Measuring spatial-temporal change of physical conditions in neighborhoods with street view imagery

Daniel Chen, Tingyan Deng, Evelyn Fitzgerald, Lijing Wang

2021 Data Science for Social Good summer program

Stanford University

Our team

Daniel Chen

DSSG fellow

MS student

Stanford University

Tingyan Deng

DSSG fellow

Undergraduate student

Vanderbilt University

Evelyn Fitzgerald

DSSG fellow

MEng recent graduate

Cornell University

Lijing Wang

Technical mentor

PhD student

Stanford University

Jackie Hwang

Faculty mentor

Professor in Sociology

Stanford University

Project Motivation

Presence of

- physical disorder
- poorly maintained properties
- vacant lots in neighborhoods

Neighborhood environmental characteristics

Negatively affect

Well-being:

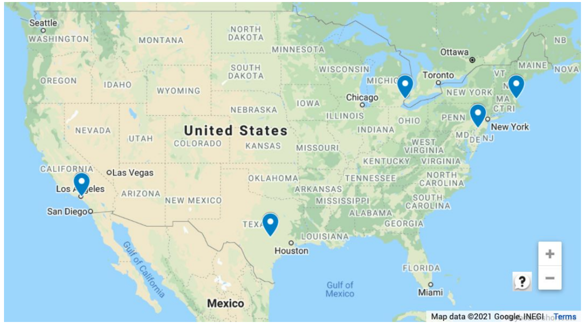
- physical and mental health
- crime and disorder
- neighborhood disinvestment

Evolution of systematic social observation (SSO)

Systematic social observation (SSO) at scale

Project Goal

Utilizing deep learning to identify building upkeep from **Google Street View** images of urban streetscapes:



- at a large scale
- over time (2007–2017)
- across multiple cities (Detroit, Boston, LA, etc.)

and quantitatively analyze relationship between building upkeep conditions and well-being characteristics

Google Street View images over time

2007

2014

2017

Training data collection

- [Input] Street View Imagery
- [Output] MTurk survey to generate *TrueSkill scores*
- Training set: 2964 Boston images, 3995 Detroit images

Left

Right

Image class labels

1. Compare pairs of images for better upkeep
2. Use comparisons to make TrueSkill score
3. Qualitatively create cutoffs on TrueSkill score to create 4 classes
4. Predict 4 classes

[Output] both class label and TrueSkill score:

