PEDESTRIAN SAFETY ANALYSIS USING COMPUTER VISION

FINAL PROJECT REPORT

by

Daniel Champlin, Lane Hanson, and Michael Lowry
University of Idaho

for

Center for Safety Equity in Transportation (CSET)
USDOT Tier 1 University Transportation Center
University of Alaska Fairbanks
ELIF Suite 240, 1764 Tanana Drive
Fairbanks, AK 99775-5910

In cooperation with U.S. Department of Transportation,
Research and Innovative Technology Administration (RITA)

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The Center for Safety Equity in Transportation, the U.S. Government and matching sponsor assume no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle	5. Report Date
Pedestrian Safety Analysis using Computer Vision	
	6. Performing Organization Code

7. Author(s) and Affiliations	8. Performing Organization Report No.
Daniel Champlin, University of Idaho Lane Hanson, University of Idaho Michael Lowry, University of Idaho	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Center for Safety Equity in Transportation ELIF Building Room 240, 1760 Tanana Drive Fairbanks, AK 99775-5910	11. Contract or Grant No.

12. Sponsoring Organization Name and Address	13. Type of Report and Period Covered
United States Department of Transportation Research and Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590	14. Sponsoring Agency Code

15. Supplementary Notes

Report uploaded to:

16. Abstract

The goal of this research project was to explore the capabilities of computer vision for pedestrian safety analysis. Computer vision, an AI application in image processing, tracks the movement of cars, bikes, and pedestrians, offering superior information about speed, trajectory, and count data for various transportation modes. The University of Idaho acquired two computer vision sensors from the startup Numina. This project funded the installation and one year of data access. In collaboration with the City of Moscow, we identified a test location, but the sensors failed to provide the necessary data. Consequently, we pivoted to the open-source computer vision package YOLOv8. Our research then focused on YOLOv8's capabilities for pedestrian safety analysis. The first task tested detection accuracy, and the second compared different model sizes, or "brain sizes," of YOLOv8, which range from smaller, faster models to larger, more accurate ones. Accuracy tests compared average detection confidence across various zones and times of day, revealing that cars in high daylight had the highest confidence levels, while objects closer to the camera and oriented perpendicularly were detected more accurately. In contrast, objects at skewed angles and farther distances had lower confidence levels. The model size comparison showed that larger models, despite requiring more time and storage, produced significantly higher-quality detections.

17. Key Words	18. Distribution Statement	
Bicycle counts, Pedestrian counts		

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified.	Unclassified.	26	N/A

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized.**

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
	NOTE: volumes greater than 1000 L shall be shown in m^3			
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	$candela/m^2$	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	$candela/m^2$	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

The goal of this research project was to explore the capabilities of computer vision to analyze pedestrian safety. Computer vision, an application of Artificial Intelligence (AI) in image processing, is used to track the movement of cars, bikes, and pedestrians. Many companies have developed proprietary computer vision sensors for traffic monitoring because this new technology provides superior information about speed, trajectory, and count data for various modes of transportation.

The University of Idaho acquired two computer vision sensors from a startup company called Numina. This research project funded the installation of the sensors and one year of data access. Collaborating with the City of Moscow, we identified a location to evaluate the sensors. However, the sensors failed to provide the necessary data. Consequently, we pivoted our research to explore the capabilities of a new open-source computer vision package called YOLOv8. This Python package, often described as the "ChatGPT of computer vision," democratizes the power of AI by making advanced image processing accessible to everyone.

We re-focused our research effort to explore the capabilities of YOLOv8 for pedestrian safety analysis. Our first task involved testing detection accuracy. The second task was to compare different model sizes, or the artificial intelligence "brain sizes," offered by YOLOv8. This package provides five model sizes, with larger models producing better detection accuracy but requiring longer analysis times.

Our test of accuracy involved comparing average detection confidence across different zones in the field of view and at various times of day (lighting variation). The results revealed that cars, particularly during high daylight hours, exhibited the highest confidence levels. Furthermore, the results reveal that objects closer to the camera and oriented perpendicularly achieved the highest confidence levels. Conversely, objects at skewed angles and farther away exhibited consistently lower confidence levels.

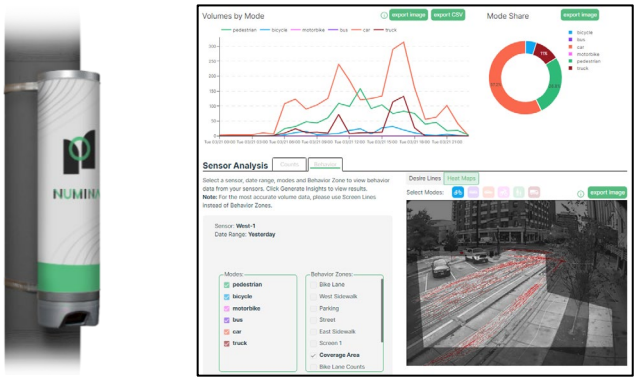
Our comparison of model size involved running the same video clip under different model sizes. The findings indicate that although larger brain sizes require more time and storage for data analysis, they produce significantly high-quality detection.

