

Geolocating and Grading Crosswalks using Deep Learning

Elliott Marsden¹, Gabe Kudirka¹, David Sacharny², and Thomas C. Henderson¹

Abstract—Many Departments of Transportation (DOTs) in the United States have incomplete accounting of the location and condition of their crosswalks. Deploying inspection crews to inspect crosswalk conditions is costly and time-consuming. A complete inventory of crosswalks and their associated paint condition would allow maintenance cycle optimization and reduce inspection costs. Using aerial imagery from the Washington, D.C. metropolitan area, a Mask R-CNN was trained to identify, measure, and localize unique object instances of crosswalks. These crosswalks will serve as the reference dataset for ongoing crosswalk quality assessment. A similar network was trained to identify crosswalk instances in consumer dashcam imagery. Crosswalks extracted from these images are used to update the quality assessment of the reference dataset. We report the data collection process, approaches to paint quality estimation, and deep learning implementations. Crosswalks were located and cataloged using aerial imagery with a precision of 0.968 and a recall of 0.949 within a test region. A quantitative paint quality metric was successfully established and calculated from street-level imagery. By attributing these scores to localized crosswalks, we are able to automatically provide city-wide quality assessments. This method serves as a promising technique for DOTs to continuously monitor the condition of their crosswalks and spare valuable department resources.

I. INTRODUCTION

Crosswalks are among the most high-traffic painted regions of roadways and are essential for maintaining safe pedestrian street crossings. Their paint is susceptible to wear from motorized vehicles, bicycles, foot traffic, roadway construction, and environmental effects. Clear and obvious crosswalks encourage pedestrian usage and prevent vehicles from stopping too far into intersections. For these reasons and more, DOTs are interested in maintaining their crosswalk populations and spend substantial department resources locating and repairing worn crosswalks. Maintaining and accounting for a city’s vast network of crosswalks is difficult given their abundance and dynamic generation and removal. Manual crosswalk paint quality evaluation is a qualitative measure, preventing an unbiased condition-based approach to maintenance scheduling. This work attempts to alleviate this burden by leveraging publicly available aerial imagery (see Figure 1) and industry-sourced dashcam imagery (see Figure 2) to locate and evaluate crosswalks in the Washington, D.C. metropolitan area.

Aerial imagery is taken on an annual basis and provides an almost orthogonal projection of the terrain. Moreover, the crosswalks in the image are generally captured in their totality and in clear weather. On the other hand, dash-cam

imagery gives a view from a camera angle almost tangent to the ground surface and thus has a gross perspective distortion. In addition, there is little control over weather, lighting, or the presence of occluding objects in the scene (e.g., vehicles, pedestrians, leaves on the road, etc.).

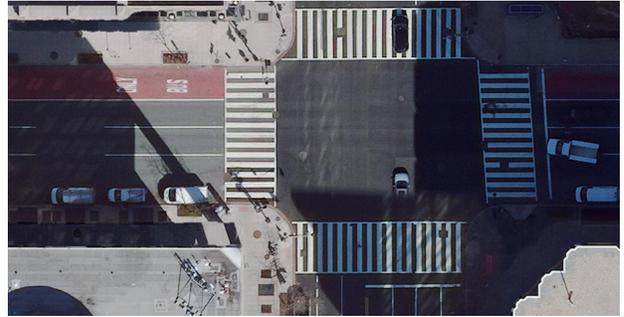


Fig. 1. An aerial image of an intersection in Washington, D.C. Four zebra-style crosswalks are seen as white stripes across each of the roadways.

This pilot-study location was selected due to data availability and a relatively small geographic area. While many crosswalk styles are utilized within the D.C. area, this work focuses on locating zebra-style crosswalks (see Figure 1). This style was selected due to the characteristic shape, color, and susceptibility to wear of such crosswalks.



Fig. 2. An example consumer dashcam image from Washington, D.C. metropolitan area. A zebra-style crosswalk is seen in the center and bottom right of the image.

