Using Convolutional Neural Networks for Parking Sign Detection

Parnia Haji Faraji, Hamid Reza Tohidypour, Yixiao Wang, Panos Nasiopoulos
Electrical and Computer Engineering
University of British Columbia
Vancouver, BC, Canada
e-mail: {parnia, htohidyp, yixiaow, panosn}@ece.ubc.ca

***Abstract*— Automatic detection and classification of parking signs play an important role in autonomous and human-driven cars as it may lead to significant traffic reduction. Existing approaches mostly focus on traffic sign detection. Although there are a few studies in recent years that focus on parking sign detection, this field of study faces a lot of challenges such as the diversity of parking signs in different countries, the fact that the size of parking signs is usually smaller than that of normal traffic signs and the difficulty of understanding their meaning, a challenge that extends even to human drivers. This paper proposes a novel method for detecting and classifying parking signs using visual information. This study is conducted on a custom dataset of nearly 16000 images of parking signs in Vancouver, Canada. We base our approach on the YOLOv7X network, which is a powerful object detection algorithm, and obtained a mean Average Precision (mAP) of 95% on the test set, a notable result compared to the existing state-of-the-art object detection algorithm.**

Keywords- Autonomous Driving; Parking Sign Detection; Object Detection.

I. INTRODUCTION

Nowadays, artificial intelligence and machine learning technologies are used in day-to-day activities, having a significant impact on our everyday lives. Machine learning is a key technology that enables autonomous driving. Object detection and classification play a vital role in autonomous driving, as they are necessary for detecting pedestrians, cars, traffic lights, and traffic signs on the roads. Automated parking sign detection is a key factor of smart city infrastructure, as it has the potential to greatly reduce traffic congestion by notifying human drivers and autonomous vehicles of available street parking spaces ahead. Researchers, governments, and industry have all taken an interest in this topic because of its significant impact on the environment and productivity.

Parking sign detection and classification may be considered similar to traffic sign detection; however, the parking signs are usually smaller and more diverse. While the traffic sign detection task has drawn a lot of attention [1]-[3], very few researches have been conducted on parking sign detection and classification.

In their study [4], Mirsharif et al. proposed a supervised computer vision method to automatically detect and locate parking signs in San Francisco. They extract the potential regions of the image that may contain the parking sign by using a sliding window and then obtain a Histogram of Oriented Gradients (HOG) for each candidate. Based on the resulting features, the Support Vector Machine (SVM) learning algorithm classifies parking sign types using a linear function. After combining potential detections from multiple viewpoints, a map of the street is created that shows the location of the signs. However, this work is limited to only detecting parking signs, falling short from identifying their content. A more recent approach is based on a YOLOv5 network to generate a real-time parking sign detection model that detects parking signs with a mean Average Precision (mAP@.5) of 96.8% [5]. However, this method is also limited to detecting the presence of a parking sign. Chau et al. [6] proposed a different approach that is also based on YOLOv5 to first detect the parking signs and then classify them by using symbol detection. This method achieved an accuracy of 96.8% for parking sign detection and 98.3% for symbol detection. Their two-stage methodology is able to both detect the signs and classify the symbols inside the signs. However, at the end of their method, there is another step needed for combining the symbols and providing a unique meaning for the detected signs and their proposed method lacks this step.

In our previous work presented in [7], we trained a YOLOv4 model on our British Columbia (Canada) parking sign dataset to detect and classify parking signs. For input frame size of 416x416, this method achieved a mean Average Precision (mAP) of 93.01% in detecting and classifying parking signs, while the accuracy (mAP) increased to 98.56% for input frames of 1080X1080 size.

In this paper, we introduce a larger and more comprehensive dataset for parking signs used in the Province of British Columbia (BC), Canada. Using this dataset, we train a YOLOv7 network to detect and classify the BC parking signs into three main categories: parking is allowed, no stopping, and no parking. The reason for this classification is that our scheme is part of a street parking availability pipeline, where we first detect whether parking is allowed or not in a specific area and then share this information with another network that detects street parking availability and combines the information to make the final accurate decision that is shared with human-driven and autonomous vehicles. We compare the performance of our proposed model with that of the state-of-the-art approach presented in [7]. We compare the accuracy and speed performance of the YOLOv7X model with the retrained (on the new dataset) state-of-the-art YOLOv4 model presented in [7]. Evaluations showed that YOLOv7X outperformed YOLOv4, with YOLOv7X achieving a mAP of 95% at a threshold of 50% overlap and

YOLOv4 reaching a mAP of 86% at the same threshold. The YOLOv7 model also performed better on generalization and inference time compared to YOLOv4.

