Predicting economic statistics from remote sensing data

Hyunjoo Yang
College of Economics, Sogang University

hyang@sogang.ac.kr

This project is supported by ESA Network of Resources Initiative

Objective of the project

- Empirical investigations of income and wealth distributions has been held back by the lack of sufficiently detailed high-quality data (Elbers et al. 2003)

- Goal of this research project is to understand which economic statistics can be extracted from compare publicly available remote sensing and web map data

NOR sponsorshop

- Contribution of the NOR sponsorship for this project
- Allowed systematically downloading large Sentinel 2 data covering South Korea
  - Sentinel 2 is the key input for this machine learning research project

Highlight of the benefits to the society from this project

- Gain further understanding on how remote sensing data can be used for predicting economic statistics
- This is crucial information and knowledge for many low-income countries which lack high-quality economic statistics
- This project provides implication that population density across the world at granular level can be robustly predicted using Sentinel 2

This presentation

1. Literature on remote sensing + economic stat.

2. South Korea with high quality ground truth data

1. Literature

Nightlight data

- Henderson et al. (2012)
- Nightlight intensity data can be used to predict GDP

Nightlight intensity predicts GDP
(Henderson et al. 2012)

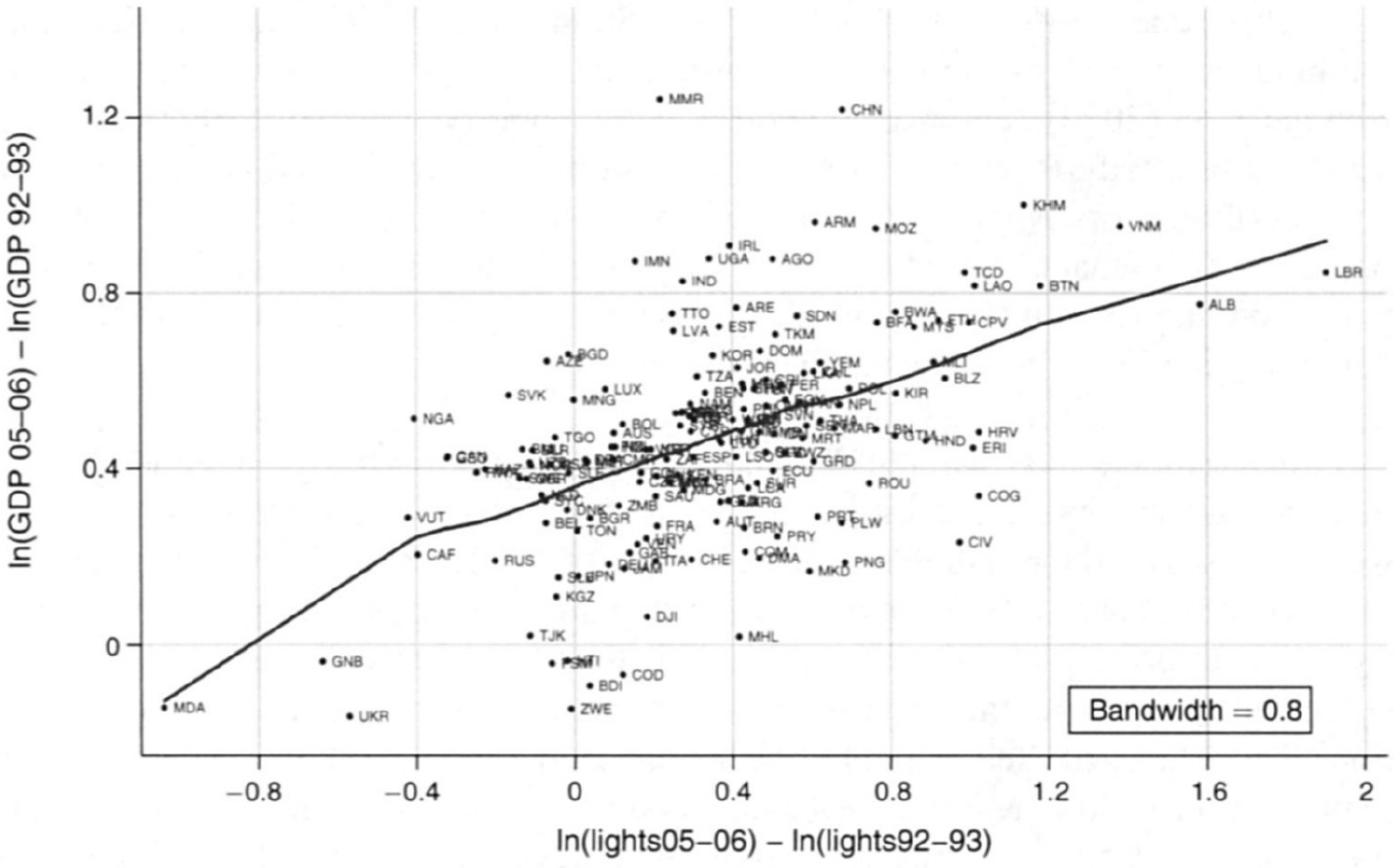
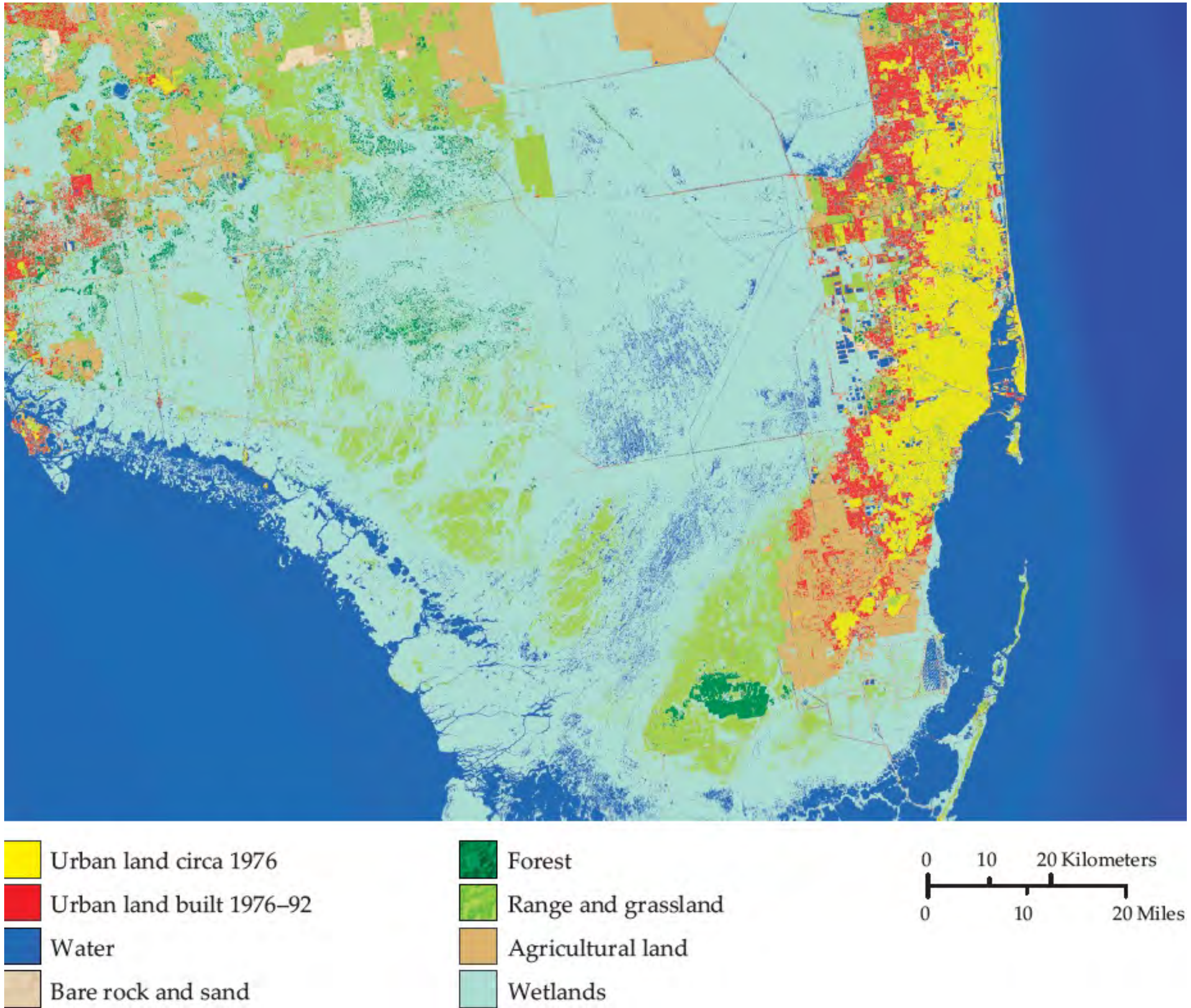


