Urbanize

Arib Syed   Jeffrey Mu   Gordan Milovac   Marcus Winter

Introduction

Perceptions of urban environments, such as how safe, wealthy, or lively a place appearsm, play an important role in city planning, social research, and understanding human visual bias. Prior work such as Deep Learning the City (Dubey et al., 2016) has taken major steps toward quantifying these perceptions by crowdsourcing large-scale pairwise ratings of Google Street View images. Their dataset captures diverse neighborhoods around the world and includes human judgments on attributes like safety, wealth, and beauty. This provides a unique opportunity to study how subtle visual cues shape our interpretation of the built environment.

Our project, Urbanize, is a significant shift from our original plan (AIsthetic), which focused on aesthetic image generation using the AVA dataset. After encountering major issues with the AVA pipeline and concerns about overlap with prior work like GANalyze, we pivoted to a more impactful and novel direction: generating urban scenes that differ in perceived wealth. To do this, we adopted the Place Pulse 2.0 (PP2) dataset and designed two GANs, one trained only on the raw images and a second conditioned on the "wealthy" ratings. The goal is for the conditioned GAN to generate images that appear more "wealthy" than the baseline model. This new direction aligns with research on human perception in urban spaces and presents a study on how machine learning models internalize visual socioeconomic cues.

Challenges

It's been difficult to settle on the architecture of the GAN, as well as to develop a baseline GAN that might serve as a good indicator for how well a model can perform on labeled data. We began initially with the AVA dataset, but then shifted to the AADB aesthetics dataset due to dataset usage and size issues. However, after discussions with our TA it seemed more novel to implement a cognitive direction that had not been previously explored, and one of them was perception of different spatial environments. After this discussion, finding the Place Pulse 2.0 dataset and understanding the literature of what has and has not been explored took significant time.

Insights

We successfully completed preprocessing of the data, in which all images were cropped from 400×300 to 300×300 and transformed for our base GAN. We trained this baseline GAN for 100 epochs and obtained image samples created by the generator, without access to human ratings of cognitive qualities. Though training the GANs themselves were a pain, we ultimately ended up obtaining relatively good quality images of urban scenes that resembled images in the training set. We also observed mode collapse in several epochs. Since pivoting from the previous dataset that had a smaller training set size, we were optimistic about being able to condition the GAN by adding heads that allow for the human rating labels, based on a file that had pairwise comparisons of different images based on which one resembles a more wealthy environment. Thus our natural next step is to develop an architecture that accounts for 'wealth', and potentially other labeled features like 'boringness' or 'safety', or a mixture of multiple features.

Plan

Our next step is to finish making and training the wealth-conditioned GAN, then qualitatively compare the outputs of it with the base GAN. We will be focusing most of our attention on implementing a few more GAN architectures that can condition on traits of the environments. We may also want to explore

potentially the combination of two cognitive steps, i.e. combining different traits or exploring minimal steps in the direction of that trait such that they have the maximal effect on the image, like changing the front door of a home could significantly increase the perceived wealth of the neighborhood.

We will also need to spend time to write up our findings in the poster and the final report, which will be a non-insignificant amount of time after training each of the models. Our shift in dataset led to a delay in our project timeline, although we still believe we will be able to complete project and create a poster by Wednesday.

Relevant Links

Github Repository
Deep Learning the City : Quantifying Urban Perception At A Global Scale