10-19-2022 10:15 AM

An Investigation Into Time Gazed At Traffic Objects By Drivers

Kolby R. Sarson, *The University of Western Ontario*

Supervisor: Bauer, Michael A., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Computer Science

Abstract

Several studies have considered driver's attention for a multitude of distinct purposes, ranging from the analysis of a driver's gaze and perception, to possible use in Advanced Driving Assistance Systems (ADAS). These works typically rely on simple definitions of what it means to "see," considering a driver gazing upon an object for a single frame as being seen. In this work, we bolster this definition by introducing the concept of time. We consider a definition of "seen" which requires an object to be gazed upon for a set length of time, or frames, before it can be considered as seen by the driver. This is done by examining consecutive frames to find those where the driver's gaze remains uninterrupted within a constant bounding box of a given traffic object over a series of frames. A time-considering approach to defining traffic objects as seen or unseen provides a more thoughtful and accurate measure of driver's perception, as we avoid the naïve assumption that gazing upon an object for a single frame is enough time for a driver to process the object gazed upon, which ultimately could prove vital to a wide array of ADAS and i-ADAS systems.

Keywords: ADAS, Computer Vision, Driver Attention, Driver Gaze, Driver Perception, Gaze Estimation, i-ADAS, RoadLab, Traffic Objects

Summary for Lay Audience

This thesis introduces a novel approach for examining the notion of what a driver "sees" in the context of Advanced Driving Assistance Systems (ADAS) where determining what a driver sees is based on images (frames) from cameras and computation of a driver's gaze. Previous research often assumes that what a driver "sees" is based on determining a driver's gaze on a single frame. "Seeing", however, is complex and is based on function of the human eye and human cognition and a computational approach that a single frame is sufficient is likely too limited. This research considers a driver's gaze and objects across a number of frames. It looks at the impacts that adjusting the number of frames considered under a driver's gaze can have on the number of objects seen. This is investigated through the utilization of various frame length thresholds as the basis for our definition of seen, allowing us to compare these thresholds. The work aims to provide conclusions on the impact of these varied frame length thresholds and whether future work on determining what a driver "sees" as part of ADAS would benefit from a more thoughtful definition of "seen".

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

1.25 million. This is the estimated number of fatalities caused by automotive accidents worldwide each year, with these accidents being the leading cause of death for young people, ranging in age from 15-29 [29]. These fatal accidents also account for approximately 3% of government GDP annually [29]. With such significant numbers, the advancement of intelligent driving systems, including both autonomous vehicles and Advanced Driver Assistance Systems (ADAS), have been explored. ADAS in particular have been derived with the purpose of reducing human driver error.

ADAS have been used for a variety of tasks to assist drivers in driving more safely. Some examples include adaptive cruise control, forward collision warnings, lane departure warnings, and traction control. Adaptive cruise control sets the speed of the vehicle to match the speed limits and surrounding traffic, maintaining safe follow distances to allow for correction in the event of a dangerous situation. Forward collision and lane departure warnings alert the driver of their respective situations which could prove disastrous. Traction control detects when sliding is occurring and will adjust the braking and turning of the wheels to account for the low surface traction in an attempt to counteract the sliding motion.

1.2 Problem

As discussed, driving has been proven to be a dangerous act. This is why in many societies we require proper training and licensing in order to partake in the activity of driving. There are a wide variety of automotive accident causes, including faulty vehicle parts causing failures, natural causes such as animals running into the road, or even human error. Human error itself could have an array of causes, and according to the U.S. Department of Transportation, National Highway Traffic Safety Association (NHTSA), lack of attention is cited as the number one cause of accidents, specifying drowsiness and distractions as sources of attention breakers [27]. Another notable source of attention breaking is drivers choosing to pay less attention due to an excessive trust in driving assistance and car safety features, Suzuki et al. [25].

When a driver is on the road it is a common expectation that they will take in and process their surroundings. But the unfortunate truth is that humans are not capable of seeing everything, all of the time, especially when drivers lose full attention. To gain a better understanding of what drivers are aware of during actual driving, we will analyze what drivers see and do not see while driving. In this context, the idea of see is limited, so we need to define what it means to see. To know if a driver has ”seen” an object we will consider two factors, whether the driver’s gaze has fallen onto the object, as well as for how long their gaze remained on said object. As for the object in question, we will utilize a predetermined set of common traffic objects, including various vehicles, traffic signs, and pedestrians.

1.3 Contribution

In this thesis, a novel idea of seeing vs gazing is considered. In the past, few works seem to touch on the idea of driver’s seeing objects beyond the high-level scope, typically considering the simplest case of; if the point of gaze lies on an object in a frame, that frame is ”seen”. We logically know that such a simple definition of seen may be sufficient for some works, but scrutiny of the definition of ”seen” could prove beneficial. Where some research may consider and object ”seen” or ”unseen” based on simplistic definitions, this contribution looks to begin the exploration of differing ideas of what it means to see, within the scope of drivers, without going as far as to consider human cognition. We will utilize some of the architecture from Shirpour et al. [22] to perform object detection on a dataset collected from drivers during actual driving activities. We then investigate the driver’s observation of objects during the driving sequence to identify those objects which are both ”seen” and ”unseen” given our self-defined and varied definitions of ”seen”.

1.4 Thesis Organization

This thesis is organized as such:

In Chapter 2 we look at some prior work in the fields of advance driver assistance systems, traffic object detection, driver’s gaze, and other related topics. In Chapter 3 we describe data collection and present the overall architecture for this work. In Chapter 4, we then present the analysis of our data and our results. In Chapter 5 we discuss our work and draw any relevant conclusions, as well as discussing future works through identifying areas of interest both by expanding the current work as well as broadening the scope of the work done.

Chapter 2

Related Work

In this Chapter an investigation into previously completed works will be conducted. The areas of interest include Advanced Driver Assistance Systems (ADAS), traffic object detection, driver's gaze, and seen versus unseen objects.

2.1 Advanced Driver Assistance Systems

Advanced Driver Assistance Systems, or ADAS, are systems implemented to "enhance, among other things, active and integrated safety", Bengler et al. [6]. A more recent movement in the field of ADAS is intelligent Advanced Driver Assistance Systems, or i-ADAS. i-ADAS are intended to reduce the feedback to the driver while behind the wheel by adding a sense of intelligence to the ADAS. This can be done through a variety of ways, but the main goal is to have the ADAS system only relay the information that is relevant to the driver or vehicle, at the appropriate time. An example of this could be only turning on surveillance measures, such as a dashcam, if the i-ADAS detects that damage may be done to the vehicle, as discussed by Lee et al. [16]. This research proposes an improved and more efficient ADAS processor, doing so by incorporating intelligence. A microchip with hardware specifications is presented in a theoretical blueprint alongside analyses of the performance. This microchip exemplifies the benefits of moving ADAS towards i-ADAS. Another example is provided by Shadeed et al. [21], where an i-ADAS is proposed involving front-lighting systems which would provide better visibility while also utilizing a glare-free highbeam lighting system. This system would provide the driver with useful information without adding any dashboard display by providing the information naturally in the environment.

In recent years there have been even more strides to produce higher quality, more efficient, and new approaches to implementing i-ADAS. After the initial foundations of ADAS were realized, research in this area shifted to the improvement and advancement of ADAS as a whole. We now know that driver assistance is possible and research has moved forward in a variety of areas, ranging from improvement of current systems, to ideas for whole new ADAS which could benefit drivers. Nidamanuri et al. [18] provides us with an exceptional review of these research areas including automotive electronics, vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication, RADAR, LIDAR, computer vision, and machine learning. This review is an excellent resource summarizing the ADAS works done in recent years for the underlying systems as well. Additionally, we are presented with a variety of ADAS which have been researched along with the areas where more research is required for each of these ADAS presented, such as collision avoidance systems, traffic sign recognition, lane change assistance, and several others.

As for current research, there are some areas that have been more problematic than others for ADAS. Some of the more substantive challenges include driver drowsiness and distraction, varied illumination, and occlusions. The problems of inconsistent illumination and occlusion of objects are sub-problems in the object detection domain. While in the drowsiness and distraction category, driver perception of objects may well benefit from the knowledge of a driver's object of attention. In the research conducted by Dong et al. [10] an investigation and proposed solution are provided for determining drowsiness or distraction of the driver. Various factors which may indicate drowsiness or distraction are collected and these factors are then searched for during testing, ultimately contributing to the classification of driver fatigue (utilizing a Random Forest) and driver distraction (utilizing an CNN). Of these factors, the driver's attention on objects, either inside or outside of the vehicle, is not considered. The closest parameters are some from the drowsiness detection, including various gaze directions (left, right, center) and blink frequency. Another attempt to recognize driver distraction is presented in the work by Banerjee et al. [3]. In this work, again, the gaze direction was utilized but divided into 6 categories instead of the 3 used in the previous work discussed. This research also looks at "down", "rearview mirror", and "instrument cluster" as gaze directions. Although this work does not handle the driver distraction itself, it does tackle glance estimation which is an important part of distraction. But similarly to the previously discussed work, this research also does not account for the driver's attention on objects.

2.2 Traffic Object Detection

Object detection is not a new breakthrough and traffic object detection is no different. There are plenty of works detailing both of these topics. Works have been done for each type of traffic object individually, including vehicle, traffic signs, traffic lights, pedestrians, etc. For traffic object detection we've seen various approaches over the years. Kuo et al. [15] provide works where we see colour and shape utilized in the detection of traffic signs. This work, although highly performing, is limited in its capabilities as it is restricted to signs either triangular or round in shape. This work does however nearly overcome the issue of illumination by utilizing HSI coloring rather than typical RGB for its specific use cases. In [8], Dalal et al. propose an approach for human detection that involves the use of locally normalized Histogram Oriented Gradient (HOG) descriptors. Coincidentally, this work also has proven performance for shape based object classes. The key discovery from this paper however was not the human detection itself, but rather the fact that smoothing actually hurts performance in human detection, whereas normally smoothing is beneficial. Wang et al. [28] show this performance by utilizing the work done in [8] to further improve upon the detection of traffic signs by applying it in a two step process, first using a HOG with a small window for classification followed by a HOG with a large window for a further classification. This work achieved nearly perfect classification across all sign categories (prohibitory, mandatory, and danger) but could benefit from an improvement to efficiency, taking over one second to detect a single object.

