

Capstone Final Project:

Detecting Vehicle Interaction for Autonomous Driving

By: Jessica Taylor

Project Members: Sadipta Das, Syed Mohammed

Advisor: Jiang Zheng | jzheng@iupui.edu

Indiana University – Purdue University Indianapolis

30 April 2023

Abstract

This project aimed to annotate motion profiles, serving as vital training data for a deep learning network dedicated to detecting vehicle interactions in the context of autonomous driving. A total of three hundred motion profiles were meticulously labeled, originating from corresponding one-minute driving videos. Vehicle detection within these videos was achieved through the implementation of the YOLO5 AI deep learning network.

The labeling process was facilitated using the Pixel Annotation Tool, an open-source software obtained from GitHub. This tool provided a robust platform for accurately annotating trajectories and interactions within the motion profiles.

Utilizing a fusion of advanced deep learning techniques and efficient annotation tools, the project successfully labeled all three hundred motion profiles. These annotated datasets now stand poised as invaluable resources, used to empower the development and training of sophisticated neural networks for the nuanced task of vehicle interaction detection in autonomous driving scenarios.

Keywords: Autonomous driving, self-driving, vehicle interaction, motion profiles, deep learning, machine learning, artificial intelligence

Introduction

In the scope of our project, we were entrusted with three hundred one-minute driving videos wherein vehicle detection was accomplished through YOLO5, a robust AI deep learning network. These videos encapsulated 14 interaction events and 4 self-actions, dynamically unfolding based on the configurations of the roads and traffic. These events were meticulously labeled on motion profiles, temporally spanning the horizon of each video at 30 frames per second.

The creation of motion profiles involved capturing YOLO5 bounding boxes, from which motion trajectories were derived, offering insights into key vehicle parameters such as position, speed, shape, size, and depth. In Figure 1.1, an illustrative snapshot of one such driving video is presented, emphasizing the delineation of the horizon in red for subsequent motion profiling. Figure 1.2 provides a visual representation of the resulting motion profile, a temporal image where 't' is a function of time. The bottom of the motion profile corresponds to the horizon at t equals zero seconds, while the top reflects the horizon at t equals sixty seconds.

Figure 1.1



Figure 1.2



Following the creation of motion profiles, the necessity arises for the generation of vehicle trajectory maps, marking a pivotal juncture in our project. This brings us to the crux of our endeavor—annotating and labeling all motion profiles. Figure 1.3 offers yet another example of a motion profile, while Figure 1.4 showcases its corresponding vehicle trajectory map. Additionally, Figure 1.5 presents a labeled motion profile, highlighting the distinctive shapes assigned to different vehicle interactions. Notably, the monochromatic nature of Figure 1.4 facilitates the annotation process, and upon completion, it will serve as a crucial component in the training phase of our project.

Figure 1.3

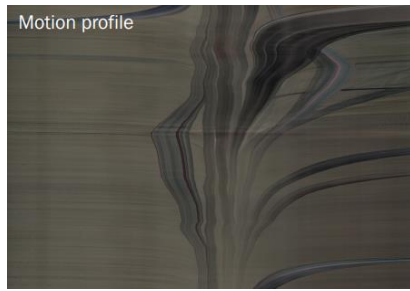


Figure 1.4

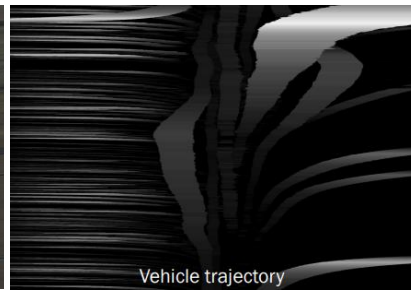
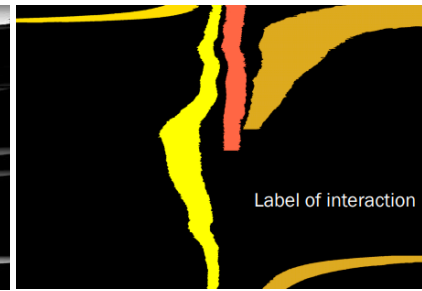


Figure 1.5



Problem Definition

The primary challenge addressed in this project centered on the acquisition of precise training data tailored for a deep learning network. The overarching objective was to equip this network with the capability to discern and accurately detect various forms of vehicle interactions. While the ultimate aspiration was the successful training of the deep learning model, the immediate focus revolved around the meticulous annotation of motion profiles, serving as the foundational input for the network.

