Supplementary Information to

# Explainable, automated urban interventions to improve pedestrian and vehicle safety

C. Bustos[a,*], D. Rhoads[a,*], A. Solé-Ribalta[a], D. Masip[a], A. Arenas[c], A. Lapedriza[a,b], J. Borge-Holthoefer[a,*]

[a]*Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Barcelona 08860, Catalonia, Spain*
[b]*Media Lab, Massachusetts Institute of Technology, 02139 Cambridge, MA*
[c]*Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain*

## S1. Data source acquisition and composition

### *S1.1. Empirical accident data*

As described in the main text, accident data were downloaded from public sources (city governments in the case of Barcelona and Madrid [34, 35], and state governments in the case of San Francisco [36]). We distinguish accidents where a vehicle and a pedestrian were involved (simply 'pedestrian' onwards), from vehicle-vehicle accidents (simply 'vehicle' onwards). In Barcelona and San Francisco, all accident records had GPS coordinates as metadata. In Madrid, only the street addresses of the accidents were available. These addresses were matched against the municipal street directory offered publicly by the city government of Madrid.

Visualizing this data spatially, as in Figure S1, highlights the distinct fingerprints of historical patterns of vehicle-pedestrian and vehicle-vehicle accidents in all three cities. In Barcelona and San Francisco, pedestrian accidents tend to be more concentrated in central areas, while vehicle-vehicle accidents are distributed across major thoroughfares. In Madrid, a single main

*Corresponding author
*Email addresses:* mbustosro@uoc.edu (C. Bustos), drhoads@uoc.edu (D. Rhoads), jborgeh@uoc.edu (J. Borge-Holthoefer)

artery concentrates the majority of vehicle-vehicle accidents, while pedestrian accidents are spread more widely in a number of separate clusters. This underscores not only that a safe street for a pedestrian is different than one for a driver, but also that pedestrian and vehicle safety can mean different meanings in different urban contexts.

*S1.2. Street-level imagery*

Street networks for all three cities were extracted from OpenStreetMap (OSM) [74], using the OSMnx package for Python. OSMnx [75] provides a simple interface for querying OSM data. Specifically, it allows for the extraction of road networks by place name (e.g., "Barcelona"). Additionally, only the driveable road network was extracted for each city (excluding pedestrian-only streets), a feature available out-of-the-box with OSMnx.

As mentioned in the main text, querying a single Google Street View image requires several pieces of information: the desired geographical location of the image, and the desired camera angle. GSV's coverage is obviously not comprehensive; requesting an image from a specific point does not guarantee a view from that precise location, and it will simply return the closest possible result. In the cities studied, we found an average distance of 15 meters between images.

Considering this situation, and in order to aggregate as many GSV images as possible, points were generated at regular intervals of 5 meters along the OSM-extracted street networks of Madrid and Barcelona (San Francisco's images were sourced from a different service, Mapillary, explained below). These points were used as query points for the GSV API. The desired camera angle was calculated considering the next two points along the path. This averaging method smooths out jagged angles on streets with curves. One-way streets were distinguished by a field available in the OSM data, and were only queried once, in the proper direction.

Imagery for San Francisco was provided by Mapillary, a crowd-sourced street-level imagery provider. 350,000 images were made available from Mapillary's San Francisco data. All Mapillary imagery was made available under the CC-BY-SA license. The imagery does not cover the entire city. Notably for the results (see Section 3.1 of the main text), the available images were concentrated mostly in medium-density areas of the city, while the high-density north-east and lower density south-west portions of the city lacked significant coverage. A map illustrating distribution of available photos in San Francisco, together with population density, can be seen in Figure S2.