In terms of vehicles, they are more complicated to detect, mostly due to their large variation in given scenes as well as possible viewing angles. This can be attributed to the fact that, unlike stationary objects, vehicles are usually not stationary; they could also be in front of, behind, or beside the vehicle, giving a large range of possibilities to account for. Sivaraman et al. [23] provide us with a survey of methods for the detection of vehicles, which use vision based approaches. This paper indicates that vehicle detection is divided into two major categories; monocular and stereo-vision. Each of these categories can be further divided into appearance-based approaches and motion-based approaches. Though these methods may have proved useful in their time, today their usefulness comes from acting as a basis to apply alongside machine learning techniques. This is exemplified by Dong et al. [9]. In this research we can also see how utilizing a Convolutional Neural Network (CNN) alongside colour and depth images can improve upon vehicle detection. This proposed approach provided high performance while nearly maintaining weather indifference for the identification of vehicles. Error and error-prone samples are focused through targeted learning to improve on the areas with the worst performance and achieve high overall performance.

2.3 Driver's Gaze

Driver's gaze is a term used to describe where a driver is looking while driving, with past approaches focusing on either vision-based methods or learning-based methods. Vision-based methods focus more on the driver and their head/eye positions to estimate gaze while learning-based methods focus on the use of machine or deep learning practices, with or without driver head/eye data to estimate gaze.

Looking at vision-based approaches, Murphy et al. [17] and Baker et al. [2] provide early works in the area of head pose estimation while Hansen et al. [12] provide a summary of eye detection frameworks as well as the gaze estimation techniques at the time, prior to 2010. More recently we have seen works such as that produced by Sugano et al. [24] proposing a gaze estimation framework which auto-calibrates through the use of saliency maps. This work provides solid groundwork in the area of gaze estimation, but is limited to stationary head position.

For more learning-based approaches, with the help of eye gaze tracking, Bublea et al. [7] produce two Deep Neural Networks (DNN) which can determine the driver's behaviour (drowsiness, distracted, etc.) during a driving sequence. It proposes two methods for eye gaze estimation, a geometric approach and an auto-keras approach, ultimately producing almost identical results. We can also see saliency maps utilized on eye imagery in [24] to estimate the gaze of users watching a video.

More recent research looks to incorporate vision and learning based approaches to estimate gaze. This is exemplified in the work done by Shirpour et al. [22], considering the position of the driver's head in relation to a forward stereo system implemented within an experimental vehicle. Gaussian process regression is then implemented using this head pose to estimate the gaze direction as a confidence interval, which is less hardware intensive and does not require eye tracking capabilities. During further investigation, Shirpour et al. integrates eye tracker data to determine a singular point of gaze in the driving scene. With these two sets of data (the gaze estimation and the eye tracker data) the eye fixation is then estimated. This final driver's eye fixation estimation is exhibited as a confidence region on the frame image.

2.4 Driver Perception

Defining ”seen” and ”unseen” objects is logically simple, either the driver sees an object, or they do not see it. Typically, any research which does make reference to objects being seen, uses very simplistic definitions, usually defining ”seen” as the driver's gaze being on a given object during a single frame, and ”unseen” being any object not falling under this definition of ”seen”.

One such example of works done which consider defining ”seen” and ”unseen” objects is that done by Shabani et al. [20]. This work looks to assess whether drivers are seeing traffic signs present during driving sequences. The definition of ”seen” used is one which looks for the driver's point of gaze and whether it intersects with a traffic sign in the corresponding frame. This definition may suffice for some research, such as that done by Shabani et al. whom use it as a feature in the determination of driver's attention, but the consideration of ”seeing” vs ”gazing upon” could benefit other research.

In the work done by Huang et al. [14], the consideration of driver focus of attention (DFoA) and True DFoA (TDFoA) is utilized to produce a driver distraction detection (DDD) method. Huang et al. developed D3DRN-AMED to act as the tool to predict TDFoA. This TDFoA is then included in the proposed DDD process along with the DFoA. This work, similarly to that done by Shabani et al., focuses on the driver's attention in a singular frame, but only considers attentional area, not whether the object is seen or not. Another example of research in this area is that conducted by Tang et al. [26], which combines driver gaze data and a car's perception of the environment to accentuate the objects of concern only if they are deemed as ”not perceived” by the driver. This research proposes a ADAS to reduce unnecessary warnings to the driver, reducing clutter and potentially avoiding over stimulation of the driver. Once again, though the research provides us with a quality ADAS system, the idea of perception is defined by a point of gaze on the environment within a single frame.

In the founding works of the DR(eye)VE project, Palazzi et al [19], looks to predict what drivers will pay attention to whilst driving. A unique approach is taken in this work, where rather than focusing on the drivers gaze in a frame, a series of frames is considered. These frames are analyzed and in the final frame of the series, a fixation map is produced to show the areas most likely to be looked at by the driver. Works in ADAS research that consider more than a singular frame when considering driver perception or attention are near non-existent, and even this piece, providing both a comprehensive dataset and an excellent predictor of driver attention, does not touch on what it means for a driver to ”see” an object. The contribution we hope to make involves an analysis of what it means to ”see” through the consideration of point of gaze over time (multiple frames). The main difference between the DR(eye)VE project and the work presented in this thesis is that the DR(eye)VE project is a prediction tool which utilizes a series of frames to predict where a driver is likely to look, whereas this work takes a series of frames and determines where the driver is actually looking, based on our given definition of seen.

2.5 Summary

The work presented in this thesis examines the parameter of time, i.e., number of frames, and its impact when considering a definition of seen. This analysis is done in the context of different types of objects important for driving activities. This concept is not present in much of the previous research around driver gaze, driver perception or ADASs, but is a topic that deserves investigation. We take the novel consideration of this parameter of time, and examine how different lengths of time can affect what is considered to be ”seen” or not and on the different types of objects considered to be seen.

Chapter 3

Data Processing & Architecture

In this Chapter we look at how the data used for our analyses was obtained and how it is organized. We also take a look at the architecture of the object detection neural net. Finally, we look the data processing and manipulation done in both preparing the data and utilizing it, as well as some of the algorithms used.

3.1 Preliminary Work

This work builds off of that done by Shirpour et al., culminating in 2021. The foundation laid by this work provided a neural network capable of producing quality object detection and recognition, as well as point of gaze data. However, the software from this work was done several years back and parts of the code and neural networks were outdated and not up to current standards, or even able to compile and run properly. This required a number of adjustments and significant error corrections to be made in order to bring the software to a usable state. The process was long and tedious, essentially involving running the code to receive an uncommon and convoluted error message, followed by the deciphering of this message to determine what needed to be changed or updated to match current standards and practices.

3.2 RoadLab

Beauchemin et al [4] developed a tool to aid in ADAS research. This tool is the Roadlab augmented vehicle, which is a vehicle equipped with a variety of sensors and cameras, as well as GPS, utilized for capturing driving data during driving sequences. These devices were used to record driver, vehicle, and environmental information in and around the vehicle. The vehicle also featured an on-board computing system with LCD displays and a disk-less cluster of computing nodes, with scalability being the foremost guiding principle. The on-board laboratory is also equipped with OBD-II (On-Board Diagnostics, 2nd generation) via the CANbus (Controller Area Network vehicle bus) protocol, which acts as an on-board diagnostic system, allowing for real-time collection of vehicle subsystem states utilizing the the CANbus communications network for interconnected components.

The Roadlab vehicle was utilized to record driving sequences from 16 different drivers. The drivers followed an identical route, producing approximately 100,000 frames of data each, captured at 30Hz. All of the participants were aged 20 to 47 and performed their driving sequence in London, Ontario, Canada. A secondary researcher was also present during each driver's sequence in the passenger's seat, acting as a navigator as well as monitoring the on-board equipment. From the sensors and cameras, the data collected includes frontal stereoscopic video, driver head position and angle, and driver ocular parameters. The OBD-II system provides various vehicle subsystem information, including "current speed and acceleration (longitudinal and lateral), steering wheel rotation, state of accelerator and brake pedals, and independent wheel speed" [4] and is captured at 60Hz. Additionally, GPS positional data was also recorded for all drivers at 60Hz.

3.3 Data Collection and Organization

As discussed previously, the data collected included 16 unique drivers routes, consisting of approximately 100,000 frames of data each. These frames provide a visual image displaying the area in front of the vehicle and the driver. In the research presented, one of these driving sequences is used in full. The driver utilized was subject 8, consisting of 101,290 frames of data in total. Due to poor gaze data for the last portion of frames, the final 70 frames were removed, leaving 101,220 frames of data. In terms of gaze data, Beauchemin et al. [5] describe the analysis of the the data gathered by the RoadLab project to determine produce a single point in frame indicating the Point of Gaze (PoG) of the driver for that frame. This PoG is presented as a coordinate (x,y) which may fall within or outside of the current frame, as the driver may not always be looking straight ahead (or in frame). To determine the Point of Gaze, cameras which focus on the driver's eyes were utilized to determine the position and direction of the eye, and then software is used to compute the gaze vector. The gaze vector is then mapped to the same coordinate system as the video images of the visual environment which is finally used to determine the point of gaze. This enables us to obtain only a single point of gaze for each frame.

The RoadLab data was gathered over a decade ago,with the data gathering taking place in 2010. This data, although old, is still very relevant and valid. The data collected consists real driving sequences and provides a solid dataset for use in research. As for the validity of the data, although it could be argued that vehicle or traffic objects styles have since changed, they remain relatively constant as well. Traffic signs and traffic lights have undergone few changes. The style of vehicles may change, but their size, shape, features (such as number of wheels, doors, windows, etc.) remain relatively unchanged. So while it is worth noting that things have changed since 2010, the dataset is very relevant considering today's driving environment.

The table below (Table 3.1) presents examples of the basic gaze data which was collected and calculated. This data includes the frame number being referenced, as well as the calculated x and y coordinates representing the PoG of the driver during this frame (a single point being gazed upon).

Frame Number	X-Coordinate	Y-Coordinate
6980	311.647888	146.613037
6981	303.729431	147.876968
6982	296.205322	150.162399
6983	284.398773	151.632599
6984	280.167969	150.933044

Table 3.1: Driver point of gaze data

3.4 Object Detection

The object detection method used for this research was based on the work from Shirpour [22], which utilized a neural net to detect bounding boxes in a driving sequence. The neural net is a compilation of two separate models, model A and model B. Each of these models provided benefits to the final neural net; model A provided better results on smaller or further objects while model B provided better results for larger or closer objects. Model A consisted of two parts, a multi-scale HOG-SVM and a ResNet-101 network, whereas Model B utilized a Faster R-CNN. The combination of these two models create a neural net object detection framework capable of 96.1%, 96.2%, and 94.8% of correct classification for traffic signs, traffic lights, and vehicles respectively.

