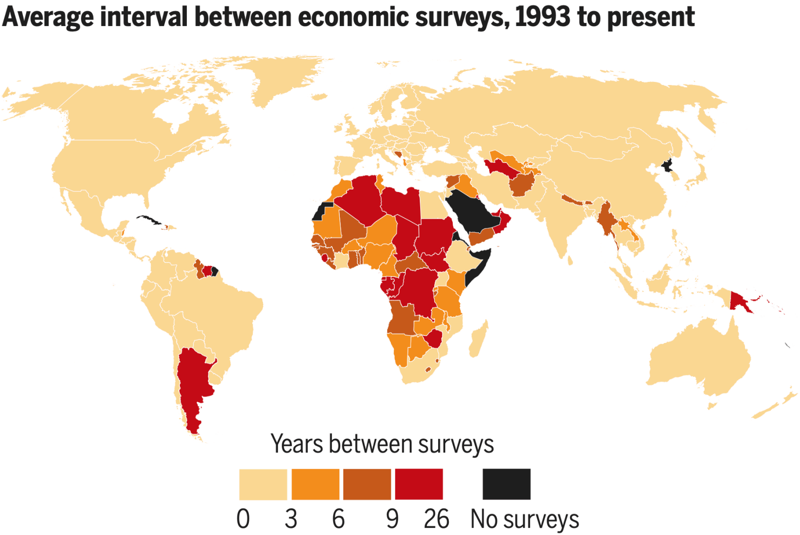
Multitask Observation using Satellite Imagery and Kitchen Sinks (MOSAIKS)

Togo Data Lab Training – UCSB, CEGA, & emLab
January, 2025

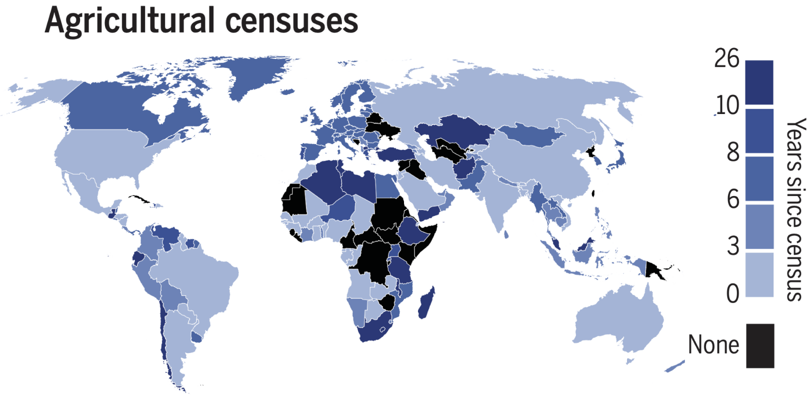
Tamma Carleton (UC Berkeley & emLab)
in collaboration with: Esther Rolf, Jonathan Proctor, Ian Bolliger, Vaishaal Shankar, Miyabi Isihara, Benjamin Recht, Solomon Hsiang

Data gaps: Economic progress



(Burke et al., *Science* 2021)

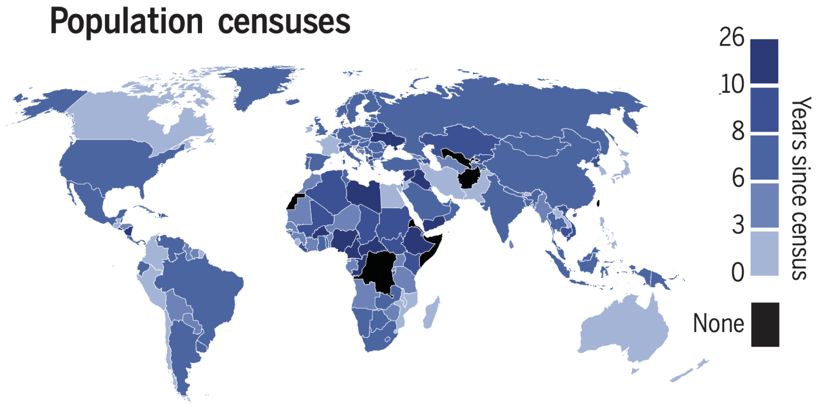
Data gaps: Agricultural losses and gains



⟶ 24% of countries have gone more than 15 years since their last agricultural census

(Burke et al., *Science* 2021)

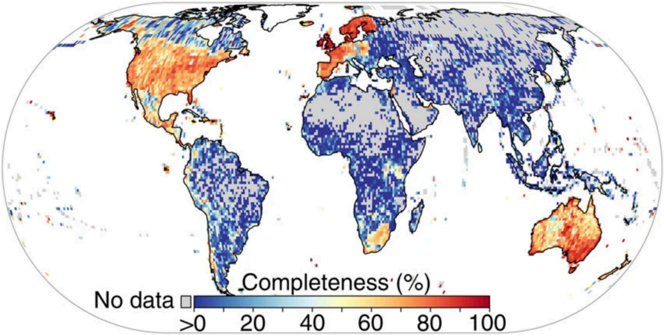
Data gaps: Demographics



⟶ 6% of countries have gone more than 15 years since their last population census

(Burke et al., *Science* 2021)

Data gaps: Biodiversity



⟶ 48% of Asian, 35% of African and 21% of South American cells have no digitally available species distribution data

(Meyer et al., *Nat. Comms.* 2015)

Disproportionate data gaps in disadvantaged communities

Traditional data collection is **expensive**

Disproportionate data gaps in disadvantaged communities

Traditional data collection is **expensive**

- One Demographic and Health Survey in one country for one year: $1.5-2 million USD (UN Sust. Dev., 2015)
- American Community Survey: >200 million USD annually (US Census Bureau, 2021)
- US Agricultural Census: $46 million USD (USDA, 2022)

Disproportionate data gaps in disadvantaged communities

Traditional data collection is **expensive**

- One Demographic and Health Survey in one country for one year: $1.5-2 million USD (UN Sust. Dev., 2015)
- American Community Survey: >200 million USD annually (US Census Bureau, 2021)
- US Agricultural Census: $46 million USD (USDA, 2022)

$\Longrightarrow$ Data gaps are largest where social and environmental challenges are most pressing

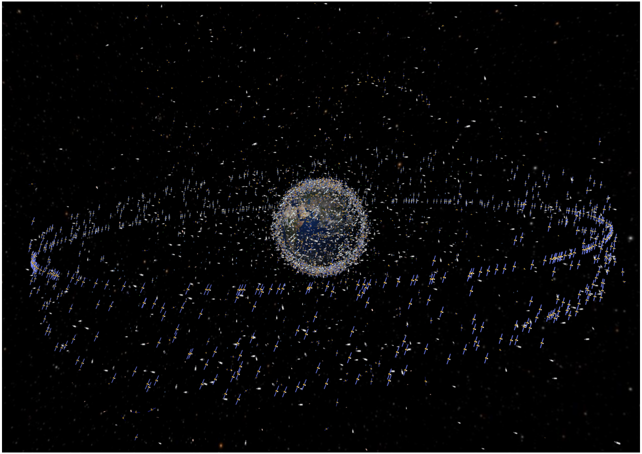
Data gaps impede social and environmental progress

"Key data are scarce, and often scarcest in places where they are most needed."
—Burke et al., (*Science*, 2021)

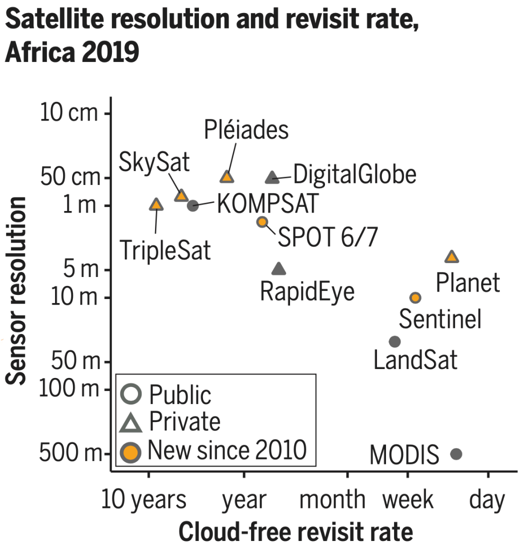
Satellite imagery: A global data solution?

There are **over 700 Earth observation satellites** in orbit.

Collectively, they retrieve >**100TB of data per day**.



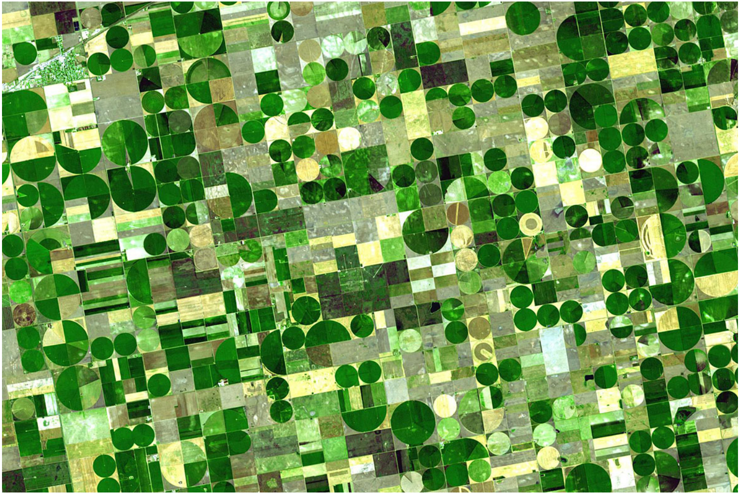
Satellite imagery: A global data solution?



(Burke et al., *Science* 2021)

Satellite imagery: A global data solution?

Measuring irrigation and crop yields



Source: NASA

Satellite imagery: A global data solution?

