



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

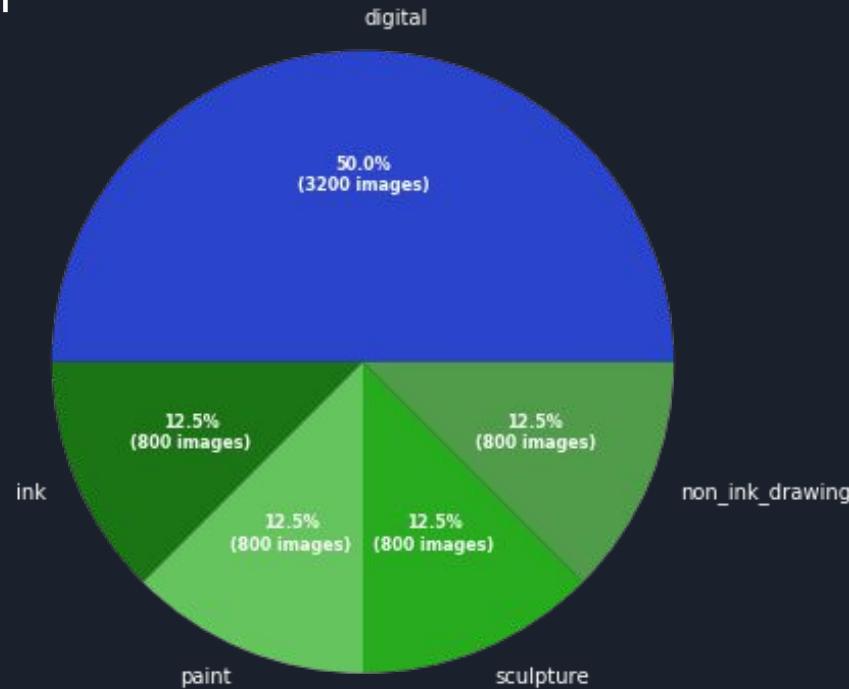


Title: Early morning at tervasaari, u/Tokaiin,
oil paint on a 45x60cm panel, 2020

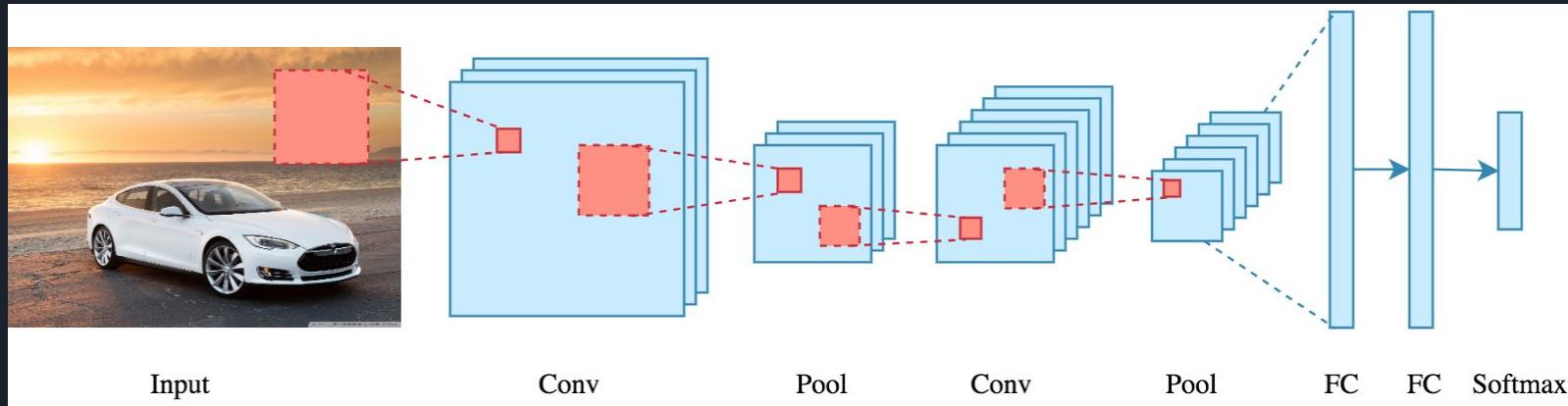
Medium: paint

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. I decided to build my set of non-digital images with an equal distribution of various traditional media



