Geospatial Data for Small Area Estimation

Machine learning for SAE

Joshua D. Merfeld
KDI School of Public Policy and Management, IZA

January 21, 2025

Introduction

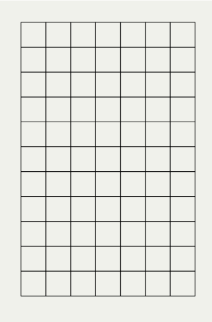
- This presentation: machine learning for small area estimation (SAE)
    - Focus of SAE will be geographic
- Will focus on convolutional neural networks (CNNs)
    - Application to satellite imagery
    - Small discussion of vision transformers

Malawi

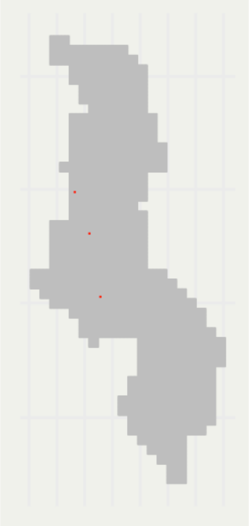
Tunduma
Njombe
Ipanya
Chitipa
Karonga
Peramiho
Mbinga
Rhumpi
Ekwendeni
Mzuzu
Nkhata Bay
Lake Malawi
Mzimba
Niassa
Lundazi
Nkhotakota
Kasungu
Lichinga
Malawi
Mchinji
Salima
Lilongwe
Dedza
Mangochi
Ntcheu
Liwonde
Zomba
Songo
Blantyre
Thyolo
Tete
Nsanje

Malawi - as grids

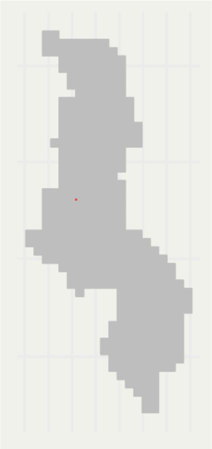
Malawi - as grids

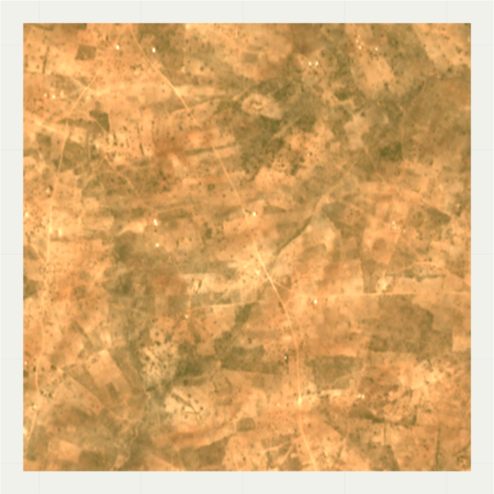


Malawi - three grid cells



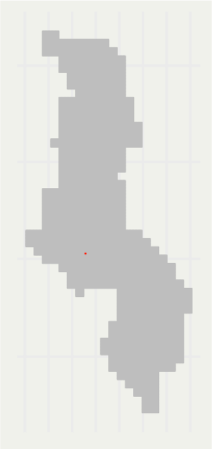
Example grid image 1





Example grid image 2

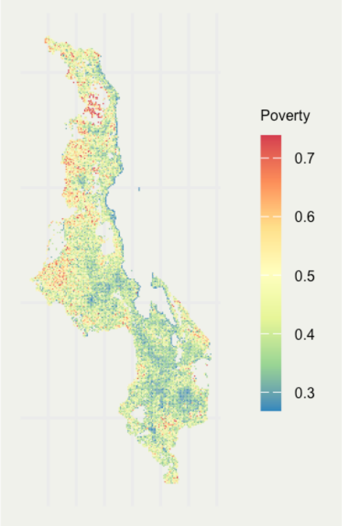
Example grid image 3



Images to predictions

- We have our images
- We will build some model
- We then predict the outcome (e.g. welfare) based on each image

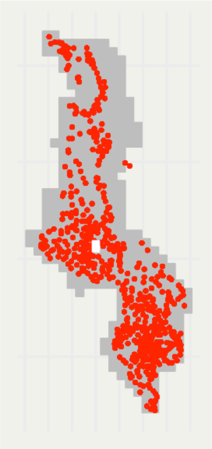
Images to predictions

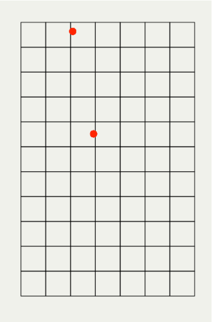


Convolutional Neural Networks (CNNs)