The objective of this work is to locate and characterize all crosswalks in the D.C. area using aerial imagery and to evaluate crosswalk paint quality using consumer dashcam imagery. While aerial imagery is preferable for both purposes, such data is collected on a yearly time scale and would be insufficient for capturing paint degradation. Dashcam imagery is routinely captured by consumer vehicles on a daily basis, and image metadata like GPS location and heading can indicate which previously localized crosswalks

¹University of Utah, Salt Lake City, UT tch@cs.utah.edu

²Chief Technology Officer, Blynscy, Inc., 175 W 200 S STE 1000, Salt Lake City, UT david.sacharny@gmail.com

are in the image field of view (FOV). This approach is unique in that it combines information gleaned from two disparate image perspectives to create a near real-time paint quality estimation for maintenance schedulers. The work presented herein is expected to drastically reduce crosswalk maintenance misappropriation and cost.

II. PREVIOUS WORK

Detection and localization of painted crosswalks from aerial imagery have a number of use cases in city planning, road maintenance, and assisting the visually impaired [1]–[5]. Crosswalk localization requires two main components, collection of aerial imagery, and crosswalk detection within these images. Both components of crosswalk localization have been the research focus for some time.

For the problem of aerial imagery data collection, several authors [6] [7] rely on annotations from publicly available satellite imagery platforms such as OpenStreetMaps to discover potential locations for crosswalks. However, from our analysis, these crowd-sourced annotations were unreliable and did not provide sufficient coverage of all possible crosswalk locations. Zhang et al. [8] poses a more thorough data acquisition method that involves using existing roadway data to collect imagery along the roadways in a given region. We implement a variation of this approach for data collection.

Earlier approaches to crosswalk detection within images rely on the use of a trained support vector machine classifier on a sliding window within a given image to detect the locations of crosswalks [7]. Barriel et al. [6] instead train a VGG-based CNN for the detection of crosswalks in images, which performs very well for the basic binary classification task of detecting crosswalks but does not locate individual instances of crosswalks in images. Zhang et al. [8] extend this approach by training a similar VGG-based CNN for object detection of instances of crosswalks in images. We improve upon these results by implementing a higher-performing model, Mask R-CNN [9], for this object detection task.

III. CROSSWALK LOCALIZATION

Establishing an initial accounting of a city’s crosswalk population requires a comprehensive view of a geographic region of interest. Fortunately, the D.C. metropolitan area has publicly available high-resolution aerial imagery that is accessible via the aforementioned image collection methodology. Aerial imagery is an excellent medium for locating crosswalk object instances because they are all at least partially visible and the metadata associated with each image allows geographic localization.

Crosswalk localization aims to locate and catalog all crosswalks in a given area, e.g. Washington D.C. This process consists of three main steps, first we collect aerial images of every potential location of a crosswalk at a high enough spatial resolution to be able to distinguish crosswalks. Then we use a trained model to detect each instance of a crosswalk within each aerial image. Lastly, with each crosswalk instance detected, we remove duplicate detected crosswalks

and recover the latitude and longitude coordinates of each crosswalk.

A. Image Collection

Initially, we approached the problem of image collection assuming no previous knowledge about roadways or intersections in the target area. For this uninformed approach, we tested two solutions. The first naive solution was to partition the entire target geographic region into a grid of $100m \times 100m$ tiles and retrieve an image of each tile. Collecting images in this way posed two issues, the first being that for larger regions, this required a massive number of images which greatly increased necessary storage capacity and computation time for downstream tasks, and second, the images taken were not centered around roadways causing crosswalks to be split across multiple images. The second solution was to collect high-resolution zoomed-out images covering large areas and use an object detection model with a high recall to identify possible locations of crosswalks. This solution requires less space and produces higher-quality images than the naive approach. However, it tends to be biased toward intersections and therefore may miss crosswalks between city blocks.

