Using Satellite Imagery and Machine Learning to Estimate Conflict in Syria and Poverty in Mexico

Jonathan Hersh (Chapman Argyros School of Business)

R Stats NYC

4/20/2018

Wild Speculations about the Future of Data Science

(from someone barely qualified)

with Two Applications

Jonathan Hersh (Chapman Argyros School of Business)

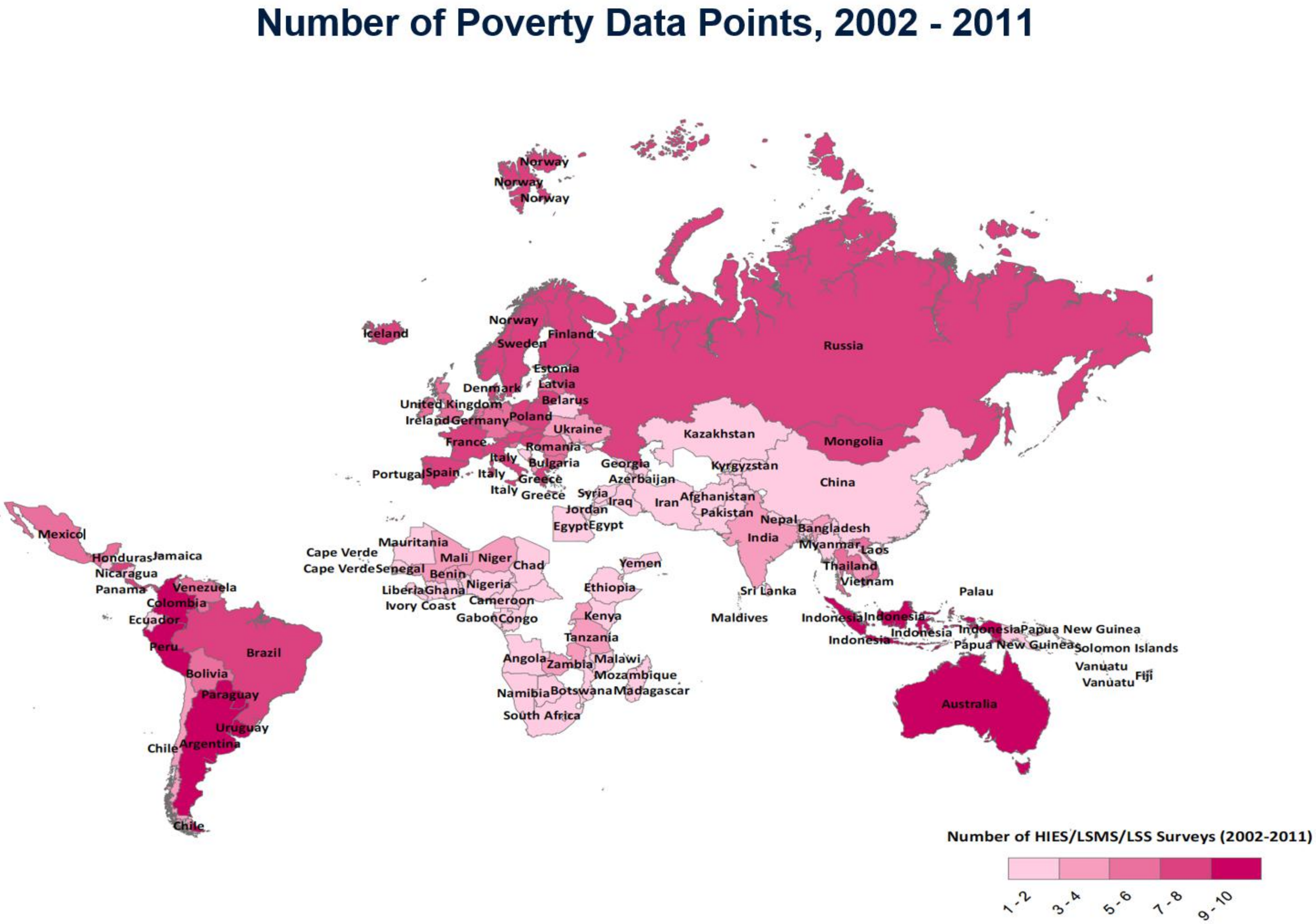
R Stats NYC

4/20/2018

Sole Qualifications: Tall and Jewish


CHAPMAN
UNIVERSITY

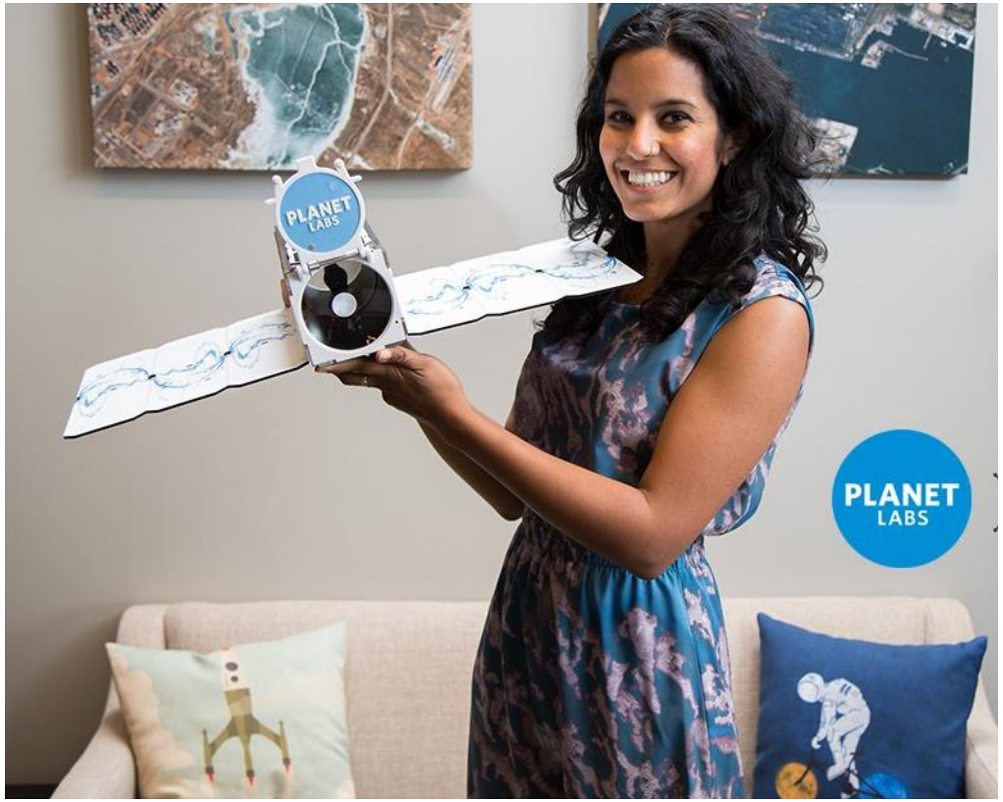
57 Countries Have Zero or One Poverty Estimate 2002-2011



