Deep Learning the City
Quantifying Urban Perception at a Global Scale
Abhimanyu Dubey, Nikhil Naik, Devi Parikh, Ramesh Raskar, Cesar A Hidalgo
naik@mit.edu

Key Contributions
- **Place Pulse 2.0 Dataset**: Contains 1.17 million pairwise comparisons for 110,988 images from 56 cities, provided by 81,630 online volunteers
- **Six perceptual attributes**: Safe, Lively, Boring, Wealthy, Depressing, and Beautiful
- **Deep network that minimizes a joint classification + ranking loss** to accurately predict perception of urban attributes

The Place Pulse 2.0 Dataset
- Goal: **Quantify the perception of urban environments**
- Helps study the relationship between a city’s physical appearance and the behavior and health of its residents
- A global dataset of human judgments in the form of pairwise comparisons of urban appearance
- Siamese-like networks, Streetscore-CNN (SS-CNN) and Ranking SS-CNN, to predict pairwise comparisons

Predicting Human Judgments

Performance Analysis
- **SS-CNN**: We calculate the % of pairwise comparisons in test set predicted correctly by
  1. Softmax of output neurons in final layer
  2. comparing TrueSkill scores from synthetic pairwise comparisons from the CNN
  3. extracting features from penultimate layer of CNN and feeding pairwise feature representations to a RankSVM
- **RSS-CNN**: We compare the ranking function outputs for both images in a test pair to decide which image wins, and calculate the binary prediction accuracy.

Table 1: Pairwise comparison prediction accuracy for standard networks fine-tuned with the Place Pulse 2.0 dataset. RSS-CNN (VGGNet) obtains the best performance

![Figure 1: User Interface for Crowdsourced Online Game](image)

![Figure 2: SS-CNN and RSS-CNN (with additional layers in light blue)](image)

Figure 3: Example results from Place Pulse 2.0 dataset.

References