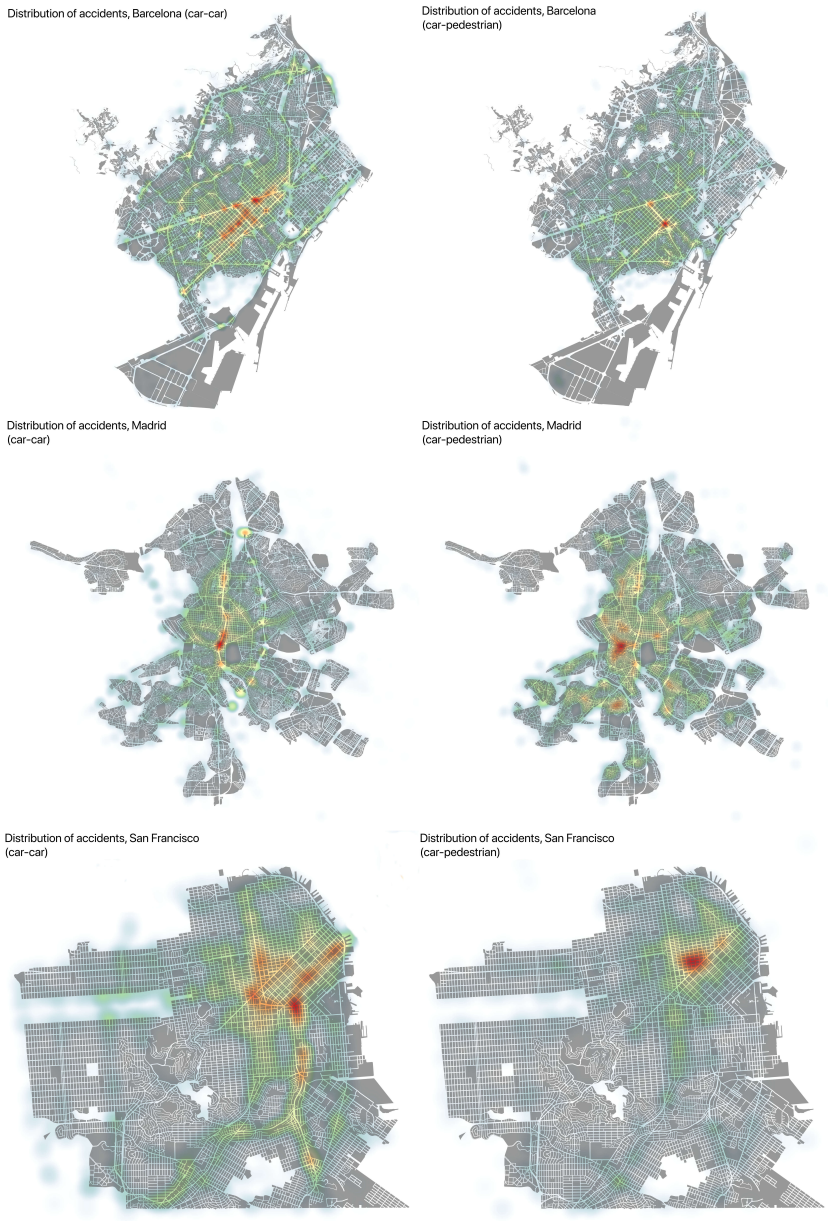


Figure S1: **Distribution of historical accident data, vehicle-pedestrian and vehicle-vehicle.** Across all 3 cities, the patterns of distribution of both types of accident are distinct. In both Barcelona and San Francisco, pedestrian hazard is heavily concentrated in the city center, while vehicle hazard is spread more widely across areas of high traffic flow. In Madrid, vehicle accidents appear funneled down the heavily-trafficked Paseo de la Castellana, while pedestrian accidents are in fact more widely distributed, concentrating in certain pockets throughout the urban core.

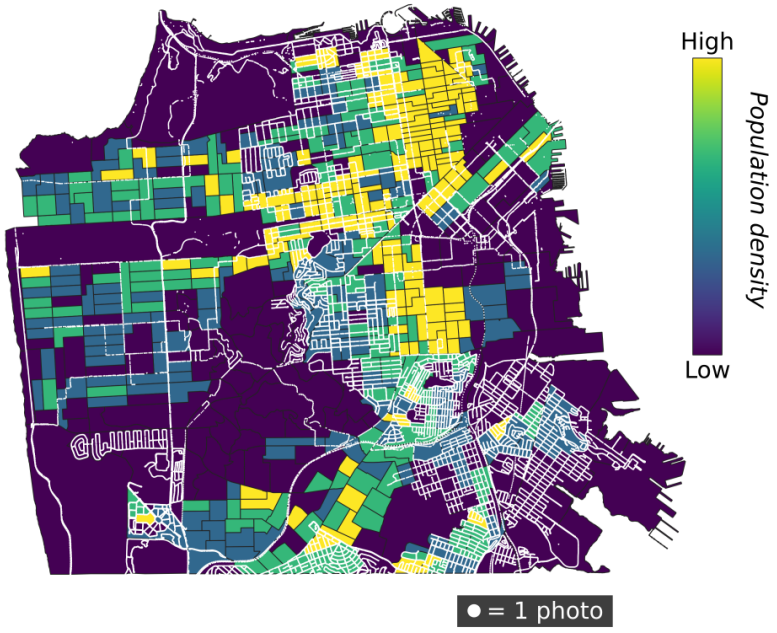


Figure S2: **Distribution of imagery available for San Francisco.** Unlike for Madrid and Barcelona, imagery for San Francisco was only available for select areas of the city. Each white dot represents an image. Census block groups are colored by their population density. Note that the majority of photos are concentrated in areas of medium population density. The imagery was available from Mapillary under the CC-BY-SA license.

Using the dataset of images from Mapillary presented distinct challenges. While GSV images are shot on standardized equipment worldwide, Mapillary imagery is crowd-sourced, and thus subject to distinct cameras, positionings of the camera, image quality, etc. Additionally, the images are not 360 degree panoramas, in contrast to GSV. This means that specific camera angles cannot be queried. In order to filter out images with angles not facing directly down the path of the street, the OSM street network data was once again used. Images whose camera angle was different from the angle of the street by more than 30 degrees were discarded.

## S2. Hazard at a sub-street segment level

As previously explained, the hazard-level estimation we develop has a high spatial resolution, assigning estimated hazard-levels to individual points separated on average by about 15 meters. This is worth highlighting, as it provides a uniquely fine-grained traffic safety measure: for a street segment of 100 meters in length (around average in Barcelona), 7 individual points can be assessed.

Figure S3 illustrates some of the interesting insights that can be gained from such a high-resolution tool. As can be seen, while some street segments have relatively constant hazard levels, others vary greatly, with some points being labelled as highly hazardous, and others being labelled as quite safe.

This same observational data can lead to other lines of investigation. For example, obvious to the naked eye (and possibly to common sense), hazard for pedestrians tends to increase at intersections.

## S3. Comparing pedestrian ($H_P$) and vehicle ($H_V$) hazard levels

The empirical vehicle and pedestrian accident data were used to estimate both vehicle and pedestrian hazard levels, $H_V$ and $H_P$ respectively, for all points in each city for which we had street-level images. Vehicle and pedestrian hazard can be compared in various ways to highlight interesting patterns and relationships.

