Night Light

Santos Volpe SCOPE Final Report

A proof-of-concept for novel data collection and analysis approach for nighttime crosswalk contrast and brightness evaluation

Maya Cranor, Rucha Dave, Allyson Hur, Daeyoung Kim, Natsuki Sacks
Advisor: Lawrence Neely

Contents

1 Executive Summary

A scalable approach to evaluating nighttime pedestrian crosswalk safety using GIS and contrast heuristics

Team Members: Maya Cranor, Rucha Dave, Allyson Hur, Daeyoung Kim, Natsuki Sacks
Advisor: Lawrence Neely
Partner: Volpe National Transportation Systems Center

Context and Motivation

Nighttime pedestrian safety is a critical concern—**76% of pedestrian fatalities occur at night.** However, most municipalities lack standardized tools to evaluate crosswalk lighting. Boston, for example, relies on a reactive 311 call-in system, which unintentionally prioritizes affluent areas due to higher reporting volume.

The current process for evaluating lighting is time-consuming, inconsistent, and often unscalable. Our goal is to build a data-driven, geographically scalable method to assess lighting conditions across thousands of crosswalks using spatial data and validation techniques.

Methodology

Data Collection

- **Crosswalks:** Polygon dataset from UMass YOLOv8 computer vision model
- **Streetlights:** Boston's Vision Zero streetlight inventory (95,000+ records)
- **Road Segments:** Street Address Management (SAM) system data
- **Field Validation:** 66 crosswalks assessed using handheld lux meters and annotated photos

Contrast and Brightness Prediction

- **Contrast Heuristic:** Calculated using distance and angle of streetlights relative to the pedestrian
- **Brightness Heuristic:** Sum of inverse-squared distances to surrounding lights
- **Classification Scheme:** Sum of inverse-squared distances to surrounding lights

Interactive Tool:

Built using **kepler.gl** displaying

- Crosswalk centerpoints
- Associated streetlights
- Direction of vehicle travel
- Predicted contrast and brightness values

Validation Process:

Validation was done through a comparison of manual ratings to model predictions. Four match types were identified: **Match, Slightly Off, Off, and Opposite** for contrast type. Any mismatches were further investigated through spatial inspection, GPS metadata, image review, and Google StreetView comparisons. Error reasons were classified into the following categories: *Threshold Sensitivity, Environmental Lighting, Dataset Mismatch, Ambiguous Perception, Software Misclassification.* Most of the major errors were due to errors in the datasets or environmental factors.

Key Results

Category	Description
Validation Coverage	66 crosswalks evaluated across Boston and surrounding suburbs
Match Rate	Majority of model predictions were either exact or within one level
Threshold Accuracy	Tuned using observed discrepancies and distribution of heuristic values
Brightness Threshold	Set at 0.03 for distinguishing high vs low brightness zones
Major Limitations	Seasonal lights, outdated streetlight data, inconsistent human perception

Table 1: Overarching results from contrast and brightness validation

Throughout this project, our main finding is that the contrast heuristic is a useful, scalable proxy for pedestrian visibility. The brightness heuristic is currently doesn't weigh lights between 6-10m enough and the function should be adjusted through experimentation and validation.

Data quality is the biggest limiting factor; outdated or incomplete streetlight data affects prediction reliability. The model's performance improves significantly with reference photos and angle-aware light classification. Validation enabled refinement of heuristic thresholds and identification of edge cases, and should be done more extensively to further refine both heuristics.

Next Steps

1. Expand validation to include at least **200 crosswalks** for stronger statistical backing.
2. Improve brightness heuristic by adjusting distance sensitivity and applying multipliers for fixture types (e.g., acorn vs rectilinear)
3. Use LiDAR data (PathPoints) for finer granularity on occlusions and real-world geometry.
4. Extend to other cities (e.g., Washington DC) with compatible GIS datasets.
5. Integrate WalkFirst-style prioritization, adding socio-economic risk factors for equitable decision-making.
6. Automate error classification and model re-training as more data is collected.

Resources

GitHub Repository: https://github.com/olincollege/night-light
Kepler Map Interface: https://studio.foursquare.com/map/public/cd85979d-db73-4a58-b17c-64dcd1544009
Methodology and Documentation: https://olincollege.github.io/night-light/explanations

2 Important Vocabulary

Luminance: the amount of light emitted or reflected from a surface in a given direction. Measured in candelas/m$\hat{2}$.

Illuminance: the amount of light falling on a surface. Measured in lux = lumen/m$\hat{2}$.

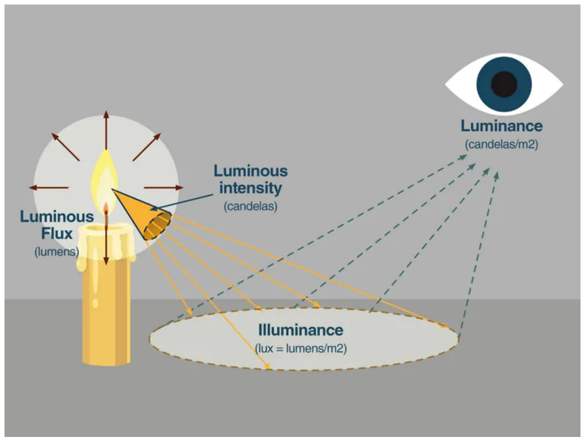


Figure 1: Graphic visualizing the various important lighting terms[1]

Contrast: the difference in brightness or color between elements, or between an element and its surroundings that creates a noticeable difference.

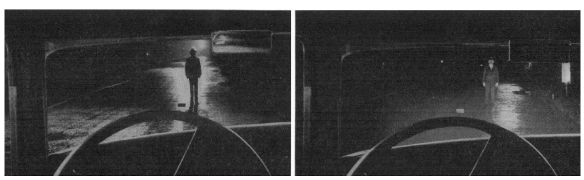


Figure 2: Left: Negative contrast. Pedestrian is dark while the background is starkly lighter. Right: Positive contrast. Pedestrian is lit while the background is starkly darker.[2]

We decided to categorize contrast using a five-level scale shown in Figure 3.

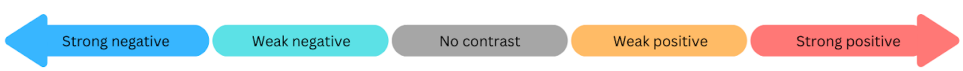


Figure 3: Graphic visualizing the contrast scale

Brightness: the subjective visual sensation of the intensity of light emanating from a surface or from a point source.

Visibility: the ability to see and identify objects, especially at a distance. Relates to how clearly something can be seen.

Heuristic: a proxy or approximation of a value used for quick problem-solving and decision-making.

[1] https://faro.es/data/webp/luminancia_e_iluminancia-9c1b90b20ea6e346e40ea7d35a2e5184.webp
[2] https://www.visualexpert.com/images3/contrastpolarity.jpg

(a) little visibility

(b) noticeable visibility

Figure 4: Difference in high vs low visibility lighting conditions

3 Problem

According to the Federal Highway Administration (FHWA), 76% of pedestrian fatalities occur at nighttime. Furthermore, the FHWA states that nighttime crossings at rural and urban intersections can be reduced by 33-38% with well-designed lighting. [3]

The current process for MassDOT and local municipalities to improve lighting is not standardized, with local and state public works departments placing and lighting crosswalks as they see fit. In Boston, the Street Lighting Department utilizes the 311 call-in system, where any resident of the City of Boston can leave a message or complaint about the lighting in a particular area. [4] The issues outlined in these calls are prioritized chronologically, since the city gets too many calls and the overhead to determine the most impactful projects would be too much.

The City of Boston is just one example. In-depth studies into lighting for pedestrian areas have only been conducted recently (within the last 10-15 years). Because lighting systems were designed based on the time and within limits of the time, re-assessment and improvements to align them with current standards is difficult. Most local governments across Massachusetts and America have difficulty maintaining and monitoring all lighting conditions across their jurisdictions, which lead to decisions that are based on assumptions and limited data. This often results in the allocation of taxpayer dollars to areas that may not have the greatest need or potential impact. By implementing a more comprehensive approach to inventory lighting conditions, local governments can ensure their investments are efficient and aligned with the needs of their communities.

Poor lighting conditions at intersections, turns, and on high-speed roads can lead to fatal crashes where better lighting could have been the difference between life and death. In this project, our work in determining a scalable method for assessing pedestrian-crosswalk lighting conditions will help government organizations like the City of Boston get a more holistic understanding of the crosswalk lighting conditions across their communities and invest in crosswalk lighting projects that will make the greatest possible impact.

Project Goal

Our goal for this capstone project was to address the lack of data on nighttime visibility of pedestrians in crosswalks. Our project demonstrates a geographically scalable approach to inventory crosswalk lighting

[3] https://www.fhwa.dot.gov/innovation/everydaycounts/edc_7/nighttime_visibility.cfm#:~:text=The%20nighttime%20fatality%20rate%20on,during%20darkness%20will%20save%20lives

[4] Donaghy, Michael. Zoom interview with co-authors Daeyoung Kim and Natsuki Sacks. 18 October 2024.

safety conditions in at least one geographic area and ideally scalable to multiple cities and states. [5]

4 Our Approach

Throughout the fall semester, we explored a number of data collection and lighting assessment methods to understand which methods would be most feasible given our limited resources, timeline, and experience.

Our main goal during this time was to uncover a data collection method that would require little onboarding/training, would be scalable to other municipalities, and would use few resources. This was intended to provide a more accessible, sustainable, and time-saving alternative to the current crosswalk lighting evaluation process standardized across the industry. This process is outlined by the FHWA[6].

Methods Explored

Below in Table 2 is a breakdown of the data collection methods we explored and their benefits and drawbacks. The avenue that we ended up deciding on was the first option below, highlighted in the table. Manual illuminance measurements were combined with existing inventory datasets to develop a prediction for contrast and brightness and validate it. We will expand upon this method throughout the paper.

When considering our options, we weighed our limited time frame, resources, and manpower. Many of these options would be viable for a multiple-year-long project, but would be unfeasible for our one year capstone. Specifically, validating our data collection method and iterating on it as we find shortcomings would be the most time consuming portion.

Table 2: Comparison of Crosswalk Data Collection Methods Explored

Method	**Pros**	**Cons**
Collecting illuminance measurements by hand at a variety of crosswalks and correlating that to the visibility of the pedestrian. **Selected Method**	• We could create a heuristic for correlation. • Would allow us to hand-pick which crosswalks we look at, ensuring we're hitting all different types of crosswalks. • Could standardize this with some sort of measurement device (something that measures distance).	• It takes a substantial amount of time to be able to get a statistically significant number of crosswalks, so the current data collection method might not be feasible within the year. • The photos that we take are not a 1:1 comparison with a human's perspective, since the camera's plane of view is larger.

Continued on next page

[5]Project Statement: Address the lack of data on nighttime visibility of pedestrians in crosswalks. Create a scalable, easy-to-use system to identify lighting conditions at various crosswalks.
[6]https://highways.dot.gov/media/4406

Table 2: Comparison of Crosswalk Data Collection Methods Explored (Continued)

Dashcam and software analysis Attach a dashcam to the inside of a car and utilize a software to analyze the contrast of the pedestrian and crosswalk lighting.	• We would be able to drive around and collect data without getting out of the car. • The software would do all of the data analysis for us, spitting out some sort of ranking or prioritization of crosswalks for our stakeholders to use in their project decisions. • If done well, this could scale to any city.	• We would still have to drive around at night for a long period of time. • There would be an issue with oncoming car headlights messing with the images, making the conditions inconsistent. • The dashcams we used had terrible video quality, making it difficult for us to analyze the videos accurately. • The video analysis working well would depend on the software being very accurate and thorough, which would likely take a couple of years. We would've had to figure out how to isolate the individual pedestrian(s), account for other external light sources like headlights, and also tune the image processing.
Raspberry Pi and GPS module Program RaspPi with a camera to take a picture when the GPS module is close to a crosswalk or if it detects a human	• We would be able to just drive around and have the hardware do the work for us; we wouldn't have to walk around with a camera and choose when to take pictures. • We could create a program that would go through the data, use image processing to determine the contrast of the pedestrian, and mark certain crosswalks as in need of attention or not. • This would definitely be scalable.	• We would still have to drive around at night for a long period of time. • There would be an issue with oncoming car headlights affecting the accuracy of the images, making the conditions inconsistent. • We would either need to artificially put a pedestrian in the crosswalk or drive around until we found one. The former would add time to the whole process and the latter would make it even more difficult to find a "perfect" situation in which there is a pedestrian in the crosswalk AND no oncoming cars. • The quality of the RaspPi camera is going to be considerably different from an actual person's perspective, possibly making our data collection a poor analogy for real-world visual conditions

Continued on next page

Table 2: Comparison of Crosswalk Data Collection Methods Explored (Continued)

Using a drone to scan crosswalks, equipped with a software algorithm that determines the level of lighting and contrast in each crosswalk	• The drone would allow for hands-free data collection. • The software processing would allow us to categorize a significant number of crosswalks. • We could encode predetermined paths for the drone to follow. The preparation process will be time-consuming (choosing crosswalks, deciding paths), but the data collection and software should be quick once everything has been iterated on to be functional.	• Onboarding process of flying the drone would take time. • Someone would likely still need to be manning the drone. • A ton of complexity is introduced – if there are cars in the road, the drone will not be able to take images. The drone will have to be a specific distance off the ground to mimic the perspective of a driver. • There would be a lot of iteration on the algorithm to ensure that it is accurate, so not feasible with our timeline.

Ultimately, we made the choice to pursue the in-person, by-hand data collection option. This aligned best with what Volpe needed and provided an option to make significant progress with this common issue. Furthermore, an interview with a potential stakeholder provided us with the resources necessary to move forward with data collection.

Turning Point: Interview with Michael Donaghy, Superintendant of Lighting for the City of Boston

We were connected to Michael Donaghy through our Volpe contacts as a potential stakeholder. We went into this meeting with the intention of figuring out what type of product an end user like Michael would benefit from, and what sort of information he looks for to make his lighting decisions.

Michael informed us about the current 311 call-in system, as described in the Context section. He also provided us with useful information involving current databases and lighting projects that were going on. He shared with us about the Vision Zero project and the City of Boston's lighting inventory, which gave us a dataset of all 7,000+ lampposts in the City of Boston.

Leveraging datasets available from the City of Boston, we were able to create a pseudo-contrast prediction using the location of each lamppost in relation to the crosswalk. We call this "pseudo-" because we are not determining contrast using the method widely accepted in the lighting evaluation industry, but have instead created a simplified version of the current time-consuming lighting data collection process. Although simplified, it could still serve as a guideline for municipal decision-makers when they are prioritizing crosswalks for lighting improvements by categorizing each crosswalk by the types of contrast present.

After generating a positive, negative, or no contrast prediction and a relative brightness prediction for each crosswalk, we could then go out into Boston and validate our data with the by-hand collection method that we explored earlier.

5 MVP Design

The MVP we designed allows for pedestrian contrast and brightness prediction for any crosswalk in the city of Boston. Every crosswalk is split into two by the centerline (one-ways are dealt with as an exception) because the contrast is different depending on which direction the vehicle approaches the crosswalk from. The MVP identifies if a pedestrian is in strong positive, weak positive, strong negative, weak negative, or no contrast for any given crosswalk center based on the presence of streetlights nearby. The background illuminance, or brightness, is similarly estimated based on the distance of streetlights nearby.

Assumptions

1. We assume that visibility is a metric made of two main factors: contrast and brightness.
2. The following datasets are available and up to date for the city of Boston. Details on the them are present in the corresponding sections. We assume that these correctly represent the street, lighting, and crosswalk characteristics in the city of Boston: Crosswalks, Lighting, Street Centerlines
3. We cannot determine whether or not lampposts are actually working from the dataset – they would have to be marked as so via data collection. We assume that the full lighting dataset is complete and correct.
4. Our current algorithm assumes all light fixtures output the same brightness. Next steps would include incorporating differences in wattage and brightness between types of light fixtures into the contrast and brightness heuristic algorithms.
5. There is one type of contrast affecting the pedestrian at a given crosswalk. Although variations in contrast occur as the pedestrian moves in the x or y direction, or even changes in height, we assume for a given side of the crosswalk, the contrast is constant.
6. We are determining contrast by assuming the worst case scenario. No headlights or background lights are incorporated into the calculation.
7. The pedestrian is not wearing all dark colors (standard is light gray).
8. Positive contrast is the best case scenario, followed by negative contrast, then finally no contrast.
9. The pedestrian is crossing down the center of the crosswalk.

MVP Steps

Based on the scope of the project, 5 steps were identified for the project:

1. Foundation: determine if positive, negative, or no contrast on pedestrian in crosswalk
2. Level 2: incorporate brightness calculations to help determine another layer of safety. If it there is no contrast and bright, could still be safe. If there is no contrast and dark, this is the worst-case scenario.
3. Level 3: incorporate one-way streets
4. Level 4: incorporate angle
5. Level 5: incorporate differences lamppost brightness and wattage

Of these, the first 4 were achieved. Level 5 as well as more directions for future work are outlined in the Next Steps section of the report.

6 Data Documentation

Data Storage

The data was stored and managed in a DuckDB database because of its lightweight, in-process nature, which makes it ideal for analytical tasks involving large datasets without the overhead of setting up a full database server. DuckDB is optimized for OLAP (Online Analytical Processing) workloads, making it well-suited for the kinds of spatial joins, aggregations, and filtering operations our project requires. Additionally it has a spatial library allowing advanced spatial querying. It integrates with Python and natively handles Parquet and GeoJSON files, which allows for a streamlined data pipeline from raw input to analysis. In order to better view the outputted parquet files we used DBeaver, which allows us to search, sort, and filter the results.

Raw Data Overview

Streetlights Database

After talking to Michael Donaghy, Superintendent of Street Lighting at the City of Boston Public Works Department, we learned that Boston completed a full catalog of their streetlight assets in 2023. This streetlights database primarily contains the locations of streetlights, with additional information about the type of bulb, height of light post, last replacement year, wattage, etc. This information is essential for developing a prediction for identifying the amount of lighting that comes from streetlights at any crosswalk.

We acknowledge that many cities may not have this data available, in which case either OpenStreetMap features could be used to roughly estimate the streetlight locations or the data could be collected by hand. While the latter will be time consuming, it will still be faster than the current evaluation method and can be done in the daytime without closing any streets.