- **Lower** trueskill scores, **higher** classes, **better** building upkeep

Building upkeep condition and TrueSkill scores

0: 33.54

0: 28.56

1: 26.58

2: 19.18

3: 17.79

Detroit

Lower trueskill scores, higher classes, better building upkeep

0: 33.44

1: 31.36

2: 24.58

2: 22.19

3: 16.96

Boston

Building upkeep condition and TrueSkill scores

0: 33.54

0: 28.56

1: 26.58

2: 19.18

3: 17.79

Detroit

0: 33.44

1: 31.36

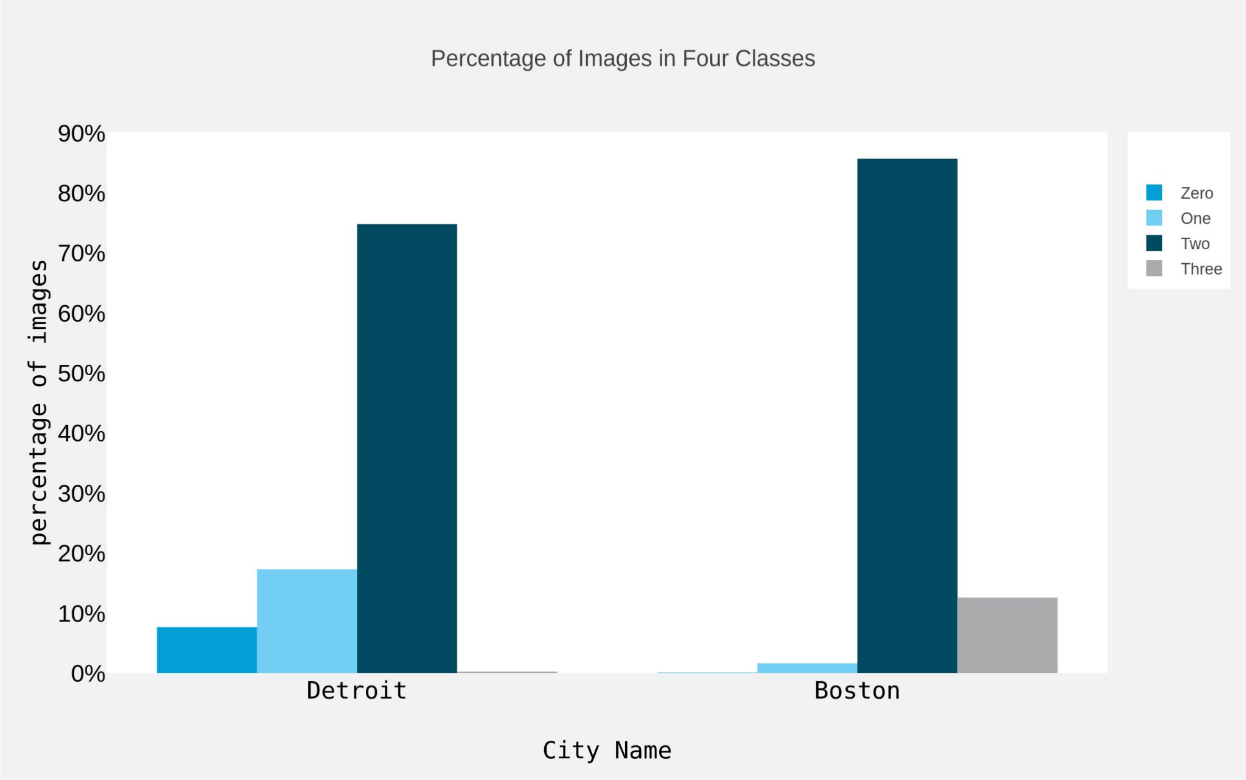
2: 24.58

2: 22.19

3: 16.96

Boston

Class imbalance



Final training data

- 2964 Boston images, 3995 Detroit images

- Each image has:
    - **Discrete class** (0, 1, 2, or 3)
    - **TrueSkill score**

Challenges in building the model

1. Regression vs Classification problem

2. How to align TrueSkill scores across Boston and Detroit?

Challenge: classification vs regression

Originally, predict discrete upkeep classes

- are these classes interpretable?

class 0

class 1

class 2

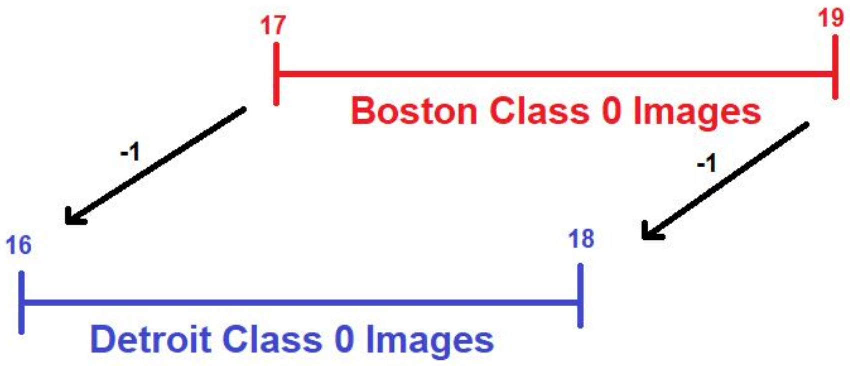
Challenge: scores across multiple cities

- TrueSkill scores derived from image comparisons
  - Comparisons are done within each city
  - TrueSkill score on different scale for each city
    - 25 in Detroit ≠ 25 in Boston

- In order to train model with both cities:
  - Need to have all TrueSkill scores at same scale