CHAPTER 1. INTRODUCTION

The goal of this research project was to explore the capabilities of computer vision for pedestrian safety analysis. Computer vision is the application of Artificial Intelligence (AI) to image processing to track the movement of cars, bikes, and pedestrians. Numerous companies have developed proprietary computer vision sensors for traffic monitoring. The technology provides several advantages over traditional methods, including robust information about speed, trajectory, and count data for various mode types.

The University of Idaho acquired two computer vision sensors from a startup company called Numina. Figure 1.1 shows a Numina sensor and the online dashboard to access the data. The person using the dashboard draws "zones" over a still image and selects the periods and desired modes for analysis (including pedestrian, bicycle, car, and truck). The output is charts, tables, heatmaps, and CSV files.

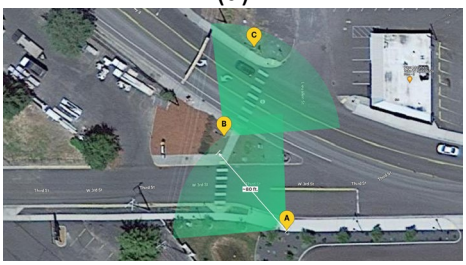
This research project funded the installation of two sensors and one year of data access. We worked with the City of Moscow to identify a location to investigate new techniques for pedestrian safety analysis. The City of Moscow identified a particularly hazardous location along State Route 8, shown in Figure 1.2, where university students cross the busy highway to get to campus.



(a) (b)

Figure 1.1 Numina (a) sensor and (b) dashboard output.

(a)



(b)

Figure 1.2 Sensor installation location (a) street view and (b) aerial view.

We intended to use the data from the Numina sensors to investigate new techniques for pedestrian safety analysis. Unfortunately, the sensors have failed to provide data. Consequently, we pivoted our research effort to explore the capabilities of a new open-source (i.e., free and modifiable) computer vision package called YOLOv8. This Python package has been described as the "ChatGPT of computer vision" because it gives everyone the power of AI (Jocher et al., 2023; Terven et al., 2023).

Our new research objective was to explore the capabilities of YOLOv8 for pedestrian safety analysis. Our first task was to evaluate detection accuracy. Chapter 2 describes the method and results for detection confidence analysis. Our second task was to compare different model sizes, i.e., the artificial intelligence "brain size." YOLOv8 offers five different model sizes. A larger model (i.e., a larger brain) will produce better detection but will take much longer to run the analysis. Chapter 3 describes the analysis we did to compare different model sizes.

CHAPTER 2. DETECTION CONFIDENCE ANALYSIS

2.1. Background

This chapter describes our analysis of detection confidence. We conducted the analysis using YouTube livestreams. For each object that is detected, YOLOv8 produces a unique identification number, determines the class (e.g., car, truck, bicycle), and displays a decimal value for the confidence level for the class determination. Figure 2.1 shows a detection example.

Figure 2.1 Example detection for a YouTube livestream in Orange, CA.

This research aimed to obtain an understanding of the program's confidence capability. We sought to evaluate confidence under different factors that alter confidence levels.

2.2. Method

The analysis was done for live YouTube traffic video feeds for Orange, California, and Derry, New Hampshire. We used the small model (i.e., artificial intelligence brain size) and the Open VINO optimizer. This allows the analysis of the videos to be completed faster but lowers the computing power.

We conducted two tests:

1. Confidence level for different hours of the day.
2. Confidence level for different zones in the field of view.

For test 1, we observed at 8:00 am, 12:00 pm, 5:00 pm, 8:00 pm and 10:00 pm PST. The analysis was completed over a 5-minute testing period for each time of day. We observed data and compared the results throughout the day to evaluate our hypotheses: that the confidence would be highest during daytime hours. Confidence will be the worst during dawn, dusk, and night. In addition, we would hypothesize that cars will have the highest confidence of all classes.

For test 2, we divided the field of view into different zones. We compared the confidence of trucks and cars passing through each zone. These trials were done over a 60-second recording period. Our hypothesis was that zones further away from the camera or at obscure angles will have lower confidence levels.

2.3. Results and Discussion

2.3.1. Confidence Throughout the Day

Table 2.1 shows the average confidence for different object classes at different hours of the day. We found that cars during 12:00 pm had the highest confidence level out of all classes and times. The class of cars always held the highest confidence level. In contrast, buses during 10:00 pm had the lowest confidence level. Motorcycles held the lowest average confidence out of all classes. When looking at some of the lower percentages, such as most of the motorcycle identifications or buses at 10:00 pm, it can be observed that these are errors in the program.

Table 2.1 Average Confidence at Different Hours of the Day

Time of Day	Car	Truck	Person	Bus	Motorcycle
8:00 am	65%	51%	34%	53%	35%
12:00 pm	67%	41%	36%	48%	28%
5:00 pm	64%	58%	40%	44%	29%
8:00 pm	61%	33%	38%	37%	26%
10:00 pm	45%	41%	30%	12%	21%
Average	60%	45%	36%	39%	28%

Errors were seen throughout the trials. Occasionally, inanimate objects would be misidentified. Even during the daytime, a sign was detected as a person, and a car was detected as a motorcycle for a split second, as seen in Figure 2.2. This is what happened with most motorcycle identifications. Cars that were partially off the side of the view or remarkably close to the edge were mistaken as motorcycles. Moreover, as can be seen, confidence for this motorcycle identification is low at 31%.

Figure 2.2 Example misclassified car as truck and motorcycle.