Value	Count
<NA>	30,738
Cobrahead	22,927
Rectilinear	17,877
Acorn	11,767
Pendant	4,879
Colonial	1,933
Flood	1,254
Wells Bach	1,140
Other	1,128
Seaport	648
Shoebox	514
Ball Globe	406
Algonquin	227
Bishop Crook	175
No Fixture	170
Wall Pack	47
Nautical	39
Copley	20
Barn Style	13
Cube	13

Figure 5: Boston streetlights type breakdown

Currently, the Boston Streetlights database has a catalogue of 95,915 streetlights. 30,738 of those street lights have a fixture type of $< NA >$. After reviewing some of these streetlights found that they were either on private property (not put up by the government, typically a parking lot), are poles with something other than a streetlight on it, or don't exist. This dataset is slightly out of date since it doesn't account for new construction, and some of the fixtures labelled $< NA >$ cause error if they aren't actually streetlights. Overall, most of the information is correct and we assumed that is good enough for our use case (see validation section for more information).

Figure 5 above shows the breakdown of streetlights in Boston based on type of light.

Crosswalks Database

To analyze every crosswalk's lighting conditions, we need a dataset of crosswalks in the area. UMass Amherst has been developing a dataset of all crosswalks in Massachusetts using a computer vision model (YOLOv8) and aerial imagery. The dataset is a good starting point for our project and has the potential to be applicable for states that also do not have a thorough catalog of their crosswalk assets. The dataset can be viewed at the UMass Crosswalk Dataset.

The dataset represents each crosswalk as a polygon (approximately a rectangle) with the coordinates being the four corners of the crosswalk. Additional data is provided regarding the district, town, type of crosswalk, and area and length of the crosswalk. There are 10,611 crosswalks currently present in the database for the city of Boston.

Road Segments Database

Finally, to identify direction of vehicle travel (which informs contrast conditions), a database of road segments is used. The City of Boston has a street segments database collected from the Street Address Management (SAM) system. This was last updated on November 14, 2024.

The dataset represents each street segment as a linestring with coordinates of the starting and ending points. Additional information is present about the location, speed limit, length, type and name of the street. Currently, there are 42,444 street segments present in the database for the city of Boston. With the dataset we found that there are 5,243 crosswalks that are a part of one way streets. It's important to note that there were rare errors in the granular shape of the street segment, which led to some misclassifications in relation to crosswalk polygons; we found in some cases, the line segment was not exactly in the middle of the road, but offset much closer to one side of the road. This impacted the model's understanding of which side of the crosswalk that it intersects with the road, and thus causing unaddressed bugs that could not be fixed without manual intervention. The dataset can be viewed at Boston Street Segment Dataset.

To apply this work beyond Boston, a similar dataset of road segments (with geometries) would be needed for the city.

Data Cleaning and Processing

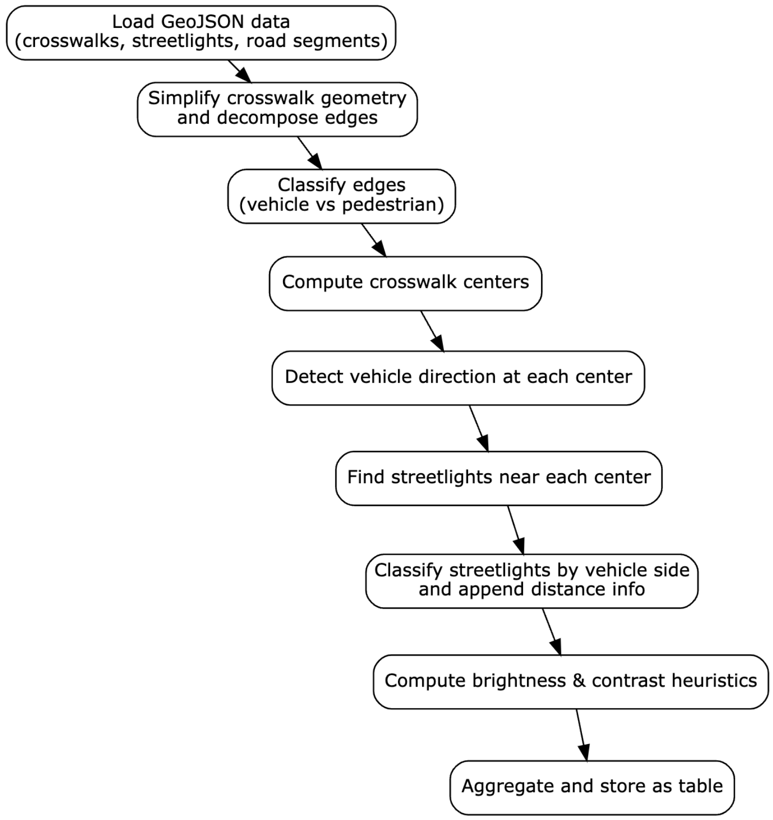


Figure 6: Breakdown of data cleaning and processing steps for contrast and brightness prediction

Figure 6 details the steps that are taken to clean and process the crosswalks, streetlights, and road segments data. The table created at the end of this process contains heuristic values for the contrast and brightness of every crosswalk in the database.

The first step involves simplifying the crosswalk geometries and decomposing the edges. The crosswalk polygons are simplified into minimum bounding rectangles with exactly 4 edges. The edges are decomposed, given IDs, and classified as being either vehicle or pedestrian edges based on whether they intersect with the street segment or not. Finally, a boolean *'is_oneway'* column is added to every crosswalk based on the intersecting street segment's characteristic. This information is stored in the *'crosswalk_segments'* table.

The next step of data processing is to identify the centers of each crosswalk. One-way streets get a single center point, which is labeled 'A' located at the centroid of the crosswalk. The two-way streets get two center points, 'A' and 'B' calculated at the interpolated midpoints between the pedestrian edges and the street segment intersections.

Car Approach Direction

The contrast at a crosswalk depends on what direction the car is approaching from and where the light is relative to the pedestrian (in front vs behind). Because of this, it is assumed that the driver always drives on the right side of the road (except for one-way streets). The next step of data processing is to identify the 'from coordinate' of every crosswalk center, or the coordinate that lies on the side of the crosswalk where the car will approach from.

For one way streets, the street segments database has a marked one-way direction which is used to determine vehicle direction.

For two way streets, we look at four different cases since crosswalks can be either tilted or straight with respect to the x axis:

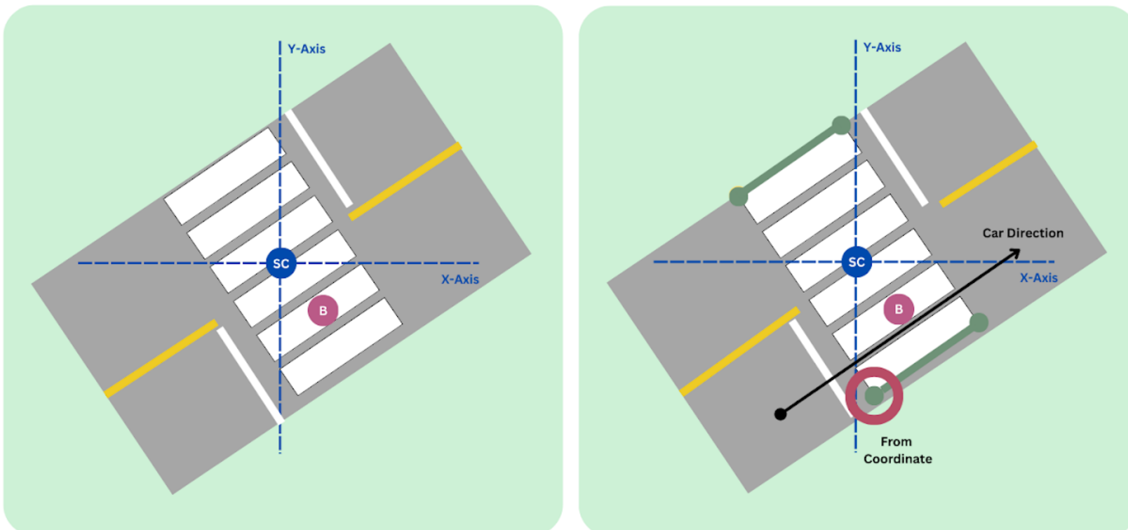


Figure 7: Diagraph showing how 'from' coordinate is identified for left-tilted crosswalks

1. As seen in Figure 7, the first case is when the x-coordinate of the crosswalk center B (in pink) is greater than the x coordinate of the street center (in blue) because the crosswalk is tilted. In this case, we look at the pedestrian edge coordinates (that were identified earlier, shown in green) and find the coordinate with the minimum y-coordinate. This is circled in red in the figure. This will be the coordinate that represents the side of the crosswalk where the car approaches from (the from-coordinate of the crosswalk), which will always be the right hand side of the road where the driver will drive.

Figure 8: Diagraph showing how 'from' coordinate is identified for right-tilted crosswalks

2. As seen in Figure 8, the second case is when the x-coordinate of the crosswalk center is less than the

x coordinate of the street center. In this case, we find the maximum pedestrian y-coordinate. This is circled in red in the figure. This will be the from-coordinate representing where the driver approaches from. Even if this coordinate is on the "opposite" side of the road, it is important to remember that it is simply being used to identify the direction of car travel and therefore will accurately depict car travel direction regardless.

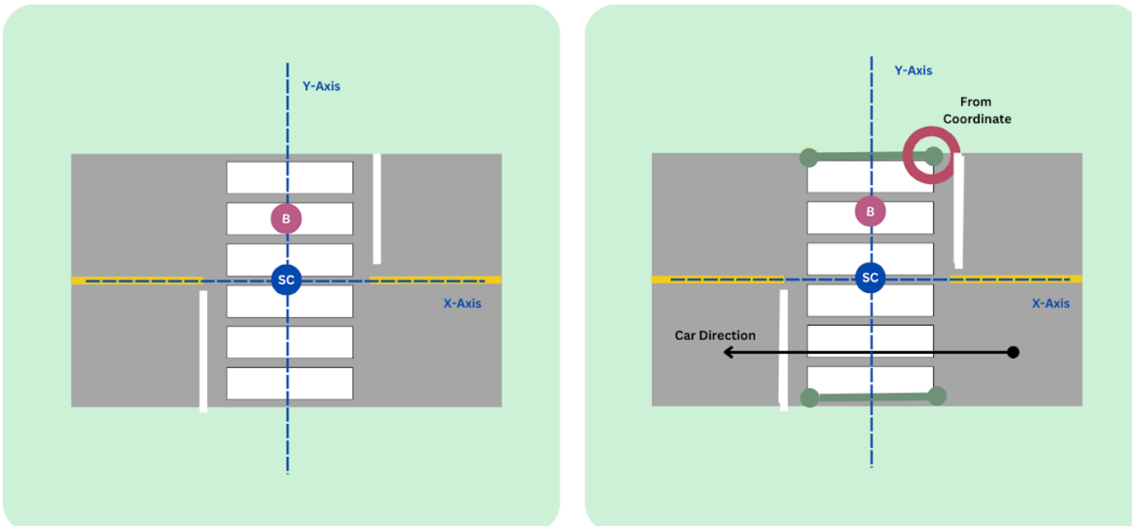


Figure 9: Diagraph showing how 'from' coordinate is identified for the top crosswalk center of a straight crosswalks

3. The third case is if neither of the above two apply (if the crosswalk is perfectly straight). In this case, we look at the y-coordinates of the crosswalk center as compared to the street center. If the crosswalk center is above the street center (higher y-coordinate), then we pick the pedestrian center with the larger x-coordinate value (to the right of the crosswalk) to be the from-coordinate. This is shown in Figure 9.

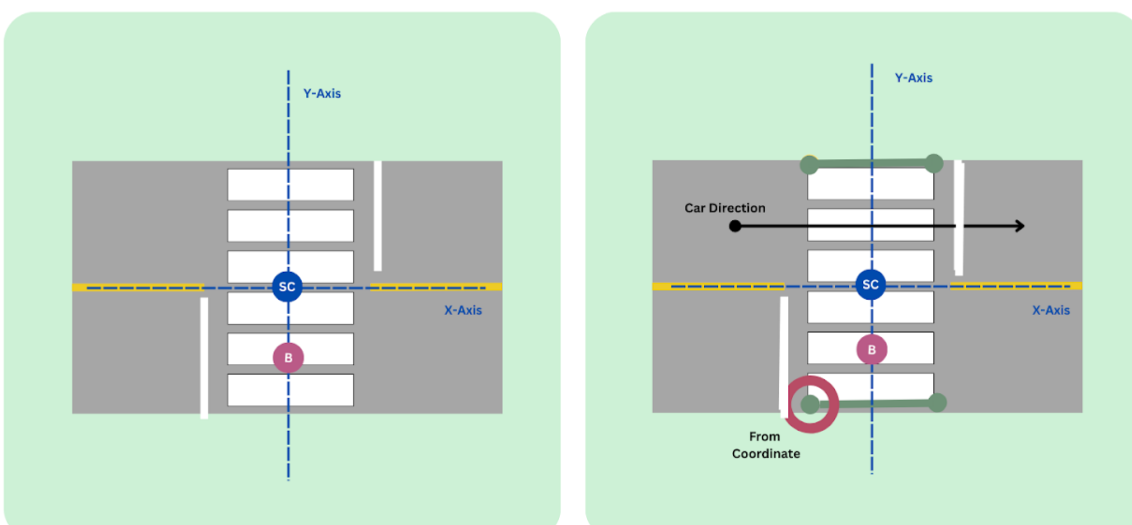


Figure 10: Diagraph showing how 'from' coordinate is identified for the bottom crosswalk center of a straight crosswalks

4. Finally, if the crosswalk is perfectly straight and the crosswalk center is below the street center (lower y-coordinate), then we pick the pedestrian center with the smaller x-coordinate value to be the from-coordinate, as seen in Figure 10

This information is saved in the *'crosswalk_centers'* table. This table includes 9 columns:

- *crosswalk_id* (ID of the crosswalk)
- *street_segment_id* (ID of the street segment)
- *ped_edge_geom* (line geometry of the pedestrian edge) →in green in diagrams
- *street_center_point* (coordinates of the center point of the crosswalk) →in blue in diagrams
- *geometry* (coordinates of the crosswalk center)→in pink in diagrams
- *center_id* (either A or B)
- *is_oneway* (boolean for if the crosswalk is on a oneway street or not)
- *from_coord* (coordinates of the from coordinate identified by the process above) →circled in red in diagrams
- *to_coord* (coordinates of the to coordinate which is the other point on the pedestrian edge line)

Streetlight Identification

After processing the essential characteristics of each crosswalk center, we must identify which streetlights are closest to each center. We currently keep streetlights that are within a radius of 20 meters from the crosswalk center using the ST_DWithin_Spheroid and ST_Distance_Sphere functions. 20 meters was a deliberately chosen metric for the threshold radius since there's a computational cost in increasing the radius and minimal returns in terms of improved accuracy since the heuristics value added through an inclusion of a given light was an inverse-square of the distance — at 10 meters away from the crosswalk, it would make $\frac{1}{10^2}$ difference, 20 meters would be $\frac{1}{20^2}$, and 30 meters would be $\frac{1}{30^2}$. It is important to note that inputs to these functions are expected to be in EPSG:4326 in latitude, longitude order.

The IDs and distances of each streetlight associated with a given crosswalk center are stored as arrays in the *'crosswalk_centers_lights'* table.

Streetlight Classification

Contrast compares the lighting conditions in front of the crosswalk to those behind it. Once impactful streetlights have been identified for each crosswalk center, they need to be separated based on what side of the crosswalk center they are present on. Figure 11 details the process of identifying an individual streetlight's side of the crosswalk.

During data preprocessing, a 'from coordinate' was identified for every crosswalk center, which represented the direction the car is approaching the crosswalk center from. In addition, the opposite pedestrian edge point was considered the 'to coordinate'.

For each crosswalk center, a vector is drawn between the two centers A and B (shown in pink). In the case of a one-way, A vector is drawn from the center to the midpoint of the pedestrian edge (which acts as a synthetic center B). A vector is also created from the 'from coordinate' to the 'to coordinate', specifying the direction of traffic (in green). Lines are drawn from the crosswalk center to ever single street light that was identified as being within 20 meters of the center (in yellow).

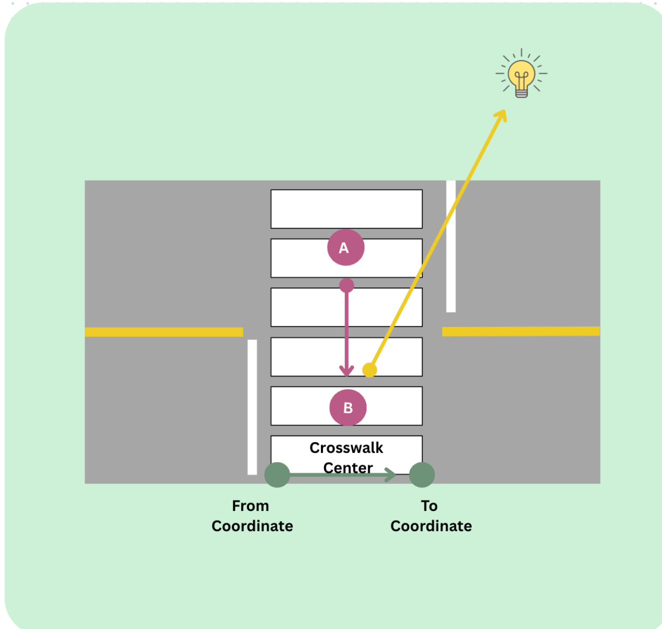


Figure 11: Diagram showing how vectors are added for light position, two crosswalk centers as a reference, and car direction

The cross product of the A-to-B vector (pink) with the from-to-to (green) vector is taken to determine the direction of the ‘from coordinate’ to the ‘to coordinate’. This acts as a reference. A similar cross product is taken of the A-to-B vector (pink) with every line between the light and crosswalk center (yellow). If this has the same sign as the reference, then this means the light is on the “to” side, or beyond the crosswalk. If it has the opposite sign, then the light is on the “from” side, on the side where the vehicle would approach from.

