Detecting Sarcasm in Multimodal Social Platforms

Are Safer Looking Neighborhoods More Lively? A Multimodal Investigation into Urban Life

Marco De Nadai
FBK and University of Trento
Trento, Italy
denadai@fbk.eu

Radu L. Vieriu
University of Trento
Trento, Italy
radulaurentiu.vieriu@unitn.it

Gloria Zen
University of Trento
Trento, Italy
gloria.zen@unitn.it

Stefan Dragicevic
TIM and University of Trento
Trento, Italy
stefan.dragicevic@unitn.it

Nikhil Naik
MIT Media Lab
Cambridge, MA
naik@mit.edu

Michele Caraviello
TIM
Trento, Italy
michele.caraviello@telecomitalia.it

Cesar A. Hidalgo
MIT Media Lab
Cambridge, MA
hidalgo@mit.edu

Nicu Sebe
University of Trento
Trento, Italy
sebe@disi.unitn.it

Bruno Lepri
FBK
Trento, Italy
lepri@fbk.eu

ABSTRACT

Policy makers, urban planners, architects, sociologists, and economists are interested in creating urban areas that are both lively and safe. But are the safety and liveliness of neighborhoods independent characteristics? Or are they just two sides of the same coin? In a world where people avoid unsafe looking places, neighborhoods that look unsafe will be less lively, and will fail to harness the natural surveillance of human activity. But in a world where the preference for safe looking neighborhoods is small, the connection between the perception of safety and liveliness will be either weak or nonexistent. In this paper we explore the connection between the levels of activity and the perception of safety of neighborhoods in two major Italian cities by combining mobile phone data (as a proxy for activity or liveliness) with scores of perceived safety estimated using a Convolutional Neural Network trained on a dataset of Google Street View images scored using a crowdsourced visual perception survey. We find that: (i) safer looking neighborhoods are more active than what is expected from their population density, employee density, and distance to the city centre; and (ii) that the correlation between appearance of safety and activity is positive, strong, and significant, for females and people over 50, but negative for people under 30, suggesting that the behavioral impact of perception depends on the demographic of the population. Finally, we use occlusion techniques to identify the urban features that contribute to the appearance of safety, finding that greenery and street facing windows contribute to a positive appearance of safety (in agreement with Oscar Newman's defensible space theory). These results suggest that urban appearance modulates levels of human activity and, consequently, a neighborhood's rate of natural surveillance.

1. INTRODUCTION

Does a neighborhood's appearance of safety affect how active it is? For decades scholars from a variety of disciplines, but mainly from urban planning, have been exploring the potential connection between a neighborhood's appearance of safety and its levels of human activity.

The modern literature connecting safety, liveliness, and architecture, can be traced back to Jane Jacobs' seminal 1961 book: *The Death and Life of Great American Cities* [22]. In there, Jacobs introduced the eyes-on-the-street, or natural surveillance hypothesis [5], which suggests that citizens can maintain the safety of their neighborhoods naturally through continued surveillance. For natural surveillance to take place, however, Jacobs argued that neighborhoods needed to have certain physical qualities, such as well lit streets and buildings with street facing windows.

Jacobs' idea that the physical quality of a neighborhood can enhance its safety was later expanded by Oscar Newman's *defensible space theory* [39]. Defensible space theory expands on the idea of natural surveillance by suggesting that neighbors will be more likely to protect an area when there are clear physical demarcations separating what is considered public and private property [23, 39]. Examples of architectural markers of defensible space are archways in the entrance of building complexes, or staircases in the entrance of townhouses. These archways and staircases do not only serve an aesthetic purpose, but also, signal the boundary between a city's public space and the private and semi-private spaces that neighbors are expected to watch and defend.

Here, we strengthen the link between Jacobs' and Newman's theories by asking whether safer looking neighborhoods are more likely to experience more human activity–and hence, experience more natural surveillance. We explore this connection, by combining computer vision methods, that can be used to measure the physical characteristics of neighbor-

MM '16, October 15 - 19, 2016, Amsterdam, Netherlands

hoods [34, 44, 45, 43], with mobile phone data, which has become a common proxy for human activity [8, 11, 18, 20, 30], for two Italian cities (Rome and Milan). The combination of computer vision and mobile phone data helps us test whether safer looking neighborhoods are more active, and therefore, if neighborhoods that look physically safer could be experiencing more natural surveillance.

Our data provides support for a connection between appearance and activity. Using spatially filtered multivariate regressions we find that neighborhoods that are perceived as safer are more active than what is expected from their population density, the density of employees, and their distance to the city center. Also, we find that the perception of safety appears to modulate the relative population of females and adults, with unsafer looking neighborhoods experiencing a lower number of female and people over 50 than safer looking neighborhoods. Conversely, we find that younger populations are disproportionately more active in unsafe looking neighborhoods. Finally, we use occlusion techniques to identify the areas of an image that trigger a positive or negative evaluation of safety in the Artificial Neural Network, finding that greenery and street facing windows tend to be associated with higher levels of safety, as suggested by Oscar Newman's defensible space theory. These observational results strongly suggest–but don't causally prove–that the appearance of neighborhoods has an effect on their levels of human activity, and potentially, on a neighborhood's level of natural surveillance.

