Identify Urban Risk Factors for Bus Crashes Using Computer Vision Tools

Keywords: transportation, crashes analysis, computer vision, pedestrian behavior, video analysis

Extended Abstract

In road safety research, bus crashes are particularly noteworthy because of the large number of bus passengers involved and the challenge that it puts to the road network (with the closure of multiple lanes or entire roads for hours) and the public health care system (with multiple injuries that need to be dispatched to public hospitals within a short time). The recent paradigm shifts of road design from primarily vehicle-oriented to people-oriented urge us to examine street and pedestrian behavioural factors more closely. Notably, the street environment is highly dynamic, corresponding to different times of the day. However, we still lack the source and tools to capture the dynamic risks factors with high resolution at a city scale.

To fill this research gap, this study leverages a rich dataset - video data from bus dashcam footage - to identify some high-risk factors for estimating the frequency of bus crashes. This research applies deep learning models and computer vision techniques and constructs a series of behavioural and street factors: 1) pedestrian exposure factors, 2) pedestrian jaywalking, 3)bus stop crowding, 4) sidewalk railing, 5) and sharp turning locations. The study area includes 244.36 km of street segments in Hong Kong, covered by 33 bus routes from morning to night. The bus dashcam videos were collected from July 2021 to March 2022. All videos obtained are also accompanied by a GPS file. The GPS file contains the latitude, longitude, timestamp, bus ID, and route ID so that each video frame can be associated with a GPS point. Further, we assign a video frame-GPS-street association for the analysis.

The **pedestrian exposure factor** (P_{kT}) is defined as the pedestrian volume at a basic street unit (BSU) (Yao & Loo, 2012) k through any given time period T (T is usually one hour). To do so, we first leverages a pedestrian tracking algorithm (Wang et al., 2020) that combines the Fast R-CNN (Girshick, 2015) and Deepsort (Wojke et al., 2017) to detect the unique number of visitors that appeared in bus j's video through BSU k during any given time Δ T (Figure 1a). Then we estimate the number of pedestrians observed by a bus j during time T as $P_{jkT} = \frac{P_{jk} \times v_p}{L_k} \times \mathrm{T} \times \mathrm{AF}$, where v_p is the average pedestrian speed (1.203 m/s), L_k is the length of a BSU. AF stands for the adjustment factor considering the variation of the bus speed through different videos. When the bus speed is much lower than the average speed, it will capture more pedestrians. P_{jkT} is adjusted downward by the ratio of v_{bk} and $\overline{v_b}$. Therefore, AF = 1 when $v_{bk} \geq \overline{v_b}/2$, $\frac{v_{bk}}{\overline{v_b}}$ otherwise. The final pedestrian exposure P_{kT} is measured as the mean of the P_{jkT} for all n videos taken. With the unique number of people captured for any given segment of street, we also derived the **pedestrian crowding** C_k as the average the average pedestrian area occupancy (APAO) within a 10-meter buffer of a given bus stop (Figure 1c).

The second task is to capture **pedestrian jaywalking**. We trained a Mask R-CNN model using 200 manually labelled images and applies the model to all videos (He et al., 2017). These labelled samples were sufficient to train the model and achieved relatively good results compared to the original Mask R-CNN coco dataset's mask mAP. Details of the model can be found in Fan & Loo (2021). The manual labelling process identifies a pedestrian's location, shape, and if they are jaywalking or not. Then this model is applied to all bus videos, and the detection results are aggregated to road segments to derive the jaywalking index.

