Laying the Groundwork for Automated Computation of Surrogate Safety Measures (SSM) for Skateboarders and Pedestrians using Artificial Intelligence

Erfan Chowdhury Shourov
Dept. of Electrical and Computer Engineering
San Diego State University
San Diego, USA
erfanchowdhury1993@hotmail.com

Christopher Paolini
Dept. of Electrical and Computer Engineering
San Diego State University
San Diego, USA
paolini@engineering.sdsu.edu

***Abstract*—The use of skateboards for transportation in pedestrian dense areas is becoming more prevalent. Increased use of skateboarding raises the probability of pedestrian-skateboarder collisions and near-collision events. Skateboarders and pedestrians can face significant injury when involved in a collision. New approaches are needed to measure the frequency and predict the potential for collision events, in real-time, as traffic conditions change due to construction or land usage. Surrogate Safety Measures can be computed to assess hazard conditions on roads and sidewalks using deep learning object detection and classification models. We developed a new dataset consisting of over ten thousand images with nearly thirty thousand bounding box annotations of pedestrians and skateboarders at eighteen different camera perspectives.We trained the Faster R-CNN and SSD models with our dataset. While both models were found to correctly classify pedestrians and skateboarders with images containing both classes, the Faster R-CNN model performed with greater accuracy than the SSD model. However, the SSD model was shown to classify at a higher video frame rate which makes SSD a candidate for edge-based detection and classification and lays the ground work for automating the calculations for Surrogate Safety Measures between skateboarders and pedestrians.**

***Index Terms*—Artificial Intelligence, Deep Neural Network, Faster R-CNN, Object Detection, Pedestrian, Single Shot Multi-Box Detector, Skateboarder, SSD**

I. Introduction

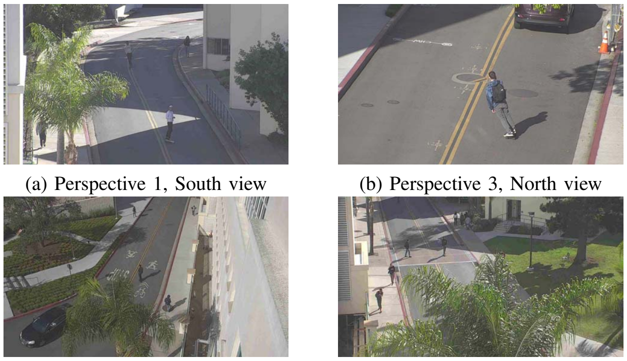
The use of skateboards for short-distance transportation is gaining in worldwide popularity. Use of skateboards in areas of dense pedestrian traffic increases the opportunities for skateboarder-pedestrian close contact or collision. Pedestrians on sidewalks are vulnerable to, and must avoid, rapidly moving skateboarders. One mechanism for estimating regions of a thoroughfare shared with multiple vehicle types are known as Surrogate Safety Measures (SSMs). These measurements provide a probability of near-collision events by measuring spatial and temporal proximity between road users. Skateboarding has been shown to be a significant cause of injury, especially among adolescents in the age range of 5-19 [1]. Relative to the walking velocity of pedestrians on sidewalks, skateboarders travel at much higher velocity, which requires them to perform maneuvers to avoid fixed and moving (e.g. pedestrians) obstacles. When transitioning between sidewalks, skateboarders will often encroach into roadways designated exclusively for vehicular traffic. Traumatic injuries often occur when skateboarders collide with pedestrians or other vehicles [2].

SSMs are numerical metrics that can be used to identify critical safety related events, such as near collision incidences that occur at specific areas on a thoroughfare. Values of SSMs that exceed a hazard threshold can be used to justify the adoption of traffic routing policies, or the redesign of roadways to lower the probability of collision or near-collision events between pedestrians and skateboarders.

