ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 11 (2024) No. 3, pp. 97-111
DOI:10.33168/JLISS.2024.0307

Moving Vehicle Logo Detection based on YOLOv7 under Complex Lighting Conditions

Aryssa Amanda Azhar, Noramiza Hashim

Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Selangor, Malaysia
noramiza.hashim@mmu.edu.my (Corresponding author)

Abstract. Vehicle logo detection is an important component of the Intelligent Transportation System (ITS) that has received a lot of attention from researchers. Its ability to provide supplementary information enhances the process of vehicle identifications, making it an essential tool in a variety of ITS applications such as parking management, traffic monitoring, toll collection and autonomous vehicles and advanced driver assistance systems (ADAS). These applications require the vehicle logo detection to work not only in an optimal lighting environment but also in complex lighting conditions as well. Our work focuses on a deep learning-based vehicle logo detection model that can detect logos under various lighting conditions such as low lighting, strong lighting, blurriness, and glares. A dataset of surveillance footage consisting of 10 common car brands in Malaysia is created for model training and testing. Furthermore, this study involves a comparative performance analysis of two deep learning models, namely YOLOv3 and YOLOv7 with image enhancement method, to evaluate their suitability for logo detection. Our findings reveal that YOLOv7 exhibits a performance improvement of 0.228% compared to YOLOv3 when applied to the complex lighting conditions dataset with the application of contrast limited adaptive histogram equalisation (CLAHE) as an image enhancement technique.

Keywords: vehicle logo detection, complex lighting, YOLO, deep learning

1. Introduction

The Intelligent Transportation System (ITS) has made a substantial impact on the growth of modern technology. This technology can be used in a variety of different settings, such as traffic management, toll collection, and parking enforcement. ITS mainly relies on two key components: licence plate recognition and vehicle logo recognition (Mao Yuxin & Hao Peifeng, 2019; Soon et al., 2019). The licence plate and the vehicle logo are both important pieces of information that can be used to identify a vehicle. The licence plate number provides the unique identification number of a vehicle, and the logo of the vehicle provides the identification of the brand of the vehicle. Just like a login system, where a user's email and password must match for access, in ITS, the licence plate and the vehicle logo should match for a vehicle to be granted entry into an access-controlled environment. The system uses these two pieces of information to authenticate the vehicle and ensure that only authorised vehicles are given access. Without both the licence plate and vehicle logo, the system would not be able to properly identify the vehicle and would not allow entry.

Many studies in the field of vehicle logo detection have demonstrated impressive results. For instance, He et al. (2017) proposed a model based on mask R-CNN that achieved high accuracy in logo detection using a dataset consisting of images captured under optimal lighting conditions. Similarly, Zhang et al. (2021) utilised a deep learning approach to detect logos in vehicle images, employing a dataset specifically collected under optimal lighting conditions. Furthermore, Jiajun Liu et al. (2019) contributed to the field by introducing VLD 1.0, a large-scale vehicle logo dataset that encompasses complex lighting conditions.

While these studies have contributed valuable insights into logo detection, their focus on normal or optimal lighting conditions raises concerns regarding the models' robustness in real-world settings. The limited availability of datasets capturing complex lighting scenarios further compounds this issue. Despite the importance of accurate logo detection in diverse lighting conditions, only a handful of studies have explored this aspect in depth. However,there is a limited availability of datasets that capture variations in lighting conditions, including low lighting, strong lighting, blurriness, and glare. This limitation hampers the development and evaluation of robust logo detection models capable of handling diverse lighting scenarios. In order to bridge this gap, this research proposes a novel dataset that encompasses optimal lighting environments as well as complex lighting conditions. Unlike existing datasets that primarily consist of static images, our dataset focuses on video sequences to provide a more realistic representation of real-world scenarios. By incorporating videos instead of static images, we aim to capture the temporal dynamics and potential variations in lighting conditions, further enhancing the evaluation of logo detection models. Additionally, the performance of state-of-the-art models like You Only Look Once version 3 (YOLOv3), You Only Look Once version 7 (YOLOv7) including the application of image enhancement methods on this dataset are compared to assess their effectiveness in handling complex lighting conditions.

2. Literature Review

The vehicle logo recognition in complex lighting is a process of recognizing and identifying vehicle logos in different lighting conditions like adverse weather, shadows, low light, and poor visibility. This process is made possible by using advanced image processing, machine learning, deep learning, and computer vision technologies. The goal of these technologies is to accurately identify a vehicle logo even in difficult lighting conditions. There have been numerous studies that have proposed various techniques for recognizing vehicle logos. These techniques can be broadly divided into two main categories: traditional machine learning methods and deep learning methods. Traditional machine learning methods typically involve using algorithms such as k-nearest neighbours, decision trees and support vector machines to classify images of vehicle logos. On the other hand, deep learning methods utilise neural networks, such as convolutional neural networks (CNN), to classify images of vehicle logos.

2.1. Image enhancement method

Image pre-processing plays a crucial role in preparing input images for logo recognition, regardless of the use of traditional machine learning or deep learning models. It involves various operations such as image enhancement, noise reduction, and image segmentation to enhance image quality and facilitate logo identification by recognition algorithms. This section explores different image enhancement techniques, including both traditional and deep learning methods, which have been utilised for vehicle logo recognition. Traditional image enhancement methods are algorithms designed to improve image quality by adjusting intensity values, reducing noise, and enhancing contrast. Examples of traditional methods include histogram equalisation, contrast stretching, smoothing, and sharpening filters. However, each of these methods has its own advantages and disadvantages.