As lighting began to decrease, the error of inanimate object classification became more common. As seen in Figure 2.3, during a 10:00 pm recording session, the program begins to mistake many signs and posts for cars and people. This leads to low confidence levels at 10:00 pm compared to other times.

Figure 2.3 Example of misclassification of inanimate objects during low lighting.

Another common occurrence during these trials was the shifting classification of specific vehicles. Seen in Figure 2.4 is an example of a singular identification changing classification. This is partially what contributes to such low rates for buses. Usually, a bus classification occurs in situations such as in Figure 24. However, for the bus trial at 8:00 am, there was a bus in the selected video frames, leading to a higher confidence level.

Figure 2.4 Example of an object changing classification from bus to truck to car.

2.3.2. Confidence Within Varying Zones of the Camera

Figure 2.5 shows the zones for Derry, NH. The results are in Table 2.2. Zone #4 has the highest confidence of the six zone options. As expected, Zones #1, #2, and #3 all hold lower confidence levels at around 50%. These zones are all located further than the zones that performed better. In #1 and #3, cars travel dramatically away or from the camera, creating harsh angles and difficult confidence identification. This concurs with our hypothesis as the zone is remarkably close to the camera and gives a good side profile to both cars and trucks. Zones #5 and #6 follow similar reasoning as to why they hold a higher average confidence than the other three zones. Even though Zone #6 is at a skew angle, it is close enough to the camera location to perform at a higher level.

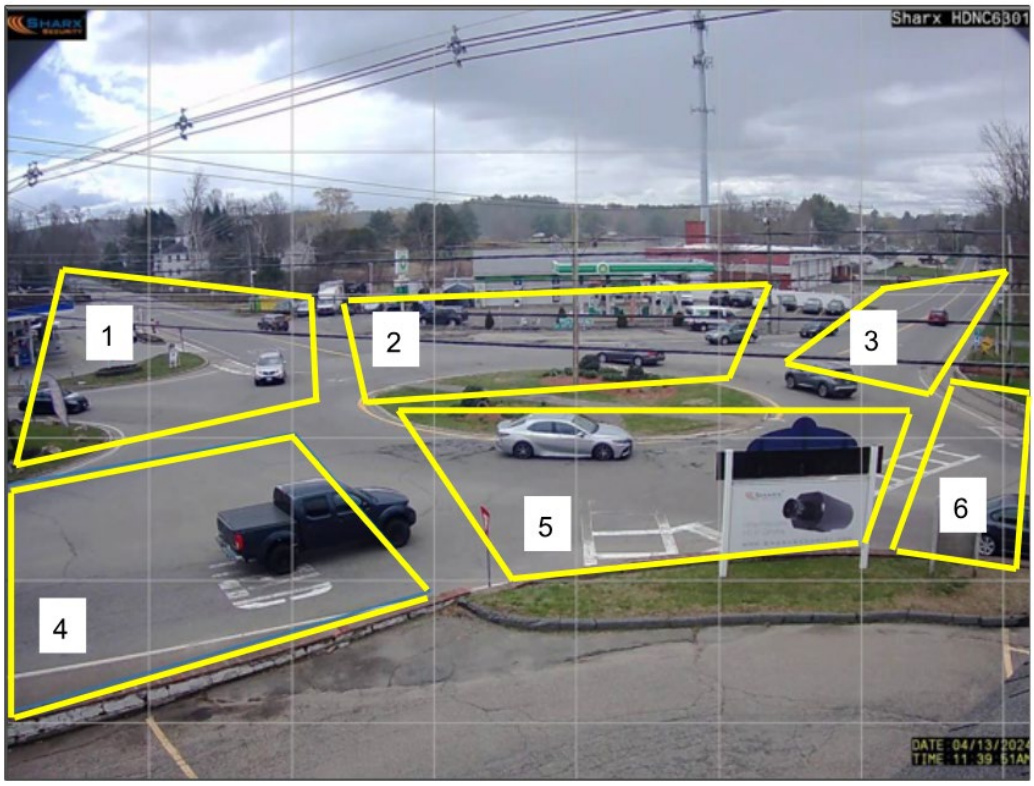


Figure 2.5 Field of view zones for Derry, NH.

Table 2.2 Average Confidence by Zone for Derry, NH

Zone	Distance	Angle	Car	Truck
1	Far	Skewed	57%	34%
2	Far	Perpendicular	55%	29%
3	Far	Skewed	51%	23%
4	Near	Perpendicular	84%	58%
5	Near	Perpendicular	73%	50%
6	Near	Skewed	72%	41%

Figure 2.6 shows the zones for Orange, CA. The results for Orange, CA are like Derry, NH. The highest-performing zones are those in closest proximity at perpendicular angles. In this case, those Zones were #2, #3, and #4. Zone 3 was the best-performing zone and is the closest and most perpendicular of all five options.