### *S3.1. Differences in spatial distribution*

Examining the information on a map, as in Figure S4, allows us to identify spatial clusters and patterns in the distribution of relative hazard for pedestrians and vehicles. We can identify some of the same patterns in the maps

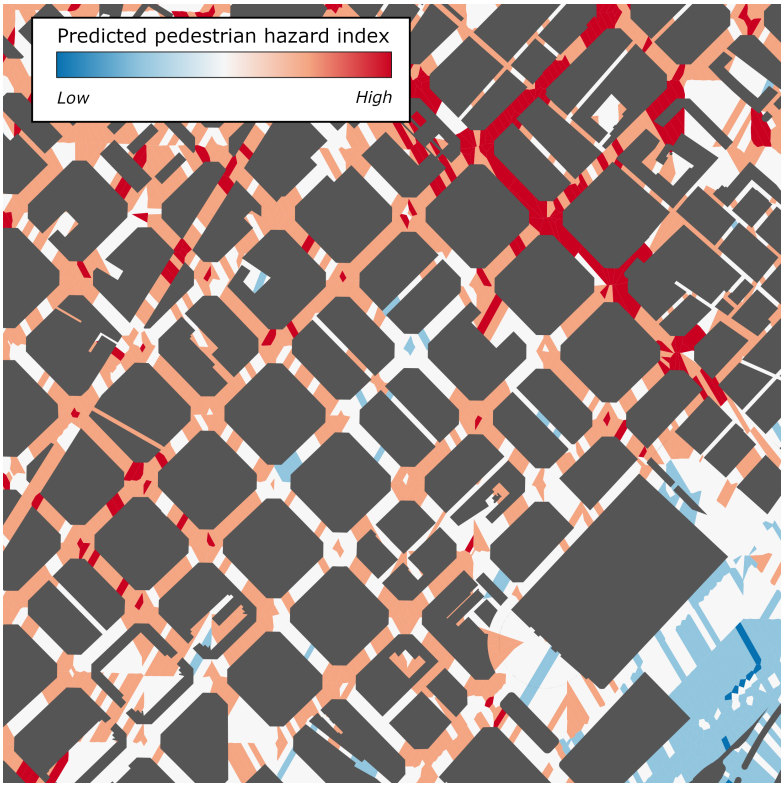


Figure S3: **Example of high-resolution hazard estimation.** An area of several city blocks in Barcelona (Poble Nou neighborhood). The colored cells form part of a Voronoi diagram of Barcelona, generated using the point locations of all street-level images available for the city. Each cell is colored according to the estimated pedestrian hazard level ($H_P$) of the image located within it. Some images within 15-30 meters of each other return highly different hazard levels.

as in the scatter plots: hazard tends to be higher in general for vehicles, and the distribution of relative hazard is similar for both groups. That being said, it is interesting to note specific areas, particularly in central locations, where hazard levels are higher for pedestrians. On the other hand, some areas of high vehicle traffic may be very dangerous for vehicles, but safe for pedestrians, simply because the lack of pedestrian infrastructure keeps walkers away. In both Madrid and Barcelona, the map of vehicle hazard clearly highlights more-trafficked roads as very hazardous, while for pedestrians, these same roads are assigned only a medium hazard level. These spatial patterns can also be compared to the distribution of empirical accident points above in Figure S1.

*S3.2. Point-by-point differences*

Figures S5c and S6c, as well as Figure 4c in the main text, show the relationship between car and pedestrian hazard for each point across the 3 cities. It can be seen that hazard is generally skewed towards vehicles (e.g., there are very few high-hazard pedestrian points that are low-hazard for vehicles, but not vice-versa). While there is variation between the pedestrian and vehicle hazard, they tend to rise together. Additionally, in line with common sense, hazard tends to be concentrated closer to the city center (not shown).

Also interesting is the comparison between the three cities themselves. While all share the patterns indicated above, San Francisco and Barcelona exhibit far more dispersed shapes, while in Madrid pedestrian and vehicle hazard seems more tightly linked.

## S4. Use of a radius to match accidents to images

As noted in the main text, we assign to a street image all accidents that are within a 50 meter radius of the image. This method, like all other possible accident assignment methods, opens up room for some error. Specifically, there is a risk that accidents from unrelated parallel streets may be assigned to an image if city blocks are particularly small in a given area. However, considering the nature of the classification problem we face, and the other options available for accident assignment, we believe the benefits outweigh the costs of this method.

The information contained in the image data used for classification is necessarily contextual, in that it takes into account not just the immediate

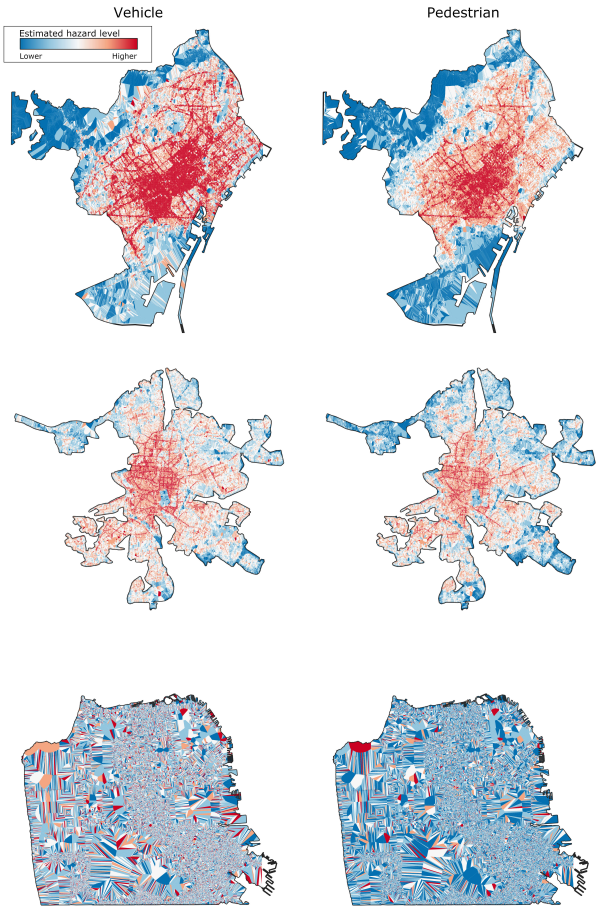


Figure S4: **Spatial distribution of estimated pedestrian and vehicle hazard index.** In the left column we see vehicle hazard levels, while the right shows those same levels for pedestrians. Here, Madrid and Barcelona, despite their distinct sizes and shapes, display similar distributions of predicted hazard for both pedestrians and vehicles. In general, hazard for both pedestrians and vehicles is much higher in the dense center. Pedestrian hazard tends to be higher in older, more pedestrianized areas, while vehicle hazard is logically higher on pedestrian-restricted roads. In San Francisco, like in the other 2 cities, vehicle hazard is noticeably higher in general than pedestrian hazard. We observe that predictions of hazard index for San Francisco is less accurate due to the precision of the Voronoi diagram constructed with a reduced (and less evenly distributed) number of images (see Figure S2 for description). Additionally, less accident data was used to train the classifier for San Francisco. Finally, it is possible that the distinct urban structures of the cities of study (Spanish cities versus a large American city) simply lead to a different distribution of hazard levels.