The traditional approach taken to asses the safety of thoroughfares with high pedestrian density is to collect and then later analyze a long history of incident data before enacting policy or design modifications. Encounters between skateboarders and pedestrians are rare events, so multiple years of data are typically required before changes in policy are implemented. Moreover, many incidents between pedestrians and skateboarders are required to gain the attention of officials before action is taken to improve safety. Each incident potentially results in trauma or musculoskeletal injury to both the pedestrian and skateboarder [3], [4]. Such a reactive approach may be hindered by modifications to roadways and sidewalks due to changes in land use, which would then impact safety analysis. More active approaches are needed to assess thoroughfare safety, in real-time, to reduce the probability of near-collision incidences and to asses the safety of pedestrians and skateboarders who utilize regions that may structurally change over time due to construction and other land-use demands. One approach taken to asses thoroughfare safety is

This research was funded by the Department of Transportation Project Number: 04-110, *Developing an Intelligent Transportation Management Center (ITMC) with a Safety Evaluation Focus for Smart Cities.*

978-1-7281-8701-3/20/$31.00 ©2020 IEEE
DOI 10.1109/AI4I49448.2020.00011



(a) Perspective 1, South view (b) Perspective 3, North view

(c) Perspective 4, South view (d) Perspective 6, North view

Fig. 1: Images were obtained from eighteen different camera perspectives configured with different pan, tilt, and zoom values. Four selected perspectives are shown in sub-figures (a) through (d).

through the use of artificial intelligence (deep learning models) to detect and classify objects in video frames captured with traffic surveillance cameras.

We have developed a new dataset consisting of 10070 images with 28718 bounding box annotations of pedestrians and skateboarders, captured from eighteen different camera perspectives of pan, tilt, and zoom. Moreover, this is the first skateboarder dataset to be ever collected and published. We have made our dataset freely available to the public [5] for use in training deep learning object detection and classification models. Annotations were made using *VGG Annotator* [6], [7] with two class labels, *pedestrian* and *skateboarder*. Images were captured using a Pelco Esprit® Enhanced Series camera. The camera was mounted on a six-story balcony overlooking an intersection of two streets with sidewalks. A mixture of road users in addition to pedestrians and skateboarders is present, including bicyclists, cars, commercial trucks, vans, scooters, and golf carts. To make the dataset usable by the public in multiple fixed camera scenarios, we captured images at eighteen pan, tilt, and zoom configurations so pedestrians and skateboarders can be detected at different perspectives. Four of the eighteen camera perspectives are shown in figure 1.

II. Object Detection and Classification Models

The Faster Region-based Convolutional Neural Network (Faster R-CNN) and the Single Shot Multi-box Detector (SSD) deep learning models were used for the classification and detection of both pedestrians and skateboarders. The Tensor Flow Object Detection (TFOD) API [8] was used to train our dataset. Faster R-CNN and SSD are two state of the art object detection models. From [9] and [10], the difference between the two architectures is evident; there is a clear trade-off between accuracy and frame-rate. On one hand, due to the high accuracy of the Faster R-CNN model, frame-rate is lower. On the other hand, SSD's accuracy decreases with its shorter inference time.

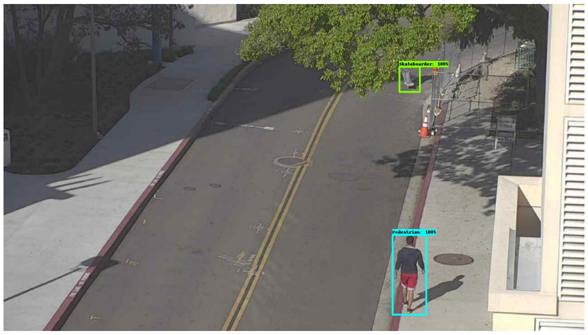


Fig. 2: Annotated frame showing the ground truth containing a portion (green bounding box) of a skateboarder obstructed by a tree and a pedestrian (cyan bounding box).