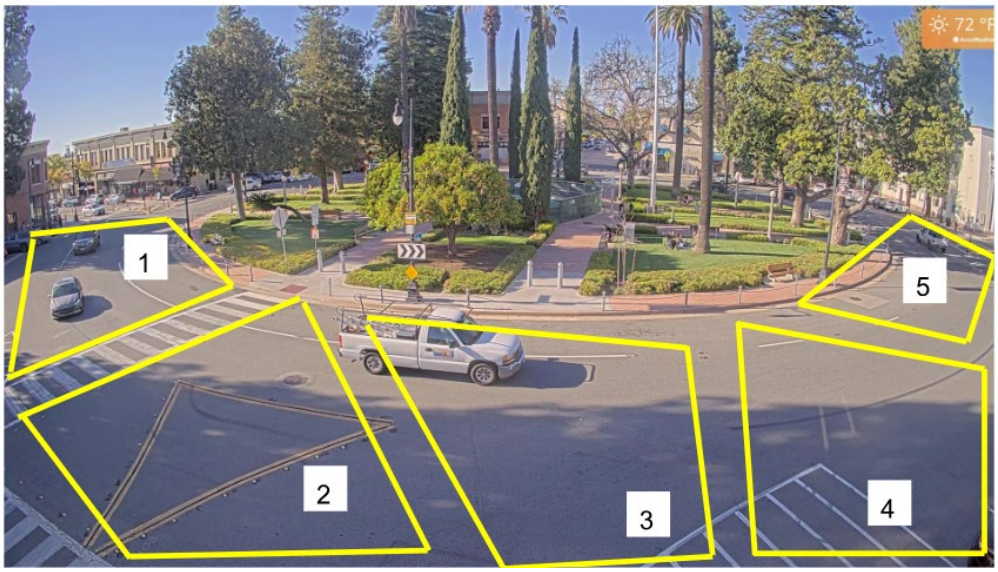


Figure 2.6 Field of view zones for Orange, CA.

Table 2.3 Average Confidence by Zone for Orange, CA

Zone	Distance	Angle	Car	Truck
1	Far	Skewed	67%	35%
2	Near	Perpendicular	75%	28%
3	Near	Perpendicular	87%	74%
4	Near	Perpendicular	87%	69%
5	Far	Skewed	58%	n/a

Matching our hypotheses again, Zones #1 and #5 had the lowest average confidence of the five options. These two zones are skewed and far away from the camera. Zone #5 could not detect any trucks due to its harsh angle. Figure 2.7 shows an example of a truck beginning to turn and in the first frame, the program begins to classify the vehicle as a car instead. This is due to the skew angle of the truck. While confidence is high in this frame, it begins to decrease as the car gets further and more skewed from the camera.

Figure 2.7 Example misclassification but with increased confidence.

2.4. Summary of Confidence Analysis

In summary, two analyses were completed on object detection confidence levels through the computer vision program. First, we compared different times of day with unique classifications to determine how they were altered as the day progressed. We then looked at different zones within two camera views to identify how confidence varied in different camera sections.

The first analysis proved that cars, particularly during high daylight hours, output the highest confidence levels of observed classes. In addition, we learned about errors in misidentification that arise mainly at night, leading to lower confidence levels.

For the second analysis, we created multiple zones within the set camera view to determine how the camera's perspective leads to confidence differences. From analyzing cameras in Derry, New Hampshire, and Orange, California, we concluded that objects closer to the camera at perpendicular orientation performed best. Objects at a skew angle and further away were found to have lower confidence levels across the board.

CHAPTER 3. MODEL SIZE COMPARISON

3.1. Background

This chapter describes our analysis of model comparison. The YOLOv8 object detection model has five different processing sizes, also known as brain size. These brain sizes range from nano, small, medium, large, and extra-large. The smallest brain size, nano, can rapidly analyze footage for object detection. The tradeoff, however, is the accuracy of the model and often, the results can be misleading due to the inaccuracies observed. Moving to the other end of the brain sizes, the extra-large brain size has increased accuracy, and experimental results can be drawn with much higher confidence. The challenge with the larger brain sizes, however, is the time that the larger brain sizes take to analyze each video and the storage required for the results. As accuracy increases, more storage is required.

The video each brain will analyze is almost nine minutes long and comes from an intersection in Tokyo, Japan, where several traffic cycles are observed. The video was recorded on April 29th, 2024, at 4 pm, just before rush hour.

Figure 3.1 shows the intersection where the video was recorded and analyzed. This intersection has specific characteristics that make it desirable for traffic identification and analysis. The volume of traffic and pedestrians and the diversity of objects to test are also high, given the location, meaning there is quite a bit of data to analyze that will aid in gathering accurate results. One of the most critical parts of this intersection is explicitly the camera angle.

Figure 3.1 Aerial view of case study intersection in Tokyo, Japan.

3.2. Method

This chapter explores the detection capabilities in identifying traffic objects using each brain size and draws conclusions from results from several experiments. The following experiments will consider the total volume of objects detected and the confidence percentage of detecting objects very heavily and use these as a metric to compare the performance of each brain size. Four experiments were conducted to test each brain's limits and capabilities, including the total count by an object in the intersection, the percent confidence of object detection, zoned detection confidence, and zoned pedestrian walking speed. The analysis does not include the large brain size due to an error in data reduction. The figures and tables of this report subsequently provide data on the other four previously mentioned brain sizes.

