ART AUTHENTICATION IN AN UNTAGGED ART DATABASE

by

Todd Dobbs

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computer Science

Charlotte

2022

Approved by:

_____

Dr. Zbigniew W. Ras

_____

Dr. Bojan Cukic

_____

Dr. Min Shin

_____

Dr. Gabriel Terejanu

_____

Dr. Heather Freeman

ABSTRACT

TODD DOBBS. Art Authentication in an Untagged Art Database. (Under the direction of DR. ZBIGNIEW W. RAS)

The identification of the artist of a painting is also known as art authentication, and the answer to this question is manifest through art gallery exhibition and is reinforced through financial transactions. Art authentication has visual influence via the uniqueness of the artist's style in contrast to the style of another artist. The significance of this contrast is proportional to the number of artists involved and the degree of uniqueness of an artist's collection. This visual uniqueness of style can be captured in a mathematical model produced by a machine learning algorithm on painting images. Art authentication is not always possible since art can be anonymous, forged, gifted, or stolen. Here we show an image only art authentication attribute marker for WikiArt, Rijksmuseum, and ArtFinder galleries. Contributions to the field of art authentication include the identification of a state-of-the-art machine learning algorithm, an extension to this algorithm, standard data sources for art galleries, standard performance measurements, standard combined measurement for accuracy and multi-class cardinality, limits to multi-class cardinality, and application recommendations for the produced models.

ACKNOWLEDGMENTS

I owe tremendous thanks to my advisor, Professor Zbigniew W. Ras. I appreciate his encouragement and guidance.

I appreciate the time and effort of all the members of my Dissertation Committee.

I am thankful to colleagues and friends, past and present, at UNC Charlotte for their collaboration and support.

While working on my PhD, I have had an opportunity to serve as an Associate Researcher at UNC Charlotte. I am grateful to Dr. Min Shin, Dr. William Tolone, and Joe Matesich for this opportunity. I have also had an opportunity to serve as an Lecturer at UNC Greensboro. I am grateful to Dr. Deng for this opportunity.

Lastly, I would like to thank my family for their patience and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1   Background

Of the online art companies surveyed in the 2021 Hiscox art trade report, the aggregate 2021 sales figures of online art is projected to grow by 72% to $13.59 billion (Table 1). Paintings consist of over 80% of these sales. Of the art buyers surveyed, 84% believe the move to online purchasing due to pandemic will become permanent method to buy art moving forward. However, of art buyers surveyed, 51% don't buy art online due to authenticity concerns, but 86% would have increased confidence with certification of authenticity [42]. With paintings being the primary driver of actual and projected sales, there is an opportunity to leverage state of the computer science techniques on painting images to help with consumer authenticity concerns.

Applying computer science techniques to art implies the need for objective observation. However, the study of art often relies on subjective observation. In the next paragraph, we briefly touch on the subjective nature of art before moving on to the objective nature of this research. Our interest in paintings is limited to objective measures we can extract from the image of a painting regardless of the convergence of the ascetics agreement.

Some paintings behave like a Veblen good meaning their value is an inverse relationship between price and demand. In other words, as the demand for the item increases,

Online art sales ($ million)

*Projected sales

13,590

7,892

4,819

4,636

4,221

3,753

3,274

2,636

1,507

| 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021* |
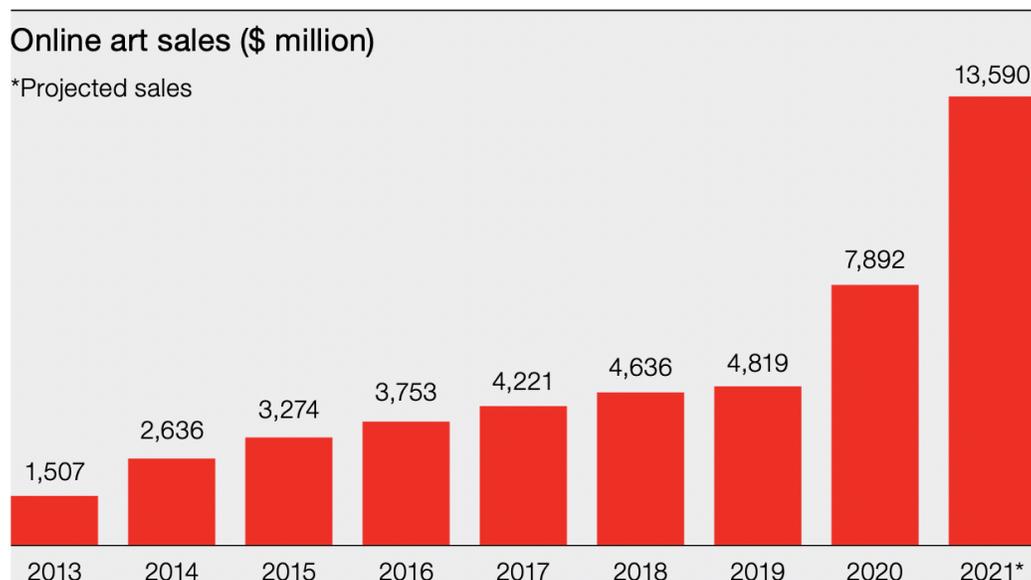
Figure 1: "A chart of increasing online art sales from Hiscox [42]."

the price does too. The drivers for the price of a Veblen good are subjective, relying on partiality, obfuscation, and notoriety [105]. The personal nature of art appreciation is further supported and scientifically measured in the growing field of Neuroscience research called Neuroaesthetics. Vessel et al.[107] define Neuroaesthetics as "A multi-disciplinary field aimed at understanding the neural basis of aesthetic experience and behavior. This includes interactions with art-objects as well as aesthetic modes of interaction with non-art objects, such as faces, natural objects, and scenes." Research in this area studies the default mode network's reaction to art. Vessel et al.[107] define the Default mode network (DMN) as "A network of brain regions typically found to be suppressed when observers engage in externally oriented tasks, which includes the medial prefrontal cortex (MPFC), posterior cingulate cortex (PCC), temporo-parietal junction (TPJ), lateral temporal cortex (LTC), superior frontal gyrus (SFG) and the hippocampus. Patterns of spatial correlation measured in the

absence of directed tasks (resting-state fMRI) support this network structure and suggest that the DMN is composed of midline hub regions (MPFC, PCC) and two subsystems." Of specific interest in our dissertation is a study performed where 16 participants are asked: "How do 109 paintings move you?". These paintings were shown in random order and were tagged somewhere in the domain of being beautiful to strange and ugly. While being scanned using fMRI, participants rated each painting. The behavioral responses were highly individual to the extent that each painting rated as highly moving by one subset of observers was rated poorly by another subset of observers (Figure 2). This study contrasts with a study by a related research group where there was a high agreement in ascetics with participants viewing real-world scenes and human faces [107]. Furthermore, eye-tracking studies performed on a variety of paintings indicate that a large variability of gazing is due to the subject's experience and knowledge, which makes the scientific study of art very complex and challenging [82].

Concerning the knowledge surrounding a piece of art, a data source is needed to conduct research. Online companies such as Artnet, Artprice, Blouin, AMR, and Sotheby's have developed online databases derived from their sales and sales provided to them from other companies around the world. A high-level review of these online art databases reveals that we can rely on some textual tags. For example, commonly available features are available such as artwork name, dimensions, and medium; artist name and biography; provenance; and in rare cases the asking and sales price. We found no evidence of features based on the image, such as dominant color, color histograms, texture, and objective artist style. However, some of these simple features
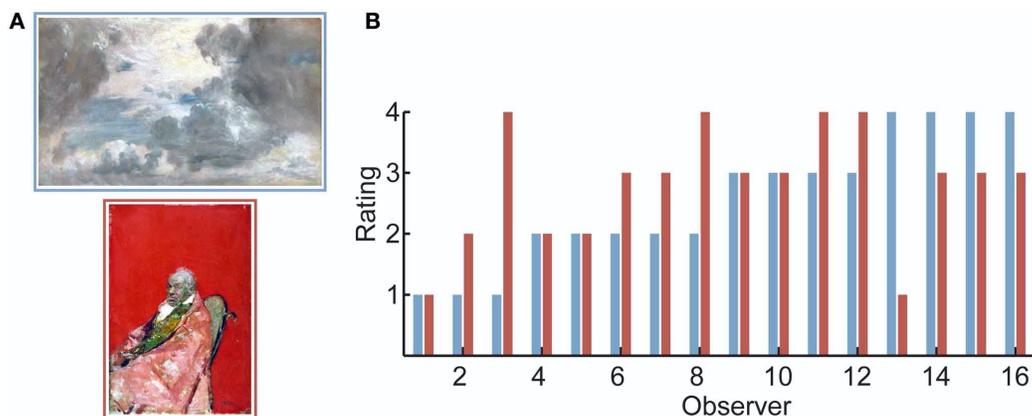
Figure 2: "(A) Two sample images from the set observers were shown. Images were reproductions of museum artworks that are not commonly reproduced. Observers rated each image for how much the artwork "moved" them on a scale of 1 (lowest) to 4 (highest). (B) Ratings of all 16 observers for the two images in (A). As was typical for the artworks used in the experiment, observers differed widely in their response to the pair of images. In particular, some observers rated the top image (blue bars) to be highly moving, while others rated the bottom image (red bars) to be highly moving [107]
." Can we find features in paintings that are objective?

would likely be available through art historians who assist customers interested in higher value works of art (Table 1) [3, 4, 10, 85, 94]. Likely, features based on fundamental and state of the art digital image processing and machine learning algorithms are not available. This dearth of information presents an opportunity for discovery. Since all online art database companies hold their information close, there is often a paywall to access a proper sample of the art information for research. Moreover, general data protection regulations and copyright information prevent information from being retrieved from these company websites in an automated fashion.

Art 500k, Linked Art, Google, MoMA, OmniArt, Rijksmuseum, Web Gallery of Art, and WikiArt are examples of organizations that provide a collection of art with open and free access. These sources are listed in table 1 and provide painting images,

Table 1: Organizations that provide painting art data

Organization	Open Source	Paywall
Art 500k	Yes	No
Art Market Research (AMR)	No	n/a
Art Price	Yes	Yes
Artnet	Yes	Yes
Artsy	Yes	No
ArtFinder	Yes	No
Blouin	Yes	Yes
Google	Yes	No
Linked Art	Yes	No
Museum of Modern Art (MoMA)	Yes	No
OmniArt	Yes	No
Rijksmuseum	Yes	No
Sotheby's	No	n/a
UGallery	No	No
Web Gallery of Art	Yes	No
WikiArt	Yes	No

related textual information, and in some cases, an application programming interface or API for free [65, 23, 73, 97, 24, 72, 36, 74].

For contemporary paintings, the primary creative and sales market is very controlled where the artist and dealer relationship solidifies via exclusive contract for a specific type and amount of work, and the dealer vets the buyer's commitment to collecting. Even the secondary market succumbs to control via the manipulation of reserve pricing and bidding. The primary and secondary markets hardly follow the free-market paradigm [95]. To navigate the evolving market, the traditional model of selling contemporary art is one of an art dealer investing in emerging artists such that they can successfully grow old and live symbiotically together. Even in times of an economic downturn such as the economic recession in 2008, similar strategies are used by galleries and dealers to navigate the change such that the contract oriented

nature of contemporary art scales by the growth in market size [113]. However, this research aims to identify objective influences in the parties of an art transaction. A painting's scene, color, and pattern may be desired for decorating, emotional, or gift-giving concerns. With budgets being a partial driver to a consumer's decision to obtain a piece of art, one may desire to take a pragmatic approach and view many pieces of art that satisfy these criteria. On the flip side, its conceivable that an artist and their representative dealers would want to be mindful when it comes to the popularity of such objective pieces of information. The opportunity here is to identify those objective influences for all parties involved.

With the advent of machine learning and its application to digital image processing and natural language processing, the use of various algorithms extract information from images. For example, short sentences and paragraphs can be generated from an image (Figure 3) [48, 49, 54]. To mine authenticity knowledge from the related image of a painting, we initially researched state of the art image tasks including classification, color analysis, object detection, and object description to determine how a class of computational visual and textual intelligence forms.

