EXPERIMENT RESULT

1. Model training

1.1 Model evaluation methods

In this study, we built a YOLOv4 neural network model using Pytorch, which provides GPU computing and speeds up the overall training process. After training the model, we input intersection images into the trained AI model for vehicle recognition. For deep learning network training, (1) precision, (2) recall, and (3) average precision (AP) are typically used as the three indicators to evaluate the model. Precision is defined as Precision = TP / (TP + FP), which focuses more on the accuracy of positive predictions. Recall is defined as Recall = TP / (TP + FN), which measures the proportion of positive samples that can be predicted correctly. AP is the average of the APs for each class.

TP (True Positive) refers to cases where the actual condition is true and the prediction is also positive, meaning the prediction is correct. TN (True Negative) refers to cases where the actual condition is true and the prediction is also negative, meaning the prediction is correct. FP (False Positive) refers to cases where the actual condition is false but the prediction is positive, meaning the prediction is incorrect. FN (False Negative) refers to cases where the actual condition is false but the prediction is negative, meaning the prediction is incorrect. These are shown in the Confusion matrix in Table 1.

Table 1 Confusion matrix

		Predicted condition	
	Total population = P + N	**Positive (P)**	**Negative (N)**
Actual condition	**Ture (T)**	**True positive (TP)**	**False negative (FN)**
	False (F)	**False positive (FP)**	**True negative (TN)**

In multi-class object recognition, a curve can be drawn for each class based on recall and precision, and AP is the area under that curve. mAP is the average of AP for each class.To evaluate the performance of the model, this paper adopts mAP (Mean Average Precision), a mainstream evaluation metric used to judge the performance of the target recognition model. In related algorithms in the field of image recognition, precision and recall are often used, and the higher the score, the better the training results. However, there may be differences in actual situations. For example, if there is only one recognition result in a photo, the precision will show 100%, but it also means that the recall rate is low. On the other hand, if the recognized results are all the same, the model will appear to have a high recall rate but a relatively low precision rate. Therefore, the precision-recall curve is usually used to evaluate the model, and AP is the area under the curve, while mAP is the average of AP for all categories. When the value is closer to 1, it means that the model has more accurate and efficient recognition ability. The steps to calculate AP are as follows: 1. Use the model to obtain predicted scores. 2. Convert the predicted scores to class labels. 3. Calculate the confusion matrix TP, FP, TN, FN. 4. Calculate precision and recall. 5. Calculate the area under the precision-recall curve. 6. Measure the average precision. The method of calculating mAP is to find the average precision (AP) for each class and then average them for multiple classes. The formula for average mean precision is as follows:

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{AP}_i$$

1.2 Model Training Results

The recall, precision, val GIOU, and mAP obtained from 900 iterations of testing in this study are shown in Fig. 11. Val GIOU is the result obtained from calculating IoU between the predicted target and ground truth. If the IoU is greater than a threshold, it is considered TP, and if it is less than the threshold, it is considered FP. The mAP is the average precision. The GIoU loss value decreases relatively as the number of training iterations increases, while the mAP increases, which indicates that the loss value becomes smaller.

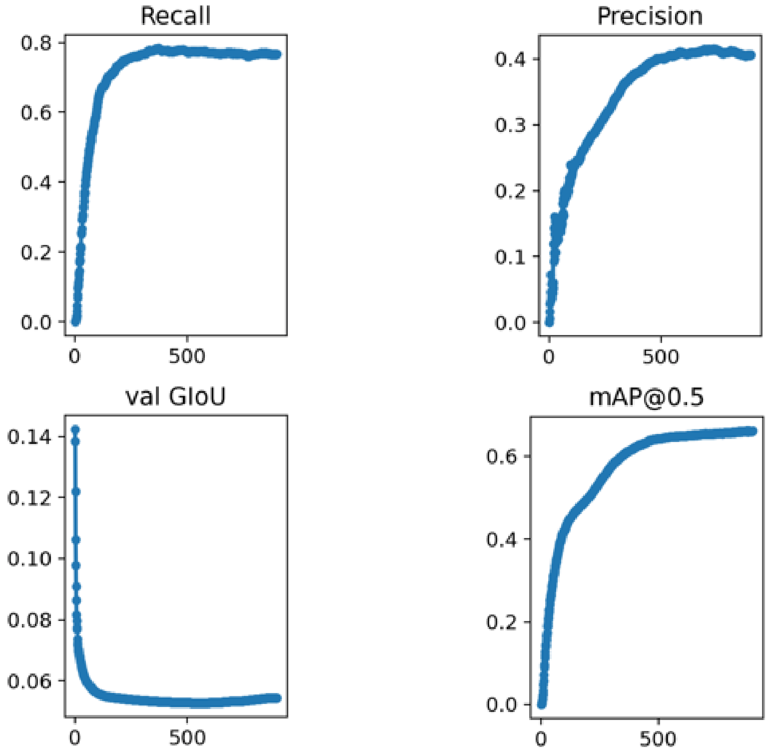

Fig. 11 Curve distribution plots of the model for Recall, Precision, validation Generalized IoU (GIoU), and mAP

2. Experimental results and analysis

Using the model trained in the previous chapter, we proceeded with verification testing. We captured three segments of optical and thermal images at a certain intersection on Section 3 of the Ring Central Road in Taichung City, totaling 543 seconds of video footage. We extracted one frame per second as test images and counted a total of 157 vehicles in the footage. The main types of vehicles identified in the images were cars, buses, and trucks. The distance of the target object was divided into two positions: position 1 at a distance of 10 meters (indicated by the red dashed line) and position 2 at a distance of 30 meters (indicated by the yellow dashed line), as shown in Fig. 12.

