Anna Boser, Ryan Stofer, Kelly Caylor
University of California, Santa Barbara

# Super-resolution of Land Surface Temperature Using Deep Learning

## Introduction

Land surface temperature (LST) data and evapotranspiration derivatives are used in:

- public health & environmental justice
- hydrology and water resources management
- ecology

The highest available resolution satellite-based LST at 70m lags behind other satellite outputs and stifles research progress. We propose downscaling the resolution of LST data using deep learning rather than traditional pixel-based approaches since they have shown significant promise in super-resolution tasks, require fewer inputs, and do not need to be retrained for each image to be downscaled.

## Methods

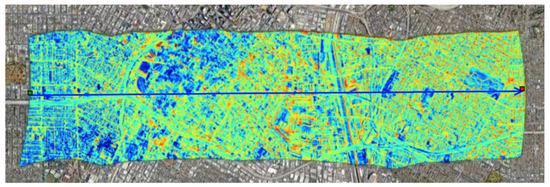
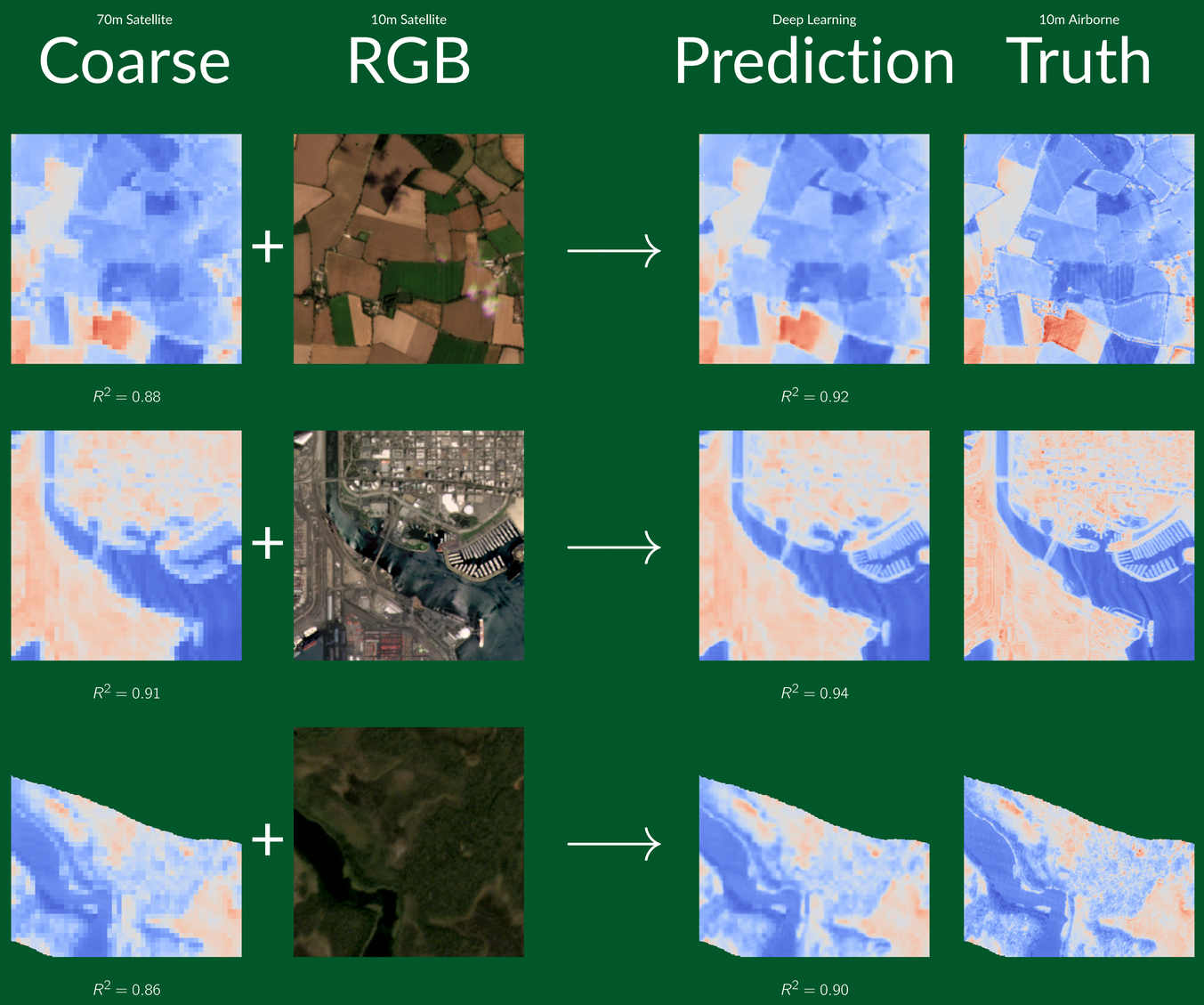


**Figure 1:** Example Flight over LA (Provided by NASA JPL)

- Collected 41 airborne HyTES LST flights and associated PlanetLabs RGB
- Locations include US and Europe
- Resampled 5525 tiles to 10m resolution with size 264x264
- Split by location and day on train/val/test to evaluate generalizibility of new locations
- Trained using a Convolutional Neural Network for 123 epochs
- Compared results to Random Forest (RF) model




## Results & Conclusion

### Model Performance:

| | $R^2$ | | | $RMSE$ | | |
|---|---|---|---|---|---|---|
| | Coarse | CNN | RF | Coarse | CNN | RF |
| Overall | 0.69 | 0.74 | 0.46 | 1.11 | 1.02 | 1.12 |
| Agricultural | 0.68 | 0.71 | 0.51 | 1.43 | 1.23 | 1.63 |
| Natural | 0.70 | 0.75 | 0.45 | 0.94 | 0.89 | 1.01 |
| Urban | 0.69 | 0.75 | 0.66 | 3.93 | 3.46 | 3.90 |

**Figure 2:** Average performance across images of the CNN and RF models compared to the baseline, coarsened 70m input (Coarse).

While the CNN predictions are closer to the ground truth than the coarsened input, the RF is unable to effectively improve the resolution of the 70m data.

### Model Runtime:

| | Initial Training | New Image Prediction |
|---|---|---|
| RF CPU | NA | 4.49 seconds |
| CNN GPU | 2 days, 14 hours | 0.41 seconds |
| CNN CPU | NA | 2.96 seconds |

**Figure 3:** Comparison of computation time for the CNN and RF. All values were computed on a GTX 1080 Ti Blower Cooled GPU and 4 3.5GHz Intel CPU cores with 128Gb of RAM, 1TB of local SSD and 4TB of local scratch space.

While the CNN requires substantial upfront computation time for model training (reported time for training 123 epochs), generating predictions for new images is significantly faster using the CNN than a RF, both on GPU and CPU.

**Conclusion:**

- Deep learning decreases the number of necessary inputs for downscaling LST data, improving LST resolution and vastly outperforming a random forest approach with only RGB inputs
- Improvements to the model such as model pre-training and adding additional bands such as NIR have the potential to further improve results

## Acknowledgements

We thank the CV4Ecology summer school, funded by the Resnick Institute for Sustainability, for helping kickstart this work. We are especially grateful for CV4Ecology mentor Benjamin Kellenberger for his substantial work and input on the model developed in this study. We thank Matthew McCabe for his help in acquiring the necessary RGB data to perform this analysis. This work is funded by the NSF Graduate Research Fellowship Program.