With little effort, humans can mine knowledge about authenticity from a painting by noticing properties and events of a painting that require semantic understanding. This seemingly simple capability remains to be a challenging problem for computer systems to address using digital image processing algorithms. Research in this area remains very active and includes image captioning, classification, labeling, and segmentation.

With the extraction of a digital image from a painting, a similar challenge exists, and in some cases, is complicated further through artistic freedom. For paintings,
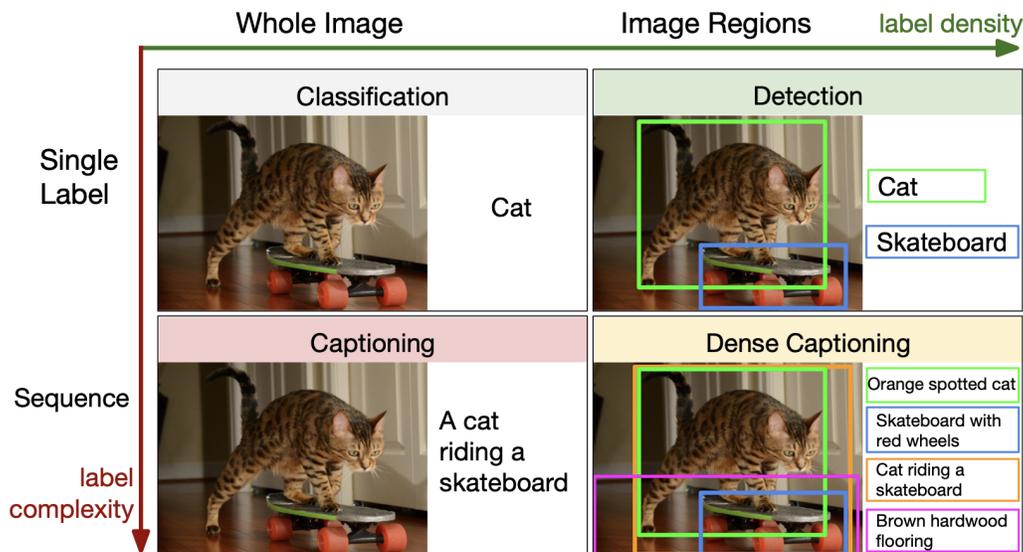
Figure 3: "We address the Dense Captioning task (bottom right) with a model that jointly generates both dense and rich annotations in a single forward pass [48]." One can apply one or more of these methods to a painting.

one can extract knowledge to classify the artist, emotion, and style of painting. Furthermore, a painting can be labeled and segmented to pull out prominent objects and events or partition regions to provide further knowledge. Much of the knowledge to be discovered will be factual. However, artistic freedom fosters the discovery of numerous illusory features. For example, consider the Nighthawks painting by George Hopper (Figure 4). From an objective point of view, it is reasonable to believe that the state of the art machine learning techniques could identify four people and segment the building, street, and sidewalk regions in the painting. We may even be able to discern that the people are sitting in a dinner, among other things. However, would the same techniques pick up that the light emitted from the dinner is likely fluorescent, which had just come into use in the early 1940s or the concept of human isolation in an urban setting? How about the juxtaposition of the fluorescent light with this isolation and the sterile emotion that can evoke?

For the remainder of this chapter, we discuss our initial work to extract knowledge from the image of a painting for authenticity purposes. From a raw digital image processing perspective, we discuss several different concepts involved with extracting information and some of the completed research in these areas. From a painting lens, we discuss some of the finished work to discover artistic knowledge from painting. We also discuss the data that supported our initial work as well as the data we use for our dissertation focus. Lastly, we discuss our initial methods and experiments as well as those that we used for our dissertation.

1.2   Related Work

Related work from digital image processing concepts of image classification, object detection, and image captioning provides a solid foundation for discovering knowledge from a painting image. Image classification's main task is to assign one label to an image. Since 2012, research produced compelling progress in this field [59, 88, 102, 69, 47]. Object detection's main task is to assign one label to each region of an image that is considered notable [35, 84, 90, 103, 47]. Lastly, the task of image captioning is to pull together the outputs of object detection into a meaningful sequence of words to describe the image [20, 21, 29, 55, 66, 108, 115, 47].

1.3   Classifying Artistic Style

Falomir et al.[30] categorize paintings in art styles based on qualitative color descriptors, global quantitative features, and machine learning (QArt-Learn). This process categorizes color in the Baroque, Impressionism, and Post-Impressionism art styles. It leverages qualitative color descriptor (QCD) (Figure 5) and its associated similarity (SimQCD)

utilizing k-nearest neighbor (k-NN), support vector machines (SVM), and machine learning techniques to classify paintings using a palette with $\approx 65\%$ accuracy. This technique can enable an artificial agent to describe a painting's color style to a human. QCD creates a relational reference graph in the Hue, Saturation, and Lightness (HSL) color space. The resulting Cartesian system describes the necessary transformations in HSL necessary to move from one node in the graph to another [30]. Likewise, Shamir et al.[91] automate the recognition of painters and schools of art. The following schools of art are labeled:

- Expressionism

- Impressionism

- Surrealism

This method focuses on determining the school of art before determining the artist. A critical challenge to this type of classification is reconciling the holistic view of creation used by an artist versus the feature-specific approach used by an algorithm. A Fisher score of the 11 algorithms used for feature extraction produces an average classification accuracy of 71% (Table 2) [91].

Cetinic et al.[16] fine-tune convolutional neural networks to classify fine art. Training strategies with a deep CNN structure automate the retrieval of metadata such as genre and style, which originate from formal elements such as color, composition, mass, shape, and texture. The impact of different weight initializations and convolutional layers pre-trained on different source domains have been investigated, and when the target dataset has many classes with fewer images per class, the pre-trained

Table 2: The 11 algorithm components used to calculate the fisher score

Algorithm
Chebyshev statistics
Chebyshev-Fourier features
Edge statistics features
First 4 moments
Gabor filters
Haralick features
Multiscale histograms
Object statistics
Radon transform features
Tamura texture features
Zernike features

model initialization influences fine-tuning performance. Moreover, they show that fine-tuning scene and sentiment recognition works better than object recognition [16].

1.4   Classifying Artistic Color

Sanz et al.[89] use qualitative color descriptions to customize for adaptability and usability and produce a model that labels a color name 93% of the time in a similar fashion to how humans label color using language and to aid in the process of classification. The remaining 7% of color labels correspond to a close color. The influence of labeling is not affected by the native language or the device used [89]. Likewise, Yelizaveta et al.[118] analyze and retrieve paintings using artistic color concepts, machine learning, and digital image processing techniques to classify painting regions with color concepts such as color temperature, color palette, and color contrast. Overall, this methodology is effective with labeling color temperature providing the best results of up to 93% accuracy [118]. In a similar study, Strezoski et al.[100, 98] have created a tool called ACE to study art, color, and emotion. This

tool uses the OmniArt dataset to draw strong correlations between time, sentiment, and color on a large scale [100, 98].

1.5   Classifying Artist

The artist who produced renaissance paintings are identified automatically by Jou et al.[51] The team uses Naïve Bayes, Linear Discriminant Analysis, Logistic Regression, K-Means, and SVMs to achieve a classification accuracy of 65% for paintings across five artists. Histogram of color and gradient features are inputs to this procedure [51]. Johnson et al.[46] detect painting forgeries by showing how features stand out when generated by Ma analysis in which contours, multiscale analysis, and local textures representing brushstrokes and patterns are key features. Using SVMs on the histograms generated from Ma analysis features, 75% accuracy is obtained for van Gogh forgeries [46]. Using machine learning for the identification of art paintings, Blessing et al.[8] obtain 85.13% classification accuracy for seven artists by leveraging machine learning techniques on a histogram of gradients (HoG) features [8]. Artist identification with convolutional neural networks produces 77.7% artist classification accuracy by Viswanathan et al.[109] with 57 artists using ResNet-18 pre-trained on ImageNet with transfer learning [109]. A similar ResNet implementation achieves 74.7% accuracy across 15 artists by Chen et al.[19] while comparing machine learning techniques for artist identification [19].

1.6   Classifying Artistic Medium

Yang et al.[116] also performed artwork stroke recognition to identify artwork medium. Leveraging WikiArt, YMSet (contemporary art dataset), and SynthSet

(synthesized dataset), the DenseNet CNN performed best with aggregate F1 scores of 79%, 89%, and 67% respectively for oil, pastel, pencil, and watercolor classification [116]. Yang et al.[117] extended this research using multiple states of the art CNNs to show better results than human trials when classifying oil, pastel, pencil, and watercolor paintings. They found that the 2017 ImageNet winner DenseNet performs best with an accuracy of 85%. This research stands out due to its experiments with the last five ImageNet winners since the last competition in 2017. Moreover, a meaningful baseline gives the comparison of human and non-photorealistic rendering (NPR), which simulates the artistic effects of art produced with a medium of oil, pastel, pencil, or watercolor [117].

1.7 Private Projects

Just as artists produce art in their spare time, art appreciators research art. Jason Bailey created the Artnome blog website as a platform to fulfill his mission "to use technology and data to improve the world's art historical record and to improve opportunities for artists from historically underserved or marginalized groups." He aims to build the world's largest analytical database of known artwork. Several of his projects include defining art analytics, machine learning applications of art, AI art, and analysis of the color palette of artists [5]. Discussions with Jason reveal that his art database sources an artist's official compendium, which is known as a catalog raisonné. "A catalog raisonné typically lists each piece's title, dimensions, date, medium, location, provenance, exhibition history, condition, and occasionally even more [1]." The Artnome blog is current as of March 2020 and contribute several

small non-academic related projects in the field of Art Analytics.

Likewise, Ahmed Hosny, who is a machine learning researcher contributing numerous academic papers to biomedical research, created a project called "The Green Canvas." This project involves private research directed at quantifying aesthetic features of visual art to determine how the artistic and statistical nature can relate to the valuation of contemporary visual art. Initial research with machine learning techniques have shown some exciting pricing and sales relationships when it comes to sales year, color, corner percentage, and exhibitions [44].

1.8   Dataset

To perform analysis on painting images, we need example painting images with associated ground truth labels to apply our methodologies in an experiment to validate this work. We were unable to find such a dataset to be available outside of a paywall. Therefore, we decided to include paintings from WikiArt Visual Art Encyclopedia. WikiArt is a private non-profit project to make the world's visual art accessible to everyone. At the time of this research, over 250,000 works are available for analysis [24].

After our preliminary experiments, this research uncovered a large-scale artistic project called OmniArt. This project has collected over 2 million photographic reproductions of art. It features 1,348,017 indexed images with full annotations and 702,000 more unlabeled images with incomplete metadata. OmniArt is linked to its sources via API if available, which keeps content current. Metadata on the collected artworks contains standard, source-specific, object level, image metadata level, and

reproduction information. The data goes through an automated cleaning step when added to the data set. Moreover, a VGG like architecture provides a classification for an artist, type, genre, school, creation period, style, object detection, and color analysis [100].

1.8.1 Data Collection

We imported a total of 163,241 pieces of art into our database from WikiArt. A total of 47,061 of these are paintings that were produced by 3,102 artists. These are all usable in our preliminary research.

1.8.2 Existing Data Features

The features in table 3 exist in WikiArt for extraction and further analysis. There is an ample domain of labeling values related to each painting for artist, period, genre, and style.

1.8.3 Discovered Data Features

The features in table 4 were discovered for extended analysis and feature discovery via MATLAB, LIRE (Lucene Image Retrieval) [63], and custom code. We do not delve into the details of each of these features for the sake of brevity. However, we discuss the details of the selected features used in our experiments in the next paragraph.

1.9 Method

The methodology used for feature extraction consists of a simple algorithm that iterates over each painting and extracts the feature vectors using custom code or

Table 3: WikiArt Features

Artist	Artwork
Birth Day	Artist
Death Day	Auction
Image URL	Completion Year
Name	Description
Wikipedia URL	Gallery Name
	Genre
	Height
	Image URL
	Last Price
	Location
	Material
	Period
	Series
	Style
	Tags
	Technique
	Title
	Width
	Year of Trade

software library. Our custom code involves various calculations on the discrete colors and textures of an image. We think it is essential to talk about the general methodology of the Fuzzy Color Histogram and Pyramid Histogram of Gradients and how they address our digital image processing methodology. Further detail about other color and texture features we extracted are in Appendix A and Appendix B.