For the majority of major metropolitan areas where the detection of crosswalks is necessary, open-source data exists which captures the locations of all roadways in the region. Using this data, we developed an approach outperforming both uninformed solutions above. The available roadway data is generally a collection of GeoJson LineString objects, each containing the coordinates of endpoints of a road segment between two intersections. For each of these recorded line segments, we save evenly spaced coordinates along its length where the spacing between each coordinate is calculated based on the desired area captured in each image. We then retrieve aerial images centered around each of these coordinates to create a data set of images that cover all of the roadways in a given region. Since crosswalks only occur on roads, this approach allows us to capture images of every potential location of a crosswalk while requiring significantly less space than the naive approach above.

B. Aerial Crosswalk Detection

A Mask R-CNN [9] model architecture was selected for crosswalk object detection due to its ability to provide a bounding box and segmentation mask prediction for each identified crosswalk object instance. The centroid of each bounding box can indicate the object’s geographic location and the predicted mask can serve as a foundation for calculating an initial paint quality assessment. 170 aerial images were collected from around the D.C. area and crosswalks were hand-labeled with bounding polygons. Obstructions to the crosswalk areas (vehicles, people, etc.) were avoided where possible. Of these 170 images, 102 were used for training and 68 were used for testing. A pretrained v2 Mask R-CNN with a ResNet-50-FPN backbone was trained for 10 epochs using Adam optimization [10] and a learning rate of $5 * 10^{-4}$. The final 68 aerial image test set average precision

(AP) and average recall (AR) results are reported in Tables I and II. All detection model evaluation metrics reported in Tables I, II, III, and IV follow the Common Objects in Context (COCO) evaluation metric standard [11]. The AP for detected objects with intersection over union (IOU) threshold scores ranging from 0.5 to 0.95 (with threshold steps of 0.05) is averaged and reported as (**mAP** and **mAR**). Alongside this general metric, AP for detected objects with a single IOU threshold of 0.5 (**AP_{0.5}**) and 0.75 (**AP_{0.75}**) are reported. AP and AR of various detected object area size thresholds are also reported for objects with small pixel areas less than 32^2 (**AP_S** and **AR_S**), medium pixel areas between 32^2 and 96^2 (**AP_M** and **AR_M**), and large pixel areas greater than 96^2 (**AP_L** and **AR_L**). Note that all values of AP and AR for pixel area thresholds are calculated using the aforementioned IOU threshold averaging used to calculate **mAP** and **mAR**.

IOU metric	mAP	AP _{0.5}	AP _{0.75}	AP _S	AP _M	AP _L
bounding box	0.823	0.956	0.926	0.837	0.812	0.836
segmentation	0.837	0.957	0.925	0.803	0.810	0.866

TABLE I
Average Precision Results.

IOU metric	mAR	AR _S	AR _M	AR _L
bounding box	0.853	0.840	0.841	0.865
segmentation	0.862	0.820	0.840	0.885

TABLE II
Average Recall Results.

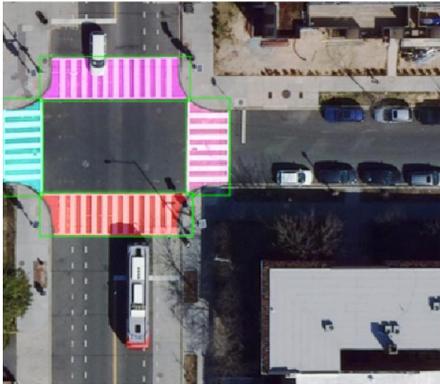


Fig. 3. An example aerial image crosswalk bounding box and mask prediction.