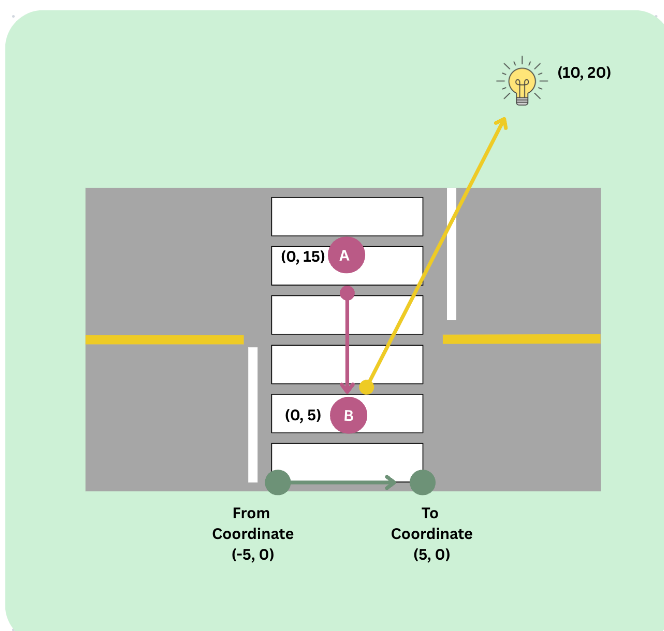


Figure 12: Example to demonstrate how streetlight sides are calculated

Following these steps for the crosswalk depicted above, we first define each of the three vectors:

$$
\begin{aligned}
A \rightarrow B &: (0,5) - (0,15) = \textbf{(0, -10)} \\
From \rightarrow To &: (5,0) - (-5,0) = \textbf{(10, 0)} \\
B \rightarrow Light &: (10,20) - (0,5) = \textbf{(10, 15)}
\end{aligned}
$$

The cross product of any two vectors A and B is defined as

$$A \times B = A_x * B_y - A_y * B_x$$

First, we take the cross-product of $A \rightarrow B$ and the $From \rightarrow To$ vector. This is our reference direction, representing the direction of the cars movement:

$$= (0 * 0) - (-10 * 10) = 100$$

Then we take the cross-product of $A \rightarrow B$ and the $B \rightarrow Light$ vector:

$$= (0 * 15) - (-10 * 10) = 100$$

Since the light has the same sign as the reference, this means it is on the opposite side of the reference and is on the "to" side of the crosswalk.

For every crosswalk center and streetlight associated with it, the location, distance and angle to the crosswalk center in question is retained along with which side of the crosswalk center it is on.

A new table *'classified_streetlights'* is created to store this information. Table 3 provides an example of a few rows from this table.

Table 3: Sample Table with Classified Streetlights

crosswalk_id	**center_id**	**streetlight_id**	**line_geom**	**geometry**	**side**	**angle_rad**	**abs_sin_**	**a_to_b**	**dist**
10	A	20	-	-	to	-	-	-	12
10	B	21	-	-	from	-	-	-	5

In this example, crosswalk center 10A has 2 streetlights close to it: 20 and 21. Each of these have separate entries on the table.

- *'line_geom'* column contains the Center→Light vector.
- *'geometry'* is the coordinates of the 10A crosswalk center.
- *'side'* represents the side of the crosswalk center the light was classified as. Here, streetlight 20 was found to be on the "to" side of 10A and 21 on the "from" side.
- *'angle_rad'* represent the angle between the Center →Light and A→B
- *'abs_sin_angle'* represents the absolute sine of that angle above
- *'a_to_b'* contains the A →B vector
- *'dist'* is the distance between the light and the crosswalk center 10A in meters. In this case, streetlight 20 is 12 meters away from 10A and 21 is 5 meters away.

Contrast Prediction

Finally, after classifying the streetlights, the contrast heuristic is calculated. The equation used is

$$\sum \frac{1}{distance^2} * |sin(angle)|$$

This heuristic accounts for both the distance of a streetlight from the crosswalk center and the angle [7]. This value is summed together for every streetlight within 20 meters of the crosswalk center. Although this is not a definite value of the contrast of a crosswalk, it works as an estimate of the contrast conditions and can be used to assess a crosswalk.

In the case that there are no streetlights identified as being near a crosswalk, the contrast heuristic is 0. Since having no light on the opposite side impacts the lighting conditions differently, crosswalks are classified differently in this case, as shown in Table 4.

Table 4: Contrast Categorization Criteria

Nonzero "to" side and "from" side contrast heuristics	**Difference in "to" side and "from" side contrast heuristic**	**Contrast Category**
Both sides have streetlights identified as being within radius (contrast heuristic != 0)	Difference <= 0.0075	No contrast
	0.0075 <= Difference <= 0.01	Weak contrast
	Difference > 0.01	Strong contrast
At least one side has no streetlights identified as being within radius (contrast heuristic = 0)	Difference <= 0.005	No contrast
	0.005 <= Difference <= 0.0075	Weak contrast
	Difference > 0.0075	Strong contrast

These threshold values were determined through the validation process detailed here.

Contrast Prediction Walkthrough

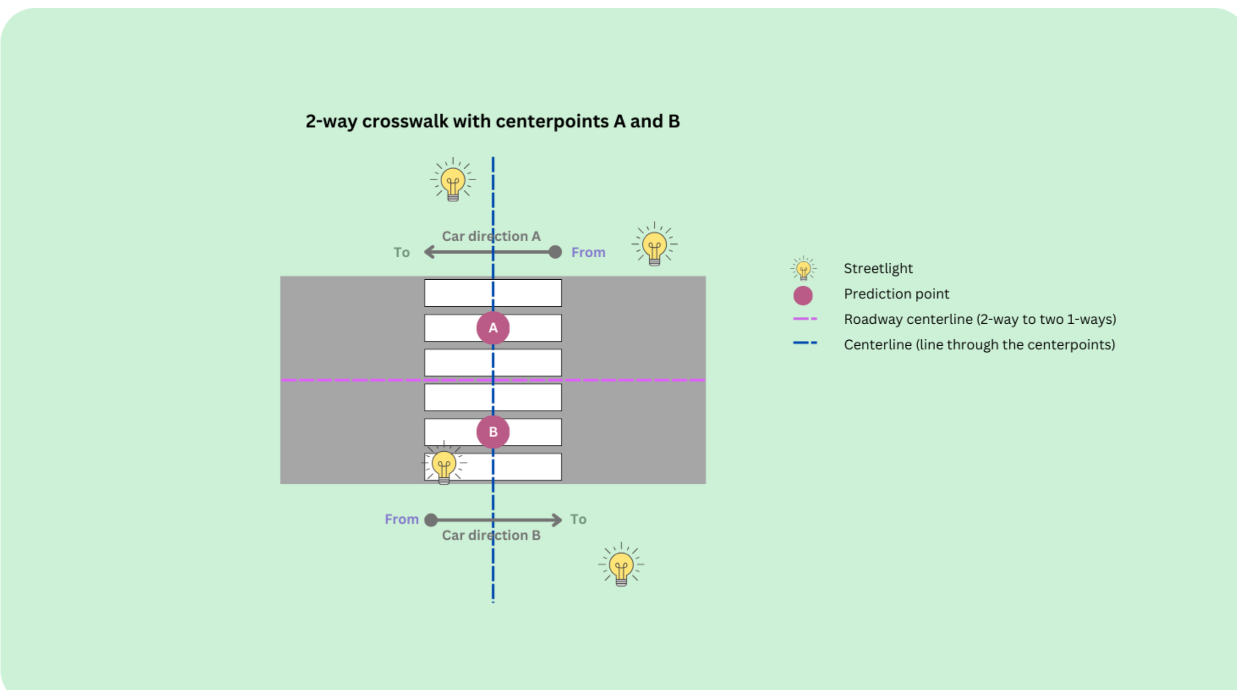


Figure 13: Sample crosswalk with center points and car directions

[7] https://www.ccs-grp.com/guide/theory-light-color/08_illuminance-properties.html

Figure 13 gives an example of a 2-way crosswalk with two prediction points A and B. There are four streetlights within the vicinity of this crosswalk.

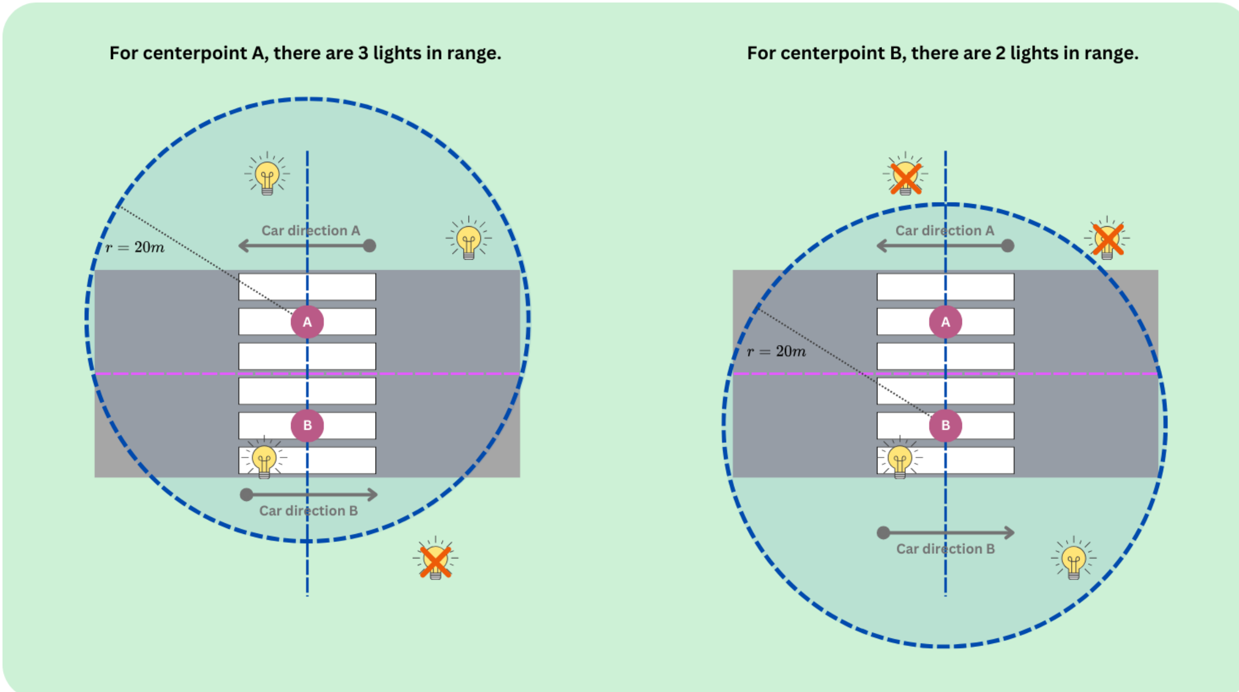


Figure 14: Sample crosswalk with lights for each center identified

Next, we apply the process detailed in the streetlight identification section to define a 20 meter limiting radius around each center point to determine which streetlights affect that respective centerpoint. For centerpoint A, there are three lights within that limiting radius. For centerpoint B, there are two lights within the limiting radius. This is shown in Figure 14.

After this, we determine which lights are on the "from" side vs. the "to" side of each centerpoint based on the car's direction of travel using the method explained in the streetlight classification section. Figure 15 shows this for the sample crosswalk. The solid circular endpoint of the car direction arrow represents the from side, while the arrow endpoint of the car direction arrow represents the to side. The areas of the "from" and "to" sides are determined by the limiting radius and by the centerline that goes through the centerpoints, which create the orange "from" side and the purple "to" side.

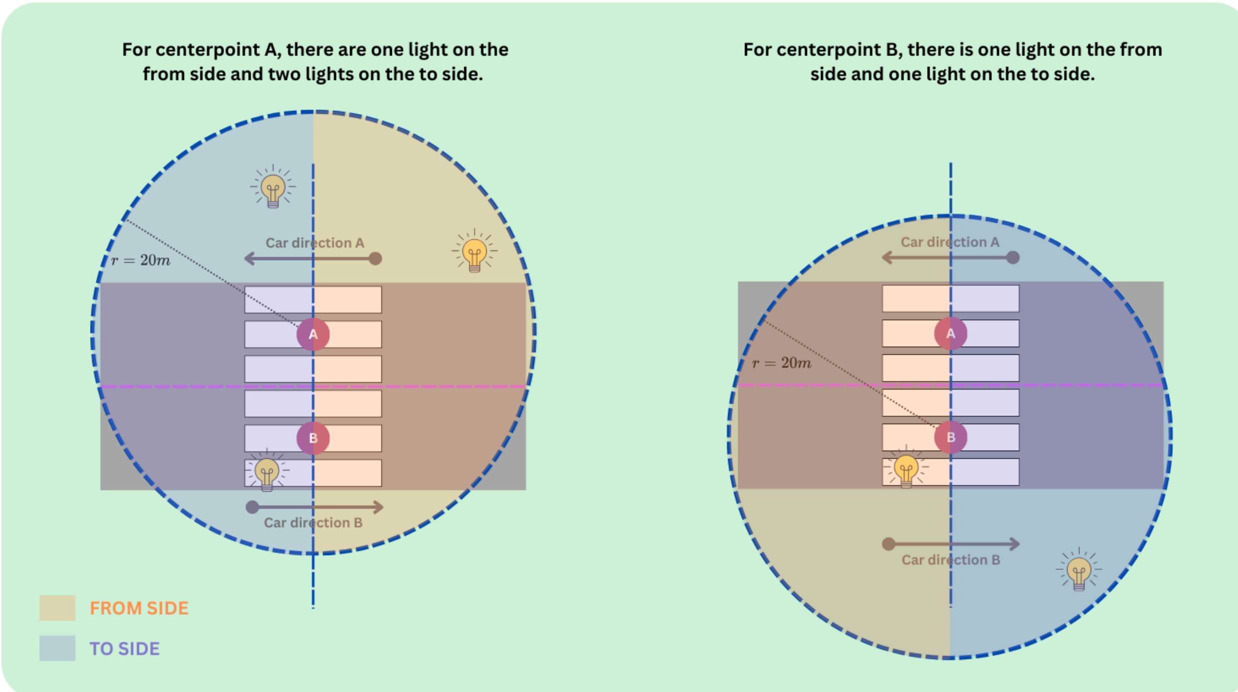


Figure 15: Sample crosswalk with lights for each center classified

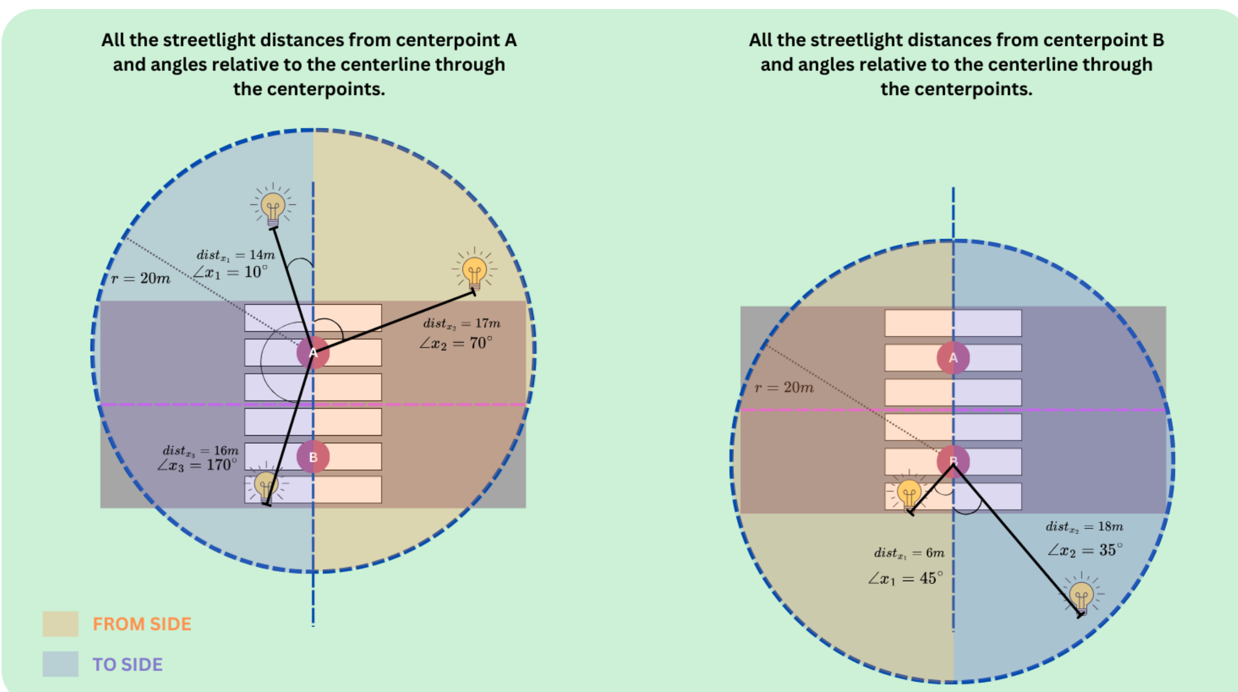


Figure 16: Sample crosswalk with light distances and angles specified

Finally, the contrast heuristic is calculated for the "from" side and the "to" side as shown in Figure 16. The equation used for this is:

$$\frac{1}{distance^2} * |sin(angle)|$$

Based off of Table 4's classification system, generally if the "from" side value is greater than the "to" side value, then the contrast is predicted to be positive. If the "to" side value is greater than the "from" side value, then the contrast is predicted to be negative.

In Figure 17, the "from" heuristic is 0.00156 and "to" heuristic is 0.00325. The difference in values is 0.00169. Based on Table 4, since this difference in values is ¡ 0.005, this crosswalk center A is classified as "no contrast".