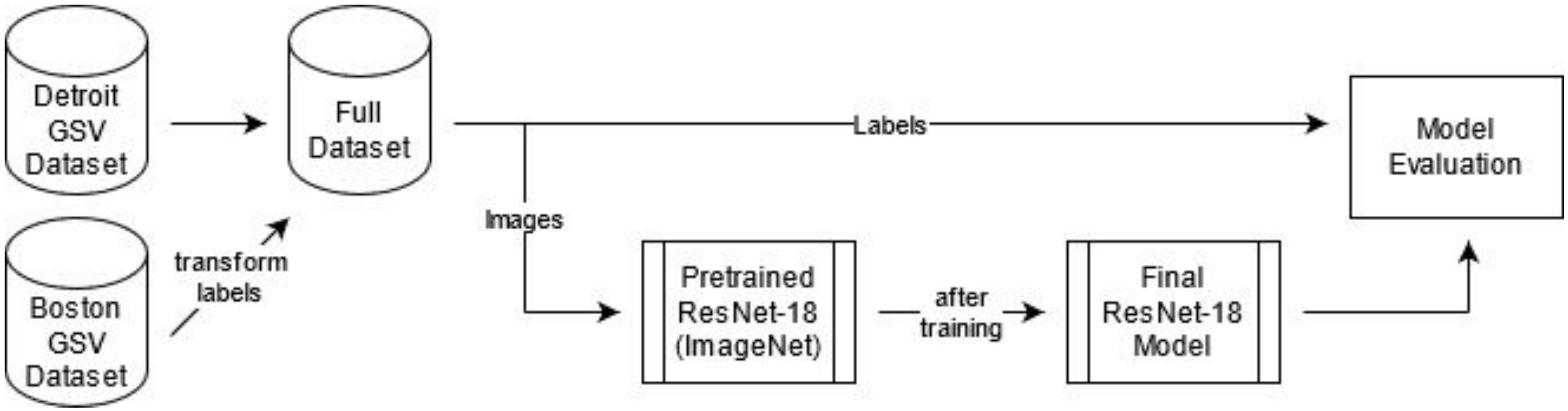
Score alignment solution

- Important note: class labels qualitatively same across cities
- Solution: transform TrueSkill scores by aligning boundary of classes