In terms of aerial instance segmentation and bounding box AP results, AP is consistently high across all IOU evaluation categories. Similar results for **AP_{0.5}** and **AP_{0.75}** indicate that most detected objects had excellent agreement between model predictions and ground truth labels. Small detected objects were predicted with a lower precision than

their larger counterparts, but still presented an **mAP** greater than 0.8. The consistently high precision amongst all object sizes is important for localization given that crosswalks vary widely in size and images taken from high altitudes will often result in small crosswalk pixel areas. As for AR results, high recall values are especially important for localization given that a missed crosswalk will be omitted from a cities inventory. The consistently high recall values for all size categories indicates that taking images of a city from a high altitude (reducing the number of images required for full region coverage) is suitable for crosswalk detection and localization.

C. Instance Post-Processing

Once we have detected each instance of crosswalks within the collected images using our trained Mask R-CNN, we then recover each crosswalk’s exact location. Using the coordinates of each image, the area captured by each image, and the pixel level bounding box of each crosswalk instance (see Figure 3), we can recover the coordinates of each detected crosswalk’s bounding box and center point.

One issue we found in our process is that collecting images along roadways produces overlapping images, causing crosswalks to be detected multiple times. To resolve duplicated crosswalks, we first remove collected images that overlap beyond a given threshold before crosswalk detection occurs. Then, after we have recovered the lat/long coordinates of each crosswalk’s bounding box, for any pair of images that overlap, if both images contain a crosswalk whose real-world bounding box coordinates overlap, we remove the detected crosswalk whose bounding box has the smaller area.

Another issue that we found is that our model would occasionally mistake other patterns of white lines as crosswalks, such as solar panels and parking spaces. To address this, we use the open source roadway data and recovered crosswalk coordinates to detect crosswalks that are predicted to be off of a road and remove them. By implementing this procedure we were able to significantly reduce the number of false positives our system detected.

IV. PAINT QUALITY ANALYSIS

The paint quality of crosswalks can deteriorate quickly over time, particularly in areas with high traffic. Because available aerial imagery is often at least a year old, relying on this data to analyze paint quality may result in outdated assessments. To address this we use dashcam imagery provided by Blynescy Inc., which is captured continuously over time to assess paint quality with higher temporal accuracy. To make these assessments, we first register a given dashcam image to the set of cataloged crosswalks that are most likely in view. Then, we use a fine tuned Mask R-CNN model to segment crosswalks in the street-level image and calculate paint quality grades for each predicted region.

A. Street-Level Crosswalk Segmentation

Street-level imagery from consumer dashcams is taken daily all over the D.C. metropolitan area. Their high-frequency capture and GPS location metadata make these

images suitable for a near real-time paint quality assessment. A Mask R-CNN model is especially useful for crosswalk object detection because it provides mask predictions for crosswalk areas that include only paint and roadway pixels (see Figure 4). These areas serve as the basis for paint quality analysis. 235 street-level images were sourced from Blyncsy Inc. and hand-annotated with bounding polygons. All visible crosswalk areas were labeled, and any obstructions were avoided where possible. Crosswalks bisected by obstructions were labeled as separate objects. Of the 235 images, 165 were used for training and 70 were used for testing. The aerial image model architecture and hyperparameters were replicated for street-level object detection. The 70 street-level image test set AP and AR results are reported in tables III and IV.

IOU metric	mAP	AP _{0.5}	AP _{0.75}	AP _S	AP _M	AP _L
bounding box	0.609	0.856	0.673	0.300	0.507	0.796
segmentation	0.617	0.866	0.714	0.269	0.523	0.806

TABLE III

Instance segmentation and bounding box AP results. The disparity in pixel area AP categories was qualitatively observed, with large crosswalks near the front of the vehicle routinely labeled well. Smaller peripheral and background crosswalks were often mislabeled, which is reflected in their relatively low AP values. Fortunately, the frequency of street-level image capture is high enough that detected objects with smaller pixel areas can be filtered and detected within future imagery.

IOU metric	mAR	AR _S	AR _M	AR _L
bounding box	0.654	0.383	0.560	0.835
segmentation	0.658	0.374	0.581	0.827

TABLE IV

Instance segmentation and bounding box AR results. The recall results follow a similar trend to the precision results reported in table III. Large area crosswalk objects are routinely detected by the model but smaller crosswalk regions are more ambiguous and difficult to consistently detect.