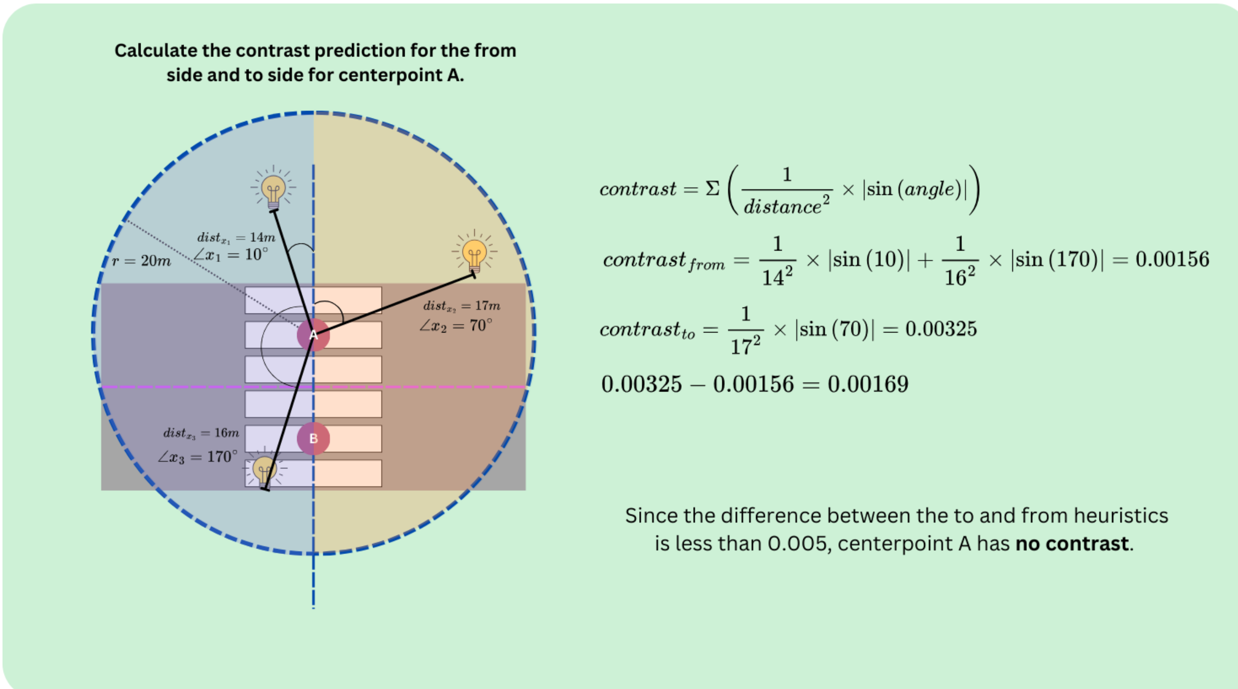


Figure 17: Contrast calculations for centerpoint A

We can use the same logic for centerpoint B. Here in Figure 18, the "from" heuristic is 0.00177 and the "to" heuristic is 0.01964. Since this value is greater than 0.01, based on Table 4, the centerpoint B is classified as having "strong positive contrast".

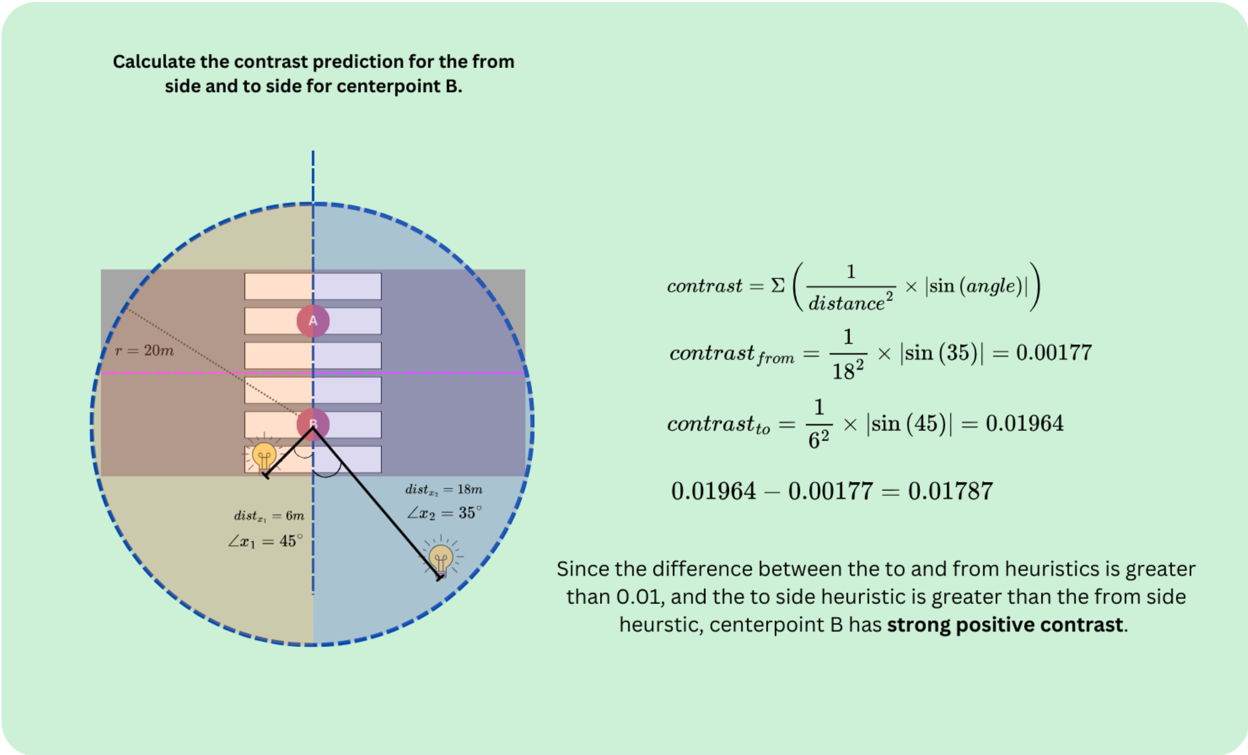


Figure 18: Contrast calculations for centerpoint B

Brightness Prediction

The brightness heuristic is calculated using the equation:

$$\sum \frac{1}{distance^2}$$

This is summed for every streetlight present near the crosswalk center (in the "to" and the "from" direction). Although this is not a direct value of the brightness of the crosswalk, it is meant to work as an estimate of the overall brightness near a crosswalk to help assess its lighting conditions.

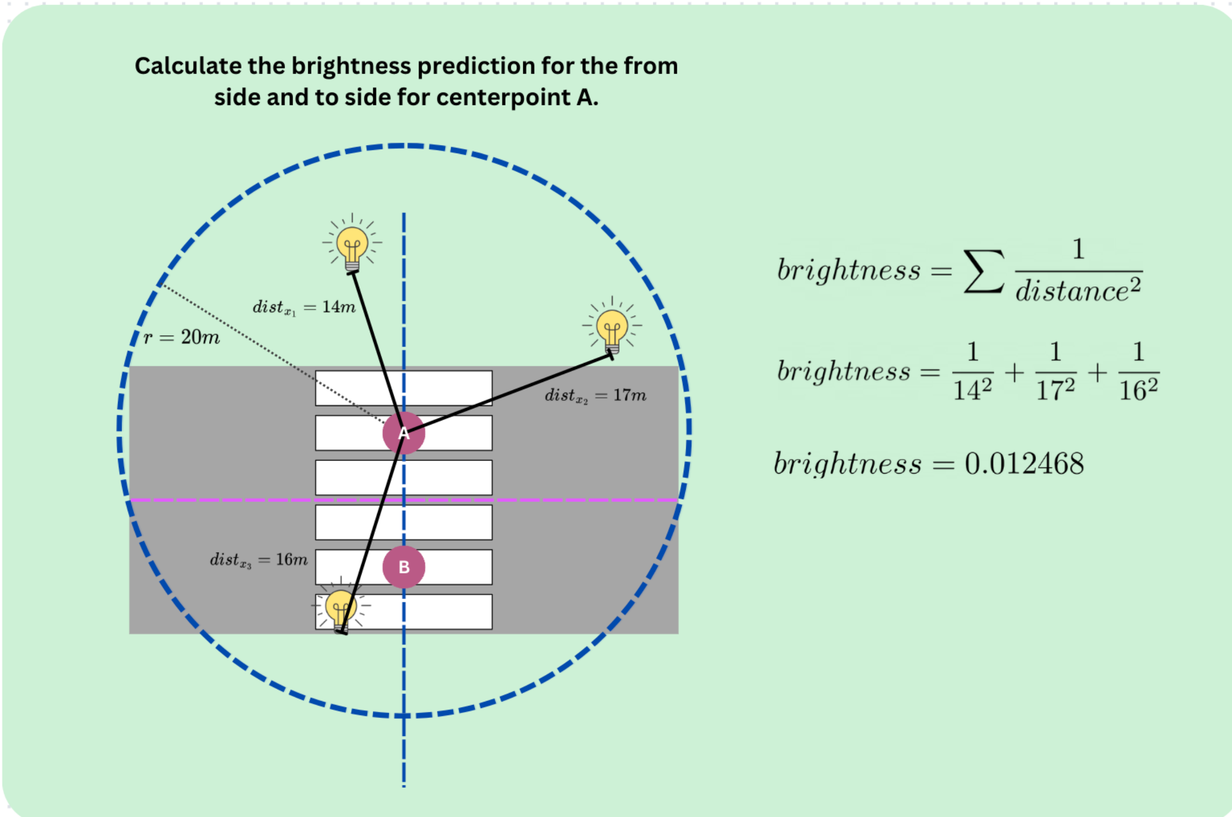


Figure 19: Brightness calculations for centerpoint A

Using the same example from the contrast calculations, Figure 19 shows an example crosswalk with the brightness heuristic calculated. Here, there are three streetlights 14, 17, and 16 meters away from center A. By the equation, the brightness heuristic ends up being 0.012468.

Final Data Table

Our final data table is now complete, labeled *'crosswalk_centers_contrast'*. Table 5 provides an example of what this table could look like. Each row contains the various heuristics calculated through the steps above for each crosswalk center in the database.

Table 5: Sample Final Table

crosswalk id	center id	to contrast heuristic	from contrast heuristic	to brightness heuristic	from brightness heuristic	contrast heuristic	light heuristic
12	A	0.0089	0.0064	0.0247	0.0175	No contrast	0.0422
14	B	0	0.0099	0	0.0339	Strong positive contrast	0.0339

As seen above, the following values are calculated for every crosswalk center:

- *To_contrast_heuristic*: this is the contrast heuristic for the "to" side of the crosswalk (the background of the pedestrian, or the direction the car is heading towards)
- *From_contrast_heuristic*: this is the contrast heuristic for the "from" side of the crosswalk (the foreground of the pedestrian, or the direction the car is approaching from)

- *To_brightness_heuristic*: this is the brightness heuristic for the "to" side of the crosswalk
- *From_brightness_heuristic*: this is the brightness heuristic for the "from" side of the crosswalk
- *Contrast_heuristic*: his is the overall contrast decided based on the "to" side contrast heuristic and the "from" side contrast heuristic, as determined by Table 4
- *Light_heuristic*: this is the brightness heuristic calculated by summing the values for the "to" and "from" sides

7 Boston Crosswalk Visibility Map

To more easily interact with our algorithm and view the data, we created an interactive map that displays crosswalks, contrast types, and more for each crosswalk present in the Boston crosswalk dataset. We used an application called kepler.gl, which is an open source geospatial analysis tool that has a mapping feature. The Boston results can be found in this kepler.gl map. The following is a walk through of how to read the data found on kepler.gl and how to use it. The final database table *'crosswalk_centers_contrast'* is put into kepler.

Kepler Walkthrough

Figure 20 provides an example of a crosswalk on the map. The left photo 20a is with the base map in satellite street mode and the right 20b is in dark mode. It can be easier to read the map in dark mode, but satellite street mode allows us to see the surrounding environment and where exactly the crosswalk and street are. Figure 21 explains how to switch between the two by toggling the map style.

(a) Kepler map in street mode *(b) Kepler map in dark mode*

Figure 20: Kepler map overview

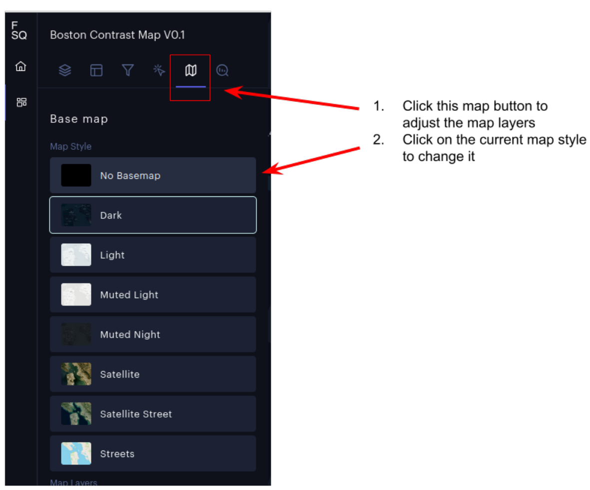


Figure 21: Map style toggle

Figure 22 provides an example of a crosswalk on a two-way street. The blue dots represent streetlights that are not within a 20-meter radius of a crosswalk center point. Next to each asset is a number in white, which is the asset's ID and can be used to look it up in the database. The yellow dots with blue outlines represent streetlights that are within 20 meters of a crosswalk center point. The red dots labeled A and B indicate the crosswalk center points. If the street is one-way, there will be only one center point labeled A, positioned in the middle of the street. The lines connecting the center points and streetlights indicate the relative location of the lights based on the direction of traffic. Red lines represent the ”to” side—cars will be driving toward the light, placing it behind the pedestrian and providing backlighting. Blue lines represent the ”from” side—cars will be driving away from the light, placing it in front of the pedestrian and providing front lighting. The text in the blue box contains the contrast evaluation for the center point located to its left.

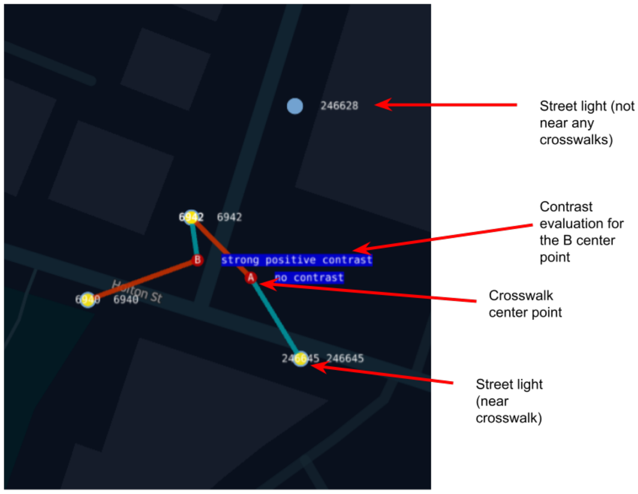


Figure 22: Sample 2-way crosswalk in Kepler

Streetlights, crosswalk center points, and the streetlight lines can all be hovered over for more information, as seen in Figures 23, 24, and 25.

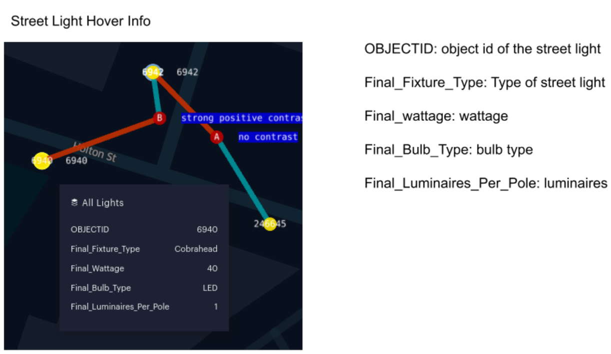


Figure 23: Example of information provided for each streetlight

Figure 24: Example of information provided for each crosswalk centerpoint

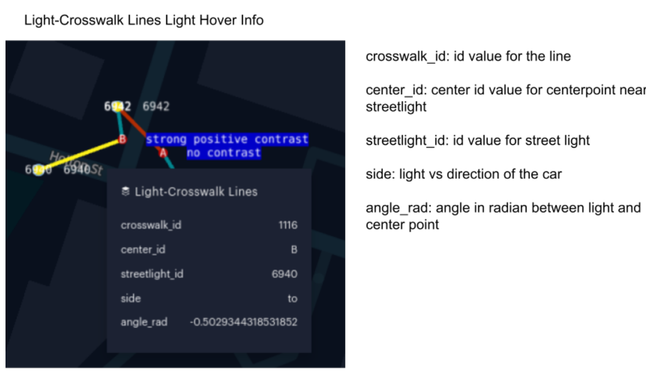


Figure 25: Example of information provided for each crosswalk line

Another important feature is the filter option. It helps find all instances of a value or find the location of a specific asset. Figure 26 and 27 provide a guide on how to use the filter button to find a crosswalk.

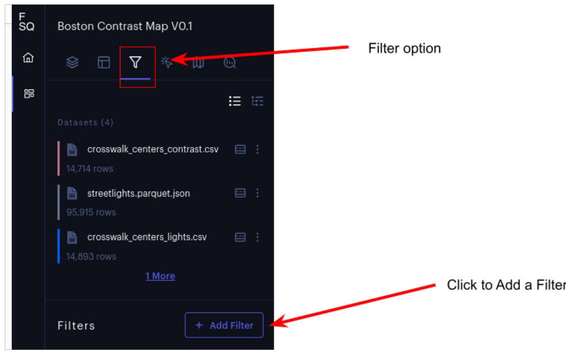


Figure 26: Description of filter buttons on Kepler

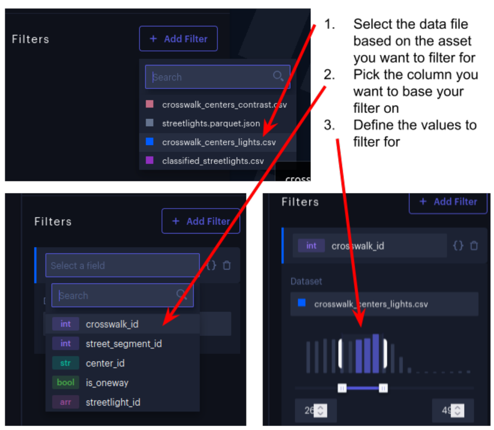


Figure 27: Steps for changing filter on Kepler

The first tab is for adding or adjusting laters. Here, more information can be added to the map and the appearance of the map can be adjusted, as shown in Figure 28 and 29.

