QUANTIFYING THE IMPACT OF **BUILT ENVIRONMENT** ON NEIGHBORHOOD **CRIME RATES**

Adyasha Maharana[1], Quynh C. Nguyen[2], Elaine O. Nsoesie[3]

1 Sciome, Durham, NC, USA
2 University of Maryland, College Park, MD, USA
3 Boston University, Boston, MA, USA

Figure: Areas in Chicago with (A) High crime and (B) Low crime

The association between the estimated and reported crime rates was higher for St. Louis (r=0.84) and Chicago (r=0.88), compared to Los Angeles (r=0.75).

Census tracts with highest crime rates were underestimated in some cases, although typically predicted as the highest in each city.

Our results are in alignment with empirical research suggesting there is a relationship between physical disorder and fear of crime and crime rates. With more interpretable predictions, such models can be used for Crime Prevention Through Environmental Design (CPTED).

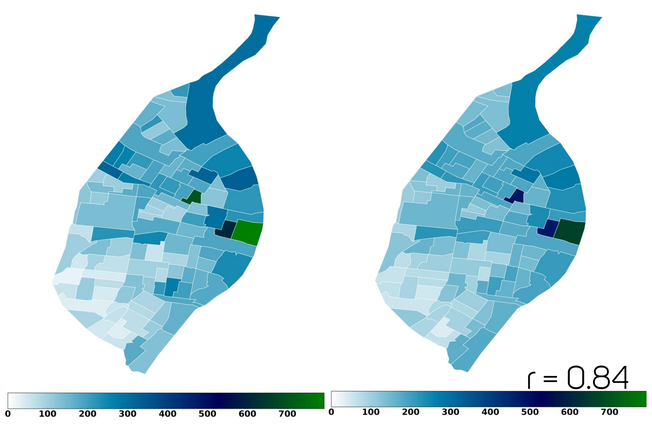


Figure: Predictions of crime rates per 1,000 persons across census tracts in St. Louis, Missouri. (A) the reported crime rates, (B) combined predictions from separate model without outliers and model for outliers

THE BUILT ENVIRONMENT HAS BEEN POSTULATED TO HAVE AN IMPACT ON NEIGHBORHOOD CRIME RATES.

However, measures of the built environment are subjective and differ across studies leading to varying observations.

We apply convolutional neural networks to 150,000 satellite images for **Los Angeles**, **Chicago** and **St. Louis** to extract features of the built environment.

We use regression models to assess the association between the built environment, sociodemographic factors and census tract-level crime rates.

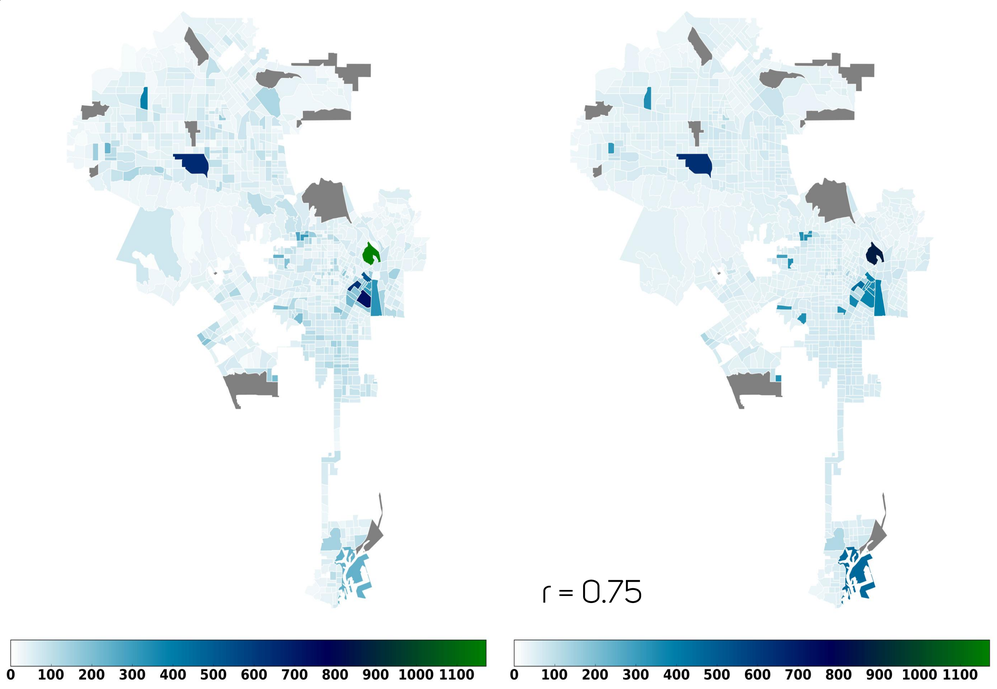


Figure: Predictions of crime rates per 1,000 persons across census tracts in Los Angeles, California. (A) the reported crime rates, (B) combined predictions from the separate model without outliers and model for outliers (i.e., census tracts with highest crime rates).