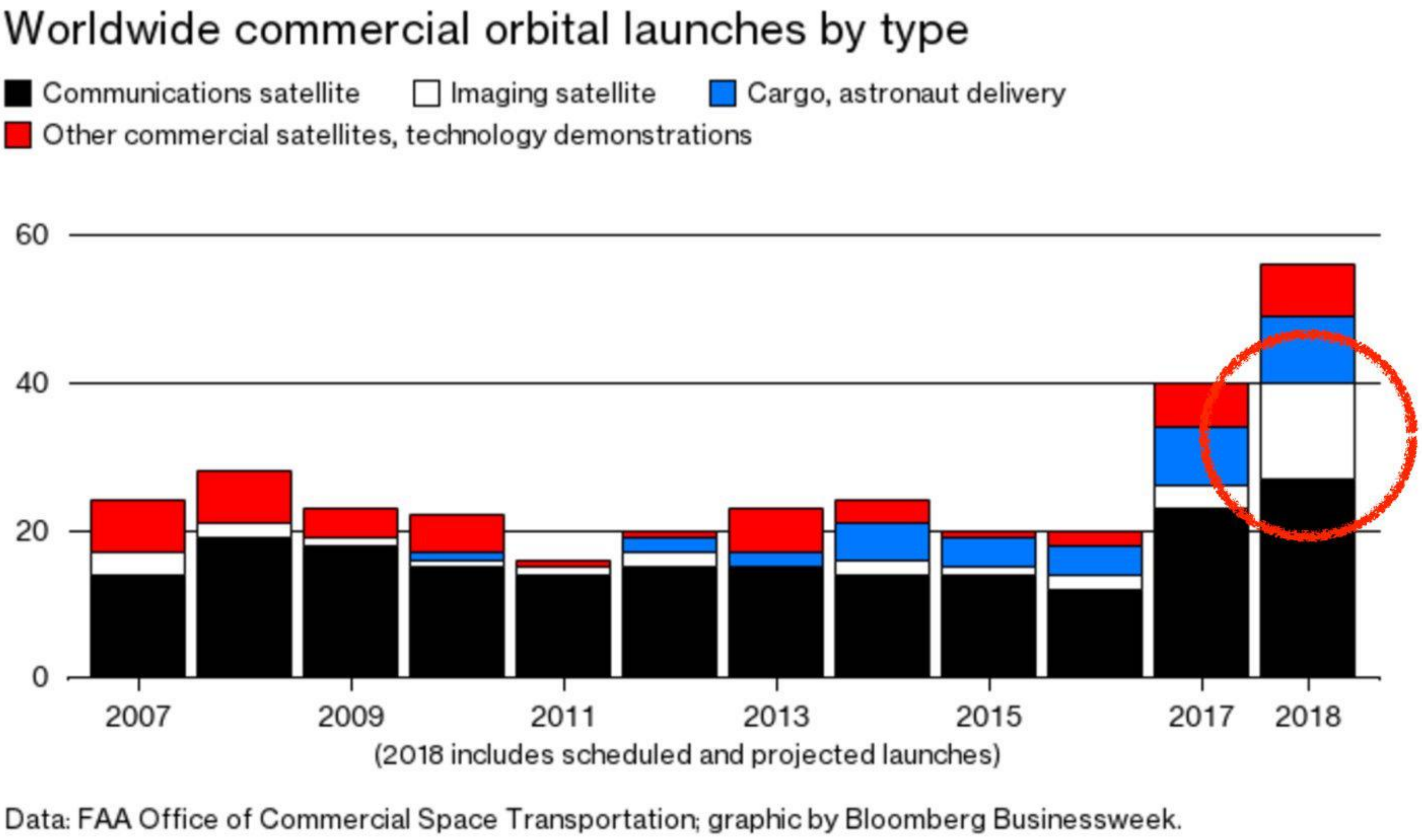
How Most Data is Collected

Lights at Night ≈ GDP

“Micro-Satellites” ~ Daily Revisit Rate



Imagery Satellites Launched in 2018



Previous Research (Engstrom, Hersh, Newhouse, 2017)
Using Intermediate Features to Estimate Poverty in Sri Lanka

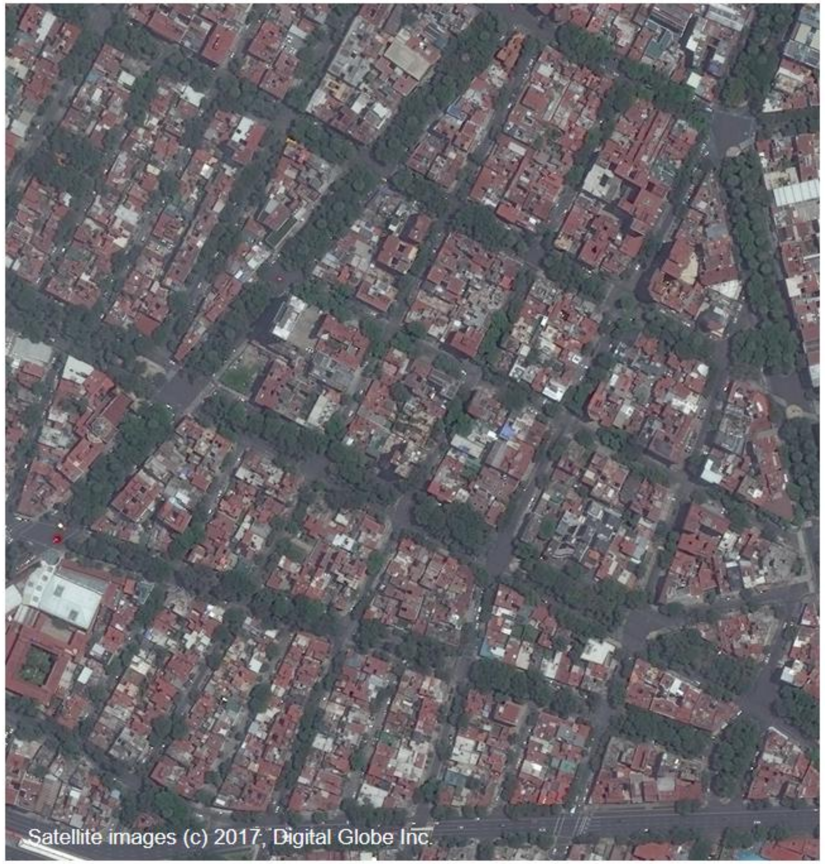
Satellite Image

Buildings

Estimate Poverty Directly from Daytime Satellite Images?

Well-off Area (Mexico City)

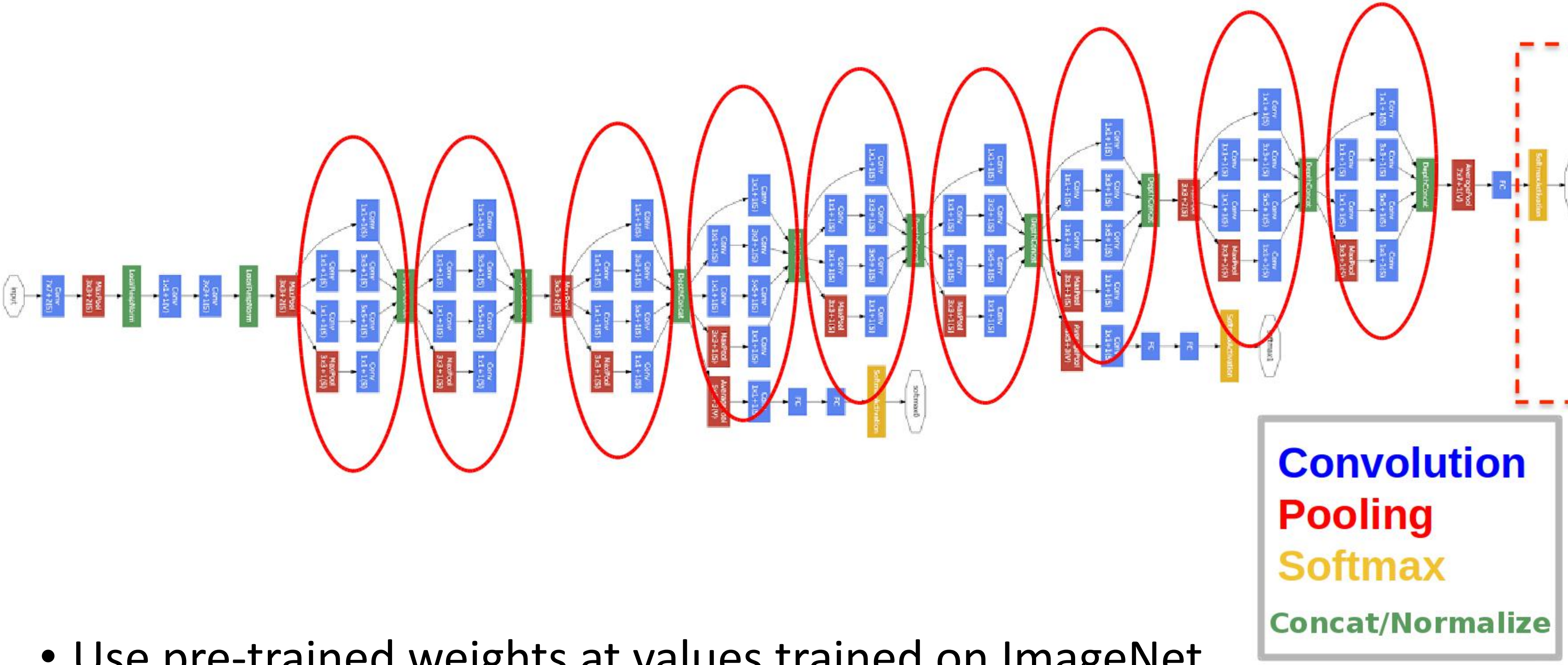
Severe Poverty	Moderate Poverty	Not Poor
16.6%	18.6%	64.8%



