Art Authentication: A Deep Learning Approach

for STATS281

Griffin Tarpenning
gritarp@stanford.edu

INTRODUCTION

Art authentication is a topic of constant interest, frequently re-injected into national discussion by the discovery of a new forged artwork. In fact, just this month the original painter was questioned for one of the most influential and expensive ($450 million) paintings ever created. Salvator Mundi, originally attributed to Leonardo da Vinci, has been downgraded by the Prado museum in Madrid to a mere copy of the original [2]. How is it possible that a copy can be sold as a masterpiece of international acclaim? After all, this painting was subject to intense study through pigment analysis, hyperspectral imaging, as well as significant expert analysis on technique [3].

To investigate the topic of art authentication and possible methods for improvement, we use a famously difficult case study: the works of Amadeo Modigliani. There are a few critical reasons why Amadeo Modigliani proves especially challenging for art authentication, and thus is an interesting case study to develop solutions for. The first, outlined by Lorena Muñoz-Alonso, is that there is no true catalogues raisonnés, or recognized collection of the artist's actual work [9]. The sheer number of fakes, in addition to a few infamous art forgers producing "Modiglianis" at the same time as Modigliani himself, casts doubts on all but a few select works. However, the second reason this case study is so interesting lies in the fact that Modigliani's art dealer, Léopold Zborowski, was possibly providing the same materials Modigliani was using to other artists. Thus, traditional materials and hyperspectral analysis of paintings, the same methods that appear to have failed in the authentication of Salvator Mundi, would fail on many Modiliani fakes.

In fact, there is no better example of the difficulty of Modigliani authentication than the 2017 Ducal Palace exhibit. There, 21 Modigliani's were shown to hundreds of thousands of guests, and eventually sold for millions of dollars after the conclusion of the show. Except, only a few months after its conclusion, a probe found that 20 of the 21 paintings were actually fake [8]! An example of a forgery that tricked audiences for years is displayed in **Figure** 1. Clearly, there is a real and immediate need for an improved method of art authentication, if not for the benefit of art scholars and historians, then for galleries and curators seeking to purchase art without risk of devaluation.

This paper investigates one crucial remaining aspect of art authentication that is rapidly gaining popularity, computational visual analysis. For the last many centuries, visual analysis of artworks has been dominated by art historians, galley proprietors, art dealers, and scholars. However, aided by massive developments in computation, specifically in the domain of artificial intelligence, visual analysis can finally be attempted by computer models. This paper implements a state of the art deep neural network approach to the pernicious problem of art authentication.

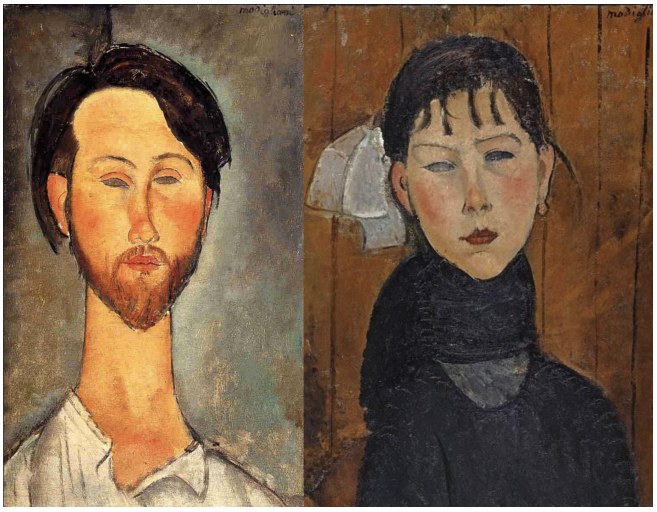

Figure 1. A genuine Modigliani picture left, and a recently downgraded fake on the right.

RELATED WORKS

As early as 2004, large computational models have been applied to art authentication [7]. Using very small swaths of paintings, and a computational method called wavelet analysis, Lyu and their team out of Dartmouth were able to effectively confirm authentication for a few specific artists. This method, however, was found to generalize poorly across artistic domains, working most effectively in highly technical pieces with clean, simple lines. These results serve as an important validation of the deep neural network (DNN) approach, as many found a consistent strength of DNN's to be improved generalizability [10].

