City-SAFE: Estimating Urban Safety Perception

Gabriel Pereira Correia Cabral da Costa
gabrielcosta@tecnico.ulisboa.pt
Instituto Superior Técnico, Lisboa, Portugal
May 2019

Abstract
The study of urban safety perception has been an important subject of focus by social scientists, seeking to draw conclusions that lead to the improvement of people’s lives. We tackle this issue in the Lisbon area by conducting a survey on street imagery using pairwise comparisons – the City-SAFE dataset. We analyze our dataset using semantic segmentation, finding evidence for the existence of a collective agreement on safety perception and providing its semantic profile. We then assign scores to locations, which represent a perceptual ordering based on the collected comparisons. This is performed through convex optimization with graph regularization on spatial and temporal similarities. To assess the performance of the proposed method, we conduct a simulation using synthetic data and compare it with the state-of-the-art. Our approach shows better performance both on noisier and fewer comparisons. We then apply our method to the City-SAFE dataset, generating scores for Lisbon, Amadora and Cascais. Finally, we build an estimator using the obtained scores as a constructed ground truth for our dataset. We use a support vector regression where image features and features constructed from a citizen reporting platform are combined via multiple kernel learning.

Keywords: Pairwise comparison, convex optimization, perceived safety, image processing

1. Introduction
Understanding whether people feel safe or not and why has been a matter of interest to social scientists for a long time. These insights are of extreme usefulness so that better policies and decisions can be made to improve citizens’ lives.

Having data that accurately describe reality and the ability to draw good conclusions is not an easy task for something like safety perception. Traditionally, this information had been collected through lengthy surveys concerning absolute judgments, which meant that relatively few people would be queried and the resulting data would contain multiple biases [1]. In recent years, the growth of accessible data combined with increasing computational power has given rise to new data-driven approaches, like ours.

In this project, we seek to estimate how people feel safety-wise across the Lisbon area by developing an estimator for perceived safety using available urban data (we used image features and features constructed on data from a platform for reporting issues like parking, the lighting of streets and rubbish collection). Since building this estimator requires training data we conduct a survey on peoples perceptions of safety in the form of pairwise comparisons using street imagery.

We use these comparisons to conduct a user analysis using image semantics, showing a general convergence on safety perception and how semantic contents are perceived in terms of safety.

We, then, solve the problem of extracting the full perception scale from these pairwise comparisons using convex optimization with graph regularization in both time and space. These measurements of perceived safety are then used as the training data for the intended estimator.

2. Background
Safety perception of citizens has been a subject of great interest for social scientists as it involves complex urban dynamics, going beyond tackling mere crime rates, which may not accurately reflect safety perception.

Q. Wilson and L. Kelling [2] proposed that if disorderly behavior remains unchecked, it leads to more disorderly behavior. Despite criticism regarding policies based on this theory, there have been studies supporting it [3]. Links have also been found between safety perception and academic achievement [4], physical activity among the elderly [5], and multiple health indicators [6].

Nowadays, available data allow for new data-driven approaches to the study of urban safety perception. Salesses et al. [7] used crowdsourced pairwise comparisons on street imagery provided by Google Street View to extract a quantitative measurement of safety perception for each image (which showed a high correlation with homicide data in New York). Naik
et al. [8] later used the same data in conjunction with TrueSkill [9] to generate safety scores which were then combined with generic image features to fit a support vector regression. This research was followed by a deep learning version where deep neural networks are assembled and tuned for feature extraction and ranking estimation with performance measured in pairwise comparison predictive accuracy at 73.5% [10]. This methodology has been followed not only for the study of safety but also beauty depression, boredom, liveliness and wealth. These predictive capabilities aren’t limited to street imagery as Najjar et al. [11] have shown a correlation between satellite images and violent crime reports from Chicago, Denver, and San Francisco.

Pairwise comparisons have shown to capture better humans perceptions than lengthy surveys on absolute judgments, has these are more prone to biases [1]. Humans naturally are better at choosing between few options than assigning absolute values to perceptions, since we build our perceptions in previous experiences from memory [12]. This approach has shown also to bypass individual biases like age, gender and location effectively [7]. Kleindessner and von Luxburg [13] have shown that only a subset of local pairwise comparisons are enough to extract an ordinal embedding and later Jain et al. [14] showed that this embedding can, in fact, be extracted with noisy measurements, as is usually the case.