Measuring urbanization and economic growth

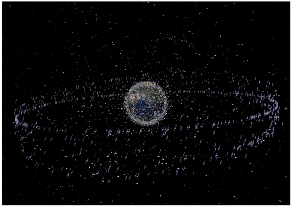
Source: Medium

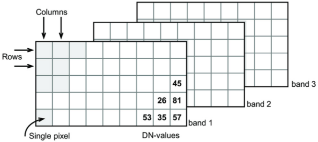
Satellite imagery: A global data solution?

Measuring the extent of flooding

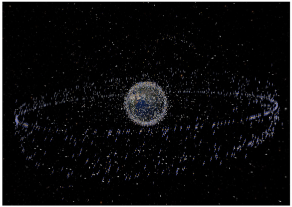
Source: GRiF

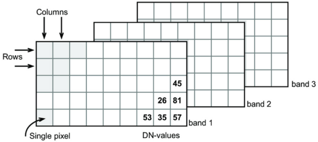
The challenge: A fire hose of unstructured information





The challenge: A fire hose of unstructured information





How do we transform unstructured pixel-level data into structured and useful information?

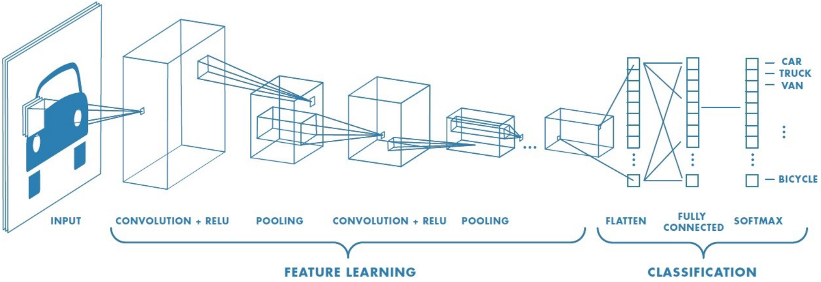
Emerging solution: Deep learning

Question: How do we transform unstructured pixel-level data into structured and useful information?

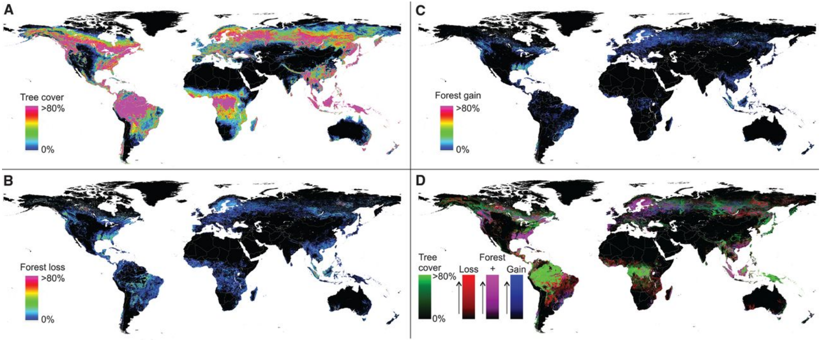
Emerging solution: Deep learning

Question: How do we transform <u>unstructured</u> pixel-level data into <u>structured</u> and useful information?

Modern answer: Deep learning (i.e., machine learning with artificial neural networks)

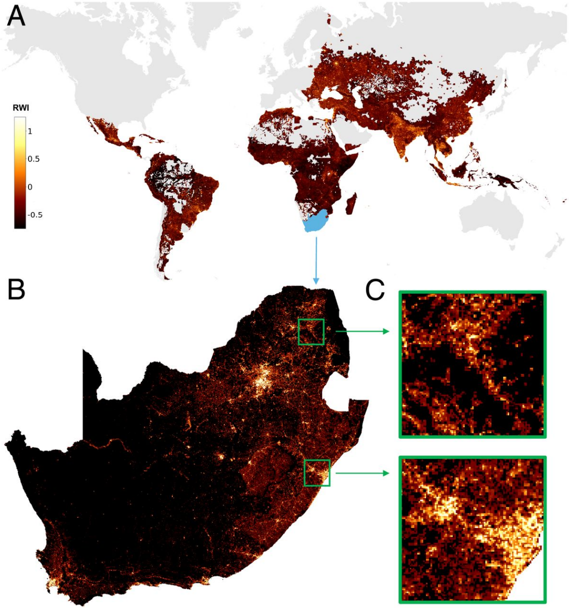


Satellite + ML measures of forest cover



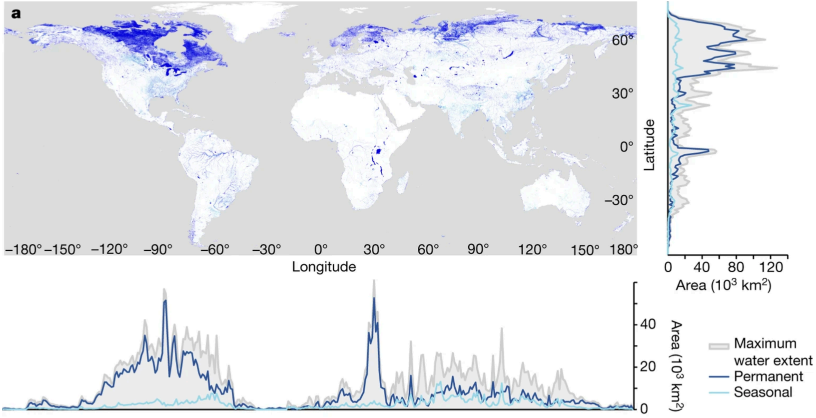
(Hansen et al., 2013)

Satellite + ML measures of wealth



(Chi et al., 2022)

Satellite + ML measures of surface water



(Pekel et al., 2016)

Each measurement is a major research enterprise

- Measuring Economic Growth from Outer Space
  Henderson et al (*AER*, 2012)
- High resolution Global Maps of 21st Century Forest Cover Change
  Hanson et al (*Science*, 2013)
- Combining satellite imagery and machine learning to predict poverty
  Jean et al (*Science*, 2016)
- Mapping local variation in educational attainment across Africa
  Graetz et al (*Nature*, 2018)
- Mapping child growth failure in Africa between 2000 and 2015
  Osgood-Zimmerman et al (*Nature*, 2018)
- Using publicly available satellite imagery and deep learning to understand economic well-being in Africa
  Yeh et al (*Nature Comm.* 2020)

Barriers to entry prevent widespread use of satellite imagery

Many people would like to combine Satellite Imagery with Machine Learning (SIML) to solve their own problem in a specific setting (domain).

Barriers to entry prevent widespread use of satellite imagery

Many people would like to combine **Satellite Imagery with Machine Learning (SIML)** to solve **their own problem in a specific setting (domain)**.

Limited access to data, compute, skills, and resources prevents most researchers and decision-makers from using SIML to tackle local and global challenges.

Barriers to entry prevent widespread use of satellite imagery

Many people would like to combine **Satellite Imagery with Machine Learning (SIML)** to solve **their own problem in a specific setting (domain)**.

Limited access to data, compute, skills, and resources prevents most researchers and decision-makers from using SIML to tackle local and global challenges.

These barriers imply that most remote sensing is conducted in developed countries (Yu et al. 2014, Haack & Ryerson 2016)

Can we make high performance SIML widely accessible?

We're developing a system that:

1. Makes SIML **easy** (a regression) and **cheap** (can be done on a personal computer)

2. Achieves **performance** competitive with leading models

3. Characterizes **uncertainty** and **sensitivity** of performance

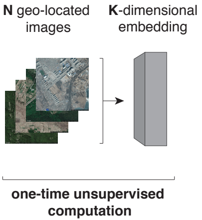
$\longrightarrow$ Problem setting: predicting properties of small regions (e.g., population density) using high-resolution satellite imagery

We hope this system will help empower diverse researchers to leverage SIML to solve their own domain-specific challenges

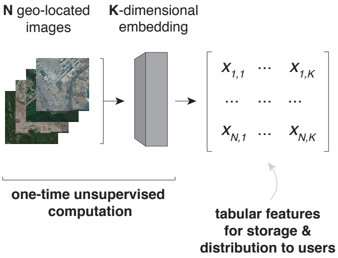
MOSAIKS: A generalizable pipeline to improve access

("Multi-task Observation using Satellite Imagery and Kitchen Sinks")

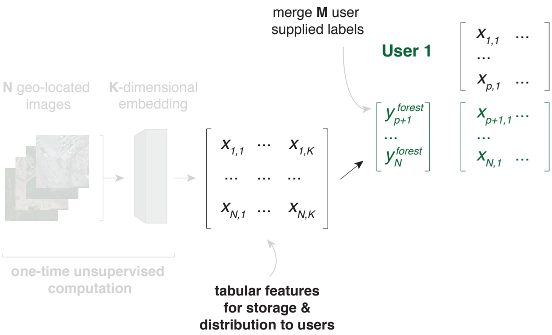
MOSAIKS: A generalizable pipeline to improve access



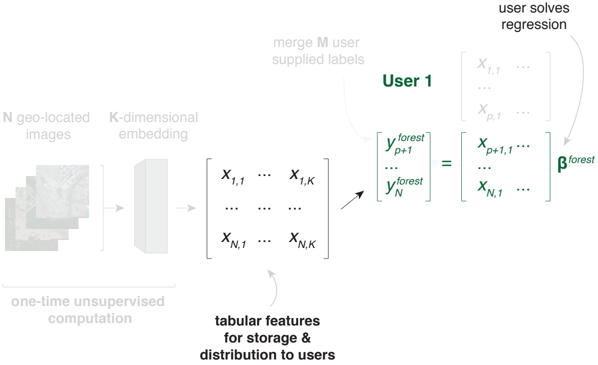
MOSAIKS: A generalizable pipeline to improve access



