

# COAL: Coal and Open-pit surface mining impacts on American Lands

Taylor Alexander Brown, Heidi Ann Clayton, and Xiaomei Wang

Group 18

CS 461: Senior Capstone Fall 2016

Oregon State University

## **Abstract**

Mining is known to cause environmental degradation, but software tools to identify mining impacts are lacking. Researchers studying this problem possess large imaging spectroscopy and environmental quality data sets as well as high-performance cloud-computing resources. This project provides a suite of algorithms using these data and resources to identify signatures of mining and correlate them with environmental impacts over time.



**CONTENTS**

|          |                           |          |
|----------|---------------------------|----------|
| <b>1</b> | <b>Purpose and Goals</b>  | <b>3</b> |
| <b>2</b> | <b>Taylor's Progress</b>  | <b>3</b> |
| <b>3</b> | <b>Heidi's Progress</b>   | <b>5</b> |
| <b>4</b> | <b>Xiaomei's Progress</b> | <b>6</b> |
|          | <b>References</b>         | <b>8</b> |

## 1 PURPOSE AND GOALS

Our project is COAL: Coal and Open-pit surface mining impacts on American Lands. The purpose of our project is to create a suite of algorithms that will identify metrics to identify and classify minerals, identify mining activity based on those minerals, correlate effects on the environment with mining, and analyze those effects over time.

The deliverables for this project are an API that contains the algorithms for the previously mentioned image processing and correlation tasks that is contained in a GitHub repository, autogenerated API documentation using Sphinx, and a website. The website will contain information about the project and its development as well as links to the GitHub repository and API documentation as well as developer profiles (which also link to our individual GitHub repositories) and background research papers.

We intend for our suite of algorithms to be used by environmental data scientists, people involved in environmental regulation, and anyone else with an interest in mineral identification and hyperspectral imagery.

## 2 TAYLOR'S PROGRESS

My contributions to the project this term include research, discussion, algorithm design, code implementation, project maintenance, document editing, and assignment submission. I have contributed to the team's basic understanding with sketches of our algorithms, extensive mailing list postings, and frequent team meetings. I implemented detailed pseudocode and skeleton code for mineral identification training and classification, mine identification training and classification, environmental preprocessing, and environmental correlation. I assisted with other team members' components as necessary.

During the first week of the term I focused on learning about machine learning. In addition to readings online I watched two MIT OpenCourseWare lectures which provided a detailed-enough understanding of neural networks to begin thinking about how to use them for classification tasks. I shared my findings in an extensive discussion on the team mailing list and I posted the first set of sketches illustrating our understanding of the problem. I also continued to review the literature provided by the client.

After learning about machine learning I turned to setting up a development environment and becoming intimately familiar with the internals of the AVIRIS data format. I repurposed an old gaming PC with a GPU at home and configured it as a development environment. I later set up the machine for remote access via reverse ssh through my personal server so that the team may use it to execute long-running tasks. I downloaded the AVIRIS files to the machine and made a detailed exploration of the file format, including the structure of the ENVI header files and the binary format of the imagery. I continued to share my findings with the team over the mailing list and evolved a basic pseudocode for mineral classification. As I began to convert the pseudocode to a more realistic format I became familiar with the Spectral Python library which implements a large set of algorithms for processing the imagery. I downloaded the USGS Digital Spectral Library 06 and found that it is also compatible with AVIRIS and Spectral Python, however I noted that there were some discrepancies between the formats such as the number of bands and the wavelength range. I supplemented my basic knowledge of Python by skimming "Python Essential Reference, Fourth Edition," which helped to demystify some of the language's unfamiliar aspects.

I continued to evolve the pseudocode by referring to the Spectral Python documentation to find methods to solve the mineral classification problem. I learned that rather than using an external library, we can take advantage of a built-in perceptron classifier (neural network) that is provided by Spectral Python. This eased development, however

we may need to optimize the classifier in the future because the current implementation does not take advantage of GPU acceleration. I posted another set of sketches illustrating each of the stages of our data-processing pipeline at a more detailed level. The team has been referring to these sketches to visualize what are relatively complex algorithms. I found a way to save memory by iterating over images pixelwise since I am unable to load the full 6+ GB data sets into memory on my development machine. After providing a detailed skeleton, I turned over the rest of the mineral classification task to Heidi who is responsible for this component.

After mineral classification I turned my attention to my own components. Although environmental preprocessing and correlation are farther down the data pipeline, the parts of the problem to do with geographic information could be implemented relatively independently. I conducted research into GIS applications and libraries and began searching for candidate data sets.

With the input of the client, we chose to use the Geospatial Data Abstraction Library (GDAL) to load and manipulate georeferenced files in both vector and raster format. This will allow us to import both GIS data and imagery using the same API. It also allows us to avoid proprietary GIS applications. I posted another sketch detailing my concept of projecting the GIS data over the image to generate correlations. I found that it made the most sense to convert to a uniform raster format which will allow application programmers to write correlation functions that operate over pixels, regions, or whole images. I contributed work-in-progress code for GIS preprocessing and environmental correlation. In addition, I wrote the initial skeleton of mining training and classification, the module for which Xiaomei is responsible. I collaborated with both Xiaomei and Heidi to contribute to their efforts with the website and the mineral classification. Our client shared another spectral library, ASTER, with the team which will provide more training samples for mineral classification. In our team meeting with the client we decided to pursue a simple analysis of mining impacts on water resources as a case study. Environmental data scientists using our library will be responsible for implementing more sophisticated analyses.

