Art Appraisal Using Convolutional Neural Networks

Rafi Ayub, Cedric Orban, Vidush Mukund*

December 15, 2017

Abstract

The appraisal of artworks has long been an esoteric process reserved for the few educated in art and art history, yet it plays an integral role for collectors, curators, and auction houses. Valuation can be influenced by numerous factors and could be swayed by personal biases and local values. Machine learning algorithms may serve as a tool to adjudicate the value of artwork due to the high subjectivity of the pieces. In order to provide a more standardized method of appraising artwork, we suggest the use of a convolutional neural network (CNN), an algorithm typically used for image classification, to allow an individual without a background in art to better determine the market value of a given piece. We aim for the model to be free of the individual bias that are present in human art appraisers because the dataset is formed from many existing appraisals, averaging over the biases of multiple localities. Our hope is that upcoming artists, art buyers, and art sellers can make more informed decisions with the help of this model, and we may be able to determine cross-cultural features of art pieces that make them inherently valuable.

1 Introduction

Art prices can vary tremendously from affordable to exorbitant even within the same auction house. The selling price for artwork is a confluence of many factors and collectives, including the social status of the dealers, the prestige of the gallery, the reputation of the artist, and even basic supply and demand. To people outside of the art market, this mixture of factors causes bewilderment and confusion at the final valuation of certain artwork. Many young aspiring artists can be discouraged by the seemingly arbitrary selection of artists that rise to fame by selling their artwork. The question then becomes if the art market places intrinsic value on the quality of the artwork. Our aim is to discover if artwork contains features that objectively make it inherently more valuable or if the valuation of art pieces is more heavily dependent on non-visual features of the piece, i.e. the artist's name and cultural significance of the piece. By finding the answers to these questions, we may be able to better inform auction houses, sellers, and dealers in trading artwork.

A rigorous approach to analyzing visual data involves borrowing techniques from machine learning. Machine learning algorithms are commonly used to find patterns and relationships in seemingly random data. Specifically, convolutional neural networks (CNNs) are a type of architecture that are designed for extracting visual feautures from image input. They are commonly used for visual feature detection and object recognition, though we will leverage its use of convolutional features to extract some underlying characteristics of an image that are difficult to pre-determine. Thus, we will apply CNNs to analyze a large dataset of various artworks to determine the relationship between the visual features of the art itself and its bidding price. The input to our CNN is a set of color images of artwork, and the output is the price range to which the given images belong to. We also built a softmax regression model that takes as input only the metadata, i.e. artist name, sell date, etc., associated with the images of art and outputs the associated price range.

2 Related Work

Though we were unable to find previous attempts at solving the particular problem of appraising artwork given only an image, we did find multiple machine learning approaches to relating images to aesthetic value or price and general image classification. Two

*Category: Computer Vision. SUNet IDs: Cedric Orban (cedrico) 06189481, Rafi Ayub (rafiayub) 06189872, Vidush Mukund (vmukund) 05874348

papers we looked at focused on the problem of building a model mapping images to values. You et al. (2017) describe the use of a recurrent neural network for real estate price estimation, using as input visual content via pictures of properties and neighborhood information of properties [8]. Their work is similar to our own in that they attempt to map visual features of properties to prices, but the difference in their approach is the use of an RNN and the access to geographic information of the properties. Suchecki et al. (2017) approached the problem of evaluating the aesthetic quality of digital pictures via a CNN modeled after AlexNet [4]. They measured aesthetic quality via the ratio of the number of views over a time period. Their problem is very close to our own in its evaluation of digital images, but they use a much larger data set (1.7 million images) and only have two classes, i.e. aesthetically pleasing photographs and aesthetically displeasing photographs.

Other papers we found focused on general image classifications problems. Krizhevsky et al. (2012) uses a convolutional neural network with 60 million parameters to classify 1.2 million images into a thousand different classes [3]. Their approach requires the use of a convolutional network that is much more resource intensive than what we could afford, but with 60 million model parameters, ImageNet was able to win ILSVRC-2012. Also, this application does not involve modeling of the valuation of the images, but rather the identification of given images.

Qayyum et al. (2017) suggest the use of convolutional neural networks for the processing and identification of medical imaging [5]. Their approach is closer to Kizhvesky,but far reduced in regards to the number of model parameters needed. The structure of the CNN used by Qayyum et al. is very similar to the architecture we used, using three convolution layers but with only on fully-connected layer. Tajbakhsh et al. (2016) also suggest the use of convolutional neural networks for the identification of medical images, particularly in radiology, cardiology, and gastroenterology [7]. But rather than training CNNs on labeled data, Tajbakhsh et al. show that its may be more effective to use deeply fine-tuned image CNNs, these pretrained CNNs performing as well or better than training a CNN from scratch on only medical image data. We thought this was a particularly interesting trick, though this direction of image classification defers from our image to price mapping.