- CNNs have been a common tool in image analysis and prediction
  - “Supervised” machine learning
- Requires labeled data
  - In our example, each grid cell needs a poverty “label”
  - Common to get it from surveys with latitude/longitude

Poverty labels

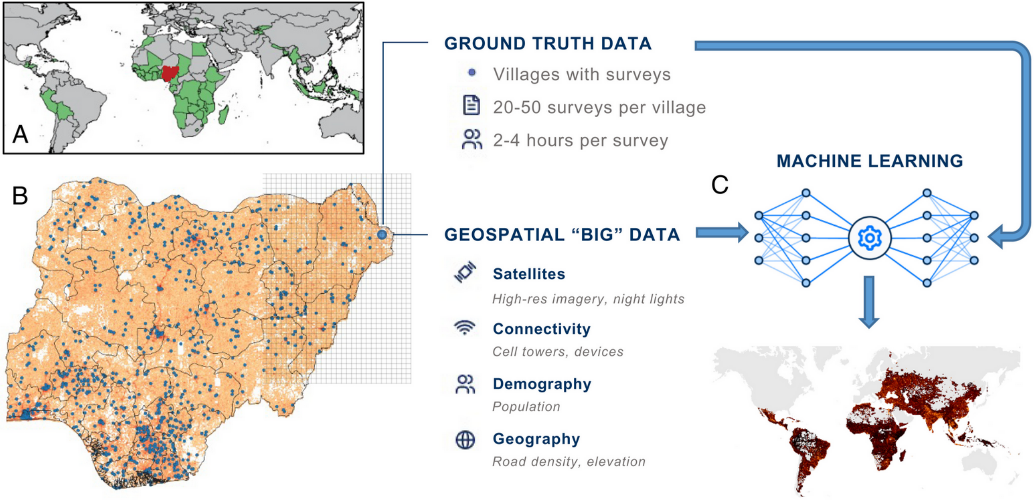




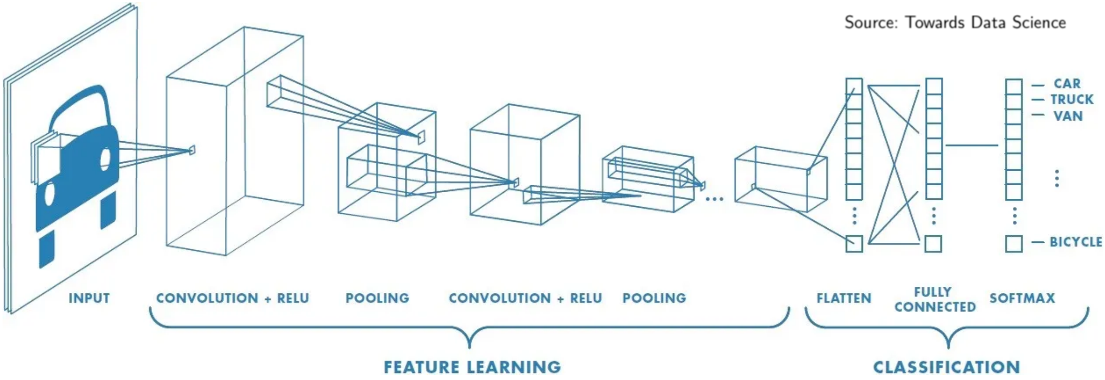
Common to combine methods

- Chi et al. (2021)
  - CNNs to extract “features” (more later)
  - Gradient-boosted regression trees to predict welfare
- Note that the use of gradient-boosted trees has been proposed for SAE before!
  - Merfeld et al. (2024)
  - Key issue is variance estimation