A considerable portion of the project is concerned with pedestrian movement within the intersection, namely pedestrian walking speed. Several zones within the intersection were created around the crosswalks to capture the necessary data for pedestrian walking speeds. These zones were made using a specific Python script that ran the video and created a resource image that could then be edited by drawing lines within the image. The figure below shows the zones of interest in this intersection. It should be noted that each zone is at a different distance. Due to the camera angle, each zone is increasingly distorted as the distance from the camera increases, which can cause additional errors when analyzing the video. Figure 3.2 outlines the zones used for the following experiments. The wide camera angle allows for a view of all four intersections and can track the paths of vehicles well, given the view of the camera. Zone 1 is the largest zone with very little distortion. The other zones are named, moving clockwise following zone 1, meaning zone 2 is farthest to the left of the image.

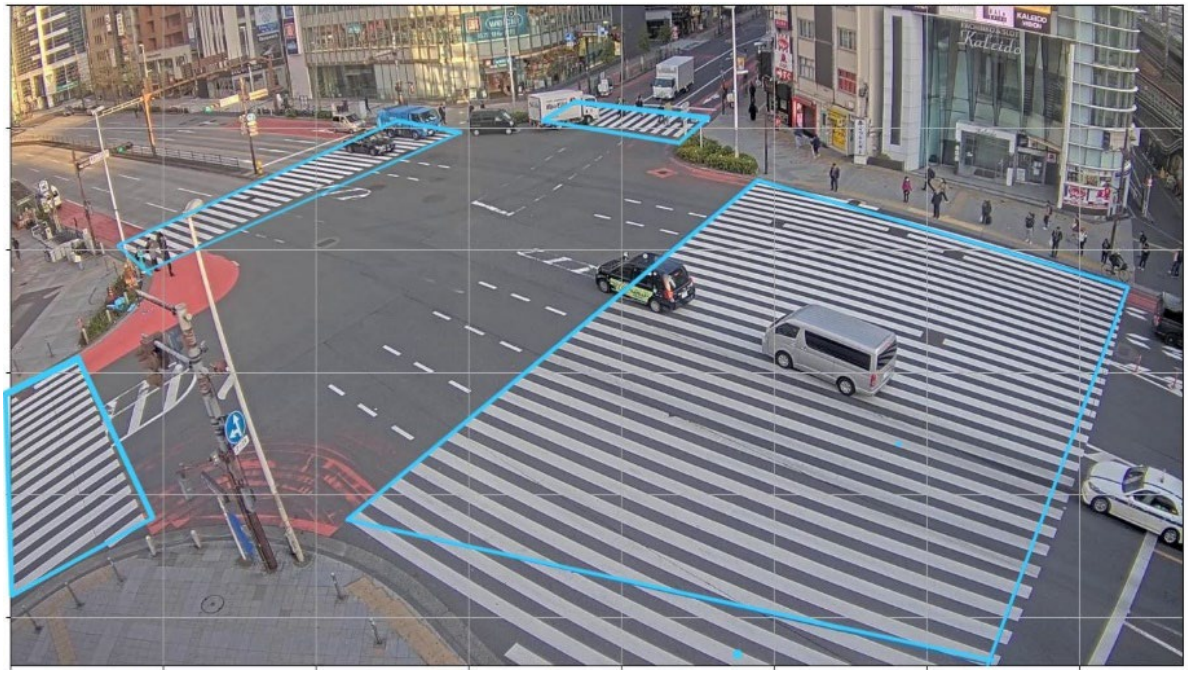


Figure 3.2 Analysis zones for Tokyo, Japan.

The first experiment was conducted to determine the total volume of objects each brain could detect within the duration of the video. The first step in gathering the necessary information was to record the

video using a Python script with the YOLO V8 package, in which the small brain size was first used to analyze the contents of the video. Once the video was recorded, it was run thrice through an analyzer Python script to allow the following brains to analyze it. Each brain then stored the recorded video and data in a resource folder, where the data was exported to a CSV file.

An individual CSV file was created for every brain size, and data were recorded by the frame with a time interval of 0.03 seconds. Since each object is in the video for a certain period that far exceeds the time interval, thousands of data points were tabulated, presenting a minor challenge in sorting the appropriate data as each object was listed hundreds of times. Microsoft Excel was used to sort out the relevant data and eliminate duplicates so that only the individual objects were listed. Then, the total count of cars, bikes, pedestrians, etc. could be determined by summing each object by class type. This same process was repeated for each brain size in a CSV file that corresponded to the individual brain size. This experiment is independent of zones.

The second experiment concerns how the confidence of object detection is affected overall for the intersection by object type. Like the previous experiment, this experiment is independent of the zones created and focuses on the entire intersection. The data that was analyzed for this experiment is the raw data from the original CSV file created in the intersection count experiment. This data can be analyzed simply by creating a pivot table in Microsoft Excel. The key components of this experiment are the brain sizes compared to the overall average percent confidence for each object as it enters the video.

Another essential factor to consider when testing the brain size's performance is the video's distance and camera angle. To test this, the zone counts, and confidence percentages will be analyzed as the zones all have differing distances and distortions as they move further away from the camera. Certain data will be impossible to obtain given the remote collection, which includes the distance from the camera and the camera angle in relation to each zone. It should be apparent that to limit distortion, the best angle for the camera is 90 degrees or straight above the intersection to track objects accurately. The camera in this experiment is not 90 degrees to any of the zones, and this will play a critical role in how the brain detects objects. The data for this experiment will also come from the CSV file. It includes the total count by zone will be compared to each zone by brain size. The corresponding brain size will track the average percent confidence of each zone to conclude how both distance and camera angle affect detection capabilities.

