Quantifying Urban Safety Perception on Street View Images

Felipe Moreno-Vera
felipe.moreno@ucsp.edu.pe
Universidad Católica San Pablo
Arequipa, Perú

Bahram Lavi
bahram.lavi@fgv.br
Fundação Getúlio Vargas
Rio de Janeiro, Brazil

Jorge Poco
jorge.poco@fgv.br
Fundação Getúlio Vargas
Rio de Janeiro, Brazil

ABSTRACT

In the last 40 years, Urban perception has become an important research area covering several fields, such as criminology, psychology, urban planning, Broken windows theory. It aims to analyze and interpret the behavior of the perception in cities. Urban perception focuses on understanding urban environments based on the characteristics of the city. With the rapidly increasing data availability and highly scalable data collection methods powered by modern web services, new techniques from other domains enabled the exploration of solutions to estimate urban perception (*i.e.*, quantify urban perception autonomously). This work presents a methodology to explore the urban perception analysis task. The work relies on the benchmark dataset, Place Pulse. This dataset is used to perform our classification tasks concerning the category of *safety* in urban perception problems.

CCS CONCEPTS

• **Urban Computing** → **Urban Perception**; • **Deep Learning** → *Perceptual Learning*.

KEYWORDS

Urban Perception, Urban Computing, Computer Vision, Perception Computing, Deep Learning, Street-level imagery, Street View, Cityscape, Perception Learning, Place Pulse, Image Pre-processing, Safety Perception, Feature Extraction

ACM Reference Format:
Felipe Moreno-Vera, Bahram Lavi, and Jorge Poco. 2021. Quantifying Urban Safety Perception on Street View Images. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI-IAT '21), December 14–17, 2021, ESSENDON, VIC, Australia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3486622.3493975

1 INTRODUCTION

Nowadays, urban perception has a vital role in society. It helps to explore several areas such as urban planning, urbanism, urban development. Numerous social-psychological efforts have advanced to understand the behavior of a city and its influence on society. “Cities are designed to shape and influence the lives of their inhabitants” [17]; the work shows the structure of the city is a consequence of their population. Multiple studies focused on the relation of the visual appearance of the city and their possible role and impact in human perception and their lifestyle [19].

Another notable example is the theory of “Broken Window” [40]. The theory asserts that visual signs of environmental disruption (*e.g.*, broken windows, abandoned cars, trash, graffiti) can induce social consequences and likely increase crime. To appraise the trustfulness of this theory, some social experiments were performed [12, 19, 33, 37] establishing whether the visual appearance can impact the behavior of the citizens or not. Other studies showed that the visual aspect of a city strongly affects the psychological state of the inhabitants [11, 17]. On the other hand, some studies showed that green areas in urban cities positively impact the safety perception [15, 31, 38].

In this study, we present a methodology to predict human perceptions of the safety category. Let us consider the case of Rio de Janeiro city which is to be said as one of the most dangerous Brazilian cities. In Figure 1 we present a pair of images in which the left-hand side image preserves from the community of Favela (known as a zone that is basically dominated by drugs traffic and not well developed urban area) while the latter one has captured in the city center preserving the organized and highly cleaned area. The main contribution of this work can be seen in three fronts: (i) we propose a concrete methodology to study and analyze Urban Perception; (ii) we explore the Place Pulse 2.0 dataset by considering the calculation of the perceptual scores and further analyze the limitations arisen from the dataset; and (iii) a classification model

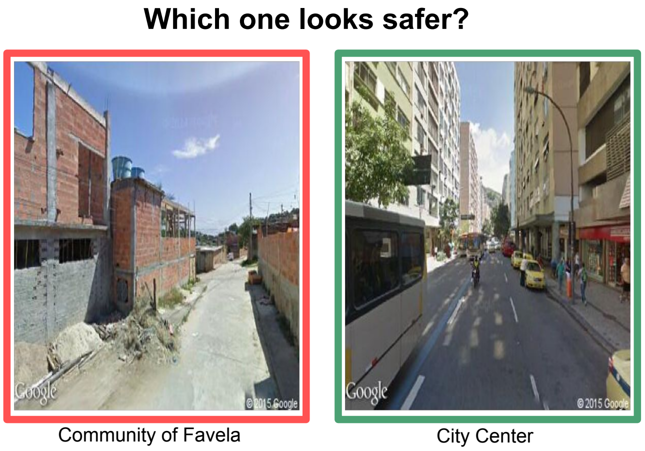

Figure 1: The image instances from the city of Rio de Janeiro, Brazil. The images were asked over the perception survey to assign the perception score whether the image looks like *safe* or *unsafe*. The images are in PlacePulse2.0 dataset [8], offering the geo coordinates for each.