The rest of the paper is organized as follows. In Section II, we present our method and dataset and briefly explain the networks we have trained. Section III presents and analyzes the visual and numerical results, and Section IV concludes the paper.

II. Our Proposed Method

A. Our New Dataset

In this paper, our first task was to extend the dataset with additional video sequences from the city of Vancouver, BC. To this end, our team captured additional video sequences under different weather and light conditions around the city of Vancouver. The resulting comprehensive dataset consists of 15838 unique frames that were extracted from all the video streams. We used the Computer Vision Annotation Tool (CVAT) to label these extracted frames. We grouped the parking signs into three main categories: no stopping, no parking, and parking allowed. Figure 1 shows examples of different signs present in frames of our dataset. Each of the labeled frames contains one or more parking signs. The labeling process draws a rectangular bounding box around each of the signs in the frame (see Figure 2). In total, our dataset consists of 7824 no stopping signs, 4091 no parking signs, and 7053 parking-allowed signs.

We exported the labels in the format supported by YOLO, meaning that for each frame there is a text file containing the class number (for example 0 corresponds to the stopping sign, 1 to the no parking sign, and 2 to the parking-allowed sign) and the coordinates of the bounding boxes. We used 12180 frames for training, 3090 for validation, and 567 for testing.

B. Our Network

We base our approach on the YOLOv7X network architecture [8]. YOLO, which stands for You Only Look Once, is a well-known family of real-time object detection networks, with the original YOLO object detector first introduced in 2016 [8]. This network architecture proved to be much faster than its peer object detectors and established itself as the choice for real-time object detection applications. Since then, different versions of YOLO have been introduced, with each one of them offering a significant increase in performance and efficiency. YOLOv7, being the latest official YOLO version, can predict bounding boxes more accurately than its peers at much faster inference speeds [9]. YOLOv7 uses an Extended Efficient Layer Aggregation network E-ELAN as the final layer aggregation, an extended version of the ELAN [10] computational block. The multiple paths supported by E-ELAN offer a shorter distance for the gradient to back-propagate through the layers, allowing the network to converge faster. Compared to YOLOv7, which increases the

(a) (b) (c)

Figure 1. Examples of (a) no parking (b) no stopping and (c) parking allowed signs.

(a)

(b)

Figure 2. Some examples of the images in our dataset.

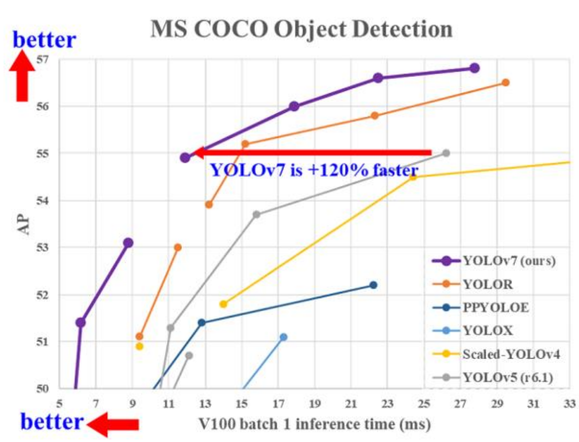

Figure 3. Comparison of YOLOv7 with previous object detection networks [9].

depth or width to improve performance, YOLOv7x introduces a compound scaling on the neck, which increases the depth and width of the entire model simultaneously, leading to improved accuracy. Figure 3 shows the improvements of YOLOv7 over the prior YOLO versions in terms of Average Precision (AP) and computational time over the Common Objects in Context (COCO) dataset [11].

III. OUR APPROACH

We compared the street parking sign detection and classification performance of YOLOv7X trained on our new dataset against that of the state-of-the-art YOLOv4 also retrained on the same dataset. Both networks were trained using Canada's Digital Research Alliance’s computing clusters [12]. We used a Tesla v100 Graphics Processing Unit (GPU) for the training and testing phases. The mAP metric is used to evaluate the performance and accuracy of the models. A model's average precision is determined by measuring its precision and recall at various Intersection-Over-Union (IoU) thresholds. Recall that IoU is equal to the ratio of the area of overlap between the predicted bounding box and the ground-truth bounding box to the area of the union between the two boxes. As YOLOv7X supports mosaic augmentation, which combines several frames into a new one, we had to disable this option for our implementation. The reason for this decision comes from the fact that mosaiced frames would introduce signs on the top, left, and bottom of the scene, thus misleading the network and reducing the accuracy.

