Labeling Land:

Land Use Classification Using Convolutional Neural Networks

Claire Miles

Introduction

Background

- Satellite data is extremely useful
    a. Bird’s eye view of the world
    b. Help us more efficiently use natural resources and solve environmental problems

- **Problem:** satellite data is massive and confusing

Can we create a model that can classify the **landscape/land use** of a given image?

This Project

- Create a land classification model using a convolutional neural network
- Tools: Python - keras/tensorflow, rasterio, gdal
- Steps:
    a. Data preparation
    b. Exploratory Data Analysis
    c. Deep Learning Modeling

Data Collection

Methods

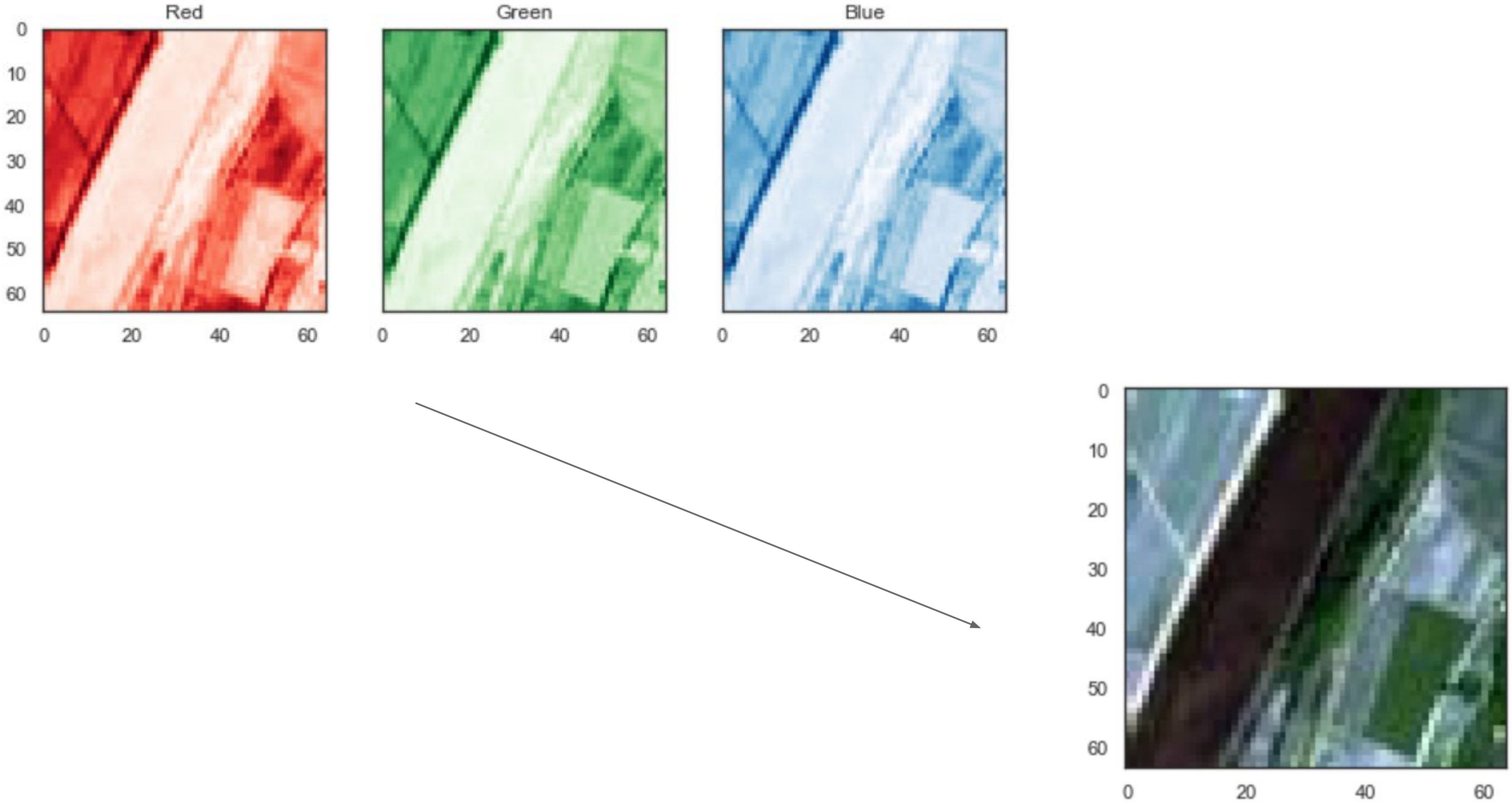
- Download data locally
- Translate files from TIF to JPEG format using gdal_translate
- Only extract red, green, and blue bands
- Separate data into training, validation and test sets