This codebase provided by Shirpour in [22] was mostly left unaltered. Although the single ”main” file needed to be altered slightly from the original for the purpose of this research. These changes mostly revolved around the compatibility to and preparation for this research. The first changes were made with regards to the output. The initial code would input a driving sequence as a series of frames, and output the same series of frames with bounding boxes (BBs) overlaid. This was great for visualization, but did not allow for analysis of the core data. To resolve this, the code was adjusted to output the various data from the frames and BBs, after each frame was overlaid with said BBs. The second set of changes made had to do with how the code was run. This did not change the code’s purpose or output, but made it possible to run the code in batches, outputting only potions of the desired final output at a time. This was done as the time and resources required to run one single iteration for all data was impractical.

3.5 Data Processing

The data processing done for this research starts with the initial driving sequence frames and outputs data revolving around whether certain objects were seen while driving. To do this, the starting point is the set of initial, unaltered driving sequence frame images, then BBs are overlaid on relevant traffic objects, as well as this BB data being input into a table. Next, the driver's PoG is overlaid on the BB images, and the PoG is also added to the now created table. Afterwards, the PoG is considered against the position of BBs to determine if the PoG falls within any BB in a given frame, keeping a count of consecutive frames where the PoG falls within any BB. Finally, with the set of all sequences of frames where the PoG fell within any BB, the sequences are checked to determine for which sequences the PoG falls within is the same BB (object) for the entire sequence.

The data collected from the RoadLab project provided a series of frames from a driving sequence, and the work done by Shirpour [22] provided a coordinate for each frame corresponding to PoG. Additionally, the work done by Beauchemin et al. [5] yielded a neural network capable of detecting bounding boxes and overlaying each frame with these detected bounding boxes.

3.5.1 Determining Bounding Boxes

The first step in the data processing approach was to both produce the bounding box images, as well as extract the bounding box data. The most bounding boxes detected by the neural net in a single frame for this dataset was 14 (Figure 3.1), which is later considered when constructing tables. This max of 14 was not a set cap but simply just the most bounding boxes seen in a single frame for this iteration. The bounding box data required for this investigation included the position of the top left corner of the bounding box, the length and width of each bounding box, and the label for each bounding box. This was done by adjusting the code utilizing the neural net to also keep record of this data and output it to a text file upon completion. Frames then had to be provided to the neural net and allow adequate time for the bounding box images to be produced, taking advantage of Compute Canada's resources to do so.

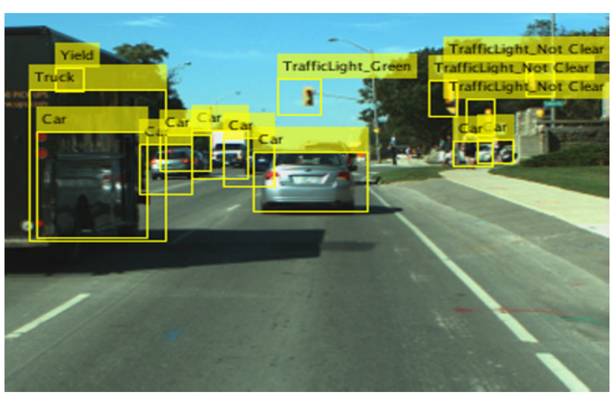


Figure 3.1: Frame with the most Bounding Boxes

The diagram presented in Figure 3.2 illustrates the overall architecture of the object detection process described above. The two neural networks described previously are integrated together to create a single neural network using non-maximum suppression (NMS) to create a single network which can perform the object detection. This network is then trained on three independent models; vehicles, traffic signs, and traffic lights to finally output bounding boxes with labels.

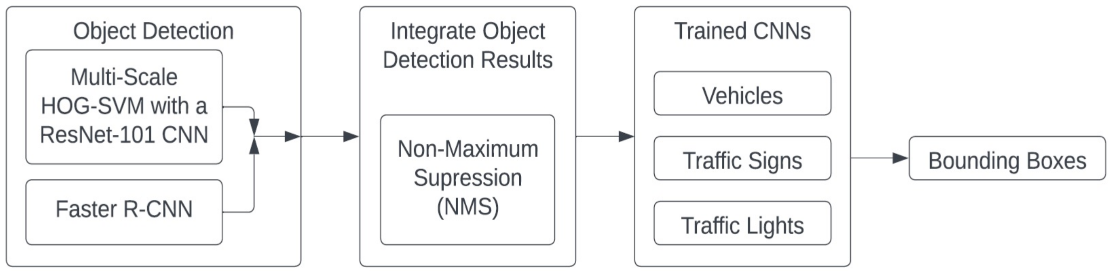


Figure 3.2: Object detection architecture

The table below (Table 3.2) presents a frame by frame representation of the data extracted from the detected images during the driving sequence. After running the driving sequence through the bounding box neural net, this table is essentially the output. For each frame looked at (which is every frame in a single driving sequence) the code outputs data about each bounding box detected. This data includes the bounding box's label, the x and y coordinate of the top left corner, the length, and the width. This data is important because with the top-left coordinate, length, and width, any other information about the bounding box that may be required could be calculated, such as any other corner, or whether a given coordinate falls within the bounding box. The table size is also dynamically calculated to be as wide as required for the given max number of bounding boxes found in a single frame, and as long as the number of frames in the given driving sequence.

Frame	Label 1	X	Y	Length	Width	Label 2	X	Y	Length	Width
10954	Car	153	82	32	29	NA	NA	NA	NA	NA
10955	Car	155	83	26	26	NA	NA	NA	NA	NA
10956	Car	156	84	25	26	NA	NA	NA	NA	NA
10957	Car	157	84	23	25	BicycleLane	12	80	16	16
10958	Car	156	82	23	25	NA	NA	NA	NA	NA

Table 3.2: Bounding box data for first two labels

Compute Canada's (CC) resources were a necessity for this project as the run time for even just one subject was unrealistic on a single device. But using these resources was not without complications. The neural net code from [22] had initially begun several years ago and as a result the code was not up to date to work with current software and resources. This was the most time consuming portion of this research as there was not much support to be found regarding the resolution of the various incompatibilities encountered. This lead to long sessions of trial and error fixing one error after another until producing a version of the code with no incompatibilities. Once the code was working locally, it then had to be made compatible with CC's resources. This was another time consuming process that eventually lead to a code version which worked both locally as well as on the CC servers. From there the code simply had to be adjusted to be able to run in portions so that an array job could be submitted to the CC server. The code ran over a few 3 hour sessions, producing the bounding box images for the chosen subject's driving sequence.

3.5.2 Determining Point of Gaze in Each Frame

The next step in the data processing approach was to visualize the PoG with the bounding box images. This was done simply by overlaying the PoG coordinate as a red cross on the bounding box images (Figure 3.2). At the same time this was being completed, PoG coordinates were also able to be added to the bounding box data, as well as adjusting the data to be more uniform, by adding a placeholder "NA" to rows (or frames) which had fewer to no bounding boxes in order to maintain a uniform row size. This would prove useful for further data processing. Upon completion of this step, there was now had a singular file containing all pertinent data. This pertinent data includes the frame number, the PoG coordinates, as well as the bounding box label and measurements (top-left coordinate, width, and height) for each bounding box in the frame.

Figure 3.3: Driver PoG represented by a red cross

The diagram presented in Figure 3.4 illustrates the architecture of how the point of gaze is determined. The gaze data is taken from that of the RoadLab project, which is then utilized to compute a gaze vector directed outward from the eye. The gaze vector is then mapped to the same visual environment coordinate system as the video, producing an intersection between the two. This intersection determines the Point of Gaze on an individual frame.

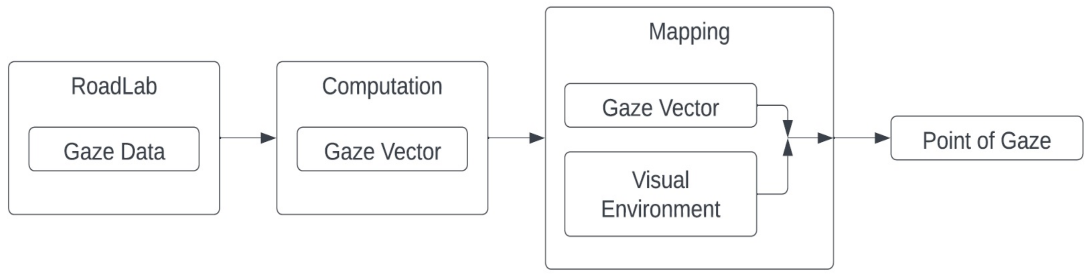


Figure 3.4: Point of Gaze determination architecture

3.5.3 Consecutive Frame Analysis

The following step was to find all series of images in which the driver gazed at a bounding box for a given number of consecutive frames. This task was broken down into sub-tasks. The first sub-task would be to find all frames in which the PoG falls within any bounding box for that frame. The next would be to look at all the images where bounding boxes were gazed upon and determine the instances where this produces a series of frames. The final sub-task would be to take the series of consecutive frames where bounding boxes were gazed upon, and extract those in which the bounding box gazed upon remained constant (the same bounding box is gazed upon during the whole series). This process is illustrated in the diagram presented in Figure 3.5.

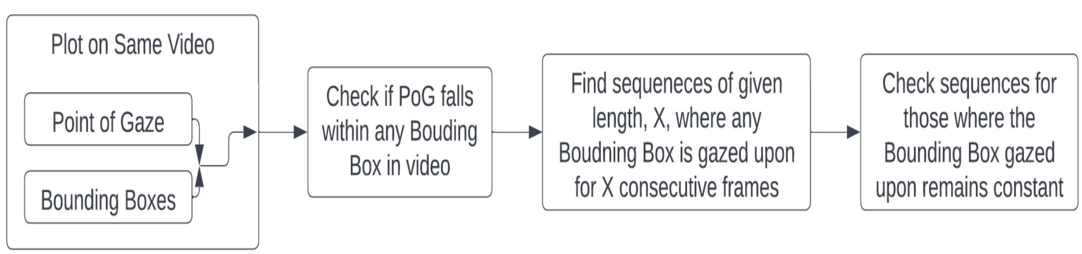


Figure 3.5: Determination of ”Seen” objects architecture

Starting with the first sub-task, using Algorithm 1, we simply have to check that the PoG falls within any bounding box in the frame. We can utilize the bounding box top left corner coordinate, as well as the height or width to find a range within each of the x plane and the y plane. We then check that the PoG x value and y value fall within these determined ranges, if so, we then can safely say that the PoG fell within the bounding box. We repeat this step for each bounding box in the frame until there remains no more bounding boxes. Each bounding box for the frame is marked with either a 1 or a 0 indicating whether or not the PoG fell within its bounds. These results were then utilized in the subsequent sub-task.