Tang et al. (2019) proposed an approach for enhancing low-light images, addressing the common issue of over-enhancement that can lead to texture and feature loss. Their approach, known as strong light weakening and bright halo suppression, employs a bright channel prior on inverted images to weaken strong light regions. It then applies a dehazing enhancement algorithm combined with superpixel segmentation to further improve the image. The authors also incorporate a revised nonlocal denoising approach to minimise noise, resulting in better visual results compared to other techniques in the field. Zhao and Guo (2022) introduced adaptive gamma correction as a method for image enhancement, specifically for vehicle logo recognition. Their method involves converting the image into the HSV colour space, adjusting the brightness of the V channel, which represents image brightness. By adapting the gamma correction process, the authors aim to enhance the visibility of logos under varying illumination levels. The processed image is then converted back to the RGB colour space for logo detection. The authors reported achieving a result of 95.86% for vehicle logo recognition after applying their image enhancement method. J. Yang et al. (2022) proposed a low-light image enhancement technique that combines the fast and robust fuzzy C-means (FRFCM) clustering algorithm and the retinex theory. This technique begins by estimating the lighting conditions in the image using the retinex theory, followed by segmenting the image using the FRFCM algorithm. The approach effectively reduces noise and unwanted elements in the original image, improving the accuracy and reliability of enhancement. Additionally, the technique enhances texture details, revealing finer patterns and structures, which further enhances image visibility and facilitates analysis.

However, it is worth considering the contrast limited adaptive histogram equalisation (CLAHE) method, which has shown promising results in various domains, including unmanned aerial vehicle (UAV) images of overhead transmission systems (Yuan et al., 2023). The use of CLAHE in image enhancement has demonstrated improved accuracy in insulator detection (Yuan et al., 2023). Furthermore, a morphological technique based on CLAHE has been proposed for dark image enhancement (Pavan A C et al., 2023). Among the various image enhancement methods explored for vehicle logo recognition, CLAHE emerges as a promising choice due to its effectiveness in addressing low-light image problems. CLAHE has shown success in improving image quality and enhancing details in various domains, including the overhead transmission system domain (Yuan et al., 2023). By leveraging CLAHE, the proposed image enhancement technique in this paper can potentially achieve improved recognition accuracy and visibility for vehicle logos.

2.2. Traditional vehicle logo recognition

Traditional vehicle logo detection methods face several limitations when dealing with complex lighting conditions. These methods heavily rely on handcrafted features such as scale-invariant feature transform (SIFT), speeded up robust features (SURF), and histogram of oriented gradients (HOG), as suggested by Chen et al. (2017) and Psyllos et al. (2010). While these features have been widely used, they may not capture all the intricate variations and complexities present in logos under challenging lighting conditions. This can lead to incomplete or inaccurate representations of the logos, hindering the detection and recognition process. Furthermore, the dependence on fixed feature extraction techniques poses challenges in handling varying lighting conditions. The illumination changes can

cause significant alterations in the appearance of the logos, making it difficult for traditional methods to accurately localise and extract the logo regions. Jatupon Benjaparkairat & Pakorn Watanachaturaporn (2018) and Psyllos et al. (2010) highlight the limitations of these methods in handling images captured under different lighting conditions, such as bright sunlight, cloudy skies, heavy rain, and low-light situations.

Moreover, the use of conventional classifiers like nearest neighbour and support vector machines, as discussed by Chen et al. (2017) and Jatupon Benjaparkairat & Pakorn Watanachaturaporn (2018), may restrict the adaptability and discrimination capability of the logo detection system. These classifiers often rely on fixed decision boundaries, which may not be flexible enough to accommodate the variations introduced by complex lighting conditions. This can lead to reduced accuracy and lower recognition performance, particularly when dealing with similar-looking logos or logos captured under challenging lighting scenarios. The limitations of traditional vehicle logo detection methods underscore the need for more robust and adaptable approaches that can effectively handle complex lighting conditions. The development of advanced feature extraction techniques and machine learning algorithms that are capable of capturing and modelling the intricate variations in logos under different lighting conditions is crucial. By overcoming these limitations, researchers can enhance the accuracy and reliability of vehicle logo detection systems, enabling better performance in real-world scenarios.