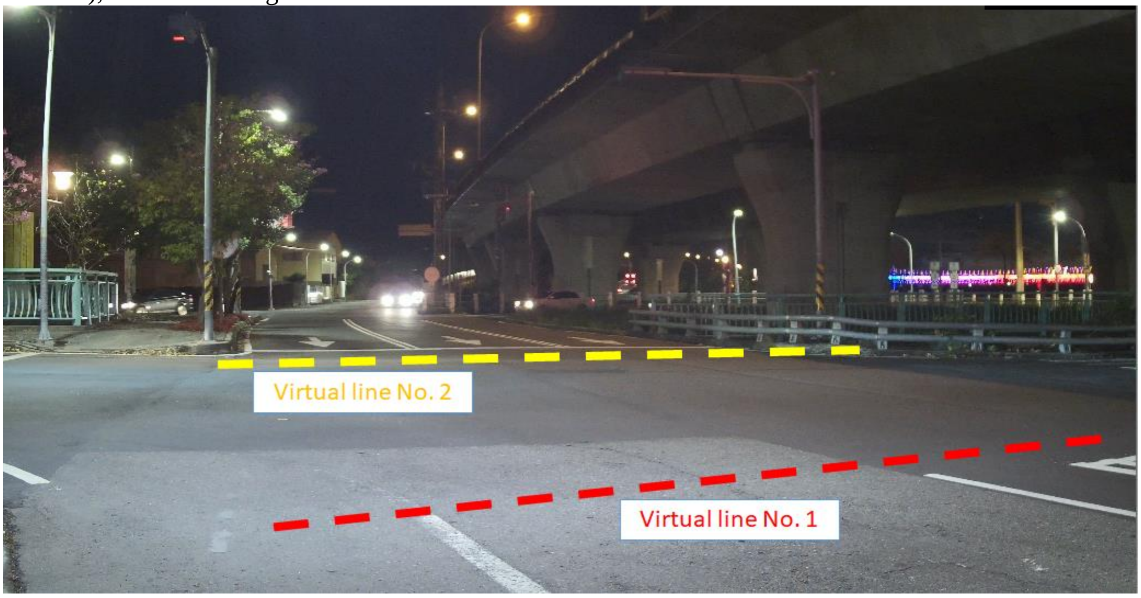

Fig. 12 Verification of the tested intersection (red dashed line represents test location number 1, yellow dashed line represents test location number 2)

According to the test analysis results, the traditional optical image recognized a total of 146 vehicles in the experiment, while the thermal image recognized 153 vehicles. The recognition rates of optical and thermal images were tested at distances of 10 and 30 meters, and the results are shown in Table 2. It can be found that the thermal image has better recognition rates for different distances and different types of vehicles. The main difference between the two in recognition is that at location 2, which is farther away, the image features are relatively blurred compared to location 1. The influence of the car lights on the optical image is much greater than that on the thermal image, which is the main reason for the difference in recognition between the two images. Additionally, we found that the recognition rate for buses was better in optical images than in thermal images, and it is speculated that the reason for this is that the training image dataset for buses in the thermal image is relatively small, which causes this phenomenon.

Table 2 The recognition rates of optical and thermal images at different distances.

Distances \ Image Dataset		Optical images	Thermal images
Virtual line No. 1 (10m)	Car	0.828	0.901
	Bus	0.913	0.917
	Truck	0.901	0.892
Virtual line No. 2 (30m)	Car	0.691	0.787
	Bus	0.724	0.691
	Truck	0.616	0.644

In addition, when the recognized object is further away from the lens, it is prone to blurring and the problem of the vehicle blending with the background, resulting in misclassification, as shown in Fig. 13. Due to the physical characteristics of thermal imaging that sense heat energy, the model is also prone to confusion when the front of the vehicle is close, as shown in Fig. 14.

Fig. 13 The optical image misclassified the background buildings as a bus

Fig. 14 Thermal imaging mistakenly identified regular cars as trucks

The distance between the vehicle and the camera can also affect the accuracy of recognition, as shown in Fig. 15. In the optical image, the accuracy of the two vehicles in front of the intersection is as high as (left) 0.87 and (right) 0.96, but the recognition rate of the rear vehicle is only 0.46. However, at the same time, in Fig. 16, there is no significant difference in the vehicle recognition rate in the thermal image, and the recognition rate of the rear vehicle at a longer distance in the thermal image is much higher than that in the optical image.

