

Evaluating Deep Learning Approaches for Out-of-Domain Land Cover Classification

Jack Xiao, Matthew Kolodner
{jackxiao, mkolod} @stanford.edu

Background

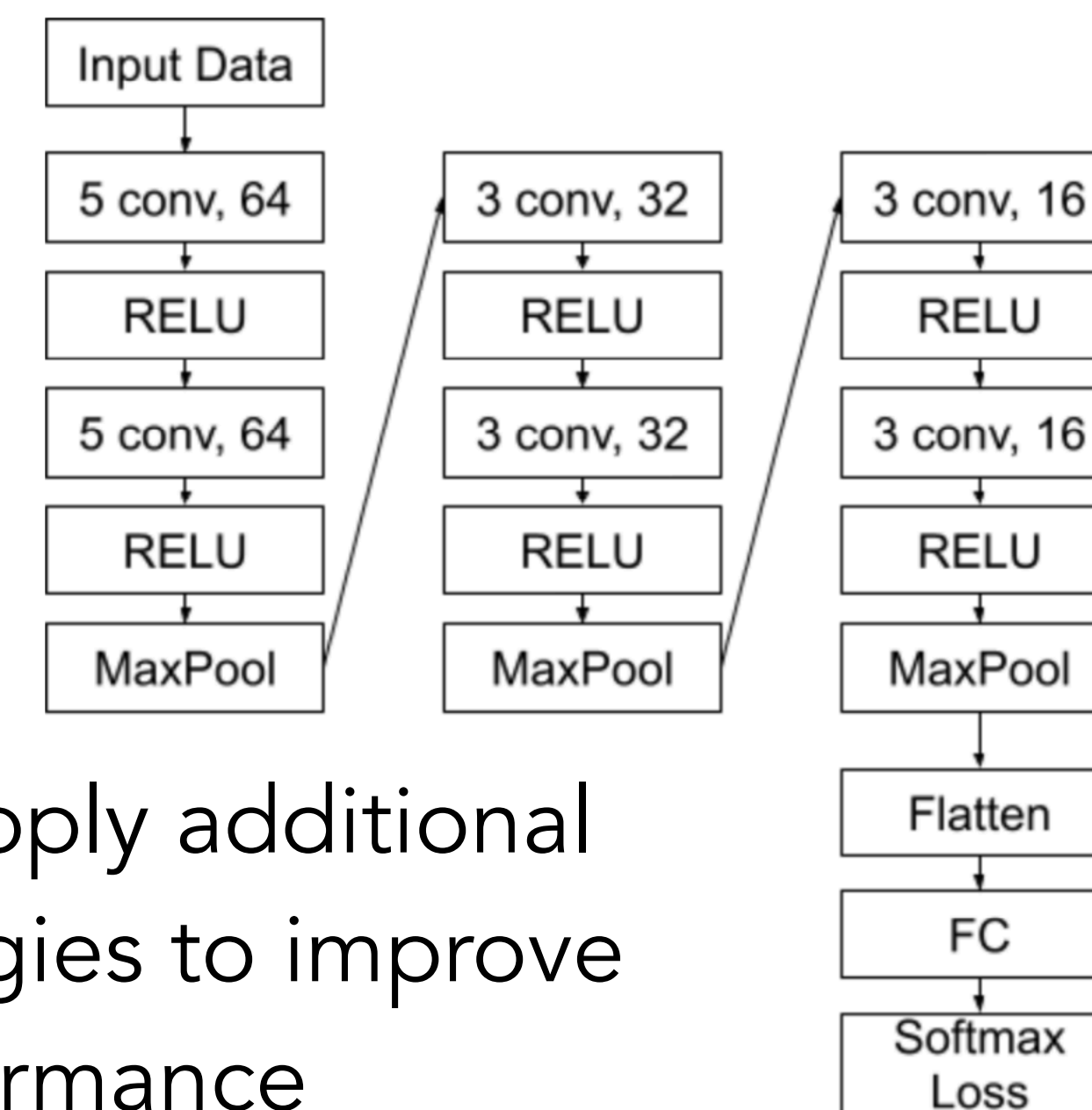
- Understanding changes to Earth's surface is key to achieving many sustainability goals
- Land cover classification and mapping provides valuable information
- Deep learning models often used for land-cover classification tasks
- Low-income regions lack good land cover data (low-resolution, not enough labeled examples)
- High-income regions have plenty of data
- Classifiers rarely perform well on out-of-domain data, especially if data distribution is different