Algorithm 1 Finding all PoGs which fall within BBs

```
numRows ← 101220 #Total Rows (frames)
currRow ← 1
numCols ← 73 #Total Columns
for currRow ≤ numRows do
    currCol ← 4
    label ← firstBBLabelinCurrFrame
    pog ← [currentPOGxcoordinate, currentPOGycoordinate]
    count = 0
    while BBlabel ≠ NA and currCol ≤ numCols do
        x ← currentBBxcoordinate
        y ← currentBBycoordinate
        xDist ← currentBBWidth
        yDist ← currentBBHeight
        if x ≤ pog(1) and x + xDist ≥ pog(1) then
            if y ≤ pog(2) and y + yDist ≥ pog(2) then
                pogInCurrentBB ← 1 #TRUE value
                count ← count + 1
            end if
        end if
        if count ≥ 2 then
            POGinMultipleBB ← Y
        else
            POGinMultipleBB ← N
        end if
        currCol ← currCol + 5 #Moves to next BB
        if currCol ≤ numCols then
            label ← nextLabelSameFrame
        end if
    end while
end for
```

The second sub-task utilized Algorithm 2 to find all series of consecutive frames in which any bounding box was gazed upon. To do so, we look at each frame and check to see if any bounding boxes were gazed upon. If so, then we begin a counter and move to the next frame. We repeat these steps until we either meet a predetermined sequence length, or reach a frame in which no bounding boxes were gazed upon. For the former, we add the first frame in the sequence to the list. In both instances, we reset the counter and move back to the frame following the first frame checked. This allows series of frames to be indicated which may overlap, but will also allow us the option to investigate series longer than that required without having to alter or rerun the code.

Algorithm 2 Counting consecutive frames where a POG falls in any BB

$numRowsPOG \leftarrow 101220$ #Total Rows (frames)
$numColsPOG \leftarrow 16$ #Columns in table holding the BB gazed flags
$counter \leftarrow 0$
$currRow \leftarrow 1$
$startingCol \leftarrow 3$
while $currRow \leq numRowsPOG$ **do**
  for $currCol = startingCol : numColsPOG$ **do**
    if $BoundingBoxPOGFlags(currRow, CurrCol) = 1$ **then**
      $count \leftarrow count + 1$ **break**
    else if $BoundingBoxPOGFlags(currRow, CurrCol) = NA$ **then**
      $counter \leftarrow 0$ **break**
    end if
  end for
  if $counter \geq consecFramesToBeSeen$ **then**
    $framesGazed.append(currFrameNumber)$
    $counter \leftarrow 0$
    $currRow \leftarrow currRow - consecFramesToBeSeen + 1$
  end if
  $currRow = currRow + 1$
end while

The final sub-task used Algorithm 3 to check which sequences of frames gazed upon had the PoG fall on the same bounding box over the entire sequence. The approach used to determine whether a bounding box gazed upon is the same in each frame was to utilize the bounding box change between frames. The bounding box was given a threshold for pixel variation. If the bounding box moved by more pixels than the threshold in either direction, then it was not considered to be the same bounding box. Given this restriction we can look at the top left corner of the bounding box for two adjacent frames, and if this point changes by less than the threshold between the two, then we can consider it to be the same bounding box.

Algorithm 3 Comparing consecutive gazed frames to determine if the BBs remains constant

---

$numRowsGazed \leftarrow 9417$ #Number of rows where PoG fell in consecutive BBs
$numColsGazed \leftarrow 16$ #Columns in table holding the BB gazed flags
$counter \leftarrow 1$
$currRow \leftarrow 1$
$startingCol \leftarrow 3$
$BBPixelThreshold \leftarrow 30$ #Number of Pixels allotted for BB to move between frames
while $currRow \leq numRowsGazed$ **do**
    $frameNum \leftarrow framesGazed(currRow)$
    for $i = 1 : consecFramesToBeSeen - 1$ **do**
        for $CurrCol = startingCol : numColsGazed$ **do**
            for $nextCol = startingCol : numColsGazed$ **do**
                if $POGinBBflag(currFrame, currCol) = 1$
                 and $POGinBBflag(nextFrame, nextCol) = 1$ **then**
                    $topLeftCurr \leftarrow topLeftBBCoordinateCurrFrame$
                    $topLeftNext \leftarrow topLeftBBCoordinateNextFrame$
                    $topLeftDiff \leftarrow topLeftBBCoordinatesDifference$
                    if $(topLeftDiff(1) \leq BBPixelThreshold$
                      and $topLeftDiff(2) \leq BBPixelThreshold)$ **then**
                        $counter \leftarrow counter + 1$ **break**
                    end if
                else if $POGinBBflag = NA$ **then break**
                end if
                if $counter \geq consecFramesToBeSeen$ **then**
                    $counter = 1$
                end if
                if $counter \geq consecFramesToBeSeen$ **then**
                end if
            end for
        end for
    end for
    $currRow \leftarrow currRow + 1$
end while

At this point a table was produced containing all of the frames which were the first frame in each sequence where the PoG remained in the same bounding box for the entire sequence, stopping when the required sequence length was reached. From this we look at the consecutive frames which appear in this table and add 1 to the sequence length for each frame following the first in a sequence. This is done because if consecutive frames appear in this table, they have to be gazing at the same bounding box, so we simplify the table by instead listing the starting frame number with its sequence length, rather than the starting frames for every sequence exactly equal to the required sequence length.

The table below (Table 3.3) illustrates the table described above. The first column shows the frame number where the sequence starts, with the following column indicating the length of the sequence found. The length of the sequence must be greater than or equal to the minimum required sequence length being considered (which is 10 frames in this case).

Frame	**Length of Sequence**
330	10
354	10
392	10
656	16
720	15

Table 3.3: Consecutive frames with length of sequence

Chapter 4

Results

In this Chapter we introduce how we define whether a traffic object is seen by the driver or not. We also take a look at and summarize the various outputs and data generated throughout this research. The data used for this analysis is that which was discussed previously, ranging from the frame numbers and bounding boxes, to the sequences and their lengths. Counts are taken of different groupings and percentages are calculated to help quantify our results. The definition of ”seen”, which we will introduce, will also be incorporated into the final calculations. Following this will be the observations drawn from these results. These include any notable points of interest or areas worth discussing and dissecting more thoroughly.

4.1 Definition of ”Seen”

Whether or not an object can be considered as seen is a complex question requiring knowledge in human cognition, and one which is worthy of discussion, but is beyond the scope of this research. For our purposes we will define ”seen” as an object which has been gazed upon for a predetermined number of frames during the driving sequence. This number must be large enough that the driver could be considered to have acknowledged the object gazed upon, but not so long that a driver, who’s eyes would be constantly scanning the environment, would never gaze at a single object for that long. Since the idea of ”seeing” is more of a cognitive concept, a broader approach to seeing is considered. A range of 3 frames to 30 frames (or 0.1 second to 1 second) is what was considered to be worth exploring. The other variable that would be important would be a threshold value for bounding box changes, which would be measured in pixel variation between frames. This variable would indicate the number of pixels a bounding box could differ by, between two consecutive frames, to be considered the same bounding box. We investigated bounding box pixel variations ranging from 10 pixels to 30 pixels.

4.2 Summary of Objects Seen

The main goal of this research is to determine the objects which were seen or unseen by drivers and whether there were differences in which types traffic objects were seen or not by a driver during a driving sequence. For this there is a variety of data which could prove useful. Knowing the number of bounding boxes detected and the number of bounding boxes seen could let us differentiate between an object type being missed or simply not being readily available in the driving sequence. It could also provide us a look into object types which are frequently or infrequently seen when available. Adding sequence count into this data could then allow us to see average sequence lengths for object types, which would translate into a length of time that object type is typically gazed upon when considered seen. With this data available it would then be possible to calculate some percentages such as the percentage of BBs seen of an object type and the percentage of total frames in the driving sequence where object types are seen.

With this in mind, we determined for each type of traffic object the total number of bounding boxes detected (Total BB Count) in the driving sequence, the total number of frames where they were deemed as seen (Total BB Seen), the total number of sequences found (Seq Count), the average length of the sequences (Avg Seq Length), the percentage of objects seen out of their total occurrences (% of BB), and the percentage of objects seen out of the total number of frames in the driving sequence(% of Total Frames).

Tables 4.1 through 4.4 present results from our analysis representing different frameWindow and bbThreshold values to illustrate the results acquired. These 4 tables were chosen to be representative of the frameWindow and bbThreshold variations, showing the impact they have on the results. Appendix A provides an additional 9 tables showing different variable combinations (frameWindows of 5, 10, 15, & 20 and bbThresholds 10, 20, & 30).

Table 4.1 presents the results of utilizing a frameWindow of 5 frames and a bbThreshold of 30 pixels. This combination of frameWindow and bbThreshold is very lenient, requiring only 5 consecutive frames for an object to be considered ”seen” and allowing for bounding boxes to move as many as 30 pixels while being considered the same bounding box. As a result, this table is one of few to show this number of seen objects. With this low of a frameWindow, some objects which do not appear very frequently or are not seen frequently, are more likely to be considered as seen. Additionally, of the total 101,200 frames, approximately 13.9% of them include a bounding box which is considered as seen.

Label	Total BB Count	Total BB Seen	Seq Count	Avg Seq Length	% of BBs	% of Total Frames
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	5	1	5	0.27	<0.01
Bus	10184	1236	141	8.77	12.14	1.22
Car	251200	11234	1206	9.32	4.47	11.10
Construction	1265	8	1	8	0.63	<0.01
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	109	11	9.91	1.76	0.11
LaneTurnsRight	745	6	1	6	0.81	<0.01
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	10	1	10	0.46	<0.01
NoTurn	4713	15	2	7.5	0.32	0.01
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	0	0	0	0	0
Pedestrian	43756	179	22	8.14	0.41	0.18
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	5	1	5	0.62	<0.01
Stop	48	0	0	0	0	0
TrafficLightAhead	659	5	1	5	0.76	<0.01
TrafficLight_Green	19044	13	2	6.5	0.07	0.01
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	70	11	6.36	0.51	0.07
TrafficLight_Yellow	4621	31	3	10.33	0.67	0.03
Truck	20633	1095	124	8.83	5.31	1.08
WatchForPedestrian	1002	0	0	0	0	0
Yield	2009	0	0	0	0	0
Totals	**406645**	**14021**	**1528**	**N/A**	**N/A**	**13.87**

Table 4.1: Summary table for frameWindow of 5 frames and bbThreshold of 30 pixels

Table 4.2 presents the results of utilizing a frameWindow of 10 frames and a bbThreshold of 20 pixels. This combination of frameWindow and bbThreshold is pretty central in the data, requiring a reasonable number of consecutive frames as well as not allowing to the bounding box to move an excessive amount while being considered the same. Some of the objects which were barely seen in Table 4.1 are no longer considered seen, such as BicycleLane and Construction, thus exemplifying the effect that changing frameWindows can have. It is also worth noting that most objects overall are seen less than in the previous table. Looking again at the percentage of frames seen, a decrease from 13.9% to 12.5% is seen between Tab;e 4.1 and Table 4.2.

