
Temporal Analysis of Regional Sustainability Using CNNs and Satellite Data

Xinyuan Cheng, xyz@stanford.edu
Nikita Namjoshi, nikitamn@stanford.edu
Rubi Rodriguez, rubi1rdz@stanford.edu

Abstract

We utilized public satellite data to build a deep learning classification model that describes the land use for a geographic region. Our model architecture is Resnet50 and achieves a 96% test set accuracy. Additionally, we built a pipeline for downloading new satellite data, and introduce a methodology to quantify the biocapacity of a geographic region.

1 Introduction

With a problem as colossal as climate change, it can often feel like no matter how many steps we take, the variables and outcomes are on a scale too large to examine quantitatively. Greenhouse gas reports are expensive, and often inaccurate; in fact, when counting carbon, being within an order of magnitude is considered a good estimate. Cities have a difficult time understanding if proposals passed actually had an impact, and nations struggle to prioritize initiatives, and understand how their footprint has changed over time. One of the metrics used to understand the environmental impact of such changes is the biocapacity of a region. Our work leverages machine learning and satellite imagery and proposes a quantitative methodology to track changes in biocapacity over time.

We leverage public satellite data from the Sentinel constellations of satellites launched by the European Commission and European Space Agency's Copernicus program in 2014. Specifically, we use imagery from Sentinel Level-2A, which measures orthorectified atmospherically corrected surface reflectance at a resolution of 10 meters for the RGB spectral bands. The input to our algorithm is a 64x64x3 image, which is fed to a Resnet50 model, and the output is one of ten classes describing the type of land use.

2 Related work

The field of machine learning and satellite imagery is growing rapidly, as researchers are moving from pixel by pixel models, to CNN architectures that can leverage spatial information to aid in classification. Kussul and Lavreniuk (2017) applied deep learning to classify crops from multitemporal images of Sentinel-1 satellites using a fully connected multilayer perceptron (MLP) and the most commonly used approach in the remote sensing community, random forest, and compared them with CNNs. Their research shows that CNN outperformed MLP yielding a target accuracy of more than 85% for all major crops.

Nijhawan and Joshi (2018) looked into land cover classification in India using satellite images. Their framework approach uses an ensemble of CNNs where each layer is fed with multispectral and geospatial satellite bands provided by Sentinel-2 satellite imagery. Their results show that CNNs outperformed with classification accuracy of 89.43%, compared to support vector machines, K-nearest neighbors, and random forest.

Helber and Bischke researched patch-based land use and land cover classification approaches using Sentinel-2 satellite images. Their research presented a geo-referenced dataset called EuroSAT. Benchmarks were provided for the dataset using CNNs to achieve an overall classification accuracy of 98.57% leveraging transfer learning. Thus, our research uses the learned lessons from previous research by choosing to use a CNN for Satellite image classification and builds upon the work from Helber et al, by using their annotated dataset EuroSat published earlier this year. We also report precision, recall, and F1 scores for our model, while they report only accuracy. Additionally, as pointed out in their paper, there are numerous environmental applications for a model trained on this dataset. Our work extends upon their research by demonstrating one type of application.

3 Dataset and Features

We used the EuroSat dataset provided by Helber et al. This dataset includes 27,000 64x64 images with 10 classes that describe land use. There is an RGB version of each image, as well as a version containing 13 spectral bands. The ten classes can be split into two broad categories: Natural (forest, river, sea/lake, herbaceous vegetation) and man made (annual crop, residential, highway, industrial, pasture, permanent crop). Although Sentinel’s coverage is global, the dataset only contains images from European regions.

4 Methods

We used the Resnet50 architecture, trained in batches of size 64 on an Nvidia Tesla v100. We appended the Resnet50 architecture with an additional Global Average Pooling layer, a Dense layer with 1024 units, Relu activation and L2 regularization, and then final Dense layer with softmax activation for 10 classes. The model was trained with the Adam optimizer and learning rate of 0.0001 with the standard categorical CrossEntropy loss. Tensorboard was utilized for model evaluation, and the Keras ReduceLROnPlateau callback used to tune the optimize the learning rate (see final_model.py script).

5 Experiments/Results/Discussion

The 27,000 images were randomly shuffled and split 80/20/20 (see build_dataset.py script), resulting in 21,600 images for training, 2,700 for the dev set, and 2,700 for the test set. We evaluated different models on the dev set accuracy and report the final metrics for the test set.