FIGURE 6

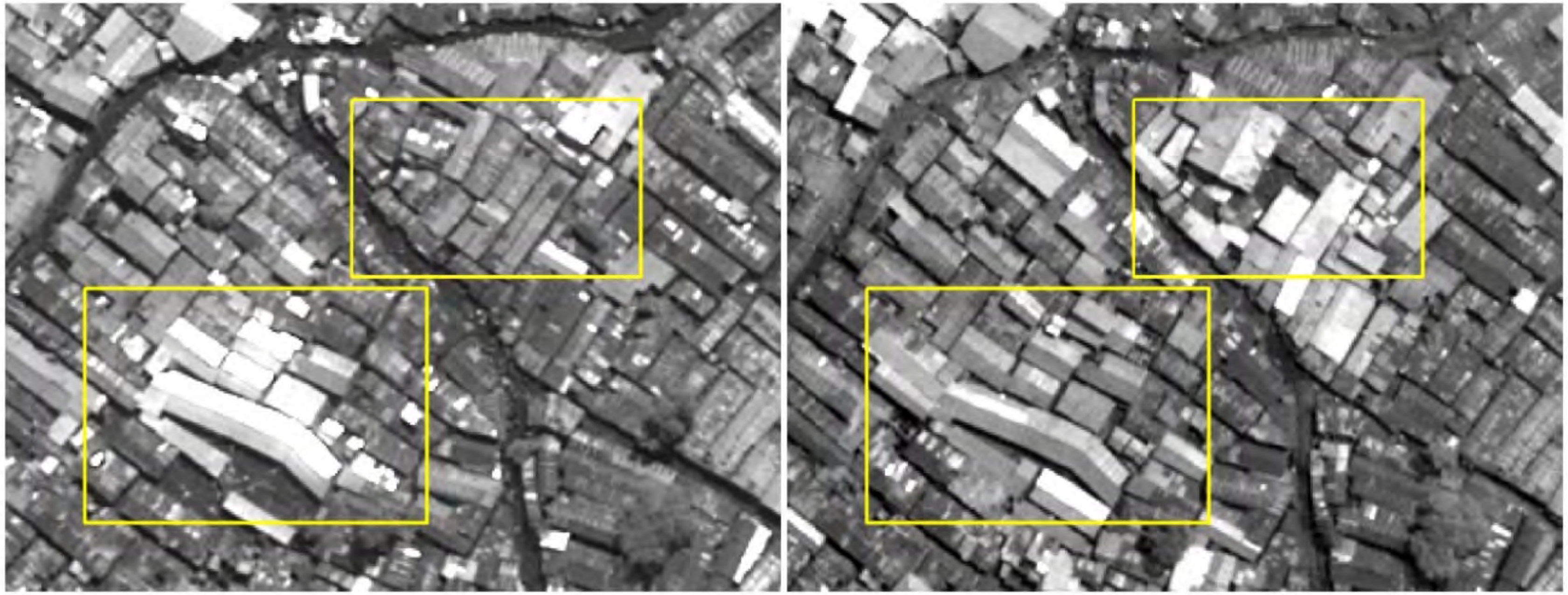
Burchfield et al. 2006



Marx et al. (2019)

- Roof types as a proxy for poverty

Appendix Figure A2: Old and New Roofs in Kibera



Note: Both pictures are taken over the same area of the slum with the same resolution (0.5 meters panchromatic).
The picture in the left panel was taken in July 2009 and that in the right panel in August 2012.
The yellow rectangles highlight clusters of roofs that markedly evolved over the period.
Roofs highlighted in the bottom rectangle degraded while roofs within the top rectangle were upgraded in the same timeframe.
The picture area is approximately 175 meters long and 140 meters wide.

Varshney et al. 2015.

- Roof types as a proxy for poverty

FIG. 1. Sample satellite image patches of the region with **(a)** thatched roofs in the center and **(b)** a metal roof in the center.

Varshney et al. 2015.

- Regional variation by roof types

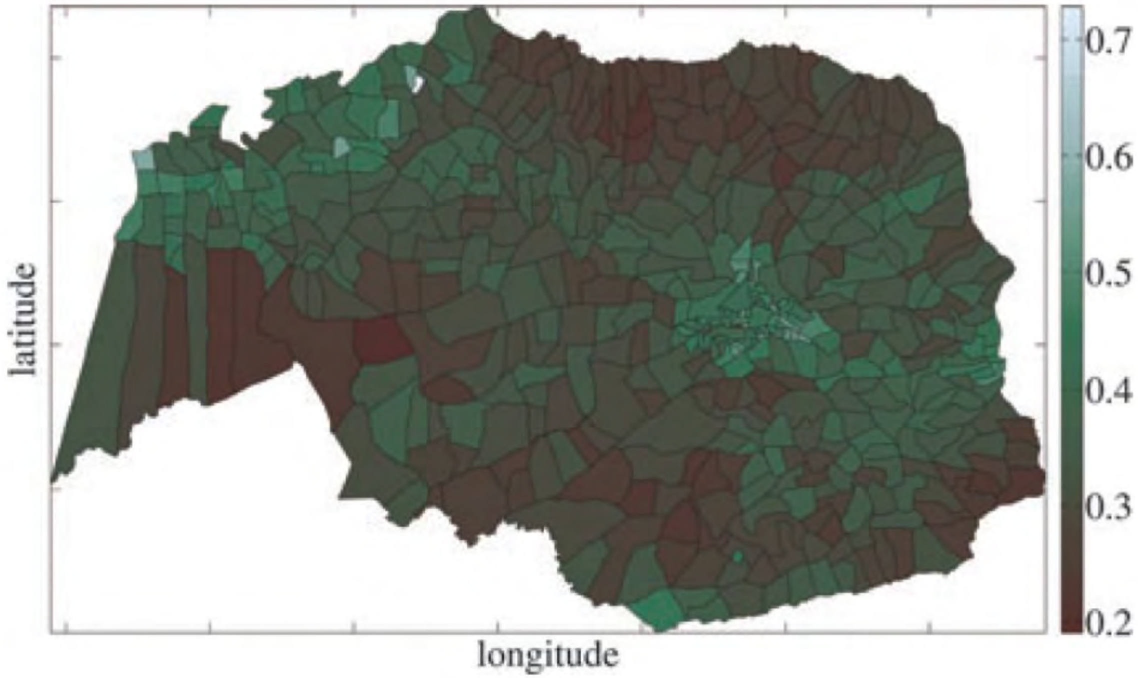


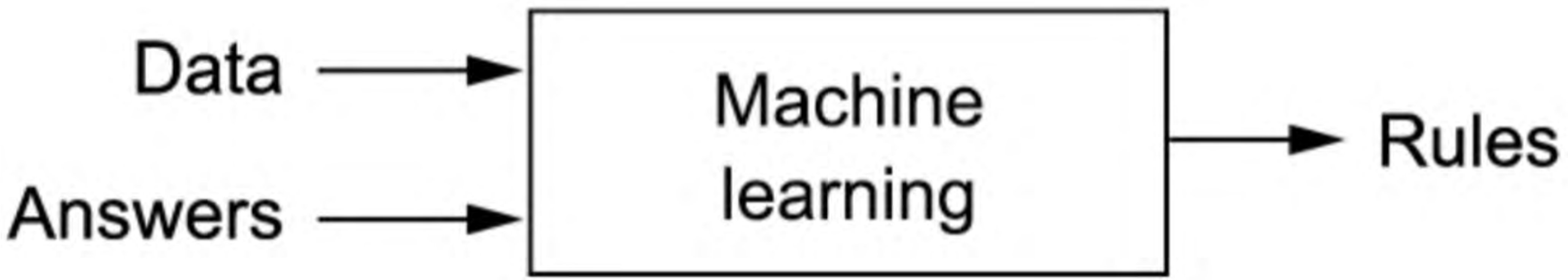
FIG. 3. Estimated proportion of roofs that are metal per village.

Recent development
machine learning + daytime satellite images

Use satellite images to predict:

- population (urbanization)
- income and consumption
- poverty
- vessel activity

Supervised computer vision deep learning



Science

RESEARCH

RESEARCH ARTICLES

ECONOMICS

Combining satellite imagery and machine learning to predict poverty

Neal Jean,[1,2*] Marshall Burke,[3,4,5*†] Michael Xie,[1] W. Matthew Davis,[4]
David B. Lobell,[3,4] Stefano Ermon[1]

Reliable data on economic livelihoods remain scarce in the developing world, hampering efforts to study these outcomes and to design policies that improve them. Here we demonstrate an accurate, inexpensive, and scalable method for estimating consumption expenditure and asset wealth from high-resolution satellite imagery. Using survey and satellite data from five African countries—Nigeria, Tanzania, Uganda, Malawi, and Rwanda—we show how a convolutional neural network can be trained to identify image features that can explain up to 75% of the variation in local-level economic outcomes. Our method, which requires only publicly available data, could transform efforts to track and target poverty in developing countries. It also demonstrates how powerful machine learning techniques can be applied in a setting with limited training data, suggesting broad potential application across many scientific domains.