The final performance experiment will attempt to estimate the walking speeds of detected pedestrians within the first crosswalk zone, as this zone yielded the highest counts of pedestrians throughout the video analysis. To calculate the pedestrian walking speed in ft/s, the raw data from video analysis was tabulated in the original CSV file for each brain size. A column of data recording run time was recorded for each file that pertains explicitly to this experiment. By filtering the data only to show pedestrians, that data can be reduced to determine the walking speed of pedestrians. This column tracks the frame of when each object is inside of the zone. Each frame is 0.03 seconds long. The summation of this data is the total time each pedestrian was in the crosswalk. For this experiment, this is regarded as a pedestrian walking time to cross the first zone. Knowing that the zone is 110 feet long, an estimated walking speed can be determined for the pedestrians when entering the zone by dividing distance by time. Doing this for each pedestrian detected should determine an average value for the desired crosswalk.

3.3. Results and Discussion

3.3.1. Intersection Object Count

The first experiment resulted in data that was to be expected. It is sensible to think the larger brain size would pick up more objects within the intersection. Table 3.1 shows all the values that each brain size could identify. Each class name clearly shows that with increasing brain size, increased objects of each class were detected. One of the categories that stand out is the pedestrian count. The data shows a considerable margin, with every other brain size outperforming the nano brain size. The nano brain size also indicates detecting a train within the intersection. Given the intersection's infrastructure, it should be well understood that this is not possible, which raises concerns that the nano brain is not only detecting a few objects but is quite possibly misidentifying objects entirely. For instance, the data trend indicates that the small brain size should find more cars than the nano brain. This is not the case with the data provided. Is this because the nano brain failed to identify the correct objects and mislabeled those objects as cars?

Table 3.1 Object Count by Brain Size

Class Name	Nano	Small	Medium	XL
person	36	91	316	676
bicycle	0	1	0	6
bus	7	13	17	17
car	266	250	403	413
motorcycle	2	4	8	16
Truck	16	11	23	31
Train	1	0	0	0
Total	328	370	767	1159

Compared to other brain sizes, only the nano brain detected what it thought was a train. This calls into question the accuracy of nearly all the subsequent results of the nano brain in this report and provides a strong argument that although this brain size is very fast, it provides untrustworthy data that can seldom be used for critical traffic data analysis. Speed is a significant factor in all data analysis. However, when it comes to the cost of accuracy to this extent, it is a compelling argument that larger brain sizes should always be used to provide reliable results.

3.3.2. Object Confidence by Brain Size

Similarly to the experiment's results above, the confidence on average should be expected to be higher with larger brain sizes. Figure 5 below shows how each brain size compares to one another for each object. As expected, there appears to be a linear fit that increases for many of the trendlines linking object confidence with brain size. One exciting piece from Figure 3.4 is how the nano brain size outperforms the small brain size on multiple occasions. For most data points, it should be expected that increasing brain size will always have a higher average confidence percentage. The data raises an interesting discussion on whether detecting fewer objects leads to potentially higher confidence

percentages in certain conditions. This means that although the nano brain sees fewer objects overall, when it sees a certain object - mostly buses and bicycles- it is more confident of the detection status. Although this is a direct contradiction to the data of the previous experiment, it is something to consider moving forward. The YOLO V8 software allows for faster detection of certain objects that are easily distinguishable, such as bikes and buses. However, this argument loses most traction when compared to the medium and extra-large brains, where the nano brain outperforms every time, and it provides even more of an argument to use the larger brain sizes.

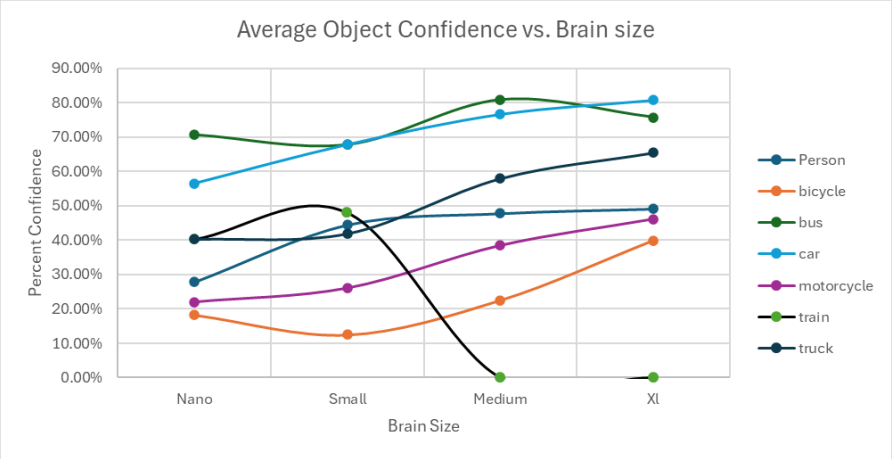


Figure 3.3 Average confidence by brain size.

One last thing to notice about the data is train confidence. Both the nano and small brain were confident that a train was detected. From the previous experiment, this is not possible. This again calls into question the reliability of the data from the smaller brain sizes. It is also very peculiar that the small brain did not record seeing a train from the first experiment yet has a considerable confidence percentage of train detection, a noteworthy discrepancy from which it is hard to conclude. The medium and XL brains have a zero-confidence percentage on seeing trains, once more adding significant support that these brain sizes produce more reliable data.