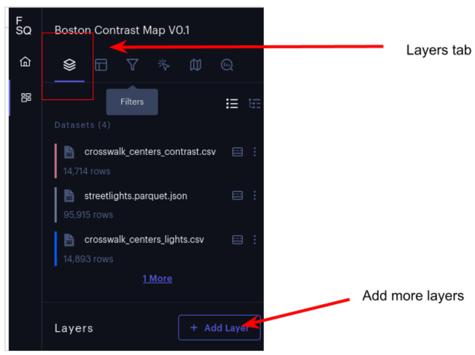


Figure 28: Description of layers buttons in Kepler

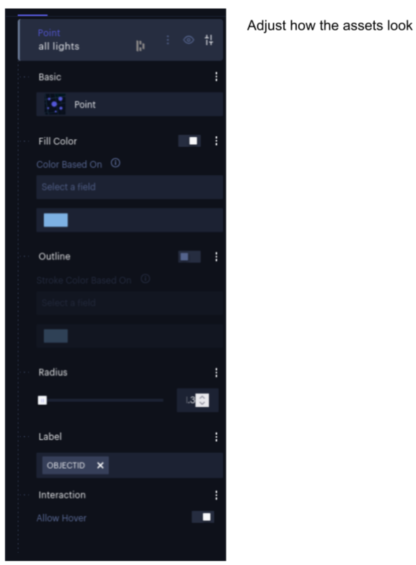


Figure 29: Tab to adjust assets

8 Data Validation Methodology

We conducted a validation study to assess the accuracy of the contrast heuristic model against field-collected, perceived contrast data. The goal was to identify discrepancies in the contrast and brightness heuristic calculations and understand their root causes for areas of possible improvements to the algorithm.

Contrast Validation Steps

Table 6: Contrast Validation Match Table

Match Type	Description	Example
Match	Exact agreement between model & perception	Weak Negative vs Weak Negative
Slightly off	One step off, but **same polarity**	Strong Negative vs Weak Negative
Off	Two steps off, but **same polarity**	Strong Positive vs No Contrast
Opposite	**Polarity mismatch (positive vs negative, excluding "no contrast")**	Strong Positive vs Weak Negative

The contrast validation process began by comparing the model's predicted contrast classifications with those gathered from field observations. The algorithm classified 5 possible contrast types: strong positive, weak positive, no contrast, weak negative, and strong negative. Each case was categorized based on how closely the model's prediction aligned with human perception – whether it matched exactly, was slightly off, or showed a complete mismatch. These are explained in Table 6. A "match" occurs when both contrast ratings are identical, signaling perfect agreement between the two columns. When the ratings are one step apart – such as "no contrast" versus "weak negative" – it's considered "slightly off" as a divergence still within the same polarity. A two-step gap – for example, "strong positive" versus "no contrast" – is labeled "off," indicating a more substantial mismatch while still sharing overall polarity. Finally, any pairing of a positive rating with a negative one (excluding "no contrast") is classified as "opposite," marking a direct reversal in polarity. The final model contrast accuracy distribution is shown in Figure 30.

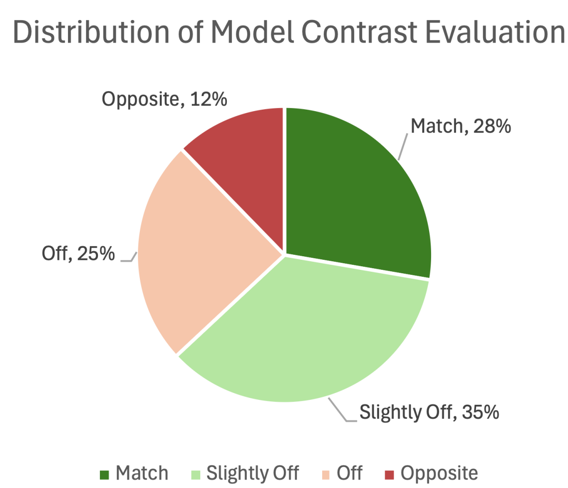


Figure 30: Pie Chart detailing rate of match/mismatch in contrast validation

As seen above, 63% of contrast predictions were acceptable. Cases with major mismatches were selected for deeper investigation. For these, we reviewed the streetlight configuration around the crosswalk using spatial data and visual maps. We checked if the model's output made sense given the distribution of streetlights and evaluated whether the underlying contrast heuristic values from each side were just above or below key thresholds. This helped determine if fine-tuning the threshold values might resolve the disagreement.

In parallel, we verified the integrity of the crosswalk and streetlight identification process by checking the catalog records and inspecting the GPS metadata embedded in the field images. These were compared against Google StreetView for additional visual source of data. Visual inspection of the photos and StreetView also helped clarify whether the mismatch was due to lighting conditions not captured in the dataset.

Throughout this process, we documented whether errors stemmed from limitations in the data, the model, or external conditions. Each case was assigned one or more labels such as environmental factors, light dataset mismatch, threshold sensitivity, ambiguous perception, or software-related issues.

Table 7: Contrast Validation Error Table

Error Type	Description	Example
Threshold sensitivity	Heuristic value near decision border	Contrast heuristic is 0.007 when the weak positive contrast threshold is 0.0075
Environmental	Temporary conditions not factored in dataset	Christmas decoration lights on the street, storefront lights, etc.
Light dataset mismatch	Incomplete or outdated lighting catalog	Streetlight mapped in dataset not present in physically
Ambiguous perception	Human interpretation varied or unclear	Multiple types considered reasonable for perceived contrast type
Software issue	Bugs in crosswalk/light classification algorithm	Incorrect vehicle direction, one way classification, etc.

Contrast Validation Walkthrough

Here's an example of a walk through for a given crosswalk in the Newbury street area in Boston:

Table 8: Contrast Validation Error Table

Error Type	Description	Example
Threshold sensitivity	Heuristic value near decision border	Contrast heuristic is 0.007 when the weak positive contrast threshold is 0.0075
Environmental	Temporary conditions not factored in dataset	Christmas decoration lights on the street, storefront lights, etc.
Light dataset mismatch	Incomplete or outdated lighting catalog	Streetlight mapped in dataset not present in physically
Ambiguous perception	Human interpretation varied or unclear	Multiple types considered reasonable for perceived contrast type
Software issue	Bugs in crosswalk/light classification algorithm	Incorrect vehicle direction, one way classification, etc.

Contrast Validation Findings

Table 9: Example Contrast Validation Table Entry

Crosswalk ID	Center ID	Perceived Contrast Type	Model Predicted	Map Perceived Contrast Type	Evaluation
4484	A	Strong Negative	Weak Negative	Negative	Slightly Off

*Note: "Map Perceived Contrast Type" was reduced to no contrast, negative, or positive, since more refined prediction (with weak or strong) wouldn't be very meaningful.

Since the model outputted a different contrast type than what we had originally identified during the field survey but still within the same polarity (strong negative vs weak negative), we started by checking the heuristic values for each direction of the street. In this case, the values were 0.01766 and 0.00944 for to-side and from-side, respectively, making the contrast heuristic value -0.00821 (= 'from-side' - 'to-side'). Since the threshold for strong negative contrast was set to -0.01, the crosswalk qualified for weak negative contrast instead. The threshold value was sufficiently far away from the output value, so this wasn't considered a threshold issue.

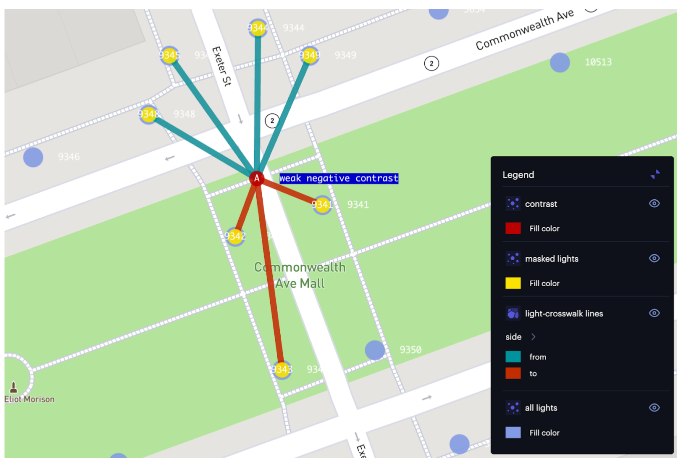


Figure 31: Sample crosswalk 4484A for deeper analysis on Kepler

We also looked at the map with the features populated (crosswalks and streetlights) as shown in Figure 31, and tried to give it a rating based on 2-dimensional configuration in bird's eye view. Even though there were more lights on the from-side, there were lights much closer to the crosswalk on the to-side, so we gave it a negative contrast type. Algorithmically, it aligned with what we would have predicted.

We then tried to identify the cause of mismatch by analyzing the image taken during the field survey directly. Looking at the image shown in Figure 32, we noticed that the street was much brighter on the left and right side of the street than expected.

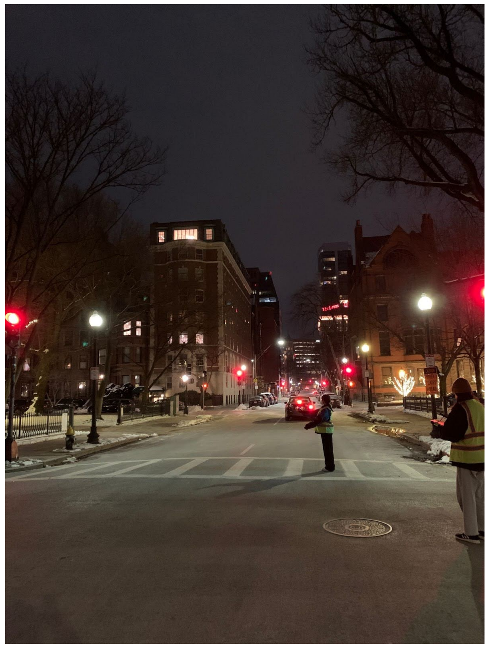


Figure 32: Field image of crosswalk 4484A

Looking at the street from a different angle, as shown in Figure 33, we realized that the decoration lights for winter season were skewing the background illumination. With the additional light source, the background was much brighter, and had led us to classify it as strong negative contrast during the field survey, while that information was absent in the lighting dataset. Therefore, we categorized this issue as environmental – as an issue that cannot be addressed without more accurate real-time lightning data.

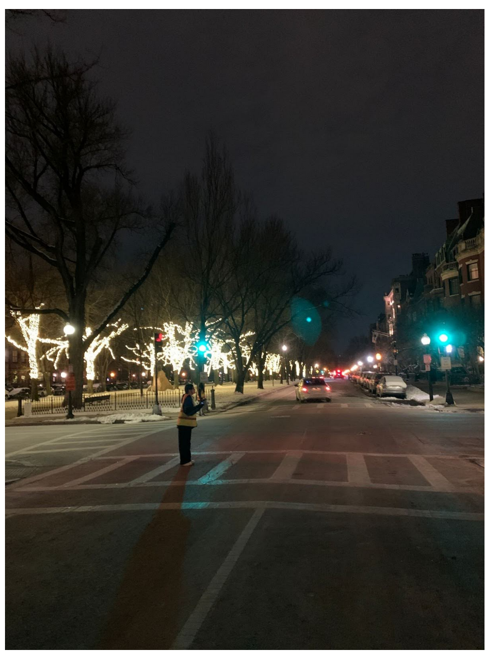


Figure 33: Field image of crosswalk 4484A

Additionally, we'll walk through another example of an error. The following crosswalk was located near the Harvard Business School area. As shown in Table 10, we found that the model predicted a strong positive and negative contrast for each side of the vehicle direction, whereas the perceived was no contrast for both.

Table 10: Example 2 Contrast Validation Table Entry

Crosswalk ID	Center ID	Perceived Contrast Type	Model Predicted	Map Perceived Contrast Type	Evaluation
32242	A	No Contrast	Strong Positive	Positive	Off
32242	B	No Contrast	Strong Negative	Negative	Off

From the visual inspection of the map shown in Figure 34, it becomes clear that the algorithm would output a strong contrast type because there is a streetlight extremely close to the crosswalk center points. Physically, this implies that the light will cast a very strong shadow in front or behind the pedestrian. But looking at the location of the streetlight, it is at the center of the road. This is unusual for a typical street configuration, and so we continue inspecting by turning to Street View images.

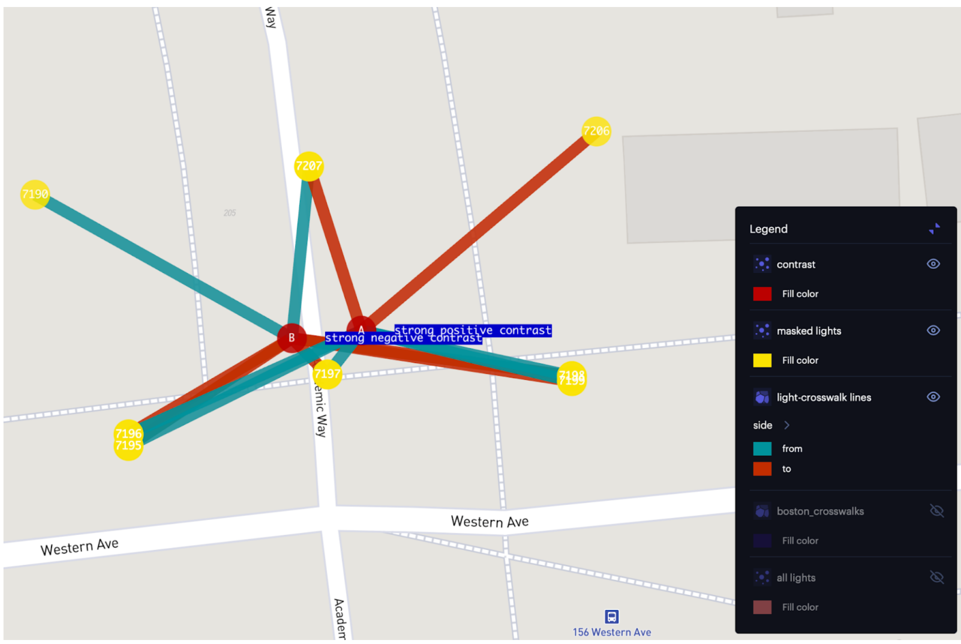
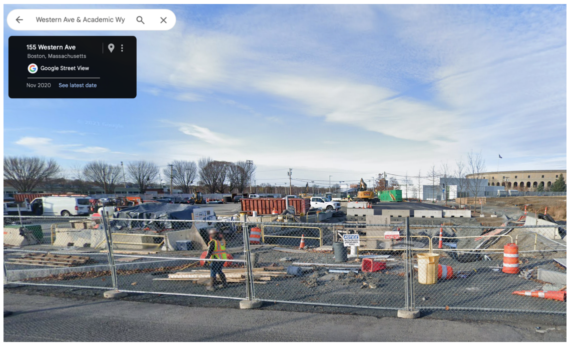
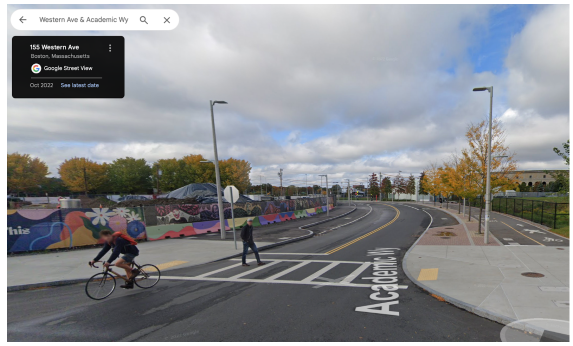


Figure 34: Map image of crosswalk 32242

The images shown in Figure 35 were taken in 2020 and 2022, respectively. What we saw during our field survey was similar to the image from 2022. According to the streetlight dataset, there is supposed to be a streetlight in the middle of the road on Academic Way, but as seen on Google Street View, it doesn't exist there today. It is possible that the streetlight used to exist at this prescribed point but might have been removed during the construction that is shown in the image from 2020.



(a) Street view of construction near the crosswalk



(b) Street view showing the light doesn't exist at shown location

Figure 35: Google Street View images of crosswalk 32242

This illustrates the importance of the accuracy of the source datasets that the model takes in and how the accuracy of the model is directly impacted by the accuracy of the sources. This situation is out of the model's predictive scope and must be addressed through more meticulous data cleaning and verification prior to the prediction synthesis. The discrepancy here was labeled as a light dataset mismatch.

We found that some types of errors – such as threshold sensitivity and algorithmic misclassifications – can be addressed through further development and refinement. Others, like environmental lighting changes or incorrect light data, are outside the system's control. This validation helped isolate which areas of the model could be fine-tuned to improve reliability and which issues must be accepted as limitations of the current data or scope.

The most extreme mismatches were from incorrect datasets. The model currently works very accurately for simple scenarios but struggles to identify contrast for more complex settings and edge cases.

Brightness Validation Steps

Similar to the validation of the contrast heuristic, we also spent time validating our brightness heuristic. Starting off, we added a *'map perceived brightness'* value for every crosswalk based. This was a perceived brightness value between 1 (low brightness) and 5 (high brightness) estimated based on the streetlights located on the map close to the crosswalk. We checked if this perceived brightness aligned with the heuristic calculated by the model. We also compared the heuristic to a 'perceived visibility' value, which ranged between 1-5. This value was estimated while manually collecting data on each of the crosswalks. Visually checking the pictures of the crosswalks helped verify whether these were accurate or not.

For any crosswalks where the perceived or map perceived visibility didn't align with the model's prediction, we further investigated what the reason for this could be and recorded the results.