2. RELATED WORK

The connection between urban perception and human activity speaks primarily to two streams of literature. The first one is the stream of literature focused on the environmental factors contributing to crime, which has a long tradition in criminology and urban sociology. While our paper does not focus on crime per se, the connection between the physical appearance of neighborhoods and natural surveillance suggested by Jacobs and Newman makes our results relevant to that stream of literature [22, 23, 39]. The second one is the stream of literature using surveys, and more recently, computer vision methods to quantify people's perception of urban environments [35, 37, 45].

2.1 Neighborhood appearance and crime

Beyond Jacob's and Newman's theories, the most widely known theory suggesting a connection between urban perception and crime is the *broken windows theory* (BWT) of Wilson and Kelling [26, 55]. The BWT is the hypothesis that urban incivilities, such as broken windows and litter, promote criminal activity. The classical mechanism used to justify the theory says that urban incivilities signal lawlessness, and may cause the offenders of small incivilities to scale their criminal behavior to more predatory forms of crime if they are not reigned in. The policy implications of the BWT, however, vary from community policing—the promotion of ties between police officers and their communities—to zero-tolerance polices, which promote cracking down on all minor offences to deter more serious forms of crime.

Evidence in favor of the broken windows theory has been presented by Kelling and Coles [26], who looked at data and stories from New York to Seattle to argue that community policing is an effective way to deter more serious forms of crime. Kelling and Sousa [27] provide additional evidence by using an extensive dataset on crime, demographics, and economic data from New York City. More recently, Corman and Mocan [3] used New York City data on the policing of misdemeanors (as a proxy for broken windows policing), and on robbery, car theft, and grand larceny, to provide evidence in support of broken windows policing.

The broken windows theory is also supported by a few field experiments, such as those conducted by Keizer *et al.* in The Netherlands [25]. In six experiments, Keizer *et al.* intervened environments by spraying graffiti on walls, or leaving supermarket carts unattended and studied the behavior of subjects, in both the presence and absence of disorder, to see when people broke norms (such as littering). Their data showed a significant increase in people's norm breaking behavior when they were in the presence of disorder.

But not all of the evidence collected to test the BWT, and its policy implications, is favorable to it [13]. In a 2006 paper, Bernard Harcourt re-analyzed the data presented by Kelling and Sousa [27] and found no evidence of the effectiveness of the broken windows policing [14]. More recently, Harcourt and Ludwig used a dataset of more than fifty thousand marijuana related arrests to provide evidence that community policing is not only ineffective, but that it also unfairly targets minorities [15].

Moreover, the BWT has been criticized by work showing that the social and ethnic context of a neighborhood may matter more than urban disorder. In Sampson *et al.* [48] and Sampson and Raudenbush [46, 47], community data from Chicago was used to argue that racial and economic context were more predictive of disorderly behavior than physical disorder. To help bridge their results with the literature, Sampson and Raudenbush [46] proposed an alternative interpretation of the theory, where both neighborhood disorder and crime, are manifestations of a lack of informal forms of control within disengaged and distrusting communities. The reframed theory, therefore, interprets the link between disorder and crime as a manifestation of the lack of informal forms of social control and organization.

2.2 The social and computational image of the city

The second literature our paper speaks to is the literature measuring urban perception and understand its social and economic implications. The original literature on this topic can be traced back to seminal work by the urbanist Kevin Lynch [31], who interviewed people in Boston, Jersey City, and Los Angeles, to understand the large scale image of cities that people made in their heads. This work on a city's imageability was later continued by social psychologists like Stanley Milgram [33], and by urbanists like Jack Nasar [37, 38], who created evaluative maps of cities, also using survey methods.

More recently, however, this literature started leveraging crowdsourcing [45] and computer vision methods [35, 43] to improve the scale, precision, and resolution of the evaluative maps created.

On the data collection side of this literature, Salesses *et al.* [45] created a large crowdsourced visual perception survey to measure people's perception of streetscapes, and to create comparable evaluative maps for New York, Boston, Linz, and Salzburg, which they also used to measure the segregation and inequality of experiences in these cities, and to show

that violent crime correlates with the variance of appearence of safety in an area.

These new sources of crowd-sourced data gave rise to studies looking to understand the features of an image that explain how streetscapes are perceived. On a recent study, Quercia *et al.* [44] investigated which visual aspects of London neighborhoods make them appear beautiful, quiet and/or happy. A related study by Porzi *et al.* [43] identified the visual elements that contributed to an image's perceived level of safety.