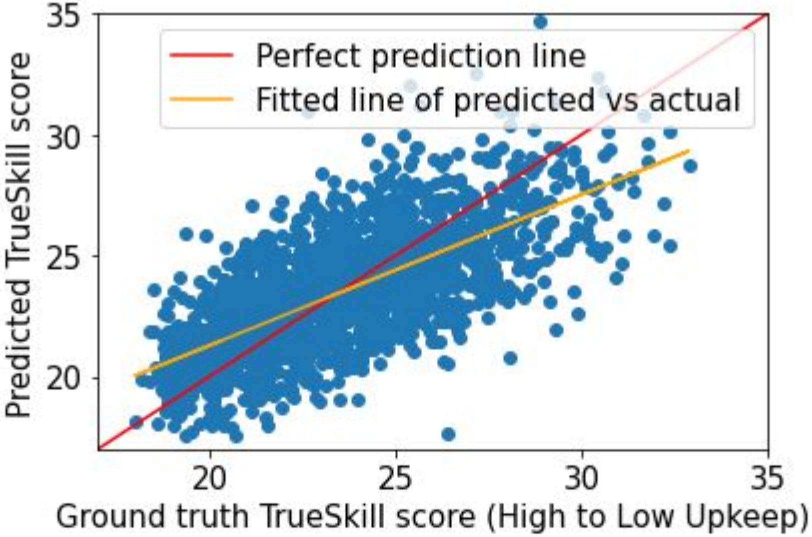


- Using this method, piecewise linear transformation on TrueSkill

Model architecture

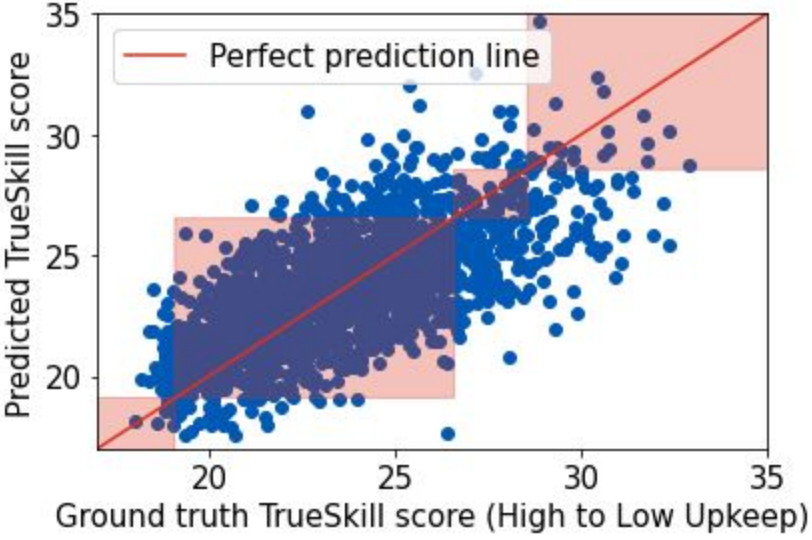


Result: regression and classification



$R^2 = 0.484$, MSE = 4.487

Result: regression and classification

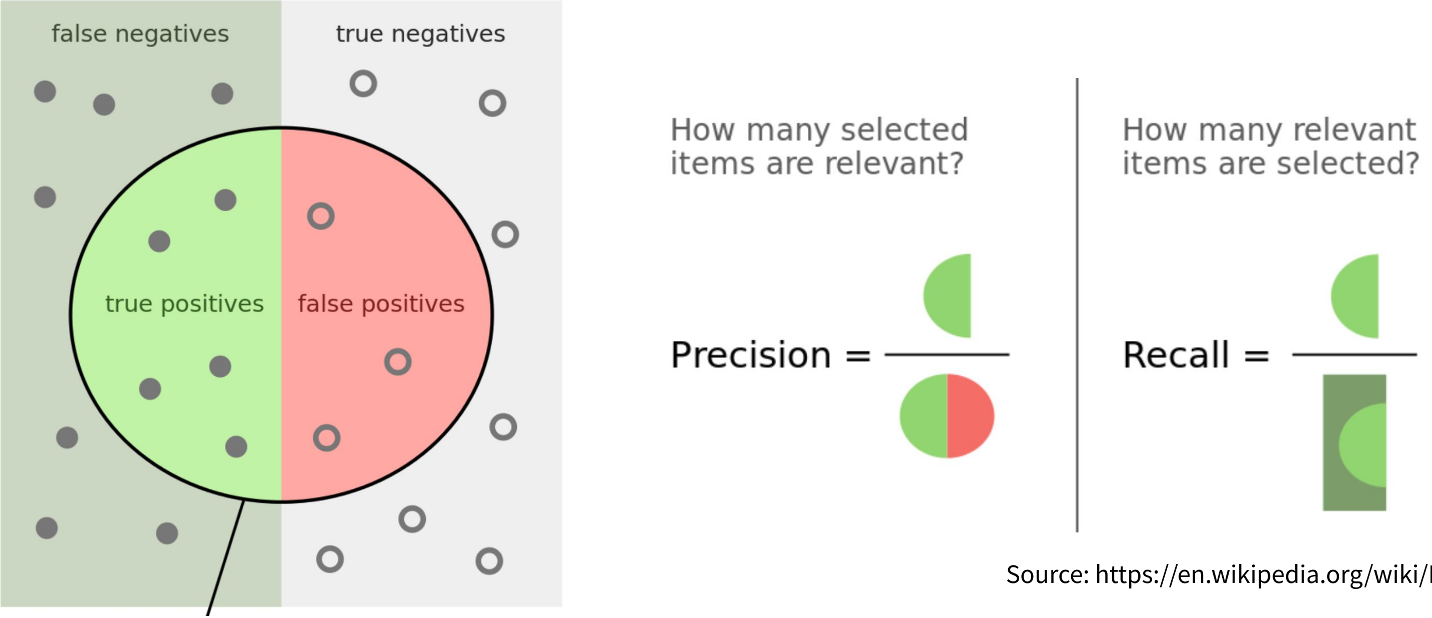


$R^2 = 0.484$, MSE = 4.487

Result: comparing to naive model

How do these results compare to naive model?

- Model that randomly predicts classes based on class frequency



Source: https://en.wikipedia.org/wiki/Precision_and_recall

Result: comparing to naive model

Naive model

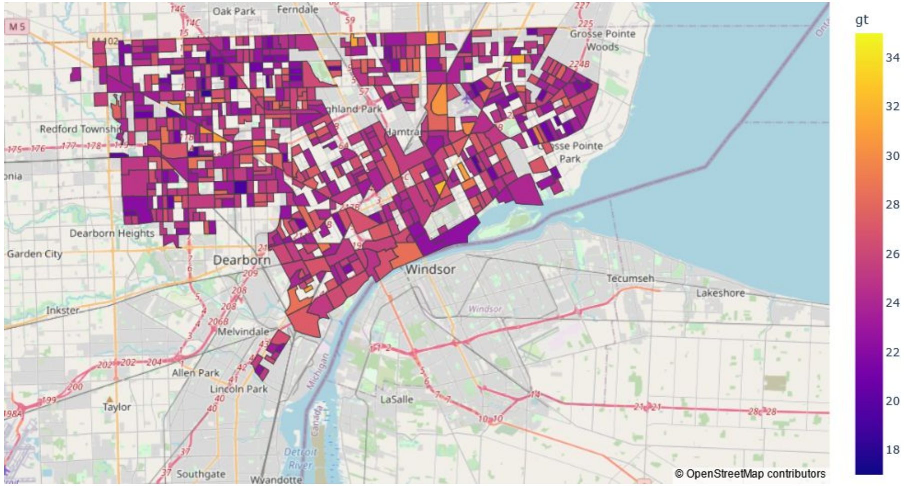
class	precision	recall	N
0	**0.05**	**0.05**	86
1	**0.09**	**0.09**	164
2	0.81	0.81	1416
3	**0.04**	**0.04**	75
accuracy			0.67