Label	Total BB Count	Total BB Seen	Seq Count	Avg Seq Length	% of BBs	% of Total Frames
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	0	0	0	0	0
Bus	10184	1597	99	16.13	15.68	1.58
Car	251200	9784	621	15.76	3.89	9.67
Construction	1265	0	0	0	0	0
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	82	5	16.4	1.32	0.08
LaneTurnsRight	745	0	0	0	0	0
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	10	1	10	0.46	<0.01
NoTurn	4713	0	0	0	0	0
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	0	0	0	0	0
Pedestrian	43756	156	8	19.5	0.36	0.15
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	0	0	0	0	0
Stop	48	0	0	0	0	0
TrafficLightAhead	659	0	0	0	0	0
TrafficLight_Green	19044	16	1	16	0.08	0.02
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	14	1	14	0.10	0.01
TrafficLight_Yellow	4621	25	2	12.5	0.54	0.02
Truck	20633	941	63	14.94	4.56	0.93
WatchForPedestrian	1002	0	0	0	0	0
Yield	2009	0	0	0	0	0
Totals	**406645**	**12625**	**801**	**N/A**	**N/A**	**12.47**

Table 4.2: Summary table for frameWindow of 10 frames and bbThreshold of 20 pixels

In Table 4.3 we see the results of utilizing a frameWindow of 15 frames and a bbThreshold of 20 pixels. This combination of frameWindow and bbThreshold is where the object types seen really begin to dwindle, largely due to the frameWindow increase. In this table there remains only 6 unique object types which are seen, with half of these being seen very rarely. This demonstrates that requiring too high of a frameWindow will lead to objects not being considered seen which might have been seen if a more appropriate frameWindow were selected. Again, we also note that the overall number of objects seen decreases when comparing the results in this table to Tables 4.1 and 4.2; compared to these tables the percentage of frames seen further decreases from 12.5% to 10.8%.

Label	Total BB Count	Total BB Seen	Seq Count	Avg Seq Length	% of BBs	% of Total Frames
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	0	0	0	0	0
Bus	10184	1600	61	26.23	15.71	1.58
Car	251200	8627	332	25.98	3.43	8.52
Construction	1265	0	0	0	0	0
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	27	1	27	0.43	0.03
LaneTurnsRight	745	0	0	0	0	0
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	0	0	0	0	0
NoTurn	4713	0	0	0	0	0
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	0	0	0	0	0
Pedestrian	43756	41	2	20.5	0.09	0.04
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	0	0	0	0	0
Stop	48	0	0	0	0	0
TrafficLightAhead	659	0	0	0	0	0
TrafficLight_Green	19044	0	0	0	0	0
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	33	1	33	0.24	0.03
TrafficLight_Yellow	4621	0	0	0	0	0
Truck	20633	604	25	24.16	2.93	0.60
WatchForPedestrian	1002	0	0	0	0	0
Yield	2009	0	0	0	0	0
Totals	**406645**	**10932**	**422**	**N/A**	**N/A**	**10.80**

Table 4.3: Summary table for frameWindow of 15 frames and bbThreshold of 20 pixels

In the fourth table, Table 4.4, we see the results of utilizing a frameWindow of 25 frames and a bbThreshold of 20 pixels. This combination of frameWindow and bbThreshold gives an output showing the lack of objects seen when the frameWindow begins to get too large, requiring almost a full second of attention from the driver. Again, other object types are no longer seen and the overall number of seen objects decreases further. This table is intended to reiterate the impact of implementing a frameWindow that is too long. Another decrease in the percentage of objects seen is observed from 10.8% to 8.75%.

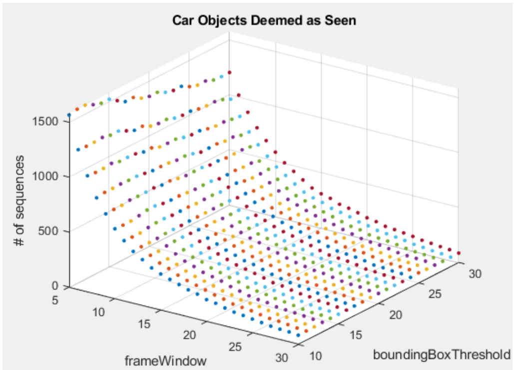
Label	Total BB Count	Total BB Seen	Seq Count	Avg Seq Length	% of BBs	% of Total Frames
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	0	0	0	0	0
Bus	10184	1360	34	40	13.35	1.34
Car	251200	6683	134	49.87	2.66	6.60
Construction	1265	0	0	0	0	0
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	27	1	27	0.43	0.03
LaneTurnsRight	745	0	0	0	0	0
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	0	0	0	0	0
NoTurn	4713	0	0	0	0	0
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	0	0	0	0	0
Pedestrian	43756	0	0	0	0	0
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	0	0	0	0	0
Stop	48	0	0	0	0	0
TrafficLightAhead	659	0	0	0	0	0
TrafficLight_Green	19044	0	0	0	0	0
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	48	1	48	0.35	0.05
TrafficLight_Yellow	4621	0	0	0	0	0
Truck	20633	739	14	52.79	3.58	0.73
WatchForPedestrian	1002	0	0	0	0	0
Yield	2009	0	0	0	0	0
Totals	**406645**	**8857**	**184**	**N/A**	**N/A**	**8.75**

Table 4.4: Summary table for frameWindow of 25 frames and bbThreshold of 20 pixels

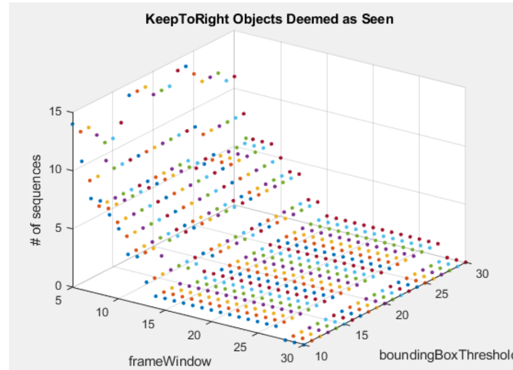
4.3 Observations

At first glance it is apparent that only a few specific types of objects are seen much more frequently than the very common ones, namely, vehicles - buses, cars, and trucks. This trend extends to all 588 combinations of variables tested. Upon graphing each of the 28 traffic objects comparing the number of sequences found (see Figure 4.1 on the following page), the frameWindow variable, and the bbThreshold variable, there is a common trend for each type of traffic object. The frameWindow has a negative trend whereby increasing the value will decrease the number of sequences found. This trend continues until a sufficiently low number of sequences are found. Also, when the sequences found is sufficiently high, the bbThreshold will also have a negative trend with the number of sequences found, but only having a slight impact. This trend with bbThreshold is also only seen with the vehicle objects, presumably due to these objects being the only ones with a considerable number of sequences seen. This can be seen in Figure 4.1 which presents four graphs showing traffic objects with varying numbers of sequences found. All 28 graphs look similar to one these four depending on the number of sequences which were found for their respective object.

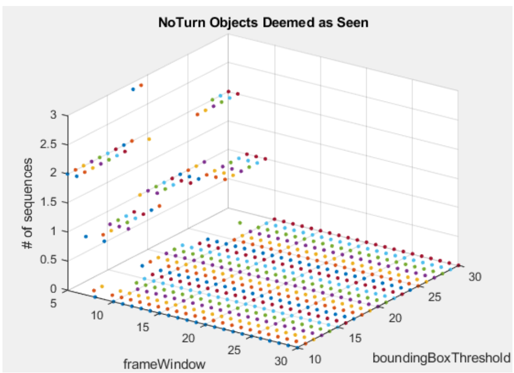
Another observation is that for many object types the number of objects seen is very low or even zero. These objects included those which were smaller (parking), uncommon (railroad crossing), or may not have been necessary to look at if the information is known (speed limit signs). This was a common trend among all variable inputs, with very low frameWindows having the highest number of sequences found for these objects. Even with very low frameWindows, the sequences found, though more than other frameWindow values, were still overall very low. Looking at Figure 4.1a and 4.1b below, we can see that even a very low frameWindow of 5 frames, KeepToRight (an object typically seen very infrequently) still only reaches a peak of less than 15 seen. In Table 4.1, which represents a very lenient variable selection of 5 frames for frameWindow and 30 pixels for bbThreshold, 13 of a total 28 object types went completely unseen, with another 12 objects being seen in less than 0.2% of frames in the driving sequence. Of these 12 objects, 11 of them also only have less than 1% of their total bounding boxes being seen. Additionally, when looking at 4.1a, we can see that the impact of increasing the frameWindow by 1 additional frame has a much less drastic impact once we reach a frameWindow of approximately 12-15 frames. Thus, considering the frameWindows of up to 15 frames is sufficient.



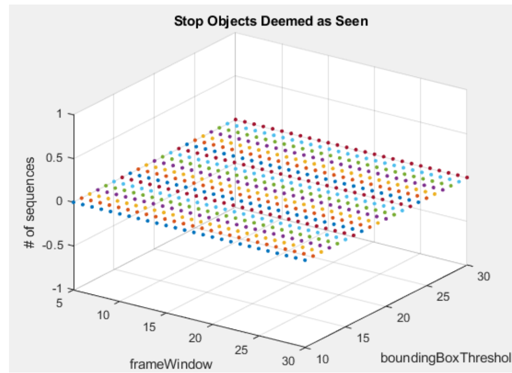
(a) Car



(b) Keep To Right



(c) No Turn



(d) Stop

Figure 4.1: Various graphs showing the effect of changes in frameWindow and bbThreshold

Another interesting point is the distribution of BBs and the number of BBs seen. As would be expected, the objects with higher numbers of BBs detected were more likely to have higher numbers of BBs seen, in general. There are some instances though where this is not the case, for example, when looking at pedestrians, trucks, and busses. Under our lenient variables, Table 4.1, we can see that pedestrians have approximately twice as many BBs detected as trucks and nearly four times as many detected as busses, with each having similar average sequence lengths between 8 and 9 frames. But even being much more commonly detected, the total BBs actually seen is much lower, approximately 6 to 7 times less.