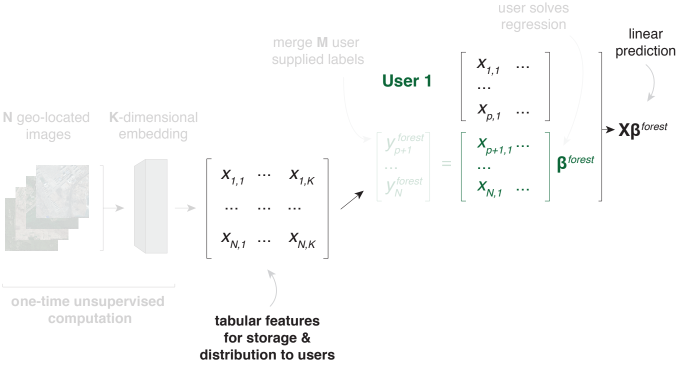
MOSAIKS: A generalizable pipeline to improve access



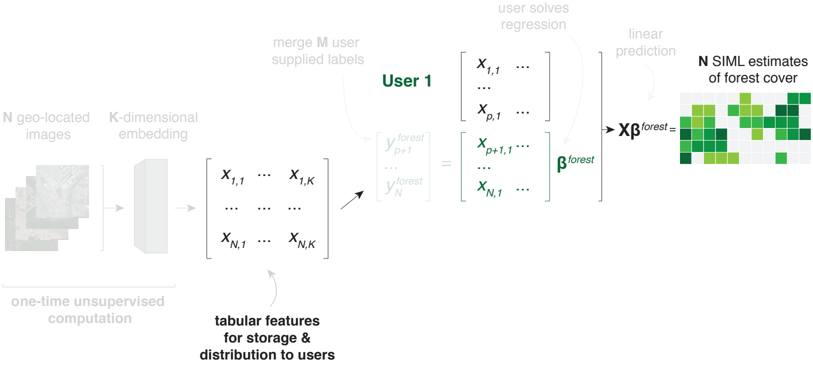
MOSAIKS: A generalizable pipeline to improve access



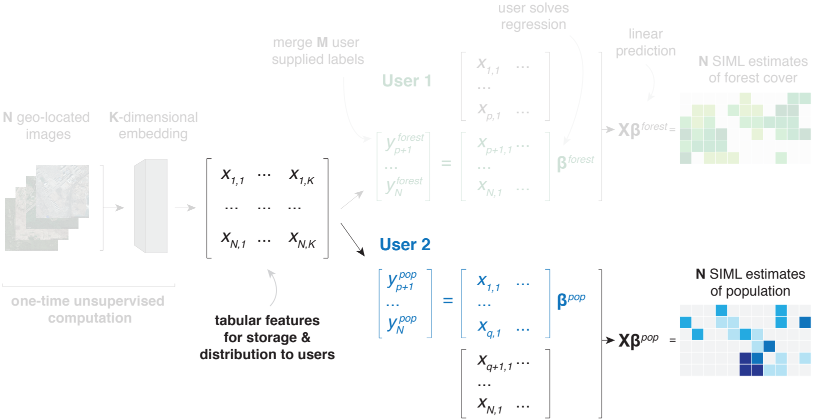
MOSAIKS: A generalizable pipeline to improve access



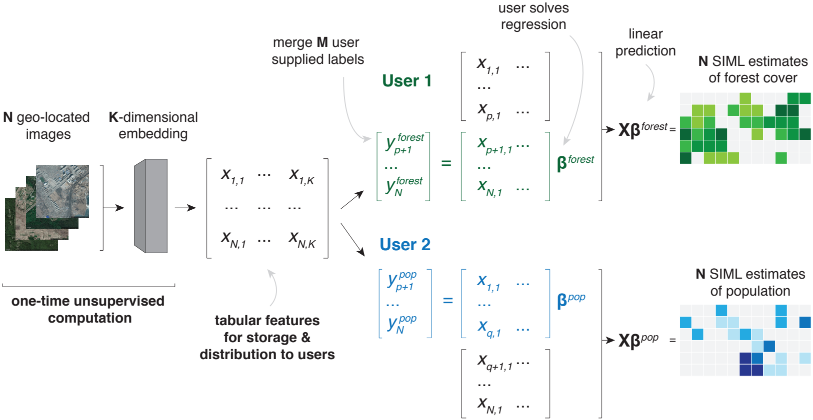
MOSAIKS: A generalizable pipeline to improve access



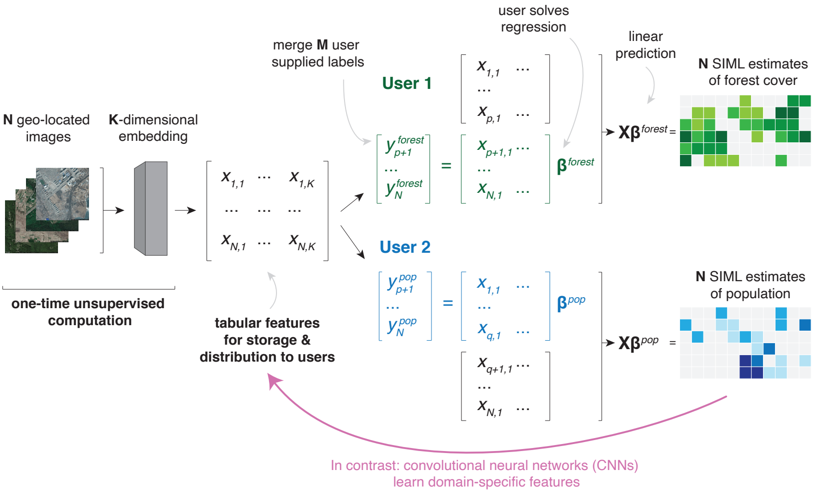
MOSAIKS: A generalizable pipeline to improve access



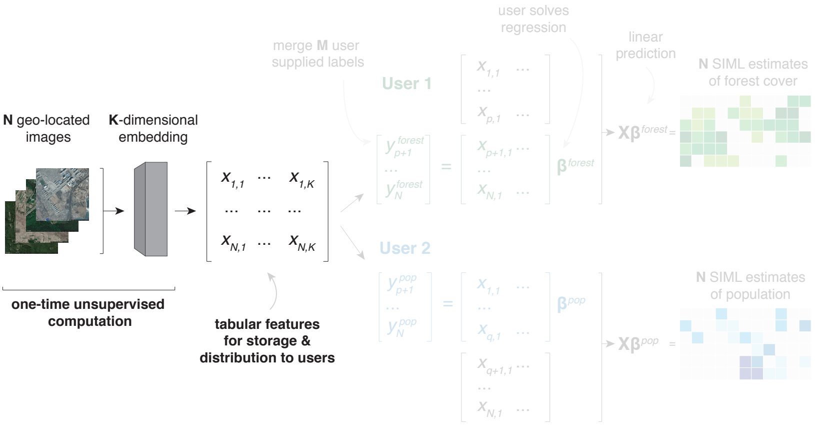
MOSAIKS: A generalizable pipeline to improve access



MOSAIKS: A generalizable pipeline to improve access



MOSAIKS: A generalizable pipeline to improve access



Needs a single summary of satellite imagery that can, without modification, accurately predict many different ground conditions.

Research question

Many unsupervised featurization approaches exist:

- GIST descriptor
  (Oliva & Torralba, Int. J. Comput. Vis., 2001)
- Scale-invariant feature transform (SIFT) descriptor
  (Lowe, Int. J. Comput. Vis., 2004)
- Histogram of oriented gradients (HOG) descriptor
  (Dalal & Triggs, Int. Conf. Comput. Vis. Pattern Recognit., 2005)
- Autoencoder
  (Hinton & Salakhutdinov, Science, 2006)
- Using pre-trained CNN as featurizer
  (Gu et al., Applied Sciences, 2019)
- Tile2Vec
  (Jean et al., AAAI, 2019)

Yet: few demonstrations that a single set of features can performance as well as deep-learning methods across multiple tasks.

Research question

Many unsupervised featurization approaches exist:

- GIST descriptor
  (Oliva & Torralba, Int. J. Comput. Vis., 2001)
- Scale-invariant feature transform (SIFT) descriptor
  (Lowe, Int. J. Comput. Vis., 2004)
- Histogram of oriented gradients (HOG) descriptor
  (Dalal & Triggs, Int. Conf. Comput. Vis. Pattern Recognit., 2005)
- Autoencoder
  (Hinton & Salakhutdinov, Science, 2006)
- Using pre-trained CNN as featurizer
  (Gu et al., Applied Sciences, 2019)
- Tile2Vec
  (Jean et al., AAAI, 2019)

Research Question

Can a single set of features achieve state of the art performance across a variety of SIML tasks?

