Journal of educational equipment: Applied research, Volume 2, Issue 321 (September 2024)
ISSN 1859 - 0810

License plate recognition algorithm based on deep learning techniques in self-driving cars

Pham Thi Loan*

**Faculty of Information Technology - East Asia University of Technology*
Received: 4/9/2024; Accepted: 11/9/2024; Published:20/9/2024

Abstract: *When subject vehicle look at the images or videos, it can easily identify and locate the objects of its interest within moments, thereby helping the subject vehicle make decisions and travel behavior on the road. We just focus on the YOLOv5 algorithm, which is a single deep Convolution Neural Networks for detecting, identifying, classifying the object. YOLO is made from combination between convolutional layers and connected layers. YOLO can split the input image into a set of grid cells, unlike image classification or face detection, grid cell in YOLO algorithm will have an associated vector in the output that tells us if an object exists in that grid cell, the class of that object, the predicted bounding box for that object. Video test results show the effectiveness of using the proposed algorithm.*

Keywords: *Autonomous vehicles, object detection, YOLOv5, deep learning, convolution neural network.*

1. Introduction

Currently, traffic accidents are a significant concern in many countries around the globe, leading to substantial property damage and posing safety risks to pedestrians who may also fall victim to these incidents. In contrast, the ability to recognize objects, such as humans and traffic signals, has proven to be an effective solution. As a result, we propose a strategy to prevent and address incidents promptly. One of the most efficient methods for identifying individuals and objects is through video surveillance and tracking systems.

With the swift progression of artificial intelligence technology, deep learning has evolved significantly. It can learn and retain specific features through neural networks that are tailored to their unique characteristics, such as video, audio, and other types of data [1]. Consequently, deep learning is effectively meeting the objectives of target detection within the realm of artificial intelligence [2].

Typically, these systems are equipped with cameras or video recording devices mounted at key surveillance points within the vehicle, complemented by specialized sub-modules and software designed to capture images, process them into sequences, detect motion, and accurately track subjects. Furthermore, applications are utilized for data storage and alerts regarding potential incidents. Recognizing moving objects within the video sequence is a fundamental and essential task in video surveillance, traffic monitoring and analysis, as well as human detection and tracking.

To enhance the system's accuracy and response speed, deep learning techniques and artificial intelligence technologies are applied. The rise of convolutional neural networks (CNNs) has facilitated the development of object detection methods that leverage these networks, resulting in substantial improvements over traditional approaches. For instance, R-CNN [3], a widely recognized method, utilizes an algorithm known as selective search along with CNN features to achieve impressive results. Additionally, adaptations such as Fast R-CNN [4] and Faster R-CNN [5] have demonstrated exceptional precision, establishing a benchmark in object detection. However, to the authors' knowledge, previous research has yet to attain real-time performance.

This paper presents an approach that incorporates the YOLOv5 model along with a sample dataset comprising Vietnamese traffic images, which we have collected and classified to align with our training algorithm

2. Problem for formulation

YOLO (You Only Look Once) is a Convolutional Neural Network (CNN) model designed for object detection, identification, and classification. It integrates convolutional layers with fully connected layers, where the convolutional layers extract features from images, and the fully connected layers

predict both the probability of the objects and their locations.

A crucial element in these algorithms is the use of metrics, which are typically evaluated against a ground truth. The ground truth refers to the reference information that defines the locations and characteristics of objects within the given dataset. In the context of an object detection system, the ground truth consists of images annotated with bounding boxes that enclose each object.

Assuming that the original image and the ground truth annotations correspond to the image shown above, the training and testing data are processed in a similar manner. The model generates numerous predictions; however, many of these predictions may have low confidence levels. Therefore, only those predictions that exceed a specified confidence threshold are considered. The original image, processed through the model and the object detection algorithm, yields the locations of the objects based on this confidence threshold. The image above illustrates the model's predictions alongside a comparison of these predictions with the ground truth for the objects.

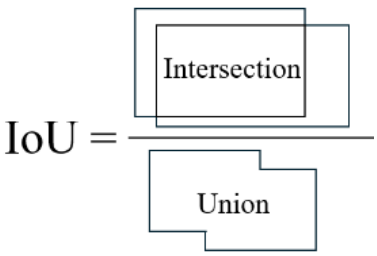

Fig 2.1. How to calculate IoU

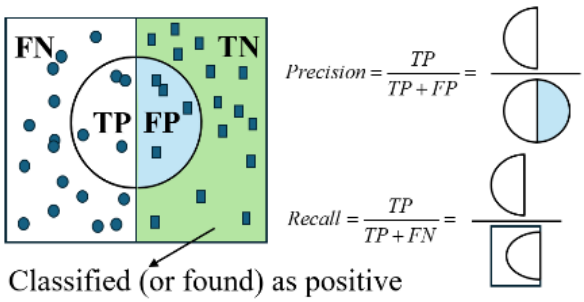

Fig 2.2. How to determine Precision and Recall

To evaluate the accuracy of the bounding boxes, we use Intersection over Union (IoU). As illustrated in Fig 2.1, IoU can be applied whenever there are both ground-truth and predicted bounding boxes. It is a ratio that measures the extent of overlap between the two boxes to determine if they intersect. This ratio is calculated based on the area of overlap between the two boxes divided by the total area covered by both boxes, including areas of overlap and non-overlap.