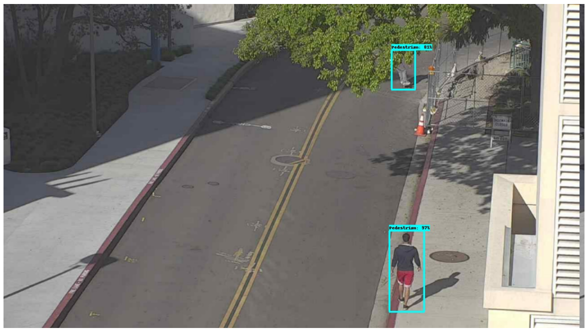


Fig. 3: Faster R-CNN is unable to correctly classify an obstructed view of a skateboarder at the initial stages of training. A skateboarder is misclassified (cyan bounding box) as a pedestrian with probability 81%.

III. Simulation and Results

This section analyzes the simulation environment in which both the models have been trained and evaluates the findings in terms of their respective performance metrics.

Figure 2 illustrates the ground-truth bounding boxes of a pedestrian and skateboarder that was supplied to the Faster R-CNN architecture. Figure 3 shows the detection of the Faster R-CNN model. It can be observed that during the initial stages of training, the network fails to label properly and incorrectly classifies a skateboarder as a pedestrian. However, by the end of training session, the network learns to differentiate between the two class labels. Figure 4 shows the correct detection and classification with a confidence level of 99%. Similar images can also be obtained for the SSD architecture with both the ground-truth level and the network detection from the Tensorboard Visualization Tool, a feature supported by the TFOD API.

The Faster R-CNN model was trained for a total of 200K steps. Figure 5 compares the training loss and the evaluation loss of the testing set. The magnitude of the training loss oscillates around 0.10. The evaluation loss is also around 0.11. Nearly the same magnitude of the training loss and the evaluation loss indicates that there has been no over-fitting in

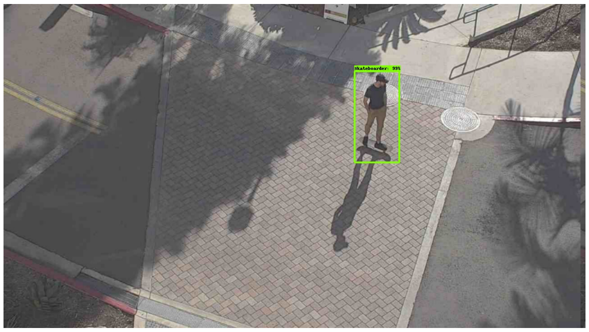


Fig. 4: Faster R-CNN is able to correctly classify an obstructed view of a skateboarder.

the network architecture. The same conclusion can be drawn after observing figure 6, where the training loss is around 0.25 and the evaluation loss is approximately 0.3. The SSD model was trained for a total of 120K steps. The details of the parameters and hyperparameter tuning values are given in table I.

TABLE I: Parameters/Hyperparameters Comparison

Parameters/ Hyperparameters	***Faster R-CNN***	***SSD***
No. of Classes	2	2
Batch Size	1	16
Number of Steps	200000	120000
Initial Learning Rate	0.0003	0.0079
Scheduled Learning Rate	0.00003 after 90K steps, 0.000003 after 120K steps	N/A
Score Converter	Softmax	Sigmoid
Image Resizer	keep_aspect_ratio _resizer (600 x 1024)	fixed_shape _resizer (640x640)
Momentum Optimizer Value	0.9	0.899
Feature Extractor	Faster R-CNN Resnet101	SSD_Mobilenet _v1_FPN
Metrics Set	COCO_detection _metrics	COCO_detection _metrics

Figure 7 demonstrates the mAP of the Faster R-CNN model at 0.5 and 0.75 IOU thresholds (orange and blue lines), respectively. Figure 8 graphically represents the mAP of the SSD architecture on our dataset also at 0.5 IOU (in orange) and 0.75 IOU (in blue). The values for both architectures have been tabulated in table II.

