Performance Analysis of Semantic Segmentation with YOLOv8: A Robust Approach for Autonomous Vehicle

Fussy Mentari Dirgantara and Soo Young Shin*
Department of IT Convergence Engineering,
Kumoh National Institute of Technology (KIT), Gumi, South Korea
Email: {dirgantara, wdragon*} @kumoh.ac.kr

Abstract

This research analyzes the application of YOLOv8, a state-of-the-art object detection algorithm, in the domain of semantic segmentation within image processing. Focusing on refining and optimizing this critical aspect of computer vision, our study explores YOLOv8's architecture and its adaptation for semantic segmentation for autonomous vehicles. We analyze the algorithm's accuracy, speed, and recall performance through experimentation and evaluation across diverse datasets and pedestrian scenes. The findings contribute valuable insights into YOLOv8 as a tool for semantic segmentation, offering a basis for advancements in real-world applications such as autonomous vehicles, medical imaging, and robotics. This research addresses the increasing need for accurate object segmentation, offering a relevant analysis of YOLOv8's contribution to improving semantic segmentation accuracy and efficiency.

Ⅰ. Introduction

Given an image, a human's vision immediately identifies objects very well. However, when it comes to computer vision, one must bin the image into important segments, each representing an object. Image segmentation is the process of partitioning an image into multiple segments, and this partitioning is widely used in computer vision applications. In the past decade, rapid progress in deep learning has been very successful in numerous computer vision applications. For example, You Only Look Once (YOLO) is one of the most popular and advanced object detection frameworks that has evolved due to speed and accuracy of object detection than other models. Nevertheless, the high-level design of YOLO can be detrimental for detecting tiny objects and for duplicated instances, because traditional YOLO merely offers the computed boundary box, which excludes important information for accurate labeling when the instances are close to one other [1].

Semantic segmentation, however, resolves the issue since it is not only for object detection, but it also refers to the act of precisely describing each pixel in the image. This information is vital in tasks such as object recognition in various scenes, autonomous vehicles, and much more identification in medical images [2, 3, 4].

This paper delves into the possibility of YOLOv8, focusing mainly on how YOLOv8 applies to segmentation. YOLOv8 realizes better results from all perspectives, accuracy, better performance at inference, and better scalability. We exploit it further on how semantics of it can help conduct object segmentation on images after capturing them with the computer.

Our research is based on a YOLO-based architecture proven by high accuracy, weightlessness, and suitability for reading at the edge, making it more suitable to industrial requirements. YOLO-based architecture is well-established and has been used in industrial applications. In this statement, the authors make the following rational contribution to this topic: the authors evaluate the effectiveness of the YOLOv8 model with the most well-known benchmark datasets of object segmentation by evaluating its complexity vis-à-vis existing state-of-the-art semantic segmentation techniques and describe the effect when YOLOv8 acts as an object segmentation [5]. This elaborate research allows the author to strive to prove and demonstrate the effectiveness of YOLOv8 as a concept for object segmentation in images.

Ⅱ. Method

1. Image Segmentation

In YOLOv8, the regular backbone feature extraction network was CSPDarknet. Therefore, this network should be tuned for segmentation tasks, e.g. by adding decoding modules such as upsampling layers, which would reconstruct the original spatial information. The head network of YOLOv8 is a network specifically designed to be able to make detections. However, in our case, this network won't be able to directly predict the object by class and obtaining tis bounding box. It should be tuned to predict the segmentation masks. The final layers of this network should be replaced by other modules which should include convolutional layers that output a probability map for every pixel to belong to a certain object class.

Training strategy that is used for this research is to measure the difference between the predicated segmentation mask and the ground truth mask using loss function. Techniques like random cropping, flipping, and color jittering also applied to training images to improve model generalization and robustness to variation in real-world data.

Image segmentation using YOLOv8 has benefit inherent speed advantage translates to faster segmentation compared to traditional segmentation models. The YOLO method has advantages in terms of swift processing compared to traditional segmentation models, precise accuracy, and robust generalization capabilities.