In cases where the dataset for a classification problem consists of classes that are significantly different from one another (unbalanced), precision and recall are commonly employed for evaluation. In Fig 2.2, precision is defined as the ratio of correctly predicted positive scores by the model to the total number of positive predictions made by the model. A higher precision indicates that a greater proportion of the predicted positive scores are accurate. If precision equals 1, it means that all points predicted as positive by the model are indeed positive; conversely, it may also indicate that none of the negative points were correctly identified, leading to false positives.

$$Precision = \frac{TP}{(TP + FP)} \quad (1)$$

Recall is defined as the ratio of correctly predicted positive points by the model to the total number of actual positive points (or the total number of points originally marked as positive). A higher recall indicates that fewer positive points are missed. When recall equals 1, it means that the model has successfully identified all points labeled as positive.

$$Recall = \frac{TP}{(TP + FN)} \quad (2)$$

3. Model of moving vehicle detection and identification system

The dataset shown in Fig 2.3, used for training, consists of footage captured from a car-mounted dash cam and a stationary camera. In preparing the training dataset for the model, videos are converted into images with a resolution of 1280x720.

The proposed method involves several steps: First, the input video data is segmented into frames and resized to 1280x720, which is the optimal resolution for achieving a balance between speed and sufficient quality for object identification. The extracted frames are then compared against the pre-trained model and processed through the YOLOv5 object recognition algorithm. The output includes the coordinates of detected objects, with each object being circled and labeled appropriately. The system generates a real-time video display of the results, which is saved in .mp4 format.

Fig 2.3. The dashed yellow rectangle represents a manually annotated FV, and the red solid rectangle is the weakly labeled FV based on the LP (in blue)

4. Results and discussion

4.1. Object detection from fixed camera

The experimental results depicted in Fig 2.4 indicate that in a busy intersection near a bridge, the recognition system successfully identifies all objects. The accuracy rate achieved is 71.4%, with an average frame processing speed of 16 FPS at a resolution of 1280x720. The system exhibits stability, with no instances of frame stacking. Additionally, it displays bounding boxes and associated probabilities for each detected object.

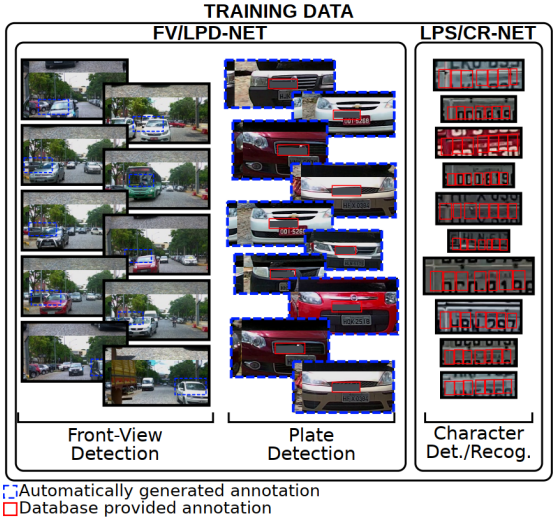

Fig 2.4. The result of object detection by using Yolov5 model

4.2. Object detection from the dashboard camera mounted on a car

When the vehicle navigates through a complex environment filled with various objects that introduce noise, such as trees, buildings, and unfamiliar items along the road, the system achieves an accuracy of 81.8% and maintains an average frame processing speed of 17 FPS at a resolution of 1280x720.

5. Conclusion

YOLOv5's rapid and accurate object detection capabilities have significantly aided automotive manufacturers and pioneering tech companies in the advancement of autonomous vehicles. By integrating YOLOv5 technology, these vehicles not only offer greater convenience but also contribute to a safer traffic environment for all. The application of the YOLOv5 model in autonomous vehicles reduces the likelihood of accidents caused by human error, enhancing efficiency and ensuring the safety of both drivers and pedestrians.

With its exceptional benefits, it is clear that the future of autonomous vehicles employing the YOLOv5 model looks very promising. Furthermore, the ongoing research and development of future iterations of YOLOv5 will be instrumental in elevating autonomous vehicle technology to new heights, presenting boundless opportunities for the automotive industry in the future.

References

[1]. Zhang S., Tuo H., Zhong H., Zhong L. J.. (2021), *Aerial image detection and recognition system based on deep neural network. Aerospace Systems*, 2021, 4.2: 101–108.

[2]. Dawson M., Perez A., Sylvestre S.. (2020), *Artificial Neural Networks Solve Musical Problems With Fourier Phase Spaces. Scientific Reports*, 2020, 10.1: 2–4.

[3]. Girshick R., Donahue J., Darrell T. and Malik J.. (2014), *Rich feature hierarchies for accurate object detection and semantic segmentation.* Proceedings of the IEEE conference on computer vision and pattern recognition, 2014: 580-587.

[4]. Girshick Ross. Fast r-cnn (2015), *Proceedings of the IEEE international conference on computer vision*, 2015: 1440-1448.

[5]. Shaoqing Ren, Kaiming He, Ross Girshick and Jian Sun (2015), *Faster R-CNN: Towards Real Time Object Detection with Region Proposal Networks*. Advances in neural information processing systems, 2015, 28.