1.9.1   Texture

Analyzing the striking differences of intensity in an image produces the texture of an image and is a primary method for texture extraction. From this analysis, calculated features including edges, corners, and localized gradient provide a feel, appearance, or consistency to the surface. Textures can be represented by distilling

an image to the black and white pixels that produce a sketch of an image or by partitioning an image up into bins that generalize one or more gradients of the image in that partition [22].

1.9.2   Pyramid Histogram of Gradients

The Pyramid Histogram of Gradients or PHoG is similar to the Histogram of Gradients, which uses an overlapping grid of multiple localized gradients utilized for object detection in digital image processing [25]. The Pyramid variety of this methodology uses a pyramidal kernel and can increase performance by 10% [11]. Figure 6 shows a visualization of HoG features of two people in the Nighthawks painting.

1.9.3   Color

Color is a dominating factor when humans view a painting. The continuous intensity of light is the source of color, making a human's perception of color the visible spectrum of light. The visible spectrum of light is analog by nature. However, for the ubiquitous computing system, this source of light needs special processing for consumption. This unique processing consists of measuring and transforming into discrete values for processing. Discrete values are commonly known as pixels, and the tessellation of these values, commonly known as a raster, is perceived by the human eye as an image. The pixels used in our methodology break down into three numbers or intensity channels of red, green, and blue or RGB for short. As the max value of these channels increases, more colors are available for representation. One bit gives two color options, and 24 bits yields close to 17 million color options.

When analyzing color, these numerical values are presented in a mathematical matrix to facilitate further processing. While it is possible to break a pixel down into a bit distribution or sub-pixels, we stop at the pixel as an atomic measure for our processing methodologies.

With the development of the multimedia content description standard or MPEG-7, the term Color Space defines a set of continuous numerical values that define a color. As we have already mentioned, RGB is one type of Color Space. Transformations can be applied to RGB to generate YCbCr, which is a simple linear transformation from RGB; Hue, Saturation, and Value or HSV; Hue Max-Min Difference or HMMD; and Monochrome, which is just the Y component of YCbCr. All of these define a part of the MPEG-7 Color Space standard. For discrete processing, a color quantization transformation drives digital analysis in any of these color spaces. For our methodology, we leverage a discrete Color Space of varying intensities based on the RGB distribution [75, 22].

1.9.4 Fuzzy Color Histogram

The Fuzzy Color Histogram or FCH leverages the concept of the Fuzzy Set, which allows for partial set membership. FCH mitigates common issues of brightness and dimension in Content-Based Image Retrieval or CBIR tasks. The WEB-CRAWLED database is an internationally used database to benchmark CBIR tasks. Using a statistically significant sample from this database, FCH outperformed CCH tasks. For queries expecting nine matches, Recall for FCH tasks ranged from 78% to 89% for 7 to 8 average best matches, respectively, where CCH performs at 55.5% for

five average best matches. For the same match expectations and images with a 43%

decreased brightness, FCH recall is 55% to 58%, where CCH performs at 33% [6, 38].

Our confirmation of these results drives the use of the Fuzzy Color Histogram feature

vector in this research.

1.9.5 Neural Pattern Recognition

To mine for knowledge at a higher level, we wanted to demonstrate an artificial

intelligence methodology in this dissertation. This methodology takes an artwork

painting as input and retrieves the related color and texture features from a database

where those features have been pre-calculated. We then plan to use these features to

classify the paintings using existing feature labels. Since we have extracted several

different types of color and texture features, we reduce the dimensions considered

by picking popular selections from existing research [51, 46, 8, 109, 19]. We reduce

features to fuzzy color to address the color component and histogram of gradients to

address the texture component. The resulting feature vector contains 575 elements.

We ensure balanced inputs during the data extraction phase. This reduction and

balancing produce a feature vector to be analyzed by MATLAB's Neural Pattern

Recognition application (Figure 7). We use the default hyperparameter settings

for the Neural Pattern Recognition tool other than the input features and output

labels, which are specific to the experiment (575 and 20, respectively). The MATLAB

model is a two-layer feed-forward network. We use the default of 10 neurons in the

first layer's hidden network with a sigmoid activation function. We use the default

softmax output neuron in the second layer. We use 70% of the input for training,

15% for validating to prevent overfitting, and 15% for testing. We run for a max of 1000 epochs, and we focus on the output confusion matrix of the test results where f-measure score, precision, and recall reside. We repeat the test with all 19 training algorithms provided by MATLAB. The table 5 lists these algorithms [68].

1.10   Experiment

As mentioned in our methodology section, we use MATLAB's Neural Pattern Recognition application to classify a label defined in WikiArt by fuzzy color and gradient histograms calculated beforehand. We do not pre-process images via clipping or blurring, but the magnitude of each feature vector is the same. Therefore, we are performing classification by running a neural network over the feature vectors that make up the color and texture quadrants of paintings in hopes of modeling an artist's style. We decide to start with artists that have a thousand paintings and decrease this count until we find 80% F-measure for classifying 20+ labels. The experiments execute on an iMac with 3.2 GHz Quad-Core Intel Core i5 processor, 32 GB RAM, and AMD Radeon R9 M380 2 GB GPU.

Our next step in the experiment is to determine what labels to classify. Given the results of existing research, we decide to attempt similar experiments of artist classification to see if we can reproduce or improve results [51, 46, 8, 109, 19]. A complete set of metrics is saved for each experiment. For brevity, we only show F-measure per algorithm until we reach our goal, where we show the confusion matrix with all metrics.

Our initial experiment to classify the eight artists that have 1000 paintings in the

database yields an F-measure of 77.78% using Cyclical order weight/bias training (trainc). Table 6 lists all training results. While the F-measure is favorable, the number of classifiers is too small.

We halve the number of paintings in our next experiment, which yields 22 artists to classify. Cyclical order weight/bias training (trainc) still performs best. However, the F-measure is quite low at 44.76%. Table 7 lists all training results.

Again, we halve the number of paintings required for artist cutoff to 250, providing our experiment with 59 artist classifiers. Scaled conjugate gradient backpropagation (trainscg) provides the best F-measure with superb performance. However, the F-measure is very low at 25.23%. Table 8 lists training results.

With the reduction of painting count threshold and the increase of artist classifiers, our F-measure performance is proportional to painting count and inversely proportional to the number of artist classifiers. We decide to run one more test of the best artist classifiers with 250 pieces of art. With this experiment, we approach very close to our goal with an F-measure of 79.40%, albeit with the top-performing artists from the previous experiment. Table 9 lists all training results. The full confusion matrix for this run can be seen in table 8.

1.11   Discussion

This concludes our initial work with extracting knowledge from digital images extracted from paintings. We explore both the intrinsic features such as image dimension and medium, which are readily available for extraction and features which require simple and advanced mathematical and algorithmic methodologies. While our

experiments with the methods built into MATLAB provide overall promising results, we learned that our models overfit as indicated by the relationship of testing and validation scores. Given these results, we wrap this chapter up by discussing two experiments to show why our data set supports continued research, the challenge of reproducing existing research, and how we move forward with this research.

To ensure our test data set provides an adequate variety of paintings for analysis, we run a simple visual experiment to represent this variety. Our main experiment in this chapter randomly pulls 250 pieces of art for the 59 artists who have 250+ pieces of art in the Wiki Art database. From these 59 artists, we select the top 20 performing artists as a basis for this work. This supporting experiment simply displays an nxn centered region of a random painting from each of the 20 artists and produces a tessellation of artwork. We reviewed multiple instances of this generated tessellation, and visible results indicate that there are no patterns to the artwork selected that make this a trivial problem. Figure 9 shows the results of a generated tessellation.

The visual test in the previous paragraph provides some anecdotal evidence for a solid foundation in a variety of art. However, we wanted to conduct an additional experiment to provide actual numbers to measure the similarity between all artists and all pieces of art to support our painting variety claim. To achieve this, we calculate the Manhattan distance between the fuzzy color and HoG features that we extracted for each piece of art. A smaller similarity score between the two paintings indicates a closer similarity between the two. For example, a similarity score of zero indicates nearly identical paintings. We use these pixel feature components to ensure a one to one mapping. This analysis yields results of an average similarity between paintings

of 1,857 with a standard deviation of 252. As a rule of thumb, this indicates that most of our data have similarity scores within two standard deviations of the mean, which means most of the similarity measures between paintings are not outliers, and the analysis considers these numbers. In raw numbers, this leaves seven (3.68%) similarity measures as outliers. Figure 10 shows a graph of similarity. Figure11 shows the most similar and dissimilar pieces of art between all artists in the random sample.

Our experiments were mainly influenced by state of the art research with artist classification and convolutional neural networks. 57 artists with 300 paintings and an F1 score of 77.1% is the highest combined results of all the research reviewed. There are no links in the research to the code and data for reproduction [109]. We found the painter by numbers challenge in Kaggle, and again, there were no results in the leader boards for this challenge by the author [53]. Moreover, the author's email address is no longer valid, so there appears to be no way to contact other than a potential social network connection. Having no way to reproduce these results from the existing source code caused some concern. Therefore, we started with LIRE and MATLAB tools to minimize the effort to see if we could produce some results to indicate whether we should move foreword. While our results with no transfer learning outperform this paper's results, albeit, with 37 fewer artists, we feel like they are good enough to move forward with more detailed research. Also, there could be a discrepancy with the art and artist used. Even though the root source for this work is Wiki Art, our feed shows 59 artists with 250+ paintings, which do not foot with the author's 57 artists with 300 paintings.

Given these experiments and findings from initial research, we moved forward with

new experiments to serve as a basis for high-level feature extraction to support art authenticity in this research. First, we continue research on 59 artists from WikiArt leveraging transfer learning and ResNet from the ImageNet challenge [96]. This work has a significant impact on our continued research from a research and tools perspective as it produces the results that surpass ImageNet solutions applied to the medium classification from Yang et al. [117].

The goal of this dissertation is to develop and apply algorithms using state of the art machine learning algorithms to assist in automated art authentication using painting images. We plan to discover knowledge from paintings to produce an artist's style model that can be leveraged by artists, consumers, and dealers. We answer the following questions:

- Which state-of-the-art machine learning algorithm does the best job classifying artists given images of their artworks?

- How can we document the datasource for experiments such that experiments can be reproduced and improved?

- What measure does the best job to determine the performance of an algorithm given the unbalanced nature of artwork data?

- How can the management of the algorithm selected increase performance?

- How can we interpret the fitness of the selected algorithm from an experiment performance and multi-classification cardinality perspective?

- What are the multi-classification cardinality limits to the selected algorithm?

- How can classification results be applied to an artist's catalogue raisonné and as an art authentication attribution marker?

1.12   Dissertation Outline

This dissertation is organized into three significant sections which are represented by chapter two, chapter three, and chapter four. Each chapter represents discovering and applying a state-of-the-art technique for art authentication on a new datasource which represents a unique set of paintings.

1.12.1   WikiArt and the Catalogue Raisonné

In chapter 2, we discuss the catalogue raisonné which is compiled by art scholars and holds information about an artist's work such as a painting's image, medium, provenance, and title. The catalogue raisonné as a tangible asset suffers from the challenges of art authentication and impermanence. As the catalogue raisonné is born digital, the impermanence challenge abates, but the authentication challenge persists. With the popularity of artificial intelligence and its deep learning architectures of computer vision, we propose to address the authentication challenge by creating a new artefact for the digital catalogue raisonné: a digital classification model. This digital classification model will help art scholars with new artwork claims via a tool that authenticates a proposed artwork with an artist. We create this tool by training a machine learning model with 90 artists having at least 150 artworks and achieve an accuracy of 72.96%. We use the ResNet Convolutional Neural Network to improve accuracy and number of artists classes over state-of-the-art artist classification experiments using the WikiArt database. We address inconsistencies in the way scholars approach

artist classification by providing a consistent method to recreate our dataset and providing a consistent method to calculate performance metrics based on imbalanced data.