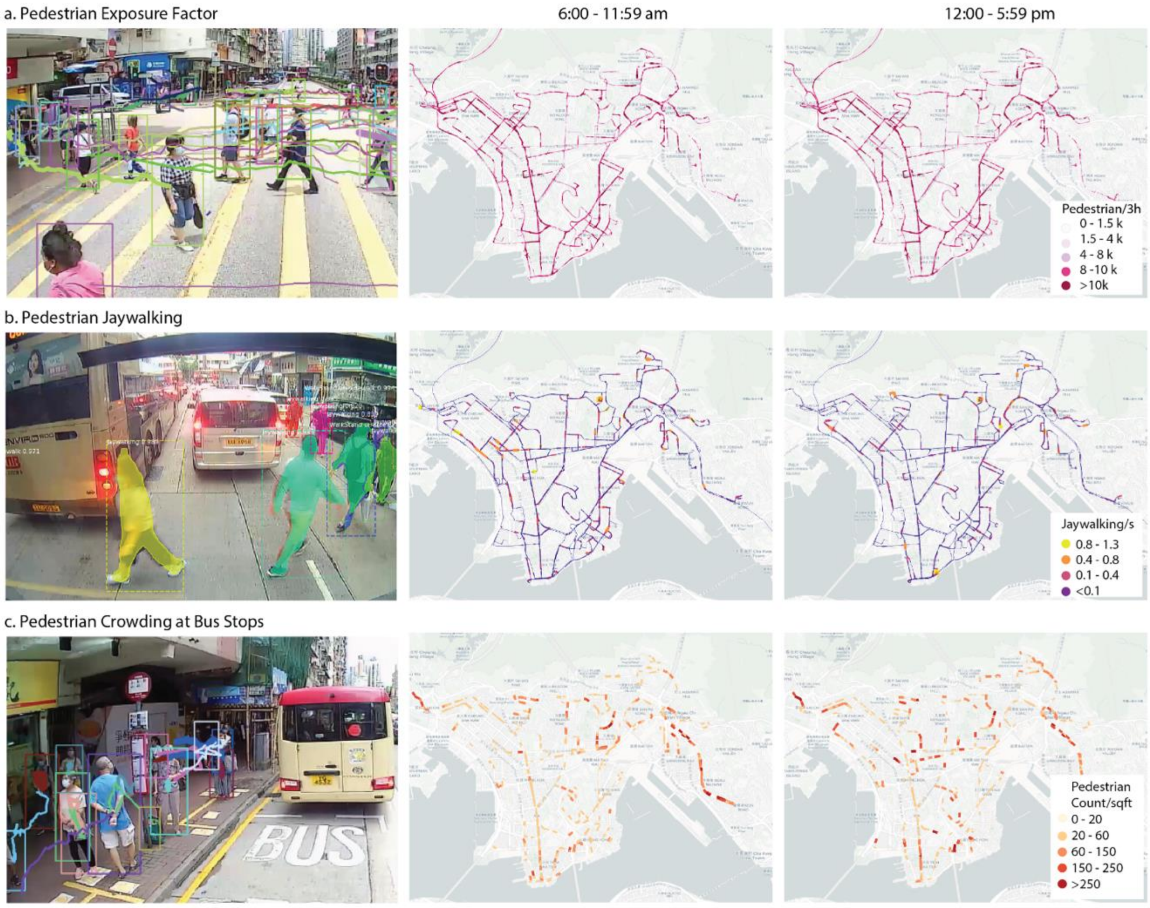

Figure 1 Three behavioural risk factors constructed from the dashcam video. A) Pedestrian exposure. B) Pedestrian Jaywalking. C) Pedestrian crowding at bus stops. Maps of distribution are plotted accordingly at different times of a day.

Then, the **sidewalk railings** are captured via image segmentation model trained on ADE20K dataset (Zhou et al., 2017). **Sharp-turn** locations are directly measured from the road network. The place with fewer than 90 degree indicates a sharp turn associated with the road.

To systematically estimate the importance of the five risk factors, we construct three models to predict bus crashes frequency in Hong Kong. Road crash data from 2015 to 2019 are collected and aggregated to the street segment. A total of 77,060 crashes happened in Hong Kong. This study used 12,679 crashes that were bus-related. Among all bus crashes, 1,408 were fatal or severe, and 11,189 were slight crashes. The three models are 1) Negative Binomial (NB), 2) Random Forest (RF) (Breiman, 2001), and 3) XGBoost (Chen & Guestrin, 2016). Here, NB is used as a baseline to compare with the performance of the other two models.