Chi et al. (2022) Figure 2

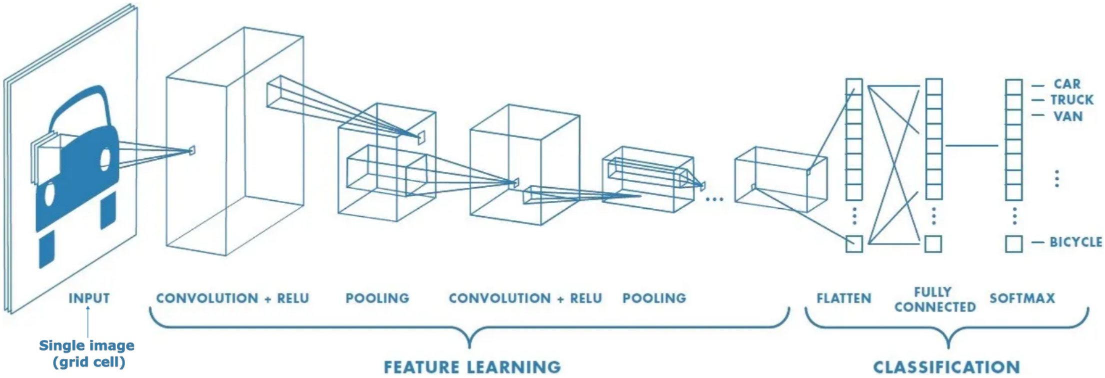


CNNs

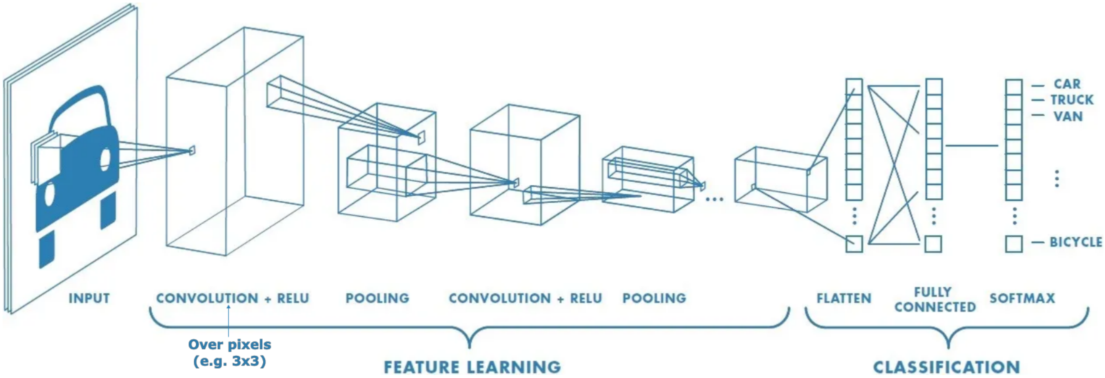


Source: Towards Data Science
CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

CNNs

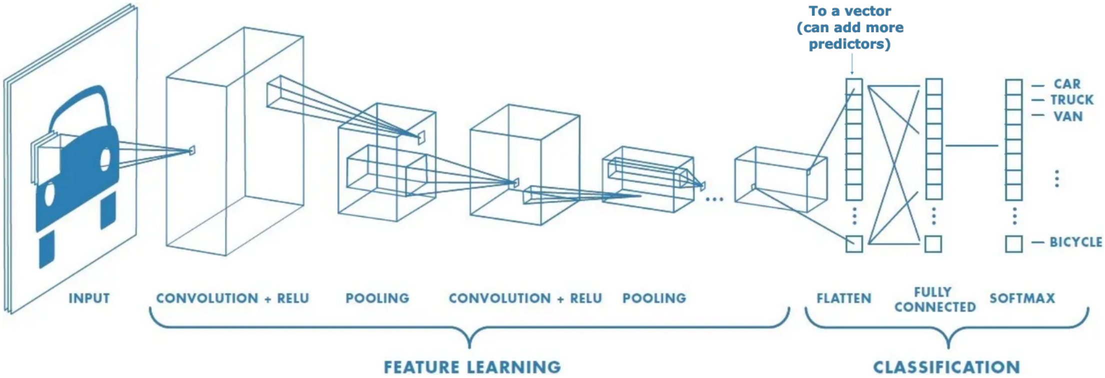

INPUT
Single image
(grid cell)
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FEATURE LEARNING
FLATTEN
FULLY
CONNECTED
SOFTMAX
CLASSIFICATION
CAR
TRUCK
VAN
BICYCLE

