Week 13 Writing Assignment

Matthew Kutugata

Part 1:

Abstract Dissection

- Introduction – Increase in CO_2 effect coccolithophores characteristics that can in turn effect larger ecological systems.
- Methods – use modeling technique to observe the effect of increasing levels of CO_2 on coccolithosphores
- Results – increases in CO_2 concentration cause coccolithosphores to grow less in most areas while calcification increases in other areas.
- Discussion – Global coccolithophores calcification decreases as CO_2 increases.
- Conclusion – Increases in CO_2 result in increases in abundance but a decrease in calcification of coccolithophores at a global tipping point.

Introduction Dissection

Paragraph 1:

- Subject sentence – "These primary producers…perform both calcification and photosynthesis."
  - Coccolithophores are abundant and important in ocean system functioning
  - Their calcification and photosynthesis processes and convert CO_2 to organic matter
  - Coccolithophore processes, if changed, have larger system consequences.
- Broader purpose – Provide background information on coccolithophores, what they do, and why they are important.

Paragraph 2:

- Subject Sentence – First sentence
  - Studies using Coccolithophores as a proxy for understanding OA are abundant but are sometimes conflicting.
  - Efforts made to advance "culturing" problem
  - Large multispecies data set showed increases in CO_2 led to lower calcification
- Broader Purpose – Contradictory effect of increasing photosynthesis and decreasing calcification show

Paragraph 3:

- Subject Sentence – First sentence
  - The importance of looking at coccolithophores
  - Introduces the fact that other models are lacking
- Broader Purpose – Introduces the motivation for her work, the problem of a lack of specific model, and description of the core of her study.

Paragraph 4:

- Subject Sentence – First sentence
  - Other attempts have been made to model subject
  - Others studies have found it important to model changing carbon chemistry and effects on larger earth systems
  - Results of other studies
  - Defines specifics of project – testing the conclusion of previous studies and building by looking at coccolithophores specifically
  - Define main goal
- Broader Purpose – Looks at background of what others have done and specifies main goal that builds on others

Paragraph 5:

- Subject Sentence – First sentence
  - Perform sensitivity studies to look at effect of increasing CO2 on coccolithophore growth and calcification
  - Specifics on simulations
  - Results of the simulations
- Broader Purpose – describes what authors does and briefly describes the results from her simulations.

Part 2:

Abstract Division:

- Introduction – any machine learning task starts with cleaning and organizing data
- Methods – Create a data input pipeline using Pytorch
- Results – We expect that the results of this method will be a streamlined process
- Conclusion – Collecting, preprocessing, and inputting data in the correct format is the first step in using powerful AI tools like Pytorch-based convolutional neural network architectures.

Introduction Bullet Points

- Why use it – Data input pipelines are developed to standardize the procedure of bringing in data for your neural network and to allow users to bring in custom datasets
- Custom dataset versus built-in formats – Many machine learning libraries contain built-in data sets that can be imported into your session
- Goal of project – The goal of this project is to build a data input object designed for complex convolutional neural network architecture.

Abstract

A major part of a machine learning problem involves organizing, cleaning, and transforming data that can be used for complex data analysis techniques. Open-sourced frameworks or libraries used to develop convolutional neural network architectures require specific data input pipelines that standardized model creation and allow users to apply custom datasets. This project plans on creating a data input pipeline using both built-in Pytorch Dataloader functions and user defined to formats multi-labeled images that can then be used in to apply semantic segmentation using a cully-convolutional network. We expect that the results of this method will be a streamlined process that allow the implementation of a custom dataset. Our goal is to complete the first step in using powerful artificial neural networks - collecting, preprocessing, and inputting data in the correct format.

Introduction:

Artificial intelligence is being used in various disciplines as a tool to solve problems that involve complex unstructured data. Digital image analysis in precision agriculture in one such discipline that is using ultra-high resolution images and convolutional neural networks to monitor crops and scout for pests and diseases. These tools, however, require users to package their data in specific formats. The process of organizing, cleaning, and transforming data required by Python-based frameworks is known as a creating a data-input pipeline.

Data input pipelines are developed to standardize the procedure of bringing in data to train complex neural network frameworks. These object-like packages also allow users to feed in custom datasets. Ultimately, learning to build custom data input pipelines gives users greater flexibility in implementing neural networks for specific problems that cannot be addressed using built-in datasets.

Many machine learning libraries contain built-in data sets that can be imported into your session. For example, the Boston Housing dataset can be brought in using the Sci-Kit Learn library (Harrison and Rubinfield, 1978). When doing so, the user brings in a specific type of data format that contain not only a pandas dataframe, but also data that is split into training and target data, feature names, a description and filename. This Bunch, a dictionary-like object, contains everything needed to use a fairly simple data set by both the user and by the accompanying module - Sci-kit Learn.

The goal of this project is to build our own 'Bunch' or data input object. Our object, however, will be more complex as the data inside will be multi-labeled images for semantic segmentation and designed for complex fully convolutional network architecture.

Referecens:

Harrison, D. and Rubinfeld, D.L. `Hedonic prices and the demand for clean air', J. Environ. Economics & Management, vol.5, 81-102, 1978.