WI-IAT '21, December 14–17, 2021, ESSENDON, VIC, Australia

ACM ISBN 978-1-4503-9115-3/21/12…$15.00
https://doi.org/10.1145/3486622.3493975

is developed to predict the urban safety perception on street view images.

The remainder of this paper is organized as follows: Section 2 studies the related works in the literature; Section 3 presents our prediction methodology on urban perception task; Section 4 reports our experimental evaluations and discusses the results; and finally, Section 5 concludes this work.

2 RELATED WORKS

Urban Perception is an important field in urbanism and urban planning. This area of research aims not only to create a highly accurate prediction model [32, 36], but also to understand the city environment and its impact on the population [3, 39]. The major goal is to develop a model to possibly map the city's visual appearance along with non-visual attributes such as crimes, house prices, and perception surveys. Many works attempt to identify and differentiate the city's appearance (*e.g.*, "What makes Paris look like Paris?" [6], "what makes an outdoor space beautiful?" [34], and "What Makes London Look Beautiful, Quiet, and Happy?" [29]). Other works consider additional data (crimes, robbery) to map over different city area considering visual appearance and statistical rates [2, 9, 24]. Another approach is to analyze the influence of the presence of graffiti in large city cities and then compare it with the human development index [1, 5, 37].

One major work in line with urban perception analysis is [30] that introduces the Place Pulse dataset. The dataset comprises comparisons between pairs of images in various categories (*e.g.*, safety, lively, wealthy). This work was studied to perform an urban mapping method using the urban perception score and further localize it inside the target city[25, 27]. Some feature extractors like GIST, DeCAF [7], and ImageNet[4] were obtained to train the image representations along with the respective perceptual scores. Others seek to extract more information about the visual appearance of the image using complex methods like convolutional neural network (CNN) [8, 18, 28] and analyze greenery areas in cities [13–15, 21]. Alternatively, object segmentation techniques also considered analyzing the object presence and their correlation with safety perception[23, 42]. Also, some works use machine learning interpretation to understand the explanation between models and human perception [20, 22, 41].

3 METHODOLOGY

This section first discusses the street view imagery dataset utilized in this work and explains its settings and configurations. It then explains our methodology in the perceptual score prediction task.

3.1 Dataset

Place Pulse has two versions: PlacePulse1.0 [30] and PlacePulse2.0 [8]. Both of them are composed by comparisons between two images providing the latitude, longitude, and the respective winner (or draw). (i) **Place Pulse 1.0**: released in 2013, it contains a total of 73,806 comparisons, 4,136 images from 4 cities: New York (including Manhattan and parts of Queens, Brooklyn and The Bronx), Boston, Linz, and Salzburg. Besides, they evaluated three types of comparisons: *safe*, *wealth*, and *unique*; and (ii) **Place Pulse 2.0**: released in 2016, it contains around 1.22 million comparisons, 111,390 images from 56 cities in 32 countries across the 6 continents (dividing North and South America) and six categories: *safe*, *wealth*, *depress*, *beautiful*, *boring*, and *lively*.

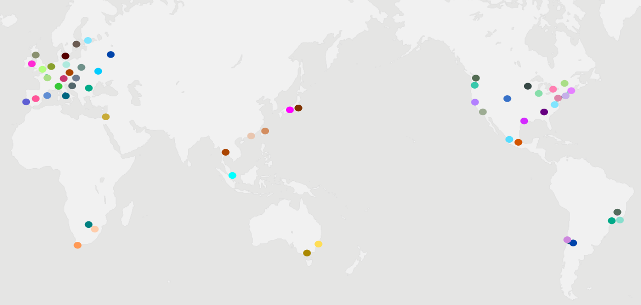

Figure 2: World wide geographical distribution overall the 56 cities involved in the Place Pulse 2.0 dataset. The majority of the cities contributed from Europe and North-America.