Bradley and Terry [15] proposed a mathematical model on how objects ratings determine the probability of each winning in pairwise comparisons. This model was the base for the Arpad Elo system [16], popularized by its use on chess players’ ratings and accounted for ties where each player was modeled with a rating estimation and uncertainty on this estimation. Glickman [17] later provided a modification where the uncertainty on the player’s rating depended on the time since his last match. More recently, Microsoft TrueSkill [18] built on this model and applied it to online games where a factor graph models players’ skills, performances, their arrangement into teams providing Microsoft with the ability to match teams as competitively as possible. This algorithm was used in the Streetscore project [8] where stable scores are achieved after about 16 comparisons per image for about 4000 images.

Building an estimator from the extracted scores with the images presented to participants in our survey requires that image features be extracted. Xiao et al. [19] showed that generic image features work well in capturing relevant information from scenes, which led Naik et al. [8] to use the combination of GIST, geometric texton histograms, and geometric color histograms. In recent years, deep learning has boosted the performance of computer vision challenges in multiple fields. Deep convolutional neural networks have shown to outperform traditional methods in feature extraction with the output of the last fully connected layer being commonly as a set of image features [20] and also by aggregating the outputs of convolutional layers [21], which represent local image features.

3. Data Collection and Exploratory Data Analysis

The data used in the scope of this thesis came from three different sources: Google Maps, the CitySAFE website (https://smartcity.isr.tecnico.ulisboa.pt/CitySAFE/) and the Na minha rua database of citizens’ reports (https://naminharualx.cm-lisboa.pt/). Other data sources such as Twitter and local police reports were explored and analyzed but have not yet been integrated.

3.1. Google Maps Platform

Street View Static API In order to study citizen’s safety perceptions from urban visual data, an extensive dataset of pictures is required. The Street View Static API allows us to query for almost any street scene in the Lisbon area (here we collect data concerning Lisbon, Amadora, and Cascais).

The queried points were chosen by overlapping a uniform grid of step 0.0005° on the city boundaries (a polygon of points extracted from OpenStreetMap). All images collected were filtered down to 2032 from Cascais, 2056 from Lisbon and 1007 from Amadora. All images were collected with default heading, field of view and radius, size of 600x300 and pitch of 10°.

Since not all pictures collected were in good condition for future use, we removed very dark images (images with average channel L below 50% in CIELab color space) and also removed indoors images (using a WideResNet18 network pre-trained on the places365 dataset by [22] for scene classification).

Distance Matrix API Throughout this thesis, distances between points are considered for geographical coherence on scores attributed to each image (the closer the images, the closer their scores), but since there are natural urban obstacles like rivers and buildings, walking distances represent much more accurately the real distance from a pedestrian point of view than the Euclidean distance. For this reason, we use the walking distances provided by the Distance Matrix API between different points.

3.2. User survey based on pairwise comparisons

This thesis builds on extracting information from pairwise comparisons. This task was carried out by crowd-sourcing comparisons through an online survey we created where participants are presented the question “Which place looks safer?”, two random pictures side by side and given the option to choose either the left or the right one as safer, or instead consider them equally
Figure 1: Number of comparisons provided by user. Most people provide less than 20 comparisons but there is a very relevant number of users that provided more than 100 comparisons.

To keep track of each participants comparisons we set a cookie with a unique ID with which all comparisons submitted would be connected. We consider that each cookie corresponds to one user.

We collected approximately 19k comparisons so far and most pictures were compared 5 to 8 times and we have about 3.5 times the number of comparisons to the number of images. The equal option was the least chosen at 26.3% while left and right were at 37.6% and 36.1%, respectively. A histogram of the number of users that submitted a certain amount of comparisons can be seen in Figure 1.

3.3. Na minha rua data

This dataset comprises reports made by citizens regarding problems they find in their neighborhoods that need solving, such as garbage collection, parking issues, green areas or lighting problems. All reports have a description and are geo-tagged with a latitude-longitude pair as well as some other metadata such as parish, report type, status, ID and address.

In total, there are 26406 unique reports of which only 125 had their status changed and only one was flagged as solved. In a parish breakdown, Lumiar had the most reports, followed by Estrela and Benfica but on a per capita analysis the most participative parishes are Ajuda, Arroios, and Areeiro. The most common problems reported by citizens are related to urban hygiene, and roads and signaling, for all parishes in general.