2. Training

Given the purpose of deploying this model on edge devices, specifically Intelligent Unmanned Aerial Vehicles, we have chosen the YOLOv8-seg (image segmentation) package with the simplest architecture. The pre-trained network consists of 80 classes and has a 50:50 data split for training and validation. The YOLOv8n-seg model consists of 261 layers, with a total of 3,409,968 parameters and 3,409,952 gradients. Investigation for a method to deploy this YOLO model for the selected environment was conducted. Many attempts were made to integrate trained models through different DNN frameworks such as Onnx and Pytorch. Each model presented here represents a pre-trained model using YOLOv8n-seg architecture that is fine-tuned on our data for 30 epochs.

3. Result Implementation

After performing training as described above, we observe results that the differences between predicted and true boxes loss of 0.775, the accuracy of the classes in each detection of 1.692, also the detected object loss 1.084 A visualization of the convergence of the results can be found in Figure 1.

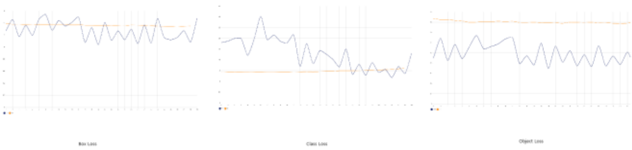

Fig 1. Training losses observed for all experiments for 30 epochs

The result of image segmentation is presented in Figure 2. For the selected model, we examine the important regions responsible for the classification of pedestrian class. These samples are selected from the map and differentiate between VGG Unet model [6] and YOLOv8n-seg model.

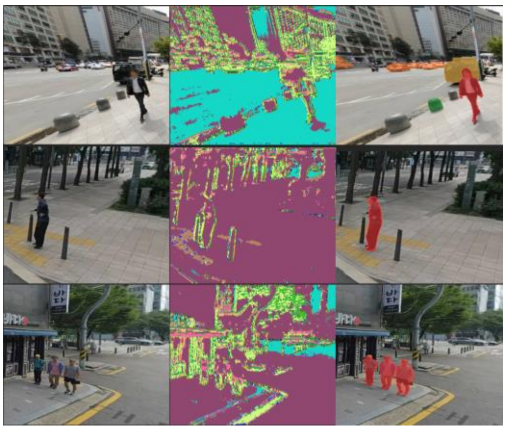

Fig 2. Qualitative results on the pedestrian image segmentation

Ⅲ. Conclusion

Despite 30 epochs of training YOLOv8n-seg for semantic segmentation, a drop in class loss indicating the model learned to classify pixels, the high box loss suggests it struggled with segmentation boundaries. Analyzing object loss alongside class loss and evaluating metrics like IoU can provide a clearer picture. Further training or hyperparameter adjustments might improve boundary accuracy.

ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-RS-2023-00259061) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation). This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2018R1A6A1A03024003)

REFERENCES

[1] H. Wang, C. Liu, Y. Cai, L. Chen and Y. Li, "YOLOv8-QSD: An Improved Small Object Detection Algorithm for Autonomous Vehicles Based on YOLOv8," in IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-16, 2024, Art no. 2513916, doi: 10.1109/TIM.2024.3379090.

[2] LI, Yan; MA, Yue; LONG, Ying. “Protocol for assessing neighborhood physical disorder using the YOLOv8 deep learning model.” in STAR protocols, 2024, 5.1: 102778.

[3] B. Zheng, C. Huang, X. Chen and Y. Luo, "Toward Head Computed Tomography Image Reconstruction Standardization With Deep-Learning-Assisted Automatic Detection," in IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-14, 2024, Art no. 2502814, doi: 10.1109/TIM.2023.3329102.

[4] KANG, Ming, et al. “ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation.” in Image and Vision Computing, 2024, 105057.

[5] Ultralytics, “YOLOv8 Docs,” retrieved from https://docs.ultralytics.com/, 2023. Accessed April 10, 2024.

[6] Gupta, D. “Image segmentation keras: Implementation of segnet, fcn, unet, pspnet and other models in keras.” in arXiv preprint arXiv:2307.13215, 2023.