Poor Area (Santiago)

Severe Poverty	Moderate Poverty	Not Poor
37.0%	38.9%	24.1%

CNN Architecture: GoogLeNet with Transfer Learning



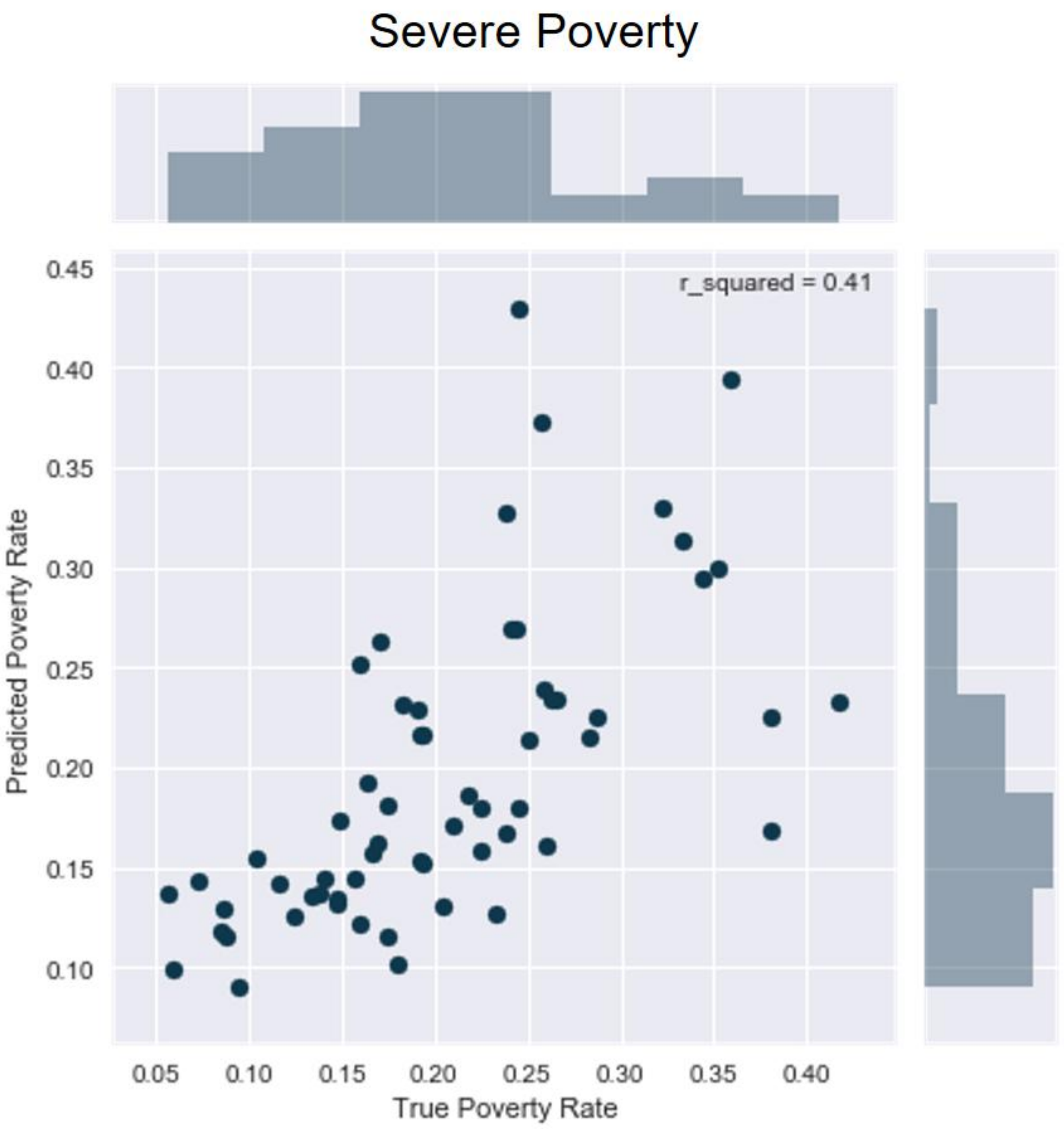
- Use pre-trained weights at values trained on ImageNet

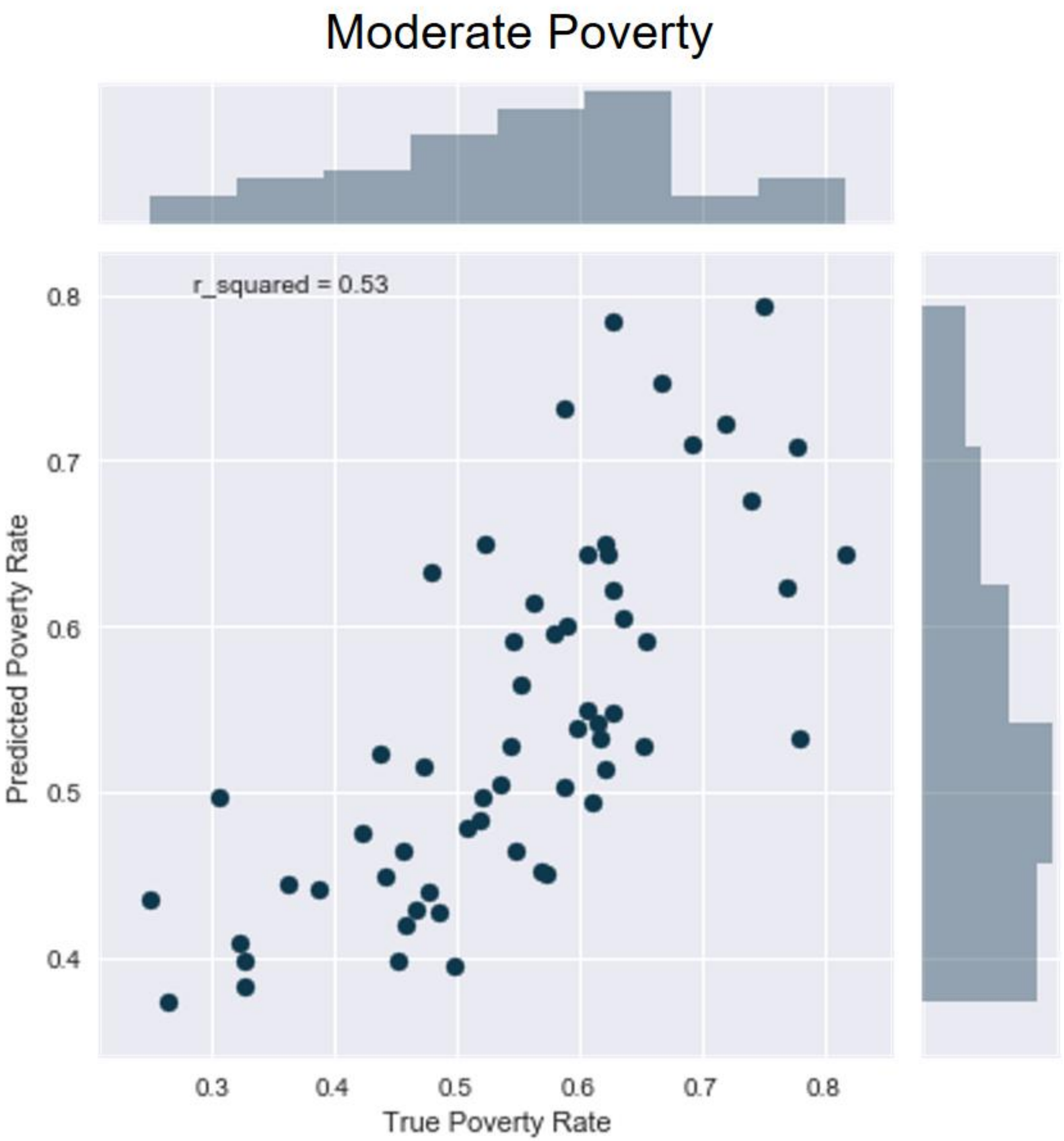
Mexico CNN: Training Details

- **Poverty Levels**
  - Use two poverty levels: **severe and moderate poverty**
  - Calculate fraction in each municipality falling in this category

