Explainable, Automated Urban Interventions to Improve Pedestrian and Vehicle Safety

Cristina Bustos[1], Daniel Rhoads[1], Albert Solé-Ribalta[1], David Masip[1], Alex Arenas[3], Agata Lapedriza[1,2], and Javier Borge-Holthoefer[1]

[1]Universitat Oberta de Catalunya
[2]Massachusetts Institute of Technology
[3]Universitat Rovira i Virgili

Abstract

Increased interactions between pedestrians and vehicles in current, crowded urban scenarios gives rise to a negative side-effect: a growth in traffic accidents, leaving pedestrians as the most injured. Here, we combine public data sources, street level imagery and a convolutional neural network (CNN) to approach pedestrian and vehicle safety with an automated and simple data-processing scheme. The steps involved include the training of a interpretable classifier to determine a hazard index for each given urban scene and the segmentation of the scene. The outcome of this approach is a fine-grained map of hazard levels across a city, a detailed analysis of the most influential urban objects in traffic safety and an heuristic to propose interventions to improve urban safety.

1. Introduction

The uncontrolled rise in urban automotive mobility has gone hand in hand with the degradation of other modes of transportation, like walking, which has suffered the most, due, in large part, to the amount of the streetscape allotted to vehicles which invades and interferes with the pedestrian space [10]. One logical consequence is the increased level of interaction between pedestrians and vehicles on common or adjacent spaces such as roads, sidewalks, and zebra-crossings [5]. Such increase gives rise to an important, negative side-effect: a growth in pedestrian injuries and fatalities [24]. Traditionally, pedestrian safety research has focused on the impact of structural [30, 26, 19, 18, 9], socio-behavioral [23], and demographic factors [20]. Nonetheless, crashes that involve motor vehicles and pedestrians are understudied, and, much less at the micro outside intersections [13]. However, nowadays with the combination of increasingly available street-level imagery sources and city open data portals, together with advances in the field of computer vision and availability of larger training datasets [39, 38], we are open to promising new opportunities for facing challenges in urban science [14]. Examples include the quantification of physical change and pattern identification in cities [21, 2, 27], the prediction of human-perceived features of street scenes [22, 17], the automated estimation of demographic variables [11, 29], or the beautification of urban images [15]. Turning to transportation research, however, computer vision has addressed mostly traffic control and surveillance [8], and automatic collision prevention [33, 34] for autonomous vehicles. Outside scene analysis, the deep learning paradigm has been exploited mostly on motor traffic [25, 32, 35, 31, 36], leaving aside, so far, its potential to tackle pedestrian safety.

In this work, we address the complexities of vehicle-to-pedestrian interaction combining the structural (scene elements) aspects of the problem. First, we exploit street level imagery data to train a CNN that estimates the degree to which urban scenes may be hazardous. Then, we map these hazard predictions to specific elements of the urban scene. Next, we deploy an automated heuristic to recommend interventions (i.e. changes in scene configuration) at a given location which could make that location safer. This study is carried out on data of two Spanish cities: Madrid and Barcelona.

2. Data Collection

The present study is based on the combination of historical accident statistics and street-level urban imagery. For both cities, accident records for the years 2010-2018 are available from the open data portals of the respective municipal governments [4, 1]. The Barcelona dataset was made up of 86,414 accidents, (11.8% vehicle-to-pedestrian and

