**Final Project Draft01**

Title: Deep Neural Networks for Landcover Classification of the Colorado River Watershed

Notice: Dr. Bryan Runck

Author: Jake Ford

Date: 10/28/2022

**Project Repository:** https://github.com/ThisFord/GIS5571-arc1.git

**Google Drive Link:** n/a

**Time Spent:** 10hrs

**Abstract**

Machine Learning and the subset of algorithms known as Deep Neural Networks are increasingly used in image analysis and classification programs. This project uses the powerful computational advantages and predictive modelling capabilities of Deep Nerual Networks (DNNs) in combination with the resources available through the Minnesota Supercomputing Institute (MSI) to produce a high-resolution Land Cover and Land Use dataset from multiple input data layers. The Chesapeake Bay Conservancy has produced a model trained on satellite, lidar and planimetric data from counties within the Chesapeake Bay Conservancy Watershed; this project uses the CBC model in an automated deterministic workflow to create a reproducible model for a new input data set: a target county in the Colorado River Watershed. Data is tested for accuracy against a manually classified subset of the input imagery, which is isolated from the testing workflow. The resulting data set is a 1m resolution raster with land cover and land use classification for the entire county. Producing an accurate dataset will demonstrate potential scalability; with future plans scaling the model up to a watershed wide dataset with automated updates.

**Problem Statement**

High resolution spatial and temporal data is essential to modern conservation efforts. Current high resolution spatial data is cost prohibitive or nonexistent. The current National Land Cover Dataset has a 30m spatial resolution based on Landsat Imagery. Using object image analysis deep neural network classification models on NAIP imagery, in combination with lidar and other planimetric data, can produce accurate 1m resolution classified data in a programmable workflow, making change analysis possible at fine spatial and temporal scales. As climate change and drought conditions continue to impact the systems reliant on the Colorado River watershed this type of accessible and repeatable data becomes more and more necessary. This project will seek to replicate the Chesapeake Bay Conservancy Land Cover and Land Use workflow on data sourced from a county within the watershed with available lidar and NAIP imagery, producing a semantically segmented dataset with pixel-by-pixel classification for the entire county. Proving the workflow works on unique input data will allow for scaling to a watershed wide analysis.(Claggett et al., n.d.; Marcos et al., 2018; Zhu et al., 2017)

Table 1. Materials needed

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **#** | **Requirement** | **Defined As** | **(Spatial) Data** | **Attribute Data** | **Dataset** | **Prep** |
| 1 | Satellite Imagery | NAIP 1m resolution imagery of target area | Georeferenced Raster |  | [USGS](https://www.usgs.gov/centers/eros/science/national-agriculture-imagery-program-naip-data-dictionary) | Clip to target area |
| 2 | County boundaries | Authoritative county boundary data | Boundary lines and coordinates |  | Target municipality | download |
| 3 | Lidar Groundcover Imagery | Point cloud data for surface and object analysis for county | Coordinate point cloud | Density, reflectivity, surface type, cover type | Target municipality | Clip to target area |
| 4 | Object Identification DNN Model | Deep Neural Network model created by the Chesapeake Bay Conservancy for object identification from lidar, NAIP, and planimetric data |  |  | Chesapeake Bay Conservancy Model |  |
|  | Planimetric data | County wide planimetric attribute data for quality control and LULC prediction | Geolocated polygons of structures and human made infrastructure | Labeled and classed objects ie structures, roads | Target municipality |  |
| 5 | ArcGIS Pro License | Industry standard GIS analysis software package |  |  |  |  |
| 6 | Jupiter Labs | Python Notebook Interface |  |  |  |  |
| 7 | Supercomputing Server Access | MSI servers to run DNN on large data set |  |  |  | DNN model and script plus data sets |

**Input Data**

The classifier will use input data from multiple sources to create accurate classifications. The following data sources include planimetric, lidar, imagery, and areal information. The various data structures are used in preprocessing to create a data rich, multiband input raster that includes elevation, heights and vector boundaries along with digital numbers from the electromagnetic bands collected with the imagery.