Our model

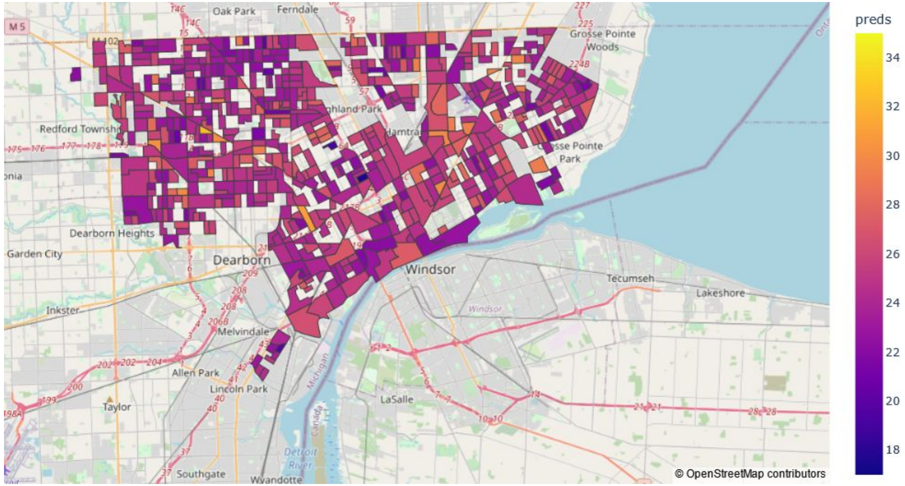
class	precision	recall	N
0	**0.42**	**0.27**	86
1	**0.25**	**0.22**	164
2	0.86	0.90	1416
3	**0.22**	**0.16**	75
accuracy			0.77