3.3.3. Zone Detection Confidence and Count

Like the prior experiments, the trend of larger brain sizes and more accurate results is clear (See Figure 3.5). The first zone has an extraordinarily strong correlation of increasing count from nano to XL which should be expected. Then, when considering the rest of the zones, the data drops off severely. This is certainly due to each zone's distance and camera angle. Since the first zone is so close to the camera and much larger, it is highly likely that each brain size, although identifying different numbers of objects, is more capable of picking out objects in the first zone. Compared to the first zone, all other zones performed similarly regardless of brain size. This is the first data from this report that suggests there is a factor that contributes to the decline of capabilities across brain sizes. Considering the size, camera angle, and distance from the camera, the evidence suggests that these factors play a critical role in object detection of any brain size and will provide a considerable challenge to experiments moving forward.

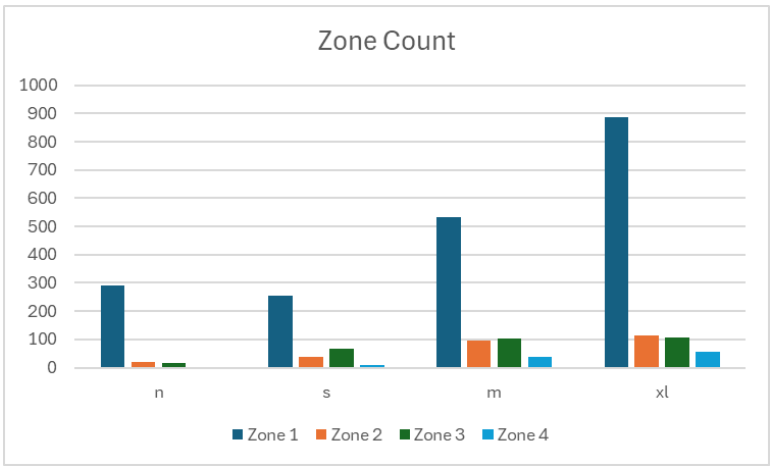


Figure 3.4 Zone count by brain size.

Figure 3.6 shows that the nano and small brain have a decline in confidence as the zones move farther away from the camera with an increasingly narrow camera angle. This would support the conclusion that the distance and angle affect the data, assuming confidence and count go hand in hand when analyzing the data. However, the medium and extra-large brain display interesting data in Figure 7 that do not necessarily support what the nano and small brain sizes do. Recall that in 4.3.2, a potential argument was made that as the brains detect fewer objects, they do so with a higher confidence percentage. It is tempting to make this assessment using what is shown in Figure 3.6, which shows brain sizes medium and XL. However, an essential factor to remember is that these zones simply do not see the same rate of objects. As a result, these zones cannot provide accurate average confidence percentages as the data for these zones is too few. This assessment aligns more with what is observed from the previous experiments. Further, it supports the conclusions of accuracy by average confidence percentage due to angle and distance and increasing brain size dominance, knowing that smaller brain size results are often hard to interpret.

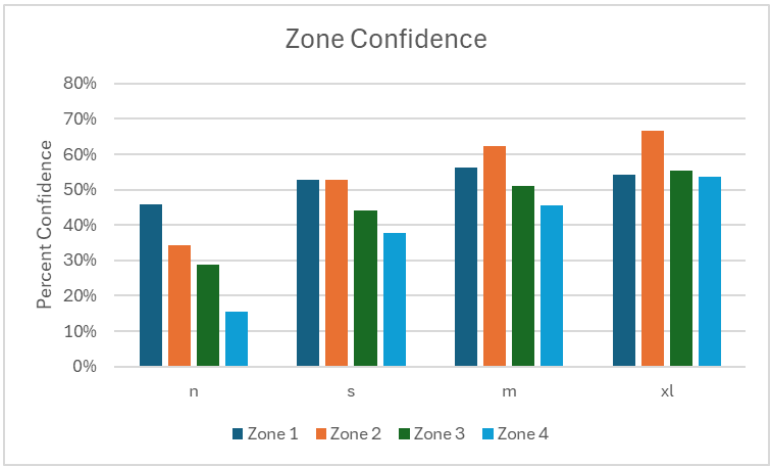


Figure 3.5 Zone confidence by brain size.

3.3.4. Pedestrian Walking Speed Using Zones

The final experiment posed concerning results regarding the average walking speed and raised concerns about the capabilities of YOLO V8 for estimating pedestrian walking speeds. The data from Table 3.2 is very puzzling. All brain sizes have similar walking speeds. The issue is that typical walking speeds are not anywhere near the proposed values above. Typical average freeway speeds are about 70 mph or 102.9 ft/s. The small brain indicates that pedestrians are walking at 90 mph. Knowing that pedestrian walking speeds are far less than estimated, the error comes from within the method used to calculate walking speeds given the YOLO V8 model. Consider Figure 3.7. This is the XL brain analyzing a snapshot of the video. Object detection in Zone 1 works well, with a certain number of pedestrians not being detected.