Comparing the permutation importance (MSE-based) from all three models (Figure 2), we find that the Pedestrian Exposure Factor is constantly the most important in estimating vehicle-pedestrian crash frequency. Nonetheless, Bus Stop Crowding ranks higher for slight bus crashes, while least important for serious bus crashes. Where bus stops are crowded, waiting bus pedestrians and passengers are more likely to overflow to roads and give rise to various slight injury bus crashes. The Railing Index is of much higher importance in predicting serious bus crashes. It suggests that the functions of railings in protecting pedestrians are especially important in case of serious and fatal bus crashes. Jaywalking, while often ignored in road safety analysis, is important in accounting for other crashes other than bus-related crashes. Last but not least, it is also observed that other than the Pedestrian Exposure Factor and Jaywalking, these other factors play minor roles in predicting non-bus crashes. The results highlight the value of differentiating crash analysis by transport modes.

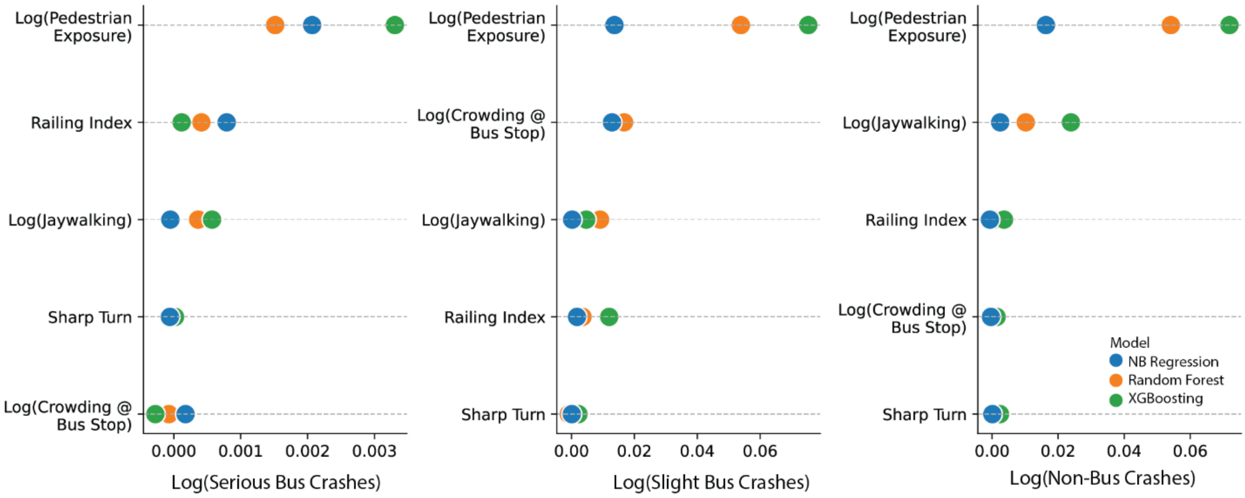

Figure 2 Permutation Importance (based on negative MSE score predicting the log-transformed dependent variable) from three models predicting serious bus crashes, slight bus crashes, and non-bus crashes.

To conclude, this study presents a method to extract pedestrian behavioral and environmental factors at street segment level for a city. We reveal that data collected by public transport companies primarily for operation reasons, such as bus dashcam videos, contains rich and valuable information for in-depth research and can help transport planning and urban management. Factors such as pedestrian exposure, pedestrian jaywalking, pedestrian crowding, and railings are operationalised factors in a dense urban environment that urban planners could manage through many pathways.

References

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.

Fan, Z., & Loo, B. P. (2021). Street life and pedestrian activities in smart cities: Opportunities and challenges for computational urban science. *Computational Urban Science*, *1*, 1–17.

Girshick, R. (2015). Fast r-cnn. *Proceedings of the IEEE International Conference on Computer Vision*, 1440–1448.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969.

Wang, Z., Zheng, L., Liu, Y., & Wang, S. (2020). Towards Real-Time Multi-Object Tracking. *The European Conference on Computer Vision (ECCV)*.

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *2017 IEEE International Conference on Image Processing (ICIP)*, 3645–3649.

Yao, S., & Loo, B. P. Y. (2012). Identification of Hazardous Road Locations for Pedestrians. *Procedia Engineering*, *45*, 815–823. https://doi.org/10.1016/j.proeng.2012.08.244

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ade20k dataset. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 633–641.