Satellite Data

- EuroSAT Sentinel-2 images from the German Research Center for Artificial Intelligence
- Ten different land use classifications
- 13 bands
  - Only RGB used in this project

Sentinel-2 Bands	Central Wavelength (µm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

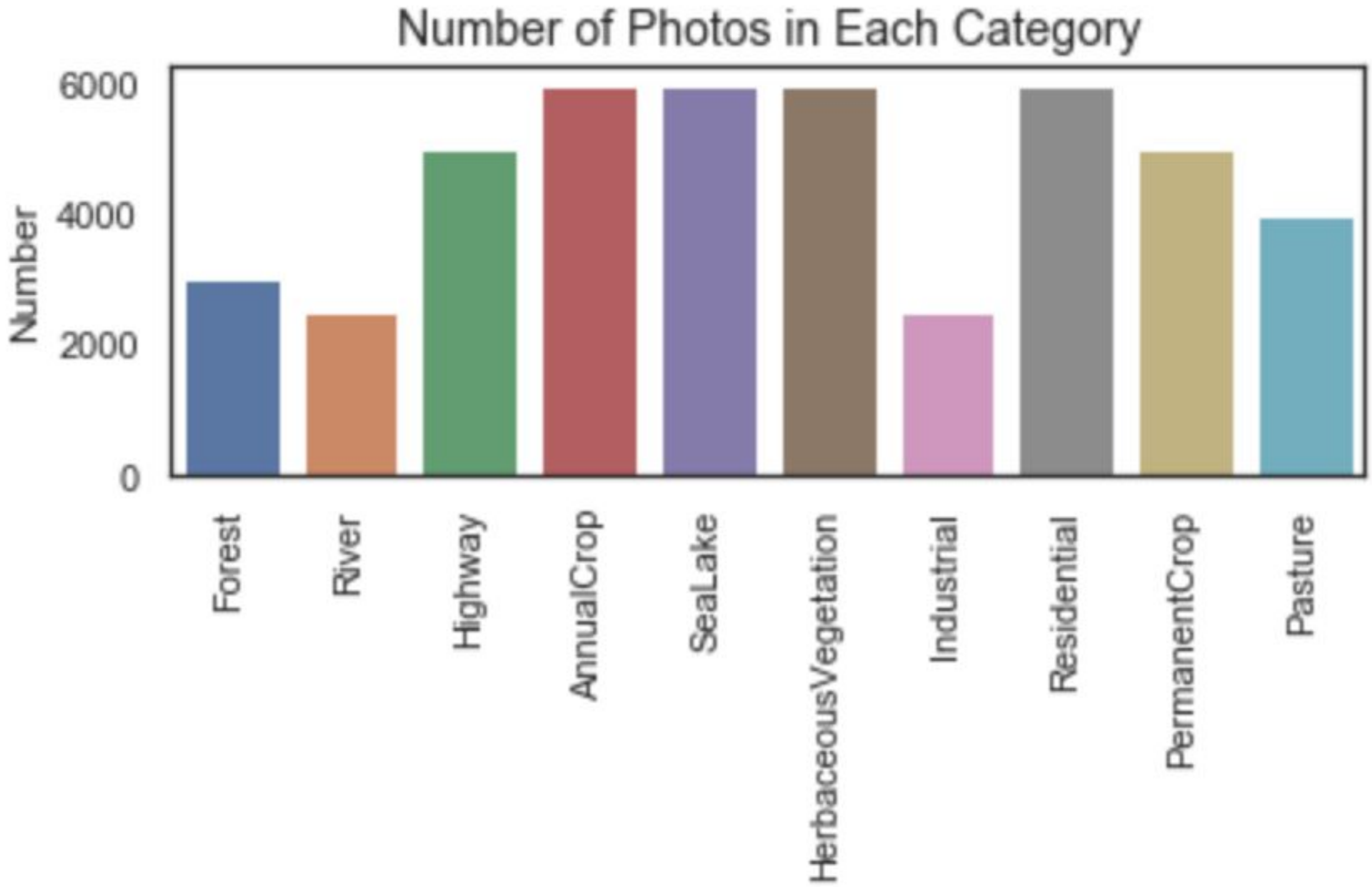
Sentinel-2 Bands, Wavelength, and Resolution