Calculating the Perceptual Scores

In order to pre-process all comparisons in the dataset we follow these steps: for each compared image i with other images j many times in different categories, we define as the *intensity of perception* of any image i as the percentage of times that the image was selected. This intensity of j affects i intensity. Due to this, we define the positive rate W_i (1) and the negative rate L_i (2) of an image i corresponding to each target category:

$$W_i = \frac{w_i}{w_i + d_i + l_i} \tag{1}$$

$$L_i = \frac{l_i}{w_i + d_i + l_i} \tag{2}$$

where, w_i is the number of wins, l_i number of loses, and d_i draws; From the equations 1 and 2 we can calculate the perceptual score associated for each an image i called *Q-score* with notation $q_{i,k}$ in a category k:

$$q_{i,k} = \frac{10}{3}(W_{i,k} + \frac{1}{n^w_{i,k}}(\sum_{j_1}^{n^w_{i,k}} W_{j_1,k}) - \frac{1}{n^l_{i,k}}(\sum_{j_2}^{n^l_{i,k}} L_{j_2,k}) + 1) \tag{3}$$

The Equation 3 is the perceptual score of the image i to be ranked, where j_x is an image compared to image i, n^w_i is equal to the total number of images i beat and n^l_i is equal to the total number of images to which i lost. Besides, j_1 is the set of images that loses against the image i and j_2 is the set of images that wins against the image i.

Finally, referring to previous studies in visual assessment [26, 30], the perceptual score Q is scaled to fit the range 0 to 10. This process is performed by adding the constant value 1 and multiplying the equation 3 by $\frac{10}{3}$. In these scores, the value 10 represents the highest possible score for a given question. As an example, if an image receives a calculated score of 0 for the question "Which place looks safer?", that means the image has been perceived as the least safe image in the dataset.

In Table 1, we show the statistics from Place Pulse 1.0 dataset in the three categories evaluated and the 4 cities studied. We note that the highest mean score values are presented in Boston City except in *wealthy* category. Different from Place Pulse 1.0, with the Place Pulse 2.0 dataset we can generate deep statistics summarized in Table 2. The sub Table 2(a) shows the information organized by continent (note that we divided North America and South America), number of countries, number of cities, and number of images; In Figure 2 we show the geographical distribution of the cities included in this dataset. In the sub-Table 2(b) we show the number of comparisons, the number of images compared, and the mean score for each category. We also remark that the *safety* category has the highest values respecting the other categories.

Place Pulse 1.0				
City	# images	*safety*	*wealthy*	*uniquely*
Linz	650	4.85	5.01	4.83
Boston	1237	4.93	4.97	4.76
New York	1705	4.47	4.31	4.46
Salzburg	544	4.75	4.89	5.04
Total	4136			

Table 1: Data summary on Place Pulse 1.0 and its respective mean score per category.

Place Pulse 2.0			
Continent	#countries	#cities	#images
Europe	19	22	38,747
North America	3	17	37504
South America	2	5	12,524
Asia	5	7	11,417
Oceania	1	2	6,097
Africa	2	3	5,101
Total	32	56	111,390

(a)

Place Pulse 2.0			
Category	# comparisons	# images	*mean*
Safety	368,926	111,389	5.188
Lively	267,292	111,348	5.085
Beautiful	175,361	110,766	4.920
Wealthy	152,241	107,795	4.890
Depressing	132,467	105,495	4.816
Boring	127,362	106,363	4.810
Total	1,223,649		

(b)

Table 2: Statistics obtained after process all comparisons from Place Pulse 2.0, reporting information about images per cities in each continent and the mean score for each target category.