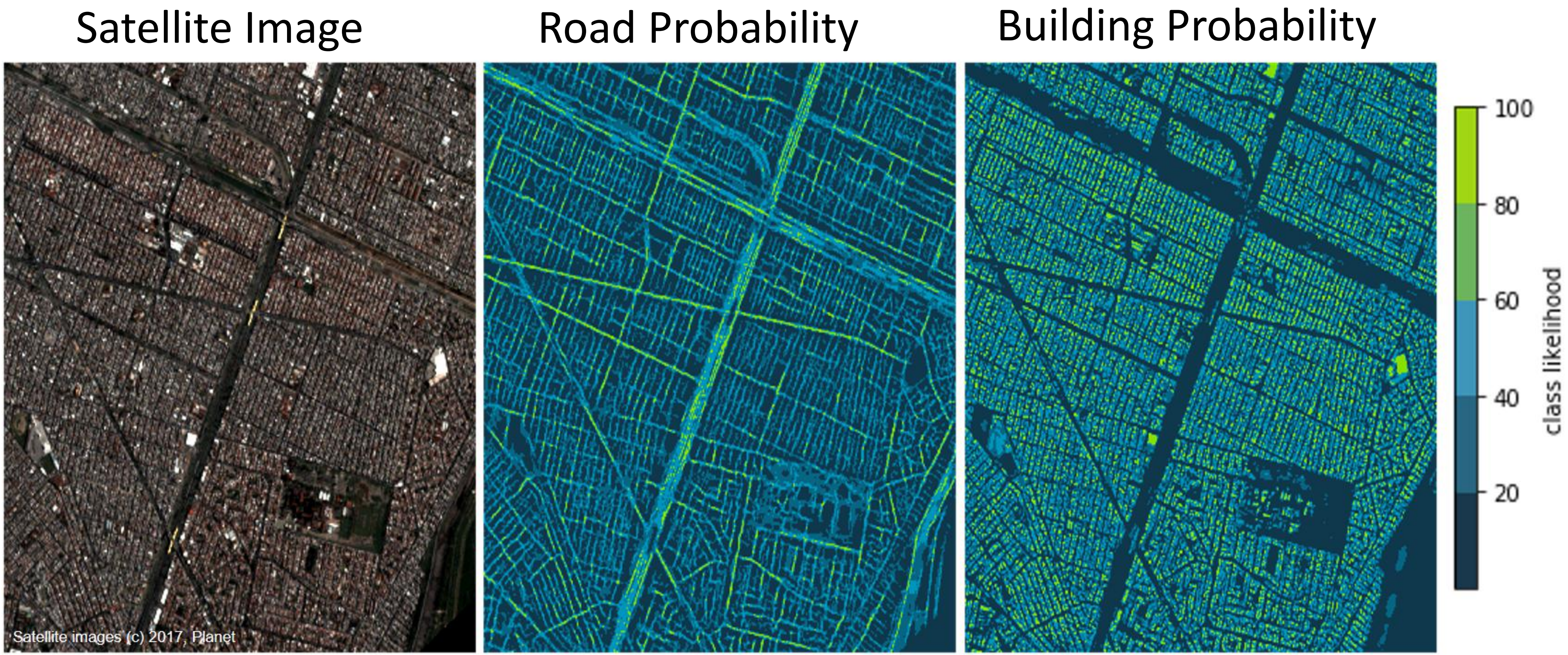
- **Training Details**
  - Withhold 10% of municipalities (90 out of 896 in MCS-ENIGH) to validate results
  - Divided remaining 806 into training and test samples
  - Train CNN → optimize using test sample → predict into validation sample