Brightness Validation Results

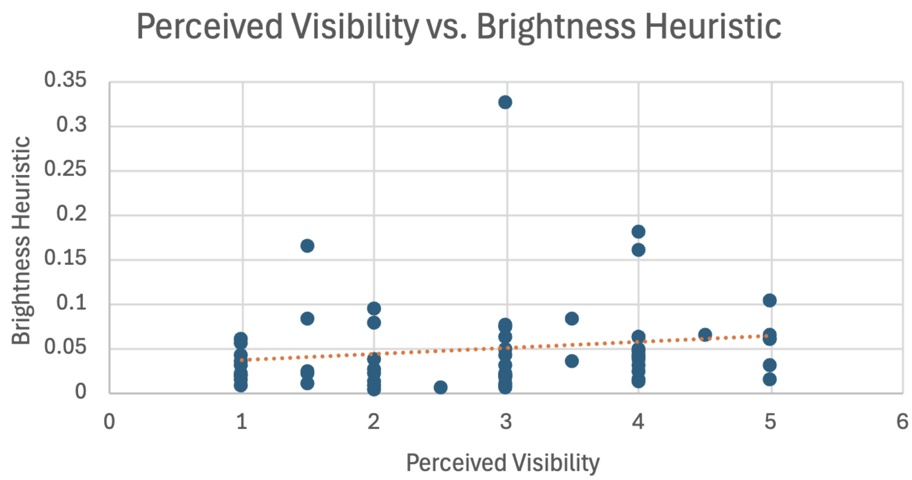


Figure 36: Brightness Heuristic vs Perceived Visibility

As seen in Figure 36, there does not appear to be any correlation between perceived visibility and the brightness heuristic. The heuristic values are scattered across the graph pretty much randomly. A reason for this is that the perceived visibility was recorded while on site collecting data. This meant that any background light coming from non-streetlight sources was also included. Also, each crosswalks was only investigated once. Towards the beginning of data collection, the observers judged the brightness more harshly, focusing on the smaller nuances between each one. But later, as we visited other locations, we realized these crosswalks were even worse. Because of this, the crosswalk evaluations changed based on the order of crosswalks with those at the end being closer to the real value than those at the beginning. Using the field observations was a less reliable way of benchmarking the visibility.

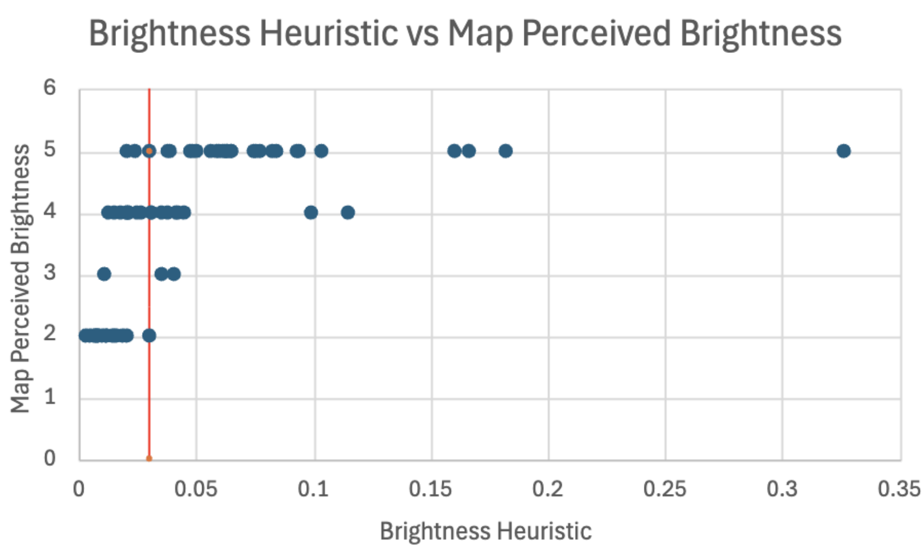


Figure 37: Brightness Heuristic vs Map Perceived Visibility

In comparison, Figure 37 shows the brightness heuristic compared to the map perceived brightness. This was the value that was estimated by looking at both the images from data collection and the Kepler map (identifying streetlights located close to the crosswalk). This method did not account for any background light sources and stayed consistent from the beginning to the end. As seen here, most crosswalks with a brightness heuristic above $\tilde{0}.03$ were classified as being a 4 or 5 in brightness, which meant that they had high brightness. In comparison, most crosswalks below $\tilde{0}.03$ were categorized as a 2 (the lowest brightness).

Based off of this validation, we set the threshold for high vs low brightness heuristic to be 0.03. Any crosswalks that were wrongly classified by the algorithm were investigated thoroughly. The majority of these cases were instances when crosswalks we had classified as being a 4 or 5 in brightness based on the map had a low brightness heuristic value. When each of these were investigated, we realized many cases were instances where there was an individual streetlight very close to the crosswalk (within 6-10 meters). However, with the current method of calculating the brightness heuristic, even lights around 6-10 meters away from the crosswalk had a heuristic value below 0.03 and were classified as being low brightness wrongly. Lights close to the crosswalk were not influencing the brightness predictions upwards as we wanted and expected.

Two immediate steps were identified for further exploration: changing the current heuristic function and adding a multiplier for light fixture types, both of which are explained in detail in the Next Steps section.

Validation Learnings

Collecting validation data in person was a time-consuming process, so we tried to streamline it as much as possible to maximize the amount of data we could collect (see the Appendix for more details on this process). Overall, more validation data is needed to gain a stronger sense of the model's accuracy.

About half of the data we collected involved our subjective ranking of crosswalk characteristics. As we went through this process, we became more consistent at evaluating general brightness and contrast. However, since there isn't a definitive "true value" for brightness and contrast, it was difficult to obtain consistent validation metrics. To address this, we included lux values and phone pictures alongside our subjective rankings.

It's important to note that contrast can vary depending on what a person is wearing. For example, someone may have high contrast on their upper body and low contrast on their lower body. We focused more on how shadows were cast across the crosswalk, but it is often difficult to assign a single type of contrast (positive, negative, none) to one scenario. For future data collection, a strong understanding of brightness and contrast will be crucial.

Once the validation data was compiled, analyzing the results was still time-consuming. We found success by manually reviewing each crosswalk and identifying discrepancies. Having reference photos of each location was extremely helpful, allowing us to see environmental differences between real life and what the dataset represented. From these photos, we discovered that the lighting dataset was slightly outdated due to recent construction, inconsistencies in the streetlight dataset, and other environmental factors affecting light levels.

In the future, the validation process should be more automated—for example, by automatically calculating the percentage of crosswalks correctly identified. However, at the time of this analysis, the model was one iteration behind needing that level of automation.

9 Lessons Learned

Compared to current methods, our solution is quicker and more scalable. The current method of manual data collection is accurate but the most time-consuming, involving taking multiple measurements at an individual crosswalk in a precise manner at nighttime. Our algorithm automates much of the evaluation process, and by utilizing both methods, we can create a final product that is both scalable and accurate.

A key component of our method is the effective use of GIS Data, which provides valuable insights into crosswalk visibility. We used the streetlight locations, crosswalk locations, and road segments to develop our algorithm that calculates our contrast and brightness heuristics based on the distances and angles of nearby streetlights relative to each crosswalk.
There were several helpful tools we used for data handling and visualization. DuckDB is a geospatial database tool that enables us to build and query our datasets. DBeaver is an application we used to view and manage parquet files. For data visualization, we used Kepler.gl, a prebuilt geospatial visualizer, to create our final contrast map.

The reliability of our algorithm is highly dependent on how updated the datasets are, since outdated datasets will produce misleading results. During our validation, we found several errors in the dataset when comparing them to Google Street View and the phone photos, such as missing streetlights. In our experience, the datasets were more out of date than applications like Google Maps. Therefore, we recommend determining the reliability of the available geospatial datasets before using them.

Another factor to keep in mind is although lighting experts use fairly standardized terminology, differentiating between concepts with similar names but distinct meanings can be confusing. One example is differentiating between illuminance and luminance. Luminance describes the amount of light emitting, passing through or reflected from a surface, while illuminance describes the amount of light falling onto a given surface area. Brightness is actually not a metric that can be objectively measured since it's based on perception.

Since brightness is harder to define, we found it difficult to isolate brightness as a factor independent of overall visibility just based on human perception. With that said, we noted that adequate brightness can compensate for a negative contrast crosswalk and increase the visibility of the pedestrian. On the other hand, low brightness can make a positive contrast crosswalk unviable. Thus, both contrast and brightness are important factors in determining the visibility of a crosswalk.

One of the key takeaways we found while manually collecting data was that the more time that is spent streamlining the data collection process beforehand, the easier and more efficient it is to be out in the field. Moreover, a clear and defined procedure minimizes risk of data mishandling. The procedure that worked for us was first writing out the order of the crosswalk centerpoints to evaluate (with the crosswalk ID) and which direction the data will need to be taken from. We found that these steps removed the need to choose which crosswalk to travel to and to identify the crosswalk ID after data collection.

One of the key takeaways we found while manually collecting data was that the more time that is spent streamlining the data collection process beforehand, the easier and more efficient it is to be out in the field. Moreover, a clear and defined procedure minimizes risk of data mishandling. The procedure that worked for us was first writing out the order of the crosswalk centerpoints to evaluate (with the crosswalk ID) and which direction the data will need to be taken from. We found that these steps removed the need to choose which crosswalk to travel to and to identify the crosswalk ID after data collection.

We also learned during our validation process that the contrast at a crosswalk can be influenced by factors such as the color of clothing the pedestrian is wearing. During our collection process, we tried to have the pedestrian wear at least one lighter color so that the contrast will be easier to determine in the images. We also noted that, under certain lighting conditions, a single crosswalk environment can exhibit multiple contrasts across the pedestrian. However, to simplify our analysis, we assigned one contrast type per crosswalk. Nevertheless, it is generally agreed upon that positive contrast, where the pedestrian is lighter than the background, has the best visibility while no contrast is the worst case scenario. We also went through all our incorrect predictions to identify the causes behind the inconsistencies. We found that referring back to the original photographs taken during our field data collection proved to be extremely useful, and we were able to explain the discrepancies observed in our contrast heuristic. This process also provided deeper insight into how the algorithm was functioning and how it could be improved.

Our GIS map solution would make it possible to get a better sense of the distribution, architecture, and lighting of thousands of crosswalks in a much shorter time than it currently takes to inventory them and with fewer resources. Currently, after much practice, collecting lighting data for one intersection takes 20-30 minutes and requires at least a portion of the road to be closed down. Our solution avoids impeding traffic and provides relevant stakeholders with a holistic overview of how bad/good lighting is in an area. This enables local lighting personnel to narrow down their range when determining the worst lit areas, hopefully speeding up the entire process of choosing and validating the impact of new projects.

10 Conclusion and Next Steps

After two rounds of data collection, we've compiled information on 66 different crosswalks. There are a total of 10,611 crosswalks recorded in the Boston area.

With our documentation and UI, Volpe could allocate resources toward further iterating on our Contrast Predictor. Two main aspects of our predictor can be explored further: 1) the contrast heuristic and 2) the brightness heuristic.

For the contrast predictor, the most immediate next step could be validating and iterating on our angle feature, where we are considering the impact of angles on the contrast heuristic at a certain crosswalk. We have successfully implemented our first iteration of the angle feature in our contrast heuristic. When we compared the contrast heuristic to the manually collected data we already had, we were able to find the source of every discrepancy that came up, which was mostly due to an inaccuracy in the streetlight data or outside sources of light. If more data is collected, then we can reevaluate the accuracy of our angle implementation in our Contrast Predictor.

In addition to the angle implementation, there are other factors of our contrast heuristic that can be validated further. We currently have thresholds for strong positive contrast, weak positive contrast, no contrast, weak negative contrast, and negative contrast, as shown in Figure 38

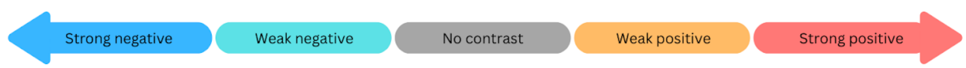


Figure 38: Graphic visualizing the contrast scale

We've set these thresholds based on the two rounds of validation data we collected. If more data is collected, then these thresholds can be fine-tuned even further.

The second aspect of our predictor that could be further explored is our brightness heuristic, since there is currently no strong correlation between the brightness heuristic and perceived visibility. There are two possible avenues to explore in regards to the brightness heuristic, the first being the brightness heuristic function. Our heuristic for a crosswalk is currently equal to the sum of the inverse-squared distances of the nearby streetlights, shown in Figure 39. However, this function causes the brightness heuristic to be quite low for streetlights that are 4 to 8 meters away, even though that range is considered quite close. Reevaluating this function to make it better fitted to our validation data could improve the performance of our Contrast Predictor.

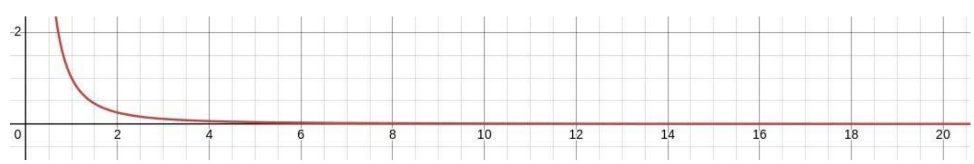
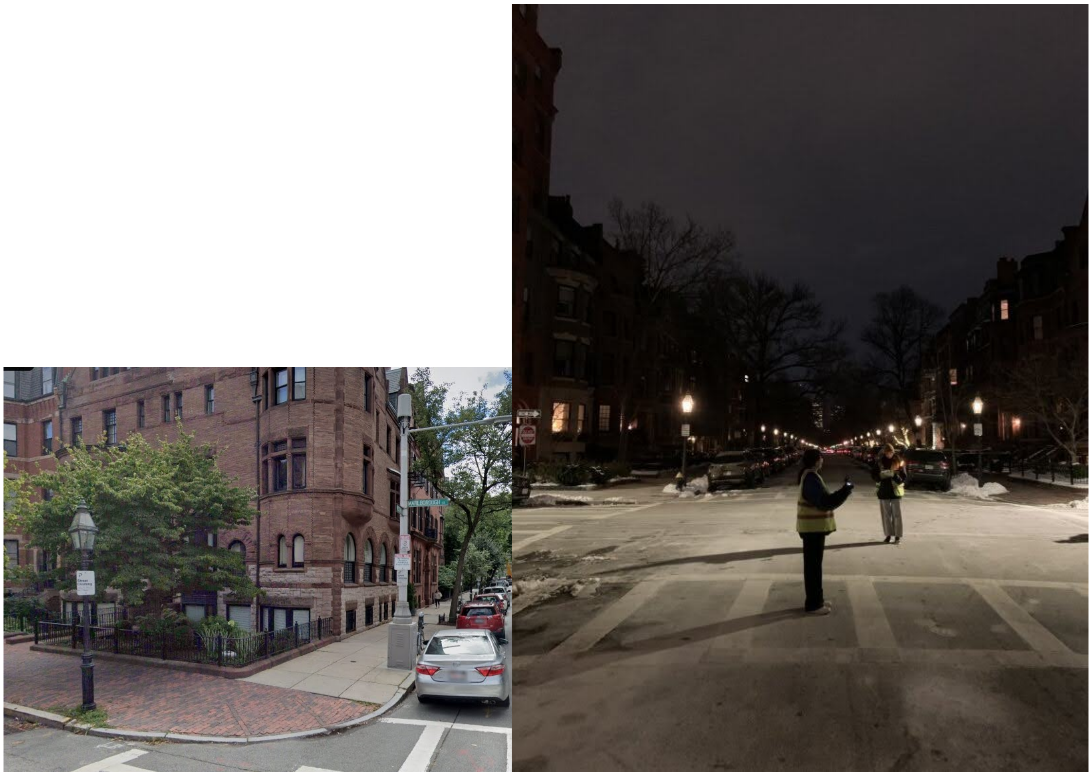


Figure 39: Current brightness heuristic function (the sum of inverse-squared distances of nearby streetlights). The y-axis is the brightness heuristic, the x-axis is the distance of the streetlight from the crosswalk.

The second avenue to explore regarding the brightness heuristic is adding a multiplier depending on the type of light fixture. While validating our data, we learned that although some crosswalks would have a large number of neighboring streetlights, the perceived visibility of the crosswalk would still be quite low. After further analysis, we concluded that the light emitted by the acorn and ball globe light fixtures was weak due to its omnidirectional light design, as seen in Figure 40. On the other hand, crosswalks with rectilinear light fixtures were very bright, likely due to its focused and direct light design.



(a) Acorn light during daytime *(b) Acorn light during nighttime*

Figure 40: Acorn light strength

There's a total of 95,515 streetlights in the Boston Vision Zero streetlight dataset. There are 18 types of light fixtures listed in the database, excluding the lights listed as NA, Other, and No Fixture. Approximately 32% are listed as NA (not available), 12% are acorn lights, and 19% are rectilinear. If the brightness of different light fixtures could be determined, a multiplier for light fixture type could be added to the algorithm so that the Contrast Predictor is more accurate.

After exploring one or both of these avenues, we can more clearly define the thresholds for strong versus weak brightness. We've determined that any light with a brightness heuristic of 0.05 or higher is definitively bright. Lowering the threshold to 0.03 captures most crosswalks with an estimated brightness of 3 or above—except for those circled in red in Figure 41. Each of these crosswalks were incorrectly classified as being of "low" brightness even though during validation they were expected to be "high" brightness. Two hypotheses for this are either because the current algorithm for brightness calculation weighs streetlights within 6-10m too low or because differences in wattage between light types is not accounted for.

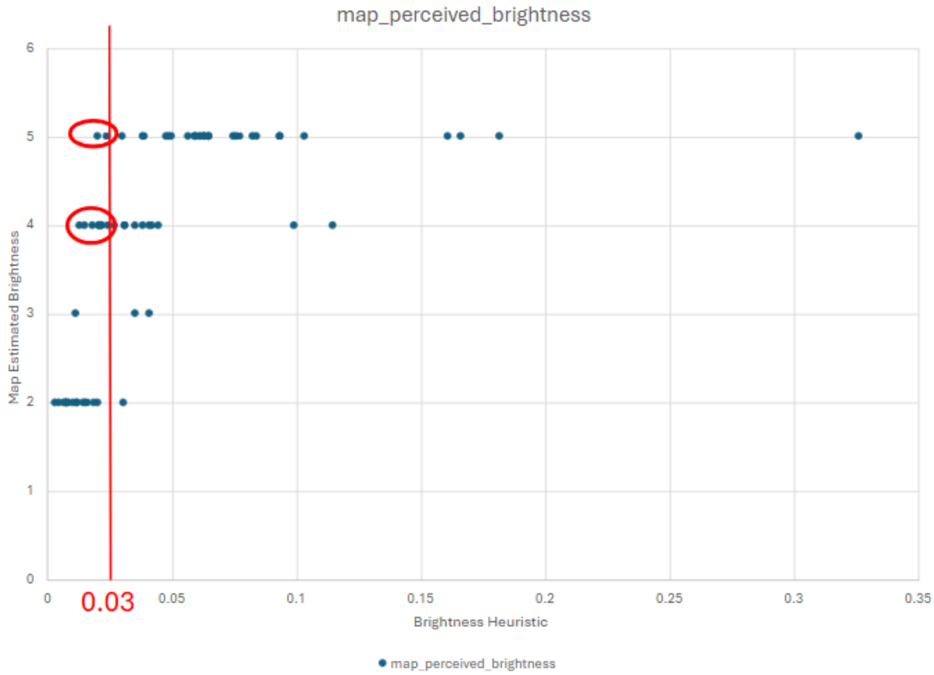


Figure 41: Current threshold for brightness is at 0.03. The points circled in red are crosswalks that are supposed to be bright but have a low brightness heuristic value.