2.3. Deep learning vehicle logo recognition

In recent years, there has been a significant advancement in logo detection through the rapid development of computer vision and deep learning algorithms. These algorithms have become the mainstream approach, offering automated identification and extraction of high-level features without manual feature extraction. Deep learning object detection methods can be categorised into classic CNNs, two-stage detectors, and one-stage detectors. Classic CNNs have been explored for vehicle logo recognition and detection. For instance, Mao Yuxin & Hao Peifeng (2019) proposed a VGGNet-based CNN model for vehicle logo recognition, achieving an average recall of 94.8% by incorporating licence plate locations and vehicle symmetry features. However, this method was not extensively tested across various illumination conditions. Another study by Soon et al. (2019) introduced a CNN-based vehicle logo recognition model with a whitening transformation technique, attaining accuracy rates of 99.07% and 99.13% (with transformation). Nonetheless, the training time was relatively long, taking 7 hours. Huang et al. (2015) presented a CNN-based vehicle logo recognition system that utilised pre-training strategies based on PCA, resulting in 90% and 97% accuracy under different lighting conditions. Huan et al. (2017) combined the Hough transform with deep belief networks (DBNs) for vehicle logo retrieval, achieving an average accuracy of 90%. However, this approach faced challenges when dealing with vehicle logo images with simple textures or similar colours to the background. In summary, classic CNN methods exhibit promising performance in vehicle logo recognition and detection, leveraging features such as symmetry, pre-training strategies, and the Hough transform. Nevertheless, limitations include the lack of testing across various illumination conditions, long training times, and difficulties with simple logo textures or similar colours to the background.

In the domain of object detection, two-stage and one-stage detectors are commonly used approaches. Two-stage detectors, like the cascaded deep convolutional network (CDNN) proposed by Yu et al. (2021), offer precise localization and classification of objects, including vehicle logos. These methods employ a region proposal network (RPN) in the initial stage to identify potential object regions, reducing the search space and enhancing efficiency. The subsequent stage, the detection network, refines and categorises the region proposals, resulting in improved accuracy. The Faster R-CNN architecture, studied by Arinaldi et al. (2018), excels in detecting objects under challenging conditions, such as overlapping or low light night-time scenarios, making it suitable for traffic video analysis. However, two-stage detectors can be computationally expensive and time-consuming. Additionally, the selective search process used in Faster R-CNN may entail unnecessary processing, impacting efficiency and performance in low-resolution images or images with similar objects. Mohammad Wahyudi Nafi'i

et al. (2019) proposed mask R-CNN for detection of vehicle brands and types. The problem the author tried to solve was detection on similar looking vehicle logos. They employed a bounding box within the bounding box method, where separate bounding boxes were created for the vehicle and the logo. If the logo bounding box was detected within the vehicle bounding box, the brand and type of the vehicle can be established based on the identified logo class.

On the other hand, one-stage detectors, such as You Only Look Once (YOLO) and Shot MultiBox Detector (SSD), offer faster and more efficient object detection compared to two-stage detectors. These methods perform detection in a single pass of a convolutional neural network (CNN), making them well-suited for real-time applications. The Shot MultiBox Detector (SSD) has been enhanced for fast vehicle detection in traffic scenes by replacing the VGG16 backbone with MobileNetv2 and incorporating a channel attention mechanism and a Deconvolution Module (Z. Chen et al., 2022). This improves the model's efficiency and feature extraction capabilities. However, further optimization is needed to accurately detect small vehicles, which can be achieved through increasing input resolution, improving data quality, and implementing self-learning anchors. Another variant of the SSD model was proposed by Zhang et al. (2022) for multi-object detection in night traffic scenarios. The framework utilises DenseNet as the feature extractor, providing advantages such as parameter efficiency and implicit supervision. However, the model may struggle to detect objects accurately under unfavourable conditions where objects are obscured. These advancements in the SSD model offer benefits such as faster inference and improved feature extraction. Nevertheless, challenges remain in accurately detecting small vehicles and handling adverse conditions in object detection tasks. YOLO variants, as exemplified by Yang et al. (2019) and Zhou et al. (2020), have demonstrated promising results in vehicle logo recognition, even in complex scenes and motion-blurred conditions. Additionally, there is a traffic impact assessment system by Jin Jie Ng et al. (2023) that employs YOLOv5 for tasks such as detecting vehicle logos, counting vehicles, and classifying them. Most of the existing research on vehicle logo detection has primarily focused on utilising the YOLO framework, particularly YOLOv3 as proposed by Zhou et al. (2020), Yang et al. (2019), and Zhao & Guo (2022). However, this particular study aims to address the complex lighting conditions by implementing the latest version, YOLOv7, to assess and evaluate the performance of the models.

3. Proposed Method

3.1. Dataset

In this paper, we introduce a novel dataset called the Malaysia Vehicle Logo Complex Lighting Dataset (MVLCL), specifically designed for vehicle logo detection purposes. In order to capture a diverse range of lighting conditions, the dataset includes footage obtained from a housing community, which recorded vehicles entering and exiting the area. The surveillance footage captures various lighting scenarios throughout the day and under various weather conditions. As a result, the MVLCL dataset contains footage from daylight, nighttime, and even rainy conditions, providing a comprehensive representation of the challenges posed by real-world lighting variations. The dataset includes both frontal and back views of the vehicles, resulting in a collection of 4000 videos. During the dataset creation process, each video was carefully examined to select the relevant footage suitable for this project specifically focusing on videos with frontal view vehicles. Then, removed any irrelevant footage that did not feature vehicles, instead capturing moving objects such as animals, flags, or leaves due to the motion-based operation of the CCTV cameras.