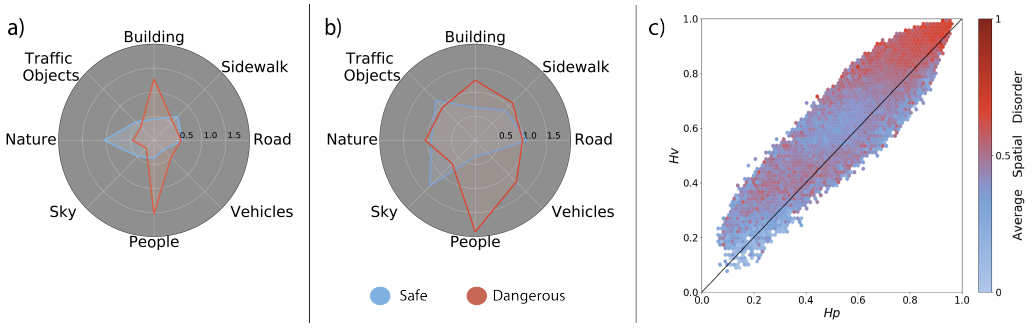


Figure S5: **Hazard level interpretability of Madrid**. **Top:** Radar plots showing the level of object fixation of the CAM model for pedestrian (a) and vehicles (b). For both, the blue area corresponds to images classified as safe ($H < 0.33$), while scenes classified as dangerous ($H > 0.66$) are mapped on the plot as red. To build these radars, each individual image is mapped to the radar categories (a relevant subset of those detected by the segmentation task), and the average of such mappings is shown. **c** The plot shows the triple relationship between $H_P$, $H_V$ and the color-coded level of disorder, which increases towards warmer colors as the levels of hazard increase. The plot corresponds to Madrid.

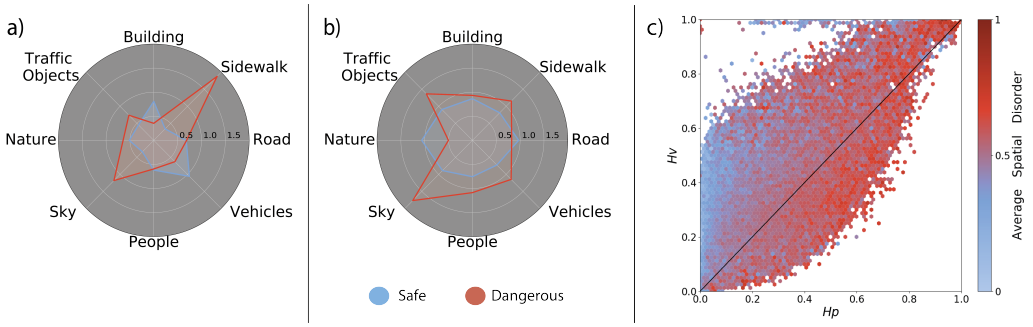


Figure S6: **Hazard level interpretability of San Francisco**. **Top:** Radar plots showing the level of object fixation of the CAM model for pedestrian (a) and vehicles (b). For both, the blue area corresponds to images classified as safe ($H < 0.33$), while scenes classified as dangerous ($H > 0.66$) are mapped on the plot as red. The patterns for San Francisco differ from those in Madrid or Barcelona, specially with regard to the role of the category "People" in dangerous scenes. **c** The plot shows the triple relationship between $H_P$, $H_V$ and the color-coded level of disorder. Unlike the clear patterns of the other cities under study, the plot for San Francisco reveals that high spatial disorder mostly dominates the region $H_V < H_P$.

location of the image, but also any part of the surrounding area that is visible,

| $r$ | Recall | Precision | Accuracy | F1-Score | FP | TP | TN | FN |
|---|---|---|---|---|---|---|---|---|
| $20m$ | 0,04 | 0,96 | 0,49 | 0,09 | 0,1% | 2,5% | 49,94% | 47,46% |
| $30m$ | 0,28 | 0,90 | 0,61 | 0,43 | 1,51% | 15,08% | 46,05% | 37,36% |
| $40m$ | 0,39 | 0,88 | 0,65 | 0,54 | 2,75% | 20,91% | 44,81% | 31,52% |
| $50m$ | 0,86 | 0,72 | 0,75 | 0,77 | 17,79% | 45,45% | 29,72% | 7,05% |
| $100m$ | 0,98 | 0,61 | 0,67 | 0,75 | 31,82% | 51,46% | 15,74% | 0,97% |
| $200m$ | 0,99 | 0,53 | 0,54 | 0,69 | 45,25% | 52,42% | 2,30% | 0,02% |

Table S1: Results of the Deep Learning approach for accident prediction, considering different radii values.

including intersecting streets and blocks beyond the current one. The visible space of the image can all be considered connected from the point of view of determining the hazard levels, since fast-moving car traffic can cover the entire area in a matter of seconds. For this reason, it is important to relate images and accidents on a similarly contextual basis.

