State-of-the-Art: End-to-End Deep Learning for Art Appraisal

Alexander Verge (averge), Ishaan Singal (ishaan)
(Presentation Video Link)

Stanford Computer Science

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

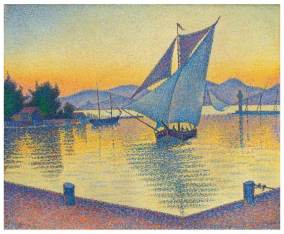
- Collect a dataset of artwork with valuations
- Develop a model capable of accurately predicting an artworks realized auction value

Stanford University

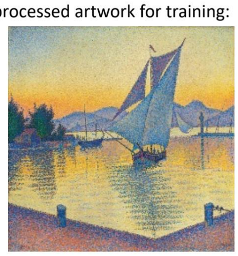
Dataset

- 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 (α)

Multilayer Perceptron
- Learning rate (α), with early stopping
- Number of hidden layers (l)
- Number of units per layer (n_l)

Vanilla CNN
- Learning rate (α)
- 64 filters with kernel size (k)

AlexNet
- Learning rate (α), with early stopping
- Dropout keep probability (p_k)

VGG-16
- Learning rate (α), with early stopping
- Number of unfrozen layers (n_u)
- Number of added layers (n_a)
- Number of units per layer (n_l)

Results

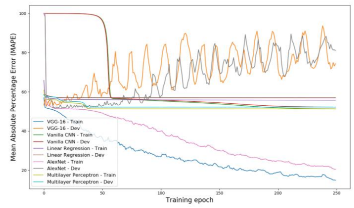
Linear Regression
- $\alpha = 0.01$

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

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

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

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

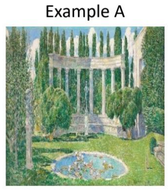


Model	Train MAPE	Dev MAPE	Test MAPE
Linear Regression	55.72	56.96	56.12
Multilayer Perceptron	51.30	52.13	51.38
Vanilla CNN	51.29	52.24	54.76
AlexNet	51.43	52.01	51.46
VGG-16	51.82	52.52	51.78

Final model performance under optimal hyper-parameter configurations

Example predictions

Example A



Artist: Childe Hassam
Value: 170,500 USD

Example B

Artist: Andy Warhol
Value: 56,165,000 USD

Model	Example A	Example B
Linear Regression	79869.80	168104.11
Multilayer Perceptron	154893.34	167464.44
Vanilla CNN	150813.73	317116.41
AlexNet	144910.42	159760.98
VGG-16	170875.59	180655.05

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