1.12.2   Rijksmuseum and Annealing

In chapter 3, we discuss art authentication which assures that a piece of art is created by an artist. A certificate of authenticity created from proper art authentication significantly increases the value of a piece of art which impacts all parties in an art transaction. The models produced by machine learning algorithms provide an objective measure to authenticate an artist to their artwork collection. In the past ten years numerous machine learning algorithms have been used to address art authentication on a variety of datasets. This work extends art authentication with residual neural networks and the Rijksmuseum data set. Our results show contributions is four key areas: A performance increase of 11.35% over the baseline for 34 artists; A new baseline for 1,199 artists; A standard methods for recreating the Rijksmuseum data set; and A standard method for measuring results from imbalanced data for the Rijksmuseum data set.

1.12.3   ArtFinder and Large Classification

In chapter 4, we discuss how the identification of the artist of a contemporary painting answers the question who painted the artwork. This is also known as art authentication, and the answer to this question is manifest through art gallery exhibition and is reinforced through financial transaction. Art authentication has visual influence via the uniqueness of the artist's style in contrast to the style of

another artist. The significance of this contrast is proportional to the number of artists involved and the degree of uniqueness of an artist's collection. This visual uniqueness of style can be captured in a mathematical model produced by an ML algorithm on painting images. However, art authentication is not always possible for contemporary art since art can be anonymous, forged, gifted, or stolen. Here we show an image only art authentication attribute marker of contemporary art for a very large number of artists. We found that it is possible to authenticate contemporary art for 2,368 artists with an accuracy of 48.97%. These results come from a model generated from a contemporary art database of 170,056 paintings and tested on 42,514 paintings from the same artists but unbeknownst by the model. Our results demonstrate the largest effort for image only art authentication to date with respect to the number of artists involved and the accuracy of authentication.

1.13 Previously Published Work

Published bodies of work related to digital image processing and machine learning comprise a wealth of previous research. Machine learning and its applications have been very popular lately, and the results from research have been auspicious. Each of the three main chapters in this dissertation will delineate information regarding previously published work.

1.14 Participants

The participants involved in this research are the artists who produced the paintings, which are analyzed. Textual information related to the artists and their paintings is not involved. Where features extracted from painting images are not in the public

domain, results that would make artists and their artwork identifiable are anonymized. Therefore, clearance from the Institutional Review Board or IRB is not required.

In this dissertation, the participants involved in this initial research are those whose art has been collected by the ArtFinder, Rijksmuseum, and WikiArt websites. This domain of art-related data including number of artists and artworks is described in further detail in chapter two, chapter three, and chapter four. In this domain of data, selection criteria are painting classification and sufficient painting sampling size by the artist. In other words, this research does not consider works of art that are not considered a painting or artists who have not produced a sufficient sample of paintings deemed usable for learning their style.

Figure 4: "Edward Hopper said that Nighthawks was inspired by "a restaurant on New York's Greenwich Avenue where two streets meet," but the image—with its carefully constructed composition and lack of narrative—has a timeless, universal quality that transcends its particular locale. One of the best-known images of twentieth-century art, the painting depicts an all-night diner in which three customers, all lost in their own thoughts, have congregated. Hopper's understanding of the expressive possibilities of light playing on simplified shapes gives the painting its beauty. Fluorescent lights had just come into use in the early 1940s, and the all-night diner emits an eerie glow, like a beacon on the dark street corner. Hopper eliminated any reference to an entrance, and the viewer, drawn to the light, is shut out from the scene by a seamless wedge of glass. The four anonymous and uncommunicative night owls seem as separate and remote from the viewer as they are from one another. (The red-haired woman was actually modeled by the artist's wife, Jo.) Hopper denied that he purposefully infused this or any other of his paintings with symbols of human isolation and urban emptiness, but he acknowledged that in Nighthawks unconsciously, probably, I was painting the loneliness of a large city [34]."

What are the literal and artistic pieces of knowledge to be extracted?
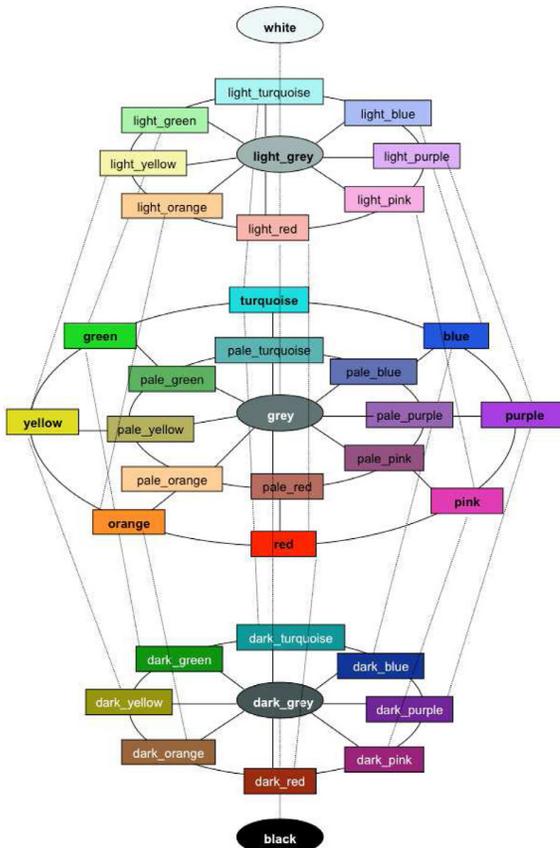
Figure 5: "The Qualitative Color Descriptor or QCD represents a relational reference graph in the Hue, Saturation, and Lightness (HSL) color space. The resulting Cartesian system describes the necessary transformations in HSL necessary to move from one node in the graph to another.[30]"
Qualitative Color Descriptor

Table 4: Custom Features

Feature	Custom	LIRE	MATLAB
Brightness Dimension Average	x		
Brightness Contrast	x		
Canny Edge Features	x		
Color and Edge Directivity Descriptor (CEDD)		x	
Color Layout		x	
Entropy - Red	x		
Entropy - Green	x		
Entropy - Blue	x		
Entropy	x		
Fuzzy Color		x	
GIST			x
Haar Wavelet Histogram		x	
HoG (Histogram of Gradients)	x	x	x
Hue Arithmetic Average	x		
Hue Circular Average	x		
Kaze Edge Features	x		
Lightness Arithmetic Average	x		
Local Binary Patterns			x
MPEG-7 Simple Color Histogram	x	x	x
MPEG-7 Scalable Color	x	x	x
MPEG-7 Edge Histogram		x	
PHoG (Pyramid Histogram of Oriented Gradients)		x	
Saturation Arithmetic Average Cylinder	x		
Saturation Arithmetic Average Bicone	x		
SIFT (Scale-invariant Feature Transform)		x	x
SIFT - Dense			x
SIFT - Sparse			x

Figure 6: Each cell contains an aggregate of the gradients calculated within. A star-like pattern in the cell visualizes the aggregate.
Histogram of Gradients (HoG) of two people in the Nighthawks Painting
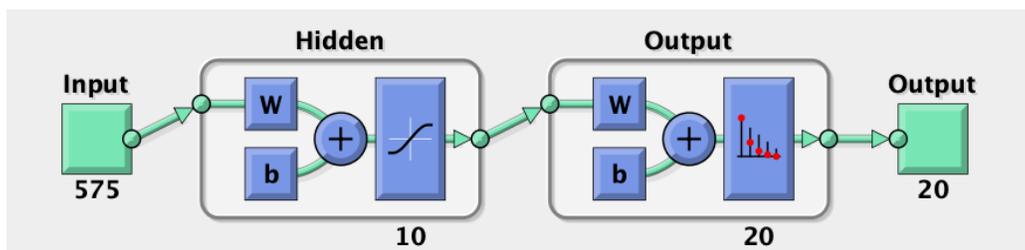
Figure 7: "Pattern recognition uses the standard network, which is a two-layer feed-forward network, with a sigmoid transfer function in the hidden layer, and a softmax transfer function in the output layer. The default number of hidden neurons is 10. The number of output neurons is 20, which is equal to the number of elements in the target vector (the number of categories) [68]
MATLAB's Neural Network Tool."

Table 5: MATLAB Training Methods

Key	Description
trainb	Batch training with weight & bias learning rules
trainbr	Bayesian Regulation backpropagation
trainbfg	BFGS quasi-Newton backpropagation.
traincgf	Conjugate gradient backpropagation with Fletcher-Reeves updates
traincgp	Conjugate gradient backpropagation with Polak-Ribiere updates
traincgb	Conjugate gradient backpropagation with Powell-Beale restarts
trainc	Cyclical order weight/bias training.
traingd	Gradient descent backpropagation
traingdx	Gradient descent w/momentum & adaptive lr backpropagation
traingda	Gradient descent with adaptive lr backpropagation
traingdm	Gradient descent with momentum
trainlm	Levenberg-Marquardt backpropagation
trainoss	One step secant backpropagation
trainr	Random order weight/bias training
trainrp	RPROP backpropagation
trainscg	Scaled conjugate gradient backpropagation
trains	Sequential order weight/bias training
trainbu	Unsupervised batch training with weight & bias learning rules
trainru	Unsupervised random order weight/bias training

Table 6: Classifying eight artists with 1000+ paintings

Training Function	Training Time (sec)	F-measure
traingdm	14	32.23%
trains	72	32.69%
trainb	51	33.29%
traingd	14	35.89%
traincgf	2	51.89%
traingda	2	52.18%
trainrp	1	56.46%
traincgb	6	58.56%
trainscg	1	60.30%
traingdx	2	61.00%
trainr	125	62.84%
trainbfg	6,033	63.39%
traincgp	2	65.03%
trainoss	4	65.89%
trainc	5,796	77.78%

Table 7: Classifying 22 artists with 500+ paintings

Training Function	Training Time (sec)	F-measure
traingd	24	5.17%
traingdm	24	7.13%
trainb	78	7.93%
trains	122	8.36%
traingda	4	19.28%
traincgb	6	25.75%
traincgf	4	28.88%
trainrp	2	31.60%
traingdx	5	32.23%
trainbfg	5,124	33.59%
trainscg	3	35.61%
trainr	436	36.25%
traincgp	5	38.60%
trainoss	15	39.11%
trainc	7,993	44.76%

Table 8: Classifying 59 artists with 250+ paintings

Training Function	Training Time (sec)	F-measure
traingd	100	1.23%
trainb	263	1.51%
traingdm	104	2.08%
trains	415	2.09%
traingda	18	5.60%
traingdx	24	13.93%
trainoss	70	18.79%
traincgf	30	21.44%
traincgb	27	21.86%
trainr	3,510	21.90%
trainbfg	1,793	23.06%
trainrp	26	23.76%
traincgp	38	23.86%
trainc	10,617	24.80%
trainscg	35	25.23%

Figure 8: "Confusion Matrix for all runs for the top 20 performing artists who have at least 250 paintings [68]."
Confusion Matrix