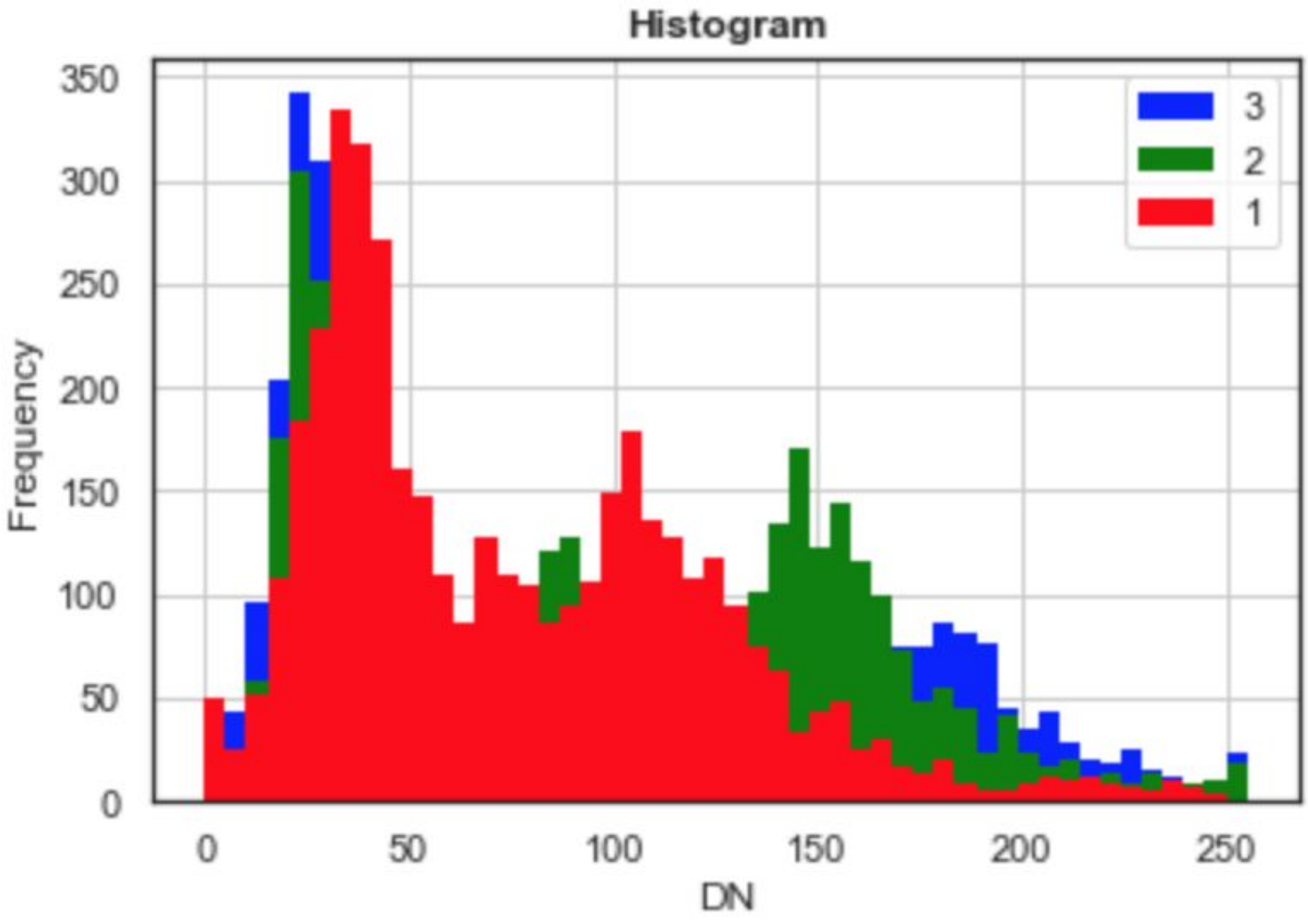

Red
Green
Blue

Exploratory Data Analysis

Methods

- Use histograms and bar plots to visualize data
    - Differences between categories
    - Differences between color bands