The final point of interest is noticed when looking at the percentage of total frames, which represents the percentage of frames in the driving sequence in which the driver has seen an object of one of our classes. This total ranges from 8.75% to 13.9% in Tables 4.1 through 4.4. Even in the most lenient case, a maximum of 13.9% is reached, which seems very low as one might expect that drivers would see more during their time driving. However, as noted, this analysis only considers the traffic objects in those classes considered in this work. The remaining 86.1% of frames not accounted for can likely be attributed not only the objects where gaze is directed for fewer than the required number of frames, but also for all other objects not examined in this work, including, but not limited to, the road, power lines, road markings, buildings, advertising signs, and many more.

4.4 Implications

When performing analysis at a higher level, there are a couple more notable points of interest. The first of these being that the most commonly seen objects are vehicles (cars, buses, and trucks), with pedestrians being the next most seen objects. These objects could be considered as some of the most critical objects to see with regards to safe driving. Additionally, while drivers are focused on vehicles most of the time, that also means that drivers' attention is on the road most of the time since, unlike signs or traffic lights, vehicles are usually on the road. This itself is very useful information for ADAS because it tells us that a driver's gaze is typically ahead of themselves and on the road. To the inverse of this, knowing driver's gaze is typically on the road tells us that their gaze is not as frequently on signage or similar. With this in mind, it could imply that advancement in sign information processing could be more beneficial than vehicle information processing, as vehicles are already more commonly seen.

Another interesting implication of these results has to do with the sequence length, or the frameWindow. By the time the frameWindow reaches 15 frames (0.5 second) the number of objects seen is sufficiently low. When considering cars in particular, which was the most frequently seen object, looking at frameWindows from 5 frames to 15 frames we see the number of cars seen declining anywhere from 60% to 80% depending on the bbThreshold used. This information tells us that drivers can see objects in shorter amounts of time, with 12-15 frames, or 0.4-0.5 seconds, being an appropriate maximum to set. Anything beyond 15 frames seems to have little impact on the results comparatively to frameWindows less than 15 frames.

Finally, based on the data collected we see consistency between different combinations of variables. Although decreasing the frameWindow size decreases the total number of objects seen, the proportions between which objects are seen remain consistent throughout. Vehicles are always seen the most, KeepToRight is always seen with much lower frequency, objects that are not seen at low frameWindows are also never seen at high frameWindows. This consistency reiterates the implications of the results as we know the implications drawn from one set of variables can be seen in any other set, assuming the frameWindow is not excessively high to the point where minimal data is seen.

Given the outcome of the results and the data gathered, there are several recommendations to consider. The first would be to select a single bbThreshold. As apparent in the results, the bbThreshold does not seem to play a significant role in the final results, only slightly decreasing the the number of sequences seen at sufficiently high frameWindow values. As for the precise bbThreshold to use, the value should be low enough that two separate bounding boxes of the same object will not be considered the same bounding box if nearby one another. This value should also be high enough that objects nearer to the vehicle, which logically could change by more pixels per frame, would still be caught and considered as once bounding box. This value would be specific to each project dependent on the size of the bounding box images, but within the same environment used here, where the images were sized 486x365 pixels, a bbThreshold of 20 pixels is recommended.

As for the frameWindow, given that this variable has a great impact on the number of sequences seen, a range of values is preferable to a single value. As mentioned previously, it would be beneficial to have a lower maximum value than was used here, ideally setting the maximum as some value between 12 and 15 frames. As for the bottom of the range, we have a drastically increasing number of sequences seen at lower values for frameWindow, so a significantly low frameWindow minimum could prove beneficial. With that said, human vision requires cognition and setting a value too low may allow for the eye to gaze upon an object, but not allow the time for the brain to process this information, leaving the object seen by the eye but not truly seen by the driver. Thus, a minimum value between 3 and 5 frames should suffice. Based on the research reported here, one recommendation would be to further investigate frameWindows in the range of 3 frames to 12 frames (0.1s seconds to 0.4 seconds).

4.5 Limitations

As with any research, certain choices in approach must be made. For this specific research, some of the key decisions made were; how to implement object recognition, how bounding boxes would be determined, how driver's gaze would be determined, the definition of seen to be used, and the dataset used. All of these decisions could lead to different results and here we will discuss the limitations of the decisions made.

To start with, the decisions of how to implement object recognition and how bounding boxes would be determined faced the same logic. With a preceding colleague having done a lot of related work on the same datasets, the result was a neural network which could produce 96.1%, 96.2%, and 94.8% of correct classification for traffic signs, traffic lights, and vehicles, respectively. Given a strong performance coupled with the ease of access of utilizing a colleagues work for this step, this was the easy choice. As a result, Shirpour's work done in [22] was the clear choice to handle the object detection and bounding box determination. Because the work of a colleague was used, the exploration of object detection methods to be utilized was not as thorough since ease of access played a large role in this decision. It is quite possible that a higher-performing method could have been found and utilized assuming ease of access was not an issue.

The next decision to discuss has to do with driver's gaze. Again, Shirpour's work done in [22] provided us with a method for determining driver's gaze. The gaze method utilized was that of a Point of Gaze, or a singular point indicating where a driver is looking at any given moment. This might be the most limiting decision made as using a single point restricts the driver's ability to see. When a single point is used it limits the driver to only being able to see a single object per frame. The human eye, however, can see more than a single object at a time. It is also capable of seeing objects it is not directly looking at, through the use of peripheral vision. Having our driver's gaze represented as a PoG was the most limiting decision made.

Another decision which led to limitations was the definition of "seen" to be used. In the initial work, a few key variable values were selected to represent the whole of the data. The definition of seen would utilize a few combinations of these variables to view key values and would be defined using specific values. This was later changed to accommodate for the limitations of only looking at a few key values, which could lead to missed trends or other observations as the whole picture isn't there. It was later decided that using a definition of seen which is not restricted by specific values would prove more beneficial.

The final decision which led to some limitations was the dataset to be used. For the purposes of this research, the decision to use a single driving sequence, approximately 100,000 frames of data, was made. Given complications and time constraints, it was decided that one driving sequence would be sufficient. But with a limited dataset means limited ability to compare. In the first summary table, table 4.1, there are many entries which seem much lower than would logically be expected. But given the limited dataset chosen, it is not possible to compare against other driving sequences, which could ultimately lead to a more accurate hypothesis as to why.

Chapter 5

Conclusion

In this Chapter we present our conclusions, as well as possible future work based on the research presented. This Chapter is divided into two sections, Testing and Broad Scope. In the first section, Testing, we will look at future work that can be done to alter and/or improve upon the work already presented. In the following section, Broad Scope, we will look at work that can be done which does not require any alteration of the work already presented.

In this contribution we presented a novel concept for consideration in works involving the driver's observation of a driving scene. Rather than considering driver observation on a frame by frame basis, we examine a definition of seen which accounts for time (in frames) and note the impact on the objects which and how many are deemed as seen. The forefront observation made at the culmination of this work is that the overall percentage of objects of our specific categories seen by drivers when taking this approach is low, with no more than 15% of all frames having an object considered as seen by the driver. This indicates a just how much of an impact the length of gaze has on objects seen. This approach emphasizes the need to account for the part that time plays in human vision and recognition.

We utilized the works of Shirpour et al. [22] for our object detection. We then overlay PoG onto these frames with included bounding boxes. Next we consider where this PoG overlaps with bounding boxes to indicate if the driver has glanced at an object before finally considering if this gslance remains on the same bounding box over consecutive frames. We consider a range of inputs for both the length of gaze, and tolerance of bounding box movement, to establish various definitions of seen spanning both shorter and longer attention to objects.

5.1 Testing

Given time constraints and complications faced with utilizing aged code, there were areas of investigation which were not fully explored. There were three main areas in which variables could be altered to produce further results, the frame window (number of consecutive frames required for a bounding box to be considered seen), the bounding box threshold (number of pixels a bounding box could change by between frames to be considered the same bounding box), and the point of gaze (a single point in the frame the driver is considered to be gazing at).

The first variable to look at is the frame window. This variable acted as a part of the definition of ”seen” used in this work. This frame window was considered in the range of 3 to 30 consecutive frames, equating to 0.1 to 1 second of time. Although a frame window in some facet is required for any definition of seen, the range explored in these works was overextended and future works should consider a smaller range. As discussed previously, a maximum of 12-15 frames should be sufficient for this variable.

The next of these variables is the bounding box threshold used. In these works, a bounding box threshold (in pixels) was utilized to identify whether a bounding box was the same object or not between frames. Simply put, if a bounding box with the same label between frames did not move by more than the threshold number of pixels in either direction, then it would be considered as the same bounding box. There are two approaches in which this could be further expanded, still utilizing a threshold but under different conditions, or using a different approach to determining bounding box constancy altogether. For the former, further work may consider different ranges of pixels thresholds, which would also be required for different sized images, or probably more appropriate, we could consider switching the pixel-based measure to a percentage-based measure. For the latter option of using new approaches, one such method may be to investigate the area of object tracking algorithms which can track moving objects. The work done in [1] proposes a novel bounding box tracking approach which may act as a good starting point for this research.

The final item worth further examination is the utilization of a driver’s point of gaze. In this work a single point is found in each frame representing the driver’s gaze. This point is then used alongside the frame window and bounding box threshold to determine if an object will be considered as seen. If this point of gaze falls within the same bounding box for a predetermined number of consecutive frames then that object will be considered as seen. A further investigation into other other measures of driver gaze rather than a single point could prove useful. A single point of gaze leaves no indication of what could be seen in peripheral vision and also limits the driver to only seeing one object per frame. Considering a cone of vision or confidence interval representing the driver’s gaze could improve upon the results discussed and account for the two issues just mentioned.

5.2 Broad Scope

Outside of the work presented in this paper, there are other spaces worth exploring, the main one being the dataset used. The dataset used in this work was a subset of the RoadLab dataset presented in Section 3.1. This dataset contained driving sequences from 16 unique drivers each completing the same route. In this work we simply looked at one full driver’s data, one driver with approximately 100,000 frames present. The data explored could be expanded in one of two ways, ideally following one another.