YOLOv7X supports input images of 640x640 and 1280X1280 pixels. Since parking signs are rather small objects, the detection and classification tasks become more difficult for low-resolution images, so we decided to use 1280X1280 input images. Note that we also use the same resolution for the YOLOv4 network for a fair comparison.

Table I below shows the results of the validation set for the two models at 0.5 IoU threshold. We observe that the YOLOv4 model scored a mAP of 91% while YOLOv7X reached a mAP of 97% for the same dataset, a significant increase in performance.

The models were then tested on 567 unseen frames from our test dataset. The total number of bounding boxes for signs used as ground truth was 720 (several frames have more than one sign). Table II shows the results of precision, recall, and mAP on the test dataset at a threshold of 0.5 for YOLOv4 and YOLOv7X. As it can be seen, YOLOv7X achieves a mAP of 95% while YOLOv4’s mAP is only 86%. In terms of precision and recall, YOLOv7X and YOLOv4 have the same precision (93%), while YOLOv7X outperforms YOLOv4, yielding a recall of 92%, which is almost 10% higher than that of YOLOv4. From the above, we can conclude that our YOLOv7X model significantly outperforms the state-of-the-art model in detecting and classifying street parking signs.

Regarding the complexity and inference time reduction reported in previous studies, YOLOv7 needs 75% fewer parameters while it requires 36% less computation in comparison with YOLOv4. In our experiments, using the same hardware setup, the inference time for YOLOv4 was approximately 530 milliseconds, while YOLOv7X completed

TABLE I: VALIDATION RESULTS

Model	mAP @ 0.5
State-of-the-art (YOLOv4)	0.91
Ours (YOLOv7X)	**0.97**

TABLE II: TESTING RESULTS YOLOV4 AND YOLOV7X FOR PRECISION, RECALL, AND MAP ON THE TEST DATASET AT A THRESHOLD OF 0.5

Model	Precision	Recall	mAP @ 0.5
State-of-the-art (YOLOv4)	0.93	0.81	0.86
Ours (YOLOv7X)	**0.93**	**0.92**	**0.95**

TABLE III: TESTING RESULTS OF YOLOV4 AND YOLOV7X FOR THE THREE DIFFERENT CATEGORIES AT A THRESHOLD OF 0.5

Model	No stopping	No parking	Parking allowed
State-of-the-art (YOLOv4)	0.85	0.90	0.85
Ours (YOLOv7X)	**0.90**	**0.98**	**0.91**

(a)

(b)

Figure 4. Visual results on a test image for (a) YOLOv7X and (b) YOLOv4.

 ISBN: 978-1-68558-071-1

the same task in 25 milliseconds, making the latter a better candidate for real-time implementation for sign detection.

Table III shows the mAP at 0.5 for the three different classes for both YOLO models. It is obvious that for all three categories of parking signs, YOLOv7X performs better than YOLOv4. In both models, the highest mAP belongs to the No Parking class. YOLOv7X had 98% accuracy and outperformed YOLOv4 by 8%. YOLOv7X achieved a mAPs of 90% for the No stopping class and 91% for the Parking allowed class, again exceeding YOLOv4's performance. Overall, the results suggest that YOLOv7X significantly exceeds YOLOv4's performance in the detection and classification task.

Figure 4 shows the same frame of a scene processed by YOLOv7 (Figure 4 a) and YOLOv4 (Figre 4 b) for visualization purposes. We observe that YOLOv4 did not detect the signs, while YOLOv7X detects and classifies both signs with high confidence (0.84 and 0.94). Note that here we show only one frame, but YOLOv4 failed to detect the signs for the entire duration of this scene.

Figure 5 shows a very rare image sample in our dataset, where there are 4 signs in the frame, whose visual quality is rather poor. It may be seen that despite that, YOLOv7X detected and classified all the signs correctly. On the other hand, we observe in Figure 6 that YOLOv4 only detected one of the four signs in the frame and failed to detect the rest of the signs. This proves that YOLOvX is more capable of detecting parking signs in images with poor visual quality.