At present, experiments leveraging the existing dataset yield an accuracy of 80% for critical events in real-time scenarios. The challenges hindering a non-deep learning approach include the intricacies of the algorithm, variable filter scales, and the inherent difficulty in fine-tuning thresholds. Even with the integration of YOLO shape detection as a precursor, the non-deep learning approach falls short, exacerbated by the substantial disparities in road configurations and vehicle characteristics.

Recognizing these limitations, our advisor advocated for exploring a deep learning paradigm to address the intricacies of interaction recognition. This approach holds promise in overcoming the challenges posed by scenes with varying speeds, diverse road structures, and a wide spectrum of vehicle sizes. The transition to a deep learning framework signifies a strategic shift aimed at enhancing the robustness and adaptability of the interaction recognition system.

Solution

To ensure the accurate labeling of each motion profile, a set of predefined rules was established. Our advisor furnished us with a comprehensive guide (as depicted in Figure 2.1), delineating distinct colors for different types of vehicle interactions. This guide not only specified the visual representation

expected in the motion profile for each interaction type but also assigned a corresponding color for trajectory labeling.

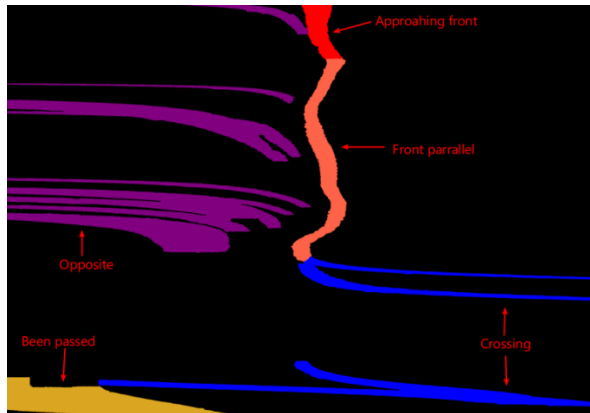
Figure 2.1

Other vehicle locations	Interaction	Motion: Visual trajectories with image velocity v	Vehicle shape	Class color
Next lane vehicle into driving lane	Cut in vehicle from side lane (close range)	Large size and large inward v at front edge	Rear view	(255, 140, 0)
	Lane changing in safe space	Inward v , small size	Rear view	(255, 165, 0)
Driving lane vehicle	Approaching of front vehicle	Expanding width of trajectory at image center	Rear view	(255, 0, 0)
	Leaving of front vehicle	Shrinking trajectory at image center	Rear view	(240, 128, 128)
	keep distance of front vehicle	Approximate parallel trajectory at image center	Rear view	(255, 99, 71)
Next lane vehicle driving	Parallel of side lane vehicle	Stable size at side, v stable	Rear/side	(255, 255, 0)
	Passing of side lane vehicle	Inward v from side	Side-rear	(255, 215, 0)
	Being passed side lane vehicle	Outward v of front edge at side	Rear-side	(218, 165, 32)
Ramp vehicle	Merging vehicle from right	Small trajectory with inward v (slow)	Side-rear	(60, 205, 128)
	Collision in merging	Enlarged trajectory/size with zero v	Side-rear	(0, 128, 0)
	Yielding vehicle at right	Small trajectory with slow outward v	Rear-Side	(60, 179, 113)
Opposite vehicle	Opposite vehicle	Fast outward	Front	(128, 0, 128)
Crossing road vehicle	Crossing vehicle	Inward v from side	Side	(0, 0, 255)
	Left turning of opposite vehicle	Rightward motion from center	Front-side	Label it as turning away (0, 255, 255)
Turning away	Turning away vehicle	Outward v of sideview of vehicle	Side face	(0, 255, 255)
Parked vehicle and BG	Parked vehicle and BG	Same as background with outward v	Side face	(0, 0, 0)
Other Ego-vehicle actions	Changing lane	Short displacement in v	Rear, side	Label it as next lane vehicle change into driving lane, (255, 165, 0)
	Turning	All scenes moving opposite direction v	Front, side, rear	Case 1: no front vehicle. Label the turning period as BG. (0, 0, 0) Case 2: front vehicle exists. Label front vehicle trajectory as turning away, next lane vehicle driving or driving lane vehicle (based on speed and position)
	Stop at signal, stop-sign, and traffic jam	All static scenes have parallel trajectories along time axis, $v=0$ absolutely.	Side, rear	front vehicle exists and stops. Label the trajectory as gray (128,128,128)