The first step would be to expand the research with data which is already collected and compiled, as presently only 1 of the 16 driver sequences are utilized. This step is easy enough as the data has already been collected and exists and simply needs to be used. This data could then go through the same preparatory steps including passing through the neural net and having the bounding boxes determined and recorded.

The second step would be to expand the dataset with new data, which could be done in a few different ways. The first expansion of the dataset could be to expand to more drivers and compile more instances of the same driving sequences. The second expansion of the dataset would be to compile driving sequences under various different conditions such as weather or light levels. And the third expansion of the dataset would be to compile driving sequences that look at different routes and locations, allowing for different traffic objects to be present in differing numbers and proportions. Expanding the dataset in one or multiple of these fashions could help uncover a clearer understanding of objects seen and unseen by drivers.

5.3 Integration

Looking more at the out-of-lab advancement of this work, the integration of the work provided into a real vehicle could open up more opportunities for further research as well. The work presented here provides information that could prove useful in the development of or research into i-ADAS.

As is, this work provides information about which objects during the driving sequence are seen or unseen, by our definition. Beginning with the objects unseen, these objects should be considered high priority for an i-ADAS to alert the driver to, as these objects are likely to have been missed by drivers. This also applies to objects which were rarely seen or objects low percentages seen by drivers. Conversely, objects deemed as seen by drivers by require less attention the i-ADAS as alerting a driver to something which they have already seen is redundant and, in situations where fractional time is vital, could hurt the i-ADAS ability to assist in a detrimental way.

In i-ADAS vehicles, processing time can make or break the effectiveness of the system. Driving can turn fatal in a fraction of a second and there are many other driving events which may also occur in marginal time. As such, having an efficient and quick processing time is necessary. The results presented in this work indicate the effects that different sequence lengths (number of frames to process) have on objects seen or unseen, so taking another step to narrow down the possible sequence lengths and determining exactly what sequence length should be used that is long enough but not excessive is a viable step towards better processing performance. Additionally, the machine learning algorithms used in this work, a combination between a Faster R-CNN and a Multi-Scale HOG-SVM with a ResNet-101 network, performs well, but may not be the fastest possible high-performing algorithm. Since processing time is so important, it is worthy of investigation into other possible bounding box detection algorithms to step closer to integration into i-ADAS vehicles.

5.4 Recommendations

We have discussed many changes and directions that could be taken with regards to this work. Here I will outline what I would recommend as the logical next steps. Firstly, incorporating the ability for multiple objects to be seen per frame would be the best change, whether that be through a confidence interval, a cone of vision, or some other consideration of peripheral vision. The next important change to be made would be to consider an object tracking algorithm for the determination of whether an object gazed upon is the same object in consecutive frames, rather than a measure utilizing changes in the bounding box as an independent calculation. Finally, expand the dataset to include more drivers from the RoadLab project. I have already discussed the validity of this dataset, and there are still additional driving sequences which could be considered. Expanding the dataset from one sequence of approximately 100,000 frames, to several sequences each of approximately 100,000 frames is an easy step forward that could paint a broader, clearer picture.

Appendix A

Label	Total BB Count	Total BB Seen	Seq Count	Avg Seq Length	% of BBs	% of Total Frames
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	0	0	0	0	0
Bus	10184	1423	209	6.81	13.97	1.41
Car	251200	10806	1566	6.90	4.30	10.68
Construction	1265	8	1	8	0.63	<0.01
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	103	14	7.36	1.66	0.10
LaneTurnsRight	745	0	0	0	0	0
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	0	0	0	0	0
NoTurn	4713	14	2	7	0.30	0.01
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	0	0	0	0	0
Pedestrian	43756	124	18	6.89	0.28	0.12
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	0	0	0	0	0
Stop	48	0	0	0	0	0
TrafficLightAhead	659	0	0	0	0	0
TrafficLight_Green	19044	60	6	10	0.32	0.06
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	64	10	6.4	0.47	0.06
TrafficLight_Yellow	4621	17	2	8.5	0.37	0.02
Truck	20633	1041	155	6.72	5.05	1.03
WatchForPedestrian	1002	0	0	0	0	0
Yield	2009	0	0	0	0	0
Totals	**406645**	**13660**	**1983**	**N/A**	**N/A**	**13.5**

Table A.1: Summary table for frameWindow of 5 frames and bbThreshold of 10 pixels

Label	Total BB Count	Total BB Seen	Seq Count	Avg Seq Length	% of BBs	% of Total Frames
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	5	1	5	0.27	<0.01
Bus	10184	1386	177	7.83	13.61	1.37
Car	251200	11193	1362	8.22	4.46	11.06
Construction	1265	8	1	8	0.63	<0.01
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	97	13	7.46	1.56	0.10
LaneTurnsRight	745	0	0	0	0	0
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	10	1	10	0.46	<0.01
NoTurn	4713	6	1	6	0.13	<0.01
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	5	1	5	0.18	<0.01
Pedestrian	43756	153	19	8.05	0.35	0.15
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	5	1	5	0.62	<0.01
Stop	48	0	0	0	0	0
TrafficLightAhead	659	0	0	0	0	0
TrafficLight_Green	19044	33	4	8.25	0.17	0.03
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	59	9	6.56	0.43	0.06
TrafficLight_Yellow	4621	31	3	10.33	0.67	0.03
Truck	20633	1117	141	7.92	5.41	1.10
WatchForPedestrian	1002	5	1	5	0.50	<0.01
Yield	2009	0	0	0	0	0
Totals	**406645**	**14113**	**1735**	**N/A**	**N/A**	**13.97**

Table A.2: Summary table for frameWindow of 5 frames and bbThreshold of 20 pixels

Label	**Total BB Count**	**Total BB Seen**	**Seq Count**	**Avg Seq Length**	**% of BBs**	**% of Total Frames**
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	0	0	0	0	0
Bus	10184	1531	104	14.72	15.03	1.51
Car	251200	9668	635	15.22	3.89	9.55
Construction	1265	0	0	0	0	0
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	82	6	13.67	1.32	0.08
LaneTurnsRight	745	0	0	0	0	0
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	10	1	10	0.46	<0.01
NoTurn	4713	0	0	0	0	0
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	0	0	0	0	0
Pedestrian	43756	151	9	16.78	0.35	0.15
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	0	0	0	0	0
Stop	48	0	0	0	0	0
TrafficLightAhead	659	0	0	0	0	0
TrafficLight_Green	19044	16	1	16	0.08	0.02
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	21	2	10.5	0.15	0.02
TrafficLight_Yellow	4621	29	2	14.5	0.63	0.03
Truck	20633	882	58	15.21	4.27	0.87
WatchForPedestrian	1002	0	0	0	0	0
Yield	2009	0	0	0	0	0
Totals	**406645**	**12390**	**818**	**N/A**	**N/A**	**12.24**

Table A.3: Summary table for frameWindow of 10 frames and bbThreshold of 10 pixels

Label	Total BB Count	Total BB Seen	Seq Count	Avg Seq Length	% of BBs	% of Total Frames
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	0	0	0	0	0
Bus	10184	1512	90	16.8	14.85	1.49
Car	251200	9708	585	16.59	3.86	9.59
Construction	1265	0	0	0	0	0
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	83	5	16.6	1.34	0.08
LaneTurnsRight	745	0	0	0	0	0
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	10	1	10	0.46	<0.01
NoTurn	4713	0	0	0	0	0
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	0	0	0	0	0
Pedestrian	43756	147	6	24.5	0.34	0.15
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	0	0	0	0	0
Stop	48	0	0	0	0	0
TrafficLightAhead	659	0	0	0	0	0
TrafficLight_Green	19044	0	0	0	0	0
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	25	2	12.5	0.18	0.02
TrafficLight_Yellow	4621	25	2	12.5	0.54	0.02
Truck	20633	980	60	16.33	4.75	0.97
WatchForPedestrian	1002	0	0	0	0	0
Yield	2009	0	0	0	0	0
Totals	**406645**	**12490**	**751**	**N/A**	**N/A**	**12.33**

Table A.4: Summary table for frameWindow of 10 frames and bbThreshold of 30 pixels

Label	**Total BB Count**	**Total BB Seen**	**Seq Count**	**Avg Seq Length**	**% of BBs**	**% of Total Frames**
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	0	0	0	0	0
Bus	10184	1491	60	24.85	14.64	1.47
Car	251200	8531	328	26.01	3.40	8.43
Construction	1265	0	0	0	0	0
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	47	1	47	0.76	0.05
LaneTurnsRight	745	0	0	0	0	0
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	0	0	0	0	0
NoTurn	4713	0	0	0	0	0
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	0	0	0	0	0
Pedestrian	43756	45	2	22.5	0.10	0.04
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	0	0	0	0	0
Stop	48	0	0	0	0	0
TrafficLightAhead	659	0	0	0	0	0
TrafficLight_Green	19044	0	0	0	0	0
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	38	1	38	0.28	0.04
TrafficLight_Yellow	4621	0	0	0	0	0
Truck	20633	706	26	27.15	3.42	0.70
WatchForPedestrian	1002	0	0	0	0	0
Yield	2009	0	0	0	0	0
Totals	**406645**	**10858**	**418**	**N/A**	**N/A**	**10.73**

Table A.5: Summary table for frameWindow of 15 frames and bbThreshold of 10 pixels

Label	Total BB Count	Total BB Seen	Seq Count	Avg Seq Length	% of BBs	% of Total Frames
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	0	0	0	0	0
Bus	10184	1534	61	25.15	15.06	1.52
Car	251200	8197	311	26.36	3.26	8.10
Construction	1265	0	0	0	0	0
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	27	1	27	0.43	0.03
LaneTurnsRight	745	0	0	0	0	0
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	0	0	0	0	0
NoTurn	4713	0	0	0	0	0
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	0	0	0	0	0
Pedestrian	43756	17	1	17	0.04	0.02
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	0	0	0	0	0
Stop	48	0	0	0	0	0
TrafficLightAhead	659	0	0	0	0	0
TrafficLight_Green	19044	0	0	0	0	0
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	38	1	38	0.28	0.04
TrafficLight_Yellow	4621	0	0	0	0	0
Truck	20633	677	23	29.43	3.28	0.67
WatchForPedestrian	1002	0	0	0	0	0
Yield	2009	0	0	0	0	0
Totals	**406645**	**10490**	**398**	**N/A**	**N/A**	**10.38**