As the footage was reviewed individually, the relevant footage is organised into two separate folders based on the lighting conditions: optimal lighting and complex lighting based on the vehicle brands. The footage covers a range of up to 19 car brands commonly used in Malaysia. However, not all brands have a significant number of videos. Therefore, the car brands with a significant number of videos are included in this dataset. This approach ensures that the dataset maintains a sufficient number of instances for each car brand, enabling a comprehensive analysis of logo detection performance across

different brands. Furthermore, the relevant footage was trimmed to extract specific frames capturing the exact moments when vehicles passed through the gate. This process resulted in shorter video clips, focused on the critical instances relevant to logo detection tasks. Trimming the footage allows for a more targeted analysis and reduces unnecessary data volume, making the dataset more efficient for research purposes. All the videos in the dataset are in the MP4 format, with a resolution of 720p (1280x720). The videos have a frame rate of 30fps and their durations range from 1.1 seconds to 10 seconds

Data augmentation has been applied to the dataset to enhance the dataset's diversity and improve the models' generalisation ability. These techniques introduce additional variations to the images, expanding the dataset and enhancing its representation of real-world scenarios. In this study, the following data augmentation methods were employed:

- **Flip/Rotation:** The images were flipped and rotated within a specified range of angles, which are horizontal flipping, rotation of 20 degrees, rotation of -20 degree and rotation of 30 degree. This augmentation simulates vehicles with different orientations and helps the models learn to detect logos from various angles.
- **Cropping:** The images were cropped at scales of 60% and 75% to increase their size. This augmentation allowed the models to identify vehicle logos at different distances and scales, enhancing their ability to generalise to real-world scenarios.
- **Weather condition:** Synthetic rain effects were applied to the images, simulating challenging weather conditions. This augmentation introduces visual noise and occlusions, mimicking rainy environments and testing the models' ability to accurately detect logos under adverse weather conditions.

As shown in Figure 1, the MVLCL dataset is divided into two subsets, addressing complex lighting challenges: optimal lighting and complex lighting conditions. The optimal lighting subset ensures optimal lighting for clear logo visibility, while the complex lighting subset includes real-world scenarios with low lighting, strong lighting, blurriness, and glares. This allows realistic evaluation of model robustness and performance in complex lighting situations. The MVLCL dataset encompasses logo samples from 10 prominent Malaysian vehicle brands: BMW, Honda, Hyundai, Mazda, Mercedes, Nissan, Perodua, Proton, Toyota, and Volkswagen, with 60 videos per brand which sum up to 1200 videos in a dataset.

(a) (b)

Fig.1: Example of images in the MVLCL dataset showing (a) optimal lighting and (b) complex lighting subsets.

Figure 1 illustrates two examples of datasets with different lighting conditions for vehicle logo identification. In (a), the dataset represents optimal lighting conditions, where the logo is clearly visible. However, in (b), the dataset demonstrates complex lighting conditions, including low lighting, strong

lighting, blurriness, and glares, which make it challenging to identify the logos.

3.2 Image enhancement

The contrast limited adaptive histogram equalisation (CLAHE) is a method utilised to enhance the contrast of images. It is an enhanced version of the adaptive histogram equalisation (AHE) method, which has a tendency to produce over-exaggerated contrast enhancement. CLAHE aims to mitigate this issue by limiting the amount of contrast enhancement applied to the image. The idea behind CLAHE is to divide the image into small regions, called tiles, and then equalise the histogram of each tile independently. This helps to preserve the local details and prevent the over-enhancement of the contrast. The contrast enhancement is limited by restricting the difference between the highest and lowest pixel values in each tile. In this way, CLAHE enhances the contrast of the image while avoiding the excessive contrast enhancement seen with AHE.

The implementation of CLAHE involves converting an RGB image into a grayscale image, as CLAHE is designed to operate specifically on grayscale images. This is a crucial step in the process as the algorithms and calculations used in CLAHE are optimised for grayscale images, rather than RGB images. The conversion to grayscale ensures that the results produced by CLAHE are more accurate and relevant to the image being processed. The CLAHE took the grayscale image and divided the entire image into smaller regions called tiles to help correct the contrast in smaller regions instead of the whole image. Next, for each region, a histogram is generated to represent the distribution of the pixel intensities in the tiles. A mapping function is generated based on the cumulative distribution function (CDF) of the histograms. The purpose of the mapping function is to enhance the contrast of the image. This function is then applied to each of the image regions to improve the contrast. Finally, to ensure a smooth transition between adjacent regions, bilinear interpolation is utilised to merge the tiles together. This results in a more natural-looking image with improved contrast and minimal loss of detail as shown in Figure 2.

(a) Original image (b) CLAHE image

Fig.2: Image enhancement given original images (a). The enhanced images are shown in (b) after CLAHE is applied.

Figure 2 showcases the effect of image enhancement on a dataset with complex lighting conditions. In (a), the example represents the original image under challenging lighting conditions. However, in (b), the image enhancement technique, CLAHE, has been applied, resulting in an improved dataset with enhanced visibility and clarity.

3.2 Vehicle logo detection model

The YOLOv7 implementation by Wang et al. (2022) is the newest version of YOLO that can efficiently predict video inputs ranging from 5 fps to 160fps. The YOLOv7 has 75% fewer parameters and 36% lesser computational time compared to YOLOv4 which also uses a trainable bag of freebies to increase the model accuracy. The overall architecture of YOLOv7 is that it uses extended efficient layer aggregation networks (E-ELAN) as the backbone. The layer aggregation techniques allow the combination of multiple networks into a single network. It also allows the model to extract more complex features from the input image. The aggregation layer for YOLOv7 consists of VoVNet, CSPVoNet, ELAN and E-LAN networks as in Figure 3. Each network processes the input image at a different scale, with the smaller scales focusing on detecting smaller objects and the larger scales focusing on detecting larger objects.