This intuition is expanded on in the domain of art classification, where models attempt to classify artworks by their style, age, and creator. DNN's have been shown to improve baseline accuracy here significantly, in some cases reaching as high as 80% across wide domains of artists[5]. In a similar task, DNN's were used to featurize images of art, identifying moments within the image of significant aesthetic importance

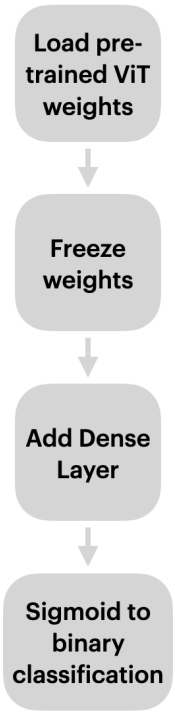

Figure 2. Model Design

[11, 1]. A common theme of these methods is their use of convolutional neural networks, a previously state of the art method of extracting useful features out of images [4]. However, recent developments in AI for visual analysis suggest that convolutional networks have drawbacks, remedied by a new state of the art architecture: transformers [12]. Thus, it seems plausible that applying this new transformer architecture to the task of art authentication could prove a worthwhile experiment in the hunt for an effective discriminator of art.

METHOD

This paper implements three experiments in the pursuit of enabling automatic art authentication. The first and second experiments can be seen as stepping stones to the ultimate goal in experiment three, which is to identify fake or forged Modigliani paintings from originals.

Model Design

The model is generally the same across the three experiments, with the primary changes being task and dataset design. The structure is illustrated in figure 1. First, a transformer model, pretrained on the 1.7 million image + caption pairs in the COCO dataset, and with over 75 million parameters, is loaded in. These weights are frozen, the existing classification layer detached, and replaced with a blank dense linear layer designed for classification. Finally, the outputs from the final dense layer can be mapped with a sigmoid function to our binary classification task. This structure is laid out in **Figure 2**.

There are a few advantages and disadvantages that should be considered with this model design. The first major disadvantage is that we are not actually training the entire transformer model through our experiments. What then, is the point of including this huge pretrained model? The purpose of the transformer is to effectively featurize our image input into a synthesized and intelligent feature space, which can then be used as input to a linear binary classifier. While it would be ideal to use the entire transformer model, due to our extremely small dataset (0.035% the size of the original pretrain task), the amount of data run through the model to see any change at all in the transformer accuracy would likely correspond to severely over-fitting the training set. By freezing the seven encoder and seven decoder layers, we ensure that the only weights updated are the classification dense layer, which are orders of magnitude more sensitive to our tiny dataset.

Experiment 1

The goal of experiment one is to classify artwork as either in the style of Modigliani or in a different style. In this way, we completely avoid the confounding problem of identifying whether or not an artwork is a fake or an original, as in this experiment both would be considered in the style of Modigliani. Thus, we construct a database that consists of all artworks that at any time have been attributed to Modigliani. Both positive and negative samples for this dataset were collected from WikiArt, with an ultimate breakdown described in **Table 1** . Negative samples were collected from a variety of artists that painted in a roughly similar style to Modigliani, some highlights being: Edgar Degas and Jozsef Ronai.

	All	Modigliani	Non-Modigliani
Total	631	158	473
Train	471	113	358
Eval	130	35	95
Test	30	10	20

Table 1. Database breakdown

The results of the model applied to this classification task are reported in **Figure 2**. With over 98.34% accuracy on the evaluation set, we can be confident that the model is, indeed, able to effectively identify the Modigliani style. Through training we see that loss is reduced to only one image on the training set, and two images on the validation set. Strong results here suggest that learning the style of an artist through transformer featurization is a valid approach, and provides compelling evidence that transformers can provide the level of fidelity when featurizing an image to be useful in art authentication.

Experiment 2

In this experiment, we create a new dataset of images that have been doctored to appear in the style of Modigliani. To do so, we implement a version of the recently-released, state of the art, style transfer network LapStyle. This deep neural network uses a laplacian pyramid scheme to efficiently take a source image, understand the style of the image, and then apply it to other images. In this way, we train the LapStyle network on a few different Modigliani images, then apply them to a few hundred image portraits in various settings [6]. **Figure 3** provides an example of said transfer. Crucially, this

Figure 3. Sample of style-transferred portraits

task should be more difficult for the model, as it now must contend with imposter images that appear at least in some surface level way to be similar to the original Modiglianis.

First, we use the same model trained in Experiment 1 to test the difficulty of this task. The accuracy, along with the first few misses, is reported in **Table 1**. As expected, this task proves more challenging for the model, only achieving 88.23% accuracy at determining real Modigliani-style paintings from these style-transferred imposters. As an additional test, we include only two of these imposter images, labeled as imposters, into our training set and retrain the model. With the inclusion of just two negative samples, the model is now able to complete the experiment with 100% accuracy. To a human, the style transferred samples are very obviously not Modigliani, so while this is perhaps expected, it is still encouraging.