Urban Municipalities: Predicted vs True



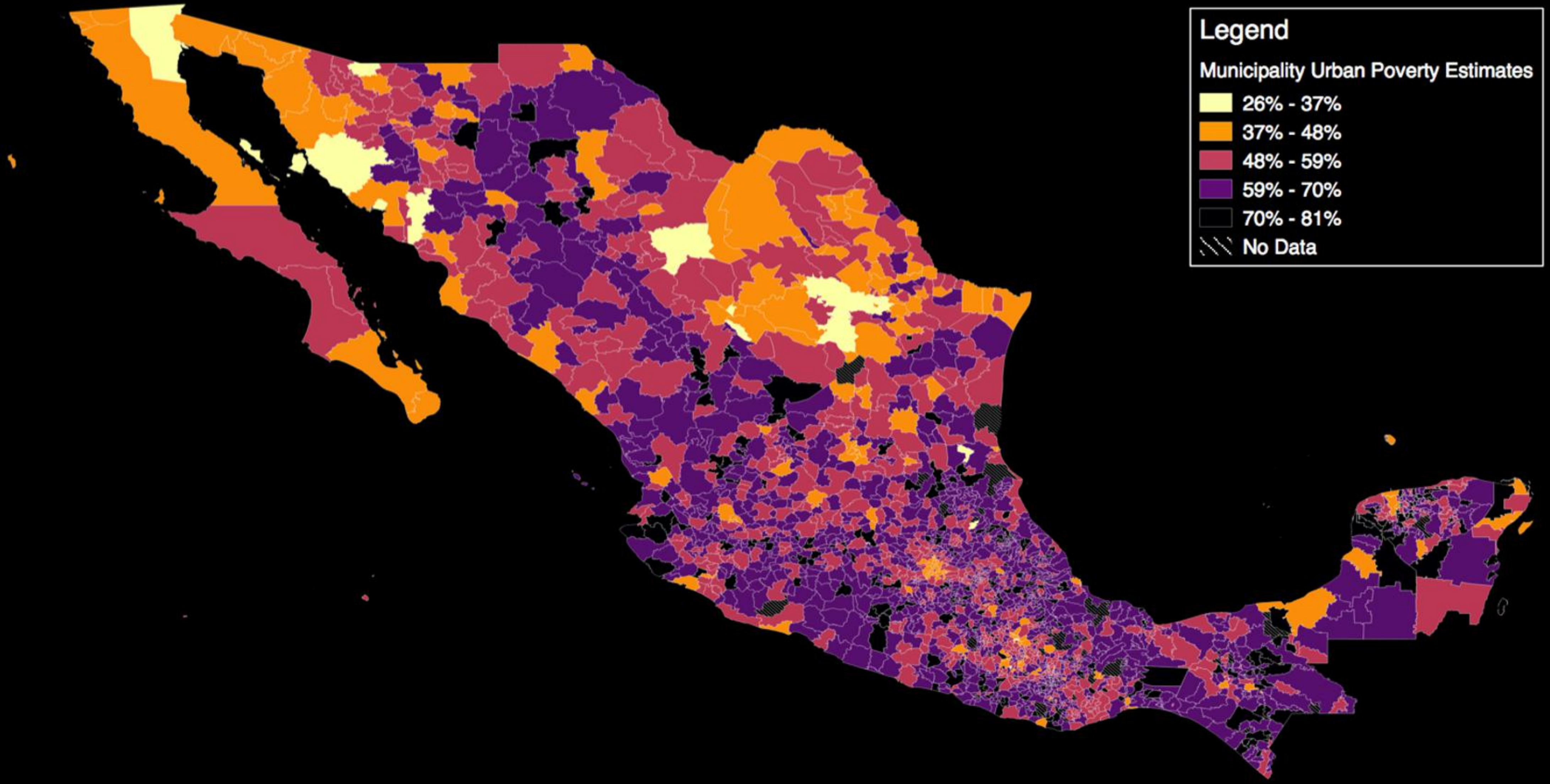


Intermediate Land Use Classification



Every Road in Mexico

Predicted Poverty Rates: Urban Municipalities



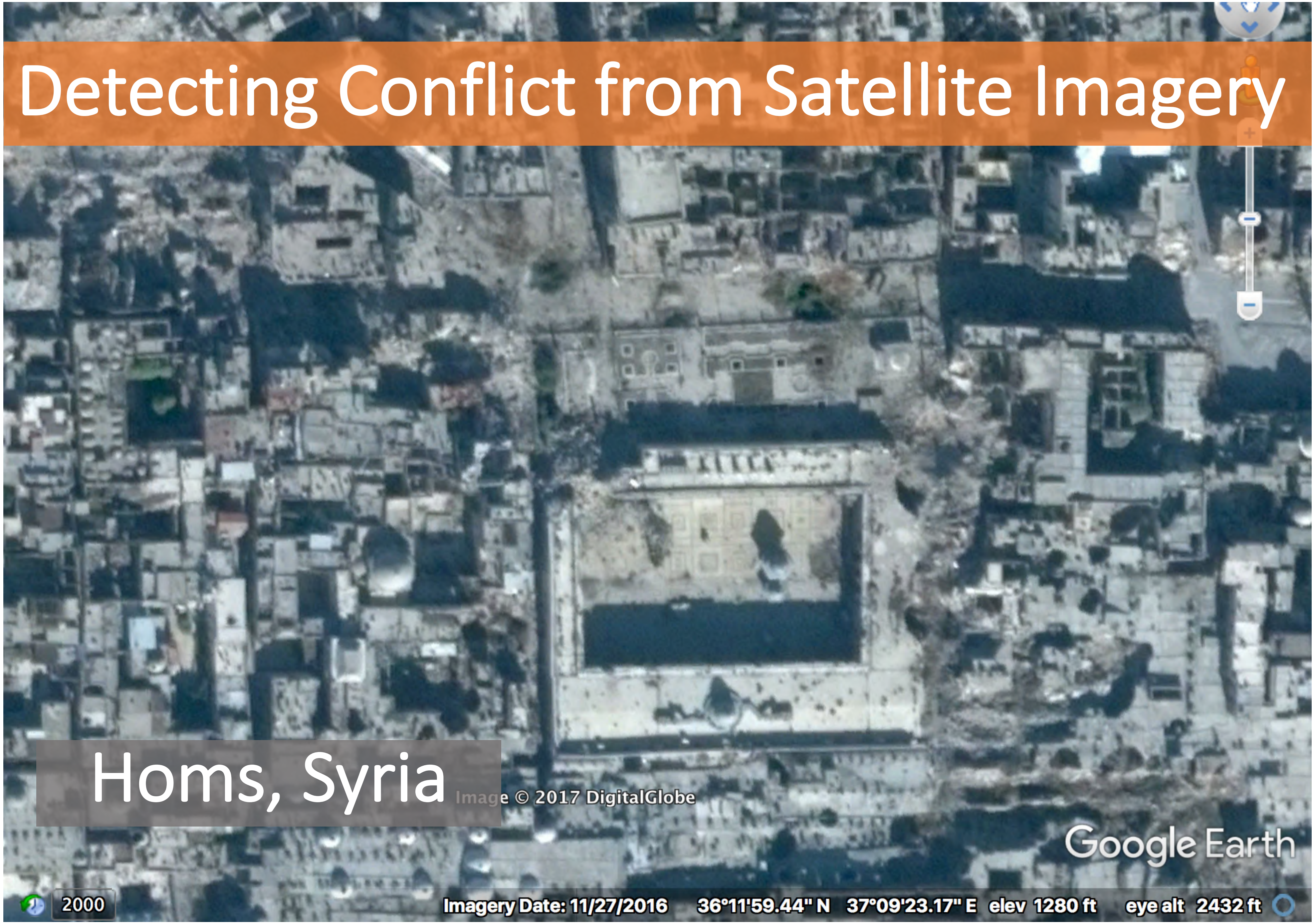
Lesson 1:

Using Off-the-Shelf Data is for Suckers

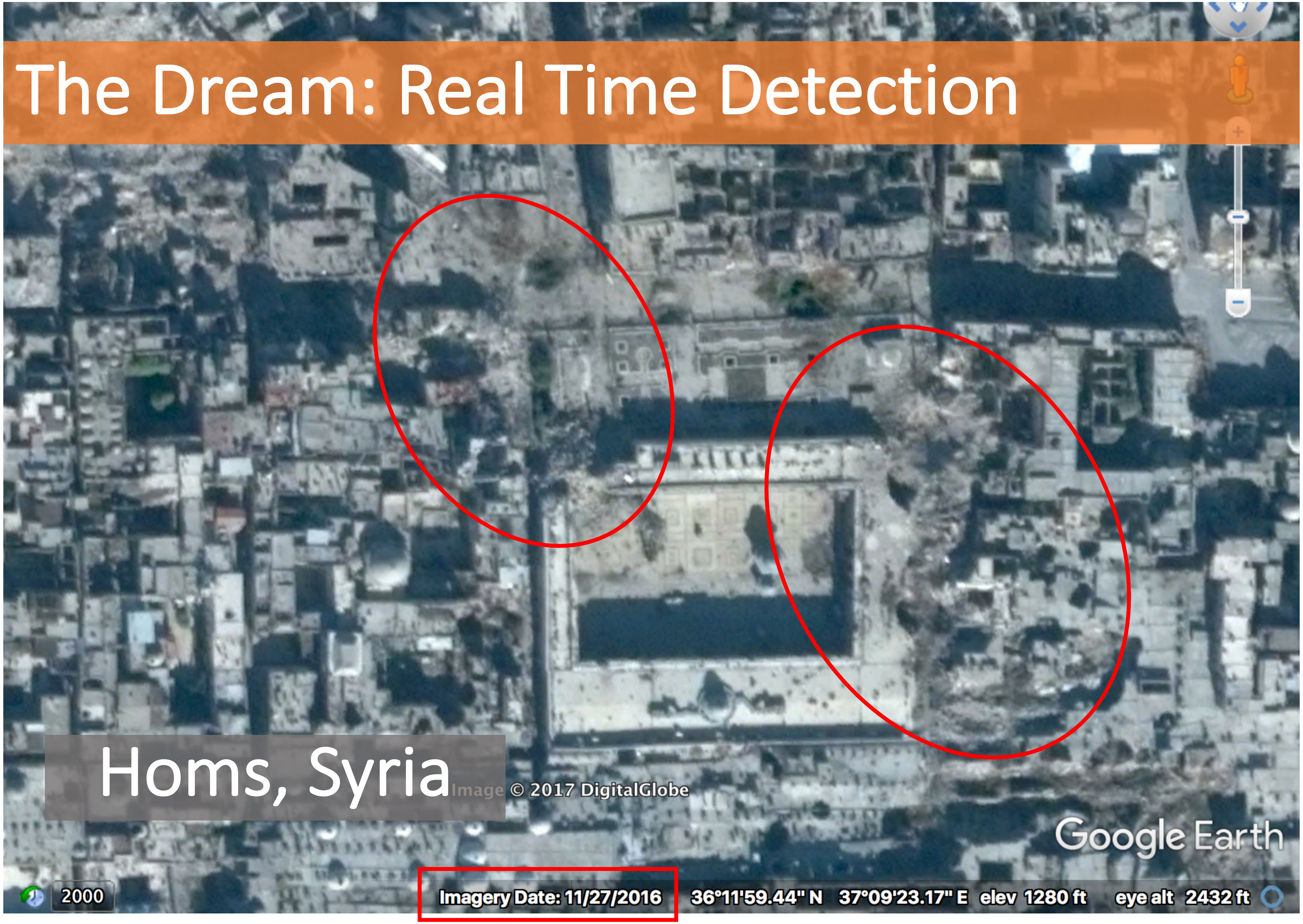
Use Machine Learning to Generate Data No One Else Has