For the midterm report I completely rewrote my sections of the design document to reflect our evolved understanding of the problem and algorithms. I also updated my tech review sections to describe the choices we had made. I also collaborated with Xiaomei and Heidi on the slides. I took responsibility for merging changes from our collaborative editing space, Overleaf, into version control. I also took responsibility for adding images to the slideshow, collecting the team's recordings and combining them with the slides to make the video, and making the final project submission. Heidi has taken care of posting revisions to OneNote.

At this point in the project we are slightly behind schedule. It would be nice to have a more complete implementation of mineral training and classification, but that turned out to be one of the most complex parts of the project. Although it has been possible to independently implement some of the later modules such as the geographic information manipulation for training the mining classifier and conducting environmental correlations, it will soon be necessary to work with some classified data to implement the mining classifier and environmental correlation steps. Another challenge has been to locate suitable sources of geographic information about mining permit boundaries and environmental data sets. Now that we have settled on a case study it has been possible to narrow our search. I intend to redouble my efforts to implement the components for which I am responsible as well as assist with development anywhere else it is needed. I look forward to a quality, if not fully mature, implementation at the end of the project. The need to understand impacts of mining on watersheds is especially relevant today, and I hope that this library will provide environmental data scientists at JPL and beyond with a quality tool for research.

### 3 HEIDI'S PROGRESS

During winter break, we did not work on our project. However, as soon as winter term began, our mailing list became active again like we were never gone. The first three weeks of the term were spent on researching the overall problem in order to get a more thorough understanding of how exactly we could implement our library. Taylor found a very helpful series of MIT lectures on neural networks that helped us a lot in understanding how neural networks are used and confirmed that they would be a good choice for this project.

After the initial research stage was finished, I was able to get started on the requirements that I am in charge of. I am in charge of the feature extraction and classification stages as well as maintaining the documentation. Currently, most of our code is in the development stage and isn't entirely functional, so there isn't a lot of documentation to maintain at this point. However, there are some examples of Sphinx-compatible Python docstrings in my personal development branch in the mineral classification file [1]. Once I get a working version of the feature extraction and classification, I plan on working on the documentation more and adding new pages, e.g. a getting started page.

As far as the feature extraction and classification of minerals, it is still a work in progress. Taylor put together a skeleton and I have been working on making it work and filling in the specific details. We decided that we will stick with Spectral Python for the implementation of these features, so that is what I am using to work on my requirements.

As per the technology review and design document, I am going to implement my requirements using a neural network. At first, we were looking at many different libraries to try and use with Spectral Python using an interface. Since last term, we were looking at neural network libraries such as TensorFlow [2] and Caffe [3]. However, Taylor discovered that Spectral Python has built-in neural network functionality, so this saved me a lot of headaches down the line.

We also decided we were going to stick to using the USGS Digital Spectral Library in order to train the neural network when looking at the spectral signatures of minerals and other land surfaces. When looking into this, we discovered that there was a compatibility issue with the library in the format we downloaded and the AVIRIS data we were going to be processing. There was a discrepancy in the representation of the data. AVIRIS uses integers and the USGS library we downloaded used floating point numbers. There was also a difference in the number of layers of data for each mineral in the USGS library compared to the AVIRIS pixels. Luckily, we were able to resolve this easily as there ended up being an alternative format of the library that is more compatible with AVIRIS images. The only remaining compatibility issue is that our USGS file represents wavelength in micrometers and our AVIRIS images represent wavelength in nanometers. Unit conversions should not present much of a challenge if the rest of the code is well-written, so it is not something we are terribly worried will cause problems as long as it is addressed.

So far, it seems that the working implementation of the feature extraction and classification of minerals will not be a terribly complicated file, as it is mostly composed of calling the Spectral Python library functions and manipulating data as needed, but so far I have been having issues with using and understanding the correct parameters needed, so that has been a bit of a setback.

It is very important that I get my requirements satisfied soon as other parts of the project are dependent on parts of its implementation. Through the end of the term, we will be meeting Mondays and Fridays at the very least and during those times, we will be helping each other out with our individual sections. I work well with occasional paired programming, so with this change I am optimistic that we will not fall far behind.

The main challenges for me have just been difficulty in getting the Spectral Python functions to work the way I want

them to. It was also a challenge at the beginning of the term to meet with our client because he was so busy, but we have started weekly meetings again and they are going smoothly. Another challenge for me was trying to keep up with our project while balancing my classes. As a result, I ended up dropping a class. So far, I am happy with that decision as that class was not very interesting to me and now I have more time to dedicate to working on this project. I was also ill during weeks five and six which made development slower on my end.