3 Dataset and Features

We gathered the digital images of 100,000 artworks from the database findartinfo.com using a scraper we wrote in Python. This dataset contains images of artworks rendered in multiple mediums and created over a span of centuries. In addition to the images and prices, we collected each listings relevant metadata, such as whether or not a signature is present on the piece, the age of the piece, and the artist.

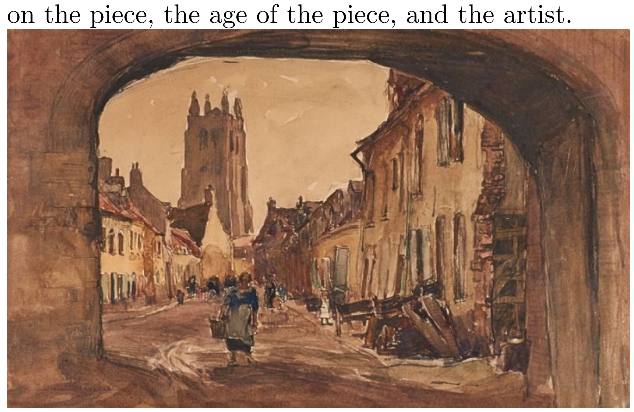

Figure 1: An example of an input artwork, *Le Route Petain*. The sell price for this artwork was $283. The associated metadata passed to the softmax regression model would include the artist (Edmond Lesellier), the medium (watercolor), the approximate dating of 1900, and that the piece was signed.

The sizes of the digital images varied for each artwork, so each image was resized to a default size of 400x400 pixels. 400x400 was chosen to not only ensure that images were manageable sizes, but also as a trade-off to preserve high resolution details present in the artworks. This was also done as opposed to padding the smaller images so that we do not overfit to the higher resolution images and risk a high generalization error to images with lower resolutions. The input to the CNN was a 3D tensor of 32-bit floats representing the raw RGB pixel values of each image. Before being passed as input to the CNN, images were batch normalized to reduce the internal covariate shift.The art prices were arranged into ten bins. The ranges of the bins were chosen so as to place the same number of artworks in each bin so as to not bias any single class of price range. Our price ranges were \$0 to \$101, \$101 to \$196, \$196 to \$321, \$321 to \$499, \$499 to \$760, \$760 to \$1202, \$1202 to \$2033, \$2033 to \$3857, \$3857 to \$9643, and \$9643 to \$60130038.

In the case of the softmax regression, we considered a subset of the full dataset that only contained the top 250 artists of all artists ranked by the number of artworks created. The features used for the softmax regression model were the name of the artist, an approximate dating of the piece, the medium in which the piece was wrought, and whether the artwork was signed by the artist. Due to the requirement for the artwork to have all of the afore mentioned metadata features, the dataset was much smaller than the data set used for ArtNet, 700 artworks used to train the softmax model. We transformed categorical metadata features into 1-of-k binary representations. For example, we used a binary value for each of the 15 mediums that the artwork could be rendered in.

4 Methods

For our first CNN implementation, ArtNet, we constructed a neural network with three convolutional layers and 2 fully connected layers. The first layer had 32 5x5 filters with strides of 1 with a 2x2 max pooling filter of stride 2. The second layer had 64 5x5 filters with strides of 1 with a 2x2 max pooling filter of stride 2. The third layer had 16 5x5 filters with strides of 1 with a 2x2 max pooling filter of stride 2. The first fully connected layer had 1024 neurons and the last fully connected layer had 10 neurons.

A random 10% of our dataset was a set aside as a test set, and for each training chunk of 30k images, a consistent 10% was used as a validation set. ArtNet was trained on an Amazon p2.xlarge instance. To ensure the neural network did not overfit the training set, a dropout rate of 50% was used on the first fully connected layer while training [6]. Loss for ArtNet was calculated via the cross-entropy loss like the softmax classifier. The Adam (adaptive moment estimation) optimization algorithm was used to minimize this loss function [2]. Adam maintains per-parameter learning rates that are adapted not only on the average of magnitudes of the gradients for the weight but also on the variance of the gradient, thus optimizing the learning rates via the first and second order moments.