Visualization on validation dataset: Detroit

Actual score

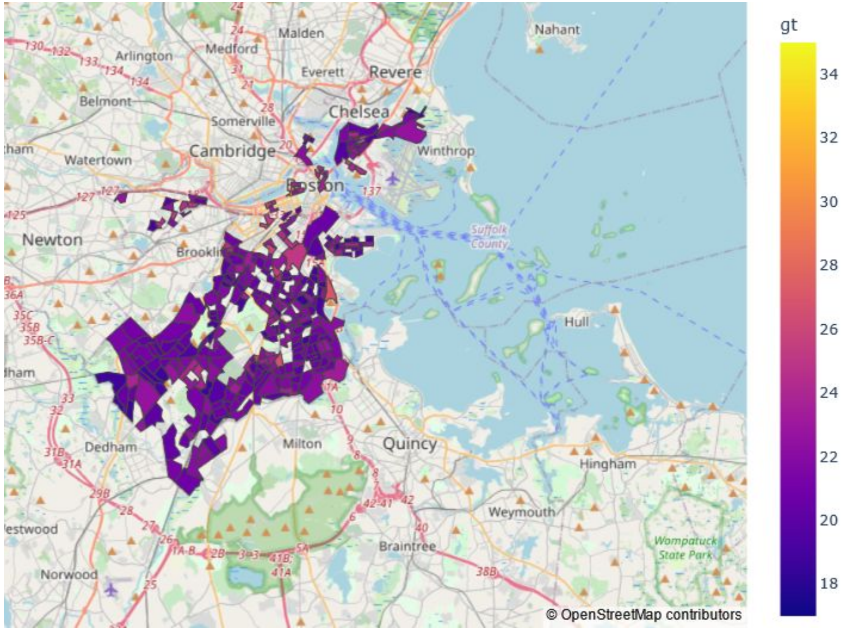


Predicted score

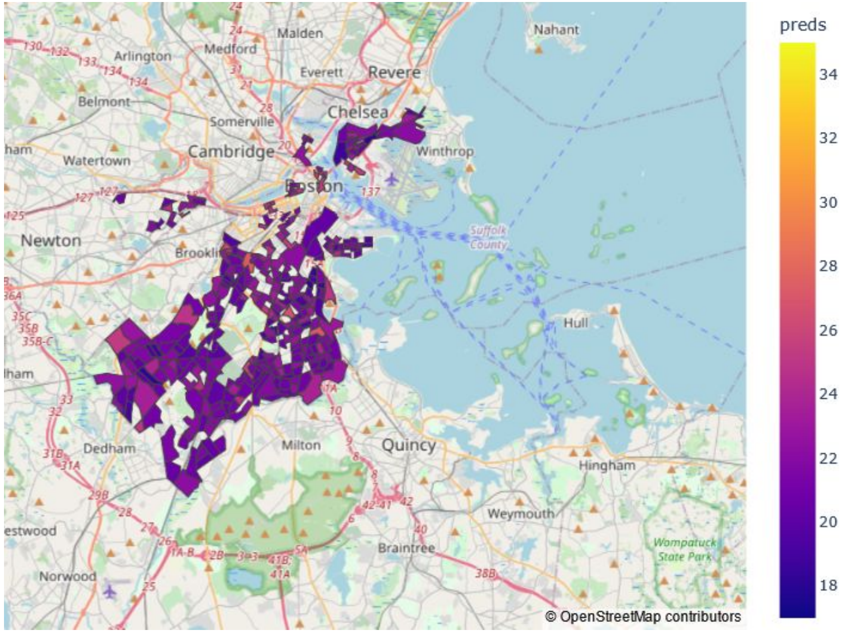


Visualization on validation dataset: Boston