Table 2. Data required

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Title** | **Purpose in Analysis** | **Link to Source** |
| 1 | NAIP Imagery | Raw input dataset for image object analysis | [USGS](https://www.usgs.gov/centers/eros/science/national-agriculture-imagery-program-naip-data-dictionary)  (*National Agriculture Imagery Program (NAIP) Data Dictionary | U.S. Geological Survey*, n.d.) |
| 2 | County boundaries | Clipping mask for input data | ArcGIS [OpenData](https://data-cdphe.opendata.arcgis.com/datasets/colorado-county-boundaries/explore?location=38.973275%2C-105.550600%2C7.55) |
| 3 | Lidar Groundcover Imagery | Object and surface attribute classification and identification | [Colorado Water Conservation Board](https://coloradohazardmapping.com/LidarDownload)  (*Colorado River Basin GIS Open Data Portal*, n.d.) |
| 4 | County Planimetric Data | QC and classification comparison | [Denver Open Data](https://www.denvergov.org/opendata/search?tag=planimetric)  (*Denver Open Data Catalog Search*, n.d.) |
| 5 | Denver County regional lidar | 1 ft resolution lidar data for object based image classification, elevation and height | [Denver Regional Council of Governments](https://drcog.org/services-and-resources/data-maps-and-modeling/regional-lidar-project) |
| 6 | Denver County regional planimetric | Vector data for infrastructure footprints | [Denver Regional Council of Governments](https://drcog.org/services-and-resources/data-maps-and-modeling/regional-planimetric-data-project) |
| 7 | Denver County land use landcover project | Verification and training examples | [Denver Regional Council of Governments](https://drcog.org/services-and-resources/data-maps-and-modeling/regional-land-use-land-cover-project) |
| 8 | Colorado River basin county boundaries | Study Area Boundaries | [Colorado River Basin Open Data Portal](https://coloradoriverbasin-lincolninstitute.hub.arcgis.com/datasets/lincolninstitute::colorado-river-basin-county-boundaries/explore?location=37.114380%2C-110.573590%2C6.86) |
| 9 | Chesepeake Bay Conservancy Project | Trained Classifier Weights (array) used for transfer learning on new input data | [GitHub](https://github.com/stactools-packages/chesapeake-lulc) |
| 10 | Eurosat Data | Well established labeled training data set for use setting up the Classifier Workflow | [Tensor Flow Datasets](https://www.tensorflow.org/datasets/catalog/eurosat) |

**Methods**

The Chesapeake Bay Conservancy has developed a robust methodology to develop high resolution 1m Land Use/Land Cover data for the Chesapeake Bay watershed from NAIP imagery using a predictive DNN model.(*Chesapeake Bay Program Land Use/Land Cover Data Project*, n.d.) This project will be replicating the CBC methodology for a single county in the Colorado River watershed. This includes assembling lidar and planimetric data to train the DNN model and set classification attributes, combining this data with 1m NAIP imagery and classifying pixels with the DNN, based on pre-determined landcover and land use classifications. (Claggett et al., n.d.)

**Diagram

Description automatically generated**

*Figure 1. Data flow diagram of a machine learning workflow for image classification. (Hagerty, 2016)*

*Graphical user interface

Description automatically generated*

*Figure 2. An early iteration of the classifier using Google Colab and Tensorflow with Eurosat data.*

**Text

Description automatically generated**

*Figure 3, an example of a submit script to run the classifier on the MSI clusters*

**Results**

The resulting data set is a 1m resolution raster with land cover and land use classification for the entire county.

Map

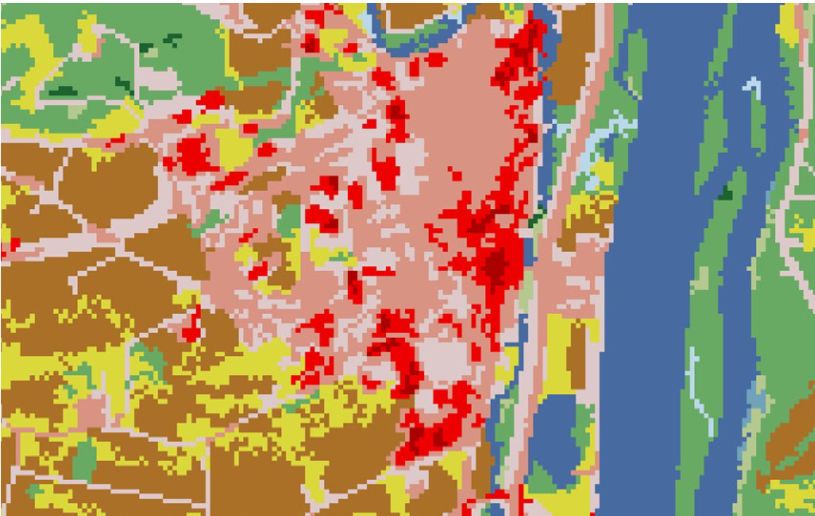
Description automatically generated

*Figure A: Example land cover/land use classification map from the Chesapeake Bay Conservancy*

*Graphical user interface, text, application

Description automatically generated*

*Figure B: an example of custom landcover classification symbology* (*Chesapeake Bay Program Land Use/Land Cover Data Project*, n.d.)