Jean et al. 2016. *Science*

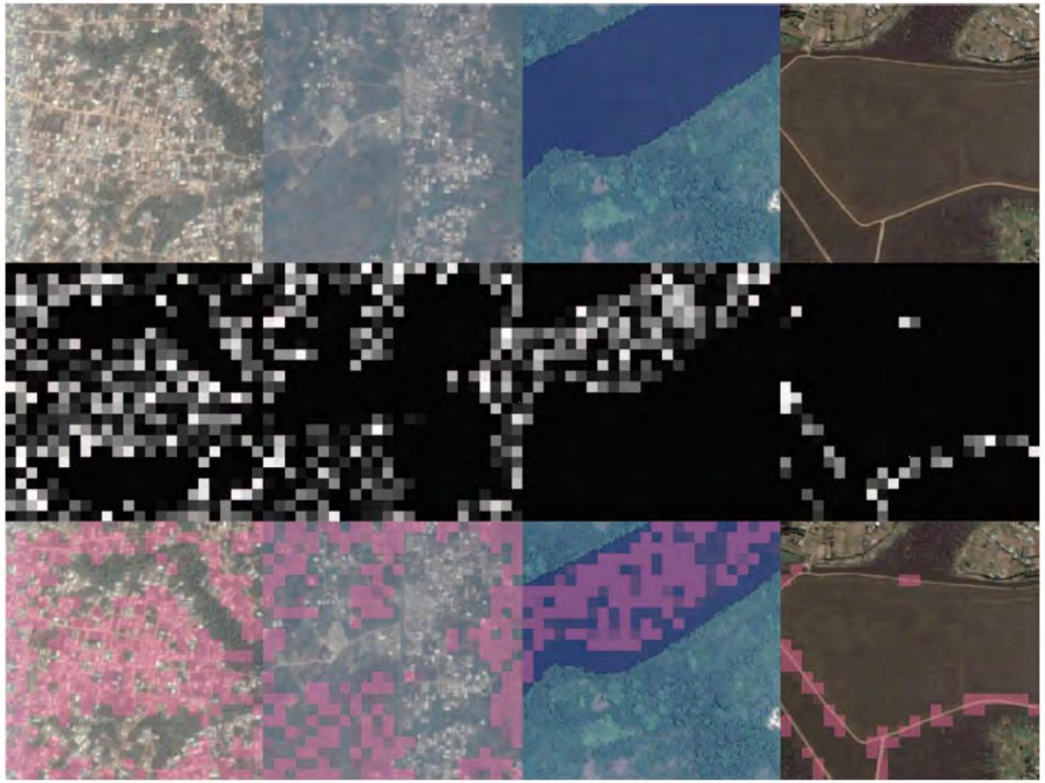
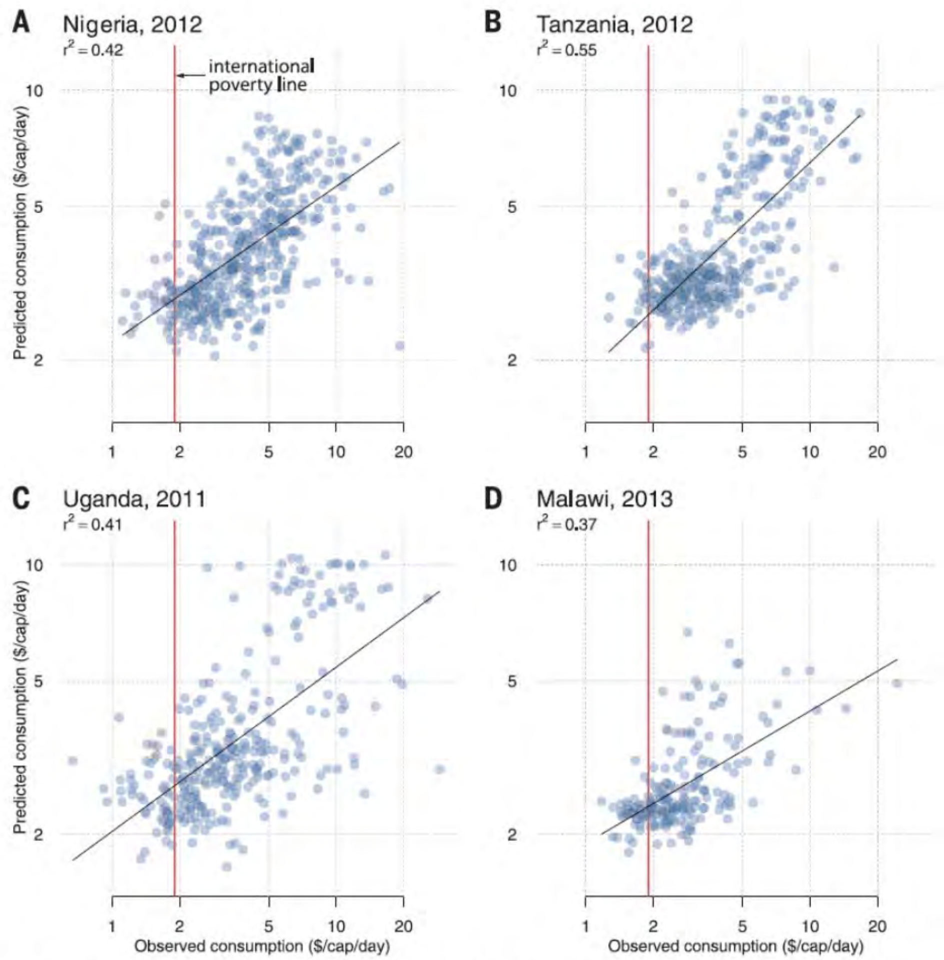
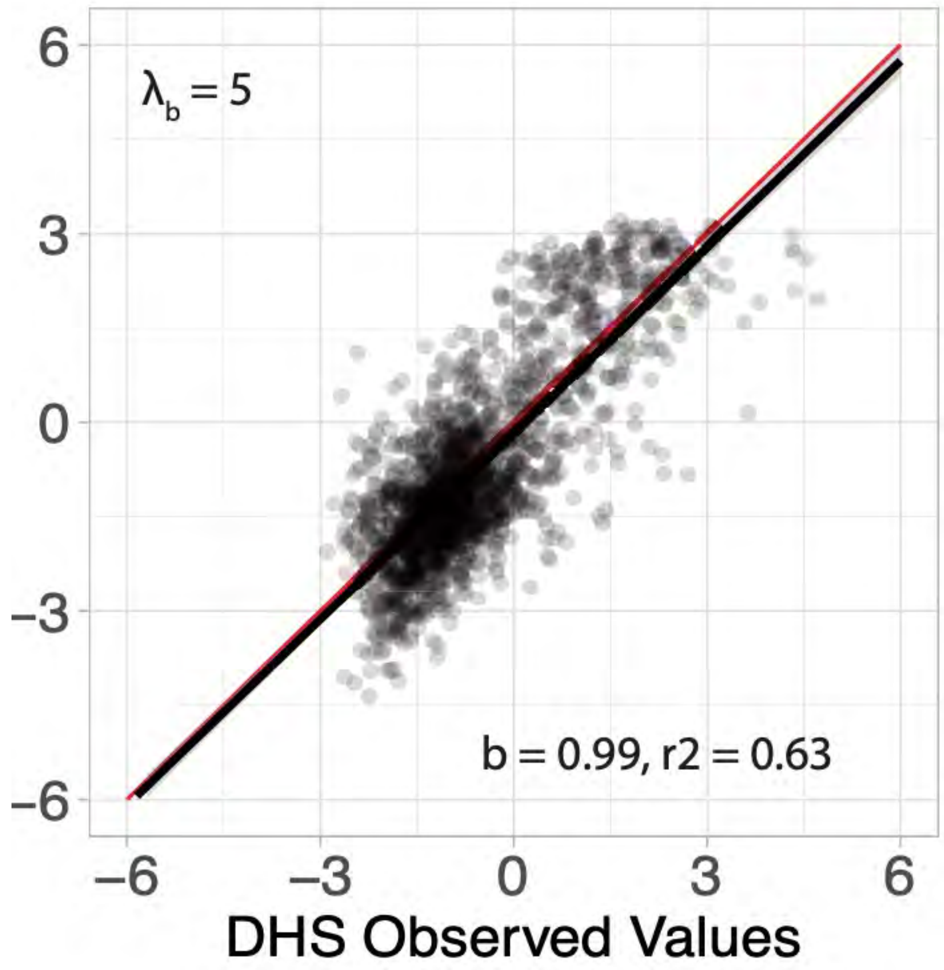


Fig. 2. Visualization of features. By column: Four different convolutional filters (which identify, from left to right, features corresponding to urban areas, nonurban areas, water, and roads) in the convolutional neural network model used for extracting features. Each filter "highlights" the parts of the image that activate it, shown in pink. By row: Original daytime satellite images from Google Static Maps, filter activation maps, and overlay of activation maps onto original images