Applying a simple radius is the most effective way to include this kind of contextual information, without manually assigning accidents on an image-to-image basis, and without including to many assumptions to an automated assignment process. Another possible option would be to assign accidents to the image's road segment. However, this leaves out information from intersecting streets. Beyond that, many common representations of road networks, like the ones we used from OpenStreetMap, present other issues that could lead to relevant accidents not being assigned to an image if only accidents from the image's segment were considered. For example: (1) complex intersections such as traffic circles tend to have many small segments which are disconnected from longer ones, (2) non-regular urban grid patterns can split single segments into smaller pieces, (3) some roads (usually arterial) are divided into separate, parallel line segments representing different lanes/traffic directions. We note that some of these problems were considered for some special cases that were easier to filter, such as grade-separated highways, whose images were not assigned accidents from parallel streets (and vice versa).

*S4.1. Results with different radii of accident assignment*

In order to construct a labeled dataset of images (dangerous/safe), we choose a radius that links a documented accident with a street-level image (or few of them, depending on the radius). For our experiments, several radii

were tested initially in Barcelona (Table S1). In this city, the best results were obtained for a radius $r = 50m$, and the same radius was applied to San Francisco and Madrid. We used the following measures: recall, precision, accuracy and F1-measure. Recall refers to the fraction of samples detected as dangerous over the total number of dangerous samples in the dataset (TP over TP+FN). Precision is the fraction of the true real danger points detected over the number of points detected as dangerous by the Network (TP over TP+FP). Accuracy measures how good is the system detecting danger points (TP+TN over all the samples). The F1-measure is a particular case of the F-measure, that combines precision and recall in a single score as:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{1}$$

These metrics were considered together to optimize the classification network's identification of dangerous and safe scenes.

The reason why radii below or above the optimal one render worse results is related to the balance between the proportion of dangerous and safe places in the training dataset. Using small radii, only a small set of points is labelled as dangerous, particularly for the pedestrian accident dataset; while the vast majority are labelled as safe. As a consequence of this, the neural network only has a few points in the class "dangerous" to learn from, tending to classify the majority of points as safe ones. Accordingly, the network learns to identify the most representative instances of dangerous points (precision is close to 1), but recall is very low: most dangerous spots are misidentified as safe. The extreme disparity between recall and precision is progressively corrected as $r$ increases. Indeed, as the radius value increases, more locations are labelled as dangerous, and the network is able to learn more features from such dangerous points, discriminating better from safe ones.

Beyond the optimal radius, the same effect occurs in the opposite direction. Now, the classifier has a large dataset of dangerous points, and only a few safe ones. Recall is much higher than precision, and the network overestimates the proportion of dangerous spots.

Finally, a mid-range radius of 50 meters yielded optimal accuracy and F1-score. Along the lines of the above paragraphs, the reason behind this good result is a balanced dataset between safe and dangerous locations. The histograms in Figure S7 show the number of points against the number of accidents occurring in those points, for Barcelona, and for several radii.

We are aware that several techniques can be used to correct imbalances

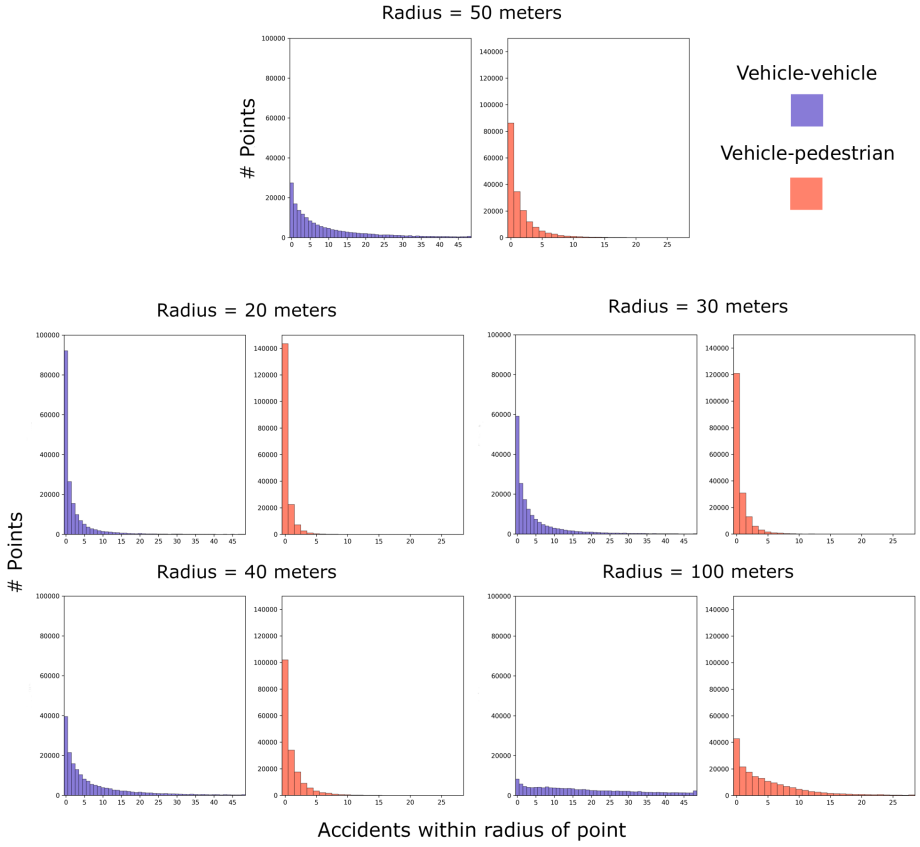


Figure S7: **Distribution of accidents per point with various radii (Barcelona).** The radius chosen for our study, 50m, is at top. Note the predominance of points with 0 accidents for pedestrians in lower radii, and the wide distribution of points with many accidents for vehicles at higher radii.

in the size of training datasets, and two of them (weighted loss function and minor class oversampling) were actually implemented, for the sake of completeness. As expected, overall results for radii below and above $50m$ improved, but never surpassed, the ones reported in the main text and in Table S1. Faced with similar results, we believe that the proposed pipeline is simpler and yields a more natural interpretation beyond the technical details. In particular, street-level images are, on average, separated by about 15 meters (as stated in the main text). Using a medium-sized radius, like $r = 50m$, allows us to include the relationship between neighboring points on a street segment, while preventing $r$ to encompass a too-large portion of a city block. In practical terms, our choice enables the inclusion of the local features, mostly within a segment, excluding at the same time irrelevant information from other blocks.

