PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

DEEP NEURAL NETWORK MODELS WITH EXPLAINABLE COMPONENTS FOR URBAN SPACE PERCEPTION.

ANDRÉS CÁDIZ VIDAL

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Advisor:
HANS LÖBEL

Santiago de Chile, July 2020

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

DEEP NEURAL NETWORK MODELS WITH EXPLAINABLE COMPONENTS FOR URBAN SPACE PERCEPTION.

ANDRÉS CÁDIZ VIDAL

Members of the Committee:
HANS LÖBEL
PATRICIO DE LA CUADRA
MEMBER B
MEMBER C

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Santiago de Chile, July 2020

Gratefully to my parents and siblings

ACKNOWLEDGEMENTS

Write in a sober style your acknowledgements to those persons that contributed to the development and preparation of your thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The abstract must contain between 100 and 300 words. The abstract must be written in English and Spanish. In the case of doctoral theses, the layout of the abstract page is different, so please check the template provided by the OGRS.

Keywords: thesis template, document writing, **(Write here the keywords relevant and strictly related to the topic of the thesis)**.

RESUMEN

El resumen debe contener entre 100 y 300 palabras. El resumen debe ser escrito en inglés y español. En el caso de tesis de doctorado, el formato de la página del resumen es distinta, por favor verifique la plantilla entregada por la Dirección de Postrgrado.

Palabras Claves: plantilla de tesis, escritura de documentos, (**Colocar aquí las palabras claves relevantes y estríctamente relacionadas al tema de la tesis**).

1. INTRODUCTION

Urban perception is a feeling held by people about a location. These feelings can be and are often related to a particular characteristic, like happiness or beauty, or also inherently negative ones, like insecurity or fear (Ordonez & Berg, 2014). Understanding the cause of these feelings is a complex task, since unique social and psychological aspects of each individual affect how they perceive and the spaces they observe (Nasar, 1990).

Visual urban perception is responsible for a large parte of the experience that people go through while being at or using an urban space, this not only affects how much the spaces themselves are used (Khisty, 1994) but also the use of related means of transport (Antonakos, 1995). Other studies have also found correlations between urban perception, crime statistics (Ordonez & Berg, 2014) and wealth, and therefore used it as a proxy measure of inequality (Ordonez & Berg, 2014; Salesses, Schechtner, & Hidalgo, 2013; Rossetti, Lobel, Rocco, & Hurtubia, 2019).

On the other hand, being able to understand a community's need and perception of a city at scale is something of key importance on developing cities, so that the limited resources of local governments can be applied more efficiently (Santani, Ruiz-Correa, & Gatica-Perez, 2018).

Traditional methods for obtaining this type of data, consist of hand made polls about specific locations making systematic evaluation of perception an extremely costly and hard to escalate task (Nasar, 1990; Clifton & Ewing, 2008). Other approach consist of surveys based on computer generated images of simulated spaces, this is more scalable, but is limited to experimental design and it doesn't apply to a real space (Laing et al., 2009; Iglesias, Greene, & de Dios Ortúzar, 2013).

Currently, thanks to the great volumes of data generated by web platforms (Salesses et al., 2013) and to modern deep learning (DL) and computer vision techniques (LeCun, Bengio, & Hinton, 2015), new solutions for estimating urban perception have become feasible, and some previous studies have achieved significant results, either by applying traditional deep learning (Dubey, Naik, Parikh, Raskar, & Hidalgo, 2016) or by combining it with other approaches (Rossetti et al., 2019; Zhang et al., 2018). The solutions consist mainly of training convolutional neural network models (LeCun et al., 1989) with datasets of urban images that have some sort of label that is used as an estimator for the perception of that urban space. Most of the research is based on the place pulse dataset (Dubey et al., 2016), which consists of pairs of images along with labels that indicate which of the images is more representative of a particular attribute.

However, current deep learning methodologies, have the disadvantage of being ”black boxes”, in other words, they lack a direct or systematic way to explain or interpret the obtained results. This problem comes from the end to end nature of the neural network models and from the millions of learnable parameters they contain. Many of the problems in which these models are used would greatly benefit of more human understandable explanations of the results, making this a very important area of research for the deep learning field (Adadi & Berrada, 2018).

For the particular case of urban perception, explainability of the results is highly relevant, since the added information is valuable for the design of public policy, for example, it could be use to better discriminate which locations would be better recipients of an intervention, and which elements to modify so it convenes an effective improvement of perception.

Current research in explainability is primarily moving in two directions: one is to design novel neural network architectures and training methods so the models are more interpretable, such as the work by Dong, Su, Zhu, and Zhang (2017), the other direction is to create post-hoc algorithms (Adadi & Berrada, 2018) that analyze the results given by the neural network, these algorithms sometimes use other machine learning models, including neural networks, such as the work by Ghorbani, Wexler, Zou, and Kim (2019).