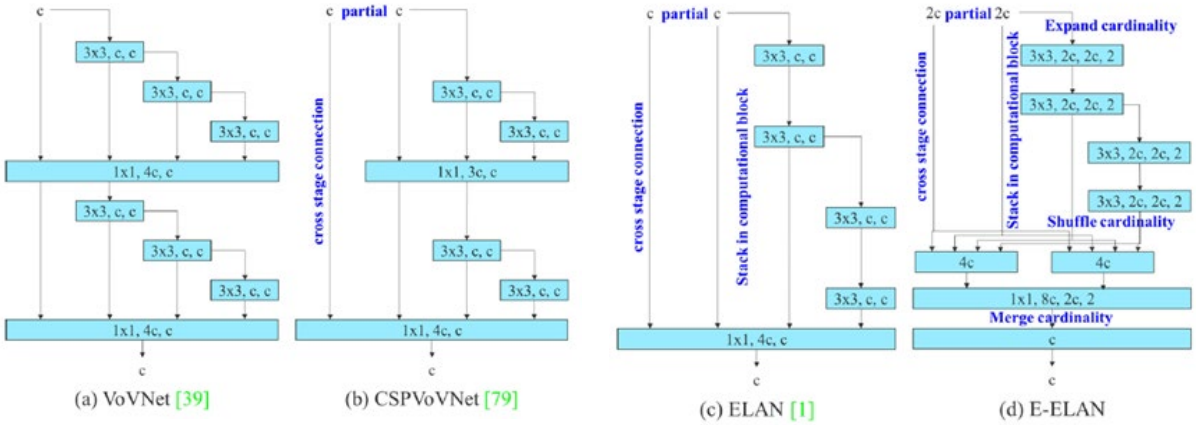

Fig.3: Extended efficient layer aggregation networks (Wang et al., 2022)

Once the output from each network is obtained, YOLOv7 introduces model scaling for concatenate-based models to combine the output into a single network. The purpose of model scaling is to ensure that the outputs of the different networks are at the same scale before they are concatenated and processed by the prediction layer. To achieve this, YOLOv7 uses a technique called "scaled concatenation" which adjusts the output of each network by applying a scale factor to the output before concatenation. Scale factors are dependent on resolution, depth, width, and stage. In YOLOv7, the architecture uses a type of convolution called RepConvN, which is a re-parameterized version of traditional convolution. The key difference between RepConvN and traditional convolution is that RepConvN does not use an identity connection. An identity connection is a type of connection that allows the output of a layer to be directly added to the input of the next layer, it is a type of shortcut connection. In YOLOv7, the idea behind using RepConvN is to replace a residual or concatenation convolutional layer with a re-parameterized convolution, eliminating the need for the identity connections that were previously used. The use of RepConvN in YOLOv7 is intended to improve the model's performance by avoiding the use of identity connections which can lead to overfitting. By using RepConvN, the model can take advantage of the strengths of different networks while reducing the number of parameters and computation cost, which helps to improve the model's efficiency.

Moreover, a label assigner mechanism was also introduced to handle the problem of label assignment in object detection tasks. This mechanism uses a combination of network prediction results and ground truth to assign soft labels to training data. Traditionally, label assignment in object detection tasks is done by directly referring to the ground truth and generating hard labels based on fixed rules. Hard labels are binary values such as 0 and 1 indicating the presence or absence of a particular class or label. However, YOLOv7 uses a more sophisticated approach, considering both the network predictions and the ground truth. It assigns soft labels, which are numeric values that represent the degree of confidence or certainty that a particular class or label applies to a given input data. The soft labels are calculated and optimised using methods that take into account not only the ground truth, but also the quality and distribution of the prediction output.

4. Experiment and Analysis

The experiments in this study utilise YOLO-based models and evaluate their performance using the MVLCL dataset. The dataset used for training and testing was annotated using Roboflow, an annotation tool that simplifies the annotation process and provides labelled data for model training and evaluation.

- **Data split:** The dataset is divided into a training set of 1000 images and a testing set of 200 videos. The training set consists of 500 samples each from ideal lighting and complex lighting images, while the testing set includes 100 samples each from ideal lighting and complex lighting videos.
- **Training:** The models are trained using optimal lighting images with input images having resolution of 256 x 256 x 3. The training is conducted for 100 epochs, with a batch size of 8 and a weight decay of 0.0005.
- **Testing:** The models are tested using three types of data: optimal lighting videos, complex lighting videos, and complex lighting videos with the application of image enhancement, CLAHE. The testing is conducted with a confidence level set at 0.5 and an Intersection over Union (IOU) threshold of 0.65.
- **Evaluation:** Several metrics are used to assess model performance, including precision, recall, F1-score, inference time, and NMS time.

