

# Milestone 1 Report

## Current State of the project

We have gathered about 4,700 images of paintings and their prices from 4 art bidding sites (liveauctioneers.com, christies.com, artsy.net, artsper.com), and created csv files for their labels. Our data should be ready for training a machine learning model.

From our data, we have removed images that are duplicates, or are of low quality, such as blurry images, images containing multiple paintings, or images that are not even of paintings/drawings.

We have also created a barebone version of our website's frontend. It can accept image uploads and display a price.

For our machine learning model, we have started experimenting with transfer learning using ResNet and other models. So far, we have found that more complicated models like ResNet50 tend to overfit quickly, while none of the models give a satisfactory test loss.

## Feature Changes

We have decided to limit our model to train with only paintings that are \$20,000 or less. This is because for paintings that are more expensive, their prices are more likely affected by the reputation of the artists, but our model aims to predict the values of art pieces using only their aesthetics. We chose \$20,000 since it is not too high in terms of art prices, and most of the data that we have gathered is within this range.

Another change is that our model will now predict the prices of all kinds of 2-dimensional contemporary art instead of just paintings. Since it is sometimes hard to tell the difference between paintings, drawings, and other mediums like computer drawings. Another benefit is that we would have more data to work with.

## Current Challenges

One challenge is training a good price prediction model. From our initial testing, our models could not accurately predict art prices and would easily overfit. Using simpler models helped with overfitting but in general none of our models were good enough for accurate price prediction. We are going to have to figure out ways to train a model other than simply changing the fully connected layer of a pretrained model. Some ways we can address overfitting include, increasing weight decay, adding dropout, adjusting learning parameters such as momentum, learning rate etc., trying different optimizers such as AdamW, and experimenting with a different scheduler. Additionally, we could augment our dataset, by perhaps randomly cropping or horizontally flipping images.

## Tasks Done and Underway by Each Member

Jared:

- Gathered data from liveauctioneer.com
- Labeled data and downloaded images
- Created website template
- Experimenting with training a model

Jane:

- Gathered and labeled images from artsper.com
- Removed images that do not meet our requirement

Eva:

- Gathered data from christies.com
- Filter the dataset and gather bad image to a google sheet
- Remove images that do not meet the requirement
- Find images that duplicate

Dylan:

- Gathered images and price labels from artsy.net
- Helped remove bad elements from the dataset