But to scale the study of urban perception to multiple cities, and to high spatial resolutions, researchers begun developing computer vision methods to score millions of images [35]. These computer vision approaches build on research predicting human perception from visual data [12, 21, 49] and on research analyzing visual streetscapes for city understanding [1, 2, 6, 10, 28, 57]. The latter of these two lines of research has been fueled by the widespread availability of geolocated image data, such as Google Street View maps or city snapshots publicly shared on social networks (e.g. Flickr, Instagram) [1, 51, 57]. Using geotagged image data Doersch *et al.* [6] showed that geographically representative visual elements, like architecture styles, can be automatically discovered from Street View Imagery. In a dynamic study, Naik *et al.* [34] used computer vision and images from different time periods to measure urban change, and to study the factors that contribute to neighborhood improvement. Computer vision methods have also been used to show that a city's visual attributes work as proxy of social and economic characteristics, such as crime rates and proximity to local businesses [2, 28], or census characteristics, such as income and inequality [36].

This new wave of research has benefited from advances in deep learning, which have been used not only to measure appearance, but also for place recognition. Zhou *et al.* [57] introduced Places205 Dataset, a large data collection gathering more than 7 million labeled pictures of scenes. They achieve state-of-the-art classification results by training deep Convolutional Neural Networks (CNNs). More recently, Arandjelović *et al.* [1] introduce NetVLAD, a modified CNN architecture able to address large scale visual place recognition. In this work, we build on top of recent work in place recognition [1, 57] to fine-tune a deep CNN architecture and show experimentally that under scarce training data, sample augmenting helps achieve state-of-the-art results on safety prediction from streescape images.

3. DATASETS

Next, we describe the datasets used to estimate urban appearance, and human activity.

3.1 Urban appearance data

We use urban appearance data from the Place Pulse dataset[1]. Place Pulse is a large, crowdsourcing project on human perception of cities. The data collection is designed as an online game, where participants are shown two images of streetscapes and are asked to choose one image in response to an evaluative question such as: Which place looks safer? Or: Which place looks more lively? A score is later computed for each image using the TrueSkill [16] algorithm.

Place Pulse begun as Place Pulse 1.0 (PP1), which scored 4,109 images from two US cities (New York City and Boston) and two European cities (Linz and Salzburg), and was launched publicly in 2011. PP1 scored images across three evaluative dimensions: *Safety*, *Upper-Class* and *Unique*.

The current version of Place Pulse (Place Pulse 2.0), launched publicly in July 2013, extended the data collection effort to 56 cities from all continents (except Antartica) and to six evaluative questions: *Safety*, *Wealthy*, *Boring*, *Lively*, *Depressing* and *Beautiful*. For more details, please see Dubey *et al.* [7]

Here we use data from the Place Pulse 2.0 dataset for Milan and Rome (PP2-I). The data includes 3,897 images that received about 25,000 evaluations for their perception of safety[2] - corresponding to an average of 7.6 clicks per image.

3.2 Mobile phone activity data

To proxy human activity we use mobile phone billing and operation data. The data records the time of communication, and the radio base station that handled it, for various types of communication (e.g. incoming calls, Internet, outgoing SMS). The natural coarsening of this data are grid cells with an area that is proportional to the underlying coverage area of the radio base stations. The cell size is $\sim 300 \times 300$ m in the city centre of Milan and it increases up to $\sim 2,300 \times 2,300$ m in the peripheral part of the cities, where few customers are served per unit of area. For each grid cell, we count the number of people who made or received a call on an hourly basis, broken down by gender (number of males/females) and age.

Call records were provided by Telecom Italia Mobile (TIM), which is the largest mobile operator in Italy with a market share of 34%[3]. Our data are aggregated every 60 minutes, and includes both TIM customers and roaming customers in Milan and Rome, and covers the time ranging from February to June 2015.

We note that our data cannot distinguish between pedestrians and people using their phones in their homes, so our measures of activity proxy the number of TIM customers in an area, but not necessarily in the street.

4. METHODOLOGY

This paper focuses on investigating the relationship between the appearance of safety and the activity or liveliness of neighborhoods in Rome and Milan. To achieve this goal we estimate the appearance of safety for different districts of the city by spatially aggregating the safety scores obtained from the images within these areas, and then, observe people's activity using mobile phone data. To achieve enough coverage, we densify our maps by collecting additional images and scoring them using computer vision. Finally, we spatially aggregate scores including census tracts for the 2011 Italian census. In the reminder of this section we explain the methodology used to score images and to measure activity.

4.1 Measures of Urban Appearance

4.1.1 Safety Perception from Visual Data

We use deep Convolutional Neural Networks (CNNs) as our model for predicting safety perception from streetscape

[1]http://pulse.media.mit.edu

[2]We include in this sum only pairwise comparisons involving either two Italian cities or one Italian vs one non-Italian city

[3]http://bit.ly/1LtNrFY

imagery. We base this choice on the recent success of CNNs in various computer vision problems (especially object classification, object detection and scene recognition [29, 40, 50]). We show that fine-tuning a CNN on predicting the level of safety along with data augmentation leads to improved performance when compared to recent work on the same task [35, 41].