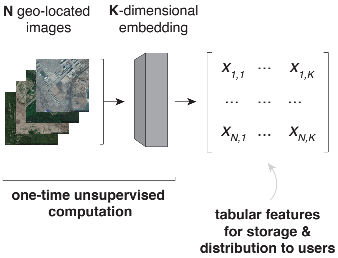
Proposed solution: Random convolutional features

Method: Rahimi & Recht (2007, 2008a,b)

1. **Key insight:** Replacing costly optimization with randomization saves time and maintains performance
2. **How?** Embed data into a high-dimensional randomly-generated feature space, run linear regression

- **Prior performance**:
    - classifying photographs (*Coates & Ng 2012*)
    - encoding genetic sequences (*Morrow et al. 2017*)
    - predicting solar flares (*Jonas et al. 2018*)
- **Speed**: replaces computationally expensive minimization with randomization (*Rahimi & Recht 2007, 2008a,b)*
- **Suitability:** to the structures of satellite imagery
    - Objects (e.g. tree, car) are generally within a few pixels.
    - Images taken from a constant distance, and (often) orthorectified

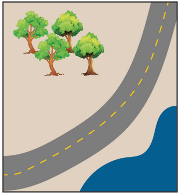
Intuition for random convolutional (kitchen sink) features



Intuition for random convolutional (kitchen sink) features

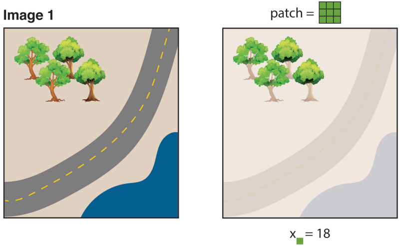
Rahimi & Recht (2007, 2008a,b)

Image 1



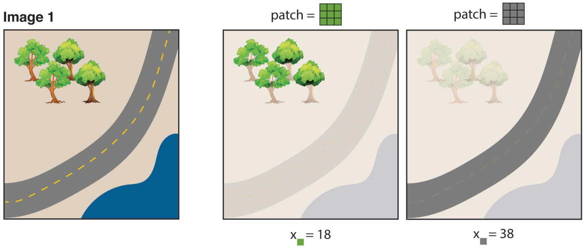
Intuition for random convolutional (kitchen sink) features

Rahimi & Recht (2007, 2008a,b)



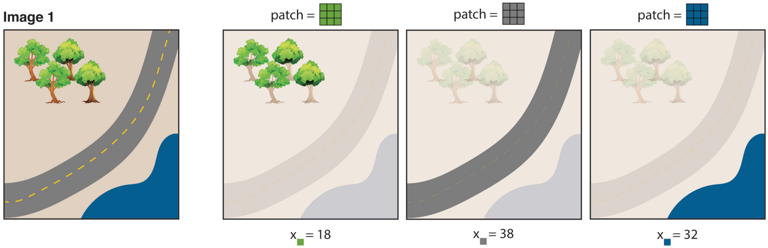
Intuition for random convolutional (kitchen sink) features

Rahimi & Recht (2007, 2008a,b)



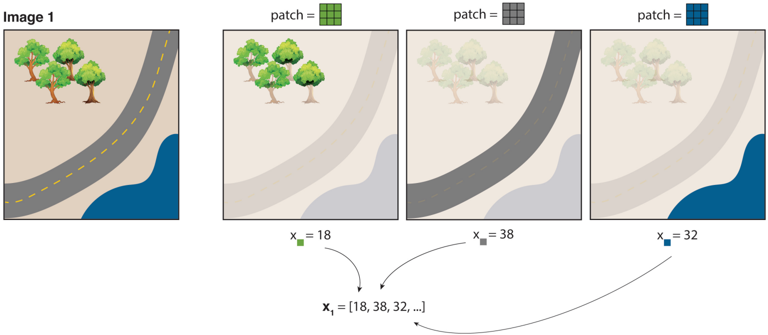
Intuition for random convolutional (kitchen sink) features

Rahimi & Recht (2007, 2008a,b)



Intuition for random convolutional (kitchen sink) features

Rahimi & Recht (2007, 2008a,b)



Intuition for random convolutional (kitchen sink) features

Rahimi & Recht (2007, 2008a,b)

Image 1

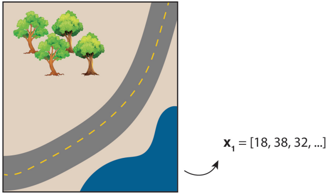
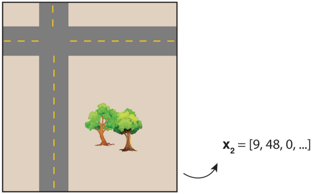


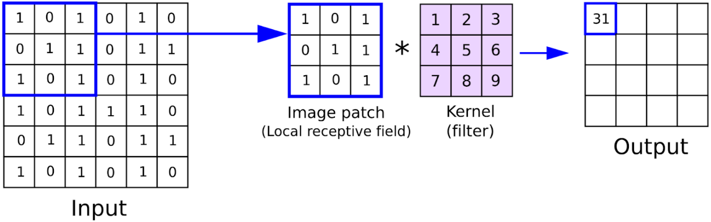
Image 2



Intuition for random convolutional (kitchen sink) features

What is a convolution?

A **convolution** is a mathematical operation comparing an image to a "filter" (here, patch). Measures the *similarity* of image and filter.



1*1 + 0*2 + 1*3 +

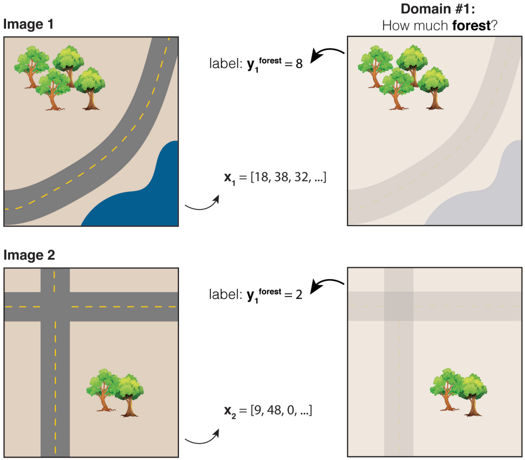
0*4 + 1*5 + 1*6 +

1*7 + 0*8 + 1*9 = 31

source: analyticsindiamag.com

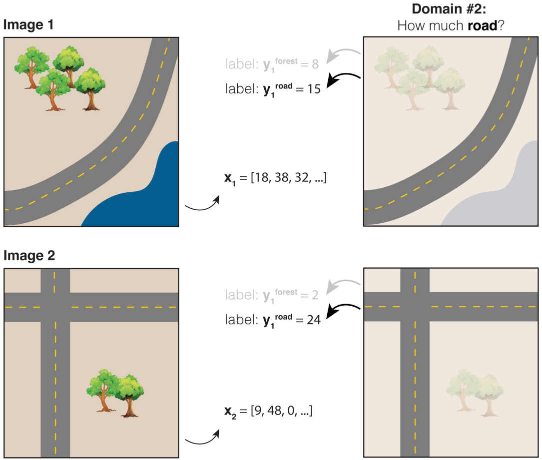
Intuition for random (kitchen sink) convolutional features

Rahimi & Recht (2007, 2008a,b)



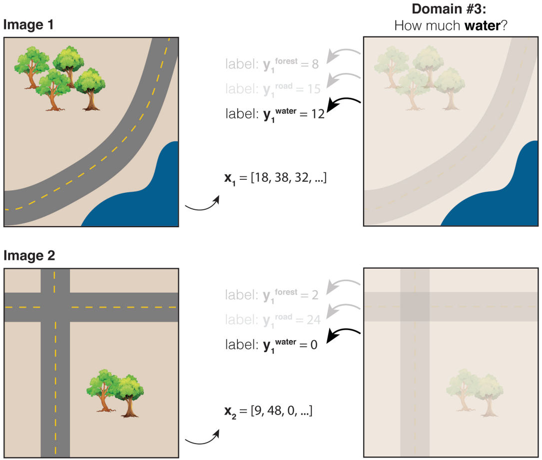
Intuition for random (kitchen sink) convolutional features

Rahimi & Recht (2007, 2008a,b)



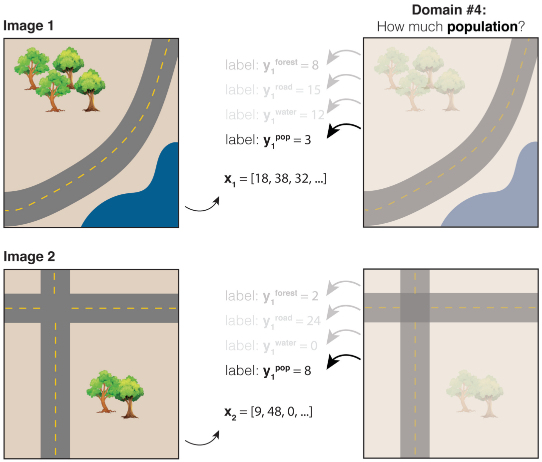
Intuition for random (kitchen sink) convolutional features

Rahimi & Recht (2007, 2008a,b)



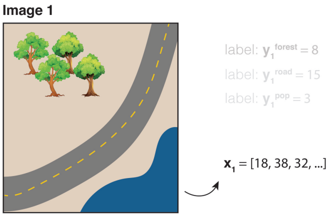
Intuition for random (kitchen sink) convolutional features

Rahimi & Recht (2007, 2008a,b)



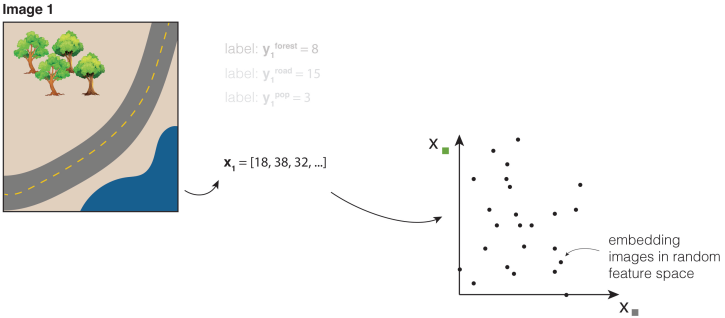
Intuition for random (kitchen sink) convolutional features

Rahimi & Recht (2007, 2008a,b)



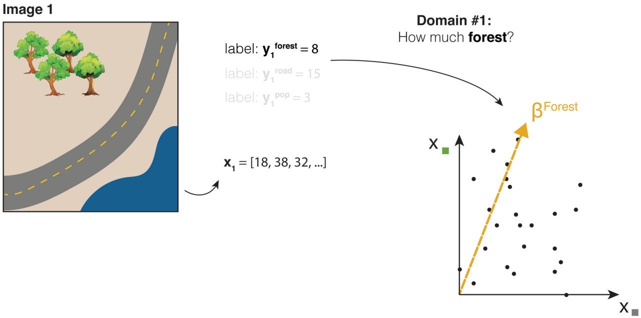
Intuition for random (kitchen sink) convolutional features

Rahimi & Recht (2007, 2008a,b)



Intuition for random (kitchen sink) convolutional features

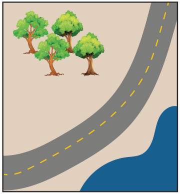
Rahimi & Recht (2007, 2008a,b)



Intuition for random (kitchen sink) convolutional features

Rahimi & Recht (2007, 2008a,b)

Image 1



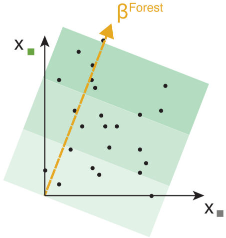
label: $\mathbf{y}_1^{\text{forest}} = 8$

label: $\mathbf{y}_1^{\text{road}} = 15$

label: $\mathbf{y}_1^{\text{pop}} = 3$

$\mathbf{x}_1 = [18, 38, 32, \ldots]$

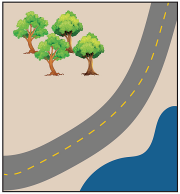
Domain #1:
How much **forest**?