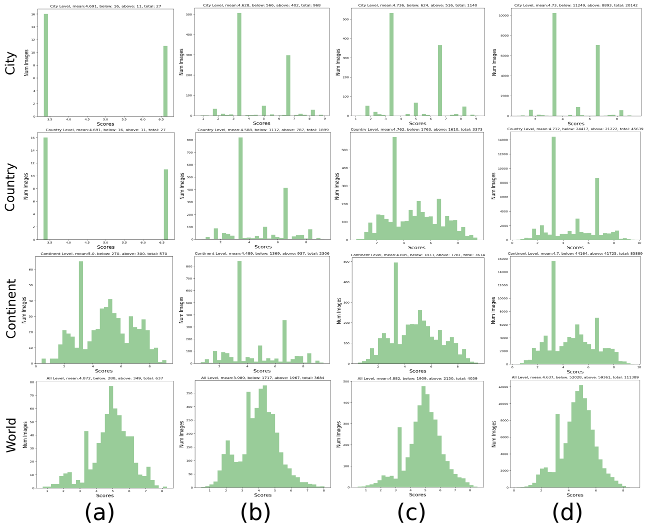

Figure 3: Perceptual score distributions in the 4 different region for 3 different cities: (a) Amsterdam-Netherlands, unique city; (b) Rio de Janeiro-Brazil with 3 cities; (c) Atlanta-USA with 17 cities. (d) World (all cities)

Generalization on Perceptual Scores

This work focuses on the Place Pulse 2.0 dataset due to the larger quantity of images. As we mentioned above, we can analyze the information dividing the world by regions like global level, continent level, country level, and city level. Following this idea, we calculate the scores using images compared with others in the same region level (*e.g.*, city-level: RJ images compared with RJ images; country level: RJ images with SP images; continent level: RJ images with Santiago (Chile) images; and global level: RJ images with Tokyo images).

Table 3 shows the impact of the region level on the perceptual score calculation. There is a decrease in the number of images evaluated. All images were compared with a random image from a random location around the world; once we attempt to filter using the Latitude-Longitude of the images compared in the same city, country, or continent, we retrieve images that were not being in compression process with any other images from the same country.

Place Pulse 2.0				
Category/Level	City	Country	Continent	Global
safety	20,143	45,640	85,890	111,390
lively	14,803	38,216	79,788	111,349
Beautiful	9,410	28,811	66,792	110,767
Wealthy	7,642	24,326	57,780	107,796
Depressing	6,556	21,171	52,504	105,496
Boring	6,148	20,931	52,031	106,364

Table 3: Quantity of images used for the calculation of the perceptual scores for each all region levels. We note that for the city level we lost a huge quantity of information (images) for each category.

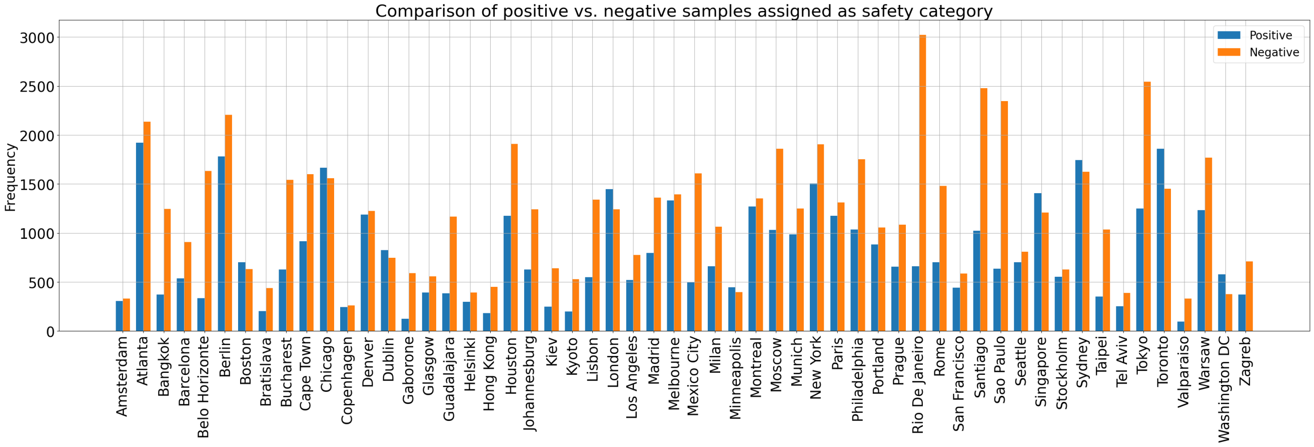

Figure 4: We keep threshold of 5.0 to assign labels as 0 (unsafety) or 1 (safety). The figure shows the imbalance in the images quantity for each class per city. We note that in most of the cities, the imbalance of classes is observed (for instance, Rio de Janeiro city holds the highest number of unsafety images).