Problem Statement

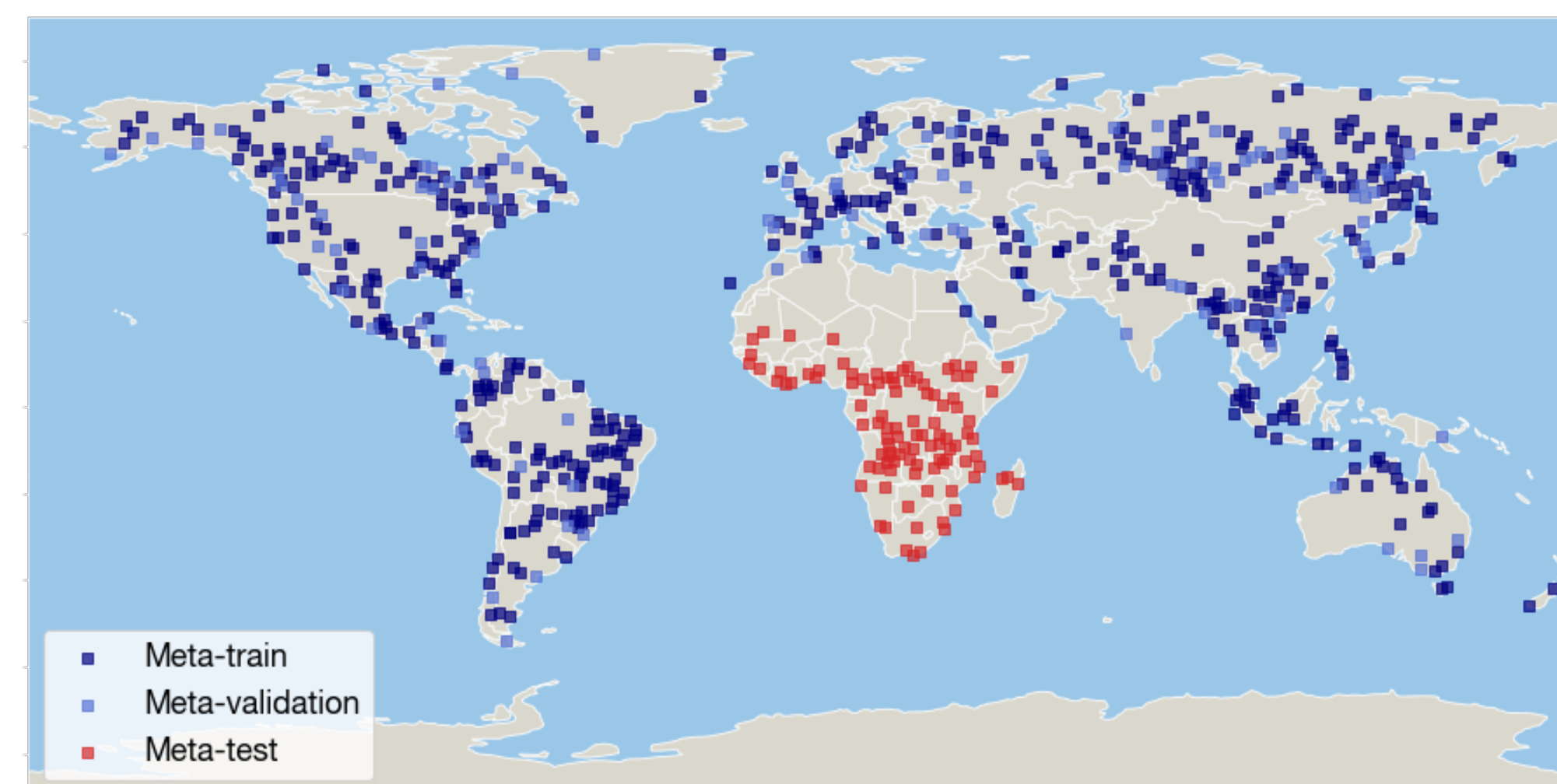
- We analyze the performance of 3 different deep learning models in classifying out-of-domain land cover data
- The inputs are time-series of pixels from satellite imagery, which we run through each model to output 1 of 17 land cover labels
- 4 metrics: Accuracy, F1, Average Precision (mAP), Macro AUROC
- Examine confusion matrices for potential patterns and specific failure cases

Methods

- 3 models (CNN, LSTM and Transformer) with individually tuned hyperparameters, KNN baseline model
- Cross-Entropy Loss
- SGD with Nesterov Momentum
- Normalize and Regularize
- Take best model, apply additional regularization strategies to improve out-of-domain performance