Although the data has 13 spectral bands, early experiments showed promising results with just RGB, so we limited our data to these three channels. Using the Keras Applications module, we tried out several architectures with different hyperparameters, and the metrics from the best configuration per architecture can be seen in Figure 1. We found that the models were very sensitive to changes in the learning rate and this was the main hyperparameter we choose to tune. For example, changing the learning rate from 0.01 to 0.001 had around a 20% increase in accuracy when using the Resnet50 architecture. We also experimented with the batch size. We tuned these parameters manually and were able to achieve good performance without a more sophisticated hyperparameter search method. We also experimented with data augmentation techniques by horizontally and vertically flipping images. This augmentation did not change the overall accuracy or precision/recall, but when we inspected the labels individually we found that augmentation consistently hurt the performance of the Annual Crop label, in one case dropping the precision by 0.17, so we decided that the augmentation was not helping our model to learn.

Model	Train Acc	Dev Acc	Test Acc
Resnet50	0.99	0.97	0.96
VGG	0.98	0.95	0.95
DeepNet	0.99	0.96	0.97

Table 1: Model Architecture Results

Our final Resnet50 model achieved a training accuracy of 0.99, validation accuracy of 0.97, and a test accuracy of 0.96. These results indicate that the model is overfitting slightly to the training set. In our