Precision measures the accuracy of the model in identifying specific objects. It considers the precision at different recall levels and calculates the average precision across all levels. The precision is calculated as the ratio of true positives to the sum of true positives and false positives.

$$\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives}) \quad (1)$$

Recall evaluates the model's ability to detect all instances of objects in the dataset. It measures the proportion of true positive detections out of all ground truth objects. The recall is calculated as the ratio of true positives to the sum of true positives and false negatives.

$$\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False Negatives}) \quad (2)$$

F1 Score combines precision and recall into a single metric to provide a comprehensive assessment of the model's performance. It is the harmonic mean of precision and recall.

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (3)$$

Inference time measures the computational cost of image processing and prompt predictions. It quantifies the time taken by the model to process an image and provide the output predictions.

4.1. Comparison of models

Table 1 : Model comparison between YOLOv3 and YOLOv7 using three different datasets: optimal lighting, complex lighting, and complex lighting with CLAHE enhancement.

Model	Data	Total videos	Precision	Recall	F1- score	Inference
YOLOv3	Optimal lighting	100	0.991	0.97	0.98	5.0 ms
	Complex lighting	100	0.521	0.239	0.328	6.1 ms
	Complex lighting + Enhancement	100	0.507	0.236	0.322	4.8 ms
YOLOv7	Optimal lighting	100	1.0	0.97	0.98	4.3 ms
	Complex lighting	100	0.598	0.229	0.33	3.6 ms
	Complex lighting + Enhancement	100	0.735	0.306	0.43	4.8 ms

Table 1 presents a model comparison between YOLOv3 and YOLOv7 using three different datasets: optimal lighting, complex lighting, and complex lighting with the application of image enhancement using the CLAHE technique. The comparison evaluates the models based on metrics such as precision, recall, f1-score, and inference time.

In the case of optimal lighting video as shown in Table 1, both YOLOv7 and YOLOv3 perform exceptionally well with high precision, recall, and F1-score values. The YOLOv7 achieves perfect precision, indicating no false positive detections. These results suggest that both models are highly accurate and reliable in detecting and localising objects in the optimal lighting videos. Since both of the models have been trained on the optimal lighting images, which means they have learned from images that have consistent lighting conditions, backgrounds, and object appearances. This training allows the models to capture and understand the patterns and features specific to the optimal lighting images, resulting in higher precision, recall, and overall performance.

When evaluating the performance of YOLOv3 and YOLOv7 on complex lighting videos, they exhibit similar performance. However, the YOLOv7 outperforms the YOLOv3 slightly in terms of the precision, recall, and F1-score with values of 0.598, 0.229 and 0.33 respectively. The precision for both of the models has slightly higher values than the recall, which indicates that they tend to produce fewer false positive detections relative to the total number of positive detections. In other words, the models exhibit a higher accuracy in correctly identifying and classifying objects, but they may have a tendency to miss some instances, resulting in a lower recall rate. Besides, YOLOv7 has a faster inference time than YOLOv3, with durations of 3.6ms and 6.1ms, respectively. It is important to note that the performance of both models drops tremendously compared to the results obtained with optimal lighting conditions. This distinct difference of performance is due to the models not being trained on various complex lighting conditions which make challenges for accurate predictions in such scenarios.

Additionally, when evaluating the performance on complex lighting videos with the application of image enhancement using CLAHE, YOLOv7 outperforms YOLOv3. YOLOv3 achieves a precision of 0.507, recall of 0.236, and F1-score of 0.322, whereas YOLOv7 demonstrates significantly improved performance with a precision of 0.735, recall of 0.306, and F1-score of 0.43. The inference time for both models remains unchanged at 4.8 ms, indicating similar overall processing speeds. Also, the precision values for both models are slightly higher than the recall values, indicating a higher accuracy in correctly identifying and classifying objects but a tendency to miss some instances, resulting in a lower recall rate. Comparing the performance on complex lighting videos with the application of image enhancement to complex lighting videos without enhancement reveals improvement in both models. This enhancement can be attributed to the use of CLAHE, which enhances object edges and contrast in an image. By applying CLAHE to complex lighting videos, the details and distinctive features of objects are enhanced, making them more distinguishable and easier to detect. Consequently, the models exhibit better precision, recall, and F1-score when confronted with complex lighting conditions and benefit from the improved visual characteristics provided by the CLAHE image enhancement technique.

The performance of YOLOv7 highlights the potential to attain a highly reliable and robust model by training it on optimal lighting images, without the requirement for complex lighting images during the training process. This is because YOLOv7 incorporates advancements in architecture design, feature extraction, and object detection algorithms, making it more effective in handling complex lighting conditions compared to YOLOv3. The feature extraction network of YOLOv7, Darknet-85, is more advanced and efficient than Darknet-53 used in YOLOv3. It incorporates additional convolutional layers, enabling YOLOv7 to extract more discriminative and robust features, enhancing its ability to handle complex lighting variations. Furthermore, YOLOv7 refines the object detection mechanism to address the limitations of YOLOv3. It introduces modifications in anchor scales, aspect ratios, and anchor assignment strategies, resulting in improved localization accuracy and precise object boundaries. These enhancements make YOLOv7 better equipped to handle objects of different sizes and aspect

ratios under varying lighting conditions. Despite these improvements, YOLOv7 retains the speed and efficiency of YOLOv3, making it suitable for real-time object detection tasks. Its optimised architecture design and efficient feature extraction and detection mechanisms contribute to faster inference times compared to YOLOv3.