Using all these data (comparisons, images and reports), we conduct a user study via semantic analysis. Secondly, we extracted an scores from the comparisons, to be then used as the ground truth target values of an estimator using images and citizens’ reports features.

4. User Semantic Study

In order to provide a meaningful estimation of people’s perceptions of safety it is relevant to determine how close are individual perceptions from each other, whether there are preferences or even clearly different profiles of perception. We achieve this using scene semantic segmentation, a very high-level representation of the image and, thus, suited for our purpose.

4.1. Semantic segmentation and image representation

To perform this analysis, we apply the semantic segmentation developed by Bulo et al. [23] which uses their InPlace-ABN on a WideResNet38+DeepLabv3 network. This implementation was chosen for being one of the top performing configurations available. The network outputs, for each image, is a pixel-wise color-coded image on the semantic class each pixel belongs to. Some classes shared the same colors so they were automatically merged (which was not a concern as they were either similar or only one of them was actually common in all cases). Since there were some unreliable classes and other classes that were too similar to be effectively differentiated; the former were discarded and the latter were merged.

Having performed semantic segmentation on all images of our dataset, we build a descriptor \( s \in \mathbb{R}^K \) for each image where \( s(k) = 1 \) if there are more than 1000 pixels of semantic class \( k \) in the image (an empirical threshold for detection by the human eye) and \( s(k) = 0 \) otherwise. This binary representation accounts for natural differences in the number of pixels that belong to each class (sky is usually far larger than bicycle).

4.2. User representation

All comparisons have an ID tag that belongs to the user that submitted it. We can use this information to build a descriptor for each user based on the comparisons he submitted. We define the comparison representation as \( c = (s_{\text{winner}}, s_{\text{loser}}) \) and the user representation, \( \tilde{c}_u \) as

\[
\tilde{c}_u = \sum_{c \in C_u} c
\]

where \( C_u \) is the set of comparisons submitted by user \( u \). This representation reflects the number of comparisons, the number of different semantic classes that were present in the winning and losing images for each user, and which semantic classes were more common in winning images and losing images.

Since some semantic classes rarely appear in our dataset, whether they are found on winning or losing images may be biased by that specific pairwise match and provide inaccurate information regarding that given class. Besides, some users provided very few comparisons which would not be an accurate representation of their perception of safety as it would be very dependent on which images were shown to them.
in those few comparisons. We discarded, for these reasons, classes that appeared less than 200 times in all comparisons and users that had submitted less than 10 comparisons.

After these considerations, our data consist of frequencies of detections concerning the observed random variables:

- $S$ — the semantic class;
- $U$ — the user;
- $W$ — the winning image;
- $L$ — the losing image.

We can then use this probabilistic formulation to study, for instance, the event $P(\neg W, L, S = \text{car}|U = u)$ which describes the proportion that for all comparisons of user $u$, the class car appears on the losing image but not on the winning image. Here we focus on $P(W, \neg L|S, U)$ and $P(\neg W, L|S, U)$ as these distributions show which classes are more frequent on winning or losing images alone, regardless of how frequently they appear, which gives a good perception of whether each class makes a difference in people’s perception of safety.

4.3. Clustering users

We cluster users according to their representations in order to look for group preferences and assess how safety perception changes throughout our dataset. This is performed using hierarchical agglomerative clustering with complete linkage and cosine distances between users. The cosine distance measurement captures the profile of each user regarding the balance of classes present in winning and losing images, and regardless of the number of comparisons he submitted. Complete clustering provides the most satisfactory clustering results by generating compact clusters of users without resulting in chaining as there are not hard dissimilarities between them. From the resulting dendrogram, we pruned it to the 6 clusters presented in Table 4.3.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>#users</td>
<td>2</td>
<td>16</td>
<td>7</td>
<td>71</td>
<td>59</td>
<td>42</td>
</tr>
<tr>
<td>#comparisons</td>
<td>22</td>
<td>278</td>
<td>109</td>
<td>8044</td>
<td>2259</td>
<td>737</td>
</tr>
</tbody>
</table>

Clusters 1 and 3 were not considered in the following analysis for representing few users.