Moreover, by calculating over the same city, we retrieve a reduction of 82% compared with the total number of images (at the global level). Since we know that the majority of countries in Europe has only one city (*e.g.*, Amsterdam - Netherlands); Brazil has 3 cities; Chile, Mexico, and Japan has 2 cities; and the USA has 15 cities. The number of cities per country and number of comparisons will impact directly the perceptual score distribution (See Figure 3).

Analysing the Imbalance data

As we describe in the previous section, we observed in Figure 3 that for a low quantity of images (city-level), the scores are accumulated in the value 3.3333 and 6.6666 (more notorious in the city level). This happened due to the number of comparisons performed in each image (maximum 3 comparisons on average). So, using the Equation 3 we know when we have the scores 3.3333 is because an image won 1 of 3 comparisons and 2 of 3 for 6.6666. So, this behavior keeps at country and continent levels as well, with the main difference that at the continent level we have a different distribution of scores. At the global level, for all cities, we finally have a good distribution of scores (last row in Figure 3). So, using the perceptual scores calculated at the global level, we will focus our analysis on the safety category to study the perception. In Figure 4 we show the imbalance for the category safety, we see that imbalance happens in the majority of the cities presented in Place Pulse 2.0 (more unsafety images); In the other hand, few cities have more safety images (*e.g.*, Washington DC, Toronto, Sydney, Singapore, Londres, Boston, and Chicago).

3.2 Predicting the Urban Safety Perception

In this work, we perform a classification task adapting the VGG16 [35] architecture to a Global Average Pool model [16] called VGG16-GAP. We modify the last layer of the block-conv5, taking the Max-Pooling layer and replacing it with a GAP layer. This modification aims to extract more informative and high-level features from input images through Global Average Pooling (see Figure 5).

Once we add the GAP layer, we remove last 2 Fully Connected layer from original model architecture after layer 13, we call the output of this layer as the features extracted from VGG-GAP. In Figure 5 we present the model used to extract our features, then we train and compare the behavior of linear and non-linear models: (i) *Logistic Regression*: $L(y, f(x)) = \sum_{i=1}^{n} log(e^{(-y_i * f(x_i))} + 1)$; (ii) *Linear SVC*: $L(y, f(x)) = \sum_{i=1}^{n} max(0, 1 - y_i f(x_i))$; and (iii) *Ridge Classifier*: $L(y, f(x)) = sgn(||y - f(x)||_2^2 + ||w||_2^2)$.

Furthermore, we can define an additional parameter called δ to control how many samples of the % top and % bottom are used to train. We prefer just to use the default $\delta = 0.5$ which means all datasets, based on results of previous works which already reported this [23, 24, 27, 30]. We evaluated our classifier model behavior using the following metrics: Accuracy (Equation 4), Area Under the

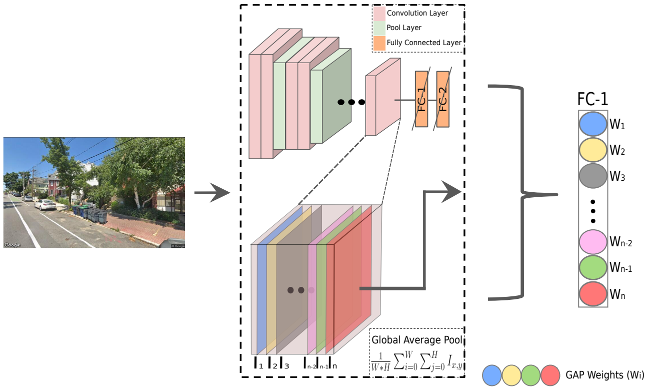

Figure 5: This Figure presents our adapted network, called VGG16-GAP. As we mention, we will use this architecture and the original VGG16 as feature extractors.

Curve (AUC), Precision (Equation 5), Recall (Equation 6), and F1 score (Equation 7). Due to the binary classification task performed, the AUC, F1, and Accuracy will tell us how well is behaving our model in the predictions. In Table 4 we report the average of the 5 cross-validations for each method proposed, dividing the dataset into 80% to train and 20 % to test. We note that VGG16_GAP-Places presents the best results of all of them.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{4}$$

$$Precision = \frac{T_P}{T_P + F_P} \tag{5}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{6}$$

$$F1_{score} = 2\frac{Precision * Recall}{Precision + Recall} \tag{7}$$

4 RESULTS AND DISCUSSIONS

This work presents a methodology to study, explore, and analyze the dataset Place Pulse 2.0, focusing on the safety category. We calculate the perceptual scores using different comparison levels as city, country, continent, and global levels. We then explore the distribution of the calculated scores noting that for the low quantity of cities evaluated, worst the score distribution. Then, selecting the global region as the better to study we use a threshold to label the scores: an image with a score above 5.0 is safety; another case, unsafety. After that, we notice the imbalance of samples in the majority of cities which is the same to say the imbalance of samples in the dataset. From this, we use the network VGG16 with pre-trained weights specialized on the ImageNet and Places365 datasets. Besides, we modify the last convolutional layer changing to a Global Average Pooling layer.