Number of Photos in Each Category
Number
6000
4000
2000
0
Forest
River
Highway
AnnualCrop
SeaLake
HerbaceousVegetation
Industrial
Residential
PermanentCrop
Pasture

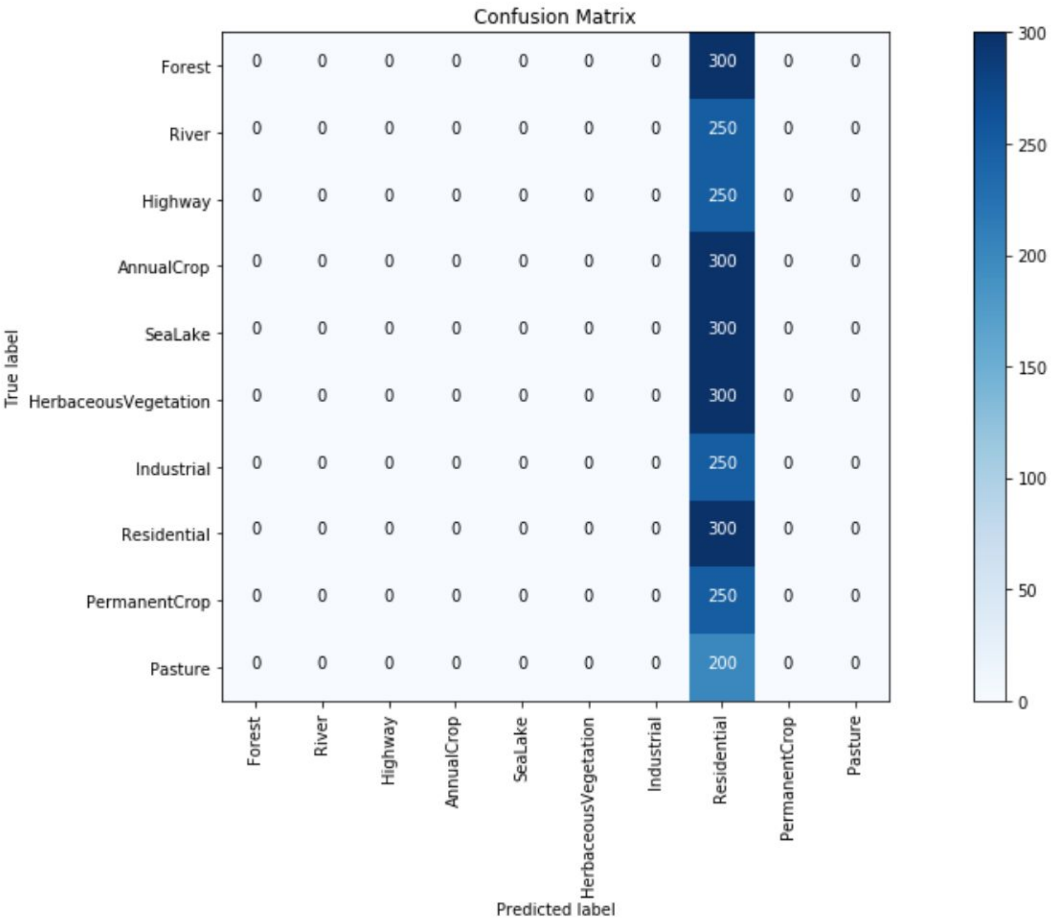


Histogram
3
2
1
Frequency
0
50
100
150
200
250
300
350
0
50
100
150
200
250
DN

Basic Convolutional Neural Network

Methods & Results

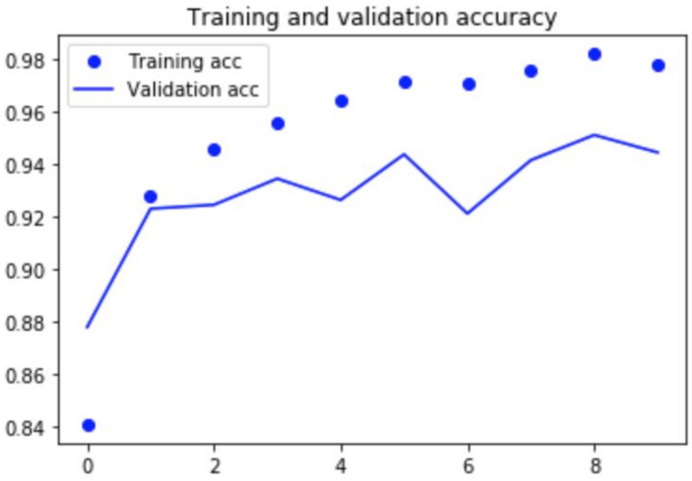
- One convolutional input layer, one flatten layer in the middle, and one dense output layer
- Good to test data preparation and parameters