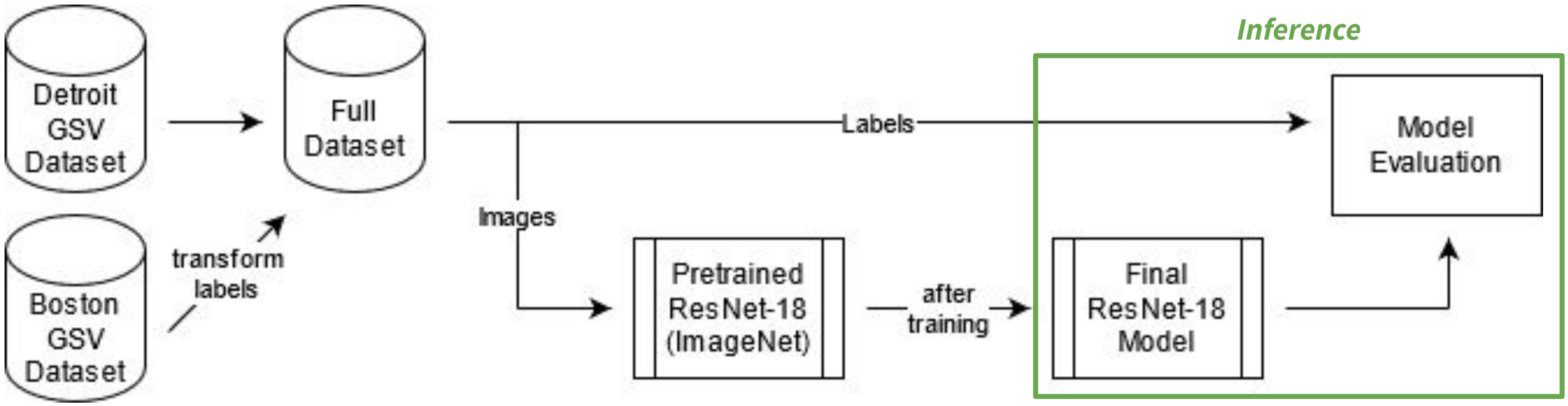
Actual score



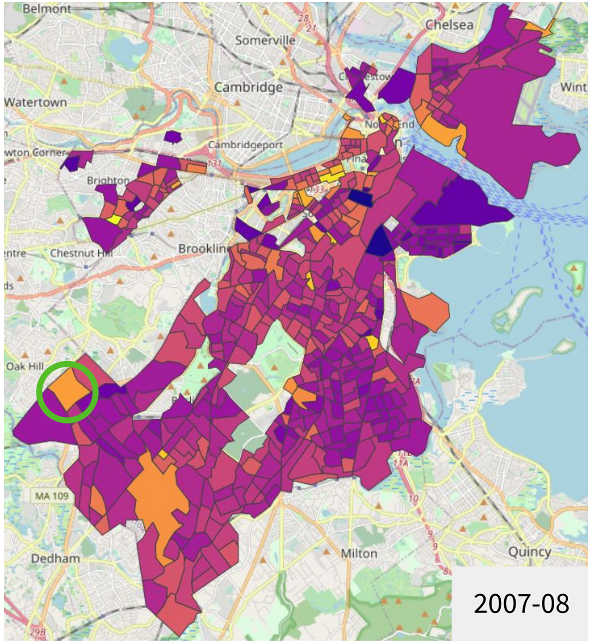
Predicted score

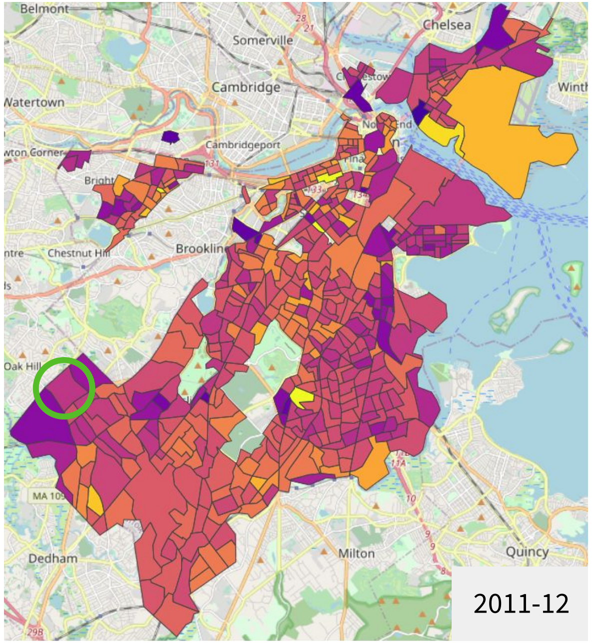


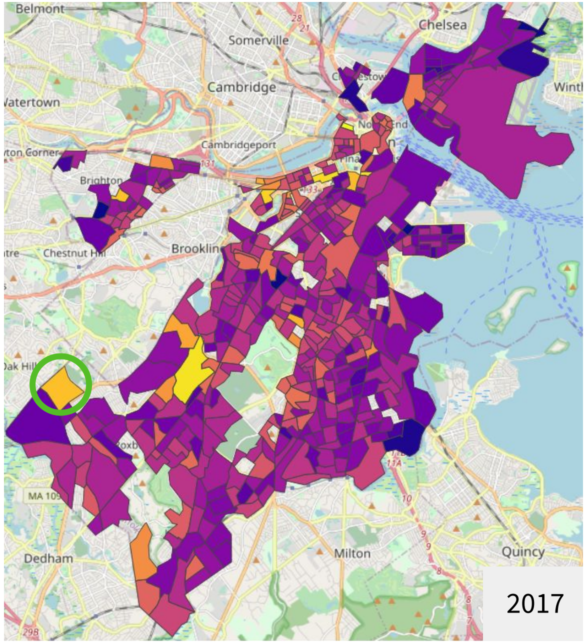
Inference on time series street view images