Model Architecture



The model used was a Convolutional Neural Network that was tuned to pick up on low level details such as pen or brush strokes

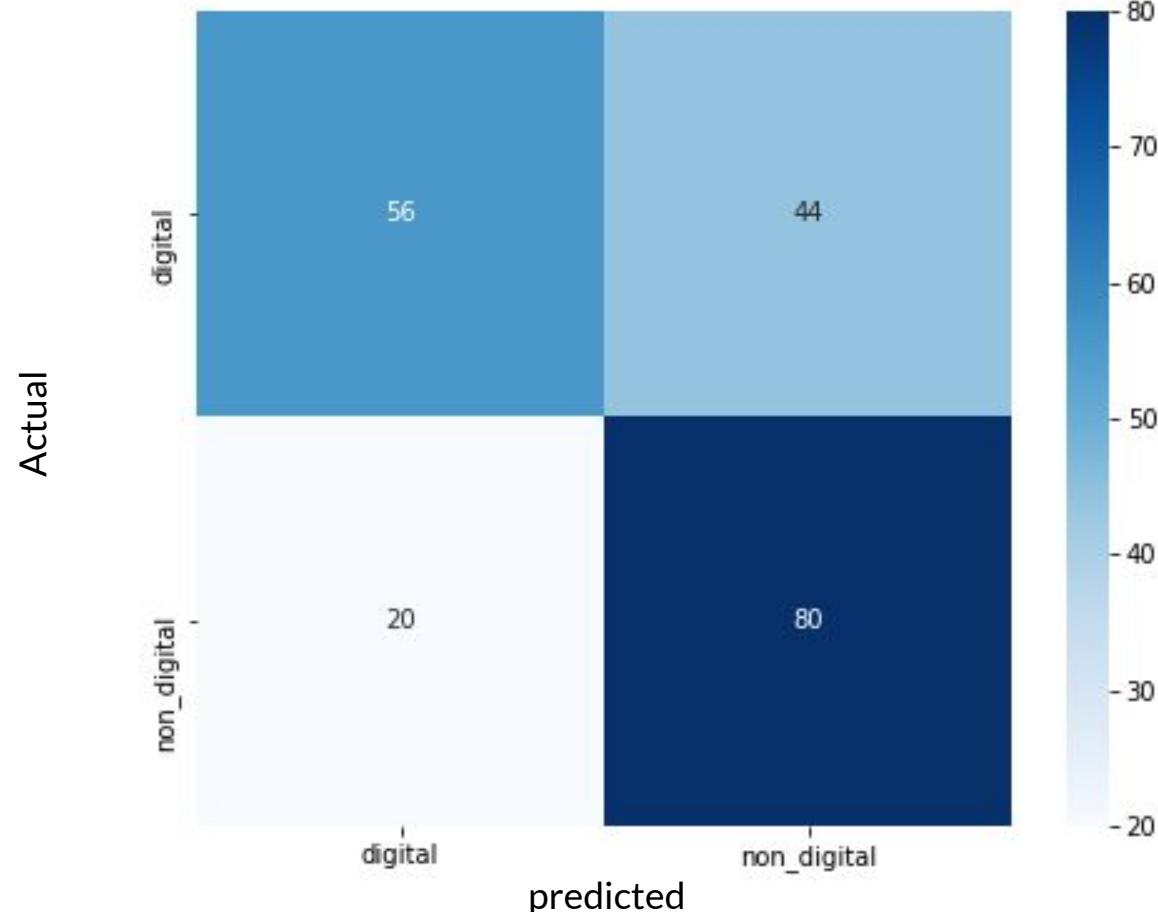
Best Model Predictions

Predictions on unseen data

Overall Accuracy: .68

Digital: .56

Non-Digital: .80





Misclassified Images



Conscience, Me, Digital,
2020

Eleven, Me, Procreate, 2020

Glassy Guitar, Me, Stained
Glass, 2020

Untitled, Me, Digital, 2020



Potential Improvements

- Segregate images that use digital mediums to mimic traditional mediums
- Increase the size of the dataset
- Create an ensemble of models



Contact Me

Matthew Williamson:

Github: <https://github.com/MWilliamson96>

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