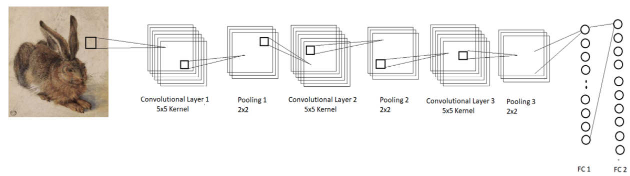

Figure 2: ArtNet CNN Architecture

Another CNN with an architecture similar to the AlexNet was used as a comparison of our custom built network's performance. For this CNN, we constructed a neural network with three convolutional layers and 2 fully connected layers. The first layer had 32 5x5 filters with strides of 1 with a 3x3 max pooling filter of stride 3. The second layer had 64 5x5 filters with strides of 1 with a 3x3 max pooling filter of stride 3. The third and fourth layers had 64 5x5 filters. The fifth convolution layer had 32 5x5 filters with strides of 1 with a max pooling filter of size 3x3 with a stride of 3. The first fully connected layer had 2048 neurons and the last fully connected layer had 10 neurons. This CNN was used to confirm if our model, when compared to a model used for general image classification, was effective at extracting underlying structures from given artwork images. Due to memory constraints, a mini-batch size of 32 images was used for both CNNs.

A softmax regression model was applied to the metadata with cross-entropy loss and l2 regularization to prevent overfitting. The Adam optimization algorithm was also used to minimize the loss for the softmax regression model. The loss function for the classifier is given below:

$$L_i = -log(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}})$$

The f term is defined as:

$$f = Wx_i + b$$

The final output of the softmax classifier is a vector containing the probabilities of the artwork belonging to each price bin.

5 Results

In regards to hyperparameters, for ArtNet, we found that the best starting learning rate to be 0.01 for our Adam optimization algorithm. For the softmax regression model, we found that the learning rate of 0.01 was optimal and that the best l2 regularization term to prevent overfitting to the training set was 0.1. Our primary metrics for measuring the efficacy of both models was comparing the training accuracy to the validation accuracy. By ensuring the training accuracy was close to the validation accuracy, we prevented the models from overfitting to the training set and possibly performing poorly on the final test set. For the softmax regression model, we also used a confusion matrix to better understand if the model, after training, was biased towards any particular classes.

With the CNN implementation described in the Methods section, we achieved a train accuracy of about 10% and a validation accuracy of 10% with both CNNs. Note that we used ten classification bins, indicating that the accuracy of the CNN is no better than pure chance.

One of our hypotheses was that the CNN may be able to find visual features of general artwork that makes the piece inherently valuable, though this appears to not be the case. It is also possible that, for artwork from very different styles and mediums, there is no one model that maps visual features to price ranges. Regardless, these results suggest that visual information of the artwork alone cannot determine its price. This motivates the use of the softmax classifier with the metadata of each artwork.

The softmax model achieved a test accuracy of 21.25%. This suggests that the artist name, signature, medium, or age of the artwork influence the auction price of an artwork.

As can be seen in Figure 3, the softmax model achieved a training accuracy of 21.67% and a validation accuracy of 21%. We found that by increasing the l2 norm regularization hyperparameter to 0.1, we prevented the model from overfitting to the training

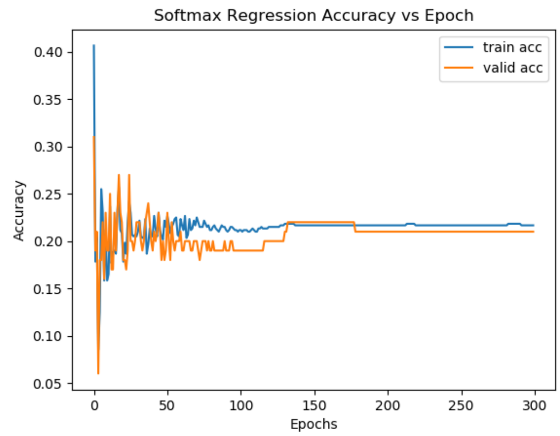

Figure 3: Accuracy of Softmax Regression over Epochs

set. By testing various hyperparameters, such as the regularization parameters and the learning rate, we found that the model was only able to achieve around 20% test accuracy at best.

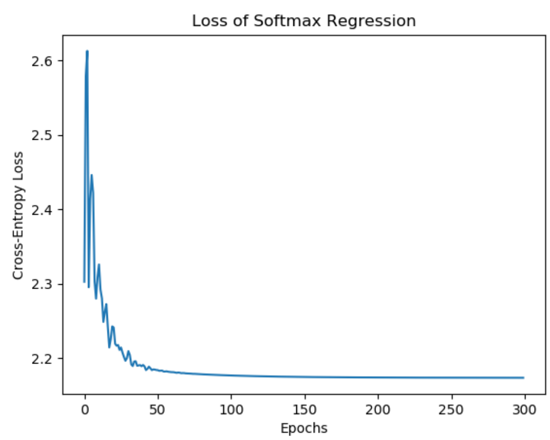

Figure 4: Loss of Softmax Regression over Epochs

The above confusion matrix shows that the softmax model is heavily biased towards clsses 0 through 4. These labels correspond to the first four price ranges described in the Dataset and Features section. This bias by the model is probably due to the fact that more artworks sold at auction in those price ranges. Since the dataset for the softmax model was heavily limited in size due to the requirement of having particular metadata features (as mentioned in