Intuition for random (kitchen sink) convolutional features

Rahimi & Recht (2007, 2008a,b)

Image 1



label: $\mathbf{y}_1^{\text{forest}} = 8$

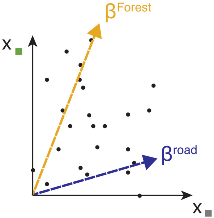
label: $\mathbf{y}_1^{\text{road}} = 15$

label: $\mathbf{y}_1^{\text{pop}} = 3$

$\mathbf{x}_1 = [18, 38, 32, \ldots]$

Domain #1:
How much **forest**?

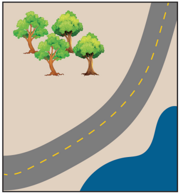
Domain #2:
How much **road**?



Intuition for random (kitchen sink) convolutional features

Rahimi & Recht (2007, 2008a,b)

Image 1



label: $\mathbf{y}_1^{forest} = 8$

label: $\mathbf{y}_1^{road} = 15$

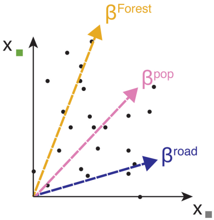
label: $\mathbf{y}_1^{pop} = 3$

$\mathbf{x}_1 = [18, 38, 32, \ldots]$

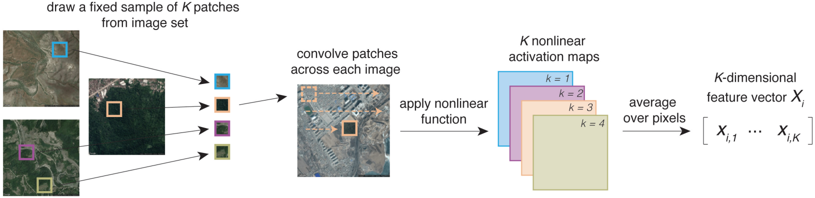
Domain #1: How much **forest**?

Domain #2: How much **road**?

Domain #4: How much **pop.**?



MOSAIKS: Applying RCF to satellite imagery



Roadmap: Rolf et al. (2021)

Can a single set of features achieve state of the art performance across a variety of tasks?

1. **Test generalization across tasks**, and compare performance and cost to existing SIML models
2. **Evaluate model sensitivity**, particularly under limited data/storage conditions
3. **Scale the analysis** globally and across the outcomes in a national survey
4. **Demonstrate additional properties** of MOSAIKS

Roadmap: Rolf et al. (2021)

Can a single set of features achieve state of the art performance across a variety of tasks?

1. **Test generalization across tasks**, and compare performance and cost to existing SIML models
2. **Evaluate model sensitivity**, particularly under limited data/storage conditions
3. **Scale the analysis** globally and across the outcomes in a national survey
4. **Demonstrate additional properties** of MOSAIKS

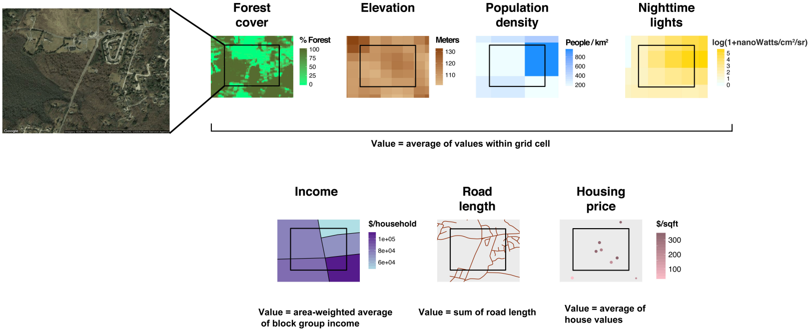
Experiment 1: test generalization across tasks

Step 1: Randomly sample 100,000 tiles (1km x 1km) from the U.S.

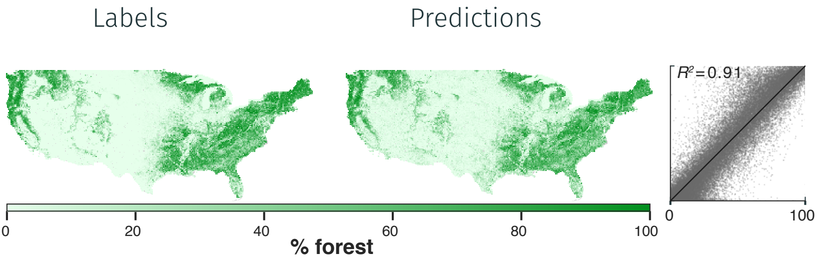
Step 2: Calculate features for each tile; Google Maps imagery (≈4m)

Step 3: Link features to labels within each tile:

Step 4: Train model using ridge regression, and test on holdout set

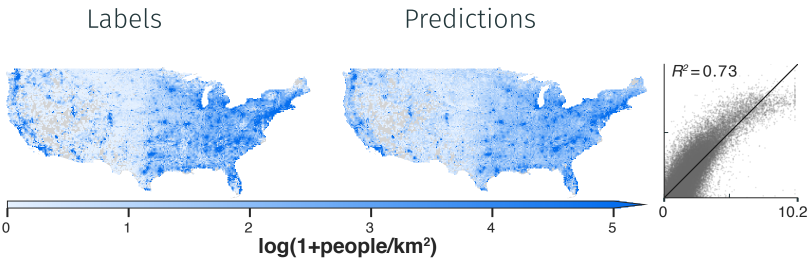


Domain #1: FOREST COVER



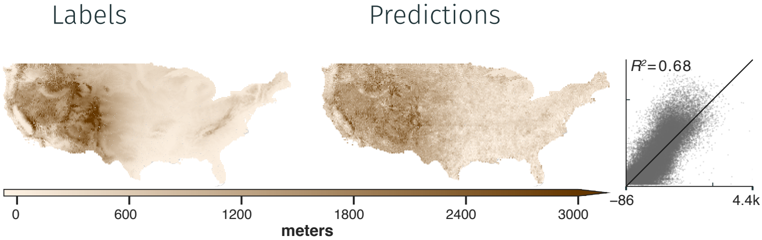

```
merge treecover.dta x.dta, by(lat lon)
ridgereg y x if insample
predict y
```

Domain #2: **POPULATION DENSITY**



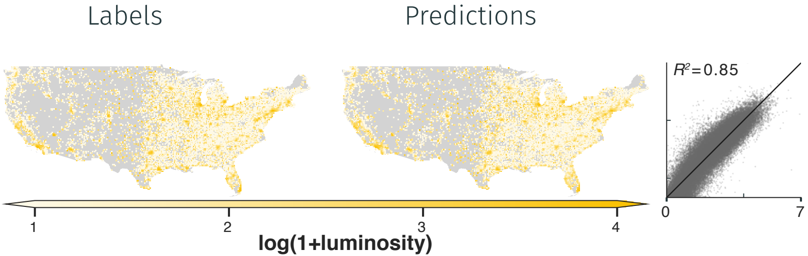

```
merge population.dta x.dta, by(lat lon)
ridgereg y x if insample
predict y
```

Domain #3: ELEVATION



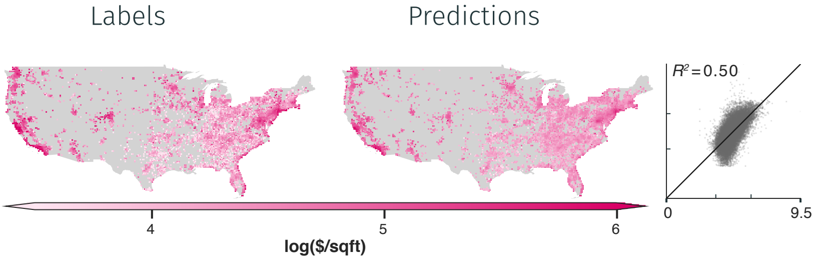

```
merge elevation.dta x.dta, by(lat lon)
ridgereg y x if insample
predict y
```

Domain #4: NIGHTTIME LUMINOSITY



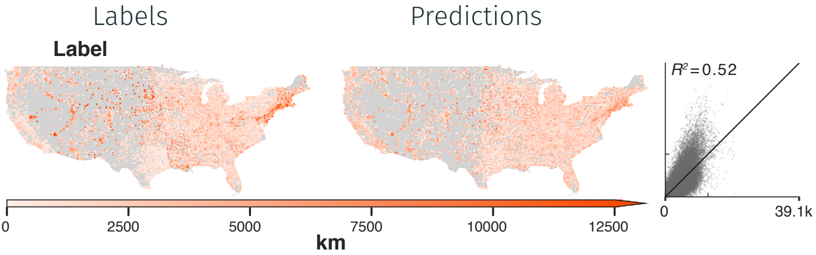

```
merge nightlights.dta x.dta, by(lat lon)
ridgereg y x if insample
predict y
```