88.2% vehicle-to-vehicle accidents). The Madrid dataset had 76,026 accidents (16.5% pedestrian and 83.5% vehicle accidents). All accidents are geolocated with their corresponding GPS coordinates. Street-level imagery was extracted from the Google StreetView (GSV) API [3]. In there, images are, on average, 15 meters away from each other. As we wanted to capture the view of the driver, we limit our queries to images facing directly down the direction of traffic of the street. For Barcelona and Madrid, there are a total of 177,645 and 704,950 street-level images, respectively. All the collected images are 640x640 pixel-size and also, contain GPS locations in their metadata, which allows us to assign each street image a binary class. We label an image as *dangerous* if one or more accidents have occurred with a 50 meter radius of its location. Otherwise, the image is labeled as *safe*. Accidents involving vehicles may happen throughout a city. However, if we focus individually on vehicle-to-vehicle and vehicle-to-pedestrian, the spatial patterns where these accidents occur are mostly non-overlapping, suggesting that the configuration of the urban scene matters. From the street-level image dataset, four different datasets were created, resulting from the combination of the two targeted cities and two accident types. In Barcelona, for pedestrian-to-vehicle accident dataset, 48.1% of the images are labeled as *dangerous*, and the rest as *safe*; for vehicle-to-vehicle dataset, 61.8% of the images are labeled as *dangerous*. In the case of Madrid, 29.1% and 48.3% are labeled as *dangerous* for pedestrian-to-vehicle and vehicle-to-vehicle dataset, respectively. For the four datasets, data was randomly split into train and test sets, containing 90% and 10% of the images respectively.

3. Methodology

To estimate a *hazard index* (H) in new, unseen images, we use a CNN based on ResNet v2 [12], pre-trained with the Imagenet [16]. We removed the connections from the last layer and we replaced it by a Softmax layer with two outputs (classes *dangerous* and *safe*). We fine-tuned the last layers of the model to predict our hazard index. To compensate class imbalance during training stage, class weights were adjusted in the objective cross entropy loss function according to inverse class frequency. In accordance with the defined accident types, we estimate two subtypes of hazard index: H_V and H_P, corresponding to the hazard indices for vehicle-to-vehicle and vehicle-to-pedestrian accidents, respectively. Therefore, we end up training 4 models in total (two per city): Barcelona P, Barcelona V, Madrid P, and Madrid V. In our data, the accuracies of these trained models are 0.75, $0, 82$, $0, 75$, $0, 75$, respectively, with precision and recall higher than 0.72 in all the cases. These 4 trained classifiers are the ones used in the rest of our study to generate all the results shown in section 4.

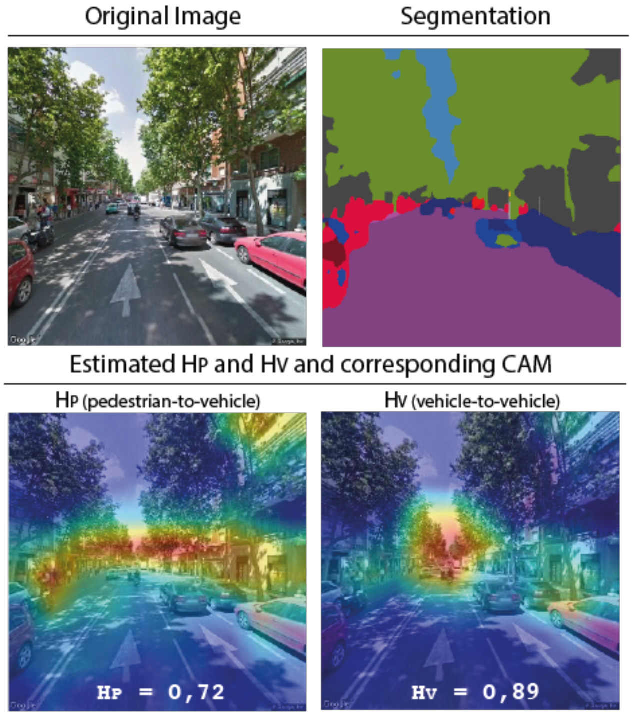
Original Image
Segmentation
Estimated HP and HV and corresponding CAM
HP (pedestrian-to-vehicle)
HV (vehicle-to-vehicle)
HP = 0,72
HV = 0,89

Figure 1. Example of an urban image with respective Hazard indexes H_P and H_V, 'Dangerous' class activation map and segmentation

To map the scene safety to the scene composition we combine scene segmentation with Class Activation Map. First, we used Pyramid Scene Parsing Network (PSPNet) [37] architecture, trained with the Cityscapes dataset [7] for urban scene segmentation. Second, we used class activation map, (GradCAM++ [28, 6]) to visually identify the most relevant image regions that activates the *dangerous* class. Since the images have been fully segmented, we can retrieve the objects that overlap with the CAM regions.