Predicting consumption and asset in Africa

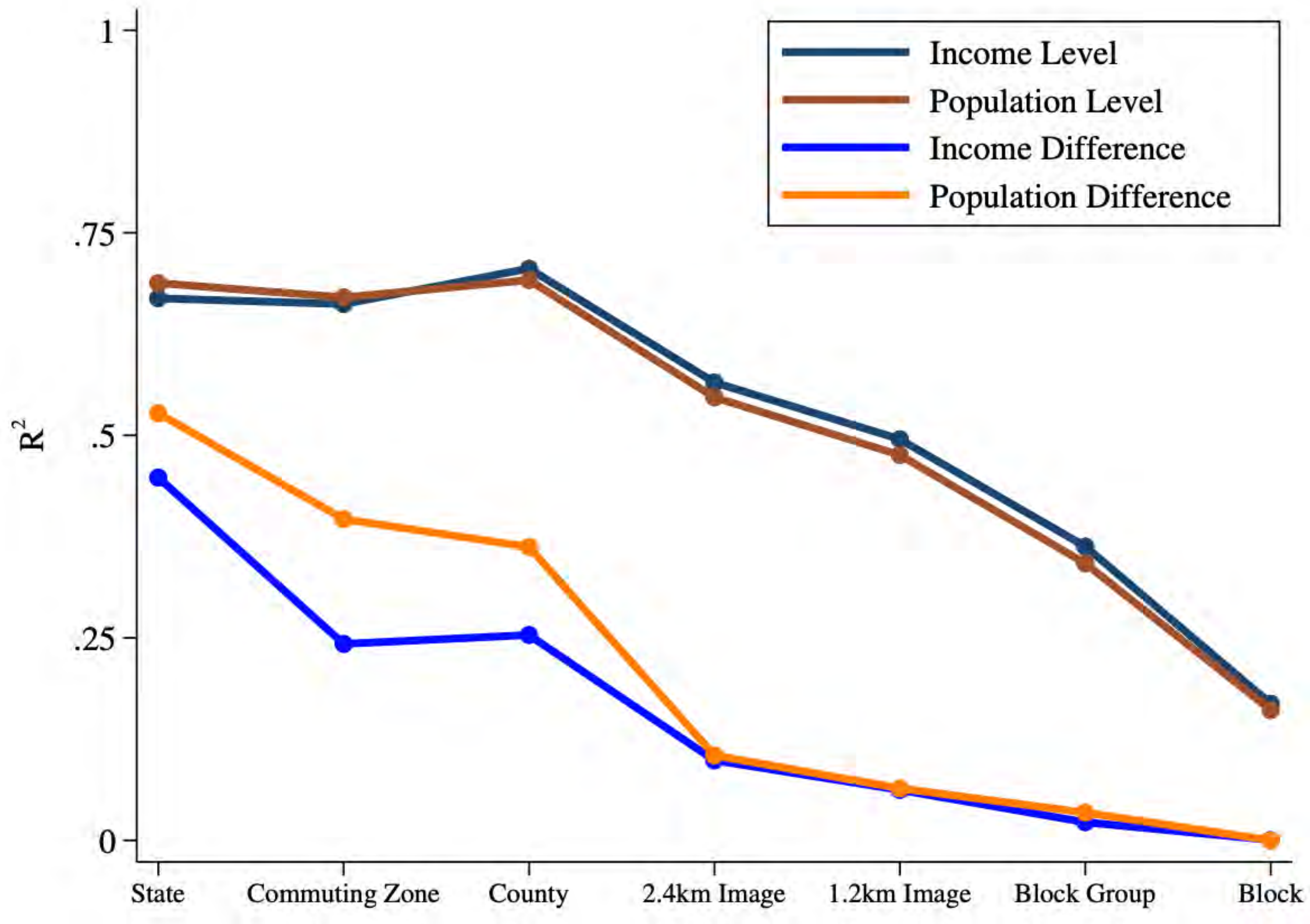


Jean et al. 2016



Ratlegde et al. 2022

Predicting population and income in the U.S.



Khachiyan et al. 2022

2. South Korea

Difference between South and North Korea

South Korea:

High quality economic data

North Korea:

Even pop. census not available

Question

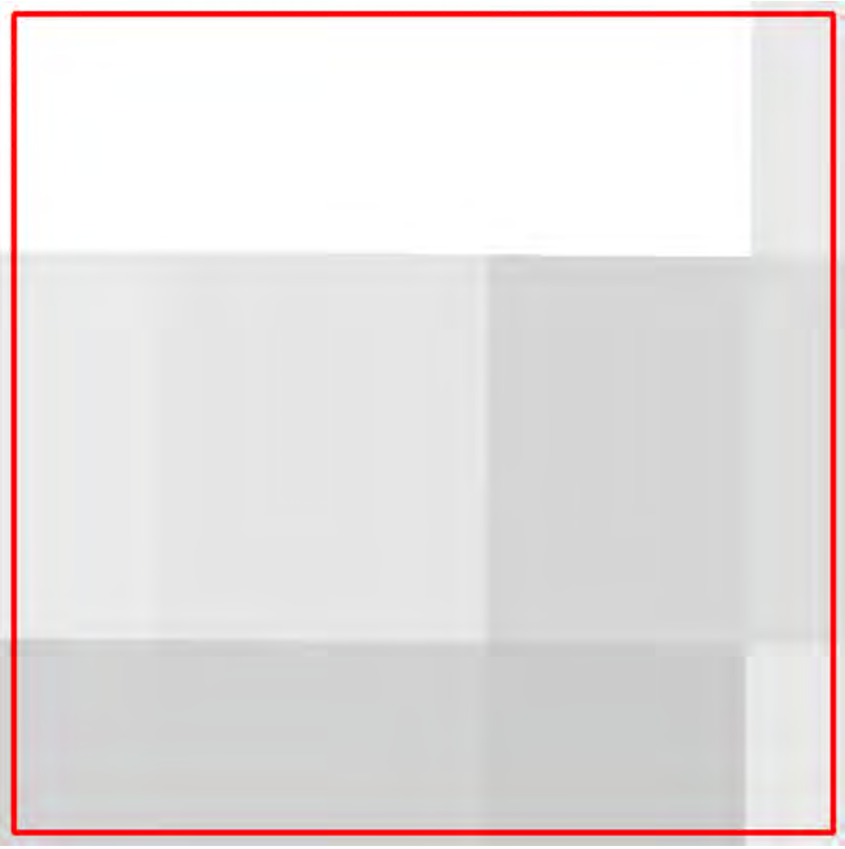
What economic statistics can we reliably extract from remote sensing data?

This measurement paper

First paper to compare three different spatial input data for predicting economics statistics

Rely on high quality ground truth data from Korea

Sentinel 2 OpenStreetMap VIIRS Night Light



This project is supported by ESA Network of Resources Initiative

Summary of results (South Korea)

Remote sensing data contain information to predict:

- population
- **total** income, consumption

But not could NOT predict:

- **average** income, consumption, credit scores
- median income

Summary of results (South Korea)

Publicly available daytime satellite images provide the best prediction performance

… compared to web maps or nightlight satellite images

Publicly available input data in 2019

- **Sentinel 2A** by European Space Agency
- **OpenStreetMap**
- **Visible Infrared Imaging Radiometer Suite** (VIIRS) by NASA/NOAA

* Sentinel and VIIRS data are annual median composite (months excluding winter: 4/1 – 11/1)

This project is supported by ESA Network of Resources Initiative

Ground truth data in 2019

1km x 1km grid-level

Population from
National Geographic Information Institute

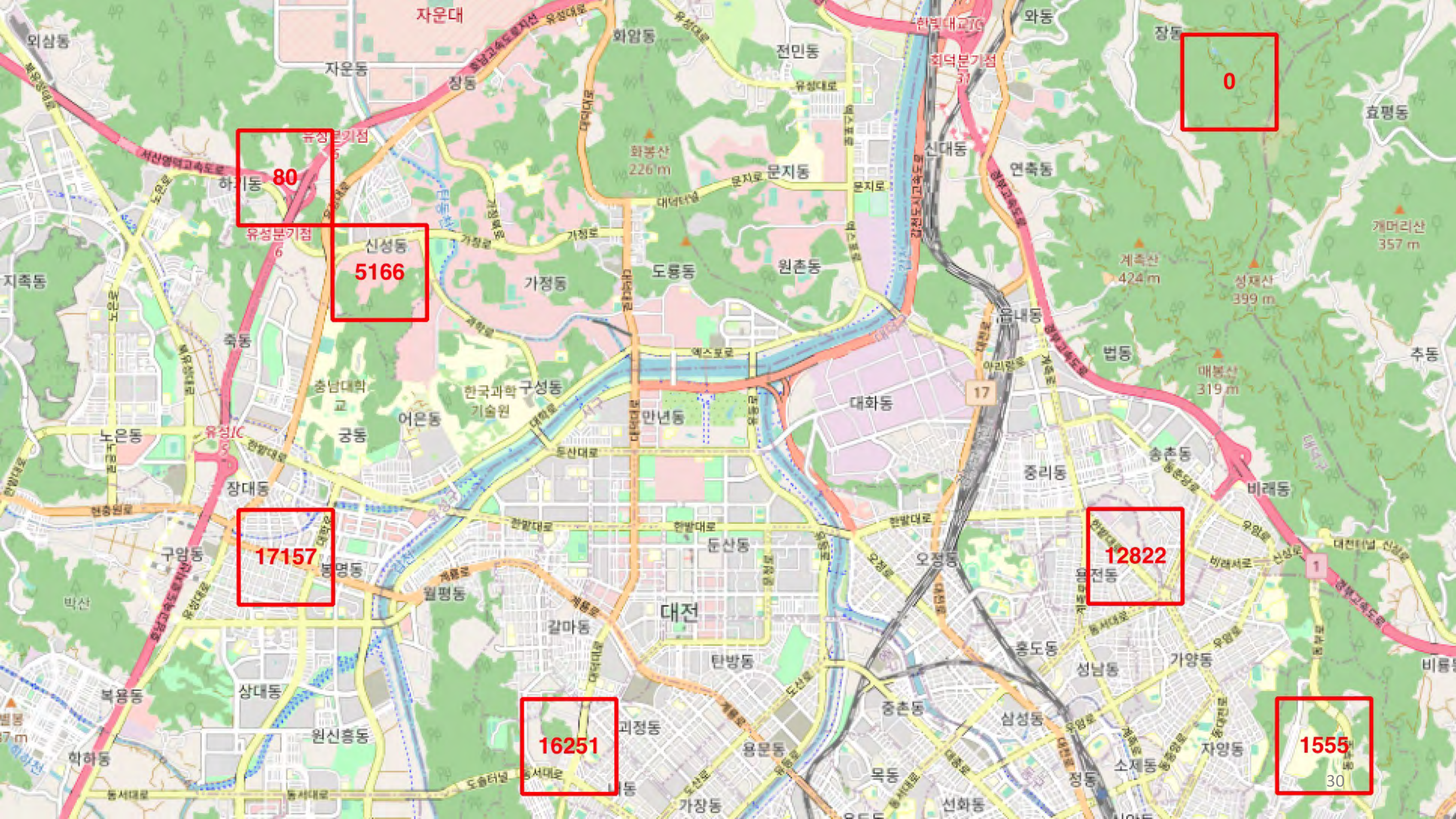
Income, card consumption, credit scores from
Korea Credit Bureau