CNNs



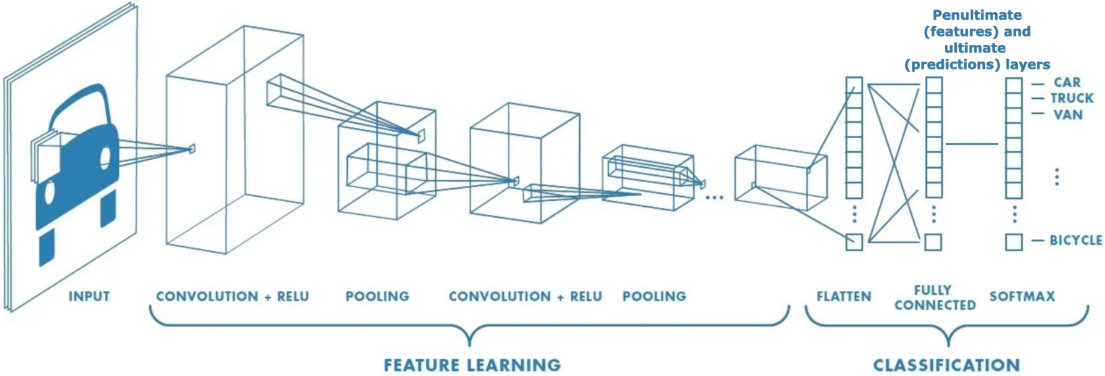
CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
Over pixels
(e.g. 3x3)
FEATURE LEARNING
CLASSIFICATION

CNNs



To a vector
(can add more
predictors)
CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY
CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

CNNs



Penultimate
(features) and
ultimate
(predictions) layers
CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY
CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

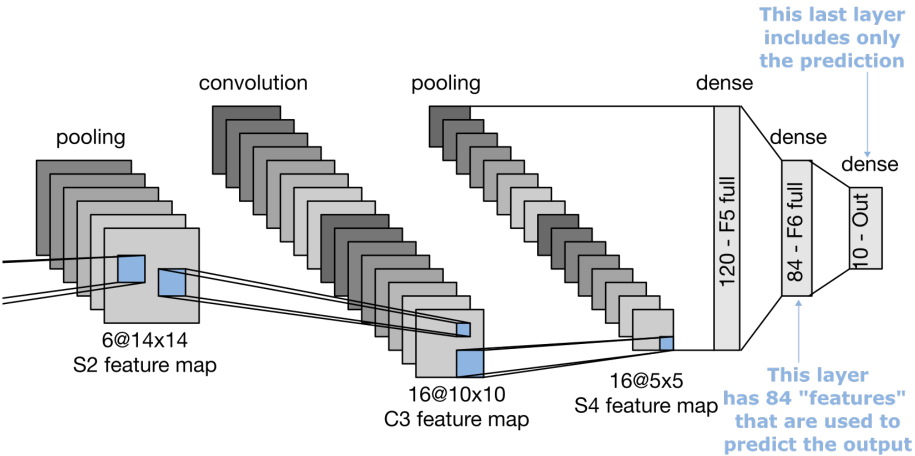
CNN architecture

- CNN “architecture” is completely flexible
  - Number of layers
  - Number of nodes in each layer
  - Number of convolutions
  - Activation functions
  - Type of output (e.g. classification? prediction?)
  - etc.
- Common to use a “pre-trained” model
  - E.g. ResNet, AlexNet, etc.
  - Transfer learning
  - “Fine-tuning”

Two ways to use CNNs for SAE

1. Just use pure predictions from the CNN itself
    - Issue: no variance estimates
2. Use some CNN output as a feature in a traditional SAE model
    - Can use prediction itself
    - Can use *features* from e.g. penultimate layer

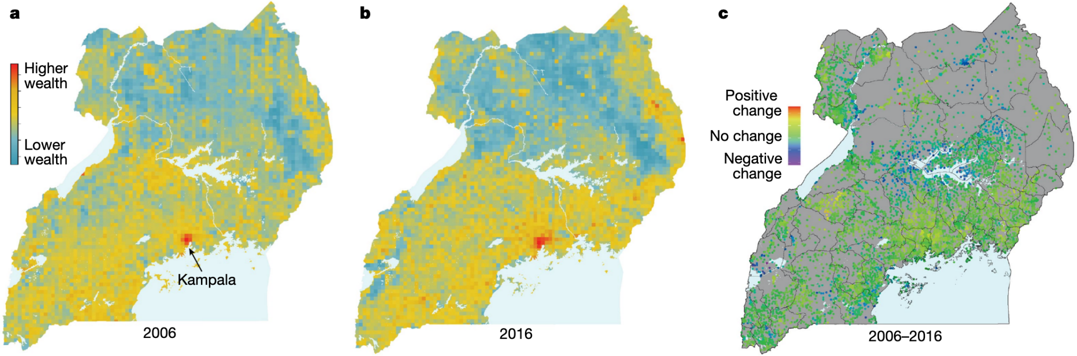
Outputs of LeNet (Source: D2L.ai)