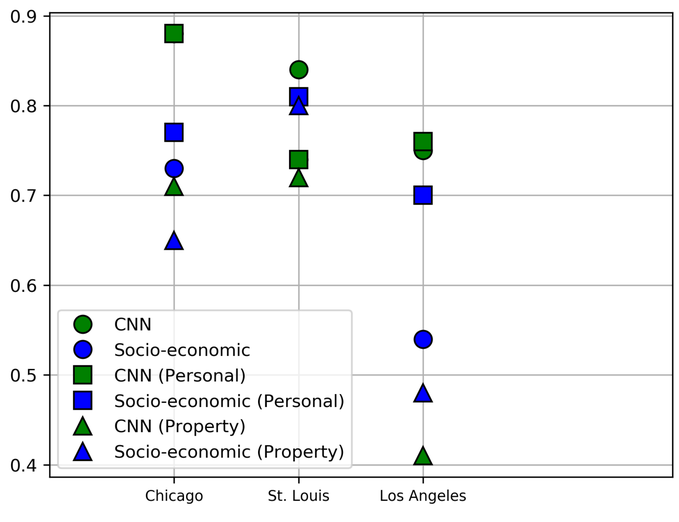


Table: Comparison of Pearson's correlation coefficient (p<0.01) for combined predictions from separate model without outliers and model for outliers, for overall crime rates, personal crime rates and property crime rates.

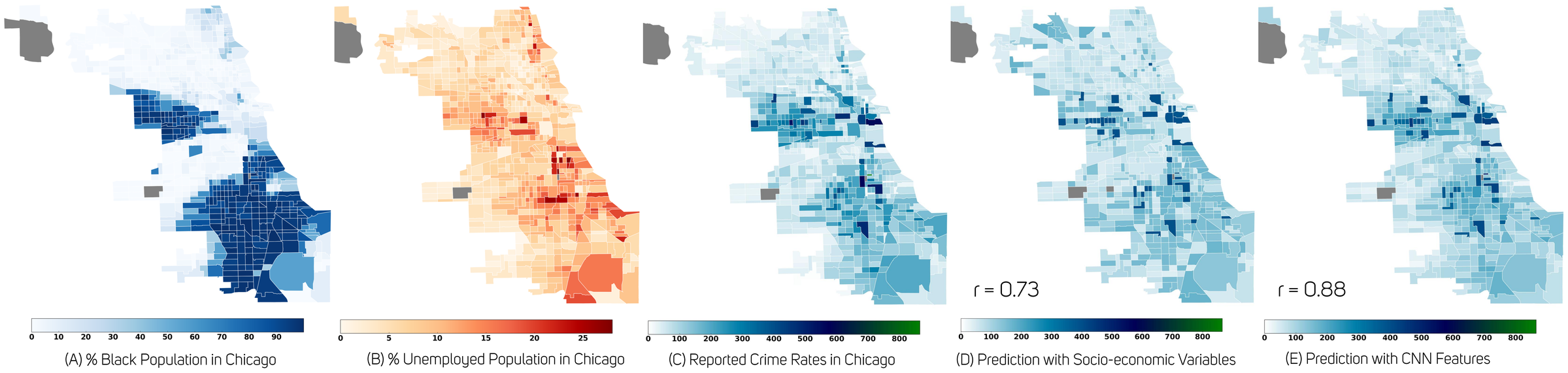
DATA COLLECTION

- Satellite imagery from Google Static Maps API at zoom level 18, based on a square grid of geo-locations
- American Census Survey (ACS) 2014 5-year estimates of socioeconomic variables for census tracts:
  - Gross Median Rent
  - Poverty status in last 12 months
  - Population by Age, Gender, Race
  - Employment status in last 12 months
  - Total Land Area
- Crime Reports from 2016, publicly available at data.gov, slmpd.org and data.lacity.org, aggregated to annual count for each census tract

METHODS

- **Fine-tuning of Convolutional Neural Network (CNN):** Pre-trained VGG-F was fine-tuned on a binary dataset of high crime and low crime area images, ending at an accuracy of ~80% on held-out test set
- **Extraction and Aggregation of Image Features:** Feature vectors were extracted from the pre-final fully connected layer of fine-tuned CNN for all satellite images and averaged over the images spanning each census tract, to produce a 4096-dimensional feature map.
- **Regression Modelling:** Elastic Net Regression was used to assess associations between:
  - CNN features and annual crime rates
  - Socio-economic variables and annual crime rates

  Separate model is built for high crime census tracts (outliers) to improve prediction.



(A) % Black Population in Chicago
(B) % Unemployed Population in Chicago
(C) Reported Crime Rates in Chicago
(D) Prediction with Socio-economic Variables
(E) Prediction with CNN Features