Table 9: Classifying the 20 best performing artists with 250+ paintings

Training Function	Training Time (sec)	F-measure
traingd	13	11.18%
trainb	40	12.04%
traingdm	13	10.14%
trains	53	11.00%
traingda	2	38.68%
traingdx	3	55.72%
trainoss	6	64.68%
traincgf	2	54.28%
traincgb	5	56.44%
trainr	290	63.08%
trainbfg	6,019	54.92%
trainrp	0	44.86%
traincgp	2	57.16%
trainc	3,612	79.40%
trainscg	2	57.26%
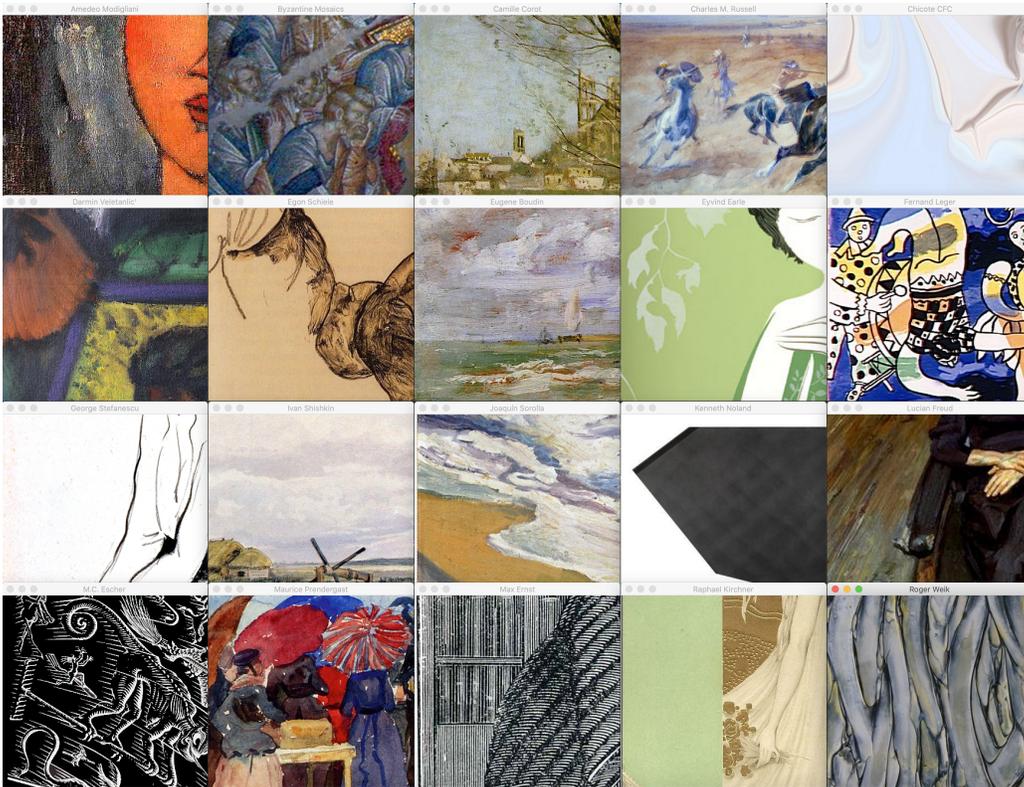
Figure 9: "This image is a sample of one painting for each of the 20 artists used in our experiments. We visually reviewed 20 such random samples from the 20 artists to verify that the paintings classified do not exhibit apparent differences, which would make classification trivial."
Painting Variety
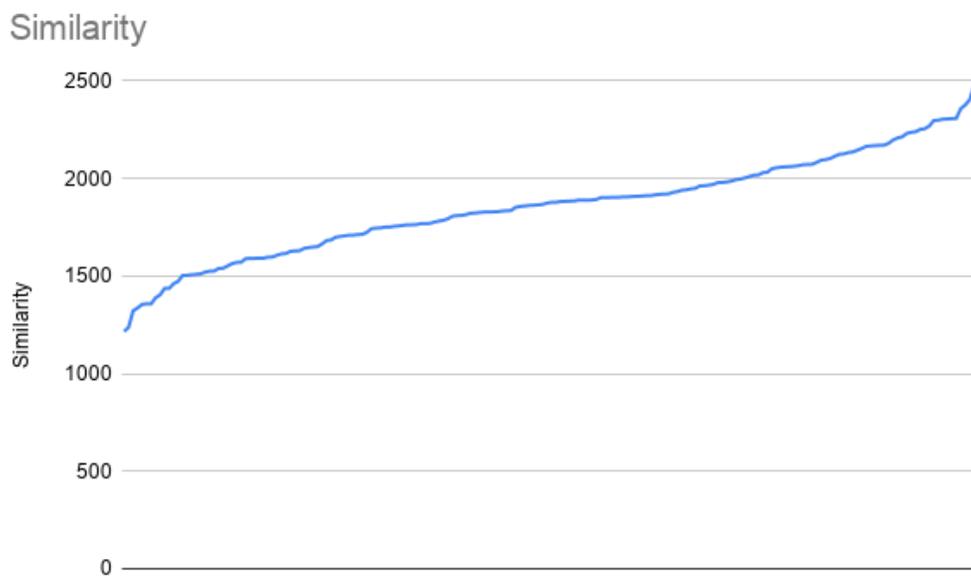
Figure 10: "Similarity scores between all artists range from 1,216 to 2,476, where similarity of zero is considered no difference. Manhattan distance of pixel component values calculates the similarity between two images."
Similarity Graph
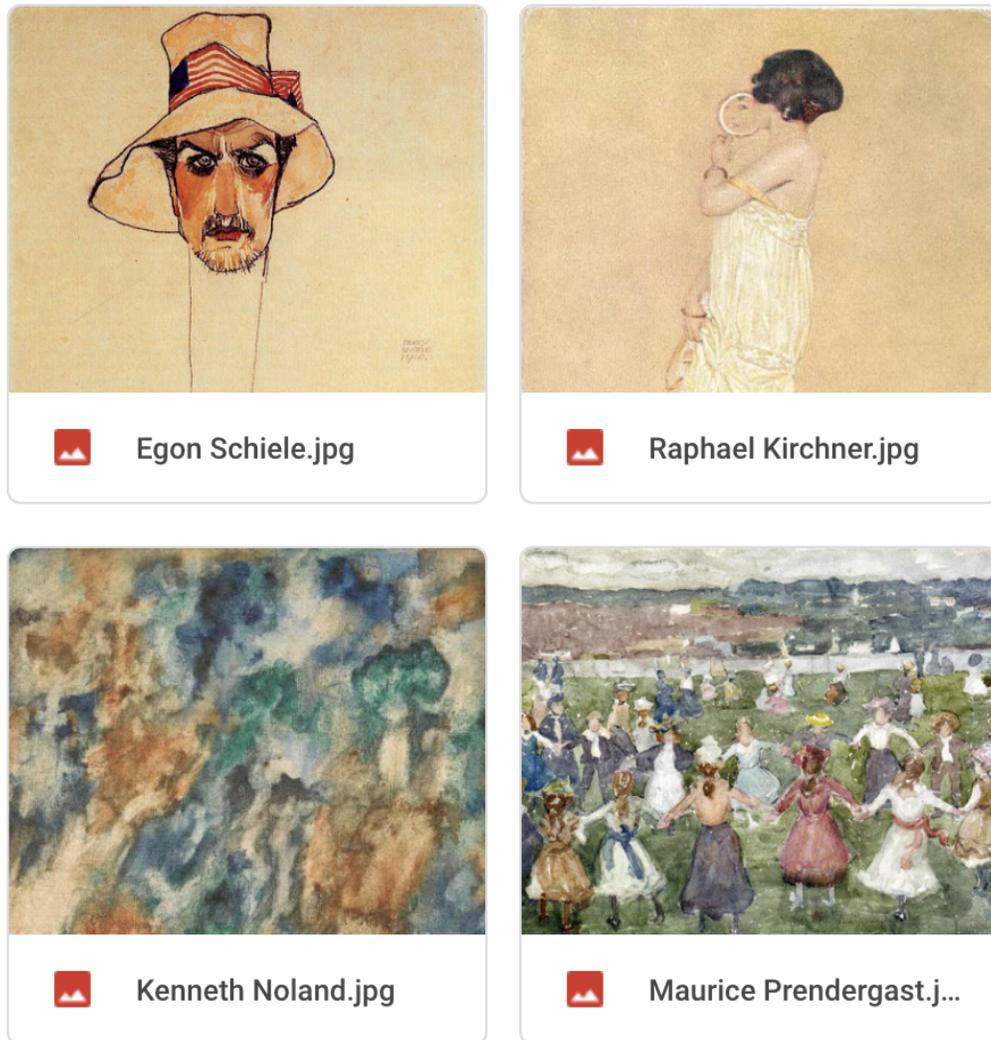
Figure 11: "The top image consists of two paintings that are most similar between the two of the artists used in our experiments. The bottom image consists of two paintings that are most dissimilar between the two artists used in our experiments." Extremes in Painting Similarity

CHAPTER 2: WIKIART AND THE CATALOGUE RAISONNÉ

2.1   The Catalogue Raisonné

The catalogue raisonné compiled by art scholars holds information about an artist's work such as a painting's image, medium, provenance, and title. The catalogue raisonné as a tangible asset suffers from the challenges of art authentication and impermanence. As the catalogue raisonné is born digital, the impermanence challenge abates, but the authentication challenge persists. With the popularity of artificial intelligence and its deep learning architectures of computer vision, we propose to address the authentication challenge by creating a new artefact for the digital catalogue raisonné: a digital classification model. This digital classification model will help art scholars with new artwork claims via a tool that authenticates a proposed artwork with an artist. We create this tool by training a machine learning model with 90 artists having at least 150 artworks and achieve an accuracy of 72.96%. We use the ResNet Convolutional Neural Network to improve accuracy and number of artists classes over state-of-the-art artist classification experiments using the WikiArt database. We address inconsistencies in the way scholars approach artist classification by providing a consistent method to recreate our dataset and providing a consistent method to calculate performance metrics based on imbalanced data.

In 1751, Edme François Gersaint created the first catalogue raisonné for Rembrandt.

This creation signifies the beginnings of a process to improve genuine art commerce and protect the amateur art collector. Fig. 12 shows an image of Rembrandt's catalogue raisonné [32]. Since the late 18th century, the catalogue raisonné has served as a complete record of an artist's work. Appraisers, artists, auction houses, collectors, curators, scholars, and students use the catalogue raisonné as a tool in their daily activities. A catalogue raisonné consists of a unique combination of information of each piece of an artist's work. Information can include, for example, image, medium, provenance, and title. Recently, advances in technology such as the digitizing of materials and cloud computing spiked an interest in recompiling and augmenting the dated and nonexistent catalogue raisonné to address the issue of art and impermanence and to add new capabilities[86]. Issues with impermanence and art surface when the medium of art deteriorates and is no longer restorable. For example, art made of organic materials may preserve indefinitely under ideal conditions. However, ideal conditions may not be possible during exhibitions or extreme happenstance such as fire or theft. Faulty painting techniques and materials can create conditions where a piece of art cracks or becomes discolored. Some artists create art such as David Medalla's columns of foam with limited life. These are but a few of many examples of the impermanence of art[13].

Regardless of the medium of the catalogue raisonné, the issue of authenticity is pervasive due to questionable artworks resulting from loss due to theft or negligence. Documentation such as certificate of authenticity, past ownership, artist signature, and other physical attributes such as dimension, medium, and title of artwork represent artifact provenance. Such attributes evolve with progress and are supporting factors

for the account of the artwork authenticity process. Authenticity is important for the account of process because the value of artwork is directly proportional to proper authentication. The opinion of the scholar who compiles a catalogue raisonné forms over time from their extensive research of an artist. In the end, it is paramount to the decision on whether a piece of art makes the cut to be included as an authentic piece of a collection. The market influences scholars with powerful clients who pressure this decision via legal action. These pressures surfaced in such events as the Warhol Authentication Board closing and the Knoedler Gallery forgeries scandal[86]. New technological capabilities enable the existence of the online catalogue raisonné and digital storage of catalogue raisonné artifacts. With the popularity of modeling digital assets with machine learning algorithms in the past ten years, we believe a unique fingerprint or model that characterizes an artist's work is a useful, novel addition to an artist's catalogue raisonné. Such a model could further support the decision to authenticate or not to authenticate a piece of art with a collection. Modeling an artist based on their work is an image classification problem. Recent advances in machine learning and imaging have outperformed humans in tasks of image classification. A key project contributing to these advances is the ImageNet project[26].

The ImageNet project organizes a vast number of online images using an ontology of images built on the WordNet lexical database. The dataset produced from ImageNet lays a state-of-the-art foundation for image classification and training[26]. In 2010, the ImageNet project formed the basis needed to start the ImageNet challenge competition, which includes a variety of classification tasks for 1000 classes. The goal is for teams to compete to create deep neural networks that can outperform expert human

annotators. The baseline human classification error to target is 5.1%[87]. In 2015, the ResNet architecture won the competition with a 3.57% error rate, thus surpassing expert human capability with image classification[39]. While the ImageNet challenge continued through 2016, we focus on the ResNet architecture due to the combination of its simplicity and the minor performance increase of ensuing winners. Table 10 shows ImageNet winners from 2010 to 2017[7].