Example 1: Ratledge et al. (2022)

- CNN to predict wealth
- Variance estimates (for treatment effects) from bootstrapping

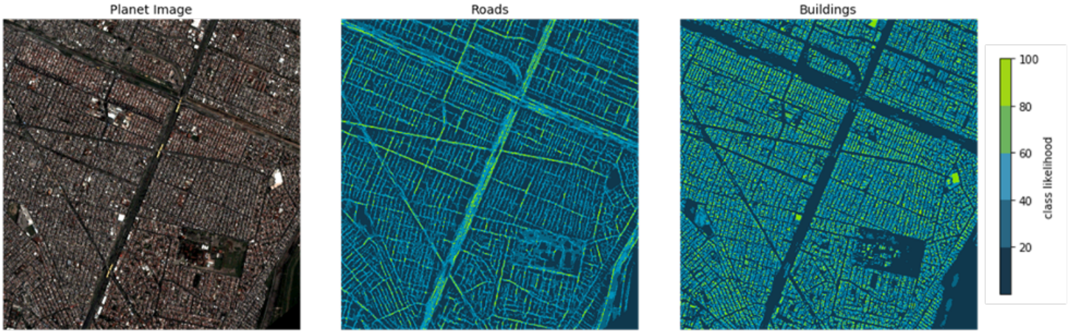
Example 1: Ratledge et al. (2022)



Example 2: Newhouse et al. (2024)

- CNN to predict poverty, land classification
- Use these predictions as features in traditional SAE model
- Performs well!

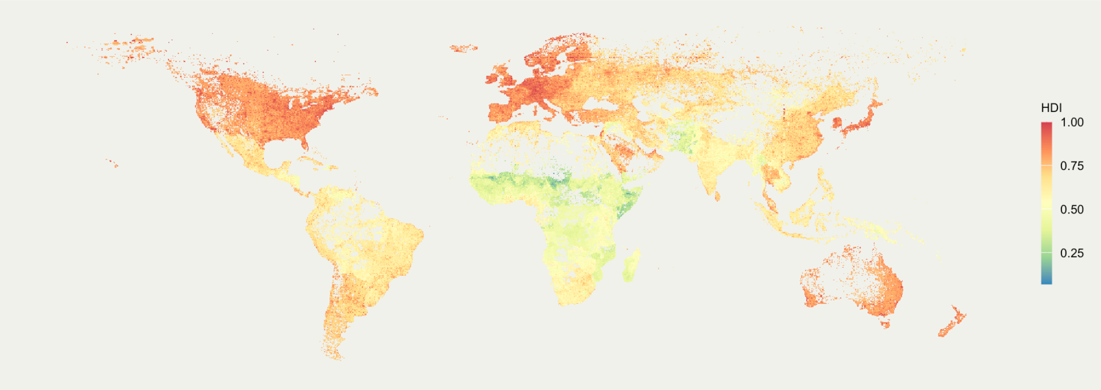
Example 2: Newhouse et al. (2024)



Example 3: Merfeld et al. (2024)

- Vision transformers (ViT), not CNNs
  - Some features in common with LLMs (e.g. ChatGPT), like “attention” mechanisms
  - Pure ViTs may perform better than CNNs (e.g. Dosovitskiy et al., 2021)
- Use features from penultimate layer as features in traditional SAE model
- Work in progress!
- Very little work of this kind

Mosaiks - Rolf et al. (2021)



Mosaiks features in SAE

- They make available all of the “features” from the penultimate layer of a CNN
    - Around 5,000 of them
- We have used these as candidate features before!
    - E.g. in Kim et al. (2024) for human capital

Thank you!

https://joshmerfeld.github.io

https://github.com/JoshMerfeld