4. Results

Urban hazard landscape: The trained models for Hazard index classification, together with the short distance intervals between consecutive images, allows us to quantify the safety of all city locations at every 15 meters approximately, independently of whether accidents have occurred at a given site or not. Figure 2 shows a visualization of the spatial distribution of hazard index in Barcelona and Madrid, respectively. Notice that, as expected, the distribution of dangerous areas for pedestrian and vehicles do not overlap and are significantly different.

Mapping safety to scene composition: The segmentation and the CAM processing steps complete the data analysis pipeline, linking hazard indices, H_P and H_V, to specific objects found in street-level images. Mapping each pixel label (e.g. "road", "sidewalk", etc.) to its corresponding

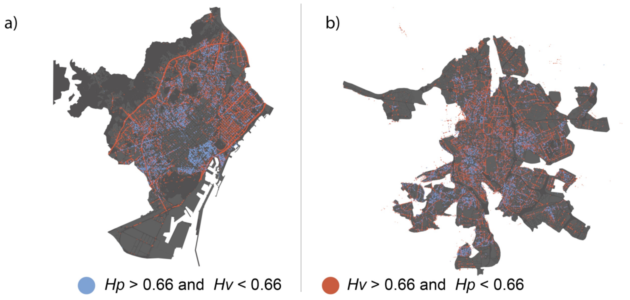

Figure 2. Distribution of high-hazard points for pedestrians and vehicles across both cities, Barcelona (a) and Madrid (b). A zoomed viewed of H_P in Barcelona (c)

activation level provides a quantification of the contribution of that pixel to the overall hazard score of the image. Thus, at the city level, we can obtain a global perspective of the object categories that most contribute to the hazard index. Specifically, we analyze the ratio between the amount of CAM fixation on a given object category (in safe and dangerous scenes), with respect to the CAM fixation on that object category across all the images of the dataset, see radar plots in Figure 3 (a) and (b). In both cases, the blue line represents safe scenes ($H < 0.33$), while dangerous ones ($H > 0.66$) are shown in red. Thus, values below 1 in the radar plots are underrepresented, while those above 1 are overrepresented. We have restricted the analysis to the city center to avoid the over-representation of natural elements (vegetation and sky) in low accident risk images. Remarkably, the presence of people in a scene triggers a dangerous classification for both vehicle-to-pedestrian and vehicle-to-vehicle predictions. Low buildings and/or wide streets (tantamount to a clear vision of the sky) render safer scenes for pedestrians, whereas the presence of buildings implies a safer environment for vehicles. Also, the absence of vegetation, such as trees, could be contributing to a safe classification for vehicles.

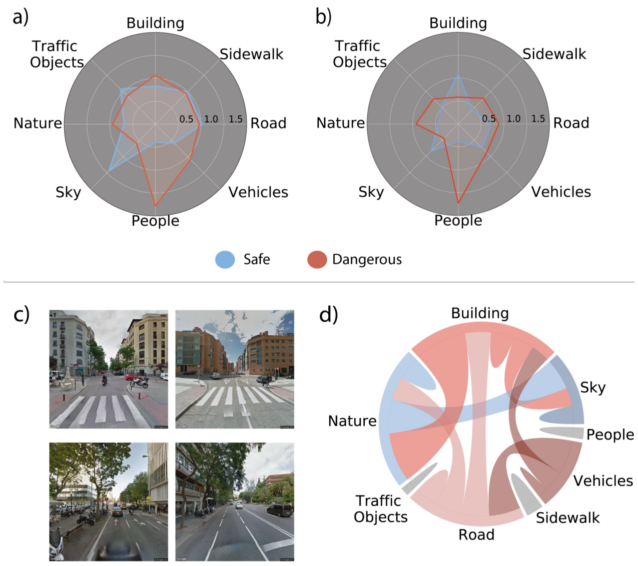

Figure 3. (a, b) Radar plots showing the level of object fixation of the CAM model for pedestrians (a) and vehicles (b); (c) Examples of original and mirror images; (d) Chord diagram representing an aggregate overview of proposed interventions (the color expresses the source of the link)