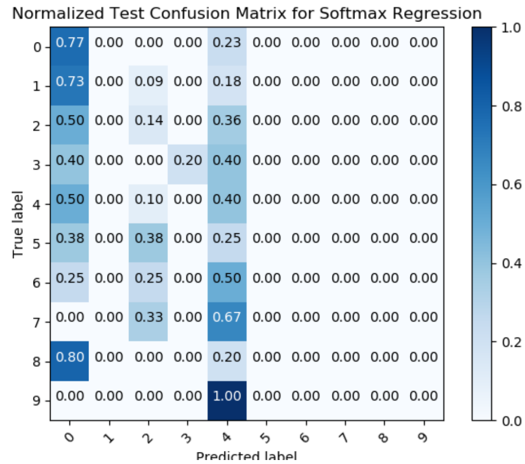

Figure 5: Confusion Matrix for the Test Set of the Softmax Model

Section 3), the class labels were more skewed to the lower end of the price ranges. This is perhaps more representative of the true distribution of auction sell prices as the majority of artworks sold belong to the price ranges corresponding to labels 0 to 4.

The low accuracies from the CNN models may suggest a couple of conclusions. One is that the auction prices may be independent of the visual content of an art piece, which is likely since the prices are completely decided by individuals who run the auction and appraise the artworks. The results of ArtNet and the AlexNet inspired CNN may then be more concrete proof that value of artwork is not tied closely to the visual features of art pieces. Another possible conclusion that art is very difficult to objectively classify given only visual content due to its inherent subjectiveness. As suggested by the higher accuracy of the softmax regression model, metadata, such as the artist's name, seem to have a far greater impact on the valuation of artworks. Given below is a table comparing the final performances of the different models. The softmax regression, with access to only the metadata, outperforms the two CNNs which were given access to only the visual information.

Model	Train Size	Test Size	Train Acc	Test Acc
Art CNN	90,000	10,000	0.106	0.101
AlexNet	90,000	10,000	0.103	0.110
Log Regr	600	100	0.2167	0.2125

Figure 6: Performance of Different Models

6 Conclusion

Our results show that art, in general, cannot be successfully appraised using convolutional neural networks on just the raw image data representing the artworks. While ArtNet performed poorly when only training on the raw images of art, the softmax model was able to achieve a test accuracy of 21.25% on the metadata alone. This is likely because their are multiple hidden factors affecting the appraisal of a piece of art (e.g. cultural, historical considerations) that cannot be captured by visual content alone. Perhaps a combination of models that take into account possible aesthetic measurement features such as the overall entropy of the piece along with cultural significance, measured perhaps via region of production or historical importance, of the piece would perform better. We also suspect that since our dataset had artworks that varied significantly in regards to artistic medium, i.e. sculpture versus painting, it was more difficult for a CNN to extract underlying structure.

Future avenues of approach include using the same models to classify subsets of the full dataset. For example, we could classify all the work by one artist or all pieces rendered in the same medium in order to eliminate the variability of the dataset. Restricting the dataset to the pieces created by a specific artist seems especially promising, given that neural networks have been used to generate images by mimicking the style of an artist [1]. Additionally, we could perform ablative analysis of the metadata features to indicate the most relevant features for sell price classification. We would first need a much larger dataset with the appropriate metadata features recorded. All of the images in the dataset scraped from the auction database we used did not necessarily have the appropriate metadata features recorded; therefore, it may be necessary to scrape information from multiple auction sites to gather enough data points.

References

[1] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic

style." arXiv preprint arXiv:1508.06576 (2015).

[2] Kingma, Diederik, and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[3] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[4] Suchecki, Maciej, and Tomasz Trzcinski. "Understanding aesthetics in photography using deep convolutional neural networks." arXiv preprint arXiv:1707.08985 (2017).

[5] Qayyum, Adnan, et al. "Medical Image Analysis using Convolutional Neural Networks: A Review." arXiv preprint arXiv:1709.02250 (2017).

[6] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." Journal of machine learning research 15.1 (2014): 1929-1958.

[7] Tajbakhsh, Nima, et al. "Convolutional neural networks for medical image analysis: Full training or fine tuning?." IEEE transactions on medical imaging 35.5 (2016): 1299-1312.

[8] You, Quanzeng, et al. "Image Based Appraisal of Real Estate Properties." IEEE Transactions on Multimedia (2017).

Contributions: Softmax classifier - Rafi Ayub. Data Set Scraping and Cleaning - Cedric Orban. Convolutional Neural Nets - Vidush Mukund.