Dataset



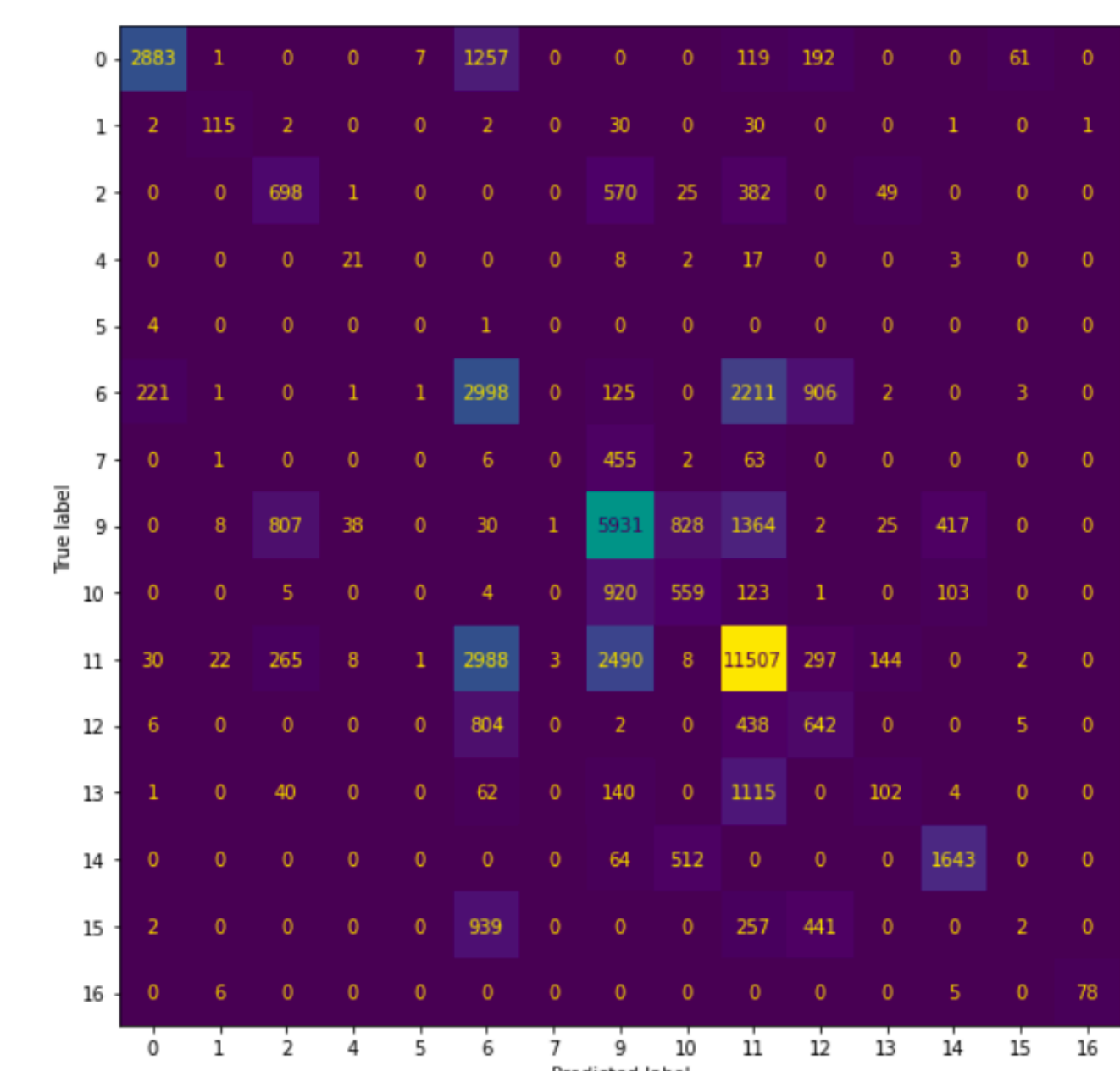
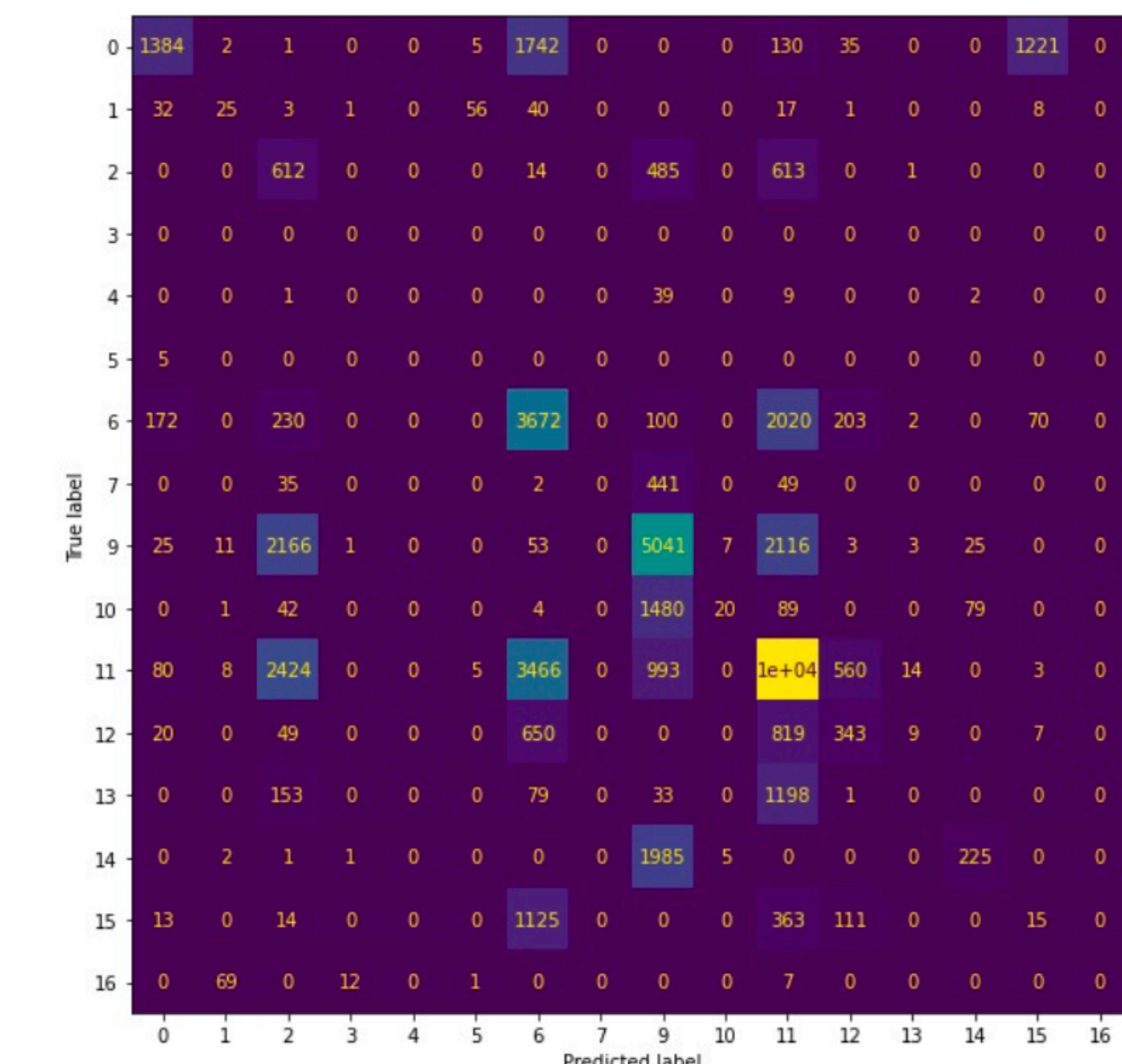
- SDG 15: Life on Land dataset from SustainBench to target UN SDGs
- Time series of pixels sampled from MODIS satellite imagery in 2018
- 7 bands (R, G, B, SWIR 1, SWIR 2, TIR, NIR)
- Total input dimension is 7x46

Results & Analysis

Model	Accuracy	F1 Score	mAP	AUROC
KNN	0.359	0.278	0.697	0.174
ConvNet	0.520	0.423	0.757	0.295
LSTM	0.517	0.467	0.791	0.324
Transformer	0.509	0.275	0.693	0.189

$p(\text{drop})$	Accuracy	F1 Score	mAP	AUROC
0	0.517	0.467	0.791	0.324
0.1	0.547	0.476	0.793	0.333
0.4	0.524	0.449	0.780	0.321

- LSTM was best overall, slightly outperforming CNN, both better than Transformer on this dataset
- Dropout notably improved performance across the board
- Clear patterns in failure cases
- Many misclassifications reflect in-domain training data distribution
- Some failure cases due to inherent similarity between types of land cover may be hard to avoid



Conclusions & Future Work

- Time-series nature of data lends itself better to models that can learn more complex dependencies
- High performance in out-of-domain data remains a challenge
- In-domain data distribution has strong influence on output
- Regularization strategies do help slightly boost test set performance
- Try additional regularization techniques (e.g. Data Augmentation)
- Representation Learning or Meta-Learning to enhance models