- Not good for model performance
- **10%** accuracy == random guess

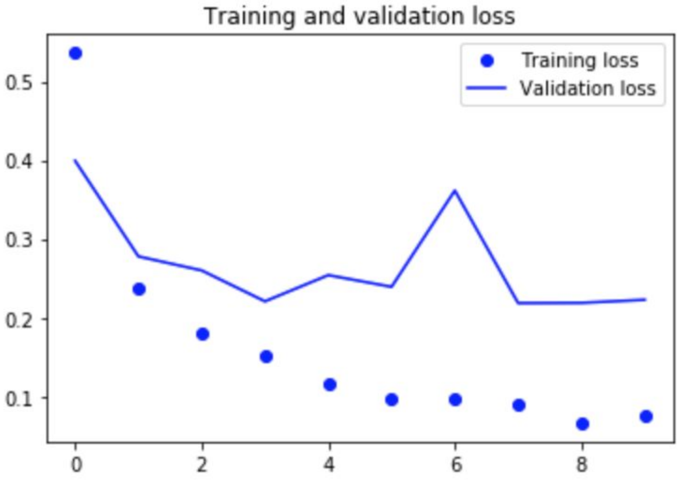

Confusion Matrix
True label
Predicted label
Forest
River
Highway
AnnualCrop
SeaLake
HerbaceousVegetation
Industrial
Residential
PermanentCrop
Pasture
0 0 0 0 0 0 0 300 0 0
0 0 0 0 0 0 0 250 0 0
0 0 0 0 0 0 0 250 0 0
0 0 0 0 0 0 0 300 0 0
0 0 0 0 0 0 0 300 0 0
0 0 0 0 0 0 0 300 0 0
0 0 0 0 0 0 0 250 0 0
0 0 0 0 0 0 0 300 0 0
0 0 0 0 0 0 0 250 0 0
0 0 0 0 0 0 0 200 0 0
300
250
200
150
100
50
0