Information on ongoing violence biased, outdated, and imprecise

Detecting Conflict from Satellite Imagery

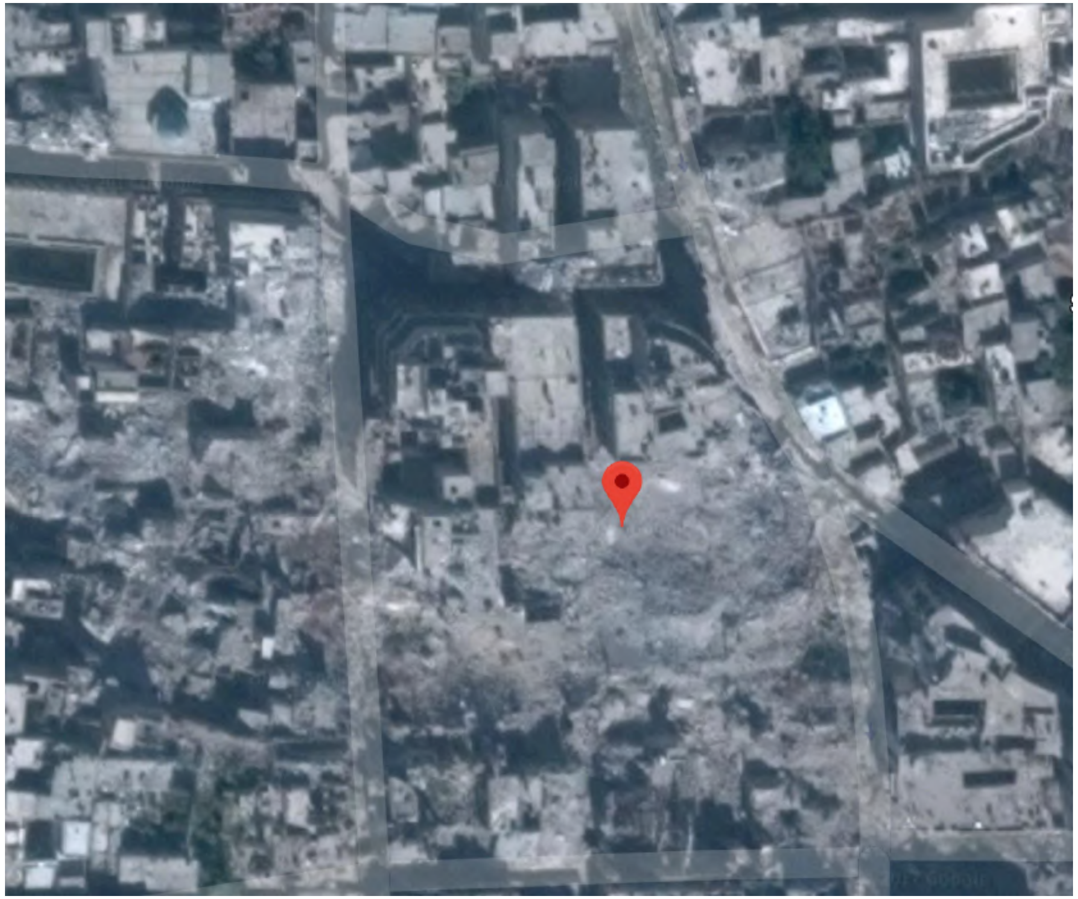


The Dream: Real Time Detection



Training sites in Syria

- Located **380 locations** in Syria with visually identified building damage
  - Eventually increase to 1,000+
- With data augmentation ~ 5,000 training images
- **Train two classifiers: one for Planet and one for Google Maps API**



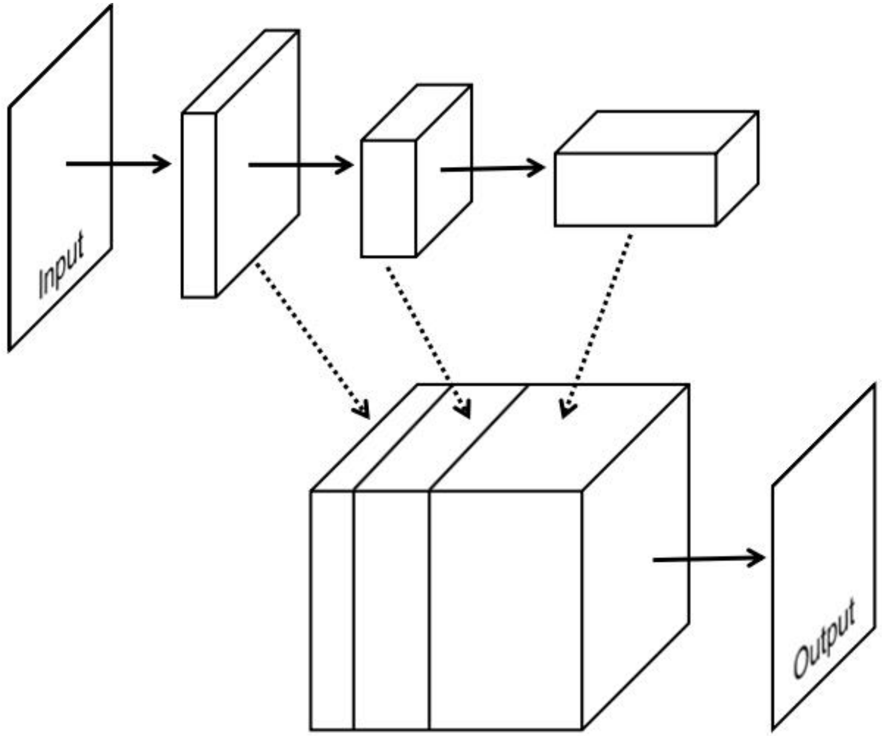
Labeling process

Satellite Image (Google Maps)

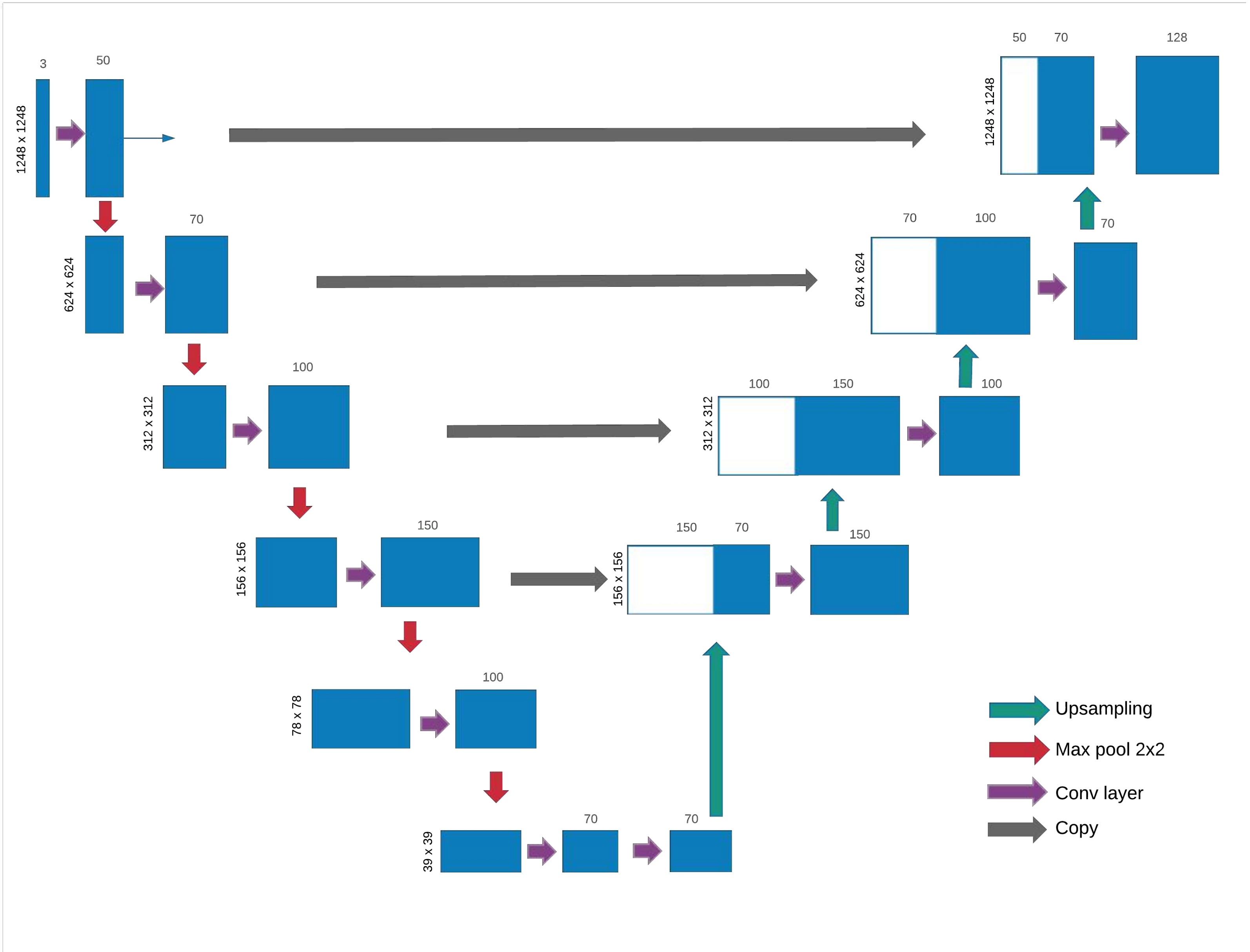
Labeled Image

CNN Architecture

- Starting point Yuan (2016)
- **7 CNNs** (with pooling in-between) with **20% dropout** between layers
- Output from 1st, 2^{nd}, and 7^{th} layer up-sampled and combined ("stacked") into feature map
- **Pixel-wise prediction** at final layer