## S5. Multi-class Classification with Ordinal Regression

For comparison purposes, we trained a multi-class classifier with four classes. For this experiment, we used the Barcelona $P$ dataset. Instead of addressing a simple multi-class classifier, we cast this problem into a Ordinal Regressor, because the created classes are not completely independent between them and follow a danger level ranking.

To construct the dataset, we have divided the images in four rating classes: 'no-danger', 'mild-danger', 'danger' and 'high-danger'. Images tagged as 'no-danger', correspond to locations with no accidents, images in class 'mild-danger' have one accident around a radius of 50m, images in class 'danger' have between 2 and 5 accidents around, and, finally, images belonging to class 'high-danger' have more than 5 accidents around. Classes distribution can be seen in Table S2.

For the Ordinal Reggresor, we used the methodology shown in [61]. We trained three binary classifiers for the four classes. For each binary classifier, the classification architecture used is ResNet, as well as in the main pipeline. As metrics, we used simple accuracy and balanced accuracy, this last takes into account the number of samples of each class. Table S3 shows every binary classifier classes and accuracies. After training, the final accuracy of ordinal reggressor is 0.47, significantly better than a simple multi-class classifier. Final classification results can be seen in Table S4.

| Class | Samples |
|---|---|
| No-danger | 85509 |
| Mild-danger | 34576 |
| Danger | 40315 |
| High-danger | 17245 |

Table S2: Dataset distribution

| Negative Class | Positive Class | Accuracy | Balanced Accuracy |
|---|---|---|---|
| No-danger | Higher than Mild-danger | 0.75 | 0.75 |
| Less than Mild-danger | Higher than danger | 0.72 | 0.67 |
| Less than danger | High-danger | 0.89 | 0.66 |

Table S3: Ordinal Regressor internal classifier results

## S6. Pedestrian Safety Improvements

Figure S8 shows the results of the application point-to-point of the exhaustive search of mirror images, as well as for the dummy k-nn regressor, for Madrid and San Francisco. Figure S9 illustrates more examples of mirror images for the 3 cities taken into this study. Finally, Figure S10 provides a visual overview of the most frequent interventions predicted by our optimization scheme for Madrid and San Francisco.

| Method | Accuracy | Balanced Accuracy |
|---|---|---|
| Simple Multi-class classifier | 0.53 | 0.39 |
| Ordinal Regressor | 0.55 | 0.47 |

Table S4: Results of Ordinal Regressor

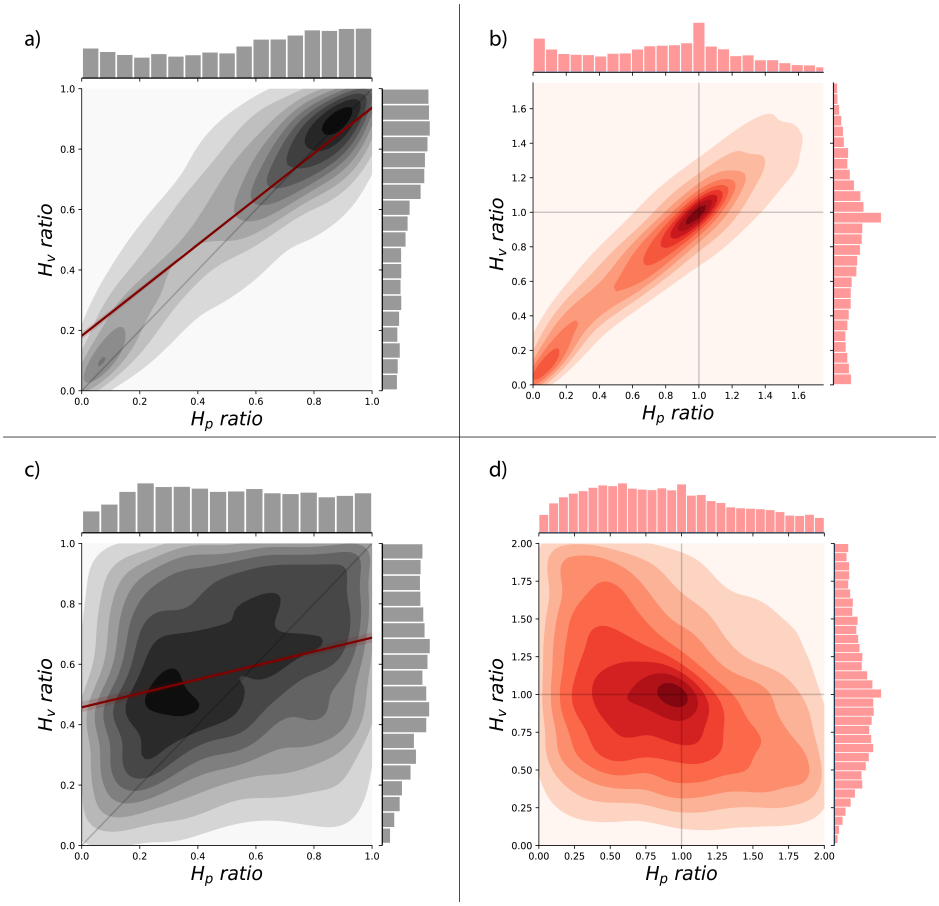


Figure S8: **Hazard reduction heuristics.** Expected improvement for pedestrian and vehicle hazards, with respect to their original values: (a) Madrid and (c) San Francisco. Grey intensity represents the density of observations in a given area of the plot. Expected improvement of a dummy k-nn algorithm that only considers similarity between images: (b) Madrid and (d) San Francisco.