Armed with this guide, the labeling process commenced. Figure 2.2 provides a tangible example of a successfully labeled motion profile. Notably, each vehicle interaction is characterized by a unique shape,

a feature vividly exemplified by the 'Opposite' label in the given instance. In the corresponding video, multiple vehicles were observed traveling in the opposite lane, resulting in a distinct cluster of trajectories marked with the 'Opposite' label.

Figure 2.2

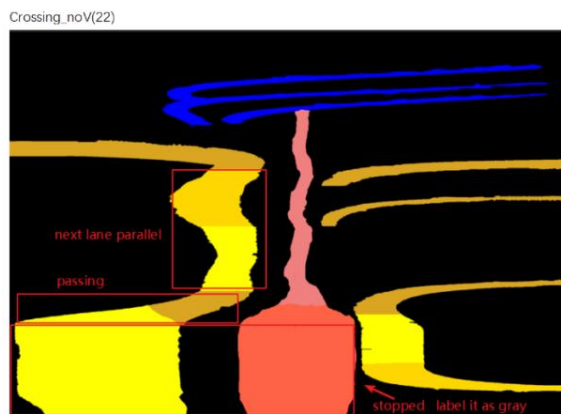


This systematic approach, guided by a predefined rule set, facilitated a consistent and accurate labeling process for the diverse range of vehicle interactions encountered across the one-minute driving videos.

Experimental Evaluation

The task of labeling each motion profile unfolded as a more intricate process than initially anticipated. We harnessed the capabilities of an open-source labeling tool, the Pixel Annotation Tool, for this purpose. Throughout this endeavor, our Teaching Assistant, Li Lin, played a pivotal role by providing detailed feedback on all motion profiles, as exemplified in Figure 3.1.

Figure 3.1



The corrective phase, in response to the provided feedback, proved to be a time-consuming aspect, partly owing to certain constraints of the Pixel Annotation Tool. The tool's limitations sometimes necessitated redoing entire profiles rather than making selective modifications.

An additional constraint surfaced as labels within the Pixel Annotation Tool could not be renamed. This constraint compelled us to consistently refer back to the initial guide to ensure accurate usage of labels. Although this restriction did not compromise the accuracy of the labeling, it introduced a layer of complexity in interpreting which color corresponded to specific types of vehicle interactions.

Despite these challenges, the Pixel Annotation Tool, coupled with constructive feedback, served as a crucial component in refining and enhancing the precision of our labeled motion profiles. The constraints encountered underscore the importance of a robust annotation tool in streamlining the labeling process for complex datasets.

Conclusion:

In culmination, our project successfully generated three hundred meticulously labeled motion profiles, forming a robust foundation for training data. While our initial aspirations extended towards delving into the realm of deep learning, unforeseen time constraints prompted a strategic pivot, with our advisor concurring that prioritizing the creation of accurate motion profiles would best serve our objectives.

The project provided invaluable insights into the intricacies of testing and curating precise datasets essential for the efficacy of deep learning networks. It underscored the substantial time and effort requisite for generating reliable training data—a critical aspect often underestimated in the development of such networks.

In essence, our project emerged as an overall success, as we navigated unforeseen challenges and remained steadfast in delivering on our primary goals. This experience not only enriched our understanding of data preparation for deep learning but also equipped us with valuable problem-solving skills, contributing to our collective growth and proficiency in tackling complex projects.

References

Z. Wang, J. Y. Zheng and Z. Gao, "Detecting Vehicle Interactions in Driving Videos via Motion Profiles," 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 2020, pp. 1-6, doi: 10.1109/ITSC45102.2020.9294617.

Breheret, A. (2017). Pixel Annotation Tool. GitHub.
<https://github.com/abreheret/PixelAnnotationTool>

Understanding Vehicle Interaction in Driving Video with Spatial-temporal Deep Learning Network. J. Y. Zheng, VISC Lab, IUPUI