# Understanding Safety Based on Urban Perception

**Anonymous Author(s)**
Affiliation
Address
email

## Abstract

Perception is the way that humans can interact and understand our environment learning new experiences or reinforcing others. Safe Urban perception is defined as how humans can feel in front of a specific context determinate by the situation, in this case, in the streets. In that sense, techniques to predict security or improve the correlation between crime datasets and the places where they were reported are proposed. Like crime maps or statistical data or even more, applications that predict the criminal tendency of the zones. In this work, we propose a method to understand machine learning models trained to predict a safety score based on street photos. For example, if an image is classified as safe, we want to know what visual features make the model predict this score.

## 1 Introduction

Currently, there is an increasing number of methods to infer is a street is dangerous or not. In some cases, the goal is to create an application to predict criminal activities [1] or show crime maps [2]. These methods are based on crime datasets or statistical data. Some works predict the relationship between human perception of safety and the visual appearance of the streets [3, 4]. Then, the following works make use of this human perception datasets to propose models inferring a safety score based on street photos. Some works use deep neural networks [5], linear classifiers [6] or visual components [7].

## 2 Methodology

In this work, we train a neural network model to extract features from Google Street View images. These images have a safety-level score based on human perception available in the MIT Place Pulse Database Version 1.0. To make the model output explainable, we used a model-agnostic technique called LIME [8].

Using street images as input, we will train images with their correspond safety-level score, using the technique applied by ordoñez et al. [9] to assign a class depending of their scores. Once we train and test the model, we can use our explainer to interpret and understand why our input image is classified as “safe" or “not safe". This technique highlights visual features that made the model to determine the output. We present in detail our entire process in the following subsections.

### 2.1 Image Classification

To train our model we use the MIT Place Pulse database which consists of a set of images from different cities (New York and Boston) and scores calculated using human answer comparing 2 different images and answering for “Which Image Looks Safer?". To do that we used a fine-tuned strategy to train our classification model (safe vs. not safe). Then we fine-tune this network via back-propagation.

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

(a) Actual score: 8.35 (b) Actual score 1.06

(c) Predicted: safe (d) Prediction: not safe

Figure 1: Images from Boston (a,b) with different scores and results of LIME explainer over the Images (c, d).

### 2.2 Model Explanation

Model interpretation method helps us to get insights and understand our learning process. In our context, we can use them to visualize which visual features might be selected or are important to infer the model output. For instance, we want to understand why our street photos are predicted as “safe" or “not safe".

In this work, we use LIME, a local interpretable model-agnostic technique that explains a black-box model by simulating local candidates close to the original prediction generating a random distribution set of possible predictions based on L2 distance called “local fidelity" taken as reference the original prediction.

As we can see in Figure 1, this technique visualizes why our model is predicting some class (green and red pixels). This is very helpful to verify what parts of our input are being selected as “important". In this way, we can see if our model learns to associate scores with image features or not.

## 3 Experiments and Discussions

Using VGG16 [10] architecture pre-trained on the ImageNet dataset, Our training data is made of 4,132 images grouped by city. We train two models, one for New York and one for Boston. Then, we split the data into 60%, 20% and 20% for training, testing, and validation respectively. Our hyper-parameters are the batch size=64, epochs=100, learning rate=0.0001, and stochastic gradient descent as optimizer obtaining a 76% of testing accuracy in Boston and 69% in New York City.

To exemplify the model explainer, we selected 2 images and show the predictions and explanation with LIME. The first image has an actual safe score of 8.35 (“safe"), the second one has an actual safe score of 1.06 (“not safe"). As we can see in Figure 1, our test images were classified correctly. LIME produces two kinds of regions, the green areas called “pros” are the positive features that help our model to predict the correct class. The red areas called “cons" determine which features do not help in the prediction (See Fig.1). In Figure 1a, we have a photo from Boston with an actual score of 8.35 (very safe place). Our classifier predict this image as safe. LIME’s result is shown in Figure 1c, in this example, “pros” areas correspond to trees, and “cons" correspond to asphalt. We could run more experiment and see verify is green areas are more prevalent in “safe" images. We can corroborate this hypothesis in the second example (Figure 1b-d).

# References


[1] Panagiotis Stalidis and Theodoros Semertzidis and Petros Daras. Examining DeepLearning Architectures for Crime Classification and Prediction. 2018

[2] US Deparment of Justice, Mapping Crime: Principle and Practice.

[3] PLACE PULSE, MIT Media Lab, `http://pulse.media.mit.edu/data/`

[4] StreetScore, MIT Media Lab, `http://streetscore.media.mit.edu/`.

[5] Zhou, Bolei and Lapedriza, Agata and Xiao, Jianxiong and Torralba, Antonio andOliva, Aude. Learning Deep Features for Scene Recognition using Places Database.Advances in Neural Information Processing Systems 27, 2014.

[6] Nikhil Naik and Jade Philipoom and Ramesh Raskar and Cesar Hidalgo, StreetScore:Predicting the Perceived safety of one million streetscapes., IEEE Conference onComputer Vision and Pattern Recognition Workshops, 2014.

[7] Abhimanyu Dubey and Nikhil Naik and Devi Parikh and Ramesh Raskar and Cesar Hidalgo Deep Learning the City : Quantifying Urban Perception At A GlobalScale, 2016.

[8] Arietta, Sean M and Efros, Alexei A and Ramamoorthi, Ravi and Agrawala, Maneesh,City forensics: Using visual elements to predict non-visual city attributes, IEEEtransactions on visualization and computer graphics, 2014.

[8] Ribeiro, Marco Tulio and Singh, Sameer and Guestrin, Carlos. Why should i trustyou?: Explaining the predictions of any classifier, Proceedings of the 22nd ACMSIGKDD international conference on knowledge discovery and data mining, 2016.

[9] Vicente Ordonez and Tamara L. Berg. Learning High-level Judgments of Urban Per-ception, European Conference on Computer Vision (ECCV), 2014.

[10] Karen Simonyan and Andrew Zisserman, Very deep convolutional networks forlarge-scale image recognition, International Conference on Learning Representations(ICLR), 2014.