Capstone Project 2 | Proposal
Claire Miles

Capstone Project Two: Land Use/Land Cover Classification Model

The Problem:
Satellite imagery can help us find solutions to the growing number of environmental problems that we face today. It allows us to not only get a bird's eye view of what's around us, but also uncovers parts of the world that are rarely seen. Tapping into the potential of categorizing land cover and land use around the world means that humans can more efficiently make use of natural resources, hopefully lowering cases of waste and deprivation. But despite its potential to be incredibly useful, satellite data is massive and confusing, and making sense of it requires complex analysis.

The Client:
This project could help a variety of stakeholders, including conservations, urban planners, and environmental scientists, survey and identify patterns in land use to see which natural areas are under threat or which areas are best for urban development. But in this case, I will tailor this project to a client that is similar to the agriculture imagery company that I currently work for. Imagery companies can use land use classification models to categorize what's in each image and optimize their efforts towards the parts of land that are important to them. For example, an agriculture company will only want to be concerned with land that's labelled as pasture, annual crop, or permanent crop. If a satellite is constantly taking images, this project would help save hours of time manually sorting through imagery.

In addition to sorting imagery, land use classification is important for identifying the parts of an image to which certain analyses are applied. For example, if the company has different crop stress algorithms for fields that are next to urban areas versus fields that are next to rivers, this project would help them to automate the process of applying these specialized algorithms and more effectively solve environmental problems.

The Data:
I will be using open source EuroSAT Sentinel-2 satellite images from the German Research Center for Artificial Intelligence, which can be downloaded locally [link]. The dataset consists of 27,000 labeled images of 10 different land use classes:

1. Annual Crop
2. Forest
3. Herbaceous Vegetation
4. Highway
5. Industrial
6. Pasture
7. Permanent Crop
8. Residential

9. River
10. Sea / Lake

Each multispectral image consists of 13 different color bands that represent different wavelengths of light/color and different resolutions. These different light bands help distinguish parts of the landscape that reflect certain types of light in particular ways. Having multiple bands of light in one image is standard for satellite imagery, so it can be assumed that the model from this dataset can be applied to data outside this particular collection.

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

(Image Source)

Method Outline:

1. Exploratory Data Analysis

Before modeling, I will familiarize myself with the image data by visualizing how many images are in each category in the dataset and look for the statistically significant differences in pixel intensity between categories and bands. Also, I plan to use random forests to analyze the importances of the images' different color bands. Hopefully combining these methods will help me to see which bands are the most distinctive in classifying an image. I will likely choose a few different combinations of bands to use in my modeling process.

Some of the resources I've seen indicate that I may need to do some preprocessing to convert the TIF files to something more compatible with some deep learning architectures. I may use the Rasterio or GDAL libraries to do this.

2. Modeling
For this project, I will be using convolutional neural networks to learn from the image data. CNNs are a particularly good neural network architecture for distinguishing the spatial arrangement of specific 2D features that can be indicative of different categories. Some of the libraries that I plan to use for deep learning include Tensorflow and it's user-friendly wrapper, keras.

I will also plan to explore transfer learning as an efficient way to create a model, especially since I don't have access to a powerful computer at all times. By only training the last few layers of a intentionally-chosen and proven model from another project, I can save a lot of time while still putting my own spin on the project. It would also be interesting to see how the out-of-the-box model performs next to my tuned one.

Deliverables:
The deliverables for this project will include the code and explanation in a series of Jupiter notebooks as well as an accompanying report, and slide deck.

Useful Resources:
Fast.ai forum and course
Land Cover Classification from Satellite Imagery
Land use/Land cover classification with Deep Learning and notebook
Land Classification w/ deep learning