An informed guide to pedestrian safety improvements: To propose interventions conducive to scene alterations that diminish at the same time H_P and H_V we propose the following strategy. We first combine information from segmentation and CAM to build a vector of characteristics $v_i \in \mathbb{R}^C$ for every image i, containing information of the relative area of each object category C in i. Second, for the target image (the one for which we intend to reduce the hazard levels), we construct an additional surrogate vector of characteristics, $\tilde{v}_i$, in which we discard those regions that contribute most to H_P, i.e. we only consider regions of i where the class activation is mild-to-low (< 0.7). Finally, we deploy an exhaustive search to find mirror images j for $\tilde{v}_i$, with their respective vectors of characteristics v_j, such that their hazard index is lower: $\text{argmin}_j ||\tilde{v}_i - v_j||_2$, given $H_P^j < H_P^i$ and $H_V^j < H_V^i$. In other words, we seek the most similar locations in the city that have smaller H_P and H_V than i. Figure 3 (c) shows qualitative results for two examples. We can observe how the interventions proposed by the heuristic seem to simplify the original image, removing objects on sidewalks. Figure 3 (d) provides a visual overview of the most frequent interventions predicted by our optimization scheme. The most notable changes point –perhaps unsurprisingly– to the need to reconfigure urban scenes towards greener and wider spaces: indeed, both categories "road" and "building" contribute largely to "nature", while the latter does the same towards "sky". Overall, the estimations and insights from the panels in Figure 3 can provide initial indications to urban planners about achieving potential reductions of a local hazard score.

5. Conclusions

We present here an automated scheme that combines interpretable classifications and scene segmentation to analyze accident data along with street-level images. With these tools, we render a holistic characterization of a city's hazard landscape. Then, we study which locations are dangerous and how the dangerous locations are related to specific objects in urban scenes and, finally, we propose a heuristic that provides actionable insights in urban safety improvement.

References

[1] Ajuntament de Barcelona. Open data bcn. `https://opendata-ajuntament.barcelona.cat/en/`, 2019. Accessed: 2019-04-20.

[2] A. Albert, J. Kaur, and M. C. Gonzalez. Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1357–1366. ACM, 2017.

[3] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver. Google street view: Capturing the world at street level. *Computer*, 43(6):32–38, 2010.

[4] Ayuntamiento de Madrid. Portal de datos abiertos del ayuntamiento de madrid. `https://datos.madrid.es/portal/site/egob/`, 2019. Accessed: 2019-04-20.

[5] R. Cervero and M. Duncan. Walking, bicycling, and urban landscapes: Evidence from the san francisco bay area. *American Journal of Public Health*, 93(9):1478–1483, 2003. PMID: 12948966.

[6] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 839–847. IEEE, 2018.

[7] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[8] Z. M. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani. State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems. *IEEE Communications Surveys & Tutorials*, 19(4):2432–2455, 2017.

[9] T. Fu, W. Hu, L. Miranda-Moreno, and N. Saunier. Investigating secondary pedestrian-vehicle interactions at non-signalized intersections using vision-based trajectory data. *Transportation Research Part C: Emerging Technologies*, 105:222–240, 2019.

[10] R. Gakenheimer. Urban mobility in the developing world. *Transportation Research Part A: Policy and Practice*, 33(7):671 – 689, 1999.

[11] T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, E. L. Aiden, and L. Fei-Fei. Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences*, 114(50):13108–13113, 2017.

[12] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.

[13] Y. Hu, Y. Zhang, and K. S. Shelton. Where are the dangerous intersections for pedestrians and cyclists: A colocation-based approach. *Transportation Research Part C: Emerging Technologies*, 95:431–441, 2018.

[14] M. R. Ibrahim, J. Haworth, and T. Cheng. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities*, 96:102481, 2020.