Use 5% random sample

1kmx1km grids

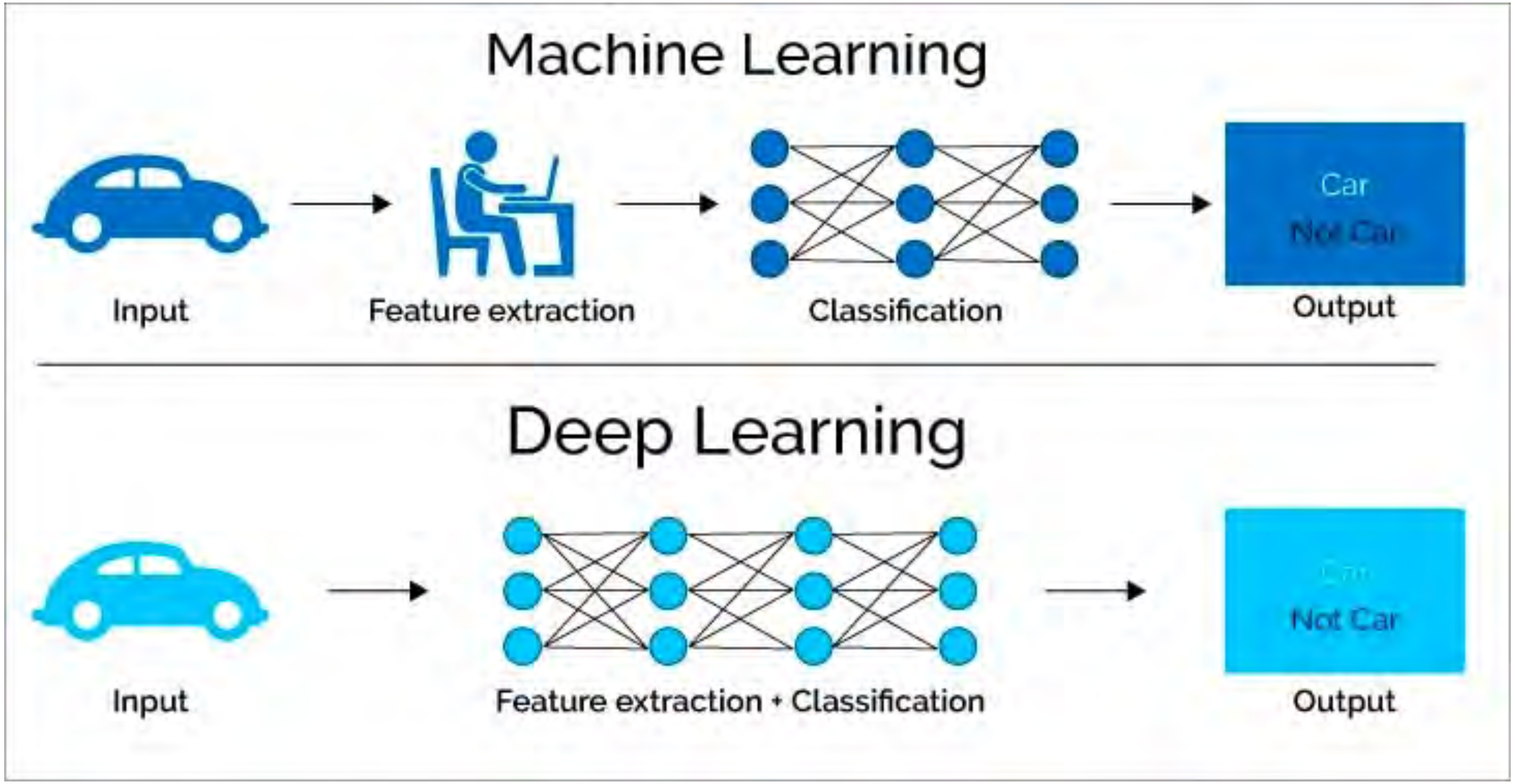
(5,353 grids)


0
80
5166
17157
12822
16251
1555


0
80
5166
17157
12822
16251
1555
대전
자운대
화암동
전민동
장동
와동
효평동
자운동
외삼동
하기동
신성동
가정동
도룡동
원촌동
문지동
신대동
연축동
화봉산
226 m
계족산
424 m
성재산
399 m
계머리산
357 m
매봉산
319 m
읍내동
법동
추동
지족동
죽동
충남대학교
한국과학기술원
구성동
어은동
궁동
만년동
대화동
노은동
유성IC
장대동
구암동
봉명동
월평동
둔산동
오정동
중리동
송촌동
비래동
용전동
갈마동
탄방동
홍도동
성남동
가양동
박산
복용동
상대동
원신흥동
학하동
괴정동
용문동
중촌동
삼성동
소제동
자양동
목동
정동
가장동
선화동
유성분기점
회덕분기점
한빛대교IC


0
80
5166
17157
12822
16251
1555

Computer vision deep learning model



Computer vision model used

CNN with pretrained network (VGG16)

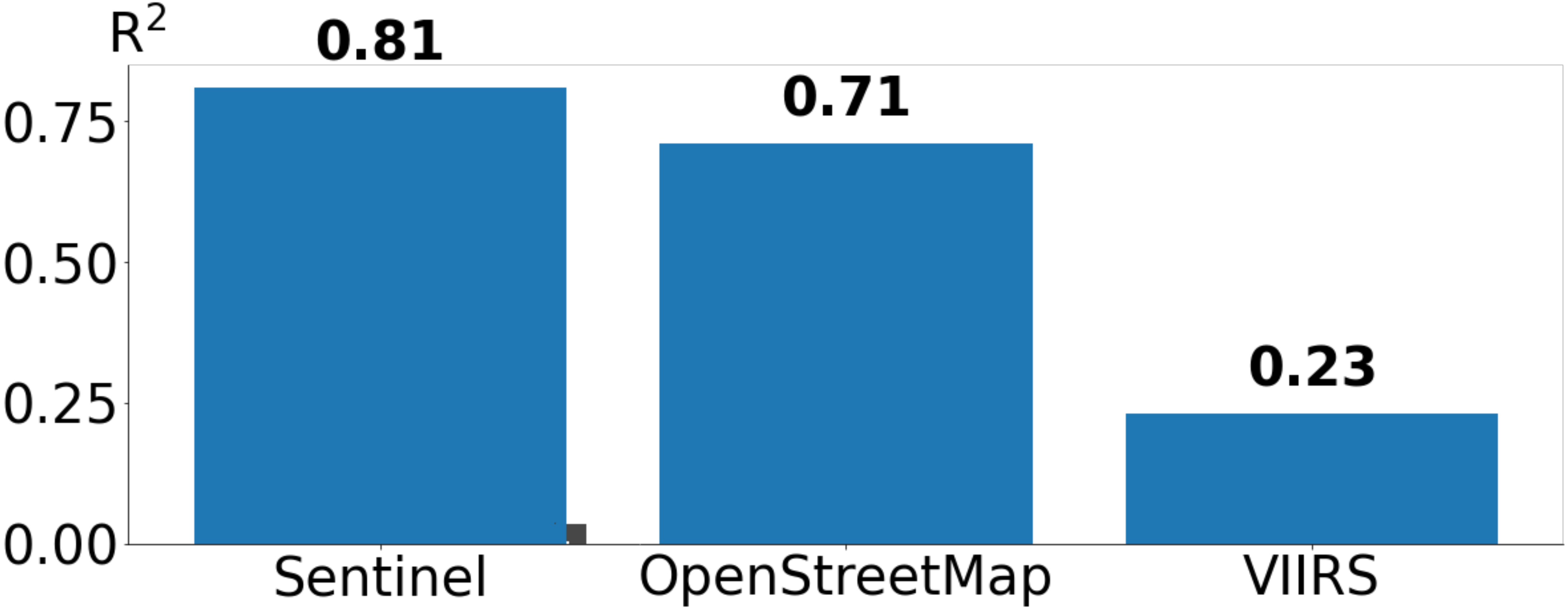
Image augmentation (random flips)