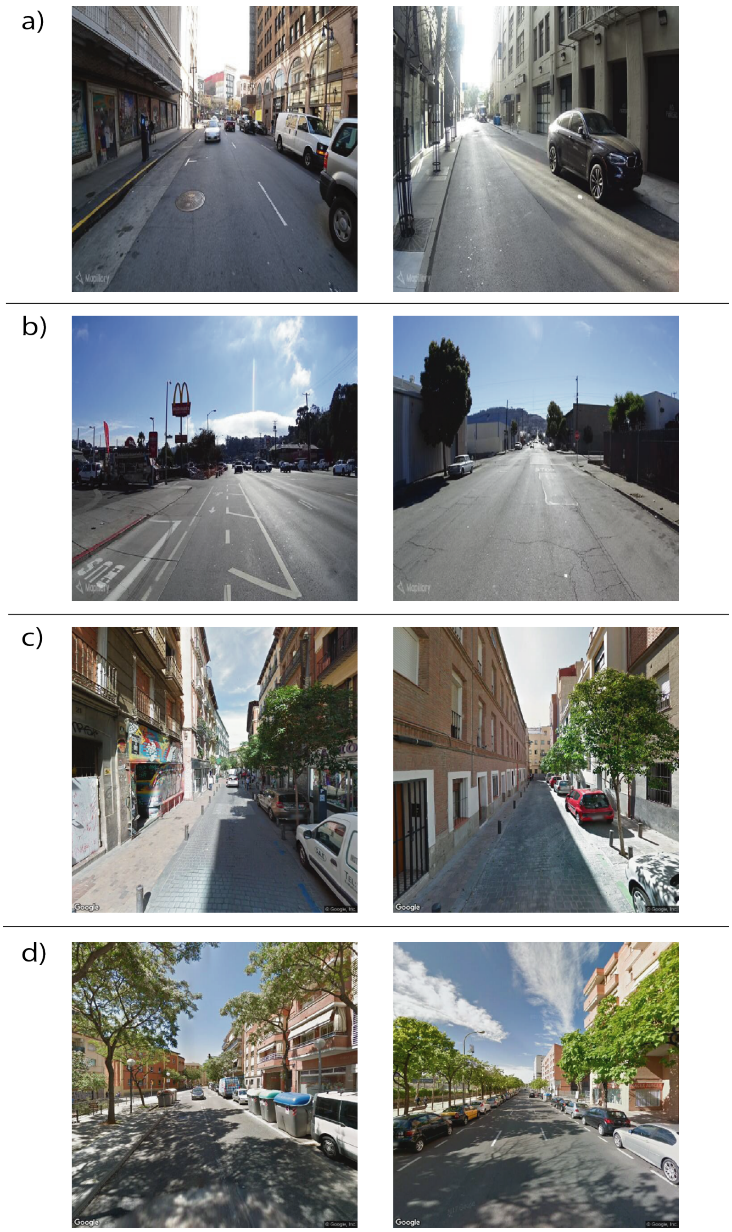


Figure S9: **Hazard reduction: selected examples.** Panels a-d show a representative selection of proposed interventions towards safer scenes in San Francisco (a and b), Madrid (c) and Barcelona (d). For each target on the left (i.e. the scene that we attempt to improve), we show its most similar, safer mirror image, illustrating some common interventions proposed by the heuristic. Visually, all of them seem to point at simplifications of the original image.

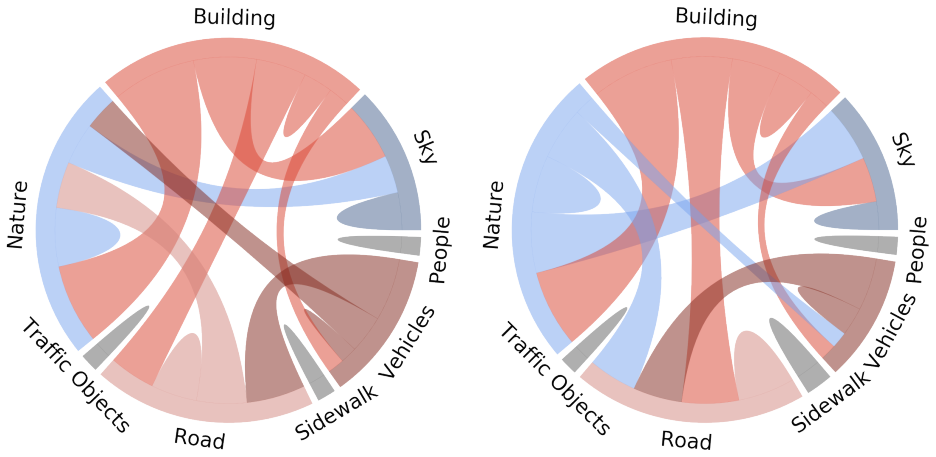


Figure S10: **Aggregate view of hazard reduction.** Chord diagram representing an aggregate overview of proposed interventions in Madrid (left) and San Francisco (right).

## References


[1] M. De Domenico, A. Solé-Ribalta, S. Gómez, A. Arenas, Navigability of interconnected networks under random failures, Proceedings of the National Academy of Sciences 111 (2014) 8351–8356.

[2] S. Jiang, Y. Yang, S. Gupta, D. Veneziano, S. Athavale, M. C. González, The timegeo modeling framework for urban mobility without travel surveys, Proceedings of the National Academy of Sciences 113 (2016) E5370–E5378.