Domain #5: AVG HOUSE PRICES



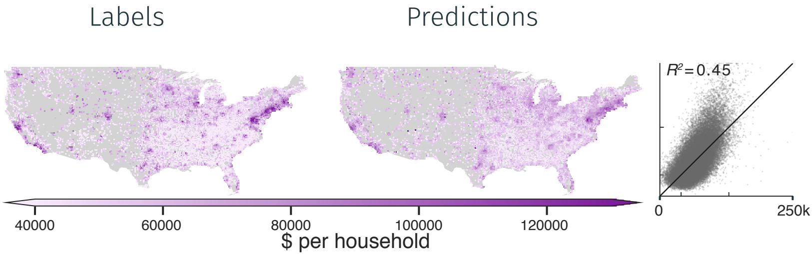

```
merge houseprices.dta x.dta, by(lat lon)
ridgereg y x if insample
predict y
```

Domain #6: TOTAL ROAD LENGTH



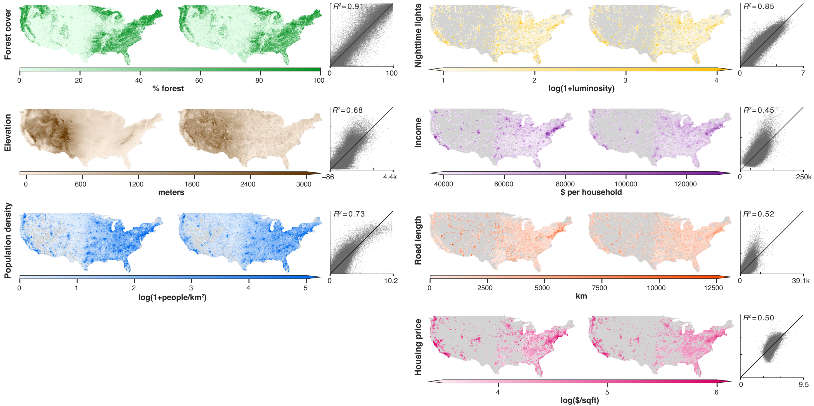

```
merge roadlength.dta x.dta, by(lat lon)
ridgereg y x if insample
predict y
```

Domain #7: INCOME PER HOUSEHOLD



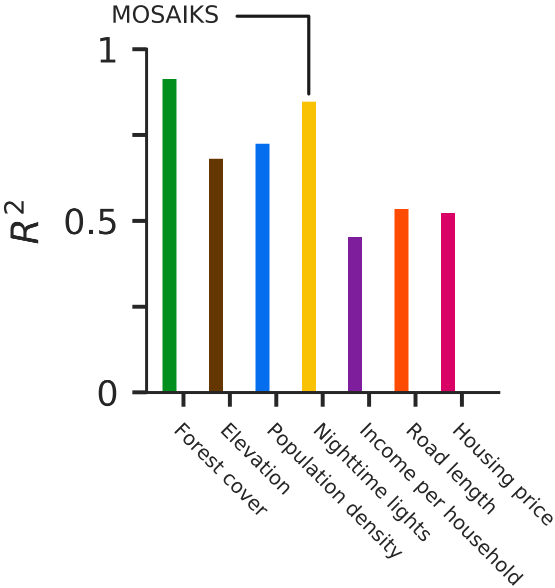

```
merge income.dta x.dta, by(lat lon)
ridgereg y x if insample
predict y
```

Pre-computed features generalize across domains

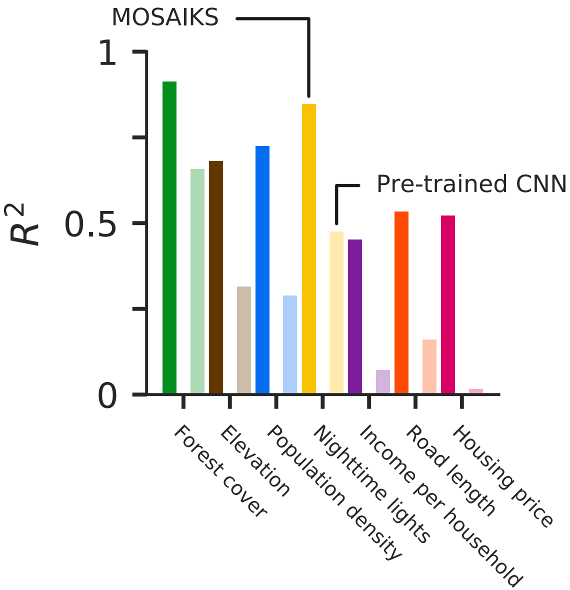


Alternative patch sizes

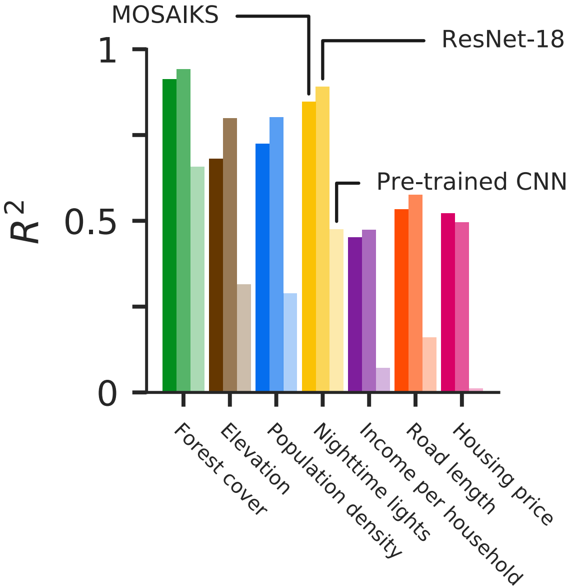
Competitive w/ deep convolutional neural network



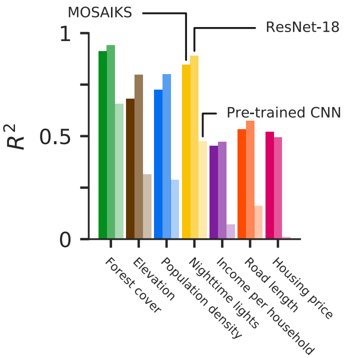
Competitive w/ deep convolutional neural network



Competitive w/ deep convolutional neural network



Competitive w/ deep convolutional neural network



- MOSAIKS is 250-10,000× **faster to train** than CNN Table
  - CNN on GPU: 7.9 hours
  - MOSAIKS on GPU: 3 seconds
  - MOSAIKS on laptop: 2 minutes
- MOSAIKS and CNN capture **similar information** from imagery Scatter
- MOSAIKS also competitive w/ nightlights **transfer learning** approach (e.g. Jean et al. 2016, Head et al., 2017) DHS

Roadmap

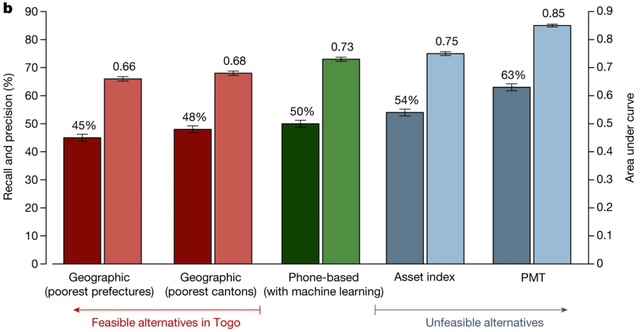
Can a single set of features achieve state of the art performance across a variety of tasks?

1. **Test generalization across tasks**, and compare performance and cost to existing SIML models
2. **Evaluate model sensitivity**, particularly under limited data/storage conditions
3. **Scale the analysis** globally and across the outcomes in a national survey
4. **Demonstrate additional properties** of MOSAIKS

Experiment 2: Evaluate model sensitivity

Consequential decisions likely to increasingly depend on (SI)ML output, such as which families should receive financial assistance

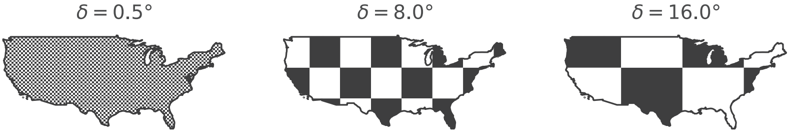
- MOSAIKS' computational efficiency enables characterization of performance and uncertainty
- Retraining deep CNNs in the same way would be computationally prohibitive



Source: Aiken et al., 2022

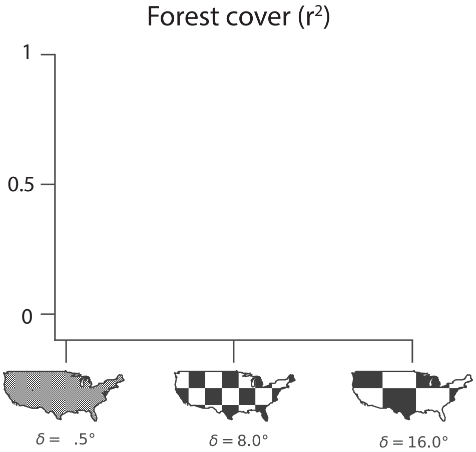
Performance: spatial extrapolation

1. Partition sample in checkerboard
2. Train on white squares
3. Test on black squares
4. Jitter checkerboard location & repeat
5. Compare to spatial interpolation of ground-truth

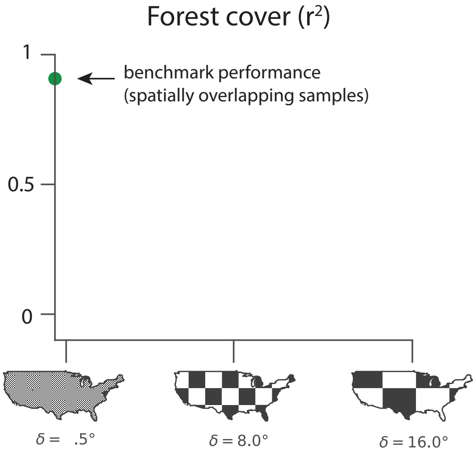


(Reference: $8° \times 8°$ = 888 km $\times$ 682 km (552 mi $\times$ 424 mi) at centroid)