Finally, the frozen weights of the Faster R-CNN and SSD models, after the trained model was used to inference an image that the network has never been trained on, is shown in figures 9 and 10. These results show the Faster R-CNN model detects all pedestrians and skateboarders with a perfect confidence level of 1.00. The SSD model, subjected to the same image, determines the confidence level of the shown pedestrians (from left to right) as 0.50, 0.90, 0.90 and 0.95, respectively, while it determines the confidence level of the skateboarder as 0.78. Even though both the networks correctly

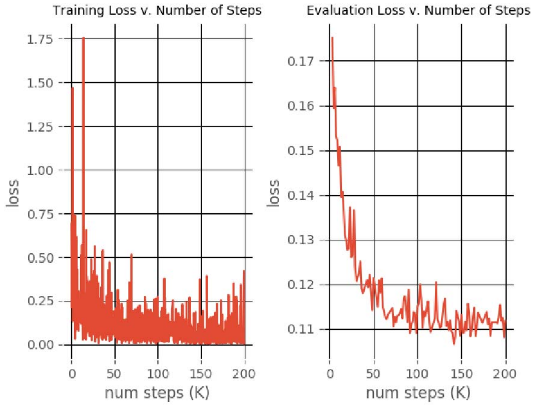


Fig. 5: Faster R-CNN training and evaluation loss as a function of iteration number.

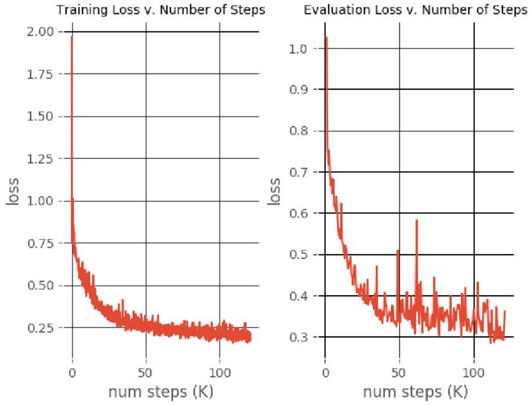


Fig. 6: SSD training and evaluation loss as a function of iteration number.

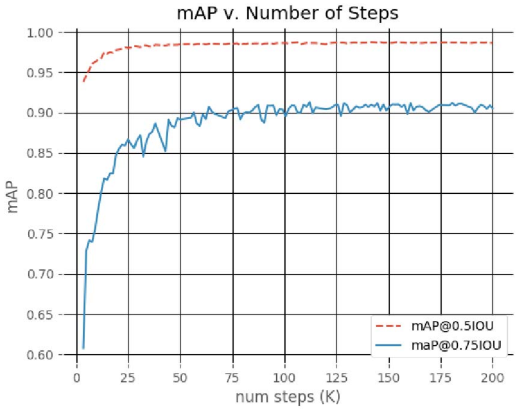


Fig. 7: Faster R-CNN mAP (mean Average Precision).

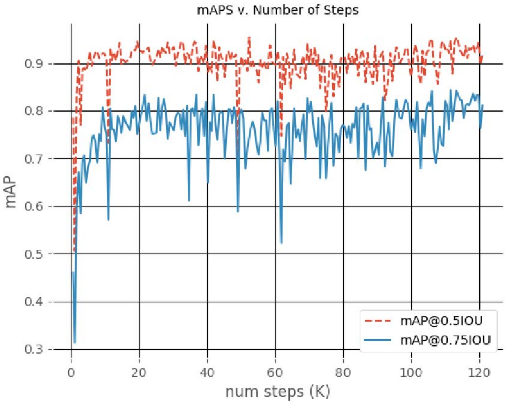


Fig. 8: SSD mAP (mean Average Precision).

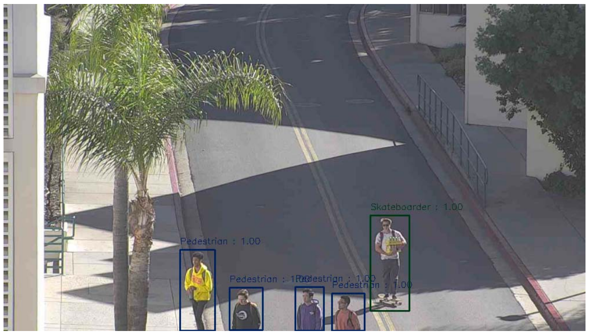


Fig. 9: Faster R-CNN example multi-classification result.

detected and classified pedestrians and skateboarders, the Faster R-CNN model outperforms the performance of the SSD model. However, the fps of the SSD is four times that of the Faster R-CNN model. There are many other pre-trained models that can be used to improve the accuracy of the SSD model, and also to gain a higher fps rate.