More validation data would be useful in adjusting the value of the threshold for adequate brightness at a crosswalk.

Another area of potential development for the Contrast Predictor is the prioritization of crosswalks. While our algorithm currently focuses on determining the visibility of a specific crosswalk through a contrast heuristic and a brightness heuristic, we've also laid the groundwork for incorporating socioeconomic data. This includes factors such as speed limit from MassDot, population density and median household income from the 5 Year American Community Results Census Data, and pedestrian generators such as schools and hospitals from Open Street Map. The project WalkFirst utilized these variables in San Francisco to provide a more holistic view of crosswalk prioritization.

To further refine our Contrast Predictor and expand its applicability, Volpe can share information with other local governments that have lamppost and crosswalk data, such as Washington DC. For the tool to function, access to datasets like the ones we used for Boston—streetlight locations from Vision Zero, crosswalk data from the UMass Artificial Intelligence Crosswalk Detection Framework, and road segments from the City of Boston Street Address Management (SAM) system—would be needed. This ArcGIS data includes streetlight locations, crosswalk locations, and road segments. Washington DC would be an ideal city to test our Contrast Predictor on since the city has similar data on streetlights, crosswalks, and road segments, which could be incorporated into the framework.

For a more complex approach, the MassDOT LiDAR PathPoints and their distance calculating feature could be utilized in place of our current streetlights database and distance calculating function. We can use LIDAR data to enhance the precision of our prediction algorithm. PathPoints overlays both geospatial and camera imagery, allowing us to build a more detailed and accurate model. We can use this data to pinpoint the type of light fixture and the exact location of each light source. Additionally, we can obtain precise measurements of streetlight heights and the widths of roads and crosswalks.

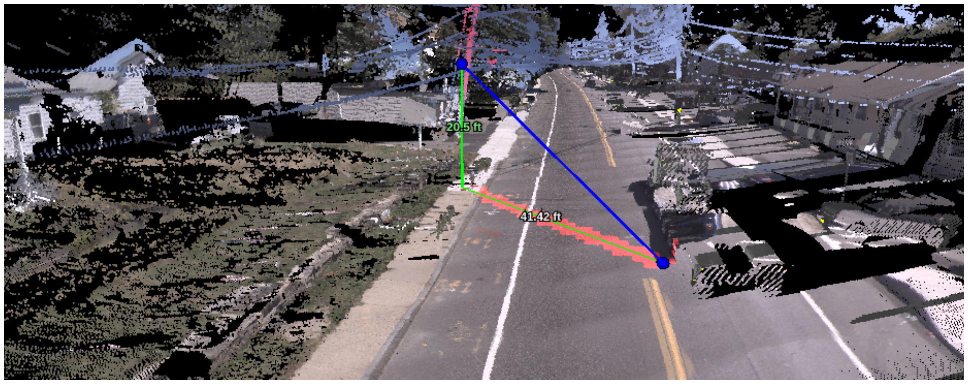


Figure 42: Lidar imaging example

In Figure 42, LiDAR data was able to identify the precise location of the light bulb or light source, the width of the road, the height of the streetlight, and the distance from the light source to the start of the roadway. If similarly extensive LiDAR coverage is available for other cities, we can also account for additional objects, such as trees, poles, or signage, that may obstruct the streetlight and impact crosswalk visibility. This added layer of spatial awareness can lead to more accurate assessments of pedestrian safety and lighting effectiveness.

11 Acknowledgements

Special thank you to all of our wonderful sponsors from Santos Family Foundation: Anne Stuart, Paul Santos, and Leonard Santos; our supportive liaisons: Alex Epstein (Volpe), Eric Englin (Volpe), Gary Baker (Volpe), Alison Link (Volpe), Alyssa Brodeur (Volpe), Derek Voight (FHWA), Paul Lutkevich (WSP), Chris Leone (WSP); and our amazing advisor, Lawrence Neeley!

Special special thanks to Alex and Eric, who have served as our main liaisons throughout this process. We appreciate all of the time, patience, knowledge, and advice you've put into this project.

We're proud of what we created and ultimately hope that it will help municipalities in the future.

12 Appendix

Github Repository

https://github.com/olincollege/night-light

This is our code repository, where the following can be found:

- Code to generate analysis on crosswalks
- Example code to run on Boston Data
- Instructions on how to use and run the code
- Website documenting the code base

Manual Data Collection Procedure

Preparation

- Decide which crosswalks to go to
- Double check that all of the crosswalks are not located within a gated community. If they are, choose other crosswalks.
- Determine which sides will be the A and B sections of each crosswalk.
- Mark all crosswalks to go to on Google Maps. We would recommend putting them all into an album to organize it all better.
- Add all crosswalks to the data collection sheet for a streamlined data collection process.
- Determine who is fulfilling which role ahead of time.

Supplies:

- Safety vests (5)
- Clipboards (2)
- Pens (4)
- Laser distance measurement tool (1)
- Illuminance meter (4)
- 2 light meters should be taped together so that both can read in opposite directions. Ensure that the buttons and sensor are not covered.
- Phone camera (2)
- Data collection sheets (4)
- Car for driving (1)
- A great team (5)

Roles (5 people)

Table 11: Data Collection Roles

Person	Role	Description
1a	Lookout	Watching for cars
1b	Guide	In charge of getting us to the right crosswalks
1c	Driver	Taking the team in between crosswalks
2a	Scribe	Listens for pedestrian's lux values and records on sheet
2b	Lookout	Watching for cars
3a	Pedestrian	Holding light meter in crosswalk
4a	Contrast/brightness determiner 1	Determines the perceived contrast and visibility
4b	Photographer 1	Photographs the pedestrian crossing from the car's perspective
5a	Contrast/brightness determiner 2	Determines the perceived contrast and visibility
5b	Photographer 2	Photographs the pedestrian crossing from the car's perspective

Procedure

Road Condition

The road must be clear of cars, and it must be safe to cross. There cannot be any excess lighting coming from oncoming vehicles and headlights to allow for a more accurate prediction solely based on street lighting.

Collecting Data

1. Role 1 identifies which side of the crosswalk is Center Point A.
2. Role 3 stands in the identified portion of the sidewalk.
3. Roles 4 and 5 stand facing the crosswalk on the side that oncoming traffic would be coming from to view the pedestrian from the car's point of view.
4. Roles 4 or 5 use the distance measurement tool to stand exactly 10 meters away from the pedestrian.
5. Roles 1 and 2 stand on the farthest side of the crosswalk from the pedestrian, ensuring that there is a clear view of the flow of traffic in both directions (if it is a 2-way street). Since Role 2 is in charge of writing down information, they are free to move wherever to still have a good viewpoint of traffic but also be able to hear out data points as they're called out.
6. The pedestrian turns the taped light meters on, extends their arm fully, and holds it out in front of them at chest level. The pedestrian presses the button on the light meter to measure the lux.
7. While the pedestrian is using the light meter, either Roles 4 or 5 should be taking a photo of the pedestrian. This is to later corroborate our contrast and brightness values that Roles 4 and 5 determine.
8. Once the image has been captured, the pedestrian reads out the lux reading for the side facing Roles 4 and 5. Role 2 writes this down.
9. The pedestrian then reads out the lux reading for the opposite side. Role 2 writes this down.
10. Roles 4 and 5 will discuss their perceived contrast and brightness to settle on an agreed upon rating. This is to ensure that both individuals are keeping each other consistent with their ratings and avoiding biases.
11. Set up Roles 4 and 5 on the other side of the crosswalk (Center Point B) from the car's POV. Repeat the procedure.

** Having two people give assessments helped ensure results were more accurate. Further work could use a more formal inter-rater reliability approach.

Lessons Learned

Preparation:
The more time you put into preparing for the manual data collection, the more straightforward the process will be. Google Maps was especially useful to mark our path. We spent hours beforehand determining:

1. Crosswalk locations and IDs
2. Most efficient routes
3. How to distinguish centerpoints A vs. B
4. Most efficient roles

During our first couple of data collection sessions, we weren't able to get as much meaningful data because we weren't going in with a full plan about which specific crosswalks we were going to and a standardized guideline of what contrast and brightness rankings we wanted to determine. This influenced the subsequent data collection tables we generated.

We also learned that it's important to determine if certain crosswalks are placed within gated communities – during our last session, we marked a bunch of crosswalks to collect data at and they were all within gated communities that wouldn't let us through.

Furthermore, we also found that certain intersections and roads were much too busy, even at 10pm on a weeknight. It would be useful to double check when selecting crosswalks that certain roadways are not main roadways and/or if the speed limit is above 25 mph to determine the feasibility of the crosswalk. We required 5 minutes at each crosswalk, so the crosswalk needs to be in a relatively quiet area.

As we learned from Paul Lutkevich at WSP, the general uniform for pedestrians in this type of illuminance measurement is all grey. We would recommend the pedestrian wearing all grey in future sessions to eliminate changes in clothing as an effect on perceived contrast or brightness.

Process: The emptiest time to collect data seems to be late at night earlier in the week. We tended to lean towards Mondays or Tuesdays starting around 9:00 pm. By 10:00 or 11:00 pm, most streets were empty enough to conduct data collection. Main roads will continue to be busy until later, though.

Phone cameras and inexpensive, off-the-shelf dash cams do not have good enough video quality and/or do not capture scenes in the exact manner that humans perceive crosswalks and darkness. If you want to utilize the dashcam and software method, you'd likely have to use a very good quality dashcam. For phone cameras, it would be best to turn off the automatic image brightening that many phone camera apps use. These effects brighten even incredibly dark images; we were in a really dark area once, and the image was edited so that the sky looked like it was almost daytime. Effects like this can skew the photograph, which can alter the accuracy when corroborating brightness and contrast.

Furthermore, we found that perceived contrast and brightness differed slightly from person to person within the group. The amount that one can see in darkness will depend on their age, eye strength, and more, so it is difficult to say precisely the lit-ness of crosswalks when determining with human perception. This should be kept in mind when using this data collection method.

We were still testing out methods for determining the best way to record photographs of pedestrians in each crosswalk section. The best way we found for our team was to send the pictures at that crosswalk to our WhatsApp group chat with a caption of the crosswalk ID. This process could definitely be improved.

Overall, five people worked well for our data collection process. The roles were well distributed and there was little down time for most of us.

One thing to improve upon would be a more straightforward data collection sheet. This could outline explicitly what the range of perceived contrast and brightness looks like in an attempt to further remove bias.

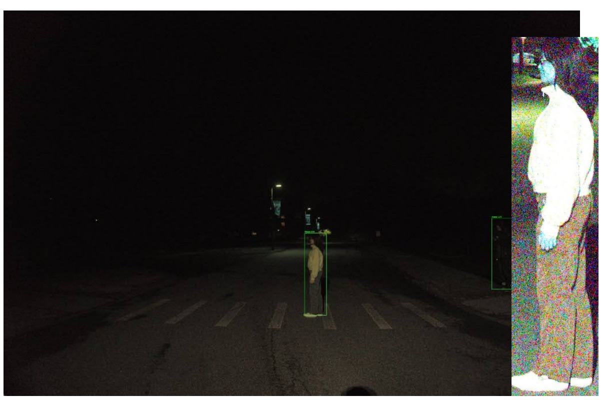
Overall Lighting:
Crosswalk lighting in the suburbs of Boston (Needham and Natick, for our purposes) could use improvement. According to our WSP liaison Paul Lutkevich, most lighting systems in the Greater Boston area were installed before crosswalk lighting recommendations were developed. It was assumed if the street was lit the crosswalk was lit. For new designs, crosswalk lighting recommendations are generally used but re-evaluation of existing lighting systems generally only occurs when safety issues are identified. Many towns are now working to catch up to these new, safer standards, but that takes time, money, and resources.

For future crosswalk lighting projects, it would ultimately be most useful to craft a standard for them to avoid having to collect this data in the future to determine which places need improvements. This is why our work is important – to get one step closer to this standard by building up data points.

Image Processing

Early in the project, the team explored how image processing techniques and computer vision could be applied to photographs of pedestrians in intersections to extract information that could help assess lighting conditions. In this section, we share some of our findings with the aim of supporting future researching who may explore the potential of image processing and computer vision for evaluating crosswalk visibility.

Using the images we compiled in September 2024, we extracted and quantified the luminance of each image in order to obtain a calculated luminance value that could be compared to our collected lux values. To achieve this, we first used the YOLOv8 model, a computer vision model, to identify and isolate the pedestrian within each image. Once detected, the image was cropped accordingly to only include the pedestrian so that external factors such as headlights and traffic lights were removed, as shown in Figure 43.



(a) Original image with pedestrian detected using YOLOv8

(b) Cropped image containing only the pedestrian

Figure 43: YOLOv8 image segmentation sample results

Then, we converted each cropped image from the RGB color space to the YUV color space by using the OpenCV library in Python. The YUV color space separates luminance (Y) from chrominance (U and V), which made it easy to analyze brightness independent of color (shown in Figure 44). After the conversion, we extracted the Y (luminance) channel and calculated the average pixel intensity across the channel to represent the cropped image's overall luminance.

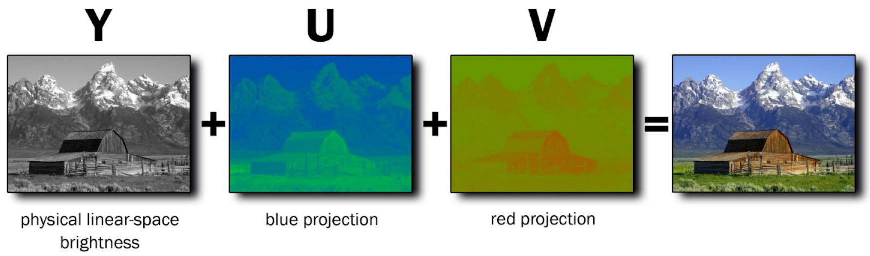


Figure 44: Channels in YUV color space [8]

When we plotted the calculated luminance values against our measured lux readings, we observed a slight positive correlation, though the points were somewhat inconsistent. This is shown in Figure 45 However, since our dataset only consisted of approximately 20 data points, we concluded that hundreds more images would be required to determine a statistically reliable correlation.

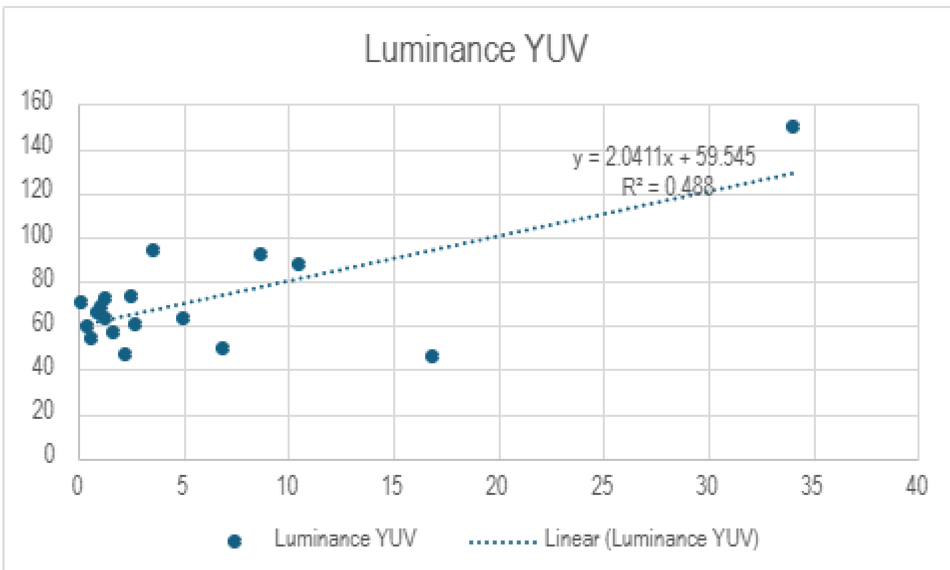


Figure 45: Calculated luminance values vs measured lux values.

An edge case we observed from using this method was that luminance values tend to be low even when the pedestrian was visible due to negative contrast, as seen in Figure 46. Therefore, this algorithm disregards images where the pedestrian is darker than the background.

[8] https://dexonsystems.com/blog/rgb-yuv-color-spaces

Figure 46: Example of edge case image with negative contrast.

To better understand visibility, we established a ground truth based on a group consensus. We evaluated the visibility of the pedestrian in each image based on a scale of 1 to 5. We then plotted the calculated lumiance values against the group consensus, then included the RMS value on a color scale as an indicator for contrast. and found that around 80 luminance was a metric for high visibility, shown in Figure 47. All crosswalks below this value were classified as being not visible, but those higher than this value were a mixed result.