The work by Rossetti et al. (2019), presents an approach to this problem for the urban perception case by using semantic segmentation of the images (Badrinarayanan, Kendall, & Cipolla, 2015) as input for a discrete choice model that estimates the perception. The approach allows for a post-hoc aggregated analysis of the results, since the weights of the model are measure of the importance of each class of the semantic segmentation in the calculation of the perception.

The objective of this work is to design and train a model for the urban perception problem, that can give explainable insights on an instance level. For that it proposes a novel solution, consisting of a neural network architecture, that is end-to-end trainable and by using semantic segmentation (Zhao, Shi, Qi, Wang, & Jia, 2016) and self attention mechanisms (Vaswani et al., 2017) can show explainable insights for each of the input images.

ESTO HAY QUE ARREGLARLO AL FINAL

The remainder of this manuscript is organized as follows, Chapter 3 summarizes relevant previous research. In chapter 4 the problem is formally defined and the proposed model is described. Chapter 5 gives details on model implementation and training. Finally, in chapter 6 presents the research results and 7 the final conclusion.

2. RELATED WORK

This are some related investigations.

3. PROPOSED ARCHITECTURE

3.1. Problem Definition

3.2. Network architecture

3.3. Loss function

4. METHODOLOGY

4.1. Implementation

4.2. Training

5. RESULTS

6. CONCLUSIONS

Nothing to say. Be happy.

REFERENCES

Adadi, A., & Berrada, M. (2018, 09). Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, *PP*, 1-1. doi: 10.1109/ACCESS.2018.2870052

Antonakos, C. L. (1995). Environmental and travel preferences of cyclists. *Transportation Research Part A: Policy and Practice*, *29*(1), 85.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, *abs/1511.00561*. Retrieved from `http://arxiv.org/abs/1511.00561`

Clifton, K., & Ewing, R. (2008, 03). Quantitative analysis of urban form: A multidisciplinary review. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, *1*, 17-45. doi: 10.1080/17549170801903496

Dong, Y., Su, H., Zhu, J., & Zhang, B. (2017). Improving interpretability of deep neural networks with semantic information. *CoRR*, *abs/1703.04096*. Retrieved from `http://arxiv.org/abs/1703.04096`

Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. *ArXiv*, *abs/1608.01769*.

Ghorbani, A., Wexler, J., Zou, J., & Kim, B. (2019). *Towards automatic concept-based explanations.*

Iglesias, P., Greene, M., & de Dios Ortúzar, J. (2013). On the perception of safety in low income neighbourhoods: using digital images in a stated choice experiment.

Khisty, C. J. (1994). Evaluation of pedestrian facilities: beyond the level-of-service concept. *Transportation Research Record*, *1438*, 45-50.

Laing, R., Davies, A.-M., Miller, D., Conniff, A., Scott, S., & Morrice, J. (2009). The application of visual environmental economics in the study of public preference and urban greenspace. *Environment and Planning B: Planning and Design*, *36*(2), 355-375. doi: 10.1068/b33140

LeCun, Y., Bengio, Y., & Hinton, G. (2015, 05). Deep learning. *Nature*, *521*, 436-44. doi: 10.1038/nature14539

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, *1*(4), 541–551.

Nasar, J. L. (1990). The evaluative image of the city. *Journal of the American Planning Association*, *56*(1), 41-53. doi: 10.1080/01944369008975742

Ordonez, V., & Berg, T. L. (2014). Learning high-level judgments of urban perception. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer vision – eccv 2014* (pp. 494–510). Cham: Springer International Publishing.

Rossetti, T., Lobel, H., Rocco, V., & Hurtubia, R. (2019). Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach. *Landscape and Urban Planning*, *181*, 169-178.

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013, 07). The collaborative image of the city: Mapping the inequality of urban perception. *PLOS ONE*, *8*(7), 1-12. Retrieved from `https://doi.org/10.1371/journal.pone.0068400` doi: 10.1371/journal.pone.0068400

Santani, D., Ruiz-Correa, S., & Gatica-Perez, D. (2018). Looking south: Learning urban perception in developing cities. *ACM Transactions on Social Computing*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *CoRR*, *abs/1706.03762*. Retrieved from `http://arxiv.org/abs/1706.03762`

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning*, *180*, 148–160.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2016). Pyramid scene parsing network. *CoRR*, *abs/1612.01105*. Retrieved from `http://arxiv.org/abs/1612.01105`

APPENDIX

A. FIRST APPENDIX