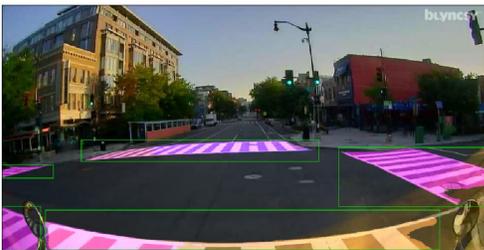


Fig. 4. An example dashcam image crosswalk bounding box and mask prediction.

B. Paint Quality Assessment

The primary goal of the paint quality assessment procedure was to provide DOTs with a simple and easily understandable metric for maintenance schedule optimization. Percent remaining (PR) was selected and is defined by the ratio

of crosswalk paint to the entire crosswalk area (paint and roadway). DOTs across the world use a variety of crosswalk geometries, and the PR of a perfect crosswalk is often municipality dependent. While unreported and relatively variable, the ideal crosswalk PR in the D.C. metropolitan area was determined to be approximately 65 percent. This approximate value was used for all PR calculations reported in this paper.

The output of the street-level Mask R-CNN model is particularly useful for calculating the PR of crosswalk object instances because its mask region predictions prevent the inclusion of obstructions (vehicles, debris, pedestrians, etc.) in the visible crosswalk regions. This leaves mask regions that only include crosswalk paint and their adjoining roadway (see Figure 5). Masks are often predicted outside of the detected object bounding boxes. To prevent the incorporation of these extraneous masked regions, any mask regions with less than 75 percent of their area within any of the bounding boxes are omitted.

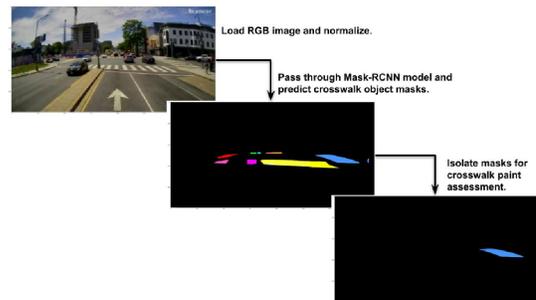


Fig. 5. Mask prediction and isolation process.

Segmenting crosswalk paint from the 8-bit RGB images was accomplished by first converting the image to the YCrCb color space. The Luma (Y) channel was then isolated and a contrast-limited adaptive histogram equalization (CLAHE) was applied to the image. The CLAHE was executed using OpenCV [12], a clip limit of 2.0, and a kernel size of 8x8. A binary threshold was then applied to the equalized image, with pixel values above a value of 170 assigned a value of 1. This process is outlined in Figure 6.

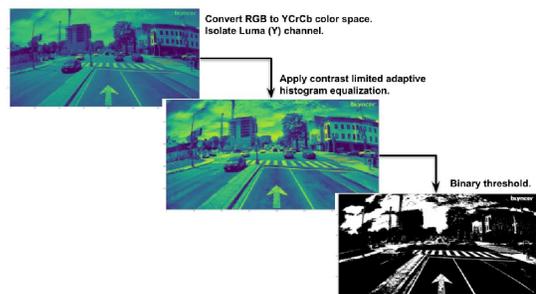


Fig. 6. Street-level image binary threshold process.

Obtaining the PR value for the detected crosswalk object instances is accomplished by combining the model predictions and the thresholded image. Each predicted mask region from the Mask R-CNN model is individually applied to the thresholded image. Division of the masked thresholded image sum and the mask pixel area provide the paint-to-crosswalk area ratio. Division by the ideal PR results in the final PR value for each respective predicted crosswalk region (see Figure 6). If the PR of a detected crosswalk region exceeds the ideal PR it is considered an invalid reading and omitted from any paint quality evaluations (see Figure 7).

