Reddit Image Classification

A model for identifying art medium

Matthew Williamson

Objective and Business Application

OBJECTIVE
Gather a large set of images using Reddit api for use in modeling medium type of artwork posted to r/Art

APPLICATION
This type of image analysis can be used for categorical image searches as well as identifying images which have been created using digital means

DATA
The process used for gathering and cleaning this data set can be generalized to a wide variety of analytic objectives

Example Images

Titles follow format:

Piece Name, Artist, Medium, Year

Title: Class 3-E, u/Zephyrextreme, Digital 3D, 2020

Medium: digital

Title: Camp, u/DingleMuffins, ink, 2020
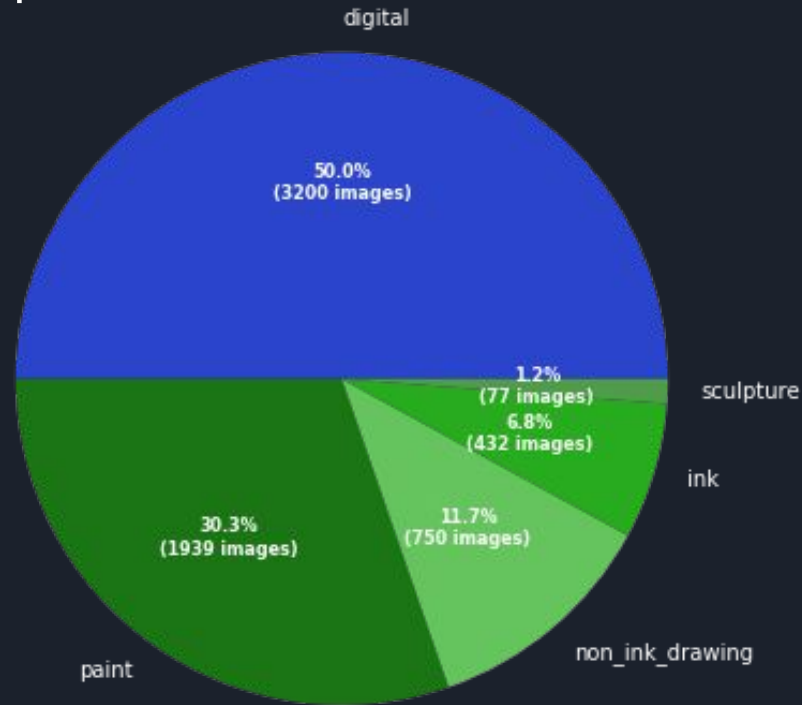
Medium: ink

Title: Early morning at tervasaari, u/Tokaiin, oil paint on a 45x60cm panel, 2020

Medium: paint

Class Distribution

This is the distribution of various different medium types in my dataset. The network was trained to identify whether an image whether or not an image was produced digitally.
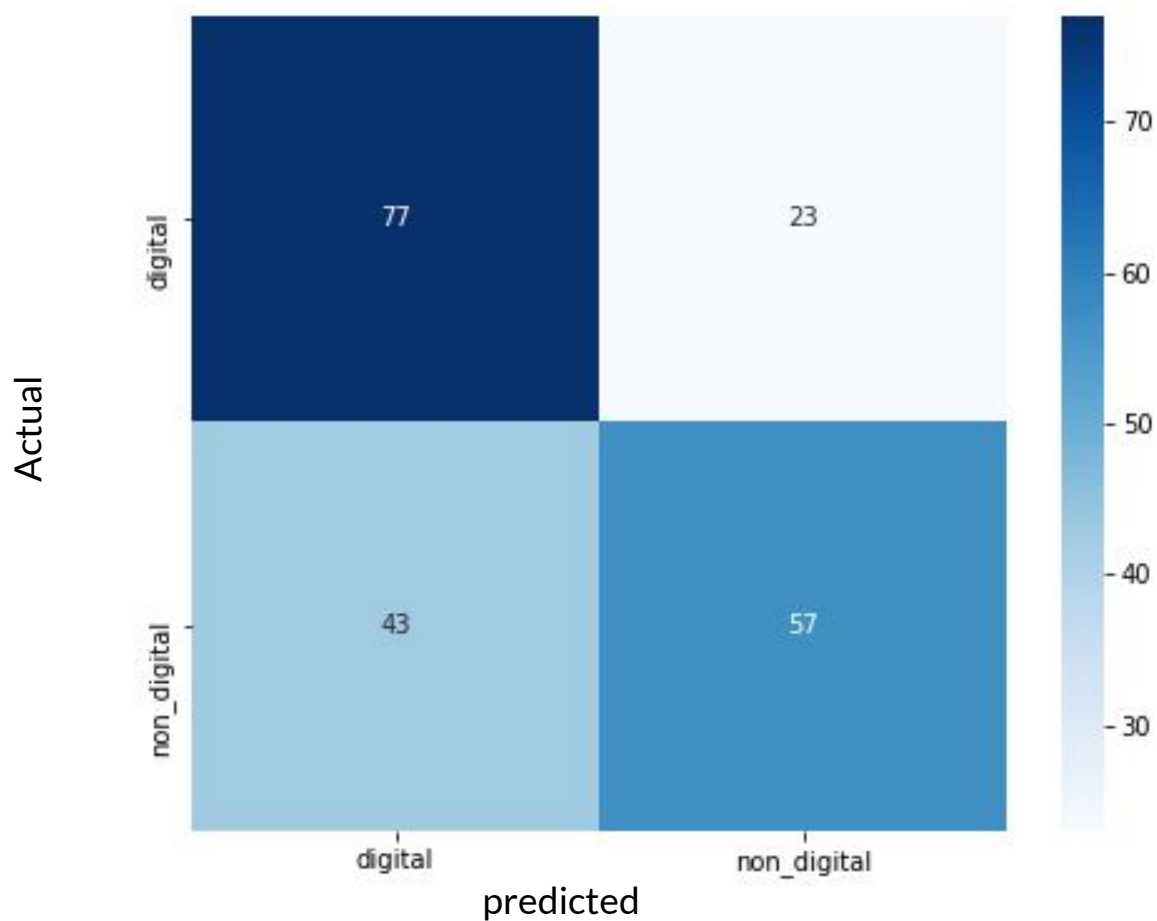
Best Model Predictions

Predictions on unseen data

Overall Accuracy: .67

Digital: .77

Non-Digital: .57

Next Steps

Look into which types of images my model gets wrong

Possible deployment

Contacts

Matthew Williamson:

Github: https://github.com/MWilliamson96

Email: matthew.l.williamson.96@gmail.com