Image classification with the MET's (Metropolitan Museum of Art) Open Access Images and CSV

Chloe Jones & Victoria Jordan

Project Summary

Our goal is to develop a machine learning model to classify artworks from the MET's (Metropolitan Museum of Art's) Open Access dataset, which consists of over 406,000 images with a corresponding CSV that contains metadata for each image. Below is an overview:

- Data extraction
- The MET's CSV includes links to each image, so we will extract the images from the links and map the metadata to its corresponding image.
- Standardize images
- Clean metadata
- Image analysis (most likely convolutional neural networks)
- Split dataset for training vs test
- Evaluate model

Problem Statement

Develop a machine learning model that accurately classifies artwork based on medium, artist, culture, or other metadata.

Dataset

MET's (Metropolitan Museum of Art) Open Access Images and CSV

Launch Date: February 2017
Content: Over 406,000 images accompanied by a CSV file containing detailed metadata for each artwork
License: Creative Commons
Metadata: Extensive, includes artist name, medium, culture, time period, classification (e.g., sculpture, painting), creation date, and more.
Number of instances (rows): 484956
Number of attributes (columns): 54
Size (rows x columns): 26,187,624