Random split of train and validation sets (8:2)

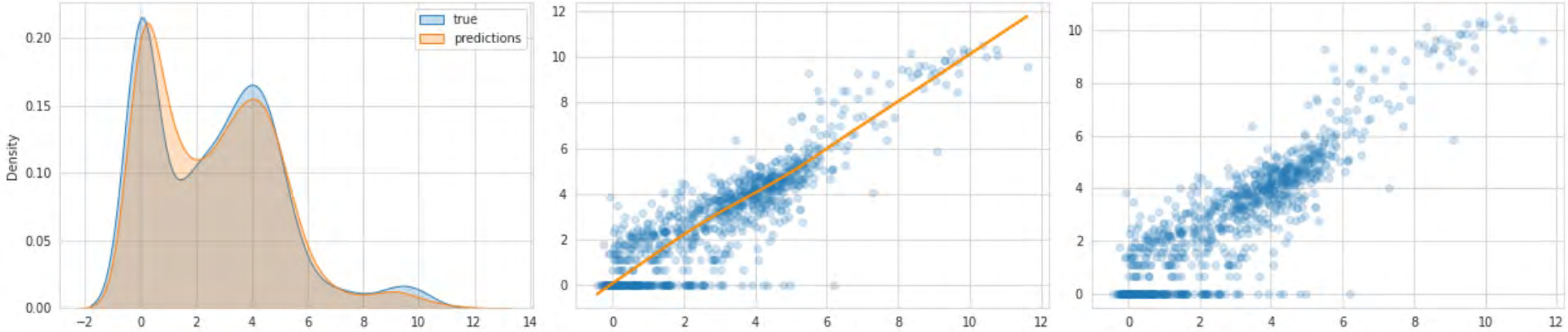
Trained on two NVIDIA RTX 6000s

Predicting population

Predicting log(**population**)

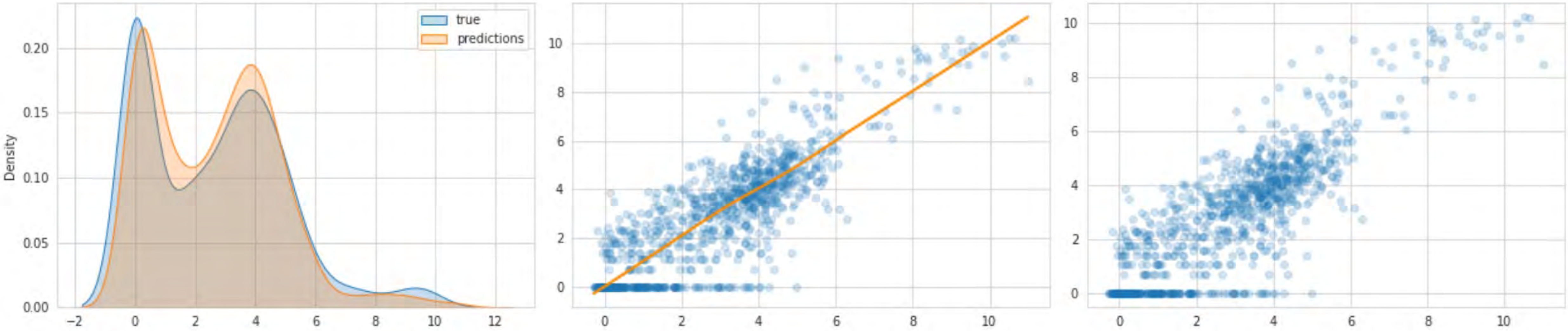


log(**population**) using Sentinel



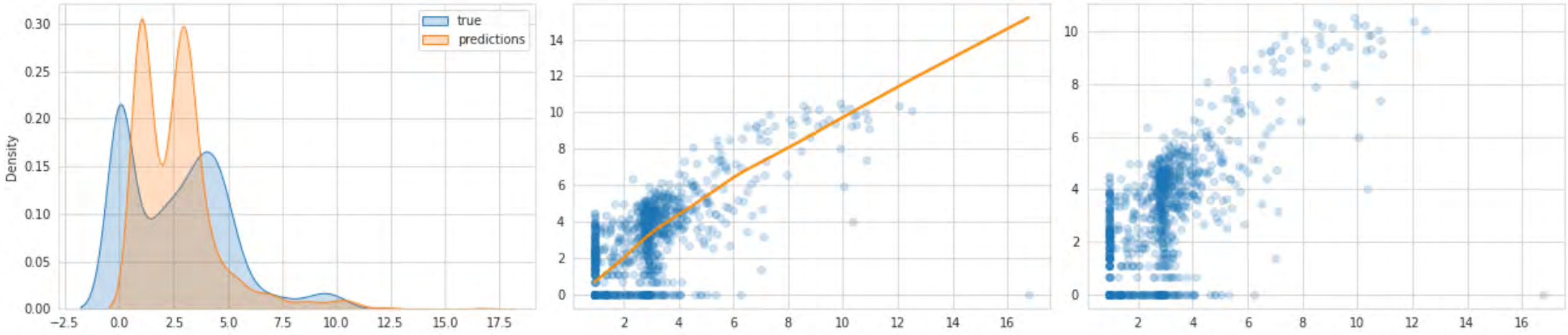

```
====================================
 - CNN v2 R2 (by lin. reg.)  : 0.826
 - CNN v2 R2 (by sklearn)  : 0.806
 - CNN v2 RMSE: 1.012
====================================
```

log(**population**) using OpenStreetMap




```
====================================
 - CNN v2 R2 (by lin. reg.)   : 0.766
 - CNN v2 R2 (by sklearn)   : 0.713
 - CNN v2 RMSE: 1.148
====================================
```

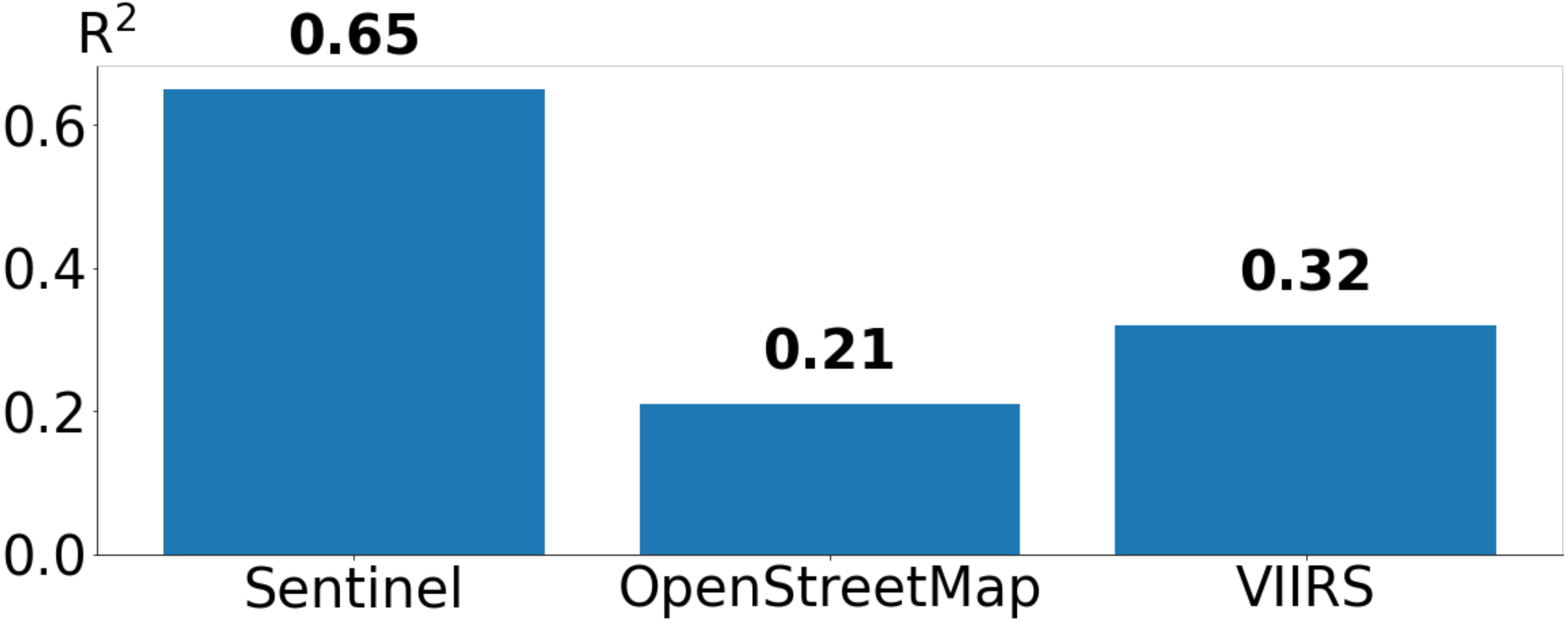
log(population) using VIIRS




```
===================================
 - CNN v2 R2 (by lin. reg.)  : 0.520
 - CNN v2 R2 (by sklearn)  : 0.228
 - CNN v2 RMSE: 1.678
===================================
```