Our main pipeline is the following: (i) pre-process the perceptual scores for each image in the Place Pulse dataset as we defined at Section 3.1 this work focuses only on the *Safety* category; (ii) we then extract features from images using the four VGG16 based models (VGG16, VGG16_GAP, VGG16-Places, and VGG16_GAP-Places); and finally (iii) we perform the classification and see the results.

In Table 4, we report the metrics for each feature extractor and each method; we note that *VGAP: VGG16_GAP-Places have the best result (the average of the 5 cross-validations) using the LinearSVC method. Since the Place Pulse dataset is composed of google street view images, this makes sense due to the nature of the pre-trained weights of Places365. Compared with ImageNet, the Places365 dataset is composed of indoor and outdoor images like streets, residential zones, markets, restaurants, landscapes, etc; Instead of objects as ImageNet. In addition, we note that VGG_GAP presents the best behavior using LinearSVC instead of Logistic (the best for VGG16 extractors) this could be due to the operation in the last Global Average Pooling layer applied which enables the network to better understand the pattern of the layer after convolution activation rather than dense layers [10, 16].

		auc		accuracy		f1 score	
FE	Method	train	test	train	test	train	test
VGG	LinSVC	63.62	56.50	68.85	65.22	54.78	**49.41**
	Logistic	60.63	**57.52**	67.25	**65.72**	51.42	49.07
	Ridge	64.72	54.75	69.44	64.38	56.50	49.34
VGAP	LinSVC	59.01	**57.93**	66.51	**66.09**	49.52	49.06
	Logistic	58.07	57.57	65.95	65.59	46.06	45.61
	Ridge	59.20	57.93	66.59	65.89	50.27	**49.76**
*VGG	LinSVC	64.44	57.14	69.48	65.79	56.39	51.20
	Logistic	61.74	**58.35**	68.16	**66.44**	53.77	**51.28**
	Ridge	65.20	55.76	69.84	64.86	57.56	50.67
*VGAP	LinSVC	60.26	**59.76**	67.38	**66.96**	51.65	51.04
	Logistic	59.40	58.97	66.81	66.62	49.16	48.90
	Ridge	60.45	59.15	67.45	66.94	52.23	**51.53**

Table 4: Each table reports classification metrics (in-percentage) over four feature extractors and the proposed methods. Glossary: FE: Feature Extractor; VGG: VGG16; VGAP: VGG16_GAP; *VGG: VGG16-Places; *VGAP: VGG16_GAP-Places; LinSVC: Linear SVC.

Dataset Limitations

Some limitations arose in this dataset. The first limitation is the construction of Place Pulse, which uses an online survey. Each volunteer chooses between two images that are the most "safe" depending on their biased personal perception criteria. The second limitation is the small number of sample images per city. Compared with another dataset which has millions of samples, our total is not above 112,000 which yields the model for poor performance when a few sample data is provided. The third limitation is that we are obligated to use all datasets, if we calculate scores based on the region's levels we will lose information; in other words, we need to use all the cities to get good behavior. And finally, we state that it is a challenging task to create a specific perceptual predictor model per city; this is highly due to the lack of sufficient training samples and the number of comparisons per each one.

5 CONCLUSIONS

In this work, we propose a methodology that allows us to explore, analyze, and understand the nature of the dataset Place Pulse 2.0. To do this, we pre-process the dataset Place Pulse 2.0 analyzing the 110K images obtained by category comparisons. we focus our study on the safety category due to the importance of urban security, but we calculate the corresponding perception scores in four different levels (city, country, continent, and global) in all categories. Finally, We conclude that our methodology is capable to learn characteristics related to the prediction of the safety urban perception in street images.