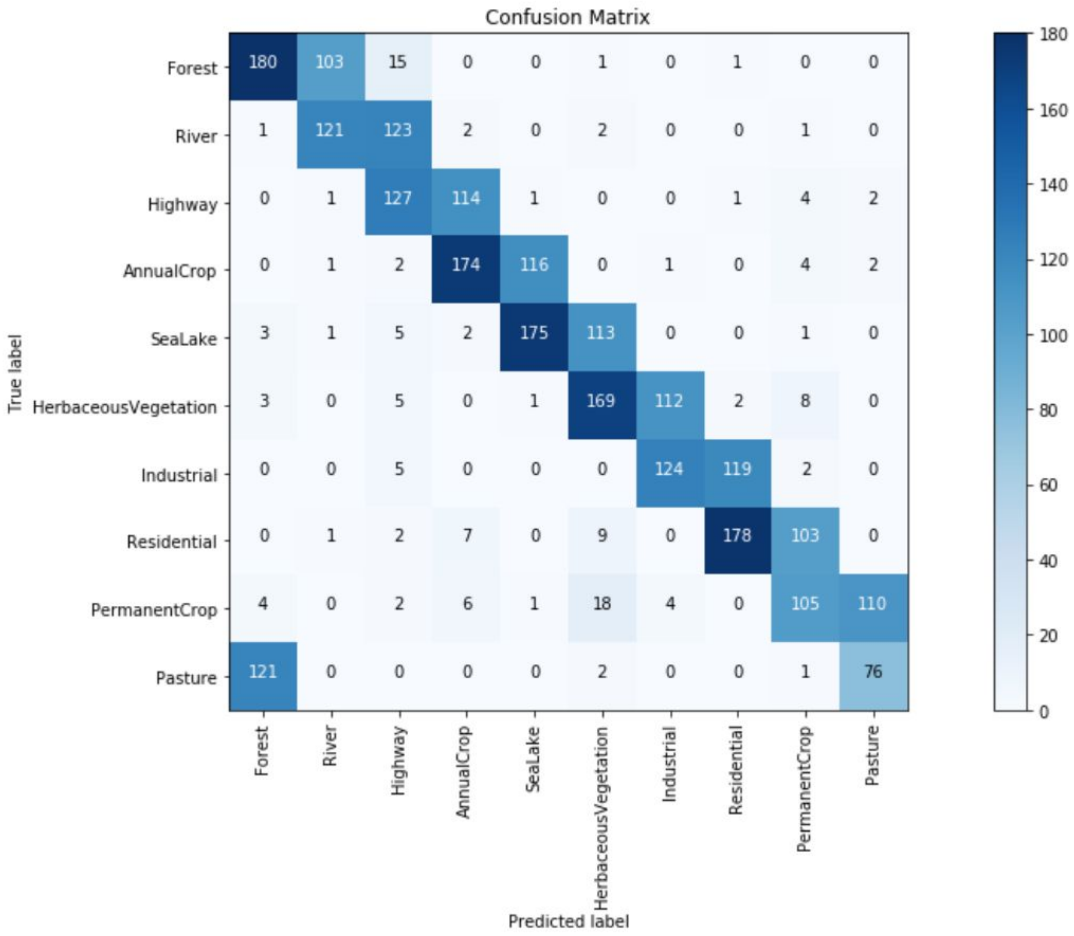
Transfer Learning

Methods & Results

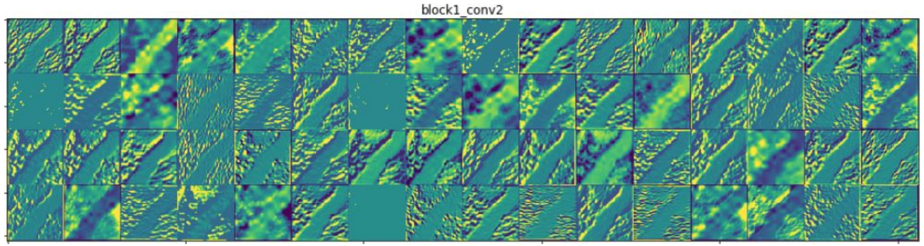
- Adapt a pre-trained model to this project
- VGG16 model, trained on 15 million images in over 22k categories
- Freeze first 12 layers, train last 10
- **94%** accuracy > random guessing


Training and validation accuracy
Training acc
Validation acc
0.98
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0
2
4
6
8


Training and validation loss
Training loss
Validation loss
0.5
0.4
0.3
0.2
0.1
0
2
4
6
8