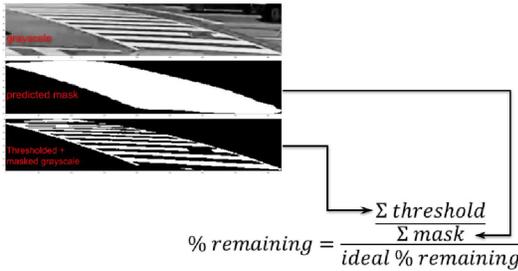


Fig. 7. Example crosswalk object instance percent remaining calculation.

To further simplify crosswalk paint quality comparisons, a color-coded wear rating scale is proposed (see Figure 8). This scale is inspired by the U.S. Forest Service wildfire risk color wheel and is intended to give an easily understandable wear rating to crosswalk object instances. An example assessment of the paint quality within a street-level image is seen in Figure 9.

Wear rating:	Color:	% remaining range:
Invalid region		>100
Low	Green	[80, 100)
Moderate	Blue	[60, 80)
High	Yellow	[45, 60)
Very-High	Orange	[30, 60)
Extreme	Red	[0, 30)

Fig. 8. Proposed wear rating chart.

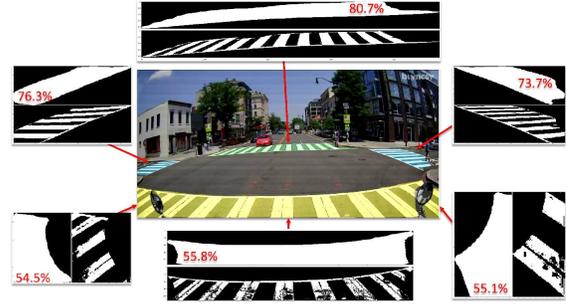


Fig. 9. Predicted crosswalk object instance masks and corresponding thresholded paint regions. Crosswalk regions are color-coded by their predicted paint quality (fig. 8) in the center RGB image.

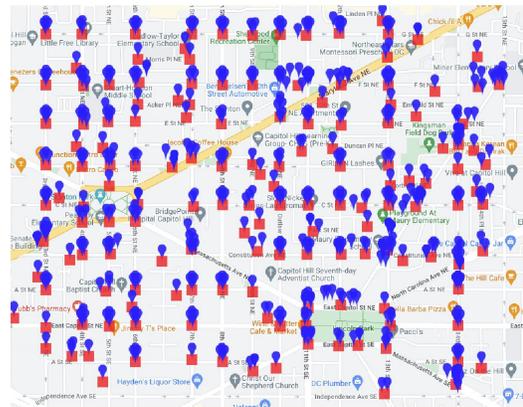


Fig. 10. All crosswalks found through our localization pipeline in our chosen region of Washington D.C. Red squares represent the clusters that nearby crosswalks correspond to.

C. Attributing paint quality estimates to localized crosswalks

Along with each street-level image, we also receive the image’s geographic location, heading yaw, and time of capture which can be used to attribute the calculated assessments to previously localized crosswalks. Initially, our goal was to register each crosswalk instance visible in a street-level image to a cataloged crosswalk from our localization procedure. We believe this could be accomplished by first using a model to predict the distance from each visible crosswalk to the camera in a street-level image, then estimating the locations of each crosswalk using the dashcam’s GPS coordinates. The resulting estimated locations could then be registered to our cataloged crosswalks. Unfortunately, the GPS readings from dashcams have inconsistent accuracy and are liable to be off by up to 30 meters which is too large of a margin of error to register individual crosswalks using the aforementioned procedure accurately.

From our sample data, we observed that crosswalks often exist in clusters at intersections, and the quality of all crosswalks in an intersection tends to degrade at a similar rate. Based on this observation, we felt it would be viable

to attribute the quality of all crosswalks visible in a given street-level image to an entire cluster of crosswalks rather than assessing the quality of each individually. Also, due to the larger area covered by an intersection, by taking this approach, we are able to attribute street-level images to full intersections with a high degree of accuracy despite the limitations of our data.