Since we have a limited amount of training data (*i.e.* a sample of a few thousands), we retrain CNNs trained on related domains (e.g. images captured in urban environments that are likely to show similar visual content). In particular, we fine-tune the well known AlexNet CNN [29], trained on Places205 Dataset [57], also known as *Places205-AlexNet*. This model was trained on 205 categories of scenes (many of which capture different areas from cities such as office buildings, churches, residential neighborhoods, shops, etc.) summing up around 2.5 million images. By retraining a CNN previously trained in a similar domain we transfer some of the knowledge contained in this network. In our case, we find that the resulting network can accurately predict the appearance of safety, as we show in the validation section (see section 4.1.2).

To increase the generalization ability of the trained CNN, we adopt data augmentation by cropping all the images used during training and testing. We find this particularly suitable for our scenario where no constraints regarding image alignment are imposed. Specifically, for every image in the training set, we generate n crops by randomly assigning values to the coordinates of the top left and bottom right cutting points, respectively. We control the size of the crops by bounding the coordinates of the cutting points to ranges proportional with the image size. Formally, given an image I of size $W \times H \times 3$, we generate points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ such that the quantities: $x_1/W, y_1/H, 1-x_2/W$ and $1-y_2/H$ are bounded by k_1 and k_2, respectively. We empirically set k_1 to 0.05, k_2 to 0.2 and n to 30 in all our experiments. Additional constraints to control diversity of the crops could be implemented, such as monitoring the *intersection over the reunion (IoU)* between pairs of crops and/or original image. During testing, we average all predictions belonging to the crops of the same test sample. The safety scores obtained at image level are then aggregated into city's districts.

Part of the standard pre-processing steps, all the images are subject to scaling to 227×227 pixels and mean image subtraction. The task is modeled as a regression problem, where the goal is to minimize the l_2 loss between the sample labels and the model predictions. As labels, we refer to the Trueskill scores computed in [35], also used in [41]. We fine-tune with *Caffe* [24] for $10,000$ iterations using a base learning rate of $1e^{-4}$. We note no further decrease in the training loss beyond this limit.

4.1.2 Validation of computer vision models

We validate the fine-tuned CNN on the two US cities from PP1 (New York and Boston), following training and evaluation protocols from [35] and [41], respectively. In the first case, we perform a 5-fold cross-validation on the 2920 US images and report the average R^2 measure, after scaling all the scores to the interval $[0, 10]$. We obtain an average R^2 of 62.2% a 9.5% relative improvement over the best result reported in [35] and 4.8% over the case of not using data augmentation during testing. In the second case, the source (train) and target (test) domains are populated by alternating US cities (e.g. NY - Boston, NY - NY, Boston - Boston, Boston - NY). As performance measure, the authors report Pearson correlation between the predicted regression values and the label scores. Table 1 shows the comparison results, where we consistently outperform [41] in all four combinations. We also observe how the difficulty of the task (for different pairs) modulates the predicting performance of both sets of models in the same way (e.g. training and testing on Boston seems to be the easiest for both our models and the ones from [41]).

Model type	Best from [41]	Our result
NY - NY	0.687	0.718
Boston - Boston	0.718	0.744
NY - Boston	0.701	0.734
Boston - NY	0.636	0.693

Table 1: Performance on the estimated level of safety (comparison with [41]). For all cases, we report Pearson correlations with $p < 0.001$.

4.1.3 Densification of Appearance data using Computer Vision

Since the distribution of the annotated images from PP2-I is sparse, on average 6.7 images/km^2, we retrieve additional images for the cities of Rome and Milan by densely sampling geo-referenced images from Google Street View. Data densification for the analysis of urban landscape has been used in previous works for generating high resolution maps of city perception [35, 41]. First, we generate a grid of points inside the area we want to cover. The granularity we choose is 100 points/km^2. To better represent the safety perception of the location, we retrieve four images from each location, which have 90 degrees horizontal field of view and different headings (north, east, south and west). By doing this we cover 360 degrees from each location, thus getting less biased safety perception of an area by averaging predicted safety scores of these four images. We developed the script which iterates through all the points and used Google Street View API to obtain four images for each location. The script discards locations where no images are available. Using this method we obtained 83,203 images for Rome and 74,815 images for Milan.

4.1.4 Validation of Densification

Table 2 reports the performances on predicting safety perception for the city of Rome and Milan. We evaluate the correlation between the aggregated original scores and (i) the predictions over the images from PP2-I and (ii) the predictions over the images from the densified dataset. In general, a slight decrease in performance is observed in the second case. We can attribute this loss in performance to the fact that images from PP2-I were retrieved in 2010, thus a variation in the urban appearance may have occurred within this time lapse. For example, the Expo Milano 2015 - a Universal Exposition - was held from May to October 2015, and Milan underwent a profound (visual) change in its northwest area to prepare for this event.