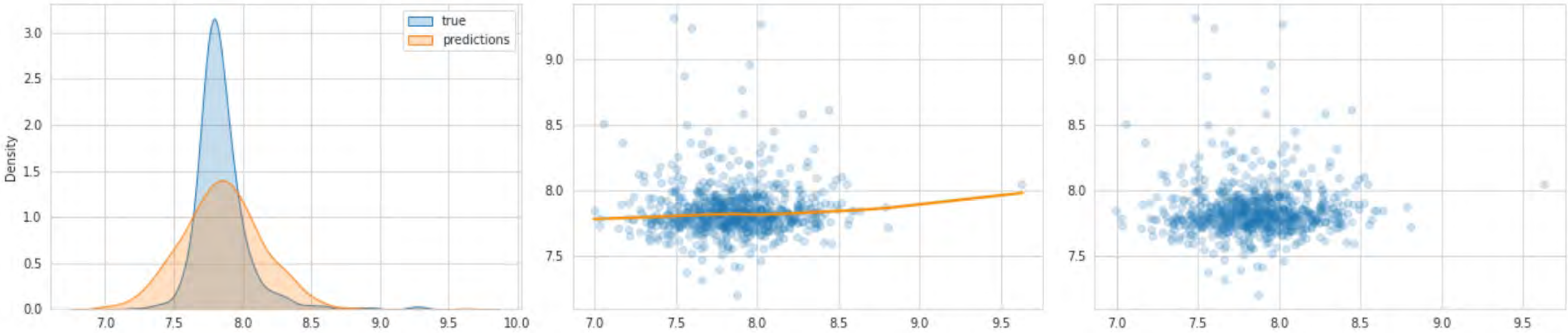
Predicting total income

Predicting log(**total income**)



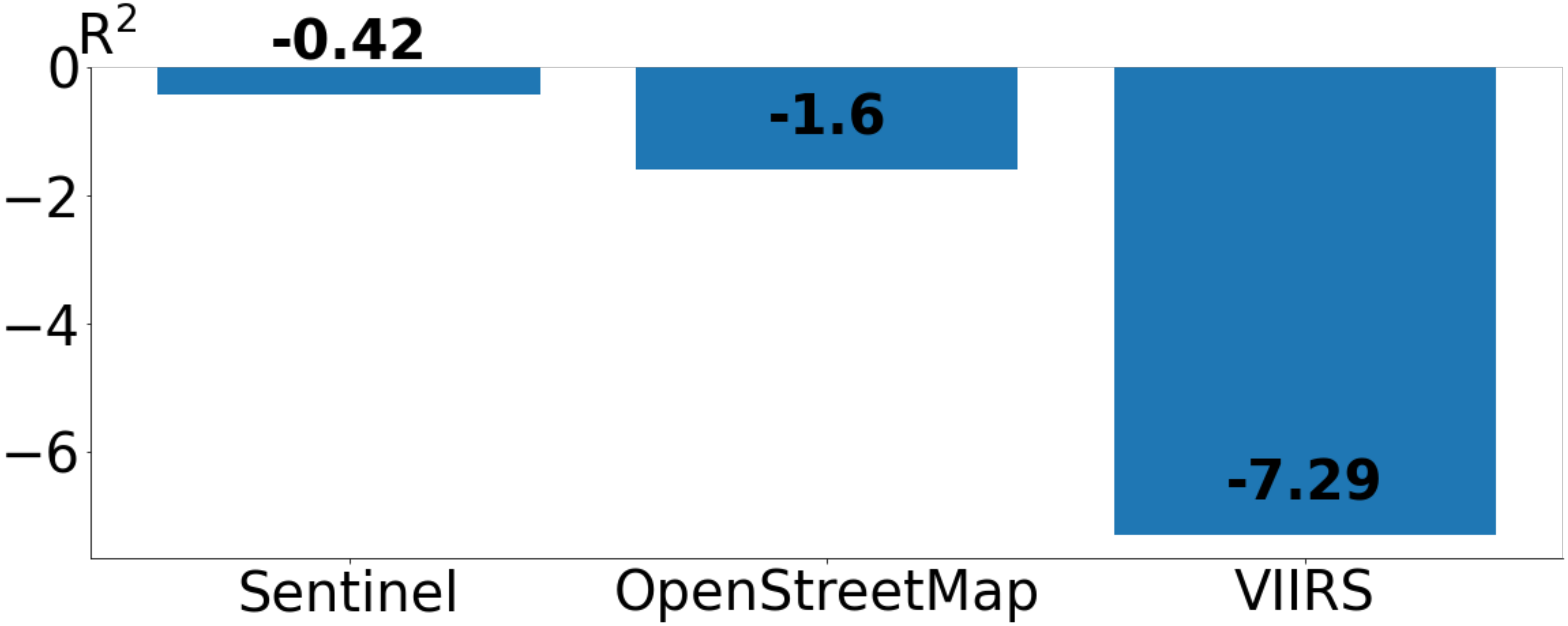
Predicting average income

log(average income) using Sentinel




```
====================================
 - CNN v2 R2 (by lin. reg.)  : 0.001
 - CNN v2 R2 (by sklearn)  : -0.417
 - CNN v2 RMSE: 0.355
====================================
```

Predicting log(average income)



Discussions

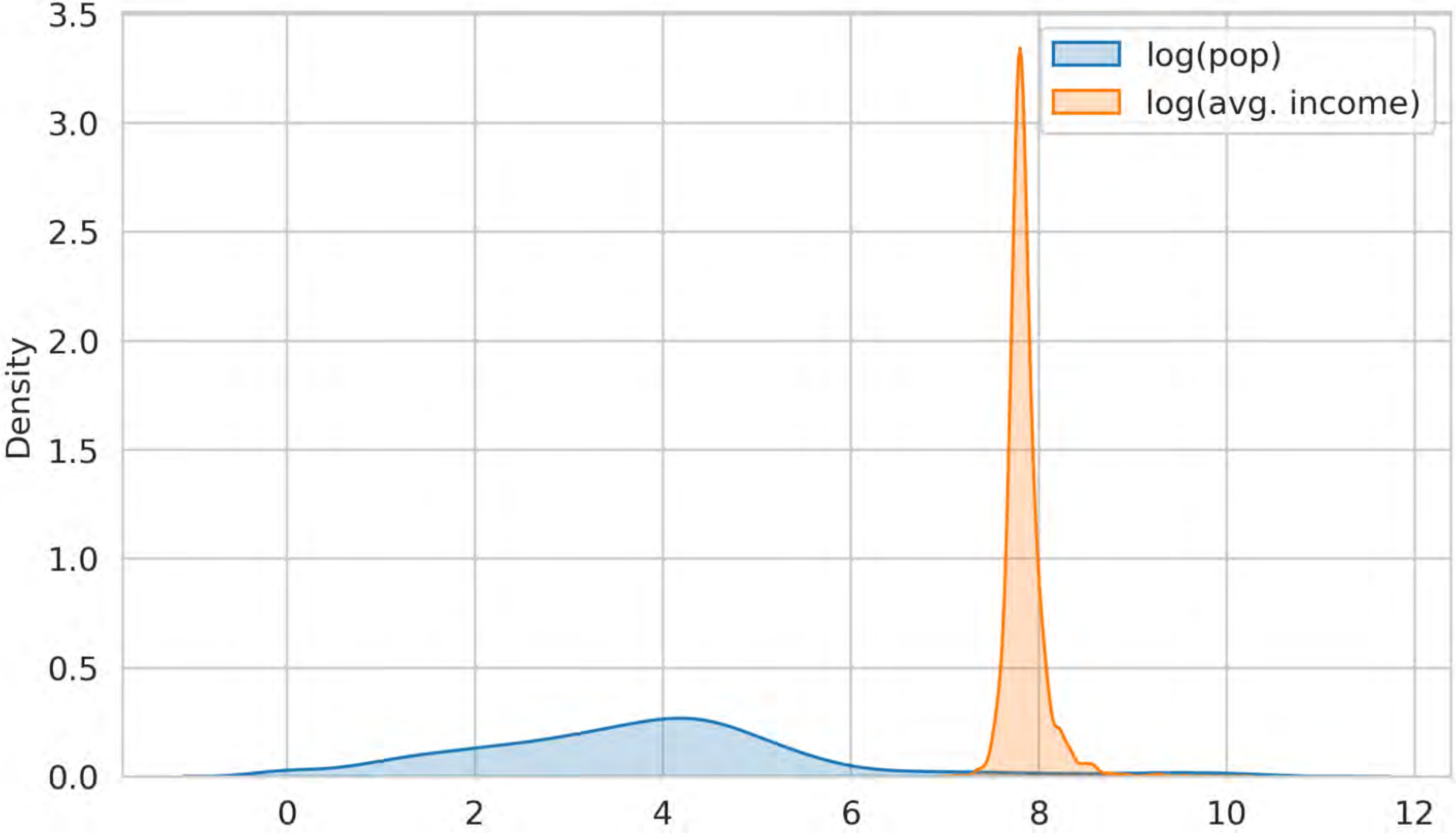
Remote sensing data can predict “quantity” (i.e. population) well.