A comparison of the proportions given by $P(W, \neg L|S, U)$ and $P(\neg W, L|S, U)$ can be seen in Figures 2 and 3, where there is a seemingly common profile among all clusters. There are, though, some notorious standouts, for example, in traffic sign and crosswalk for cluster 2 (which is also the smallest), and traffic sign and service lane for cluster 6. Despite this, in general, we see that the same classes have higher appearances on winning images (while not being on the losing ones) for all clusters (and vice-versa).

Since clusters have different average numbers of comparisons per user, are profile differences attributable to these different amounts of data? Is there actually a common safety perception as we further survey users’ perceptions? Following this hypothesis, we created a general profile of safety perception, $\tilde{c}^*$, which is the centroid of the overall dataset. We then took the most active member from each cluster and measured its cosine distance to $\tilde{c}^*$ as we consider a greater sample of comparisons. Figure 4 shows that all the users’ distances to the generic profile decrease towards zero as a higher number of comparisons is considered and, thus, their profiles also tend to one another, supporting our hypothesis.

A final exploration of the data as a single cluster shows, in Figure 5, which semantic classes appear more frequently in winning and losing images. Human-related classes such as pedestrian area, crosswalk and person are more typical of winning images while bridge, service lane, rail track and guard rail/trash can are more typical of losing images. Figure 5 also evidences that while sky and vegetation are very common classes they do not relate with safety perception.
5. Scoring Images from Pairwise Comparisons

The collection of safety perception data in the form of pairwise comparisons had been chosen for trustworthiness purposes, as it is a more natural and easy way to capture perceptions. This approach requires, though, more complex processing of the data to perform the task of attributing a score for each image that better extracts a quantitative safety perception measurement from the available data.

5.1. Convex problem formulation for pairwise comparisons

We start by defining the following convex optimization problem:

\[
\begin{align*}
\text{minimize} & \quad x^T \epsilon \\
\text{subject to} & \quad 1^T x = 0,
\end{align*}
\]

where \( x \in \mathbb{R}^M \) is the vector with the scores for all \( M \) images (the highest the score, the safer the image); \( b_n \in \mathbb{R}^M \) conveys the information of one pairwise comparison, taking value 1 in the index of the winning image, value \(-1\) in the index of the losing image and \(0\) on every other entry; and \( \epsilon \) consists of an error margin to be tolerated (which is convex because it is the composition of multiple convex functions).

This cost function was designed to penalize scores that violate more comparisons by a larger margin within the error margin \( \epsilon \) (the ramp function, \((y)_+\), enforces that only violated comparisons are considered). This can be reformulated as the linear program

\[
\begin{align*}
\text{minimize} & \quad 1^T t \\
\text{subject to} & \quad 1^T x = 0,
\end{align*}
\]

by introducing the variable \( t \).

So far, the optimization problem does not account for tied comparisons (about 25% of our dataset), which can be incorporated by penalizing differences between
tied images with the term

$$\lambda_{\text{ties}} \sum_{i \sim j \in C_{\text{ties}}} |x_i - x_j| = \lambda_{\text{ties}} \sum_n |b_n^T x|$$  \hspace{1cm} \text{(4)}

for all ties between images $i$ and $j$.

5.2. Temporal and geographical regularization

Two nearby places have naturally similar safety perceptions provided one can quickly go from one to another. This leads us to consider the walking distances between locations instead of Euclidean distances (since buildings, roads, water bodies, and other urban obstacles introduce discontinuities). Using the collected distance data from the Google Maps platform, we can add penalties for close images that have very different scores, which will introduce geographical coherence.

This is achieved through an undirected weighted graph, $G_{\text{geo}} = (V_{\text{geo}}, E_{\text{geo}})$, with vertices as images and their edges’ weights with a spatial similarity function (we used the inverse of the walking distance). With the weights, $a_{ij}$, generating an adjacency matrix, $A$, and the sum of the weights generating the vertex degrees matrix, $D$, the relationship

$$x^T L_{\text{geo}} x = \sum_{i \sim j \in E_{\text{geo}}} a_{ij} (x_i - x_j)^2$$  \hspace{1cm} \text{(5)}

allows us to introduce the desired graph regularization, where $L = D - A$.