To extend this work, we will focus on solving the issue of imbalance data, as well as, the lack of sample in the studied dataset. We believe that is possible to achieve better results using other types of models (such as self-supervised learning techniques) which can mitigate such sort of limitations.

ACKNOWLEDGMENTS

This work was supported by grant 234-2015-FONDECYT (Master Program) from CienciActiva of the National Council for Science, Technology and Technological Innovation (CONCYTEC-PERU), grant #312483/2018-0 from The Brazilian National Council for Scientific and Technological Development (CNPq-BRAZIL), and Getulio Vargas Foundation.

REFERENCES

[1] Javier Rozo Alzate, Marta S Tabares, and Paola Vallejo. 2021. Graffiti and government in smart cities: a Deep Learning approach applied to Medellín City, Colombia. In *International Conference on Data Science, E-learning and Information Systems 2021*. 160–165.

[2] Sean M Arietta, Alexei A Efros, Ravi Ramamoorthi, and Maneesh Agrawala. 2014. City forensics: Using visual elements to predict non-visual city attributes. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2624–2633.

[3] Marco De Nadai, Jacopo Staiano, Roberto Larcher, Nicu Sebe, Daniele Quercia, and Bruno Lepri. 2016. The death and life of great Italian cities: a mobile phone data perspective. In *Proceedings of the 25th international conference on world wide web*. 413–423.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.

[5] Alexandre Magno Alves Diniz and Mark C Stafford. 2021. Graffiti and crime in Belo Horizonte, Brazil: The broken promises of broken windows theory. *Applied Geography* 131 (2021), 102459.

[6] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei Efros. 2012. What makes paris look like paris? (2012).

[7] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2014. Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning*. 647–655.

[8] Abhimanyu Dubey, Nikhil Naik, Devi Parikh, Ramesh Raskar, and César A. Hidalgo. 2016. Deep Learning the City : Quantifying Urban Perception At A Global Scale. *CoRR* (2016).

[9] Kaiqun Fu, Zhiqian Chen, and Chang-Tien Lu. 2018. StreetNet: preference learning with convolutional neural network on urban crime perception. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 269–278.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), 770–778.

[11] Rachel Kaplan and Stephen Kaplan. 1989. *The experience of nature: A psychological perspective*. Cambridge university press.

[12] Kees Keizer, Siegwart Lindenberg, and Linda Steg. 2008. The Spreading of Disorder. *Science (New York, N.Y.)* 322 (12 2008), 1681–5. https://doi.org/10.1126/science.1161405

[13] Leonardo León-Vera and Felipe Moreno-Vera. 2018. Car Monitoring System in Apartments' Garages by Small Autonomous Car Using Deep Learning. In *Annual International Symposium on Information Management and Big Data*. Springer, Springer International Publishing, 174–181.

[14] Xiaojiang Li, Chuanrong Zhang, and Weidong Li. 2015. Does the visibility of greenery increase perceived safety in urban areas? Evidence from the place pulse 1.0 dataset. *ISPRS International Journal of Geo-Information* 4, 3 (2015), 1166–1183.

[15] Xiaojiang Li, Chuanrong Zhang, Weidong Li, Robert Ricard, Qingyan Meng, and Weixing Zhang. 2015. Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening* 14, 3 (2015), 675–685.

[16] Min Lin, Qiang Chen, and Shuicheng Yan. 2013. Network in network. *arXiv preprint arXiv:1312.4400* (2013).

[17] Pall J Lindal and Terry Hartig. 2013. Architectural variation, building height, and the restorative quality of urban residential streetscapes. *Journal of Environmental Psychology* 33 (2013), 26–36.

[18] Xiaobai Liu, Qi Chen, Lei Zhu, Yuanlu Xu, and Liang Lin. 2017. Place-centric visual urban perception with deep multi-instance regression. In *Proceedings of the 25th ACM international conference on Multimedia*. ACM, 19–27.

[19] Kevin Lynch. 1984. Reconsidering the image of the city. In *Cities of the Mind*. Springer, 151–161.

[20] Weiqing Min, Shuhuan Mei, Linhu Liu, Yi Wang, and Shuqiang Jiang. 2019. Multi-Task Deep Relative Attribute Learning for Visual Urban Perception. *IEEE Transactions on Image Processing* 29 (2019), 657–669.