IV. CONCLUSION

We proposed an innovative parking sign detection and classification scheme that is based on the YOLOv7X network architecture. The network was trained and tested on a custom dataset of almost 16000 frames that were extracted from videos captured by our team with a car camera in the streets of Vancouver, BC. We used a three-class strategy to classify our dataset and since the objects in our dataset were small and hard to detect, we chose the input image size to be 1280 by 1280 pixels.

Figure 5. An example of YOLOv7X's ability to accurately detect several signs even at very poor visual quality.

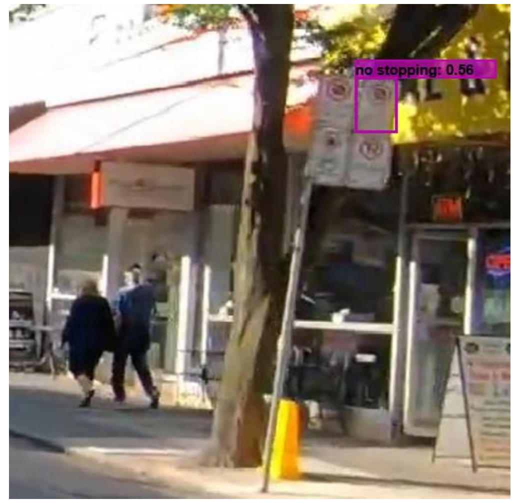

Figure 6. The exact same frame from figure 5 processed by YOLOv4 model.

Performance evaluations have shown that the YOLOv7x model outperforms YOLOv4 in terms of both accuracy and detection and is much faster than YOLOv4. Additionally, YOLOv7X is capable of detecting and classifying small and low-quality signs (i.e., weather deteriorated, low light conditions) present in video frames.

In the future, we plan to expand our current dataset by adding more videos from different parts of the city and under different weather and light conditions. We will also extend our initial three-class strategy to more classes that will include time information and different municipality related restrictions and then develop a new method that can detect and classify parking signs according to all these classes.

REFERENCES

[1] A. De La Escalera, M. Luis, E. S. Miguel Angel, and M. A. José, "Road traffic sign detection and classification." *IEEE Transactions on industrial electronics* 44, no. 6, 1997, pp. 848-859.

[2] Z. Zhu, L. Dun, Z. Songhai, H. Xiaolei, L. Baoli, and H. Shimin "Traffic-sign detection and classification in the wild." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2110-2118.

[3] Y. Yang, L. Hengliang, X. Huarong, and W. Fuchao, "Towards real-time traffic sign detection and classification." *IEEE Transactions on Intelligent transportation systems* 17, no. 7, 2015, pp. 2022-2031.

[4] Q. Mirsharif, D. Théophile, M. Sqalli, and V. Balali, "Automated recognition and localization of parking signs using street-level imagery." In *Computing in Civil Engineering 2017*, 2017, pp: 307-315.

[5] J. Yin, "Real-Time Parking Sign Detection for Smart Street Parking" A Ph.D. Thesis in the University of Washington, 2022. Available from: *https://digital.lib.washington.edu/researchworks/handle/1773/48418*

[6] H. Chau, J Yin, L. J. Li, J. Hu, and W. Cheng, "Real-Time Street Parking Sign Detection and Recognition.", “unpublished, presented at AI4AD Workshop in 2022”. Available from: https://learn-to-race.org/workshop-ai4ad-ijcai2022/assets/papers/paper_13.pdf

[7] P. Haji Faraji et al., "Deep Learning based Street Parking Sign Detection and Classification for Smart Cities." In *Proceedings of the Conference on Information Technology for Social Good*, pp: 254-258. 2021.

[8] C. Y. Wang and A. Bochkovskiy, Implementation of YOLOv7. GitHub repository (2022). Available from https://github.com/WongKinYiu/yolov7

[9] C. Y. Wang, A. Bochkovskiy, and H. Y. Mark Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors.", 2022, arXiv *preprint arXiv:2207.02696*.

[10] C. Y. Wang, H. Y. Mark Liao, and I. H. Yeh, "Designing Network Design Strategies Through Gradient Path Analysis.", 2022, *arXiv preprint arXiv:2211.04800*.

[11] T.Y Lin et al., Microsoft COCO: Common Objects in Context. CoRR [Internet], 2014, abs/1405.0312. Available from: http://arxiv.org/abs/1405.0312

[12] Digital Research Alliance of Canada’s state-of-the-art advanced research computing network. Available from: https://alliancecan.ca/en.