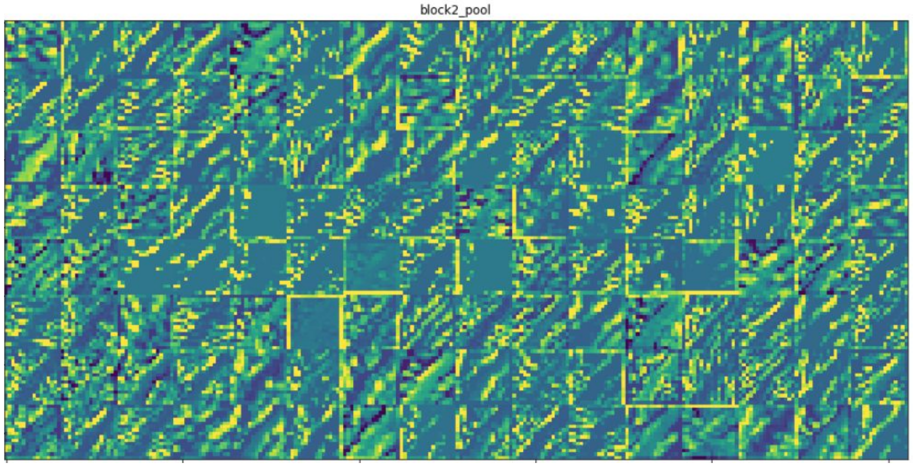


Confusion Matrix
True label
Predicted label
Forest
River
Highway
AnnualCrop
SeaLake
HerbaceousVegetation
Industrial
Residential
PermanentCrop
Pasture
180 103 15 0 0 1 0 1 0 0
1 121 123 2 0 2 0 0 1 0
0 1 127 114 1 0 0 1 4 2
0 1 2 174 116 0 1 0 4 2
3 1 5 2 175 113 0 0 1 0
3 0 5 0 1 169 112 2 8 0
0 0 5 0 0 0 124 119 2 0
0 1 2 7 0 9 0 178 103 0
4 0 2 6 1 18 4 0 105 110
121 0 0 0 0 2 0 0 1 76
180
160
140
120
100
80
60
40
20
0