This approach simplifies the overall training procedure, as developers can focus solely on gathering and utilising high-quality data with optimal lighting conditions. By eliminating the complexities associated with training on diverse lighting conditions, developers can improve the training workflow and reduce the time and resources invested in acquiring and annotating complex lighting datasets. Although the direct impact on overall processing time may vary depending on other factors, such as hardware and implementation, the elimination of training with complex lighting images indirectly contributes to improved efficiency in the processing phase. This approach allows for faster and more efficient object detection and localization tasks, as the model has been specifically optimised to excel in optimal lighting conditions. In summary, the superiority of YOLOv7, combined with the emphasis on training solely on optimal lighting images, simplifies the training process, enabling developers to save valuable time and resources. While the direct impact on processing time may vary, the elimination of complex lighting images during training indirectly enhances efficiency during inference. This approach presents a practical and efficient solution for achieving robust performance in object detection and localization tasks, promoting faster and more reliable results in real-world applications.

Fig.4: showcases the results of YOLOv7 logo detection using images captured under optimal lighting conditions.

Fig.5: showcases the results of YOLOv7 logo detection using images captured under complex lighting conditions.

Figure 4 shows vehicle logo detection with good lighting during the day, where logos are clearly visible in most of the videos. In contrast, Figure 5 demonstrates vehicle logo detection using a challenging lighting dataset, with videos taken at night under low lighting conditions causing blurriness and pixelation, and strong glares that make it difficult to see the logos.

5. Conclusion

This paper presents a comparative study involving YOLOv3 and YOLOv7, including CLAHE image enhancement for vehicle logo detection. This study aims to identify the strengths and weaknesses of these models, providing valuable insights into their robustness in handling real-world scenarios. YOLOv3 is widely used for this task, while limited research exists on vehicle logo detection using YOLOv7. Our findings reveal that YOLOv7 outperforms YOLOv3 in handling complex lighting dataset. The superior performance of YOLOv7 suggests the possibility of achieving a highly reliable and robust model by only training on optimal lighting images, eliminating the need for complex lighting images during training. By leveraging YOLOv7 and focusing on optimal lighting conditions, the overall training process becomes simpler. This approach also has the potential to save time and resources during the training phase, indirectly contributing to more efficient processing overall.

Nevertheless, it is important to acknowledge the limitations of this study. Image enhancement can be categorised as spatial domain, histogram-based technique, frequency domain and homomorphic filtering. However, in this study, we only explored the impact of histogram-based techniques, specifically histogram equalisation using CLAHE, while other image enhancement techniques could yield different results. Future research should aim to bridge this gap by evaluating and comparing the performance of various algorithms from different categories. The exploration should prioritise techniques that reduce noise, enhance contrast, and preserve crucial details to achieve optimal results in logo detection tasks under varying lighting conditions. Exploring spatial domain techniques, such as spatial filtering and edge enhancement, can provide insights into how pixel-level manipulation can

enhance image visibility and reduce noise. Frequency domain techniques, including Fourier Transform-based methods, can offer strategies for selectively enhancing or suppressing frequency components to handle specific lighting variations. Additionally, exploring homomorphic filtering can provide valuable insights into compensating for non-uniform illumination conditions and improving image quality in challenging lighting scenarios. The exploration of different approaches will contribute to a more comprehensive understanding of how to effectively handle various lighting conditions, resulting in improved visibility, enhanced contrast, reduced noise, and ultimately, more robust and reliable logo detection algorithms. By investigating and implementing different image enhancement approaches, researchers can gain a deeper understanding of their effectiveness and suitability for real-world scenarios.

Also, the study does not delve into the fine-tuning process of YOLOv7, which could potentially improve the accuracy of logo detection. Fine-tuning is important because it allows us to further refine and optimise a pre-trained model for a specific task or dataset. While pre-trained models, such as YOLOv7, provide a good starting point with general knowledge learned from large-scale datasets, fine-tuning enables us to adapt the model to our specific logo detection task and improve its performance. Moreover, fine-tuning allows us to address the limitations and biases of the pre-trained model. By fine-tuning on our specific dataset, we can refine the model's predictions and adapt it to the characteristics and challenges of our logo detection problem. It helps to overcome domain-specific challenges, such as different lighting conditions, logo variations, or specific environmental factors that may affect the performance of the pre-trained model. Further research and exploration are needed to address these limitations and expand the understanding of YOLO-based models in various scenarios.

Reference

Arinaldi, A., Pradana, J. A., & Gurusinga, A. A. (2018). Detection and classification of vehicles for traffic video analytics. *Procedia Computer Science*, *144*, 259–268. https://doi.org/10.1016/j.procs.2018.10.527

Chen, C., Lu, X., Shengqin, J., & Jiaji, S. (2017). An effective vehicle logo recognition method for road surveillance images. *2016 2nd IEEE International Conference on Computer and Communications, ICCC 2016 - Proceedings*, 728–732. https://doi.org/10.1109/CompComm.2016.7924798

Chen, Z., Guo, H., Yang, J., Jiao, H., Feng, Z., Chen, L., & Gao, T. (2022). Fast vehicle detection algorithm in traffic scene based on improved SSD. *Measurement: Journal of the International Measurement Confederation*, *201*. https://doi.org/10.1016/j.measurement.2022.111655

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). *Mask R-CNN*. http://arxiv.org/abs/1703.06870

Huan, L., Yujian, Q., & Li, W. (2017). *Vehicle Logo Retrieval Based on Hough Transform and Deep Learning*.