[21] F. Moreno-Vera. 2019. Performing Deep Recurrent Double Q-Learning for Atari Games. In *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. 1–4. https://doi.org/10.1109/LA-CCI47412.2019.9036763

[22] Felipe Moreno-Vera. 2021. Understanding Safety Based on Urban Perception. In *International Conference on Intelligent Computing*. Springer, 54–64.

[23] Felipe Moreno-Vera, Bahram Lavi, and Jorge Poco. 2021. Urban Perception: Can We Understand Why a Street Is Safe?. In *Mexican International Conference on Artificial Intelligence*. Springer, 277–288.

[24] Nikhil Naik, Jade Philipoom, Ramesh Raskar, and Cesar Hidalgo. 2014. StreetScore: Predicting the Perceived safety of one million streetscapes. *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2014).

[25] Nikhil Naik, Ramesh Raskar, and César A Hidalgo. 2016. Cities are physical too: Using computer vision to measure the quality and impact of urban appearance. *American Economic Review* 106, 5 (2016), 128–32.

[26] Jack L Nasar. 1998. The evaluative image of the city. (1998).

[27] Vicente Ordonez and Tamara L. Berg. 2014. Learning High-level Judgments of Urban Perception. *European Conference on Computer Vision (ECCV)* (2014).

[28] Lorenzo Porzi, Samuel Rota Bulò, Bruno Lepri, and Elisa Ricci. 2015. Predicting and Understanding Urban Perception with Convolutional Neural Networks.

[29] Daniele Quercia, Neil Keith O'Hare, and Henriette Cramer. 2014. Aesthetic capital: what makes London look beautiful, quiet, and happy?. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*.

[30] Mark Philip Salesses. 2012. *Place Pulse: Measuring the collaborative image of the city*. Ph.D. Dissertation. Massachusetts Institute of Technology.

[31] Robert J Sampson, Jeffrey D Morenoff, and Thomas Gannon-Rowley. 2002. Assessing "neighborhood effects": Social processes and new directions in research. *Annual review of sociology* 28, 1 (2002), 443–478.

[32] Darshan Santani, Salvador Ruiz-Correa, and Daniel Gatica-Perez. 2018. Looking south: Learning urban perception in developing cities. *ACM Transactions on Social Computing* 1, 3 (2018), 1–23.

[33] Herbert W Schroeder and Linda M Anderson. 1984. Perception of personal safety in urban recreation sites. *Journal of leisure research* 16, 2 (1984), 178–194.

[34] Chanuki Illushka Seresinhe, Tobias Preis, and Helen Susannah Moat. 2017. Using deep learning to quantify the beauty of outdoor places. *Royal Society open science* 4, 7 (2017), 170170.

[35] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)* (2014). arXiv:1409.1556 [cs.CV]

[36] Xuan Song, Quanshi Zhang, Yoshihide Sekimoto, Teerayut Horanont, Satoshi Ueyama, and Ryosuke Shibasaki. 2013. Modeling and probabilistic reasoning of population evacuation during large-scale disaster. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1231–1239.

[37] Eric K. Tokuda, Cláudio T. Silva, and Roberto Marcondes Cesar Jr. 2019. Quantifying the presence of graffiti in urban environments. *CoRR* abs/1904.04336 (2019). arXiv:1904.04336 http://arxiv.org/abs/1904.04336

[38] Roger S Ulrich. 1979. Visual landscapes and psychological well-being. *Landscape research* 4, 1 (1979), 17–23.

[39] Matthias Wendt. 2009. The importance of death and life of great American cities (1961) by Jane Jacobs to the profession of urban planning. *New Visions for Public Affairs* 1 (2009), 1–24.

[40] James Q Wilson and George L Kelling. 1982. Broken windows. *Atlantic monthly* 249, 3 (1982), 29–38.

[41] Yongchao Xu, Qizheng Yang, Chaoran Cui, Cheng Shi, Guangle Song, Xiaohui Han, and Yilong Yin. 2019. Visual Urban Perception with Deep Semantic-Aware Network. In *International Conference on Multimedia Modeling*. Springer, 28–40.

[42] Fan Zhang, Bolei Zhou, Liu Liu, Yu Liu, Helene H Fung, Hui Lin, and Carlo Ratti. 2018. Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning* 180 (2018), 148–160.