CNN Architecture: U-Net Variation

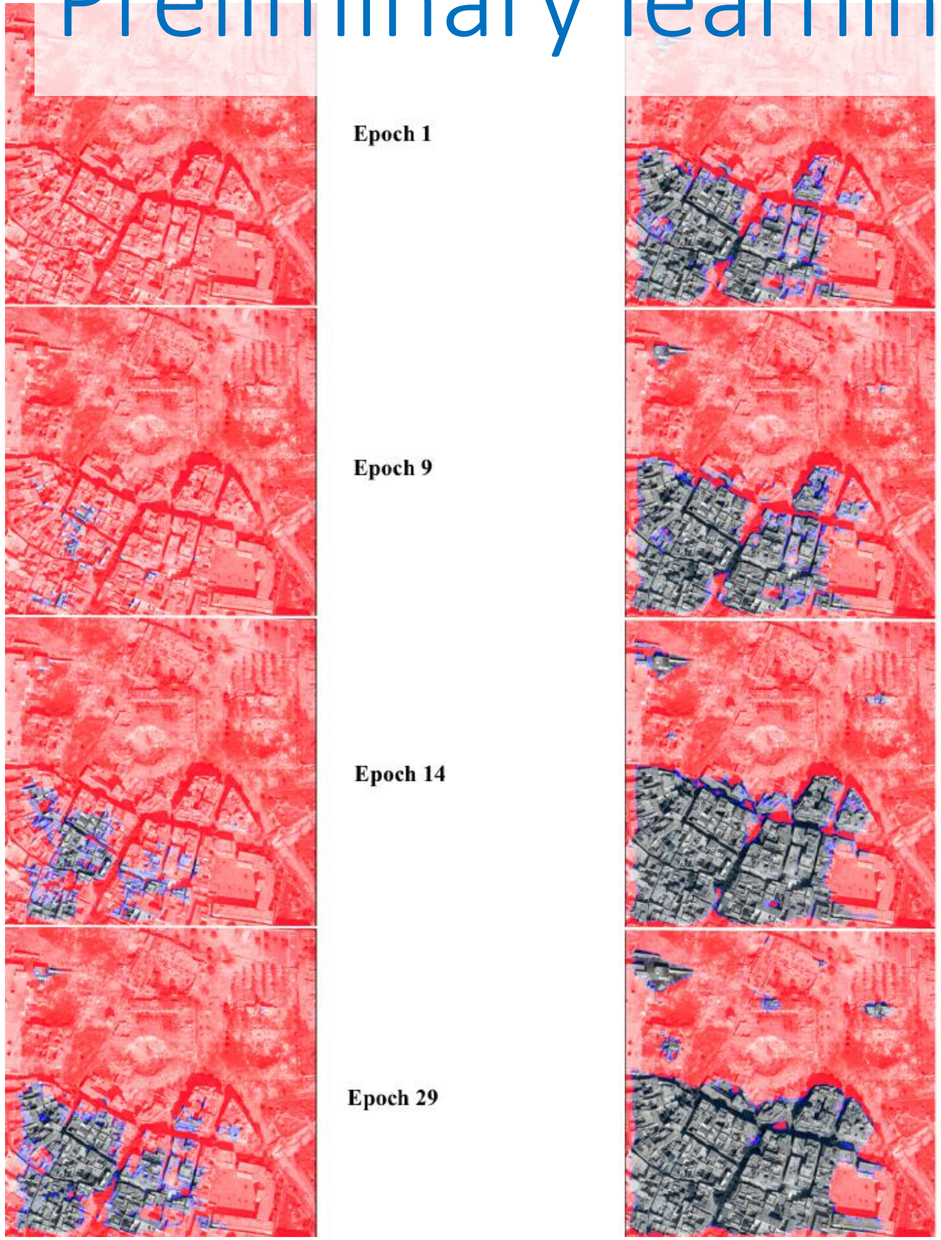


Training notes

- Implemented in Theano
- Microsoft Azure Server w/ 2 K80 GPUs
  - ~4 days
- Adam optimization
- Adjusted loss function for class imbalance
- Received $20,000 La Caixa grant on socioeconomic well-being

