Project title and student names

- **Project title:** Using satellite imagery to train a model for identifying the type of landmarks.
- **Student names** (1)
  - Leomar Durán

Revision History

Revision #	Author	Revision date	Comments
1.10.0	Leomar Durán	2022-03-25t23:57	design and implementation challenges, started anticipated project outcomes and impacts
1.9.0	Leomar Durán	2022-03-25t23:46	data science fundamentals, project objective and constraints
1.8.0	Leomar Durán	2022-03-25t22:31	finished overview of problem, flush level 1 lists
1.7.0	Leomar Durán	2022-03-25t22:19	moved description of dataset to proposed work
1.6.0	Leomar Durán	2022-03-25t22:16	added revision history
1.5.0	Leomar Durán	2022-03-25t21:47	added more about the dataset, grammatical fixes
1.4.0	Leomar Durán	2022-03-25t02:45	started bibliograph, problem statement
1.3.0	Leomar Durán	2022-03-25t02:10	motivation, flushed heading numbers into margin
1.2.0	Leomar Durán	2022-03-25t02:05	sections, title, students
1.1.0	Leomar Durán	2022-03-25t01:55	page specifications
1.0.0	Leomar Durán	2022-03-25t01:51	starting proposal

1 Introduction section

1.1 Motivation

The sciences of geomatics and land surveying interest me as hobbies. I really enjoy the idea of collecting data about the terrain, whether it be rural or urban, and working with that data to find solutions to problems or even just for fun.

1.2 The Problem

1.2.1 Overview

An image of a terrain is given to a computer which will make a decision on the fly based on the type of terrain. We will train a model that will be used by the computer to identify this terrain.

This sort of decision might be involved in deciding if the terrain would be appropriate for developing a building thereon. A preliminary sweep by a machine may save on costs of having an engineer waste time looking for land to develop. Another example of making this decision may be helpful for automatic landing software that will be used to safely land aircraft on stable terrain.

A third example is that combined with time series data, we can predict different types of natural weather-related phenomena, such as draughts, floods and earthquakes.

1.2.2 Data science fundamentals

This problem involves multi-class classification. We will classify the terrain according to features such as whether the area is urban, densely forested, mountainous, or contains water for 4 disjoint classes.

We will evaluate the results using accuracy, recall, precision and the F1 score.

1.2.3 Project objective and constraints

For this project, we hope to train a model to learn different 4 disjoint classes of terrain, and then classify a test sample.

The algorithm that we pick has to deal well with the curse of dimensionality, as there will be $(28 \times 28)\text{px/examples} \times 4\,\text{channels/px} = 3136\,\text{channels/examples}$.

Ideally the solution would also perform well for multiple clases, but this is less of an issue than dimensionality.

1.3 Related works

2 Proposed work

We are given The SAT-4 airborne dataset[1]. This data is hosted by the Louisiana State University's Division of Computer Science and Engineering[1] and can be downloaded directly from the Google Drive[2] along with the SAT-6 airborne dataset.

The dataset consists of 400,000 example tiles taken from satellite imagery originally from the National Agriculture Imagery Program (NAIP) dataset. Each sample has features representing the pixels of a (28×28)px image multiplied by the channels for red, green, blue and near infrared (NIR). According to Basu, Ganguly, Mukhopadhyay, *et al.* [1], these tiles represent "different landscapes like rural areas, urban areas, densely forested, mountainous terrain, small to large water bodies", so these as a disjoint set of landscapes would make for appropriate labels.

Our proposed solution is a multi-class logistic regression.

2.1 Design and implementation challenges

A challenge to this solution is the size of the dataset. Because of its size (about 3 Gbyte), we expect long processing times. One possible solution to this challenge may be to reduce the datasize from 400,000 to a more reasonable number such as 20,000.

Another issue that we will run into is deciding the best way to split the classes for the multi-class classification.

2.2 Anticipated project outcomes and impacts

An anticipated outcome is a model that can identify the types of terrains accurately from the given dataset.

[1] <http://csc.lsu.edu/~saikat/deepsat/>

[2] <https://drive.google.com/u/0/uc?export=download&confirm=sWVM&id=0B0Fef71_vt3PUkZ4YVZ5WWNvZWs>

3 Timeline

4 References

[1] S. Basu, S. Ganguly, S. Mukhopadhyay, R. Dibiano, M. Karki, and R. Nemani. “Deepsat - a learning framework for satellite imagery, acm sigspatial 2015.” (2015).