Unpacking aspects of *what we see* from retail premises to characteristics of the human environment

Sam Comber[1], Dani Arribas-Bel[2]
[1]Departartment of Geography and Planning

January 28, 2019

KEYWORDS: *computer vision, retail geography, autoencoders*

Abstract

Often representations of urban environments are inferred using visual audits conducted by researchers who collate data describing physical attributes of the location visible from street-level. This might include data describing the ambiance of a location, the variety of architectural styles, or the presence of undesirable items such as graffiti and litter. Yet, *offline* studies are cost-intensive, and limited in the throughput required to reconstruct how urban environments are perceived. Excitingly, however, increasing quantities of human behaviour have become digitsed, and this "accidental" side-effect has opened up opportunities for approaching research questions that were previously impractical to approach given constraints to time and cost (Arribas-Bel, 2014). Street-level imagery derived from web-based tools such as Google Street View expose a window into the visual characteristics of neighbourhoods, allowing researchers to simulate a virtual stroll down the street. This low-cost access allows researchers to access locations remotely, mitigating the cost of in-person visual audits.

This new access to ground-level imagery has coincided with a growth in computational analytics. Various computer vision techniques have become increasingly deployed for extracting the visual features of street-level imagery. Deep learning methods such Convolutional Neural Networks (CNNs) are typically the workhorse of computational social science tasks relating to visual classification (Tan et al. 2017; Seresinhe et al. 2017), and proceed by training a statistical model using human-labelled data to learn discriminative features of the image. Unfortunately, a drawback of CNNs are the large sample sizes required to train the network, which are often unfulfilled in reality. While CNNs require large number of training instances, Convolutional Autoencoders (CAEs) are neural networks that are able to extract visual features from image data without supervision.

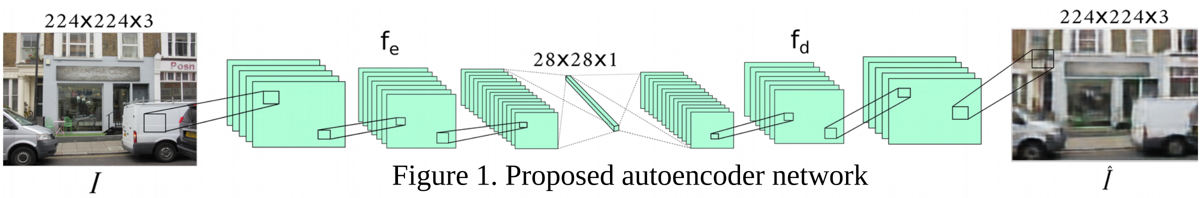

Figure 1. Proposed autoencoder network

In this article, we apply CAEs to extract the visual features of scenes for retail properties across England and Wales. In doing so, we create a compressed set of variables that jointly observe the scene properties from street-level images. By clustering these visual features obtained for each retail premise, we can begin to unpack relationships between visual cues encoded in the images and socio-economic characteristics from UK census data. In doing so, this article addresses the research

1 s.comber@liverpool.ac.uk
2 d.arribas-bel@liverpool.ac.uk

question of how we might relate the visual characteristics of retail properties to tangible features of the neighbourhood. Crucially, our study relies upon a large database of JPEG images produced by the Local Data Company (LDC) that displays the storefronts of geocoded retail premises.

To conclude this short paper submission, we note that a particular interest on street-level imagery for retail properties is because urban hierarchies are increasingly sensitive to the moderating influence of attractive retail spaces (Dennis et al. 2002). Previous studies show home-buyer aspirations are increasingly driven by a desire to reside in an environment that satisfies their wants and desires (Florida, 2008) through concentrations of leisure plazas and attractive shopping opportunities (Glaeser et al. 2001). While the literature is rich in describing the impact of proximity to retail on factors such as housing prices, the *visual* properties that influence how particular amenities are perceived is underdeveloped. Therefore, by recording the visual characteristics of commercial properties, we are able to explore which aspects of *what we see* are related to particular socio-economic variables.

References

Arribas-Bel, D. (2014). Accidental, open and everywhere: emerging data sources for the understanding of cities. Applied Geography, 49 , 45 – 53. URL: http://www.sciencedirect.com/science/article/pii/S0143622813002178.doi:https://doi.org/10.1016/j.apgeog.2013.09.012

Dennis, C., Marsland, D., & Cockett, T. (2002). Central place practice: shopping centre attractiveness measures, hinterland boundaries and the uk retail hierarchy. Journal of Retailing and Consumer Services,9 , 185 – 199. URL: http://www.sciencedirect.com/science/article/pii/S0969698901000212. doi:https://doi.org/10.1016/S0969-6989(01)00021-2.

Florida, R. (2008). Who's Your City?: How the Creative Economy Is Making Where You Live the Most Important Decision of Your Life. Random House of Canada, Limited. URL: https://books.google.co.uk/books?id=Ll4lhDMi4c4C.

Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. Journal of Economic Geography, 1 , 27–50. URL: http://dx.doi.org/10.1093/jeg/1.1.27. doi:10.1093/jeg/1.1.27.

Seresinhe, C. I., Preis, T., & Moat, H. S. (2017). Using deep learning to quantify the beauty of outdoor places. Royal Society Open Science, 4 . URL: http://rsos.royalsocietypublishing.org/content/4/7/170170. doi:10.1098/rsos.170170.

Tan, Y., Tang, P., Zhou, Y., Luo, W., Kang, Y., & Li, G. (2017). Photograph aesthetical evaluation and classification with deep convolutional neural networks. Neurocomputing, 228 , 165 – 175. URL: http://www.sciencedirect.com/science/article/pii/S0925231216312565. doi:https://doi.org/10.1016/j.neucom.2016.08.098.