- Only show results for Google Maps imagery & trained on a subset of images

Preliminary learning



Epoch 1

Epoch 9

Epoch 14

Epoch 29

Epoch 37

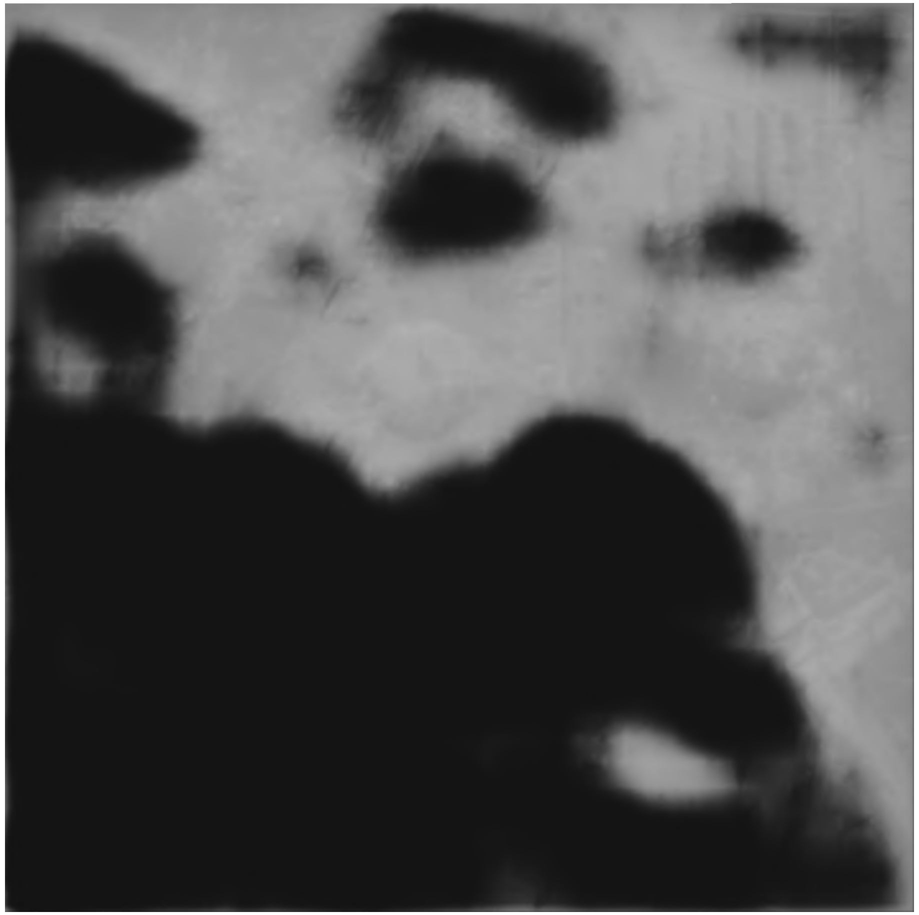
Epoch 40

Epoch 48

Epoch 53

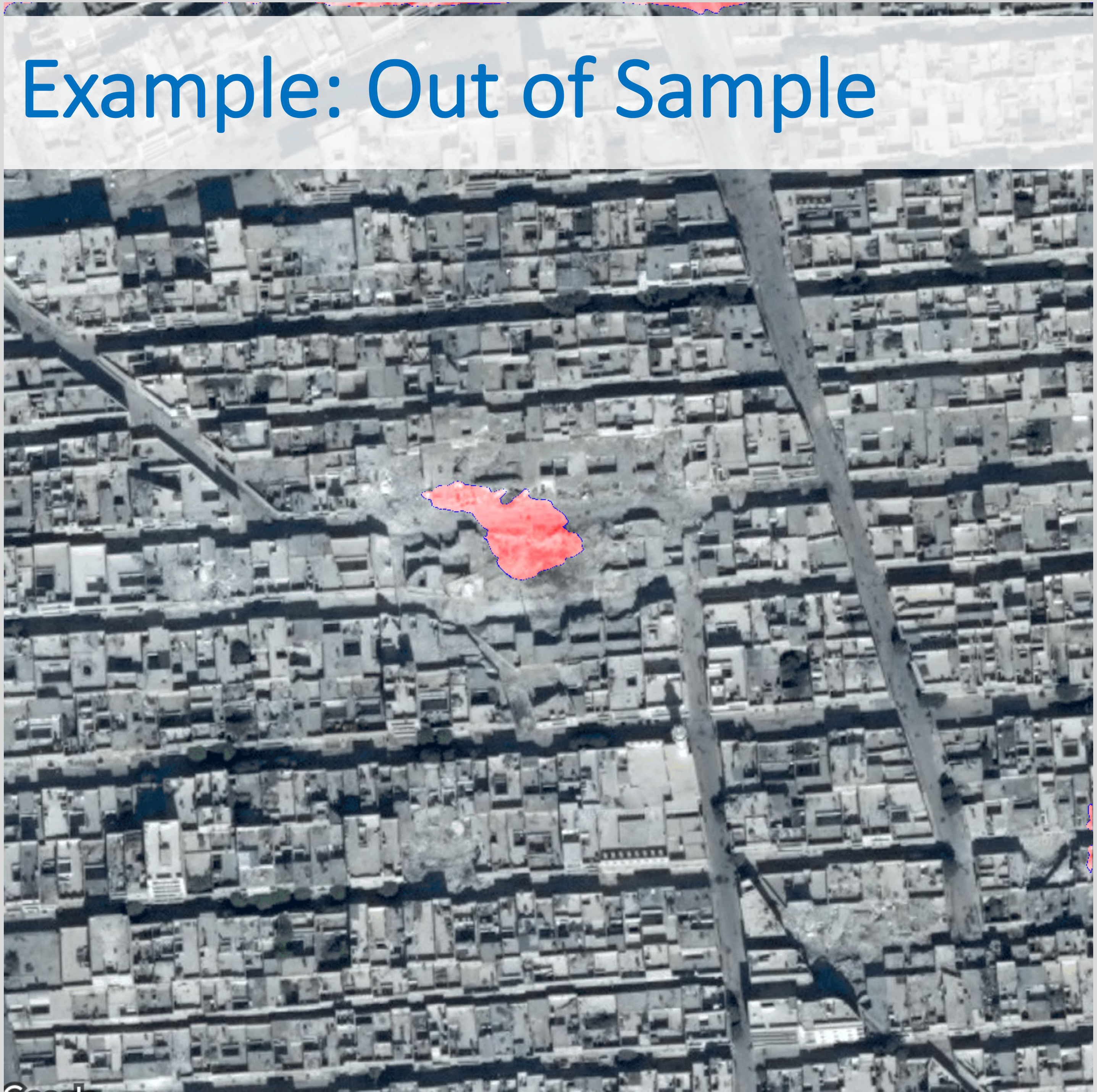
Final output: 800 Epochs

Labeled Destroyed Area

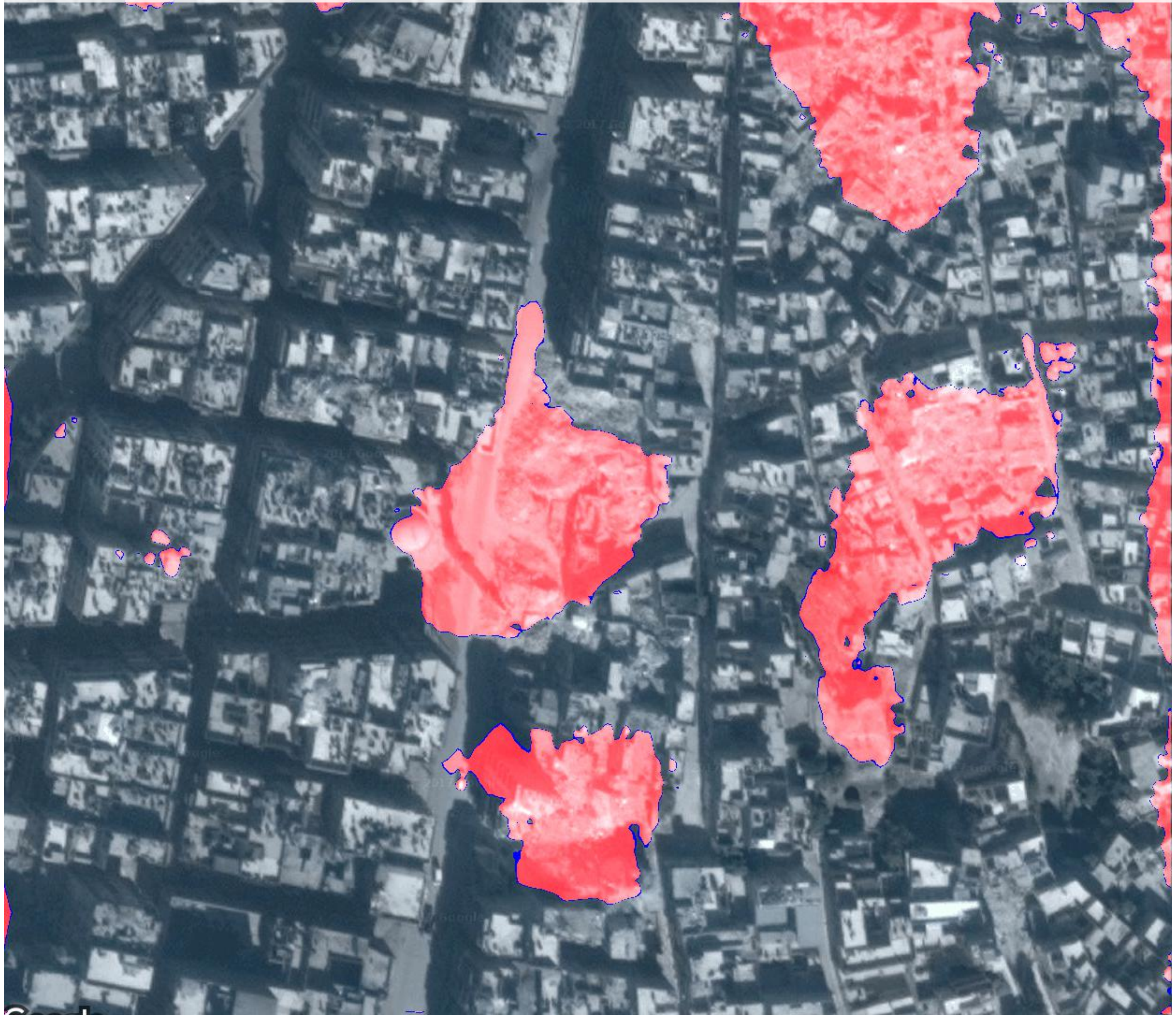


Predicted Destroyed Area

Example: Out of Sample



Example: Out of Sample



Accuracy Summary

True positive rate (TPR), Recall, Sensitivity,	False positive rate (FPR), Fall-out,
0.9847595837	0.139072130447
False negative rate (FNR),	True negative rate (TNR), Specificity (SPC)
0.0152404162998	0.860927869553

Next steps

- Planet imagery (daily revisit)
- A history of violence in Syria during the civil war (2011-2017)
- Predict refugee flows
- Analyze reporting biases (compare to Carter Center)
- Domain transfer to learn generalizable features (e.g. Yemen?)

Lesson 2:

Data flows in real time

Static models $\succ$ Dynamic models

Comments/suggestions appreciated!

Jonathan Hersh

Argyros School of Business, Chapman University

hersh@chapman.edu

(Please talk to me if you know about Hierarchical/Group Lasso)

(Or want to give me research $)