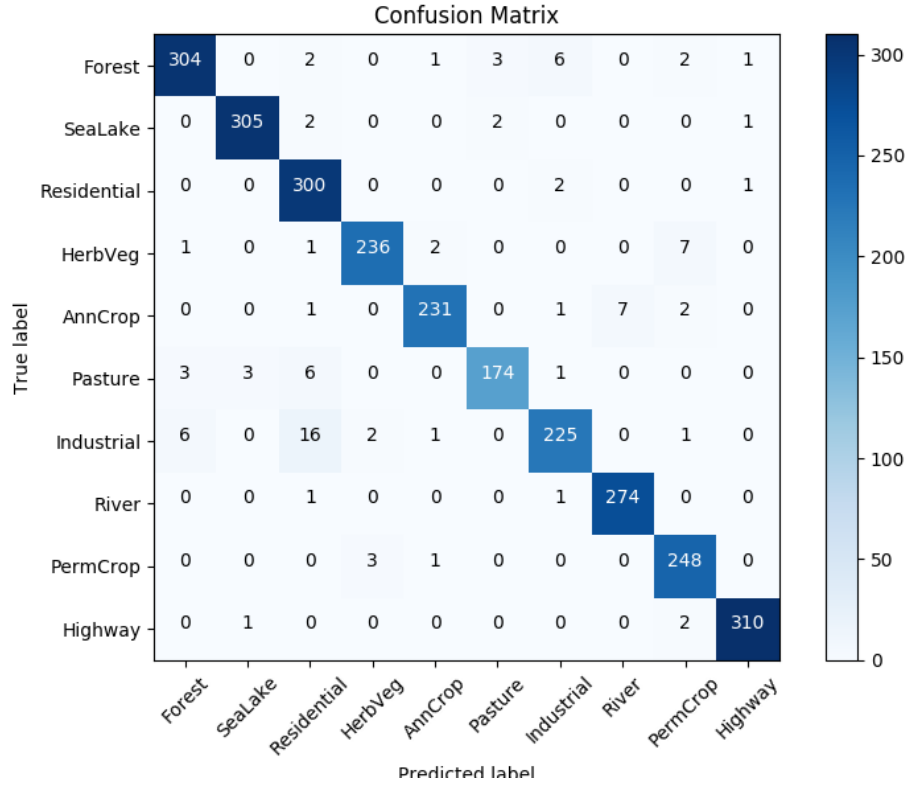


Figure 1: Confusion Matrix

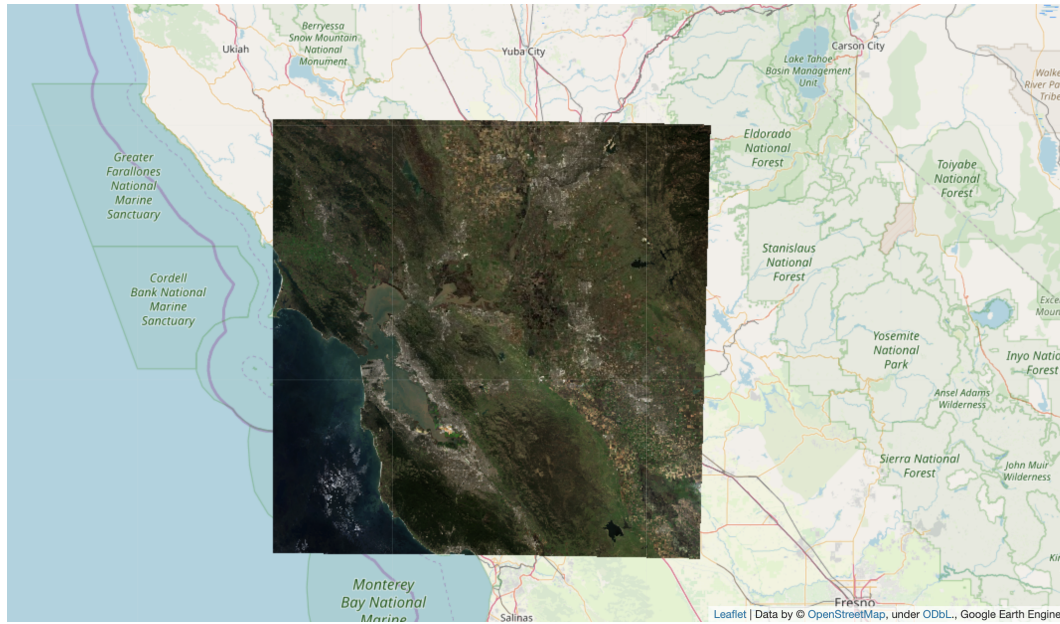
experimentation, we found that reducing the batch size helped reduce the overfitting by adding in noise to the training process. Examining the confusion matrix, it is clear that the model performs well on the test set and the majority of values lie along the confusion matrix diagonal. As evidenced by the value 16 in the confusion matrix, the models struggled a bit more to discern Industrial and Residential images, which lead to a slightly lower precision for the Residential class as seen in Table 3.

Dataset	Macro Precision	Macro Recall
Dev	0.97	0.97
Test	0.97	0.96

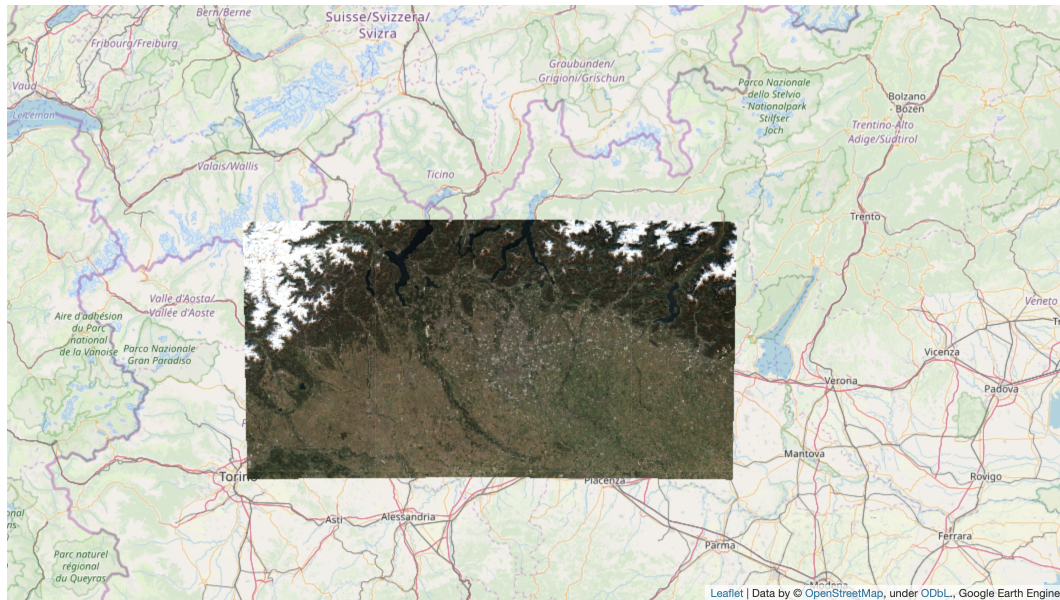
Table 2: Precision/Recall for Dev and Test

Label	Precision	Recall	F1
Forest	0.97	0.95	0.96
SeaLake	0.99	0.98	0.99
Residential	0.91	0.99	0.95
HerbaceousVegetation	0.98	0.96	0.97
AnnualCrop	0.98	0.95	0.97
Pasture	0.97	0.93	0.95
Industrial	0.95	0.90	0.92
River	0.98	0.99	0.98
PermanentCrop	0.95	0.96	0.96
Highway	0.99	0.99	0.99

Table 3: Test Set Metrics by Label



Region 1: Bay Area



Region 2: Northern Italy

6 Applications

We utilized Google Earth Engine to export new Sentinel-2A imagery as a TFRecord which we fed to our model for predictions. Sentinel has a revisit period of 2-10 days, depending on the location. That means around every week and a half, there is new imagery covering most of the world. We extracted data from two geographic regions and use our trained model to calculate their footprint. In both cases, we took all of the images taken by Sentinel-2A of that region from Jan 1 2018 to Jan 1 2019, and calculated median pixel for the True Color Image RGB bands in each image to get a composite for the entire year. The images were split into smaller 64x64x3 regions that we could feed to the model(see export_satellite_data.ipynb).