To determine clusters of crosswalks, we use the density-based clustering algorithm DBSCAN on the localized crosswalks. This clustering was performed using Euclidean distance, an EPS of 30 meters, and one as the minimum number of samples to include crosswalks that do not belong to an intersection. Based on qualitative assessment, we felt this clustering produced clusters that correspond to all crosswalks that may be in the field of view in a street-level image. An example of the clusters found in a sample region can be seen in Figure 10.

With these clusters calculated, we then register each street-level image by finding the cluster whose center point is closest to the coordinates of the image. The paint quality of a street-level image is calculated by taking the average PR of all crosswalk instances within the image, weighted by the area of each crosswalk’s mask. This weighted average is then attributed to its registered crosswalk cluster. To calculate a cluster’s estimated paint quality rating, we take a time-weighted average of all of its attributed paint quality assessments.

V. EXPERIMENTS

To test crosswalk localization, we first selected a smaller region within Washington D.C. to collect data from. We selected the area from 2nd & H St. to 15th & Independence Ave since this is a residential area that we believed would contain many crosswalks and be representative of the full city. We then manually located all crosswalks and crosswalk clusters within our region that are visible in our satellite imagery data. In total, we labeled the locations of 356 crosswalks and 125 clusters for this area. Lastly, we applied our crosswalk localization procedure to this region and compared the results.

To gauge the results of our method, we imported the predicted crosswalk and crosswalk cluster locations in ArcGis and manually inspected each predicted location to determine if it corresponded to an actual crosswalk. Using this approach, we found that of the 349 crosswalks detected by our system, 11 were false positives, and 18 of the manually labeled crosswalks were missed. We also found that of the 127 clusters this approach located, 4 were false positives, and 2 manually labeled clusters were missed.

Past approaches to crosswalk localization have assumed crosswalks to be located at intersections, ignoring other areas in their image collection. To test the validity of our approach of collecting images along all roadways, we implemented the same testing procedure on a baseline where aerial imagery was only taken at intersections. This baseline approach produced 287 crosswalk predictions with 7 false positives,

missing 76 labeled crosswalks. After clustering, this approach produced 105 clusters, where 3 were false positives, missing 23 true clusters. While this baseline does lead to higher precision, it misses significantly more crosswalks than our approach. Precision and recall values of these results are shown in Tables V and VI. Precision is calculated as the number of true positives divided by the sum of true positives and false positives. Recall is calculated as the number of true positives divided by the sum of the true positives and all missed clusters. The F1 score is calculated as the harmonic mean of both precision and recall.

Method	Precision	Recall	F1 Score
All Roadways	0.968	0.949	0.959
Intersections Only	0.975	0.787	0.87

TABLE V

Results for our crosswalk localization approach on crosswalks.

Method	Precision	Recall	F1 Score
All Roadways	0.969	0.984	0.976
Intersections Only	0.971	0.816	0.887

TABLE VI

Results for our crosswalk localization approach on crosswalk clusters.

VI. DISCUSSION

A. Limitations of crosswalk localization

The most notable limitation of our crosswalk localization procedure is that we only detect zebra-style crosswalks and ignore other types, which may also be prevalent in some regions. We have observed that these non-zebra crosswalks tend to have much less paint and, thus, fewer visual cues for our aerial crosswalk detection model. Because of this, accurately detecting these crosswalks requires a significant number of annotated images of crosswalks in this style. If more annotated data on other crosswalk styles are available in future work, our procedure could easily be augmented to account for them.

Another downside to our method is that it can depend highly on the crosswalks’ quality when the aerial imagery is taken. The large majority of crosswalks that our model could not detect were due to heavy wear on the crosswalks in the aerial imagery, making them difficult to detect even for a human. We believe this could be addressed in future work by increasing the annotated images of highly degraded crosswalks.