Predicting safety on Milan and Rome: We are interested in finding a good candidate model for labeling the densely sampled Street View images from Milan and Rome. We experimented with several model choices (including train-

City	PP2-I	Densified
Milan	0.621	0.488
Rome	0.635	0.548

Table 2: Performance on the estimated safety perception level for each city. All values are statistically significant, with $p < 0.001$.

ing on PP2-I) and discovered that, surprisingly, using PP1 for training yields the best correlation value for the two Italian cities. We attribute this result to the much inferior average number of votes per image (around 7.6 for PP2-I, compared to around 90 for PP1). For the rest of the experiments, we use only the model trained on PP1 for safety prediction. In Figure 1 we visually report the spatial distribution of the safety prediction for Milan and Rome.

4.2 Metrics for Urban Liveliness or Activity

Next, we define the metrics we use as proxies for an area's level of activity, or liveliness. Here we study only in urban areas where more than 50% of the surface is not composed by farmlands or forests. We measure activity for four populations, all people, females, people younger than 30, and people older than 50.

Formally, we measure the density of all people in district i as:

$$R_p(i, 24h) = \frac{|people_{i,24h}|}{area_i} \quad (1)$$

Next, we measure the fraction of females in a district i as:

$$R_f(i, 24h) = \frac{|females_{i,24h}|}{|people_{i,24h}|} \quad (2)$$

Additionally, we measure the population of people below 30 and above 50 as:

$$R_{<30}(i, 24h) = \frac{|people(<30)_{i,24h}|}{|people_{i,24h}|} \quad (3)$$

$$R_{>50}(i, 24h) = \frac{|people(>50)_{i,24h}|}{|people_{i,24h}|} \quad (4)$$

.

4.3 Spatial Regression

We test the connection between urban appearance of safety and activity using spatially corrected Ordinary Least Squares (OLS) regressions. Since we are dealing with spatial variables, OLS residuals are assumed not to be spatially auto-correlated; otherwise the regression model is said to be misspecified. Thus, we use the Griffith filtering [53] which extracts a set of orthogonal and uncorrelated eigenvectors from the expression:

$$(I - \frac{11^T}{n})W(I - \frac{11^T}{n}) \quad (5)$$

derived from the spatial auto-correlation Moran's I numerator, where I is a $(n \times n)$ identity matrix, 1 is a $n \times 1$ vector containing only 1's and W is a $(n \times n)$ spatial weight matrix based on topological adjacency, so-called Queen criterion: if two areas share a boundary or a vertex, the entity of the spatial weight matrix is coded as 1, and otherwise, 0. The eigenvectors obtained can be employed in a multivariate regression to account for spatial auto-correlation. However, it is clear that employing all n eigenvectors in a regression framework is not desirable for reasons of model parsimony. Thus, a subset of eigenvectors are selected in a step-wise fashion so as to minimize the sequential residual spatial correlation (Moran's I) values [53]. The final subset of candidate eigenvectors represents the *spatial filter* for the variable analysed. Before applying the regression model, the data were Z-score scaled.

5. RESULTS

After describing our methods to measure urban appearance and neighborhood activity we test whether the appearance of safety and the activity of neighborhoods is correlated. To test for this correlation we merge our data with information from the Italian census so we can control for other sources that intuitively correlate with neighborhood activity: population density (residential), density of employment (which should also proxy pedestrian density during the day), distance to city centre, and a deprivation index (to control for poverty in the neighborhood).

We begin by looking simply at the correlation between the number of people per unit of area observed in our mobile phone dataset and the appearance of safety in a neighborhood while controlling for population density, employee density, deprivation, and distance to centre.

Table 3 shows the result of a spatially corrected multivariate regression with the number of people per unit of area measured using the density of all people as the dependent variable. Not surprisingly, the strongest correlate of the number of people present per unit of area is employee density, and the number of people per unit area decreases with distance to the center. Yet, despite the strong effect of the other control variables, the appearance of safety is significantly and positively correlated with the number of people present in a neighborhood per unit area.

Presence of people (1)[1]	
Population density[1]	0.155**
Employees density[1]	0.328**
Deprivation	-0.022
Distance centre	-0.257**
Safety appearence	0.105**
Spatial Eigenvectors	11
Adj-R^2	0.91
Moran's I (p-value)	0.07 (0.08)

[1] log transformed variable.

Table 3: OLS regression model between presence of people and safety perception. The β coefficients are reported in the table. *$p < 0.01$, **$p < 0.001$.