We propose to make four contributions. First, we propose to increase the classification accuracy of artwork authentication for paintings using more classes than earlier experiments and a deeper ResNet architecture. Second, we propose to use the ResNet architecture to create a model for inclusion in an artist's catalogue raisonné to aid in the artwork authentication problem. Third, we address inconsistencies in the way scholars approach artist classification by providing a consistent method to recreate our dataset. Fourth, we address inconsistencies in the way scholars approach artist classification by providing a consistent method to calculate performance metrics based on imbalanced data.

The academic contribution of this paper is increased classification accuracy and class count using state of the art deep learning techniques for objects that a typical human observer would find difficult to discern. We also provide standard methods for recreating the data source and measuring results from imbalanced data. This research is important to an interdisciplinary audience of art scholars and computer scientists. For art scholars, a born digital model of an artist's artwork is available to help with artwork authentication claims. For computer scientists, the complexities of an algorithm map an objective measure to the abstract nature of art. How this

mapping works and can be improved supplies opportunities for continued research. For art scholars and computer scientists, we support continued research by providing standard methods for database recreation and result measurement.

In the next section, we review works relating to the problem of art identification. This includes an exploration of various art datasets used, as the data itself is critical, and existing methods of artist classification. In the methods section, we discuss the methods we use in our experiments to create a state-of-the-art model to include in a catalogue raisonné. In the results section, we explore our experiments in detail and show that our results outperform the current state-of-the-art models for WikiArt by artist count and accuracy. In the last section, we conclude with a discussion of future research related to this work.

Table 10: Human and ImageNet error rates [7, 87]

Year	ImageNet	Error Rate
2010	Lin et al	28.2
2011	Sanchez & Perronnin	25.8
2012	Krizhevsky et al (AlexNet)	16.4
2013	Zeiler & Fergus	11.7
2014	Simonyan & Zisserman (VGG)	7.3
2014	Szegedy et al (GoogLeNet)	6.7
	Human	**5.1**
2015	He et al (ResNet)	3.6
2016	Shao et al	3.0
2017	Hu et al (SENet)	2.3

2.2 Art Authentication and WikiArt

Our approach to artwork authentication for the catalogue raisonné is to create and associate an artwork model generated from a convolutional neural network (CNN). We

Figure 12: Rembrandt's catalogue raisonné from 1751[32]

generate the artwork model with 90 artists to strengthen the binary class authentication claim for the artist in question[2]. A catalogue raisonné is a comprehensive listing of an artist's known works. In a traditional sense, think of a catalogue raisonné like a book of art found on a coffee table, in a bookcase, or for purchase in a gift store of an art museum. A CNN is a complex computer algorithm inspired by visual biological processes that classify visual input. The output of a CNN is a mathematical model of classification. We ascribe this model as a digital asset with a catalogue raisonné. This model must supply state-of-the-art accuracy and number of classifiers. The rest of this section discusses the historical effort and critical importance of this paper of

compiling a digital art database and identifying artists based on their artwork using machine learning techniques.

2.2.1   Artist Database

This work uses the WikiArt dataset, a public source of data for artists and their artworks, including high-resolution images of art[79]. All artwork from the WikiArt dataset has an associated artist so no anonymous or unknown artworks exists in the dataset. The dataset contains approximately 290 different artwork styles ranging from abstract to surrealism. We discuss related work using the WikiArt dataset as our primary focus. To a lesser extent, we explore related work using the Rijksmuseum dataset, which contains images of cutlery, furniture, maps, newsprint, paintings, sculptures, text, and other pieces of art[70]. We also discuss work sourced from OmniArt, which combines data from WikiArt, Rijksmuseum Museum, and other sources[101], and anime image datasets as a way of comparison of methodology and experimentation.

OmniArt combines data from WikiArt, Rijksmuseum Museum, and other sources. While this dataset is one of the most comprehensive datasets reviewed, the related experiments performed thus far fall short. For example, a seven artists classifier with 70.9% accuracy using a CNN similar to VGG, the 2014 ImageNet winner[101]. Likewise, experiments involving anime images yield a 93% classification rate for only five artists using the ResNet50 CNN[58].

2.2.2 Artist Classification

According to C. Johnson et al.[45], the availability of high-resolution images prompted more research utilizing van Gogh paintings in 2008, thus forming the art authentication problem's foundations. This study of 101 paintings revealed that classification through machine learning is possible using the fluency, geometry, style, and texture of a painting. Of these 101 paintings, 82 are well-known van Gogh, 13 are questionable van Gogh according to experts, and six are not van Gogh. Comparing all paintings' textures using a Gabor wavelet decomposition and support vector machine (SVM) classification, four of the six non-van Gogh classified as van Gogh. Moreover, two van Gogh paintings were classified as non-van Gogh. Art experts consider this analysis of texture to detect enough dissimilarity in brushstrokes to support authenticity assessment[45]. This binary experiment is for van Gogh and a group of six artists, and the accuracy of classification is 94%.

Soon after the van Gogh experiments by Johnson et al.[45] and the WikiArt dataset creation by Pirrone et al.[79], related research continued for multiple artists in 2010 and 2011. Blessing & Wen ran experiments on seven artists and achieved 85.13% using histogram of oriented gradients (HOG) for feature extraction and SVM for classification. The data for this experiment sources from Google image search[9]. We consider this source of data closely tied to WikiArt because all artists were publicly available through WikiArt at the time of this experiment. Moreover, all these artists are part of the experiments conducted. Influenced by Blessing & Wen's work, Jou & Agrawal conducted similar experiments using histogram of oriented gradients (HOG)

for feature extraction and Naïve Bayes for classification. This approach leads to a reduced accuracy of 65% with less artists. It is important to note that the data for this experiment sources from specific websites for each artist, and two artists are not part of the Blessing & Wen experiments[52]. We consider this source of data closely tied to WikiArt because all artists were publicly available through WikiArt at the time of this experiment. Moreover, all these artists except one are part of the experiments conducted.

The number of artists in experiments using data from WikiArt greatly increases with the use of convolutional neural networks (CNN) after the ImageNet challenge starts in 2015 [87]. In 2017, Viswanathan produced the most notable of these experiments using the ResNet 18 algorithm with transfer learning to achieve 77.7% accuracy with 57 artists. For this experiment, the artists have at least 300 paintings each[110]. While this experiment does not supply an exact list of artists, the 300-painting threshold places these artists in a subset of the artists used in this experiment. Moreover, the method used in Viswanathan's experiment is closely related to the propsed experiment. We mention two related experiments using WikiArt and CNNs. While the results we are interested in pale in comparison to Viswanathan's results, they are important to mention to show the varied research in the area. First, using a variation of Viswanathan's CNN design, a 15 artists classifier with 74.7% accuracy. It is important to note that the experiments in Viswanathan's paper are geared toward a comparison between using CNNs and SVMs and the setup involved for both[18]. Cetinic et al.[17] develop an experiment using 23 artists and CaffeNet, a CNN derived from the 2012 winner of the ImageNet challenge called AlexNet. This method achieves

a 79.1% accuracy, and the team explores more classification experiments of genre, style, timeframe, and nationality[17]. The last two experiments explicitly list the artists of which all exist within the domain of artists used.

Similar experiments using data from the Rijksmuseum Museum produce promising results. In 2013 the Rijksmuseum Museum started a series of challenges to name the artist, type, material, and creation year of their art using computer science techniques. In 2014, the first experiment used SVM to classify 100 artists with 76.3% accuracy using a 96-dimension Fisher vector based on scale-invariant feature transform (SIFT). It's important to note that the algorithm uses the top 100 performing artists from an initial pool of 374 artists and an initial classification accuracy of 59.1%[70]. In 2015, Van Noord et al.[106] extends this work with a focus on paintings. Using PigeoNet, a CNN derived from CaffeNet and AlexNet, the top 78 artists in the dataset that are the least likely to be confused are classified with 73.3% accuracy[106]. In 2017, OmniArt developed a multi-task deep learning method that, when applied to the Rijksmuseum Museum challenge, produced 81.9% accuracy for the top 52 artists in the dataset that are the least likely to be confused[99].

Experiments with data sourced from OmniArt and anime produce results with good accuracy but few classes. Using OmniArt, a seven artists classifier with 70.9% accuracy using a CNN similar to VGG, the 2014 ImageNet winner[101]. Performance and number of classifiers are improved using the OmniArt multi-task deep learning method. Experiments yield 80.8% accuracy for 87 artists[99]. Likewise, experiments involving anime images yield a 93% classification rate for only five artists using the ResNet50 CNN[58].

2.2.2.1   Summary

This research aims to improve on existing work that uses a subset of the WikiArt data in our experiment. Given the related work, we take on the task of producing an experiment that will improve upon Viswanathan's work. This will create a model for inclusion in an artist's catalogue raisonné to aid in the artwork authentication problem.

2.2.3   Classifying Artists in WikiArt

Our goal is to build a system that inputs a single image of a painting and outputs an artist label. Our system must be able to handle red, green, and blue additive color model (RGB) images. Our target is to classify twice as many artists with 250+ paintings with comparable accuracy to Viswanathan's experiment, which reports a 77.7% accuracy with 57 artists having 300+ paintings. We do not plan to cherry-pick artists based on their performance to maximize accuracy because we aim to generically show the style of an artist with a random sample of artists with base proliferation. The goal is to maximize the number of artists and accuracy because both metrics strengthen the model to add to the catalogue raisonné. We carry out this goal by using a state-of-the-art CNN architecture. We show a pictorial of our method in Fig. 13. In this figure, an artist's paintings feed a CNN to create a model. An art scholar attaches the model to an artist's catalogue raisonné and uses it as a tool to aid future claims for adding new art to the catalogue raisonné.

Specifically, we implement ResNet 101 CNN with ImageNet transfer learning. ResNet 101 is the 2015 ImageNet winner and grants ease of implementation and solid

performance on detection, localization, and segmentation aspects of the challenge. We decide to bypass the implementation of the 2016 and 2017 ImageNet winners due to the increased implementation complexity, which would theoretically only allow a gain of $.6 - 1.3\%$ [7, 39].

The ResNet CNN introduces the concept of residual learning. Residual learning addresses the accuracy degradation problem that arises as the depth of CNNs increase. Research found that accuracy can diminish after making a change that should logically produce better results. On the one hand, extra layers increase the performance of the CNN. On the other hand, research shows that blindly adding layers diminishes accuracy because enduring discoveries fade due to a vanishing gradient. Residual learning addresses this problem by ensuring these discoveries persist as layers of the network are added[39].

By way of comparison, consider the activities associated with the classic shape sorting child's toy. In this activity, a child receives a variety of colored, three-dimensional, wooden shapes. The goal is to fit these shapes into a wooden box via a two-dimensional opening. There are a variety of things to consider when fitting each shape into the corresponding box opening. For example, objective considerations like shape type, shape size, shape orientation, shape velocity, shape acceleration, hole type, hole size, and hole orientation determine a fitting outcome. Other subjective considerations like color or pattern matching may exist for an added challenge. If we use a robot to perform this activity, we can map these considerations to separate learning layers of a CNN. Obviously, scenarios exist where we do not want to lose residual accomplishments as learning progresses, and a model begins to form. For example, we don't want to

lose key residual learning with respect to what is known about placing a cube into a square hole when learning the subjective measure of color as a blue cube is placed into a square blue hole rather than a square red hole.

How does this shape sorting activity relate to classifying art? Like the shapes in the sorting activity, a painting consists of color and shape or texture. Research shows that an artist's style alone contributes a significant amount to art classification. For example, through feature learning of a CNN versus feature engineering and clustering, artist classification for single and dual authorship show that a distinctive visual texture is present even in areas that appear empty to the human eye[106]. The notion that more CNN layers increase performance supports the mapping needed for the multitude of layers necessary to represent the vast number of ways to approach the style of a painting. Therefore, the concept of deeper CNNs and therefore deeper residual learning is necessary to yield greater CNN performance for art authentication.

2.3   Experiments on Wikiart

We benchmark our ResNet 101 implementation with a previously published ResNet 18 implementation[110]. Both ResNet implementations use the MATLAB deep learning toolbox[56] and use the same data from WikiArt, which uses artists with 250 or more paintings[79]. We compare precision, recall, F1 score, accuracy, and mean class accuracy (MCA) overall and at the class level for both our experiment and the baseline to evaluate the performance.