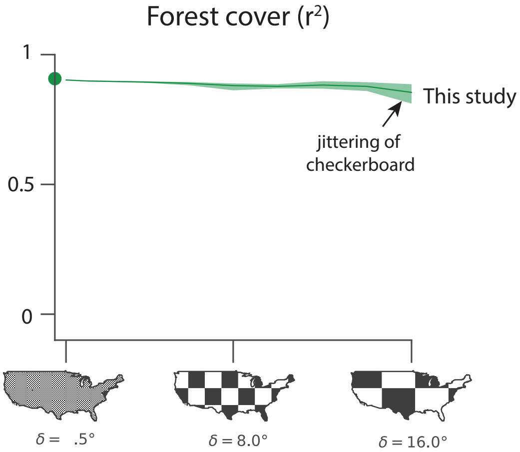
Spatial extrapolation out of sample



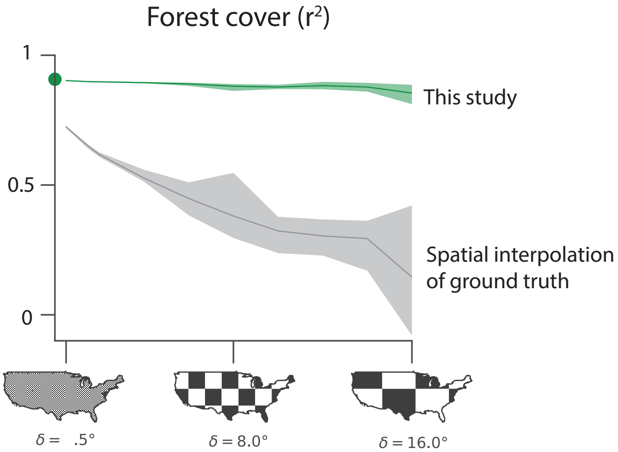
Spatial extrapolation out of sample



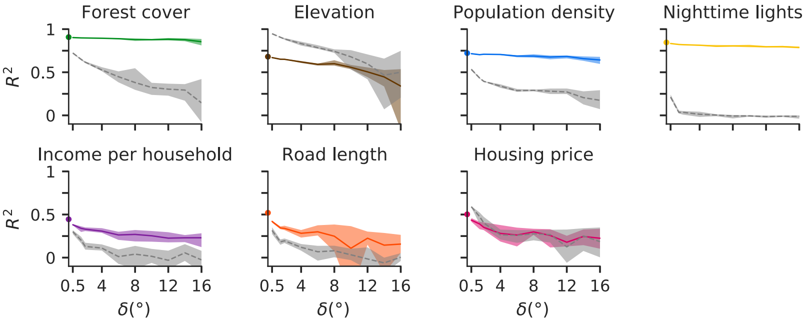
Spatial extrapolation out of sample



Spatial extrapolation out of sample

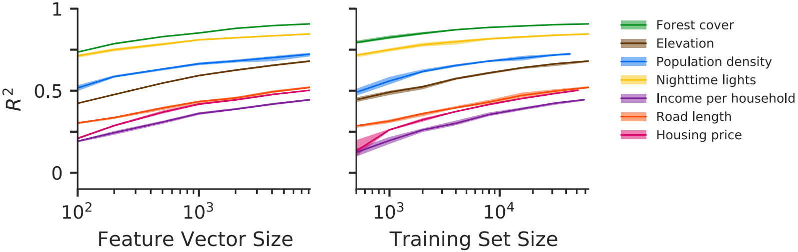


Spatial extrapolation out of sample



MOSAIKS substantially outperforms spatial interpolation across all tasks except for elevation and housing price.

Model sensitivity: number of features & sample size



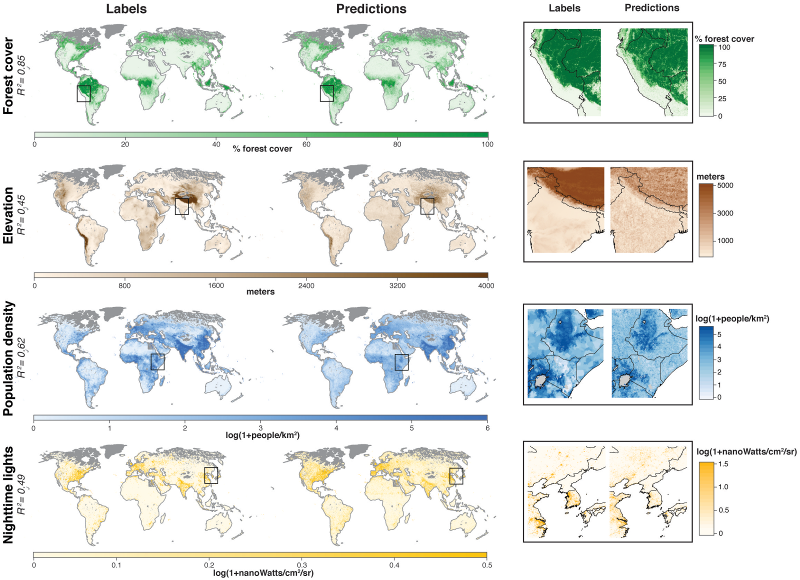
- A majority of the signal is recovered using K = 200 (vs. K = 8,192)
  - Range: 55% (income) - 89% (nighttime lights)
- A majority of the signal is recovered using N = 500 (vs. N = 64,000) for some (but not all) outcomes
  - Range: 26% (housing price) - 87% (forest cover)

Roadmap

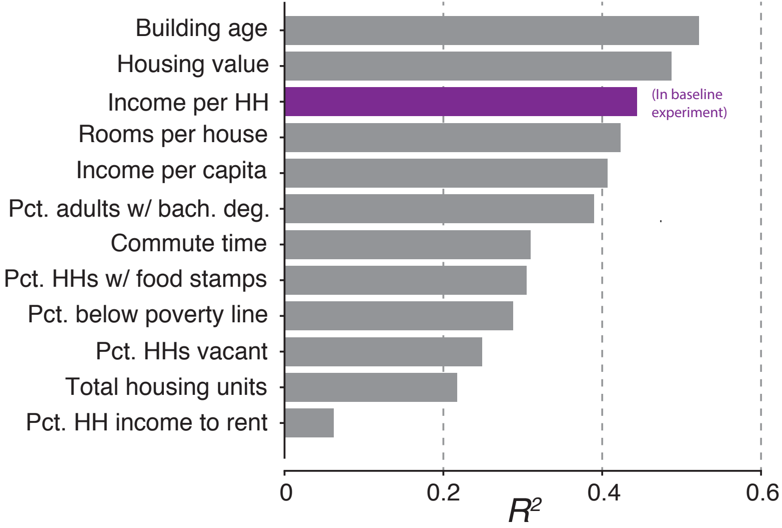
Can a single set of features achieve state of the art performance across a variety of tasks?

1. **Test generalization across tasks**, and compare performance and cost to existing SIML models
2. **Evaluate model sensitivity**, particularly under limited data/storage conditions
3. **Scale the analysis** globally and across the outcomes in a national survey
4. **Demonstrate additional properties** of MOSAIKS

Performance at global scale



Performance predicting variables from the 2015 ACS



Roadmap

Can a single set of features achieve state of the art performance across a variety of tasks?

1. **Test generalization across tasks**, and compare performance and cost to existing SIML models
2. **Evaluate model sensitivity**, particularly under limited data/storage conditions
3. **Scale the analysis** globally and across the outcomes in a national survey
4. **Demonstrate additional properties** of MOSAIKS

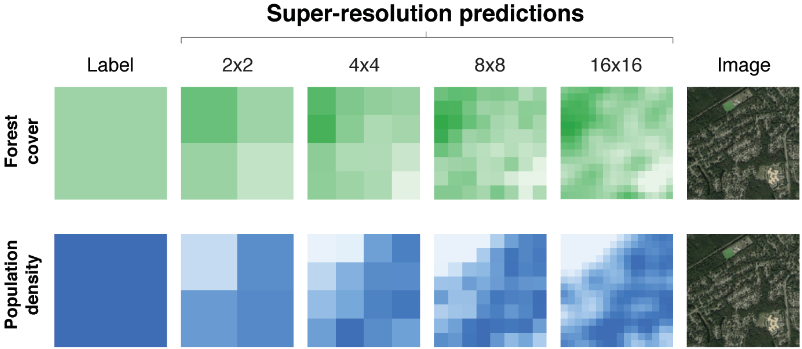
Additional property: Label super-resolution

Goal: Predict at finer resolution than existing training data.

Step 1: Train using 1km by 1km labels

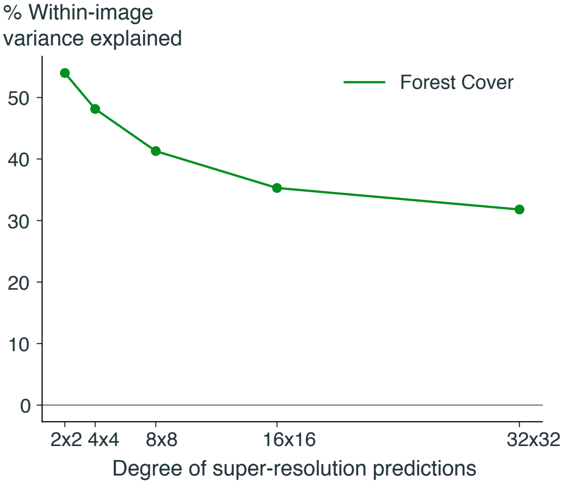
Step 2: Predict at finer resolution

Step 3: Evaluate performance using fine resolution labels



Additional property: Label super-resolution

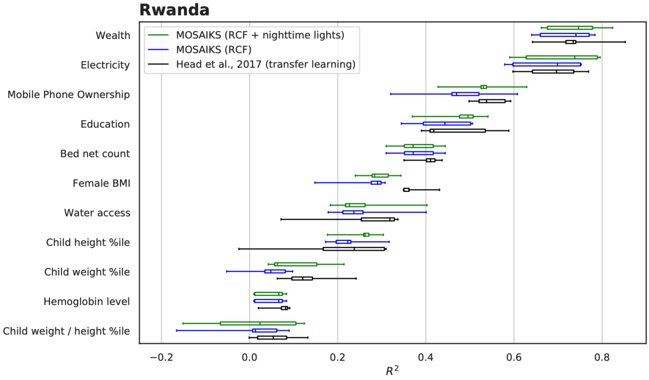
Goal: Predict at finer resolution than existing training data.