Next, we look at the fraction of females present in an area. Looking at the population of females separately is motivated by empirical research showing that women are twice as likely as men to report feeling unsafe [54], even though they have a much smaller risk of being victimized [42, 52]. This suggest that the presence of women in a neighborhood should be more strongly affected by its appearance of safety than the presence of men. In fact, Felson and Clarke [9] suggest that a high ratio of women in the street is a positive sign towards urban safety, as they act as "crime detractors," in agreement

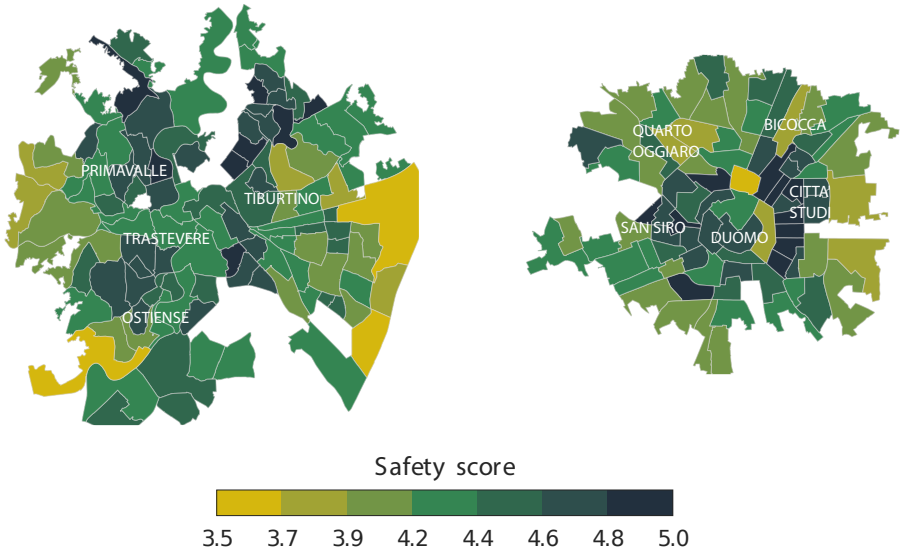

Figure 1: Spatial distribution of perceived safety in each district of Rome and Milan.

with Jacobs' natural surveillance hypothesis. These theories would suggest that the ratio of females is lower in places perceived as unsafe.

Table 4 looks at the ratio of females in the population observed in our cell phone data as the dependent variable, finding that the appearance of safety is highly significant and positive. In fact, the coefficient is roughly twice that observed for the general population.

Presence of women (2)	
% of women (residents)[s]	0.001
Deprivation	-0.005
Distance centre	-0.003
Safety perception	0.020**
Spatial Eigenvectors	12
Adj-R^2	0.65
Moran's I (p-value)	0.06 (0.11)

[s] cube-root transformed variable.

Table 4: OLS regression model between presence of women and safety perception. The β coefficients are reported in the table. *p < 0.01, **p < 0.001.

We also look at the proportion of people younger than 30 and older than 50 in an area. According to Felson and Clarke [9], a younger population is a predictor of criminal incidents in an area, as they show a higher aggression potential compared to older populations. Nevertheless, younger people, especially men, show less fear of crime [19, 32].

Table 5 looks at the ratio of people younger than 30 in the population observed in our cell phone data as the dependent variable. The variable which contributes the most to the correlation is the distance from the city centre, followed by appearance of safety which is highly significant and negative. Contrarily, when we look at the ratio of people older than 50 (Table 6), we find that the appearance of safety is highly significant and positive. This is in agreement with the theory, showing that older people are more likely to be present in places that appear safe.

Presence of people younger than 30 (3)[l]	
% of younger residents[l]	-0.001
Deprivation	0.032**
Distance centre	-0.150**
Safety perception	-0.048**
Spatial Eigenvectors	16
Adj-R^2	0.66
Moran's I (p-value)	0.07 (0.09)

[l] log transformed variable.

Table 5: OLS regression model between presence of younger people and safety perception. The β coefficients are reported in the table. *p < 0.01, **p < 0.001.

Presence of elderly people (4)	
% of elderly residents[s]	0.006**
Deprivation	-0.006**
Distance centre	0.006**
Safety perception	0.017**
Spatial Eigenvectors	14
Adj-R^2	0.64
Moran's I (p-value)	0.07 (0.09)

[s] cube-root transformed variable.

Table 6: OLS regression model between presence of elderly people and safety perception. The β coefficients are reported in the table. *p < 0.01, **p < 0.001.

5.1 Visual Attributes Determining Safety

Finally, we explore the visual attributes of the images that contribute positively, or negatively, to their appearance of safety. To identify these attributes we set up an occlusion sensitivity experiment. In this experiment, inspired by [56], we randomly generate occluding patches in images and replace them with the average pixel value. For every such altered

Figure 2: (Top) Sample images associated to a (left) low and (right) high level of safety and corresponding activation masks: highlighted areas correspond to the ones that mostly contribute to the perception of (center) unsafety and (bottom) safety.

image, we monitor the effect at the output of the predictor (did the image score higher or lower in its appearance of safety). This allows us to identify patches in an image that contribute positively or negatively to their appearance of safety.

Figure 2 shows some examples, with the original image on the (top row), followed by the areas that contribute to a low appearance of safety (middle row) and a high appearance of safety (bottom row). The images are sorted from an overall low appearance of safety to a high one. The examples, while illustrative instead of comprehensive, show that street facing windows and greenery tend to to contribute positively to an streetscapes appearance of safety. The positive effect of street facing windows is in agreement with the natural surveillance hypothesis of Jane Jacobs.