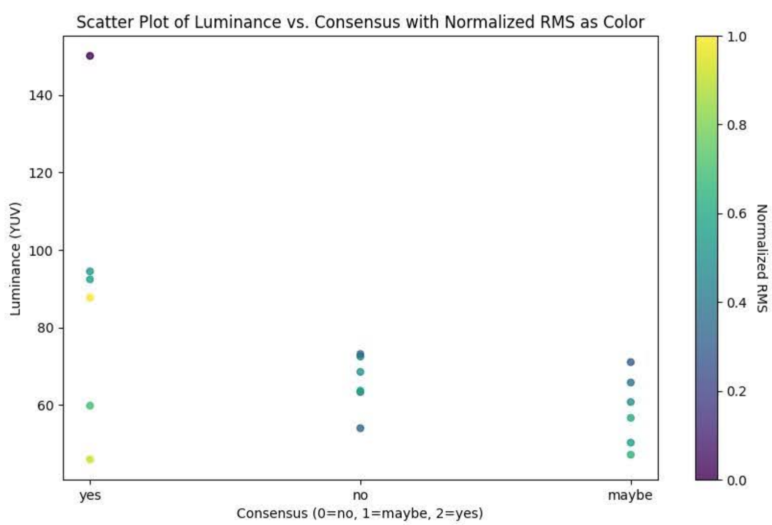


Figure 47: Scatter plot showing luminance vs. consensus of visibility

We also explored edge detection as a method for visibility assessment. Using the same computer vision cropped images, we applied the Sobel operator, a common edge detection technique. We also experimented with the Canny edge detection method, but we found that it didn't work as well as the Sobel operator, possibly due to its sensitivity to variability. We found that the Sobel operator consistently highlighted the edges of the pedestrian in well-lit conditions. For low-lighting conditions, the outlines of the pedestrian would be either be incomplete or indiscernible. This indicates that the Sobel-based edge detection can serve as a useful indicator of visibility, especially for adequate lighting conditions.

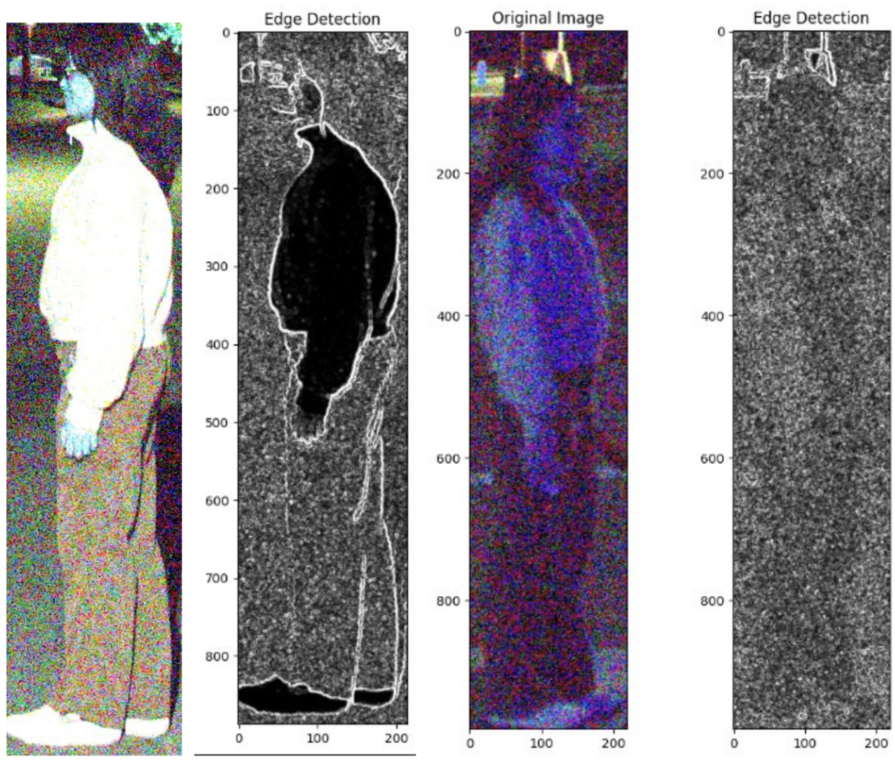


Figure 48: Left to right: (1) Cropped image A with sufficient lighting above 8 lux, (2) Image A with well-defined edges, (3) Cropped image B with insufficient lighting, (4) Image B with hardly no well-defined edges.

Lessons Learned

The edge detection method was effective for images above 8 lux since the edges of the pedestrian would be clearly highlighted. However, this method would output ambiguous results for images with lower-light conditions. Therefore, this tool can be used for confirming when there is a sufficient amount of light, but it's less effective in distinguishing between fair and poor lighting conditions.

There is also a possible correlation between the manually collected lux values and the calculated luminance values. However, due to the inconsistency in the current images, no definitive conclusions can be made yet. A larger and more diverse dataset with at least a couple hundred images would be needed to confidently confirm this correlation.

The images of crosswalks we collected were heavily skewed toward being poorly lit crosswalks. Approximately 75% of the crosswalks in our dataset fell under the recommended 8 lux threshold. This imbalance makes it challenging to prioritize and identify which crosswalks are in most urgent need of lighting improvements.

Contacts

The table below lists contact information for individuals that we worked with consistently throughout the semester. They are familiar with the details of the project if any questions may arise. Alex and Eric at Volpe Center have been our main persons of contact.

Table 12: Contacts

Organization	**Name**	**Email**	**Details**
Volpe Center	Alex Epstein	alexander.epstein@dot.gov	Advisor
	Eric Englin	eric.englin@dot.gov	Advisor
	Gary Baker	Gary.Baker@dot.gov	GIS SME
	Alison Link	Alison.Link@dot.gov	GIS SME
	Alyssa Brodeur	alyssa.brodeur@dot.gov	
FHWA	Derek Voight	derek.voight@dot.gov	
WSP	Paul Lutkevich	paul.lutkevich@wsp.com	Lighting SME Main Contact
	Chris Leone	Christopher.Leone@wsp.com	Lighting SME
Santos Family Foundation	Anne Stuart	annestuart@usa.net	Sponsor
	Paul Santos	paulsantos@usa.net	Sponsor
	Leonard Santos	lensant@gmail.com	Sponsor
Olin College Team	Natsuki Sacks		Team Member
	Allyson Hur	allysonhur@gmail.com	Team Member
	Rucha Dave	ruchadave16@gmail.com	Team Member
	Maya Cranor	mayacranor@gmail.com	Team Member
	Daeyoung Kim	corydykim@gmail.com	Team Member
	Lawrence Neeley	lawrence.neeley@olin.edu	SCOPE Advisor

Potential Stakeholders

Table 13: Contacts

Organization	**Name**	**Email**	**Details**
City of Boston	Michael Donaghy	michael.donaghy@boston.gov	Superintendent of Lighting (Conducted interview on 10/18/24.)
Boston Region Metropolitan Planning Organization	Ali Kleyman	akleyman@ctps.org	Vision Zero Program Manager (Discussed conducting an interview but the team did not have the bandwidth or time.)
MassDOT Highway Division	Jonathan Smith	jonathan.smith@dot.state.ma.us	Point person for PathPoints LiDAR data.

Database Structure

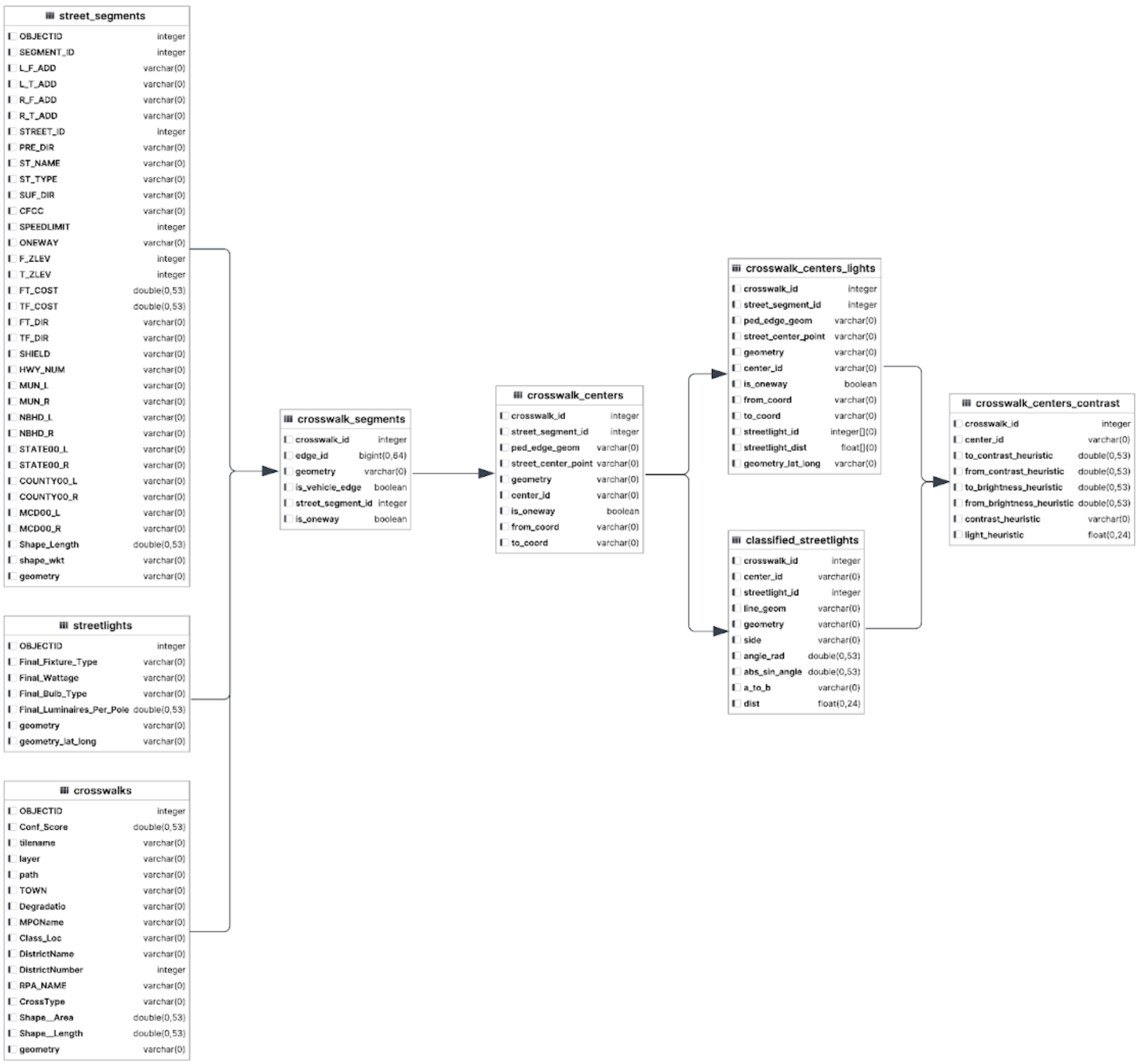


Figure 49: Database Walkthrough

WakFirst

Link to the published paper

Given that we wanted our new direction to be driven by impact, we decided to use WalkFirst as our main inspiration. This was a research project by the City of San Francisco, published in 2011, aimed at improving overall safety and walking conditions in the city. Overall, the city was using this project to reduce severe and fatal pedestrian injuries and increase the number of pedestrians walking in the city, thus increasing the number of automobiles and in turn increasing pedestrian safety.

The project has two main goals:

1. Identify key walking streets in SF.

2. Develop criteria to prioritize pedestrian improvements.

WalkFirst's process of developing criteria involves reviewing existing socioeconomic, road condition, and SF crash data to analyze risk factors that cause collisions. They also emphasize a focus on reducing inequity by taking into account factors like socioeconomic status when prioritizing pedestrian safety infrastructures.

The information that WalkFirst provides helps us to define: our own criteria for prioritizing pedestrian improvements; and pedestrian activity categories and factors.

Another reason for following this WalkFirst-inspired path was based off a user interview that we had with Michael Donaghy, the Superintendent of Lighting for the City of Boston. With the help of our sponsors, we had identified Michael as a potential end user of our hypothetical product and reached out to him to hopefully provide some insight on what sort of tool he might end up using to figure out which areas to prioritize for lighting funding.

Michael had highlighted a particularly inequitable process within the City of Boston, called the 311 call-in system. The 311 system allows Boston citizens to call into the City of Boston and voice concerns about parts of the city that need lighting attention. This system prioritizes lighting projects based on chronological order, possibly to be more "equitable." However, it has the opposite effect; the city sees a higher number of calls from more wealthy areas of Boston, shifting funding towards them. Without any analysis and informed prioritization scheme, the current system is inequitable. WalkFirst helps to fix that.

Furthermore, we found in our field data collections that a majority of places have terrible lighting. Since lighting contractors are not required to follow a standardized distance/angle procedure, there are overall inconsistent lighting patterns throughout the city. Since most crosswalks have poor lighting, we figured it would be more impactful to consider other non-physical aspects to hopefully further distinguish which crosswalk improvements would have the most meaningful impacts.

The various categories that we utilized to characterize crosswalks are listed below in Figure 50. We took these from WalkFirst, sifting through all the different factors they were considering and seeing which ones we had Boston data for.

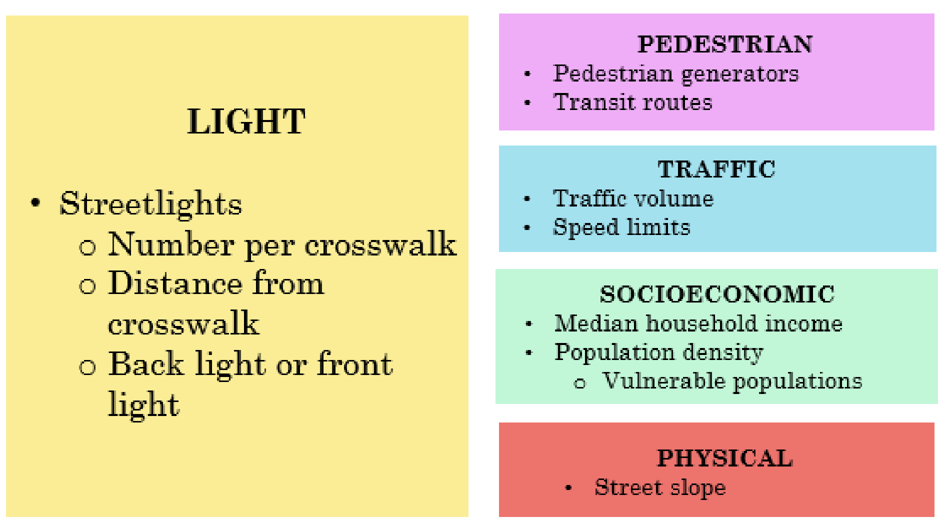


Figure 50: Dataset overview for implementing WalkFirst with Boston data

The end goal for our database was something that end users could utilize to easily understand the most impactful crosswalk projects. There would be a score that weighed all the categories differently, depending on which categories the user identified as important. This way, our tool could be customized to the use cases of different users. We would have our default weights on each of the categories to start as shown in Figure 51.

Crosswalk (lat, long)	Nearby Lamp Posts	Lighting based on Lamp Post Distance	...	Median Household Income	Speed Limit of Road	Score (1-10)*
(42.297664, -71.253976)	1	2		40000	25	8.78
---	---	---	---		---	...
(40.384033, -70.598021)	2	7	---	97000	20	2.31

Figure 51: Dataset overview for implementing WalkFirst with Boston data

We were able to find a number of useful databases for our implementation:

<u>Traffic Dataset</u>:
MassDOT provides a convenient dataset that includes average annual daily traffic (AADT) counts for most roads in Massachusetts. The counts will be used to inform the risk of a pedestrian being hit by a car at a given crosswalk. The dataset can be viewed here: MassDOT Traffic Dataset.

To estimate the danger of a road, we also need a dataset with information about speed limits per road. MassDOT provides a database (updated yearly) with speed limits for most roads in Massachusetts. This can be viewed and downloaded at this link here.

To get a list of traffic lights, which have a correlation with safer crosswalks, we look into BostonMap's GIS. Although MassDOT does provide a database for the state of Massachusetts, this is pretty sparse and less useful. The dataset used can be found at this link here.

<u>Population Density Dataset</u>:
The population density data is sourced from the American Community Survey (ACS) 5-year estimates, which provide reliable demographic information on U.S. populations. The ACS data is integrated with geographic data from census tract boundaries to compute the area of each tract and derive population density values.

Density Calculation Methodology:

1. Geographic Area Calculation: Each census tract is represented by a polygon geometry, and its area is calculated in square kilometers. To ensure accuracy, geometries are reprojected to a projected coordinate system (EPSG:3857) before calculating areas.

2. Population Segments: The dataset includes specific population segments based on age and disability status:

   - Total Population: Total number of individuals residing within each tract.
   - Senior Population (65+): The population aged 65 years and older.
   - Youth Population (0-17): The population aged 17 years and younger.
   - Disabled Population: Individuals identified as having a disability.

3. Density Calculation: Population density for each segment is calculated by dividing the total population or segment-specific population by the area of each tract in square kilometers. This results in values representing the density of each population segment (e.g., seniors per km^2).

Income Dataset:
The income dataset is also sourced from the American Community Survey (ACS) 5-year estimates. The dataset includes median household income data for each census tract within a specified state and year. The data is used to analyze the relationship between income levels and pedestrian safety, as well as to identify areas with possible infrastructure inequity.

Past Accidents Vision Zero Dataset:
We have chosen to use the pedestrian accidents dataset from the Vision Zero initiative in Boston. The dataset contains information about the location of accidents, the date and time of accident, and the severity of the accident (injury/fatality). The dataset is used to identify high-risk areas for pedestrian accidents and to inform the prioritization of crosswalks for safety improvements. The dataset can be viewed at the Vision Zero Dataset.

Vision Zero initiatives are a nationwide effort to eliminate traffic fatalities and severe injuries. Growing number of cities have contributed to this effort and collected data, which will help this project be applicable outside of Boston as well.

Pedestrian Generators:
Pedestrian generators are establishments and attractions that will draw a flow of people. We focused on making catogories that target light night foot traphic and higher risk populations. To collect the data we used Open Street Map, which has a lot of crowd sourced data. The data is not a complete set, but is a good starting place to look at this attribute. The following are the catogories that we looked for:

- Tourist Spots: Tourism areas, museams, art galleries, attractions, viewpoints, zoo, theme parks, memorials, monument, historic sites
- Schools: Schools, university, college, language schools
- Health Facalities: Hospitals, clinics, nursing homes, doctors, dentists, pharamcies, disability services, social facilities, healthcare
- Open Spaces: Parks, nauares reserve, gardens, recreation grounds, playgrounds, grass
- Shopping: Shops, markets, supermarkets, convenience stores, department stores, clothe stores, shoe stores
- Night Life: Bars, Pubs, Nightclubs, casinos, cocktail bars, beer gardens, dance centre, drinking water
- Open Spaces: Parks, natures reserve, gardens, recreation grounds, playgrounds, grass
- Restaurants: Restaurants, fast food, food court, ice cream, pizza

Here is more information on each of the tags on the OSM Website.