B. Limitations of paint quality analysis

While the aforementioned paint quality analysis is a good approximation for the PR of crosswalks, there are several limitations to the method that contribute to error. Changes in lighting conditions (day to night, shadows, etc.) make the application of a simple binary threshold inconsistent. This inconsistency sometimes produced a poor street paint segmentation, which misrepresents the actual amount of paint that exists within any given crosswalk area. Future work should include training a semantic segmentation model to

conduct the street paint segmentation task. This could help prevent missed or false paint, increasing the accuracy of the PR calculation.

Due to the perspective of the street-level dashcam imagery, crosswalk objects often appear distorted relative to their true geometries. This distortion can conceal small regions of paint deterioration and may bias clusters towards deceptively high PR ratings. The area-weighted averaging and removal of regions that exceed the ideal PR is intended to mitigate these errors, but further paint study analyses should be conducted to better characterize their effect on PR calculations.

VII. CONCLUSIONS

In this paper, we propose a method for evaluating crosswalk paint quality using satellite and continuous street-level imagery. Our approach thoroughly catalogs and localizes all crosswalks within a region without the use of crowd-sourced annotations (e.g., OpenStreetMaps). We then introduce a novel paint quality measure that can assess crosswalks that are visible in street-level images. These assessments can then be associated with cataloged crosswalk clusters. We believe this two-model approach to localizing and grading crosswalks will significantly increase the efficiency of crosswalk maintenance and reduce associated costs.

ACKNOWLEDGMENT

This work was supported in part by Blyncsy, Inc. The opinions and findings of the authors do not necessarily reflect the view and opinions of Blyncsy, Inc. In addition, we would like to thank the State of Utah for the computational and other resources provided for the Deep Learning in AI and Robotics program.

REFERENCES

- [1] A. Ballardini, A. Saz, S. Lineros, J. Lorenzo, I. Alonso, N. Parra, I. Daza, and M. Soto, "Urban Intersection Classification: A Comparative Study," *Sensors*, vol. 21, no. 6269, September 2021.
- [2] W.-C. Chang, C.-H. Hung, and L.-C. Chen, "Progressive Registration of Image Features and 3D Vector Lines for Orientation Modelling," *The Photogrammetric Record*, vol. 33, no. 16, pp. 66–85, March 2000.
- [3] D. Costea and M. Leordeanu, "Aerial Image Geolocalization from Recognition and Matching of Roads," in *British Machine Vision Conference*, January 2016.
- [4] O. Karatzaferis, "Crosswalk Detection for the Outdoor Navigation of People with Visual Impairment," Master's thesis, GIMA: University of Utrecht, TU Delft, University of Twente, March 2022.
- [5] C. Zegeer, J. Stewart, H. Huang, P. Lagerway, J. Feaganes, and B. Campbell, "Safety Effects of Marked Versus Unmarked Crosswalks at Uncontrolled Locations," University of North Carolina, Tech. Rep. FHWA Publication Number: HRT-04-100, August 2005.
- [6] R. Berriel, A. Lopes, A. de Souza, and T. Oliveira-Santos, "Deep Learning-Based Large-Scale Automatic Satellite Crosswalk Classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1513–1517, 2017.
- [7] D. Koester, B. Lunt, and R. Stiefelhagen, "Zebra Crossing Detection from Aerial Imagery Across Countries," in *International Conference on Computers Helping People with Special Needs*. IEEE, July 2016, pp. 27–34.
- [8] Y. Zhang, C. Zhang, and J. Luttrell, "An Automated System for Pedestrian Facility Data Collection from Aerial Images," University of Southern Mississippi, Tech. Rep. Final Report NCHRP IDEA Project 209, December 2021.
- [9] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," *CoRR*, vol. abs/1703.06870, 2017. [Online]. Available: <http://arxiv.org/abs/1703.06870>
- [10] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [11] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014. [Online]. Available: <http://arxiv.org/abs/1405.0312>
- [12] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.