6. CONCLUSION

In this paper we explored the question: "*Are safer looking neighborhoods more lively?*" in the context of two Italian cities: Milan and Rome. Our findings suggest that perceived safety modulates the active population in an area, with effects that depend on age and gender. The overall effect of the appearance of safety in activity appears to be positive, even after controlling for population, employment density, and distance to the city center. Yet, the effect does not appear to be universal, and depends on the demographic with the population, with females and people older than 50 appearing to have a stronger preference for the appearance of safety.

Our results, however, do not provide a causal explanation of the observed effects. For instance, our data cannot distinguish between the hypotheses that people over than 50 prefer safer looking places, or that they modify their homes and shops to make the places they live and work at look safer. Nevertheless, they provide preliminary evidence suggesting a connection between the appearance of safety and levels of human activity that is strong enough to manifest itself at the city scale.

These methods, which could be readily applied to other cities if the data were available, can help improve modern efforts to use computational methods for urban recommendations. Some recent literature has focused on developing algorithms to recommend places for new business by using data on the presence of amenities in neighborhoods [17] and other urban features [4].

7. ACKNOWLEDGMENTS

Most of the computation of this article was done using free software and we are indebted to the developers and maintainers of the following packages: python, pandas, scikits.statsmodels, pysal to mention only a few.

8. REFERENCES

[1] R. Arandjelović, P. Gronat, A. Torii, T. Pajdla, and J. Sivic. NetVLAD: CNN architecture for weakly supervised place recognition. *arXiv preprint arXiv:1511.07247*, 2015.

[2] S. M. Arietta, A. A. Efros, R. Ramamoorthi, and M. Agrawala. City forensics: Using visual elements to predict non-visual city attributes. *IEEE TVCG*, 20(12):2624–2633, 2014.

[3] H. Corman and N. Mocan. Carrots, sticks, and broken windows. *Journal of Law and Economics*, 48(1):235–266, 2005.

[4] M. De Nadai, J. Staiano, R. Larcher, N. Sebe, D. Quercia, and B. Lepri. The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective. In *ACM WWW*, 2016.

[5] H. Doeksen. Reducing crime and the fear of crime by reclaiming New Zealand's suburban street. *Landscape and urban planning*, 39(2):243–252, 1997.

[6] C. Doersch, S. Singh, A. Gupta, J. Sivic, and A. A. Efros. What makes Paris look like Paris? *Communications of the ACM*, 58(12):103–110, 2015.

[7] A. Dubey, N. Naik, D. Parikh, R. Raskar, and C. A. Hidalgo. Deep learning the city: Quantifying urban perception at a global scale. *ECCV*, 2016.

[8] N. Eagle, M. Macy, and R. Claxton. Network diversity

and economic development. *Science*, 328(5981):1029–1031, 2010.

[9] M. Felson and R. V. Clarke. Opportunity makes the thief. *Police research series, paper*, 98, 1998.

[10] E. L. Glaeser, S. D. Kominers, M. Luca, and N. Naik. Big data and big cities: The promises and limitations of improved measures of urban life. Working Paper 21778, National Bureau of Economic Research, 2015.

[11] M. Gonzalez, C. Hidalgo, and L. Barabasi. Understanding individual mobility patterns. *Nature*, 453(7196):779–782, 2008.

[12] H. Grabner, F. Nater, M. Druey, and L. Van Gool. Visual interestingness in image sequences. In *ACM Multimedia*, 2013.

[13] B. E. Harcourt. *Illusion of order: The false promise of broken windows policing*. Harvard University Press, 2009.

[14] B. E. Harcourt and J. Ludwig. Broken windows: New evidence from new york city and a five-city social experiment. *The University of Chicago Law Review*, pages 271–320, 2006.

[15] B. E. Harcourt and J. Ludwig. Reefer madness: Broken windows policing and misdemeanor marijuana arrests in new york city, 1989-2000. *Criminology and Public Policy*, 2007.

[16] R. Herbrich, T. Minka, and T. Graepel. Trueskill$^{\text{TM}}$: A Bayesian skill rating system. In *NIPS*, 2006.

[17] C. A. Hidalgo and E. E. Castañer. Do we need another coffee house? The amenity space and the evolution of neighborhoods. *arXiv preprint arXiv:1509.02868*, 2015.

[18] C. A. Hidalgo and C. Rodriguez-Sickert. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications*, 387(12):3017–3024, 2008.

[19] W. Hollway and T. Jefferson. The risk society in an age of anxiety: situating fear of crime. *British journal of sociology*, pages 255–266, 1997.

[20] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, J. Rowland, and A. Varshavsky. A tale of two cities. In *ACM HotMobile*, pages 19–24, 2010.

[21] P. Isola, D. Parikh, A. Torralba, and A. Oliva. Understanding the intrinsic memorability of images. In *NIPS*, 2011.

[22] J. Jacobs. *The death and life of American cities*. Random House, 1961.