2.3.1   Data

We acquired data for our experiments from WikiArt using a custom download tool and the WikiArt API. We query all artists and download an artist's artworks if they have 250 or more paintings. We only download RGB formatted images. In some cases, we retrieve less than 250 artworks due to invalid formats. Overall, we downloaded 45,974 paintings for 90 artists. The most paintings downloaded are for Vincent Van Gogh, with a total of 1,931. The fewest paintings downloaded are for George Grosz, with a total of 158. A select and full distribution of artists is shown in Table 11 and Table 26, respectively[79]. We share this work on GitHub to recreate our WikiArt data source and verify the artwork used in our experiments.

One challenge with this dataset is the class imbalance. The ImageNet dataset does not declare class balance as a prevailing property, but its designers mention the importance of class balance when comparing their dataset to related datasets[26]. For the ImageNet challenge, the focus is on the accuracy of classification and object detection. There are no class balance measures, which leaves the responsibility of handling class imbalance to competitors[87]. Moreover, the topic of balancing input for CNNs remains an active area of research since larger numbers of observations are encouraged for each class for performance[50]. We can address the class imbalance through input data modification or out measure calculation. From an input perspective, research shows that oversampling handles class imbalance optimally with respect to multi-class true positive rate (TPR) and false positive rate (FPR)[12]. From a measurement perspective, research shows that macro balanced accuracy based on

true positive rate (TPR) and false positive rate (FPR) is a good predictor when there is a concern for under-represented classes[37]. For this research, we choose to handle class imbalance using the macro balanced accuracy measurement. We choose this approach to learn as much as possible from each artist and for simplicity of implementation. Moreover, we found no research showing oversampling outperforms macro balanced accuracy for the CNN multi-class imbalance problem. We share this work on GitHub to recreate result measures for our experiments.

A common technique to maximize experiment results is to select the top n true positive artists from a larger class experiment. These top-performing artists feed later experiments, which boosts accuracy metrics[106]. For this experiment, we refrain from this tactic and use all artists selected for the experiment regardless of performance. We do this to explore the opportunities presented from the analysis of weaker performing artists.

Table 11: Select Artist artwork distribution along with the training, validation, and test counts used in experiments

Artist	Artwork Count	Training Count	Validation Count	Test Count
Claude Monet	1,366	956	205	205
Francisco Goya	284	199	43	42
Henri Matisse	999	699	150	150
Pablo Picasso	1,139	797	171	171
Rembrandt	765	536	115	114
Salvador Dali	1,164	815	175	174
Vincent Van Gogh	1,931	1,352	290	289

2.3.2   Training Details

Training details are identical for the baseline and proposed experiment. We use default hyperparameter values from MATLAB for the first experiment. These default hyperparameter values end up producing solid results. The only default that we change is the data split between training, validation, and test. The default splits the data set into 70% training and 30% validation. We change this to 70% training to allow for test data, and the rest splits into 15% validation and 15% testing. Training data creates a model by learning from the data. Validation data checks for accuracy during training. Test data tests model accuracy once validation accuracy is acceptable. The full distribution of artists, training, validation, and test split counts are shown in Table 11[79].

Input painting images are resized to match the network's input size, which is 224 x 224 x 3. We randomly rotate paintings between -90 degrees and 90 degrees, randomly scale paintings between one to two times the original size, and randomly reflect paintings on the x-axis. The solver used is stochastic gradient descent with momentum (SGDM) with a learning rate of 0.01 and a momentum of 0.9. Training passes through the data set 30 times (30 epochs), with validation occurring after 50 iterations. The epoch count of 30 is the default of MATLAB and gives ample iterations for validation accuracy saturation. If experimentation shows a monotonic increase of accuracy with each epoch, repeating the experiment with a higher epoch count is necessary. With each iteration, a mini batch size of 128 images processes through the CNN. The mini batch corresponds to the subset of the training data

that evaluates the gradient of the loss function and updates the weights through backpropagation. After each epoch, training data shuffles paintings to handle the situation where the mini batch size does not equally partition the data. To reduce overfitting, a weight decay regularization term with a value of .0001 adds to the loss function.

To give an example of how paintings train and cross-validate, it is helpful to review the processing of an epoch. Given that we have 45,974 paintings, we use 70% of this data or 32,181 paintings for training. Given that we process paintings in batches of size 128, the training process cross-validates every 250 iterations. Note that 250 iterations multiplied by a batch size of 128 is 32,000 paintings. However, there are 32,181 paintings for training. To account for the remaining 181, we shuffle paintings after each epoch. We continue this process for 30 iterations. We visualize this entire process, displaying the accuracies and losses over the iterations, in Fig. 15, Fig. 16, Fig. 18, and Fig. 19.

The execution environment is set to parallel, which takes advantage of multiple CPU cores and GPUs. The environment is set to process on one node in a high-performance cluster (HPC) using four cores, each of which has two GPUs. The specific hardware for this node is dual 8-Core Intel Xeon Silver 4215R CPU @ 3.20GHz (16 cores total) with 192GB RAM (12GB / core) and 8 x Titan V GPUs (12GB HBM2 RAM per GPU).

2.3.3 Baseline Experiment

The baseline experiment uses a ResNet 18 CNN architecture. This architecture has 71 layers and 78 connections. We show a visual of the layers and connections with a focus on convolutions of the architecture in Fig. 14. Note that we group similar convolutions by color and scale up in the number of convolutions performed with respect to the depth in the stack. We display residual convolutions with a dashed box and transition convolutions with a dotted box. The convolutions in Fig. 14 couple batch normalization and ReLU activation function steps, both of which remain hidden to conserve space. It took six hours and 5 minutes to train the model. The training and validation accuracy and loss are in Fig. 15 and Fig. 16, respectively. The aim of training is to maximize accuracy and minimize loss. The accuracy represents how well predictions are made, and the loss represents the errors in prediction. The blue curve represents training accuracy in Fig. 15 and training loss in Fig. 5, and the red curve represents validation accuracy in Fig. 15 and validation loss in Fig. 16. The training accuracy is a result of the specific iteration while the validation accuracy takes all iterations into account. We report on the validation numbers. We perform this experiment to compare with Viswanathan's experiment, which uses a ResNet 18 CNN architecture on 57 artists, and our proposed experiment, which uses a ResNet 101 CNN architecture on 90 artists.

2.3.4 Proposed Experiment

The proposed experiment uses a ResNet 101 CNN architecture. This architecture has 347 layers and 379 connections. From a network layer perspective, the ResNet

101 architecture has 276 more layers than ResNet 18. A visual of the layers and connections with a focus on convolutions of the architecture are in Fig. 17. Note, we do not repeat the details on the architecture because they are the same as the ResNet 18 CNN architecture mentioned above. Other than the number of convolutions, the major difference between the ResNet 18 and ResNet 101 CNN architecture is the grouping of multiple convolutions and the combination of residual and transition convolutions, which we show with a dashed and dotted box. It took seven hours and 46 minutes to train the model. The training and validation accuracy and loss are in Fig. 18 and Fig. 19, respectively. The aim of training is to maximize accuracy and minimize loss. The accuracy represents how well predictions are made, and the loss represents the errors in prediction. The blue curve represents training accuracy in Fig. 18 and training loss in Fig. 19, and the red curve represents validation accuracy in Fig. 18 and validation loss in Fig. 19. The training accuracy is a result of the specific iteration while the validation accuracy takes all iterations into account. We report on the validation numbers. We perform this experiment to show both improvement in accuracy and artist count with respect to Viswanathan's experiment.

2.3.5   Results

Tests using the baseline and proposed experiment models produce the two confusion matrices shown in Fig. 20 and Fig. 21. Both matrices have total-normalized artwork counts to account for the fact that some artists have more artwork than others (i.e., the data's imbalanced nature). The saturation of blue on the diagonal stands for the number of a true positive predictions. The saturation of red outside of the diagonal

stands for the number of a false negative predictions on the horizontal axis and false positive predictions on the vertical axis. These confusion matrices supply a high-level visual that supports the fact that our results produce more true positive results versus false negative and positive results. From the raw values of these confusion matrices, we calculate measures for all the baseline and proposed experiments listed in Table 12. We set the alpha or significance level to a typical value of .05 stating that we would like to be 95% confident that our analysis is correct. Given the macro balanced accuracy of the 90 artists using ResNet 18 and ResNet 101, we arrive at a p-value of 0.01657353173. Since our observed p-value is lower than alpha, we conclude that our results are statistically significant. By way of comparison, the unbalanced accuracy of the 90 artists using ResNet 18 and ResNet 101 provides a p-value of 0.005762296603. This p-value is lower than alpha and is statistically significant. Using balanced data calculations provides a similar p-value for our experiments. We compare these measures calculated from the confusion matrices with Viswanathan's experiment in the analysis section.

We calculate measures for multi-class classification based on a generalization of binary measures from a confusion matrix generated from the test data set and training model. Macro measures are an average of the class measures. Micro measures are a sum of the class measures before measure calculation. We add measures for error rate and the macro and micro versions of precision, recall, F1 score, and accuracy [92]. Furthermore, we add Grandini's macro and micro versions of the balanced accuracy measure to address class imbalance[37]. All future measure references will be at the micro level. We leave the macro calculations for reference. We list all the formulas

used in the next section.

2.3.6   Result Formulas

The following formulas are used to calculate result metrics.

fp = false positives

fn = false negatives

tp = true positives

tn = true negatives

μ = micro calculation

M = macro calculation

$$\text{Error Rate}_\mu = \frac{\sum_{i=1}^{l} fp_i + fn_i}{\sum_{i=1}^{l} tp_i + fn_i + fp_i + tn_i}$$

$$\text{Error Rate}_M = \frac{\sum_{i=1}^{l} \frac{fp_i + fn_i}{tp_i + fn_i + fp_i + tn_i}}{l}$$

$$\text{Accuracy}_\mu = \frac{\sum_{i=1}^{l} tp_i + tn_i}{\sum_{i=1}^{l} tp_i + fn_i + fp_i + tn_i}$$

$$\text{Accuracy}_M = \frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i}}{l}$$

$$\text{Balanced Accuracy}_\mu = \frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l} tp_i + fn_i} + \frac{\sum_{i=1}^{l} tn_i}{\sum_{i=1}^{l} tn_i + fp_i}$$

$$\text{Balanced Accuracy}_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i} + \sum_{i=1}^{l} \frac{tn_i}{tn_i + fp_i}}{l}$$

$$\text{Precision}_\mu = \frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l} tp_i + fp_i}$$

$$\text{Precision}_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l}$$

$$\text{Recall}_\mu = \frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l} tp_i + fn_i}$$

$$\text{Recall}_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l}$$

$$\text{F1 Score}_\mu = \frac{2 \times \text{Precision}_\mu \times \text{Recall}_\mu}{\text{Precision}_\mu + \text{Recall}_\mu}$$

$$\text{F1 Score}_M = \frac{2 \times \text{Precision}_M \times \text{Recall}_M}{\text{Precision}_M + \text{Recall}_M}$$

Table 12: Table showing an increase for all measures from the baseline to proposed experiments using test data[92]

Measure	ResNet 18 (Baseline)	ResNet 101 (Proposed)	Improvement
Micro Error Rate	46.51%	40.13%	-15.92%
Macro Error Rate	47.31%	41.17%	-14.93%
Micro Accuracy	53.49%	59.87%	10.67%
Macro Accuracy	52.69%	58.83%	10.45%
Micro Balanced Accuracy	69.70%	74.90%	6.95%
Macro Balanced Accuracy	67.78%	72.96%	7.09%
Micro Precision	69.70%	74.90%	6.94%
Macro Precision	67.78%	72.96%	7.10%
Micro Recall	69.70%	74.90%	6.94%
Macro Recall	69.29%	74.06%	6.44%
Micro F1 Score	69.70%	74.90%	6.94%
Macro F1 Score	68.53%	73.50%	6.76%

2.3.7   Result Discussion

2.3.7.1   Analysis

Our results indicate that there is an 72.96% chance to identify one of the 90 artists given one of the 45,974 paintings in our dataset. The probability of randomly guessing an artist is 1.11%. The best chance to randomly guess an artist is 4.2% for Vincent Van Gogh. There are 290 different styles of art in our dataset, we are confident that our proposed algorithm will produce similar results for a different set of 90 artists with their own style of creative curiosity. The algorithm works because it learns the texture and colours produced from an artist's imagination, brush strokes, and colour selection.