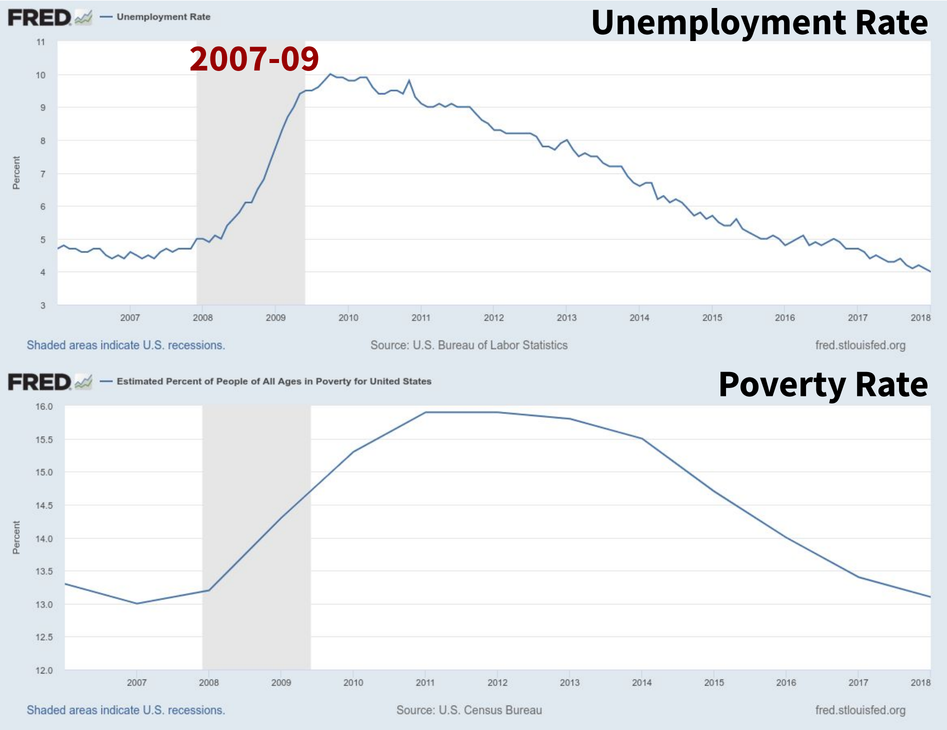
Time series analysis: Boston







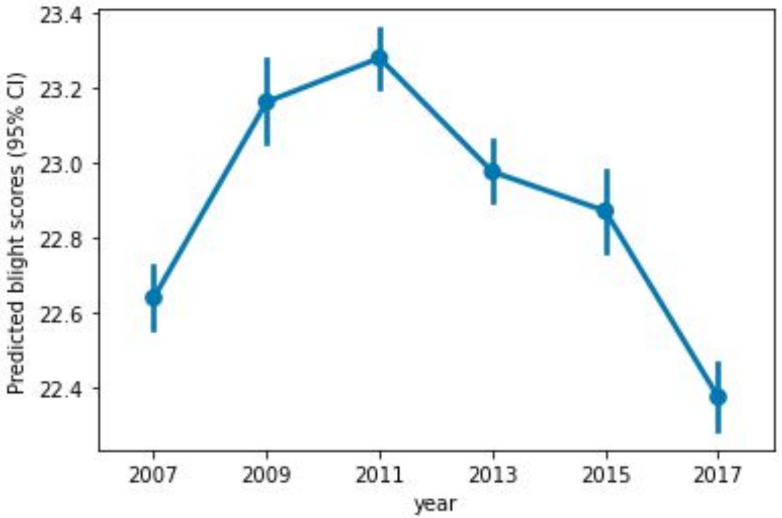
The Great Recession



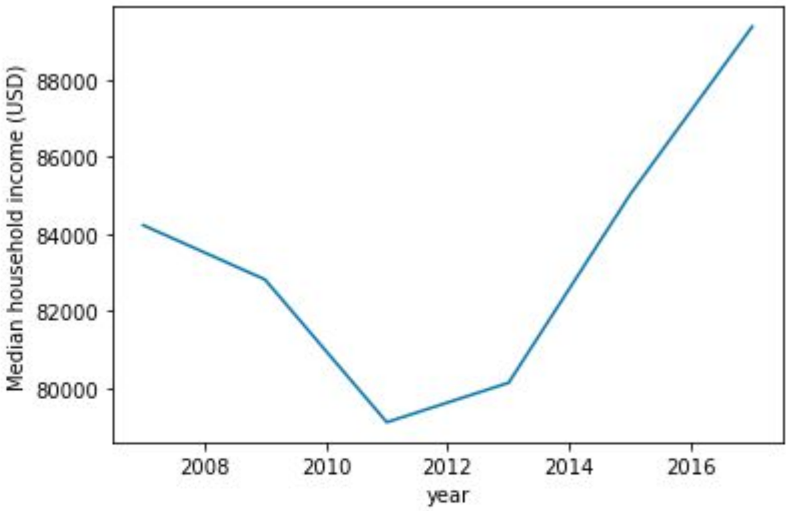
Chang, S.-S; Stuckler, D; Yip, P; Gunnell, D (2013). "Impact of 2008 global economic crisis on suicide: Time trend study in 54 countries."

Well-being analysis: Boston

Predicted blight scores (95% CI)



Median household income in Boston



Conclusion

- We trained a deep neural network to perform systematic social observation at scale.
- Our neural network can identify changes in building upkeep in different neighborhoods and cities.
- Our time series analysis shows how changes in building upkeep in Boston reflect the Great Recession and recovery.

Expected impact

- Understand how neighborhood upkeep in cities changes over time
  - Our other target cities: Austin, Detroit, Los Angeles, Philadelphia
- Understand relationship between neighborhood upkeep and well-being
  - Measures of well-being: crime rates; physical, mental health; income; subjective well-being
- Recommend policies to reduce inequities in well-being
  - e.g. targeted greening/cleaning initiatives