[3] S. Abbar, T. Zanouda, J. Borge-Holthoefer, Structural robustness and service reachability in urban settings, Data Mining and Knowledge Discovery 32 (2018) 830–847.

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

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

[6] 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.

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

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

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

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

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

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

[13] K. K. Mukoko, 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).

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

[15] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, A. Oliva, Learning deep features for scene recognition using places database, in: Advances in neural information processing systems, pp. 487–495.

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

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

[18] A. Albert, J. Kaur, 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, ACM, pp. 1357–1366.

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

[20] W. Song, S. Workman, A. Hadzic, X. Zhang, E. Green, M. Chen, R. Souleyrette, N. Jacobs, Farsa: Fully automated roadway safety assessment, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, pp. 521–529.

[21] N. Naik, J. Philipoom, R. Raskar, C. Hidalgo, Streetscore-predicting the perceived safety of one million streetscapes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 779–785.

[22] L. Liu, E. A. Silva, C. Wu, 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 (2017) 113–125.

[23] T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, E. L. Aiden, 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 (2017) 13108–13113.

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

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

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

[27] L. Zhang, L. Lin, X. Liang, K. He, Is faster r-cnn doing well for pedestrian detection?, in: European Conference on Computer Vision, Springer, pp. 443–457.

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

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

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

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

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

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

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

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

[36] Safe Transportation Research and Education Center, University of California, Berkeley, Transportation injury mapping system (tims), 2019. Accessed: 2019-06-27.

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

[38] Mapillary contributors, Mapillary - Street-level imagery, powered by collaboration and computer vision , `https://www.mapillary.com/app`, 2019.

[39] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436.

[40] J. Schmidhuber, Deep learning in neural networks: An overview, Neural networks 61 (2015) 85–117.

[41] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, in: European conference on computer vision, Springer, pp. 630–645.

[42] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, pp. 1097–1105.

[43] A. Adadi, M. Berrada, Peeking inside the black-box: A survey on explainable artificial intelligence (xai), IEEE Access 6 (2018) 52138–52160.

[44] H. Fukui, T. Hirakawa, T. Yamashita, H. Fujiyoshi, Attention branch network: Learning of attention mechanism for visual explanation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 10705–10714.

[45] J. Wagner, J. M. Kohler, T. Gindele, L. Hetzel, J. T. Wiedemer, S. Behnke, Interpretable and fine-grained visual explanations for convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9097–9107.

[46] S. Desai, H. G. Ramaswamy, Ablation-cam: Visual explanations for deep convolutional network via gradient-free localization, in: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, pp. 972–980.

[47] B. N. Patro, M. Lunayach, S. Patel, V. P. Namboodiri, U-cam: Visual explanation using uncertainty based class activation maps, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 7444–7453.

[48] Z. Wang, J. Yang, Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation, arXiv preprint arXiv:1703.10757 (2017).

[49] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890.

[50] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, The cityscapes dataset for semantic

urban scene understanding, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3213–3223.

[51] R. M. Haralick, K. Shanmugam, I. H. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics (1973) 610–621.

[52] N. Moray, Attention in dichotic listening: Affective cues and the influence of instructions, Quarterly journal of experimental psychology 11 (1959) 56–60.

[53] D. Kahneman, Attention and effort, volume 1063, Citeseer, 1973.

[54] G. A. Alvarez, P. Cavanagh, The capacity of visual short-term memory is set both by visual information load and by number of objects, Psychological science 15 (2004) 106–111.

[55] J. E. Richards, The development of attention to simple and complex visual stimuli in infants: Behavioral and psychophysiological measures, Developmental Review 30 (2010) 203–219.

[56] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2921–2929.

[57] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626.

[58] A. Chattopadhay, A. Sarkar, P. Howlader, 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), IEEE, pp. 839–847.

[59] C. Ventura, D. Masip, A. Lapedriza, Interpreting cnn models for apparent personality trait regression, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 55–63.

[60] P. Harrington, Machine learning in action, Manning Publications Co., 2012.

[61] E. Frank, M. Hall, A simple approach to ordinal classification, in: European Conference on Machine Learning, Springer, pp. 145–156.

[62] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).

[63] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.

[64] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-first AAAI conference on artificial intelligence.

[65] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861 (2017).

[66] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

[67] R. Louf, M. Barthelemy, A typology of street patterns, Journal of The Royal Society Interface 11 (2014) 20140924.

[68] D. Quercia, R. Schifanella, L. M. Aiello, The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city, in: Proceedings of the 25th ACM conference on Hypertext and social media, ACM, pp. 116–125.

[69] H. Huang, M. A. Abdel-Aty, A. L. Darwiche, County-level crash risk analysis in florida: Bayesian spatial modeling, Transportation Research Record 2148 (2010) 27–37.

[70] P. Chen, J. Zhou, Effects of the built environment on automobile-involved pedestrian crash frequency and risk, Journal of Transport & Health 3 (2016) 448–456.

[71] P. Olszewski, I. Buttler, W. Czajewski, P. Dabkowski, C. Kraśkiewicz, P. Szagała, A. Zielińska, Pedestrian safety assessment with video analysis, Transportation Research Procedia 14 (2016) 2044–2053.

[72] Y. Xu, S. Jiang, R. Li, J. Zhang, J. Zhao, S. Abbar, M. C. González, Unraveling environmental justice in ambient pm2. 5 exposure in beijing: A big data approach, Computers, Environment and Urban Systems 75 (2019) 12–21.

[73] A. Palazzi, D. Abati, F. Solera, R. Cucchiara, et al., Predicting the driver's focus of attention: the dr (eye) ve project, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (2018) 1720–1733.

[74] OpenStreetMap contributors, Planet dump retrieved from https://planet.osm.org , `https://www.openstreetmap.org`, 2017.

[75] G. Boeing, Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks, Computers, Environment and Urban Systems 65 (2017) 126–139.