Experiment 3

In this final experiment, we use the model trained in the previous two experiments to determine the authenticity of test set Modiglianis. This existing model has been trained to identify the Modigliani style, using a training set where true positives include known fake Modiglianis. Thus, we expect this model to classify both real Modiglianis and forged Modiglianis as "True." However, because forgeries and fakes come from all manner of individuals and places, with their own subtle deviations from the Modigliani style, we hypothesize that the model will be most confident when classifying a true Modigliani vs. a fake. To incorporate this into the classification approach, we define a confidence interval that maximises the number of true Modigliani's and minimizes the number of imposter Modigliani's greater than it. Initially, this confidence interval is set based on the mean confidence for true predictions in the validation set, at 95.8%. Thus, our final classification is True if the model predicts True with a confidence greater than 95.8%, otherwise we predict False. "True" being a real Modigliani, "False" being an imposter. The results of this experiment are shown in **Table 2** below.

Experiment	Accuracy (%)
Ex1	98.34
Ex2	88.23
Ex2+Ret	100
Ex3	73.1

Table 2. Experimental Results

CONCLUSION

The results from the final experiment suggest that it is indeed possible to learn elements about Modigliani to make informed predictions as to its authenticity. The model is certainly not perfect, but distinguishing between an excellent fake and a real Modigliani isn't trivial, as seen by the Ducal Palace exhibit fiasco that displayed 20 fakes alongside one real Modigliani. Achieving 73% accuracy should be seen as a major step towards creating models with the necessary accuracy required for perfect art authentication. However, perhaps the most important result of these experiments is that the learned featurizer from large, pre-trained transformer models is effective at ascertaining useful information in the domain of fine art. This is not a trivial result, as transfer learning has been shown to be extremely sensitive to domain translation, with even some of the most generalizable state of the art models failing spectacularly on seemingly simple translations [10].

DISCUSSION

One crucial point of discussion regards the simplicity of the model used. The model operated in this paper offers significant room for improvements in the future. The creation of a dataset with dramatically more samples could enable some number of the transformer layers to be un-frozen, and thus improve the featurization of the image. However, the creation of said dataset would require significantly creative design, a few possibilities follow:

1. Expand on the success of the style transfer experiment shown in the study, to stylize a significant portion of COCO, and train the transformer on that. The benefits here are that the transformer might pick up on significantly more style-related elements of the images, like brush strokes and color contrasts. However, a drawback is that no matter the sophistication of the style transfer mechanism, the resulting images will still look nothing like a real Modigliani if analyzed by even a human of middling art exposure.

2. Perhaps there are image augmentation techniques that can be applied to our small corpus of Modigliani-like images that could expand the corpus without destroying the integrity of the image. This will require a very sensitive touch, as traditional image augmentation schemes for image classification tasks involve distortion and blurring, both of which

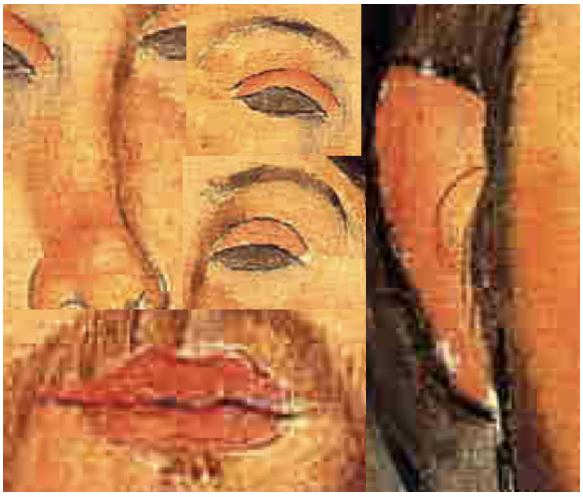

Figure 4. Example of a possible mixed patch-based composition. Patches of varies sizes are concatenated into one image here, but could be fed in separately and their outputs combined after.

could severely compromise the stylistic elements of the Modigliani style.

3. If we make the assumption that there are many distinctly Modigliani elements to an image, it is feasible that each image could be split up into many parts, and passed through the transformer as separate images. The benefit to this is that we are staying true to the style of the original, but we are potentially losing the higher-order information that might be impossible to discern from only looking at a smaller part of the image. However, the dissection of the image could be intelligently done. For example, an algorithm that identifies high and low entropy parts of the image could be used to identify patches that respect important lines/shapes.

Moreover, the model design itself could be improved, perhaps with the inclusion of additional sources of information. As discussed in the introduction, art authentication has for centuries included a wide variety of sources of information in order to make a designation. While these methods have been shown to fail spectacularly in some cases, they certainly also offer additional, if incomplete, information. One way of including this information into a model could be to concatenate to the final dense layer a few indicators of confidence currently used in authentication, like pigment analysis, hyperspectral information, location, provenance, exhibition history, etc. This would then get included in the linear classifier, and could aid in the classification process. Or, the confidence output of this model could be included in a Bayesian belief net, in combination with those indicators previously mentioned, and trained on an extensive corpus of art. Either of these methods would provide a more holistic analysis than is currently employed, and could be key in limiting the billions of dollars of fake art that has pervaded galleries, homes, and museums across the world.