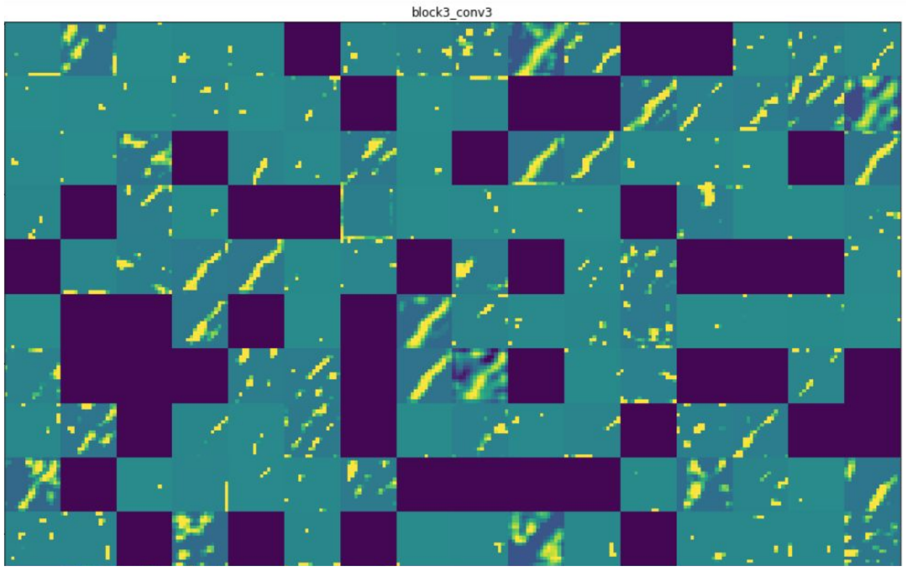


Convolutional layer early in the model

Pooling layer early in the model



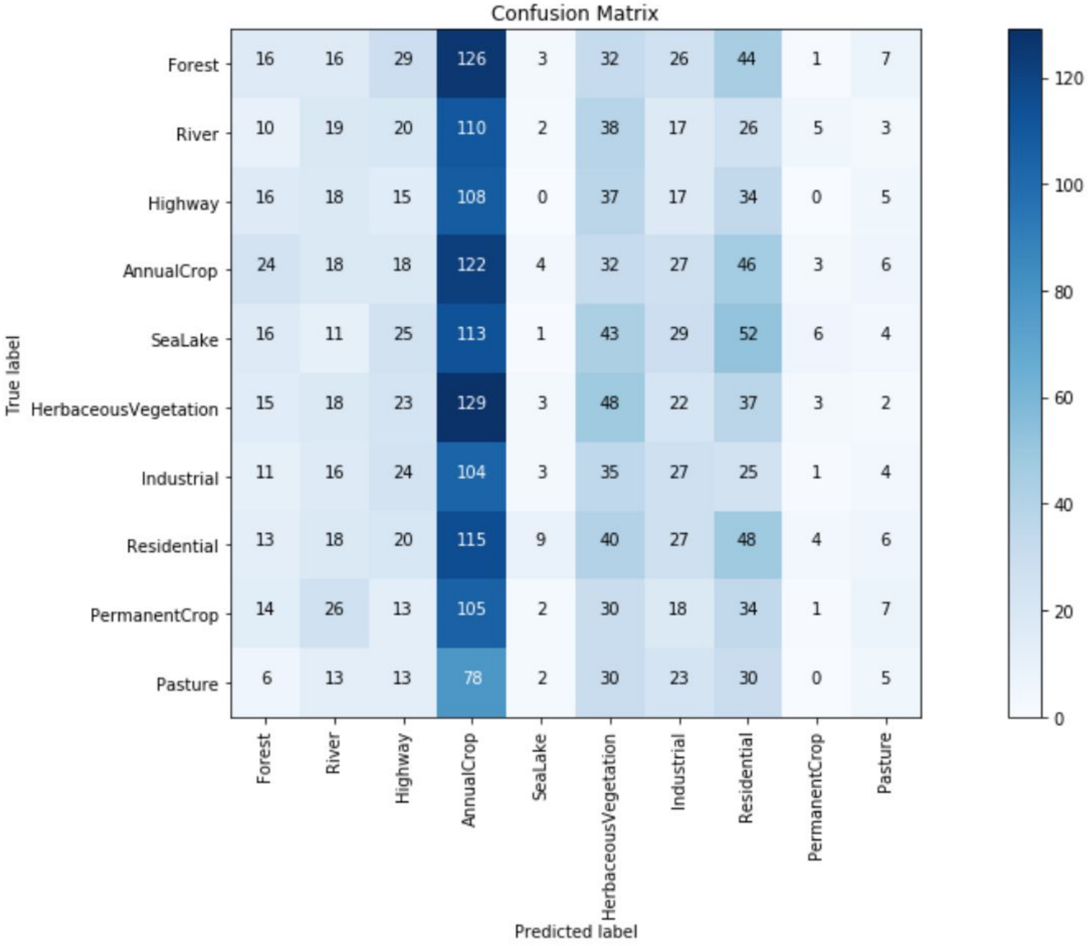
Convolutional layer
late in the model



Model Tuning

Methods & Results

- Data augmentation
    - Computational challenge to the CNN
    - Increases dataset size
- Batch normalization
    - Improves speed and performance of the neural network
- **56%** accuracy
    - Could increase with more epochs and higher learning rate


Confusion Matrix
True label
Predicted label
Forest
River
Highway
AnnualCrop
SeaLake
HerbaceousVegetation
Industrial
Residential
PermanentCrop
Pasture
16 16 29 126 3 32 26 44 1 7
10 19 20 110 2 38 17 26 5 3
16 18 15 108 0 37 17 34 0 5
24 18 18 122 4 32 27 46 3 6
16 11 25 113 1 43 29 52 6 4
15 18 23 129 3 48 22 37 3 2
11 16 24 104 3 35 27 25 1 4
13 18 20 115 9 40 27 48 4 6
14 26 13 105 2 30 18 34 1 7
6 13 13 78 2 30 23 30 0 5
0
20
40
60
80
100
120

Conclusions & Next Steps

Conclusions

1. Save on employee costs
    a. Money spent manually categorizing satellite imagery can be saved by having the model handle it
2. Use employee talents for projects that add value to the company
    a. Research and development
    b. Growth strategy
    c. Human-centered design

Future Explorations

1. Improve model tuning techniques
   a. Dropout regularization
   b. Weight initialization

2. Explore other base models for transfer learning
   a. MobileNet

3. Use satellite data from other sources to corroborate or improve the model

Thank you!
Questions?