On the other hand, not all pictures available in Google Street View were taken at the same time (they cover a span of 10 years) and places can change a lot in this time frame. This means that, even though two images are geographically close, if they are temporally very distant from each other, their scores need not be as similar as if they were also temporally close.

We use this information to build a very similar graph, $G_{\text{time}} = (V_{\text{time}}, E_{\text{time}})$, to the one presented for geographical similarity but, this time, edges’ weights are based on time similarity,

$$a_{ij} = e^{-\alpha |t_i - t_j|},$$  \hspace{1cm} \text{(6)}

where $t_i$ and $t_j$ are the timestamps of two images.

These penalties are only relevant for nearby images, so a radius $r = 350$ m was set since it covers three points in a straight line in our dataset. All image pairs that were further than this from each other had no edge in the graph representation. Besides, images that would become completely disconnected from any other image would also be removed (42 cases were found).

The full convex optimization problem becomes

$$\begin{align*}
\text{minimize} & \quad 1^T x + \lambda_{\text{ties}} 1^T |B^T x| + \frac{\lambda_{\text{geo}}}{2} x^T L_{\text{geo}} x \\
\text{subject to} & \quad 1^T x = 0 \\
& \quad \epsilon - b_n^T x \leq t_n \\
& \quad 0 \leq t_n, \ n = 1, \ldots, N,
\end{align*}$$  \hspace{1cm} \text{(7)}

adding weighting parameters to each term.

5.3. Experiments with synthetic data

We generated an artificial dataset to be able to assess the performance of our proposed method. This involved generating a map of 2544 points with scores obtained by randomly overlapping 400 Gaussian distributions across the city of New York. The radius for the geographical and temporal considerations was scaled to 1 km which was the distance covered by three points in a straight line. The temporal data was generated by taking 5 snapshots (representing 5 years) where scores randomly change following a Gaussian distribution with a standard deviation of $\frac{4}{5}$ since scores range from 0 to 10.

The next step of this process is emulating the provided pairwise comparisons collected through the website. Since there is an equal option, we assumed that people do not have a safety perception sensibility higher than 10% of the scale so we quantized scores by rounding them. Besides, comparisons are noisy because of biases, uncertainties and other factors. Following this, the closer two images are in known safety perception from our artificial dataset, the more likely it is that the winner of a given comparison is wrongly assigned. To model this, we define the function

$$P(\Delta Q = \Delta q | \Delta \tilde{q}) = \frac{f(\Delta q, \Delta \tilde{q})}{\sum_{i=10}^{10} f(i | \Delta \tilde{q})}$$  \hspace{1cm} \text{(8)}

where $\Delta Q$ is the discrete random variable “perceived score difference”, $\Delta q \in \mathbb{Z} \cap [-10,10]$ are the outcome possibilities (negative values mean that the right image is safer and vice versa), $\Delta \tilde{q}$ is the known score difference and $f \sim N(\Delta \tilde{q}, \sigma)$ (where $\sigma$ is a noise parameter).

This function is our model for the way people perceive the score difference and thus the probability of for a given set of images either left, right or equal are chosen is given by

$$P(\text{vote} = v) = \begin{cases} 
P(\Delta Q < 0) & \text{if } v = \text{left} \\
0 & \text{if } v = \text{equal} \\
1 - P(\Delta Q > 0) & \text{if } v = \text{right} 
\end{cases}$$  \hspace{1cm} \text{(9)}

This artificial dataset was useful to tune our parameters to minimize the Spearman’s rank correlation coefficient over 100 trials (as two different runs of the
Figure 6: Comparison side by side of the scores from all algorithms with the actual artificial scores with noise parameter $\sigma = 2$ and 8000 comparisons. It can be seen that taking into consideration the scores of the neighbors significantly improves the estimation of the scores, especially with fewer comparisons.

Figure 7: Cumulative density function of the ranking error for multiple amounts of comparisons at noise $\sigma = 2$ using convex optimization (green), convex optimization with comparisons only (purple), TrueSkill (red), naïve algorithm (blue), no prediction (black) and random prediction (yellow). Darker plots represent more data.

The proposed method outperforms the state-of-the-art rank-wise even at higher levels of noise and with fewer data. Since TrueSkill uses comparisons only, running our algorithm with the same information provides quite similar results, showing that taking distances in time and space between images makes a significant difference.
5.4. Application to the City-SAFE dataset
We finally applied the developed optimization problem to our dataset. The weight for the geographical regularization, \( \lambda_{\text{geo}} \), was scaled accordingly. The resulting maps can be seen in Figures 9 and 10.