A. Conclusion and Future Work

In this work we developed a new dataset consisting of over ten thousand JPEG images with nearly 30 thousand bounding box annotations of pedestrians and skateboarders at eighteen different camera perspectives. This dataset has been released to the public [5] under the Attribution 4.0 International (CC BY 4.0) license, and can be freely downloaded and used by researchers for training and testing their own deep learning-based, real-time, object detection and classification models. We trained the Faster R-CNN and SSD MultiBox models, two well-known and state-of-the-art computer vision object detection models, for detecting and classifying pedestrians and skateboarders, with the final goal of computing Surrogate Safety Measures in real-time. Since our intent is for real-time SSM computation, the limited frame rate of 4 fps of the Faster R-CNN model renders the architecture infeasible. The SSD MultiBox model was shown to classify pedestrians and skateboarders, albeit with lower accuracy, at a video frame rate of 16 fps, which makes SSD a good candidate for edge-based detection and classification applications.

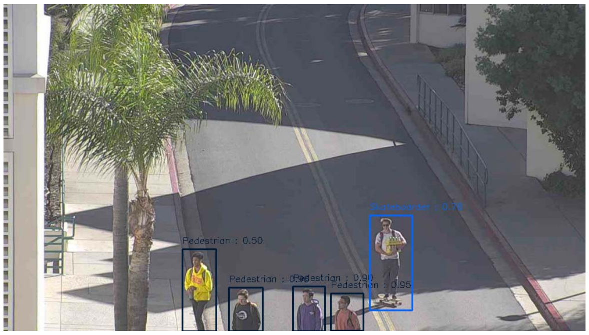


Fig. 10: SSD example multi-classification result.

TABLE II: mAP and fps Tabulated

Model	*mAP (%)*	*fps*
Faster R-CNN	98.3	≈ 4
SSD	93.8	≈ 16

In this paper we have laid the groundwork for the automated computation of SSM for skateboarders and pedestrians and propose to implement the Single Shot MultiBox Detector model on a Google Coral Edge TPU Board for faster and reliable real-time edge-detection, classification, and SSM computation. The Edge TPU Board is a small form-factor (85mm x 56mm) low power (2W) device and can be installed on intersection light posts adjacent to existing video cameras.

References

[1] L. B. McKenzie, E. Fletcher, N. G. Nelson, K. J. Roberts, and E. G. Klein, "Epidemiology of skateboarding-related injuries sustained by children and adolescents 5-19 years of age and treated in us emergency departments: 1990 through 2008," *Injury epidemiology*, vol. 3, no. 1, pp. 10–10, Dec 2016, 27747547[pmid]. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/27747547

[2] J. L. Fountain and M. C. Meyers, "Skateboarding injuries," *Sports Med*, vol. 22, no. 6, pp. 360–366, Dec 1996.

[3] S. B. Kyle, M. L. Nance, G. W. Rutherford, and F. K. Winston, "Skateboard-associated injuries: participation-based estimates and injury characteristics," *J Trauma*, vol. 53, no. 4, pp. 686–690, Oct 2002.

[4] L. Forsman and A. Eriksson, "Skateboarding injuries of today," *Br J Sports Med*, vol. 35, no. 5, pp. 325–328, Oct 2001.

[5] C. Paolini and C. E. Shourov, "Skateboarder and pedestrian dataset," https://doi.org/10.17605/OSF.IO/CQD9Z, 2020.

[6] A. Dutta and A. Zisserman, "The VIA annotation software for images, audio and video," in *Proceedings of the 27th ACM International Conference on Multimedia*, ser. MM '19. New York, NY, USA: ACM, 2019. [Online]. Available: https://doi.org/10.1145/3343031.3350535

[7] A. Dutta, A. Gupta, and A. Zissermann, "VGG image annotator (VIA)," http://www.robots.ox.ac.uk/ vgg/software/via/, 2016.

[8] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," 2015.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.