Table A.6: Summary table for frameWindow of 15 frames and bbThreshold of 30 pixels

Label	Total BB Count	Total BB Seen	Seq Count	Avg Seq Length	% of BBs	% of Total Frames
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	0	0	0	0	0
Bus	10184	1374	30	45.8	13.49	1.36
Car	251200	6395	127	50.35	2.55	6.32
Construction	1265	0	0	0	0	0
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	25	1	25	0.40	0.03
LaneTurnsRight	745	0	0	0	0	0
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	0	0	0	0	0
NoTurn	4713	0	0	0	0	0
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	0	0	0	0	0
Pedestrian	43756	0	0	0	0	0
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	0	0	0	0	0
Stop	48	0	0	0	0	0
TrafficLightAhead	659	0	0	0	0	0
TrafficLight_Green	19044	0	0	0	0	0
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	26	1	26	0.19	0.03
TrafficLight_Yellow	4621	0	0	0	0	0
Truck	20633	839	13	64.54	4.07	0.83
WatchForPedestrian	1002	0	0	0	0	0
Yield	2009	0	0	0	0	0
Totals	**406645**	**8659**	**172**	**N/A**	**N/A**	**8.57**

Table A.7: Summary table for frameWindow of 25 frames and bbThreshold of 10 pixels

Label	Total BB Count	Total BB Seen	Seq Count	Avg Seq Length	% of BBs	% of Total Frames
Arrow	397	0	0	0	0	0
Background	2950	0	0	0	0	0
BicycleLane	1870	0	0	0	0	0
Bus	10184	1353	32	42.28	13.29	1.34
Car	251200	6397	132	48.46	2.55	6.32
Construction	1265	0	0	0	0	0
DoNotEnter	920	0	0	0	0	0
ExitOnly	420	0	0	0	0	0
KeepToRight	6211	27	1	27	0.43	0.03
LaneTurnsRight	745	0	0	0	0	0
MaximumSpeedLimit	2141	0	0	0	0	0
NoTruck	2170	0	0	0	0	0
NoTurn	4713	0	0	0	0	0
NotAThroughStreet	483	0	0	0	0	0
Parking	2803	0	0	0	0	0
Pedestrian	43756	0	0	0	0	0
PedestrianCrossover	194	0	0	0	0	0
RailroadCrossing	98	0	0	0	0	0
RightLaneEndsAhead	802	0	0	0	0	0
Stop	48	0	0	0	0	0
TrafficLightAhead	659	0	0	0	0	0
TrafficLight_Green	19044	0	0	0	0	0
TrafficLight_Not Clear	11578	0	0	0	0	0
TrafficLight_Red	13729	48	1	48	0.35	0.05
TrafficLight_Yellow	4621	0	0	0	0	0
Truck	20633	728	12	60.67	3.53	0.72
WatchForPedestrian	1002	0	0	0	0	0
Yield	2009	0	0	0	0	0
Totals	**406645**	**8553**	**178**	**N/A**	**N/A**	**8.46**

Table A.8: Summary table for frameWindow of 25 frames and bbThreshold of 30 pixels

Bibliography

[1] J. An and E. Kim, ”Novel Vehicle Bounding Box Tracking Using a Low-End 3D Laser Scanner,” in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 6, pp. 3403-3419, June 2021, doi: 10.1109/TITS.2020.2994624.

[2] S. Baker, I. Matthews, J. Xiao, R. Gross, T. Ishikawa, and T. Kanade, ”Real-time non-rigid driver head tracking for driver mental state estimation,” CMU-RI-TR-04-10, Feb 2004.

[3] S. Banerjee, A. Joshi, J. Turcot, B. Reimer and T. Mishra, ”Driver Glance Classification In-the-wild: Towards Generalization Across Domains and Subjects,” 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), 2021, pp. 1-8, doi: 10.1109/FG52635.2021.9667084.

[4] S. Beauchemin, M. Bauer, D. Laurendeau, T. Kowsari, Ji Cho, M. Hunter, K.Charbonneau, and O.McCarthy ”Roadlab: An in-vehicle laboratory for developing cognitive cars.” Proc. 23rd Int. Conf. CAINE. 2010.

[5] S. Beauchemin, M. Bauer, T. Kowsari, and Ji Cho. Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality. IEEE Transactions on Instrumentation and Measurement, 61(2):391-401, 2011.

[6] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner. Three decades of driver assistance systems: Review and future perspectives. IEEE Intelligent Transportation Systems Magazine, 6(4):6-22, 2014.

[7] A. Bublea and C. D. Căleanu, ”Deep Learning based Eye Gaze Tracking for Automotive Applications: An Auto-Keras Approach,” 2020 International Symposium on Electronics and Telecommunications (ISETC), 2020, pp. 1-4, doi: 10.1109/ISETC50328.2020.9301091.

[8] N. Dalal and B. Triggs, ”Histograms of oriented gradients for human detection,” 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

[9] W. Dong, Z. Yang, W. Ling, Z. Yonghui, L. Ting, and Q. Xiaoliang, ”Research on vehicle detection algorithm based on convolutional neural network and combining color and depth images,” 2019 2nd International Conference on Information Systems and Computer Aided Education (ICISCAE), 2019, pp. 274-277, doi: 10.1109/ICISCAE48440.2019.221634.

[10] B. T. Dong and H. Y. Lin, ”An On-board Monitoring System for Driving Fatigue and Distraction Detection,” 2021 22nd IEEE International Conference on Industrial Technology (ICIT), 2021, pp. 850-855, doi: 10.1109/ICIT46573.2021.9453676.

[11] A. Doshi and M. Trivedi, ”Investigating the relationships between gaze patterns, dynamic vehicle surround analysis, and driver intentions,” 2009 IEEE Intelligent Vehicles Symposium, 2009, pp. 887-892, doi: 10.1109/IVS.2009.5164397.

[12] D. W. Hansen and Q. Ji, ”In the Eye of the Beholder: A Survey of Models for Eyes and Gaze,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 3, pp. 478-500, March 2010, doi: 10.1109/TPAMI.2009.30.

[13] Q. Hu, S. Paisitkriangkrai, C. Shen, A. van den Hengel, and F. Porikli, ”Fast Detection of Multiple Objects in Traffic Scenes With a Common Detection Framework,” in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 4, pp. 1002-1014, April 2016, doi: 10.1109/TITS.2015.2496795.

[14] T. Huang and R. Fu, ”Driver Distraction Detection Based on the True Driver’s Focus of Attention,” in IEEE Transactions on Intelligent Transportation Systems, doi: 10.1109/TITS.2022.3166208.

[15] W. Kuo and C. Lin, ”Two-stage road sign detection and recognition”, Proc. IEEE Int. Conf. Multimedia Expo., pp. 1427-1430, 2007.

[16] K. J. Lee et al., ”A 502-GOPS and 0.984-mW Dual-Mode Intelligent ADAS SoC With Real-Time Semiglobal Matching and Intention Prediction for Smart Automotive Black Box System,” in IEEE Journal of Solid-State Circuits, vol. 52, no. 1, pp. 139-150, Jan. 2017, doi: 10.1109/JSSC.2016.2617317.

[17] E. Murphy-Chutorian, A. Doshi, and M. M. Trivedi, ”Head Pose Estimation for Driver Assistance Systems: A Robust Algorithm and Experimental Evaluation,” 2007 IEEE Intelligent Transportation Systems Conference, 2007, pp. 709-714, doi: 10.1109/ITSC.2007.4357803.

[18] J. Nidamanuri, C. Nibhanupudi, R. Assfalg and H. Venkataraman, ”A Progressive Review - Emerging Technologies for ADAS Driven Solutions,” in IEEE Transactions on Intelligent Vehicles, doi: 10.1109/TIV.2021.3122898. 2021.

[19] A. Palazzi, D. Abati, s. Calderara, F. Solera and R. Cucchiara, ”Predicting the Driver’s Focus of Attention: The DR(eye)VE Project,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 7, pp. 1720-1733, 1 July 2019, doi: 10.1109/TPAMI.2018.2845370.

[20] S. Shabani, ”Assessment of Driver’s Attention to Traffic Signs through Analysis of Gaze and Driving Sequences” (2019). Electronic Thesis and Dissertation Repository. 6209.

[21] F. H. Shadeed and S. J. Wallaschek, ”Concept of an Intelligent Adaptive Vehicle Front-Lighting Assistance System,” 2007 IEEE Intelligent Vehicles Symposium, 2007, pp. 1118-1121, doi: 10.1109/IVS.2007.4290267.

[22] M. Shirpour, ”Predictive Model of Driver’s Eye Fixation for manoeuvre Prediction in the Design of Advanced Driving Assistance Systems” (2021). Electronic Thesis and Dissertation Repository. 7706.

[23] S. Sivaraman and M. M. Trivedi, ”Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis,” in IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 4, pp. 1773-1795, Dec. 2013, doi: 10.1109/TITS.2013.2266661.

[24] Y. Sugano, Y. Matsushita, and Y. Sato, ”Appearance-Based Gaze Estimation Using Visual Saliency,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 2, pp. 329-341, Feb. 2013, doi: 10.1109/TPAMI.2012.101.

[25] K. Suzuki and K. Yamada, ”Method for evaluating the collision mitigation ratio when using collision avoidance alarm at intersection: Effects of the driver’s mental situation on the integrated error of driver and system,” Proceedings of SICE Annual Conference 2010, 2010, pp. 1374-1379.

[26] R. Tang and Z. Jiang, ”Driver’s Perception Model in Driving Assist,” 2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C), 2020, pp. 237-240, doi: 10.1109/QRS-C51114.2020.00047.

[27] U.S. Department of Transportation, National Highway Traffic Safety Association, ”Critical reasons for crashes investigated in the national motor vehicle crash causation survey”, Traffic Safety Facts - Crash Stats, 2015.

[28] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, ”A robust, coarse-to-fine traffic sign detection method,” The 2013 International Joint Conference on Neural Networks (IJCNN), 2013, pp. 1-5, doi: 10.1109/IJCNN.2013.6706812.

[29] World Health Organization. Global status report on road safety 2015. World Health Organization, pages 7-12, 2015.

Curriculum Vitae

Name:	Kolby Sarson
Post-Secondary Education and Degrees:	University of Windsor Windsor, Ontario 2015 - 2020 BSc
Honours and Awards:	Director's Honour Roll 2015-2020
Related Work Experience:	Teaching Assistant University of Windsor 2016 - 2020 Graduate Assistant The University of Western Ontario 2020 - 2022 System Application Technologist Lambton Kent District School Board 2022