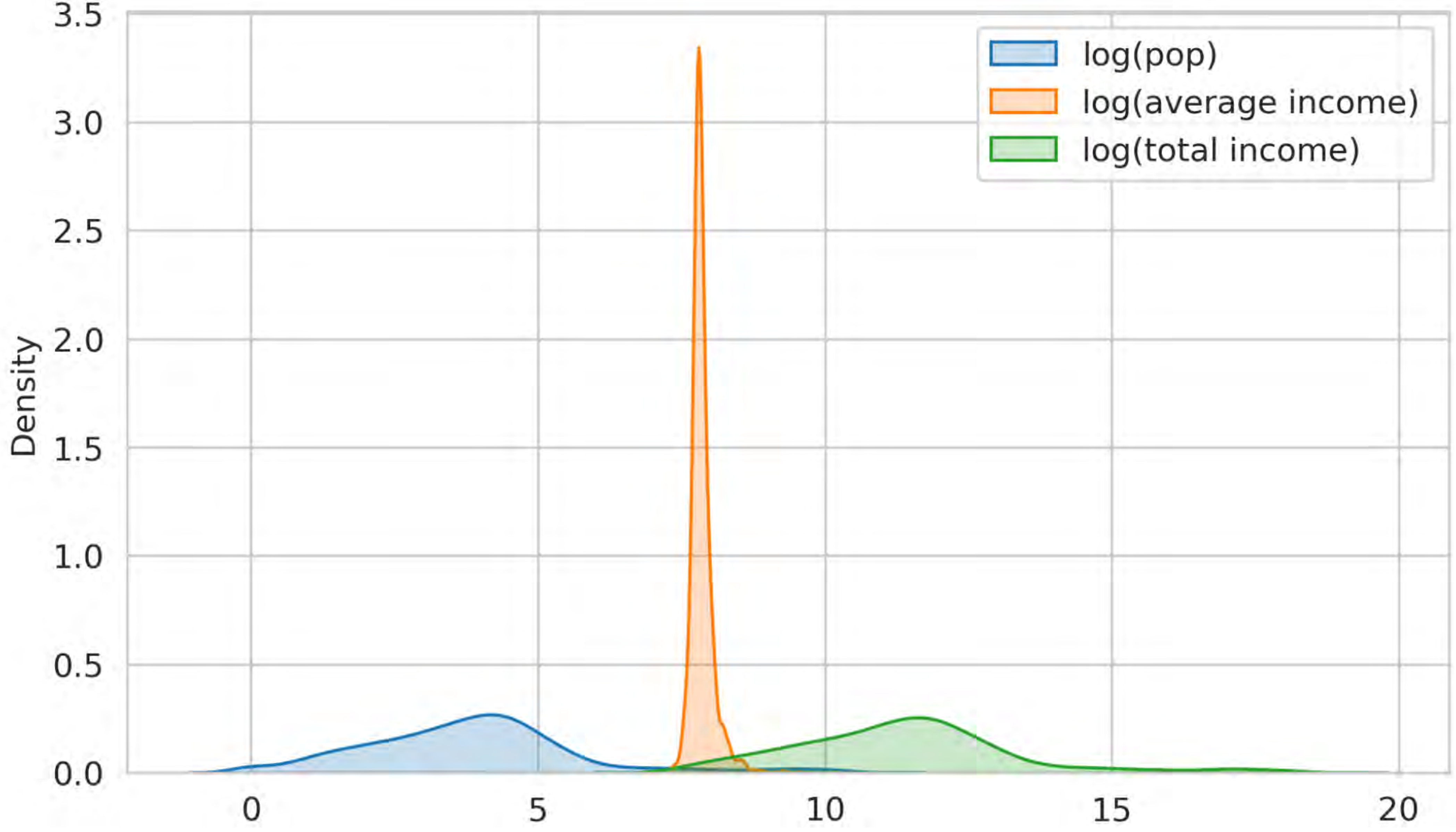
“Quality” measures (e.g., avg. income, consumption) may require additional inputs

Average income v.s. total income

$$\log(\text{total income}) = \log(\text{population} \times \text{avg. income})$$

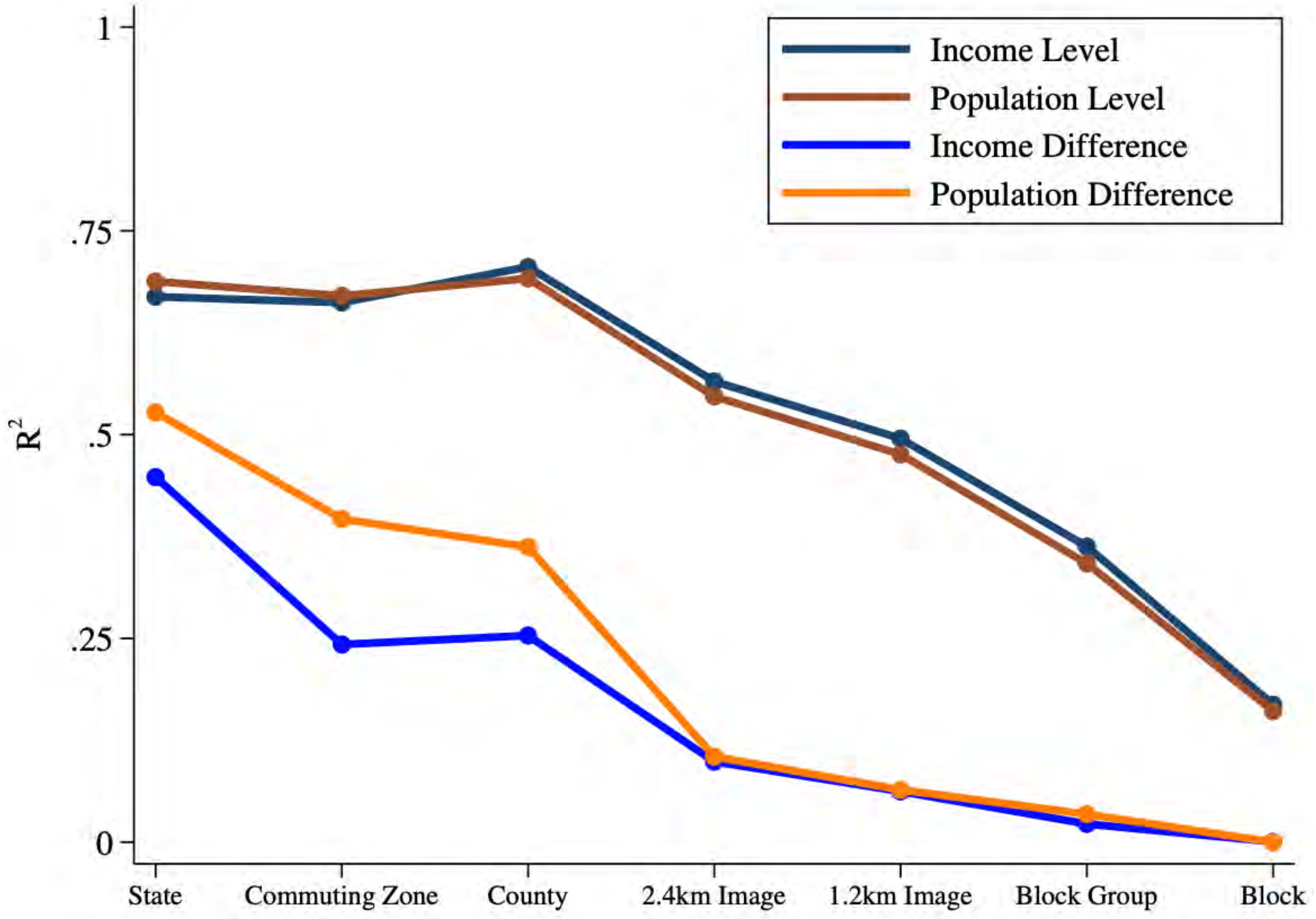
$$= \boxed{\log(\text{pop})} + \log(\text{avg. income})$$


log(pop)
log(avg. income)
Density
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
12



log(pop)
log(average income)
log(total income)
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
5
10
15
20

from Manson et al. (2020). From each sample, we use population by Census Block and total personal income, for residents ages 15 years and older, by Census Block Group.[7] Because income data are only published at the Block Group level, we interpolate income



Khachiyan et al. 2022

Conclusions

- Remote sensing data have potential to be useful
- Predicting economic statistics
- Acknowledge limitations
- Augment with other sources of information