Our first application region spans the bay area, just north of Salinas and south of Petaluma. Because the EuroSat dataset contained only images from European regions, we were curious to see how the model would generalize to other parts of the world. This region received the following label distribution:

Forest	SeaLake	Residential	HerbVeg	AnnCrop	Pasture	Industrial	River	PermCrop	Highway
894	4978	10024	35	1020	42	846	54	251	732

Table 4: Region 1 Label Distribution

Region two covers Northern Italy, including Milan and the part of the Alps mountain range. This region received the following label distribution:

Forest	SeaLake	Residential	HerbVeg	AnnCrop	Pasture	Industrial	River	PermCrop	Highway
2890	180	2229	74	850	454	3276	16	38	1

Table 5: Region 2 Label Distribution

Using the predicted label distributions from our Resnet50 model, we can now calculate the total biocapacity of the region, but first, since the Global Footprint Network classifies land uses into 6 categories, we had to allocate our 10 classes to the six categories as follows: (1) Cropland (AnnualCrop + PermanentCrop), (2) Forest Land (Forest), (3) Grazing Land (Pasture + HerbaceousVegetation), (4), Marine Fishing Grounds (River), (5) Infrastructure (Industrial + Highway + Residential), (6) Inland Fishing Grounds (SeaLake).

We then use equation (1) to calculate Biocapacity . Equivalence factors are used to convert a specific land type into a universal unit of biologically productive area, a global hectare. Yield Factors account for differences between countries and years in productivity of a given land type.

$$Biocapacity = \sum (Area(\%) * EquivalenceFactor * YieldFactor) \quad (1)$$

Land Use Type	Equivalence Factor (g-ha/ha)	US Yield Factor (ha/ha)	Italy Yield Factor (ha/ha)
Crop Land	2.5	1.33295	0.786316
Forest Land	1.28	1.24176	1.67864
Grazing Land	0.46	0.723748	1.90792
Marine Fishing	0.37	1.05823	0.897117
Infrastructure	2.5	1.33295	0.786316
Inland Fishing	0.37	1	1

Table 6: Equivalence Factors and Yield Factors.[4]

The results of our analysis shows that the Bay Area (Region1) has a biocapacity of 2.45 g-ha/ha while Northern Italy (Region 2) has a biocapacity of 1.92 g-ha/ha. Thus, we can conclude that the land represented by Region 1 is more biologically productive than the land in Region 2.

7 Conclusion/Future Work

We were able to achieve high test set accuracy, precision, and recall with our model. For future work, we would experiment with transfer learning from satellite datasets, as suggested in the EuroSat paper. Additionally, examining model performance given different atmospheric conditions, cloud cover, or non European regions would quantify its generalizability. Future work would also include a study into the changing footprint of a region over time. We calculated the average biocapacity over an entire year, but Sentinel takes new images weekly, allowing for more granular analysis. Overall, our work demonstrates a high performing classification model with a real world application that can help researchers and policy makers understand environmental impacts and changes in natural landscapes over time.

8 Contributions

- **Nikita:** Data preprocessing for RGB and multispectral data, Resnet50 model experiments and results, Google Earth Engine pipeline for exporting the new Sentinel data, report writing
- **Rubi:** Baseline and VGG model experiments and results, biocapacity analysis, report writing and poster.
- **Xinyuan:** VGG model experiments, poster making.

Link to Github repository: <https://github.com/nikitamaia/landcover-sustainability>

References

- [1] Global Footprint Network (2019). National Footprint and Biocapacity Accounts: *NFA 2019 Calculations - Equivalence Factors and Yield Factors*. <https://www.footprintnetwork.org/resources/data/>
- [2] Kussul, N., Lavreniuk, M. , etal. (2017) *Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data*, Department of Space Information Technologies and Systems, Space Research Institute, National Academy of Sciences of Ukraine and SSA Ukraine, IEEE Geoscience and Remote Sensing Letters (Volume: 14 , Issue: 5 , May 2017)
- [3] Nijhawan, R. , Joshi, D., etal. (2018) *A Futuristic Deep Learning Framework Approach for Land Use-Land Cover Classification Using Remote Sensing Imagery*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEE
- [4] Helber, P., Biske, B., Dengel, A. & Borth, D., (2019) *EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (Volume: 12 , Issue: 7 , July 2019)
- [5] *Scikit-learn*: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.
- [6] *TensorFlow*: Large-scale machine learning on heterogeneous systems,2015.Software available from tensorflow.org.