Motivation
- The appraisal of an artwork is susceptible to an expert's biases
- A robust predictive model could provide radical transparency on art valuation
- Uncovering raw image features that correlate with art valuation could provide new, insightful signal to the art community

Why Deep Learning?
- Recently there have been tremendous advancements in image processing and image understanding thanks to convolutional models
- Existing models can be applied to our problem with little modification

Objective
- Collect a dataset of artwork with valuations
- Develop a model capable of accurately predicting an artwork’s realized auction value

Dataset
- Collected 100K images from Christies
- Resized all images to 256px by 256px
- Converted all valuations to USD
- Problems:
  1. Dataset contained out of scope pieces (furniture, ceramics, etc.)
  2. Positively skewed distribution
- Solution: Modified dataset to contain only “high-value” art (defined as appraisal over 100,000 USD)
- Final dataset comprised train/dev/test allocations of 16705/2088/2088

Original artwork by Paul Signac:

Pre-processed artwork for training:

Approach

Loss function:
- Mean Absolute Percentage Error (MAPE)

Linear Regression
- Learning rate ($\alpha$)

Multilayer Perceptron
- Learning rate ($\alpha$), with early stopping
- Number of hidden layers (L)
- Number of units per layer ($n_l$)

Vanilla CNN
- Learning rate ($\alpha$)
- 64 filters with kernel size (k)

AlexNet
- Learning rate ($\alpha$), with early stopping
- Dropout keep probability ($p_d$)

VGG-16
- Learning rate ($\alpha$), with early stopping
- Number of unfrozen layers ($n_f$)
- Number of added layers ($n_a$)
- Number of units per layer ($n_l$)

Results

Linear Regression
- $\alpha = 0.01$

Multilayer Perceptron
- $\alpha = 0.001, L = 25, n_l = 5$

Vanilla CNN
- $\alpha = 0.001, k = 11$

AlexNet
- $\alpha = 0.0001, p_d = 0.5$
- Train MAPE w/o early stopping: 20.48

VGG-16
- $\alpha = 0.001, n_f = 2, n_a = n_l = 0$
- Train MAPE w/o early stopping: 14.88

Example predictions

<table>
<thead>
<tr>
<th>Model</th>
<th>Example A</th>
<th>Example B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td>79869.80</td>
<td>168104.11</td>
</tr>
<tr>
<td>Multilayer Perceptron</td>
<td>154893.34</td>
<td>167464.44</td>
</tr>
<tr>
<td>Vanilla CNN</td>
<td>150813.73</td>
<td>317116.41</td>
</tr>
<tr>
<td>AlexNet</td>
<td>144910.42</td>
<td>159760.98</td>
</tr>
<tr>
<td>VGG-16</td>
<td>170875.59</td>
<td>180655.05</td>
</tr>
</tbody>
</table>

Conclusion
- It is possible to achieve low bias and deeper, more complex models result in the lowest training set loss
- Generalization to unseen data is the most difficult aspect of the problem
- Standard techniques for variance reduction are insufficient

Future Work
- Expand dataset
- Continue architecture search
- Establish human baseline
- Discretize appraisals for categorical prediction
- Use additional information beyond raw image inputs

Acknowledgements
- Jay Whang
- Andrew Ng
- Kian Katanforoosh