[23] J. M. Jacobs and L. Lees. Defensible space on the move: revisiting the urban geography of alice coleman. In *International Journal of Urban and Regional Research 37.5 (2013): 1559-1583*, 2013.

[24] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional Architecture for Fast Feature Embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[25] K. Keizer, S. Lindenberg, and L. Steg. The spreading of disorder. *Science*, 322(5908):1681–1685, 2008.

[26] G. L. Kelling and C. M. Coles. *Fixing broken windows: Restoring order and reducing crime in our communities*. Simon and Schuster, 1997.

[27] G. L. Kelling and W. H. Sousa. *Do Police Matter?: An Analysis of the Impact of New York City's Police Reforms*. CCI Center for Civic Innovation at the Manhattan Institute, 2001.

[28] A. Khosla, B. An, J. J. Lim, and A. Torralba. Looking beyond the visible scene. In *CVPR*, 2014.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[30] M. Lenormand, M. Picornell, O. G. Cantú-Ros, T. Louail, R. Herranz, M. Barthelemy, E. Frías-Martínez, M. San Miguel, and J. J. Ramasco. Comparing and modelling land use organization in cities. *Royal Society Open Science*, 2(12), 2015.

[31] K. Lynch. *The image of the city*, volume 11. MIT press, 1960.

[32] W. Mark. Fear of victimization: Why are women and the elderly more afraid? *Social science quarterly*, 65(3):681, 1984.

[33] S. Milgram. The experience of living in cities. *Science*, 167(3924):1461, 1970.

[34] N. Naik, S. D. Kominers, R. Raskar, E. L. Glaeser, and C. A. Hidalgo. Do people shape cities, or do cities shape people? The co-evolution of physical, social, and economic change in five major US cities. Technical report, National Bureau of Economic Research, 2015.

[35] N. Naik, J. Philipoom, R. Raskar, and C. Hidalgo. Streetscore–predicting the perceived safety of one million streetscapes. In *CVPRW*, 2014.

[36] N. Naik, R. Raskar, and C. A. Hidalgo. Cities are physical too: Using computer vision to measure the quality and impact of urban appearance. *The American Economic Review*, 106(5):128–132, 2016.

[37] J. L. Nasar. *The evaluative image of the city*. Sage Publications Thousand Oaks, CA, 1998.

[38] J. L. Nasar, B. Fisher, and M. Grannis. Proximate physical cues to fear of crime. *Landscape and urban planning*, 26(1):161–178, 1993.

[39] O. Newman. *Defensible space*. Macmillan New York, 1972.

[40] M. Oquab, L. Bottou, I. Laptev, and J. Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *CVPR*, 2014.

[41] V. Ordonez and T. L. Berg. Learning high-level judgments of urban perception. In *ECCV*, 2014.

[42] C. Pantazis. 'Fear of Crime', Vulnerability and Poverty. *British journal of criminology*, 40(3):414–436, 2000.

[43] L. Porzi, S. Rota Bulò, B. Lepri, and E. Ricci. Predicting and Understanding Urban Perception with Convolutional Neural Networks. In *ACM Multimedia*, 2015.

[44] D. Quercia, N. K. O'Hare, and H. Cramer. Aesthetic capital: what makes London look beautiful, quiet, and happy? In *ACM CSCW*, 2014.

[45] P. Salesses, K. Schechtner, and C. A. Hidalgo. The collaborative image of the city: mapping the inequality of urban perception. *PloS one*, 8(7):e68400, 2013.

[46] R. J. Sampson and S. W. Raudenbush. *Disorder in urban neighborhoods: Does it lead to crime*. US Department of Justice, Office of Justice Programs, National Institute of Justice, 2001.

[47] R. J. Sampson and S. W. Raudenbush. Seeing disorder: Neighborhood stigma and the social construction of

"broken windows". *Social psychology quarterly*, 67(4):319–342, 2004.

[48] R. J. Sampson, S. W. Raudenbush, and F. Earls. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328):918–924, 1997.

[49] A. Sartori, V. Yanulevskaya, A. A. Salah, J. Uijlings, E. Bruni, and N. Sebe. Affective analysis of professional and amateur abstract paintings using statistical analysis and art theory. *ACM TiiS*, 5(2):8, 2015.

[50] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[51] J. Sivic and A. A. Efros. Urban-Scale Quantitative Visual Analysis. *Smart Cities*, page 43, 2014.

[52] R. B. Taylor and M. Hale. Testing alternative models of fear of crime. *The Journal of Criminal Law and Criminology (1973-)*, 77(1):151–189, 1986.

[53] M. Tiefelsdorf and D. A. Griffith. Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A*, 39(5):1193–1221, 2007.

[54] G. R. Wekerle and C. Whitzman. *Safe cities: Guidelines for planning, design, and management*. Van Nostrand Reinhold Company, 1995.

[55] J. Q. Wilson and G. L. Kelling. Broken windows. *Critical issues in policing: Contemporary readings*, pages 395–407, 1982.

[56] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *ECCV*, 2014.

[57] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In *NIPS*, 2014.