[15] T. Kauer, S. Joglekar, M. Redi, L. M. Aiello, and D. Quercia. Mapping and visualizing deep-learning urban beautification. *IEEE Computer Graphics and Applications*, 38(5):70–83, 2018.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[17] L. Liu, E. A. Silva, C. Wu, and H. Wang. A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Computers, Environment and Urban Systems*, 65:113–125, 2017.

[18] G. Mecredy, I. Janssen, and W. Pickett. Neighbourhood street connectivity and injury in youth: a national study of built environments in canada. *Injury Prevention*, 18(2):81–87, 2012.

[19] M. Moeinaddini, Z. Asadi-Shekari, and M. Z. Shah. The relationship between urban street networks and the number of transport fatalities at the city level. *Safety Science*, 62:114–120, 2014.

[20] K. K. Mukoko and S. S. Pulugurtha. Examining the influence of network, land use, and demographic characteristics to estimate the number of bicycle-vehicle crashes on urban roads. *IATSS Research*, 2019.

[21] N. Naik, S. D. Kominers, R. Raskar, E. L. Glaeser, and C. A. Hidalgo. Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences*, 114(29):7571–7576, Jul 2017.

[22] N. Naik, J. Philipoom, R. Raskar, and C. Hidalgo. Streetscore-predicting the perceived safety of one million streetscapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 779–785, 2014.

[23] J. Nasar, P. Hecht, and R. Wener. Mobile telephones, distracted attention, and pedestrian safety. *Accident analysis & prevention*, 40(1):69–75, 2008.

[24] National Highway Traffic Safety Administration. Fatality analysis reporting system (fars) encyclopedia. `https://www-fars.nhtsa.dot.gov/Main/index.aspx`, 2018. Accessed: 2019-06-27.

[25] N. G. Polson and V. O. Sokolov. Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 79:1–17, 2017.

[26] S. M. Rifaat, R. Tay, and A. De Barros. Effect of street pattern on the severity of crashes involving vulnerable road users. *Accident Analysis & Prevention*, 43(1):276–283, 2011.

[27] I. Seiferling, N. Naik, C. Ratti, and R. Proulx. Green streets-quantifying and mapping urban trees with street-level imagery and computer vision. *Landscape and Urban Planning*, 165:93–101, 2017.

[28] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations

from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 618–626, 2017.

[29] E. Suel, J. W. Polak, J. E. Bennett, and M. Ezzati. Measuring social, environmental and health inequalities using deep learning and street imagery. *Scientific Reports*, 9(1):6229, 2019.

[30] S. Ukkusuri, L. F. Miranda-Moreno, G. Ramadurai, and J. Isa-Tavarez. The role of built environment on pedestrian crash frequency. *Safety Science*, 50(4):1141–1151, 2012.

[31] Y. Wang, D. Zhang, Y. Liu, B. Dai, and L. H. Lee. Enhancing transportation systems via deep learning: A survey. *Transportation Research Part C: Emerging Technologies*, 99:144–163, 2019.

[32] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang. A hybrid deep learning based traffic flow prediction method and its understanding. *Transportation Research Part C: Emerging Technologies*, 90:166–180, 2018.

[33] L. Zhang, L. Lin, X. Liang, and K. He. Is faster r-cnn doing well for pedestrian detection? In *European Conference on Computer Vision*, pages 443–457. Springer, 2016.

[34] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele. How far are we from solving pedestrian detection? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1259–1267, 2016.

[35] Z. Zhang, Q. He, J. Gao, and M. Ni. A deep learning approach for detecting traffic accidents from social media data. *Transportation Research Part C: Emerging Technologies*, 86:580–596, 2018.

[36] Z. Zhang, M. Li, X. Lin, Y. Wang, and F. He. Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. *Transportation Research Part C: Emerging Technologies*, 105:297–322, 2019.

[37] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2881–2890, 2017.

[38] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(6):1452–1464, 2017.

[39] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*, pages 487–495, 2014.