Huang, Y., Wu, R., Sun, Y., Wang, W., & Ding, X. (2015). Vehicle logo recognition system based on convolutional neural networks with a pretraining strategy. *IEEE Transactions on Intelligent Transportation Systems*, *16*(4), 1951–1960. https://doi.org/10.1109/TITS.2014.2387069

Jatupon Benjaparkairat, & Pakorn Watanachaturaporn. (2018). *Vehicle Logo Detection Using Sliding Windows with Sobel Edge Features and Recognition Using SIFT Features*.

Jiajun Liu, Fei Shen, Mengwan Wei, Yuzhao Zhang, Huanqiang Zeng, Jianqing Zhu, & Canhui Cai. (2019). *A Large-Scale Benchmark for Vehicle Logo Recognition*.

Jin Jie Ng, Kah Ong Michael Goh, & Connie Tee. (2023). Traffic Impact Assessment System using Yolov5 and ByteTrack. *Journal of Informatics and Web Engineering*, *2*(2), 168–188. https://doi.org/10.33093/jiwe.2023.2.2.13

Mao Yuxin, & Hao Peifeng. (2019). *A Highway Entrance Vehicle Logo Recognition System Based on Convolutional Neural Network*.

Mohammad Wahyudi Nafi'i, Eko Mulyanto Yuniarno, & Achmad Affandi. (2019). *Vehicle Brands and Types Detection Using Mask R-CNN*.

Pavan A C, Lakshmi S, & M.T. Somashekara. (2023). An Improved Method for Reconstruction and Enhancing Dark Images based on CLAHE. *International Research Journal on Advanced Science Hub*, *5*(02), 40–46. https://doi.org/10.47392/irjash.2023.011

Psyllos, A. P., Anagnostopoulos, C. N. E., & Kayafas, E. (2010). Vehicle logo recognition using a sift-based enhanced matching scheme. *IEEE Transactions on Intelligent Transportation Systems*, *11*(2), 322–328. https://doi.org/10.1109/TITS.2010.2042714

Soon, F. C., Khaw, H. Y., Chuah, J. H., & Kanesan, J. (2019). Vehicle logo recognition using whitening transformation and deep learning. *Signal, Image and Video Processing*, *13*(1), 111–119. https://doi.org/10.1007/s11760-018-1335-4

Tang, C., Wang, Y., Feng, H., Xu, Z., Li, Q., & Chen, Y. (2019). Low-light image enhancement with strong light weakening and bright halo suppressing. *IET Image Processing*, *13*(3), 537–542. https://doi.org/10.1049/iet-ipr.2018.5505

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. http://arxiv.org/abs/2207.02696

Yang, J., Wang, J., Dong, L. L., Chen, S. Y., Wu, H., & Zhong, Y. W. (2022). Optimization algorithm for low-light image enhancement based on Retinex theory. *IET Image Processing*. https://doi.org/10.1049/ipr2.12650

Yang, S., Zhang, J., Bo, C., Wang, M., & Chen, L. (2019). Fast vehicle logo detection in complex scenes. *Optics and Laser Technology*, *110*, 196–201. https://doi.org/10.1016/j.optlastec.2018.08.007

Yu, Y., Guan, H., Li, D., & Yu, C. (2021). A Cascaded Deep Convolutional Network for Vehicle Logo Recognition from Frontal and Rear Images of Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *22*(2), 758–771. https://doi.org/10.1109/TITS.2019.2956082

Yuan, Z., Zeng, J., Wei, Z., Jin, L., Zhao, S., Liu, X., Zhang, Y., & Zhou, G. (2023). CLAHE-Based Low-Light Image Enhancement for Robust Object Detection in Overhead Power Transmission System. *IEEE Transactions on Power Delivery*. https://doi.org/10.1109/TPWRD.2023.3269206

Zhang, J., Yang, S., Bo, C., & Zhang, Z. (2021). Vehicle logo detection based on deep convolutional networks. *Computers and Electrical Engineering*, *90*. https://doi.org/10.1016/j.compeleceng.2021.107004

Zhang, Q., Hu, X., Yue, Y., Gu, Y., & Sun, Y. (2022). Multi-object detection at night for traffic investigations based on improved SSD framework. *Heliyon*, e11570. https://doi.org/10.1016/j.heliyon.2022.e11570

Zhao, Q., & Guo, W. (2022). Detection of Logos of Moving Vehicles under Complex Lighting Conditions. *Applied Sciences (Switzerland)*, *12*(8). https://doi.org/10.3390/app12083835

Zhou, L., Min, W., Lin, D., Han, Q., & Liu, R. (2020). Detecting Motion Blurred Vehicle Logo in IoV Using Filter-DeblurGAN and VL-YOLO. *IEEE Transactions on Vehicular Technology*, *69*(4), 3604–3614. https://doi.org/10.1109/TVT.2020.2969427