Additional property: Feature combinations

Goal: Combine features from multiple sources into the same model

$$y = \text{RCF}\beta + \text{NL}\gamma + \varepsilon$$



Note: Predicting household outcomes from Demographic & Health Surveys using MOSAIKS vs. transfer learning. Similar results for Haiti and Nepal.

Additional property: Feature combinations

Goal: Combine features from multiple sources into the same model

$$y = \text{RCF}\boldsymbol{\beta} + \text{CNN}\boldsymbol{\alpha} + \varepsilon$$

Task	MOSAIKS R^2	ResNet-18 R^2	Hybrid R^2
Forest cover	0.89	0.94	0.94
Elevation	0.68	0.80	0.81
Population density	0.71	0.81	0.81
Nighttime lights	0.85	0.89	0.90
Income	0.45	0.47	0.51
Road length	0.53	0.58	0.59
Housing price	0.53	0.49	0.58

Limitations

Satellite imagery + machine learning is a powerful but inherently limited combination.

Limitations

Satellite imagery + machine learning is a powerful but inherently limited combination.

- Machine learning predictions are **only as good as the ground truth data** we train on

  $\Longrightarrow$ continued collection of traditional datasets is critical (Lobell et al., 2020)

Limitations

Satellite imagery + machine learning is a powerful but inherently limited combination.

- Machine learning predictions are **only as good as the ground truth data** we train on

  $\implies$ continued collection of traditional datasets is critical (Lobell et al., 2020)

- **Algorithmic bias** is insufficiently studied in the imagery domain

  $\implies$ which populations are we misrepresenting? (Chi et al., 2022)

  $\implies$ what effect to systematic errors have on downstream research and policy decisions? (Proctor, Carleton and Sum, 2023; Angelopoulos et al., 2023)

Limitations

Satellite imagery + machine learning is a powerful but inherently limited combination.

- Machine learning predictions are **only as good as the ground truth data** we train on

  $\implies$ continued collection of traditional datasets is critical (Lobell et al., 2020)

- **Algorithmic bias** is insufficiently studied in the imagery domain

  $\implies$ which populations are we misrepresenting? (Chi et al., 2022)

  $\implies$ what effect to systematic errors have on downstream research and policy decisions? (Proctor, Carleton and Sum, 2023; Angelopoulos et al., 2023)

- **Privacy and ethical** concerns grow as imagery and algorithmic precision improve

Conclusions

- Satellite imagery + machine learning hold substantial promise for filling critical data gaps and supporting sustainable development
    - But this potential has been limited by the costs of modern deep learning approaches

Conclusions

- Satellite imagery + machine learning hold substantial promise for filling critical data gaps and supporting sustainable development
  - But this potential has been limited by the costs of modern deep learning approaches
- MOSAIKS can make this technology accessible while unlocking more information from each image
  - Many maps can be predicted using a single featurization of imagery, at low computational cost
  - API allows you to generate your own imagery-based predictions!
  - Code, input data, and output data are publicly hosted

Conclusions

- **Satellite imagery + machine learning hold substantial promise** for filling critical data gaps and supporting sustainable development
  - But this potential has been limited by the costs of modern deep learning approaches

- **MOSAIKS can make this technology accessible while unlocking more information from each image**
  - Many maps can be predicted using a single featurization of imagery, at low computational cost
  - API allows you to generate your own imagery-based predictions!
  - Code, input data, and output data are publicly hosted

What questions can satellite imagery + machine learning help you solve?

Acknowledgements

Team

Benjamin Recht, Cullen Molitor, Esther Rolf, Eugenio Noda, Hikari Murayama, Hannah Druckenmiller, Ian Bolliger, Jeanette Tseng, Jessica Katz, Jonathan Proctor, Luke Sherman, Miyabi Ishihara, Simon Greenhill, Solomon Hsiang, Jonathan Proctor, Taryn Fransen, Trinetta Chong, Vaishaal Shankar

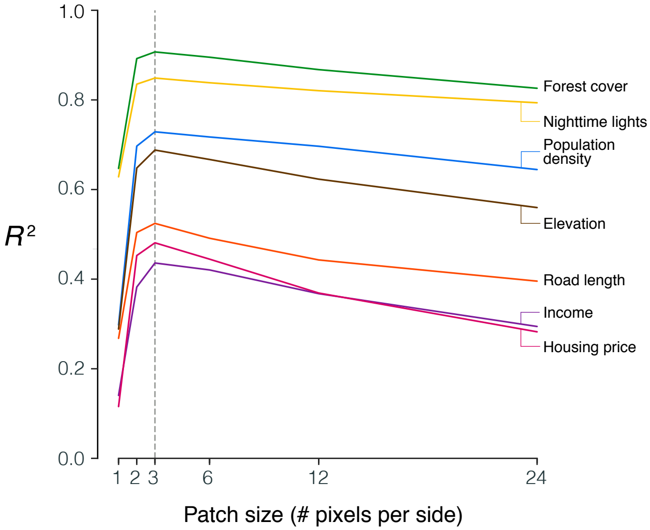
Partners and Funding

USAID, UNDP, CEGA, UCSB Bren School Environmental Data Science Masters Program

MOSAIKS API

www.mosaiks.org

Performance by patch size



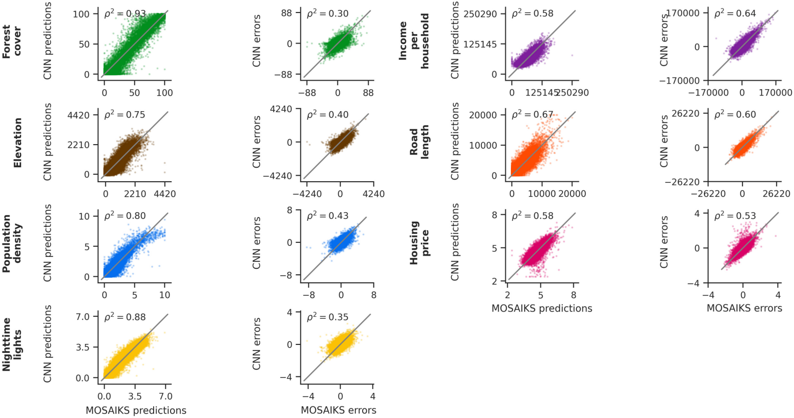
Back

Wall-clock times: MOSAIKS vs. CNN

Component	**ResNet** Time (GPU)	**MOSAIKS** Time
Training set featurization ($N = 80$k)	$\sim$ **7.9 hours**	$\sim$ 1.2 hours (GPU)
Model training		$\sim$ 2.8 seconds (GPU) $\sim$ 50 seconds (10 cores) $\sim$ **1.8 minutes (laptop)**
Holdout set featurization ($N = 20$k)	$\sim$ **40 seconds**	$\sim$18 minutes (GPU)
Holdout set prediction		$<$ 0.01 seconds (GPU) $\sim$ 0.1 seconds (10 cores) $\sim$ **0.7 seconds (laptop)**
Total cost to ecosystem	$\sim$ **7.9 hours**	$\sim$ 1.5 hours (GPU)
Total cost to user	$\sim$ **7.9 hours**	$\sim$ 2.8 seconds (GPU) $\sim$ 50.1 seconds (10 cores) $\sim$ **1.8 minutes (laptop)**

Back

RCF and CNN capture similar information from images



Back

References

Angelopoulos, Anastasios N, Stephen Bates, Clara Fannjiang, Michael I Jordan and Tijana Zrnic. 2023. "Prediction-powered inference." *Science* 382(6671):669–674.

Chi, Guanghua, Han Fang, Sourav Chatterjee and Joshua E Blumenstock. 2022. "Microestimates of wealth for all low-and middle-income countries." *Proceedings of the National Academy of Sciences* 119(3).

Hansen, Matthew C, Peter V Potapov, Rebecca Moore, Matt Hancher, Svetlana A Turubanova, Alexandra Tyukavina, David Thau, Stephen V Stehman, Scott J Goetz, Thomas R Loveland et al. 2013. "High-resolution global maps of 21st-century forest cover change." *Science* 342(6160):850–853.

Lobell, David B, George Azzari, Marshall Burke, Sydney Gourlay, Zhenong Jin, Talip Kilic and Siobhan Murray. 2020. "Eyes in the sky, boots on the ground: Assessing satellite-and ground-based approaches to crop yield measurement and analysis." *American Journal of Agricultural Economics* 102(1):202–219.

Pekel, Jean-François, Andrew Cottam, Noel Gorelick and Alan S Belward. 2016. "High-resolution mapping of global surface water and its long-term changes." *Nature* 540(7633):418–422.

Proctor, Jonathan, Tamma Carleton and Sandy Sum. 2023. Parameter Recovery Using Remotely Sensed Variables. Technical report National Bureau of Economic Research.

Rolf, Esther, Jonathan Proctor, Tamma Carleton, Ian Bolliger, Vaishaal Shankar, Miyabi Ishihara, Benjamin Recht and Solomon Hsiang. 2021. "A generalizable and accessible approach to machine learning with global satellite imagery." *Nature communications* 12(1):1–11.