6. Building a Perceived Safety Estimator
Having collected the comparisons and generated scores for each place, we now proceed to join this information with both image data and the Na minha rua data to create an estimator capable of predicting scores for new data.

6.1. Feature extraction
We represent image data using three different features to then assess which one performs the best. Firstly, we use the set of hand-crafted features used in the Streetscore project [8] which comprises the GIST descriptor, geometric color histograms, and geometric texton histograms; secondly, the set of outputs of the final fully connected layer of a VGG16 network pre-trained on the Imagenet dataset (without any fine-tuning) [24]; and the last set of features was the binary descriptor used in Section 4.

All image features were scaled following

\[
x'_i = \frac{x_i - \bar{x}_i}{\max(x_i) - \min(x_i)}.
\]  

(10)

As for the Na minha rua dataset, topic discovery using latent Dirichlet allocation (LDA) [25] was performed on the report descriptions. These texts were used to build a dictionary (discarding words with less than 15 appearances in total). Then, a model based on time frequency-inverse document frequency (TF-IDF) was generated to better capture the relevance of each word. After setting a number of topics (changing this number did not provide a significant difference in performance, so 9 was chosen), the model outputs the topic probability distribution for every report.

In order to assign a feature set to every image, we collect, for each image, the 10 closest reports (at most) and, using their topic probabilities as histograms, a mean histogram is computed. Only reports within a 200 m radius are considered and if no report is available within this radius, we discard that point for the estimation process.

6.2. Multiple kernel learning
The estimator is based on a support vector regression algorithm, where the different types of data are combined using two different kernel functions and multiple kernel learning (we implement using the shogun toolbox [26]). The parameters \( C \) and \( \epsilon \) were set at 0.03 and 0.001, respectively, following 5-fold cross-validation on the Spearman’s rank correlation coefficient.

For the image feature sets, a linear kernel is used as we do not want to over-fit the data.

\[
K_{\text{lin}}(x, y) = x^T y + c \quad x, y \in \mathbb{R}^n
\]  

(11)

The Na minha rua features were integrated using an histogram intersection kernel [27],

\[
K_{\text{hist}}(x, y) = \sum_{i=1}^{m} \min(x_i, y_i)
\]  

(12)

where \( m \) is the number of bins in the histogram (which corresponds to the total number of topics assumed).

6.3. Regression results
Figure 11 shows the cumulative density function of ranking error on the prediction for multiple sets of image features, paired with a 9-topic feature set from Na minha rua. The 4096-feature set extracted from a VGG network is the top performer, closely followed...
by the generic 5068-feature set and, then, by the 44-feature set of binary detections of semantic classes.

It is worth noting that a semantic representation of the image does a fairly good job when taking into consideration that it is only a binary descriptor for 44 classes. Using these features allows us to use the estimator afterwards without the need for image processing as one can just provide a semantic description of an image to estimate its score.

As Figure 12 shows, although the Na minha rua feature provides predictive power, using it with image features does not seem to improve the performance with regards to using the image features only.

7. Conclusions
This work puts forward the City-SAFE dataset, comprised of user-tagged pairwise comparisons for perceived safety. The performed user study shows evidence for a general agreement on perceived safety by citizens and also shows that human-related contents are seen as safer, unlike vehicle-related contents.

The extraction of scores for each image was performed using our convex optimization problem resulted in a better performing approach on noisy and few comparisons, when compared to TrueSkill (which was not designed to solve this problem and does not explore all the available information regarding images). Both score and ranking-based metrics were explored to measure performances as they capture differently the quality of the resulting scores. Improving the generation of scores using a probabilistic approach, and modeling the image score explicitly would be an interesting direction for future work in order to improve the constructed scores.

We finally developed an estimator (with a support vector regression) that uses the constructed scores as its ground truth target values to predict scores for new locations. Image features and features from a citizen reporting platform were combined using multiple kernel learning. Since our dataset is unbalanced towards the mean, the regression performance is affected. Improving this estimator would also be a future step in this work by integrating other data sources that have valuable predictive power regarding safety perception. The estimator obviously also benefits from improving the constructed ground truth scores.

References