For the rest of the term, my plan is to make sure the feature extraction and mineral classification implementations are functional and complete. Then, I will work on making the Sphinx documentation look nice and well-organized. Additionally, I will help out my teammates whenever they need it on their assigned portions of the project. I will also be doing my best to keep my portions of the senior capstone documents up-to-date so it isn't a mad scramble to get them done in the future. I am currently the official team leader and signed us up for the Engineering Expo in May, so I will also be making sure to stay on top of those deadlines in the future.

Our client also decided that we would have team leaders (not to be confused with the official team leaders that signed up for Expo) that would rotate every two weeks. We started this during week five. The last two weeks will be my turn, in which I will be making sure the project board we have on GitHub is up-to-date and that our meeting agenda is filled out in our shared meeting notes document before every Friday meeting with our client.

#### **4 XIAOMEI'S PROGRESS**

After the winter break, I spent maybe the first two or three weeks to review our project, and also the related papers and literature; this is because our group did not meet during the break and we do not have any chances to discuss about the relevances of our project. For me, I traveled through the entire winter break, so I did not have a good opportunity to dive deeper than what we already got from the last term. That is why we all spent the first two or three weeks to review. Also, We did not get in touch with our client until the Friday of the third week; that is one of the reasons why our group was kind of slow-spaced at the beginning of this term.

Right now, we divided our project into several phases: extraction and classification of AVIRIS data, mineral identification; mining identification; and preprocessing data to correlate mining with environmental impact. Until the final step listed, our project would be considered appropriate for further use. Also, our project website, which is responsible for tracking our progress and also store our data, is updated according to our needs. We are at the initial part of our plan: Taylor has been researching and incorporate the GIS data as well as the environment impact part; Heidi has been doing the mineral identification part; I have continued to catching up their progress, and also is responsible for the website development part. Everything is considered going well by now.

At the first and the second week, all of our group members watched lectures about neural networks, which is suppose to be in use in the mineral and the mining identification as we concerned. We are using Spectral Python to implement those features of mineral identification; it has a built-in neural network functionality. We have discussed multiple times though out this term, but in the end, we decided to use USGS Digital Spectral Library to train the neural network to search for signatures of minerals and other land components. Because the mineral identification is the very first part of our plan, we can start to worry about other things such as mining identification of the following once the mineral part is done. According to the technology review from the last term, I suppose to be writing the mining identification part with the help of my group members. With the skeleton of the mineral identification, the mining part should take less time; still, I cannot really predict what kind of difficulties would I face by then.

The primary segment I am in charge of in the past five weeks is our project website. The one I set up the last term is pretty basic and needs a lot of improvements. Taylor had helped to fix some text errors and also suggested more advises about the content and the background. I added our final presentation from the fall term to our main page. Also, I changed the background page to fit our project theme. One problem is that most of the images are too colorful as a website's background; until I think of a solution, I have not decided to change the background with the picture provided. In the end, I find a way to make the background blurrier so it can be more compatible with other content, especially the text.

Another task I did is to add a page that can access all the paper we wrote from the last term, so it is more convenient for our group members to access them. The function is not hard to implement; when I update our website, the styling part is the one that typically cost more of my time. The underlying skeleton of the site right now is very basic HTML; I might change it somehow to make it more efficient and easy to read as well as easy to update. Currently, since the last update, the site provide links to our project description from the course website, the links to our GitHub pages, the links to our wiki pages, the links to our Google Group's discussion page, the links to the Github pages for all of our group members, and the links to our group papers as well as the literature we read during our project development. The purposes our site serves now is to guide us to different resources quicker rather than to show our project.

One thing yet need to add to the website is the Blog function; this is the main places where our members to upload blogs to show the results as the project advanced. The Blog supposes to be one of the most important parts of our website; in order to implement it, I need to use Jekyll, a static site generator, to achieve this. I have never used Jekyll before, although it is quite easy for the blog function, I still need more time than I spent on the other features to make it available. At week five, after the meeting with our client, Taylor and I stayed at the school library to research how to use it. Basically, download Jekyll to the local computer is more convenient for the further development of the website. I need to change the file skeleton to fit the requirements that Jekyll provides.

I have encountered several challenges as the project advanced; the primary one is the learning curve of this project make me hard to follow. There are not many ways to overcome this except read more materials. I need to continue to do that for the rest of the term as well. Our client has appointed a rotational group leader for every two weeks; this is a good chance to get familiar with everything. I am expecting to handle the code about the mining identification very soon, and I hope once I got into that, things would become easier for me. Our team functions well by now, and I hope we can overcome more difficulties and finish this project well.

## REFERENCES

- [1] <https://github.com/heidiaclayton/coal/blob/mineral-id-dev/coal/mineral.py>
- [2] <https://github.com/tensorflow/tensorflow>
- [3] <http://caffe.berkeleyvision.org/>
- [4] <https://github.com/capstone-coal>
- [5] <https://github.com/capstone-coal/coal/wiki/Updates>
- [6] <https://capstone-coal.github.io/>
- [7] <https://github.com/spectralpython/spectral>