DATS-6501 Data Science Capstone

Anomaly detection in Wood Fossil

Prof. Edwin lo

Project Proposal

Pranay Bhakthula

Project Proposal

Introduction:

The secret to humanity's success is probably in how well we are able to use tools to mediate environmental change. But practically all of our present understanding of early cultural evolution comes from research on stone tools and fossilized bones discovered in the archaeological record. Due to their inherent fragility, tools constructed of plants are almost never found in the earliest records of human material culture. The fact that plant materials are used for tools much more frequently than stone in contemporary human civilizations and among non-human primate species raises the possibility that a significant portion of ancient technology is being left out of current archaeological data.
Here, we provide methods for analyzing internal and external damage patterns in living primates' percussive wooden tools. Our research demonstrates that the harm done is irreversible and may endure throughout the fossilization processes. This study provides the opportunity to examine organic artifacts, a significant but overlooked part of the evolution of technology within the primate order.

Dataset:

Fieldwork to collect percussive tools used by chimpanzees was carried out during December 2017 and March 2018 in the North Group of the Tai Chimpanzee Research Project in the Tai National Park (Cote d'Ivoire).
Dataset link : http://cdna.eva.mpg.de/Organic_Tool_Data/

Methods:

In this project we aim to find a method to train a machine to detect anomalies in the wood fossil; i.e, it can detect the damaged part of the fossil from the undamaged part.

For this problem we will analyze a few "features" of the wood fossil images to see if they can separate the damaged from the undamaged portion via unsupervised linear PCA or nonlinear AE/VAE (autoencoder or variational autoencoder).

The features that we use in this capstone project will be:

1. Haralick features
2. VGG's 4096 features normally used in transfer learning
3. Resnet50's Gram matrix used in neural style transfer

Finally, for each of the above features, we run PCA and AE/VAE to visualize data clustering, asking if the damaged crops can be somewhat separated from the undamaged crops

Some of the limitations of the project are the damage on the surface and internal structure will most likely be impacted by the physical properties of wood. These properties vary widely depending on tree species and water content at the time of use. Our study has focused on the materials that were selected by chimpanzees for nut cracking in the Tai Forest. As a result, we only investigated the damage pattern of the most prevalent wood species (Coula edulis).

The required resources for the project are usage of python libraries and cloud computing. Significant progress can be made in the following 2 months and come out with insightful findings from wood fossils.

The one real time usage of this project when fully implemented would be useful for archeologists in remote part of the world to take pictures of fossils and anomalies/unusual/important part can be detected instantly, instead of transporting the delicate fossil to labs for analysis.