Table 3.2 Walking Speed by Brain Size

Brain Size	Walking Speed (ft/s)
Nano	133
Small	115
Medium	111
Extra Large	138

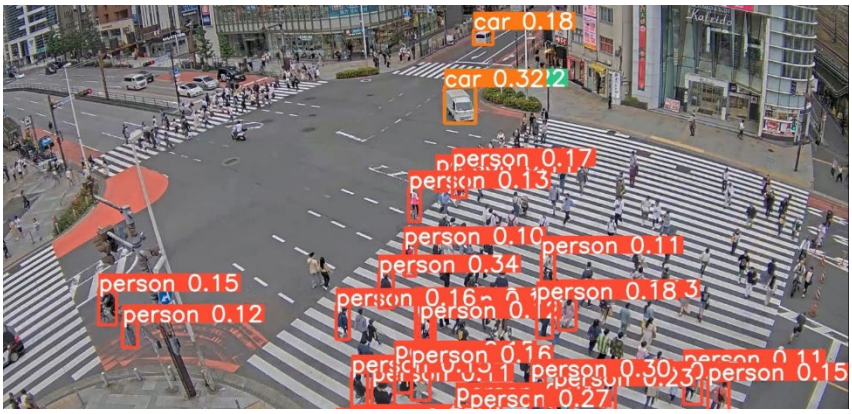


Figure 3.6 Example of a high level of pedestrian detection.

Now consider Figure 3.8. Not a single pedestrian is being detected within Zone 1. This snapshot comes just a few seconds after the snapshot of Figure 8. Recall the method in section 4.2.4. This method of pedestrian walking speeds assumes that each pedestrian has crossed the street when their respective run time has finished. Based on what is displayed in Figures 8 and 9, this obviously is not what happens. When the ID tag is gone, the object is no longer detected, and the run time ends. Assuming the last run time value for each pedestrian is the max value of our method, the pedestrian has now crossed the street, and the walking speed can be calculated. Now, considering that these ID tags come and go at a random phenomenon, the denominator of the simple equation used to calculate speed becomes unproportionally low, causing the speed to be very high. After analyzing all data for each brain, this was, unfortunately, the case for every brain size in all collected data. Had the assumption been correct that pedestrians cross the street when the running time is at a maximum value, this method would have worked with few issues.

Figure 3.7 Example of no pedestrian detection.

To solve this problem, a deeper understanding of YOLO V8 must be obtained to determine why the model stops identifying objects randomly. One potential source of the drop in identification could be the rates of occlusion observed at the crosswalk. Occlusion is the principle that an object being tracked becomes covered up by a different object, tracked or not, and hides it from view. As these rates of occlusion increase amongst pedestrians in the zone, there could be a correlated rise in the calculated walking speed using this model. To accurately determine if this is a sufficient explanation, an independent study must be conducted to correlate occlusion to both zone detection and the proposed walking speed model.

3.4. Summary of Model Size Comparison

The purpose of this chapter is to provide insight into the capabilities and performance of the YOLO V8 detection software through a range of correlated traffic experiments and to draw conclusions on how AI detection can be used moving forward in engineering applications within the field of transportation and outside of this field as well. By using a range of performance metrics such as average confidence and detection count inside and outside of specified zones, the capabilities of the brain sizes of YOLO V8 were compared to address the tradeoffs behind the accuracy of results when using these different models. The conclusions of this chapter present a strong argument that although the larger brain sizes take a considerable amount of time and storage to analyze data, the data produced is of high quality and can be used in confidence when analyzing characteristics of a desired intersection such as object count or average detection confidence. More research will be needed to fully understand the relationship of the detection model to accurately pair the brain sizes with specific characteristics such as camera angle and object distance and mitigate occlusion to aid in more accurate results across all YOLO V8 applications.

CHAPTER 4. CONCLUSION

Two analyses were conducted on object detection confidence levels using a computer vision program. First, we examined how various times of day affected unique classifications to see how confidence levels changed as the day progressed. Then, we analyzed different zones within two camera views to identify variations in confidence across different sections of the cameras' fields of view.

The first analysis revealed that cars, particularly during high daylight hours, exhibited the highest confidence levels among observed classes. Additionally, we found that misidentification errors, primarily occurring at night, led to lower confidence levels.

In the second analysis, we divided the camera views into multiple zones to assess how the camera's perspective influenced confidence levels. By analyzing footage from cameras in Derry, New Hampshire, and Orange, California, we concluded that objects closer to the camera and oriented perpendicularly achieved the highest confidence levels. Conversely, objects at skewed angles and farther away exhibited consistently lower confidence levels.

The purpose of this study was to provide insights into the capabilities and performance of YOLO V8 detection software through a series of correlated traffic experiments. By utilizing performance metrics such as average confidence and detection count inside and outside specified zones, the report compares different versions of YOLO V8 to address the trade-offs in accuracy associated with these models.

The findings indicate that although larger models require more time and storage for data analysis, they produce high-quality data that can be confidently used for analyzing intersection characteristics, such as object count and average detection confidence. Further research is necessary to fully understand how to pair model sizes with specific characteristics like camera angle and object distance, and to mitigate occlusion, thereby improving accuracy across all YOLO V8 applications.

REFERENCES

Jocher, G., Chaurasia, A., & Qiu, J. YOLO by Ultralytics (Version 8.0.0). 2023, https://github.com/ultralytics/ultralytics

Terven, J., & Cordova-Esparza, D. (2023). *A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond*. arXiv preprint arXiv:2304.00501.