*Figure C: example land use classification from Chesepeake Bay* (*CIC\_high\_resolution\_data\_services.Pdf*, n.d.)

Map

Description automatically generated

*Figure D: example land cover classification from Chesapeake Bay* (*CIC\_high\_resolution\_data\_services.Pdf*, n.d.)

**Results Verification**

Results will be verified by comparing a manually classified set of 100 random 27x27 pixel clippings that are removed from the input data set against the predicted output. The Chesapeake model was able to achieve a 91% average accuracy across its study area. Similar results are expected. (Claggett et al., n.d.; *Conservation Innovation Center*, n.d.)

**Discussion and Conclusion**

Proving that the CBC model is adaptable for other water sheds at the county scale will open the door for further work on the entire Colorado River watershed, allowing for change detection and monitoring at high temporal resolutions. The reproducibility of the process will allow for further development of multiple time periods, data essential for monitoring the impact of the continuing drought and climate change on a valuable and fragile watershed. (*R45546.Pdf*, n.d.)

**References**

*Chesapeake Bay Program Land Use/Land Cover Data Project*. (n.d.). Chesapeake Conservancy. Retrieved September 24, 2022, from https://www.chesapeakeconservancy.org/conservation-innovation-center/high-resolution-data/lulc-data-project-2022/

*CIC\_high\_resolution\_data\_services.pdf*. (n.d.). Retrieved September 25, 2022, from https://www.chesapeakeconservancy.org/wp-content/uploads/2017/02/CIC\_high\_resolution\_data\_services.pdf

Claggett, P., Ahmed, L., Buford, E., Czawlytko, J., MacFaden, S., McCabe, P., McDonald, S., O’Neill-Dunne, J., Royar, A., Schulze, K., & Walker, K. (n.d.). *Chesapeake Bay Program’s One-meter Resolution Land Use/Land Cover Data: Overview and Production*. 61.

*Colorado River Basin GIS Open Data Portal*. (n.d.). Retrieved September 25, 2022, from https://coloradoriverbasin-lincolninstitute.hub.arcgis.com/

*Conservation Innovation Center*. (n.d.). Chesapeake Conservancy. Retrieved September 24, 2022, from https://www.chesapeakeconservancy.org/conservation-innovation-center/

*Denver Open Data Catalog Search*. (n.d.). Retrieved September 28, 2022, from https://www.denvergov.org/opendata/search?tag=planimetric

Hagerty, P. (2016, September 4). *Establishing a Machine Learning Workflow*. Medium. https://medium.com/the-downlinq/establishing-a-machine-learning-workflow-530628cfe67

Marcos, D., Volpi, M., Kellenberger, B., & Tuia, D. (2018). Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models. *ISPRS Journal of Photogrammetry and Remote Sensing*, *145*, 96–107. https://doi.org/10.1016/j.isprsjprs.2018.01.021

*National Agriculture Imagery Program (NAIP) Data Dictionary | U.S. Geological Survey*. (n.d.). Retrieved September 24, 2022, from https://www.usgs.gov/centers/eros/science/national-agriculture-imagery-program-naip-data-dictionary#acquisition\_date

*R45546.pdf*. (n.d.). Retrieved September 25, 2022, from https://crsreports.congress.gov/product/pdf/r/r45546

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307

**Self-score**

Fill out this rubric for yourself and include it in your lab report. The same rubric will be used to generate a grade in proportion to the points assigned in the syllabus to the assignment.

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Description** | **Points Possible** | **Score** |
| **Structural Elements** | All elements of a lab report are included **(2 points each)**:  Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score | 28 | **28** |
| **Clarity of Content** | Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**. | 24 | **24** |
| **Reproducibility** | Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified. | 28 | **28** |
| **Verification** | Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**. | 20 | **20** |
|  |  | 100 | **100** |