Fig. 15 Differences in vehicle recognition of optical images at different distances

Fig. 16 The difference in vehicle recognition of thermal imaging at different distances

The study also found that the headlights of vehicles at night can affect the recognition rate, as shown in Fig. 17. Traditional optical images are easily affected by the halo effect of car lights, which reduces the recognition rate. However, at closer distances, vehicles can still be recognized, but at farther intersections where many car lights are flashing and overlapping, the vehicle features become blurred, and the model cannot accurately determine whether it is a vehicle or not, resulting in ineffective recognition of vehicles affected by car lights. In contrast, as shown in Fig. 18, the recognition results of thermal images can still clearly identify vehicles at farther distances, with recognition rates of 0.85 and 0.84, respectively. The thermal image is not significantly affected by the halo effect of car lights during night-time vehicle recognition, and can perform better in this scenario.

Fig. 17 Recognition results of optical images under the influence of car headlights

Fig. 18 Recognition results of thermal imaging under the influence of vehicle headlights

V. CONCLUSION

This study utilized Yolo technology, a deep learning technique within AI, to recognize vehicles during night driving, while comparing the difference in recognition performance between optical and thermal imaging. After comparing several deep learning models, the Yolo method was selected for its stronger recognition ability and real-time performance. The pytorch model library was also used to train the model, and the impact of vehicle labeling on image recognition was investigated. It was found that the labeling of vehicles had a significant impact on image recognition performance. If a model trained with optical images is used to recognize thermal images, the recognition rate will decrease, and vice versa. Therefore, this study separately labeled and trained models with optical and thermal images and used different models for vehicle recognition based on the image source. When tested at distances of 10 and 30 meters from the target vehicle, the proposed method outperformed single models. The study's results demonstrate that thermal imaging can improve nighttime vehicle recognition and can be further integrated into intersection surveillance cameras for various nighttime vehicle recognition applications such as traffic accidents and violations. The proposed method can be applied in practice.

DISCUSSION

This study found that the headlights of nighttime vehicles have a significant impact on image recognition. The principle of thermal imaging is to use thermal energy to form an image, so the impact of headlights is relatively small compared to traditional optical imaging. However, to completely eliminate the influence of headlights, future research could focus on first processing the impact into grayscale and using the texture of the image object for recognition, thereby reducing the impact of headlights on recognition accuracy at night.

REFERENCES

[1] A. M. Ghoreyshi, A. AkhavanPour, A. Bossaghzadeh, Simultaneous Vehicle Detection and Classification Model based on Deep YOLO Networks, *International Conference on Machine Vision and Image Processing*, 2020,1-6.

[2] A Bochkovskiy, Wang C, Liao, H M. YOLOv4: Optimal Speed and Accuracy of Object Detection,*Computer Vision and Pattern Recognition*.2020, 10934.

[3] C. Kumar B., Punitha R., Mohana, YOLOv3 and YOLOv4: Multiple Object Detection for Surveillance Applications., *Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2020, pp. 1316-1321.

[4] H. Wang, J. Hou, N. A Chen, Survey of Vehicle Re-Identification Based on Deep Learning, *IEEE Access*, 2019, vol. 7, pp. 172443-172469.

[5] H. K. Leung, X. Z. Chen, Chao-Wei Yu, Hongyi Liang, Jian Yong Wu, Yen-Lin Chen, A deep-learning-based vehicle detection approach for insufficient and nighttime illumination conditions, *Department of Computer Science and Information Engineering*, 2019, 9(22), 4769

[6] H. Song, H. Liang, H. Li, Vision-based vehicle detection and counting system using deep learning in highway scenes, *European Transport Research Review*, 2019, volume 11, Article number: 51.

[7] J. Redmon, A. Farhadi, YOLO 9000: Better, Faster, Stronger, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6517-6525.

[8] Janiesch Christian, Patrick Zschech, Kai Heinrich, *Machine learning and deep learning*, 2021, Electron Markets 31, 685–695.

[9] Karen Simonyan, Andrew Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, *Computer Vision and Pattern Recognition*, 2014.

[10] L. Lianqiao, C. Xiai, Z. Huili, W. Ling, Recognition and Application of Infrared Thermal Image Among Power Facilities Based on YOLO, *Chinese Control And Decision Conference (CCDC)*,2019, pp. 5939-5943.

[11] M. Teutsch, T. Mueller, M. Huber, J. Beyerer, Low resolution person detection with a moving thermal infrared camera by hot spot classification, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014.

[12] V. Sindhu, S. Vehicle, Identification from Traffic Video Surveillance Using YOLOv4, *5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2021, pp. 1768-1775.

[13] Vuong Xuan Can, Phan Xuan Vu, Mou Rui-fang, Vehicle Detection and Counting Under Mixed Traffic Conditions in Vietnam Using Yolov4, *International Journal of Advanced Research in Engineering and Technology*, 2021.

[14] A. M. Ghoreyshi, A. AkhavanPour, A. Bossaghzadeh, Traffic vehicle cognition in severe weather based on radar and infrared thermal camera fusion, *Measurement Science and Technology*, 2021, 32.