SIGNIFICANCE

There are two main areas of significance for this research. The first is what the experimental results suggest about how transformers understand fine art. Second, this work also takes a step towards real art authentication, which would prove considerably important to the art world.

Technical

Transformer models have been used as featurizers in a wide variety of domains, including fine art, but in very few cases. This paper demonstrates the considerable transfer learning potential of transformers trained on images to fine art. While not perfect, the positive results suggest that the same features that allow transformers to classify pictures are also still relevant to ascertain the artistic style of an image. This is not an insignificant finding, as object recognition and image captioning, the tasks that were used to train the transformer, are wildly different objectives to style-identification. The generalizable nature of transformers can be seen clearly, and absolutely should be investigated further. One very critical question that has not been answered is as follows: what parts of the image are most important to classification? Further research into answering this question would not only prove intensely interesting for art scholarship, it could also show unique advantages of transformer models.

Art

As mentioned previously, there are incredible quantities of fake artwork circulating galleries, museums, and private collections. While there would be considerable monetary incentives to creating a method for unequivocal art authentication, I believe this issue is perhaps less important than the scholarly gain. Many fake artworks have been attributed to masters alongside their actual work, fundamentally changing the perception and our understanding of the artist. Painting is a form of information storage! Through paintings, scholars are able to reconstruct information about history, individuals, painting styles, and the artist themselves. However, this analysis is sullied by the inclusion of fake artwork, and thus is a ripe target for an effective delineation scheme.

While the method employed in this paper did not reach a high enough accuracy to be considered a truly effective delineation mechanism, it does provide a unique source of information correlated with authenticity. It could be an interesting tool for scholars to apply to a broad corpus of works, not necessarily Modigliani, to determine which artworks are the most canonically in-style. It is also relatively swift to apply the model (only taking around a second per painting), which means it's feasible to trawl the web looking for images that are most similar to a given painting-style. While not the intended use of the model, its an interesting alternate use case, as the model has fundamentally learned the style of the artist. This might be more useful as a research tool than just doing an image-similarity search (like tineye).

REFERENCES

[1] Rafi Ayub, Cedric Orban, and Vidush Mukund. 2017. Art Appraisal Using Convolutional Neural Networks. *CoRR* abs/1707.08985 (2017). `http://cs229.stanford.edu/proj2017/final-reports/5229686.pdf`

[2] Martin Bailey. 2021. Major museum casts fresh doubt over the authenticity of $450M 'Salvator Mundi. *CNN*

(2021). https://www.cnn.com/style/article/salvator-mundi-prado-museum/index.html

[3] Jennifer Calfas. 2017. A Leonardo da Vinci Painting Just Sold for $450 Million. Here's How Experts Figured Out It Was Real. *Time Magazine* (2017). https://time.com/5028341/leonardo-da-vinci-salvator-mundi-authentication/

[4] Y. Hong and J. Kim. 2017. Art painting identification using convolutional neural network. 12 (01 2017), 532–539.

[5] Devi K.R. 2019. Art Authentication System using Deep Neural Networks. *IRJET* ISO 9001:2008= (2019). https://www.irjet.net/archives/V6/i5/IRJET-V6I51191.pdf

[6] Tianwei Lin, Zhuoqi Ma, Fu Li, Dongliang He, Xin Li, Errui Ding, Nannan Wang, Jie Li, and Xinbo Gao. 2021. Drafting and Revision: Laplacian Pyramid Network for Fast High-Quality Artistic Style Transfer. *CoRR* abs/2104.05376 (2021). https://arxiv.org/abs/2104.05376

[7] Siwei Lyu, Daniel Rockmore, and Hany Farid. 2005. A digital technique for art authentication. *Proceedings of the National Academy of Sciences of the United States of America* 101 (01 2005), 17006–10. DOI: http://dx.doi.org/10.1073/pnas.0406398101

[8] Caroline Mortimer. 2018. Modigliani art exhibited at Ducal Palace in Genoa revealed to be almost entirely fake. *The Independent* (2018). https://www.independent.co.uk/arts-entertainment/art/news/italy-modigliani-fake-show-police-investigation-art-genoa-a8154701.html

[9] Lorena Muñoz-Alonso. 2017. Why Is Modigliani Catnip for Forgers? The Fakes (and Feuds) Behind One of the Art Market's Most Dangerous Artists. *Artnet* (2017). https://news.artnet.com/art-world/modigliani-catnip-art-forgers-1033748

[10] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. *CoRR* abs/2103.00020 (2021). https://arxiv.org/abs/2103.00020

[11] Maciej Suchecki and Tomasz Trzcinski. 2017. Understanding Aesthetics in Photography using Deep Convolutional Neural Networks. *CoRR* abs/1707.08985 (2017). http://arxiv.org/abs/1707.08985

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. *CoRR* abs/1706.03762 (2017). http://arxiv.org/abs/1706.03762