We analyze the results in Table 12. First, we review accuracy. Next, we compare the ResNet 18 baseline versus ResNet 101 proposed experiments. We then show

improvement from Viswanathan's work with our ResNet 18 baseline and ResNet 101 proposed experiments with a focus on performance and class count. Lastly, we look at artists with the best and worst performance with respect to artwork count, image similarity, and mean-squared error.

2.3.7.2   Accuracy

We note that the macro and micro accuracy of the ResNet 18 baseline and ResNet 101 proposed artists are low and the relative balanced accuracies are inline with the final validation accuracy from training. We believe these accuracies are low because we are using unbalanced data, and this further supports the need to use the macro balanced accuracy measures in our analysis. Moving forward in our analysis, we use the term accuracy in place of macro balanced accuracy for brevity.

2.3.7.3   ResNet 18 Baseline vs. ResNet 101 Proposed

With this comparison, we see that all measures improved from our baseline ResNet 18 experiment to our proposed ResNet 101 experiment. This experiment is new for 90 WikiArt classes of artists, and the problem of classifying a painting is much more open-ended than that of the specific images in ImageNet. However, we expected improved results since we increase the depth of the CNN and use residual learning, both of which work together to allow for the performance increase. According to He et al.[39], the increase of 7.09% in accuracy is on par with similar depth increases shown in residual learning research[39].

2.3.7.4   Viswanathan vs. ResNet 18 Baseline

Accuracy decreases by 12.76% from Viswanathan's experiment to the baseline ResNet 18. Precision and recall decrease by 11.48% and 11.05%, respectively. This discrepancy is because the former experiment has 63.33% of the latter experiment's artists' classes. Moreover, the former experiment uses a random sample of balanced data, while the latter experiment uses all samples and is imbalanced. The source of both experiments is WikiArt, and we verify that the 57 artists used in Viswanathan's experiment is a subset of the 90 artists used in our experiment. We are unable to find the specific pieces of art to reproduce Viswanathan's experiment exactly, but the extra learning from the increased classes with the increased accuracy as evidence shows a potential for overall improvement. Moreover, in his research, Viswanathan concludes that a future experiment using the method we implement should yield an increased accuracy.

2.3.7.5   Viswanathan vs. ResNet 101 Proposed

Accuracy decreases by 6.11% from Viswanathan's experiment to the proposed ResNet 110 experiment. Precision and recall decrease by 3.74% and 3.33%, respectively. We explain this discrepancy using the same rationale in section 4.2.2. The only difference we see here is a reduction in decrease as expected. These measure improvements are a direct result of using a deeper CNN with residual learning. We use this analysis of our results as the final basis to satisfy the state-of-the-art method to provide a CNN model to assist with the artwork authentication problem for the catalogue raisonné.

2.3.7.6   Calculating Accuracy Analysis Measures

To rule out the correlation between accuracy and simple engineered features of an artist's artworks, we analyze our accuracy results for each artist by comparing with their artwork count, similarity, and estimator measures. It is important to show no correlation to support the viability of our learned models. For artwork count, we count the number of artworks used in our experiments for each artist. For similarity, we calculate the average structured similarity index (SSIM) between all the combinations of two artworks for an artist. For the estimator, we calculate the average mean-squared error (MSE) between all the combinations of two artworks for an artist. Before analysis, we augment the artwork images to the same as the input size of the experiment CNN network, which is 224 x 224 x 3.

For the similarity and estimator measures, we use the binomial coefficient formula to figure out the number of calculations needed for each artists' artworks taken two at a time. We use the following formula for each artist where n is the number of their paintings and k is 2:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

The sum of these combinations results in 17,018,158 calculations needed for each SSIM and MSE. For this number of calculations, we need to use an HPC. The calculations take 17 hours and 6 minutes to process, and the execution environment is set to process on one node using 12 cores and 128GB of RAM. The specific hardware for this node is Dual 24-Core Intel Xeon Gold 6248R CPU @ 3.00GHz (48 cores / node) 384GB RAM (8GBs / core).

2.3.7.7 Artwork Count vs. Accuracy

The purpose of this analysis is to verify that there is no major impact on artist accuracy based on an artists' number of artworks. Moreover, we want to verify that our minimum number of 158 pieces of artwork for learning is sufficient. We display the result of the artwork count versus accuracy analysis in Fig. 22. To compare artwork counts with the accuracy of each artist in the same pictorial, we normalize artwork counts. We also sort by artwork counts to aid in the visualization between the artwork count and accuracy curves. Due to space restrictions, we do not list all artist names, but we do call out the artists minimum and maximum accuracies with a black dot on both curves. The accuracy moves between the minimum and maximum accuracy values independent from artwork count, thus visually showing no correlation between the two measures. From this analysis, we are confident that there is no major impact on accuracy based on the number of artworks for each artist. Moreover, we are confident that 158 pieces of artwork are sufficient for learning an artist's style.

2.3.7.8 Mean SSIM vs. Accuracy

Li et al.[61] define SSIM as a measure that assesses the visual impact of the luminance, contrast, and structure characteristics of an image[61]. The formula used to calculate SSIM is as follows where $\mu_x, \mu_y, \sigma_x, \sigma_y,$ and σ_{xy} are the local means, standard deviations, and cross-covariance for images x and y. C_1 and C_2 are constants to prevent division by zero:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} [61]$$

An SSIM between two images with an upper bound value of one specifies that

the images are the same. The minimum value of SSIM is zero, showing a maximum difference between two images. Our goal is to obtain an average SSIM value for an artist, given all the possible combinations of an artist's paintings. We aim to show that the similarity of an artist's paintings does not significantly impact artist accuracy. We display the result of SSIM versus accuracy analysis in Fig. 23. We do not need to normalize SSIM for our analysis because the domain of SSIM values is in proportion to accuracy. We sort by average SSIM to aid in the visualization between the average SSIM and accuracy curves. Due to space restrictions, we do not list all artist names. However, we do call out the artists minimum and maximum accuracies with a black dot on both curves. Like artwork count, the accuracy moves between the minimum and maximum accuracy values independent from average artwork SSIM, thus visually showing no correlation between the two measures. There is one exception in that our artist with the highest accuracy correlates to the artist with minimum similarity. However, we note at least five other artists with high accuracy and similarity scores spaced out amongst the whole spectrum of similarity. From this analysis, we are confident that there is no major impact on accuracy based on each artist's similarity of artworks.

<div align="center">2.3.7.9   Mean MSE vs. Accuracy</div>

Pishro-Nik defines MSE as a measure that assesses the quality of an estimator[80]. The formula used to calculate MSE is as follows where x and y are the images to compare and n is the number of pixels to compare:

$\text{MSE}(x, y) = \frac{1}{n} \sum_{(i=1)}^{n} (x_i - y_i)^2 [80]$

An MSE between two images with a value closer to zero is better because it shows an overall smaller difference in the image's pixel values. Our goal is to obtain an average MSE value for an artist, given all the possible combinations of an artist's paintings. We aim to show that the estimator of an artist's paintings does not have a major impact on artist accuracy. We display the result of MSE versus accuracy analysis in Fig. 24. We normalize MSE for our analysis because the domain of MSE values is not in proportion to accuracy, which makes the visual comparison of curves impossible. We sort by average MSE to aid in the visualization between the average MSE and accuracy curves. Due to space restrictions, we do not list all artist names. However, we do call out the artists minimum and maximum accuracies with a black dot on both curves. Like artwork count and average SSIM, the accuracy moves between the minimum and maximum accuracy values independent from average artwork MSE, thus visually showing no correlation between the two measures. From this analysis, we are confident that there is no major impact on accuracy based on an estimator of artworks for each artist.

2.3.7.10   Best Performing Artist

Kenneth Noland has the best classification accuracy measure of 98.52%. We downloaded 271 of his artworks. Our model trains from 190 (70%) of his artworks, and we calculate the accuracy measure from the test data of 40 (15%) of his artworks. Given our analysis, the number of artworks, artwork similarity, and estimator does not affect accuracy in a significant way. Kenneth Noland was an American abstract painter who was one of the best-known color field painters. Kenneth Noland has

many more false negatives than false positives, meaning that these paintings are classified with other artists. Of the false negative artists, none are abstract color field painters[79]. However, several of these artists have many false negatives with the other artists in this research, which leads us to believe that there are either common missed opportunities for learning by the ResNet 101 architecture or intractable situations for learning for these artists.

2.3.7.11   Worst Performing Artist

Alfred Sisley has the worst classification accuracy measure of 72.04%. We downloaded 471 of his artworks. Our model trains from 330 (70%) of his artworks, and we calculate the accuracy measure from the test data of 70 (15%) of his artworks. Like our best performing artists, our analysis does not show that the number of artworks or artwork similarity and estimator impact accuracy in a significant way. According to Pirrone et al.[79], Alfred Sisley was a French impressionist landscape painter who rarely deviated from painting landscapes. Reviewing our experiment confusion matrix for Alfred Sisley, he was predominately confused as false positive and false negative with Camille Pissarro and Claude Monet, who are both French impressionists[79]. Out of the false classifications, Alfred Sisley's false positives are more prominent, which means that paintings by Camille Pissarro and Claude Monet classify incorrectly as Alfred Sisley rather than the other way around. Both false classifications have two to three times as many artworks. However, Pyotr Konchalovsky has a similar artwork count to Camille Pissarro and Pierre Auguste Renoir has a similar artwork count to Claude Monet, and both artists have two false classifications with Alfred Sisley. Therefore, there is

no correlation between the number of artworks for an artist and false classification count.

2.3.7.12   Conclusion

We introduce the idea to include a born digital classification model to the catalogue raisonné to aid art scholars with the artist authentication and impermanence problems. We improve artist classification using WikiArt data with a model that improves on earlier work from an accuracy and number of classes perspective. Specifically, we increase accuracy by 7.09% to 72.96% and the number of classes by 57.89% to 90. We use the ResNet 101 CNN to carry out this increase in performance. We also show that the number, similarity, and estimator characteristics of an artist's artworks do not have a major influence of the accuracy of our trained models. These improvements supply an academic contribution for art scholars and computer scientists to use and extend. Art scholars obtain an object born digital which will bolster the denial or support of claims, and the computer scientist discovers a new application and research opportunity for an algorithm which improves classification accuracy and class count measures for otherwise indiscernible objects. Lastly, we share code artifacts and methods to recreate our data source and result performance measurements.

In future work, we would like to aid art scholars with improved accuracy and number of classes using a deeper CNN and a CNN with augmented layers beneficial to painting classification. In showing how artwork count, similarity, and estimator aspects do not have a major impact on accuracy, we would like to conduct experiments that give a better understanding of the salient features that aid in learning. We also

believe adding style as a decision attribute in addition to our model attribute will increase classification performance. Moreover, we believe that it is possible to increase accuracy by creating a binary classifier for each artist with respect to the remaining group and adding these binary classifiers into a composition for classification. In addition to using the WikiArt collection, we would like to apply this work to Rijksmuseum data and contemporary art collections. Lastly, we would like to work closely with art historians to figure out the best number of artists' classes and classification accuracy for model usefulness in a catalogue raisonné.

This work supports the future trends we see emerging as AI applies to art history collections. Primarily, we see applications to authentication, generative art, style transfer, and born digital artefacts. From an authentication perspective, we believe further analysis of results as the accuracy and class count increases will help explain what aspects of an artist's paintings are most helpful with classification. As the confidence of generated models increases with art historians, we expect these models to be ubiquitous as part of an art scholar's decision, but not as a full replacement. As we glean a better understanding of artist classification, we expect aspects of the generated model as useful with addressing issues when generating new art or transferring the style of artist to an existing piece of art. Lastly, we are optimistic that authentication through AI will foster the catalogue raisonné as a born digital artefact by default.
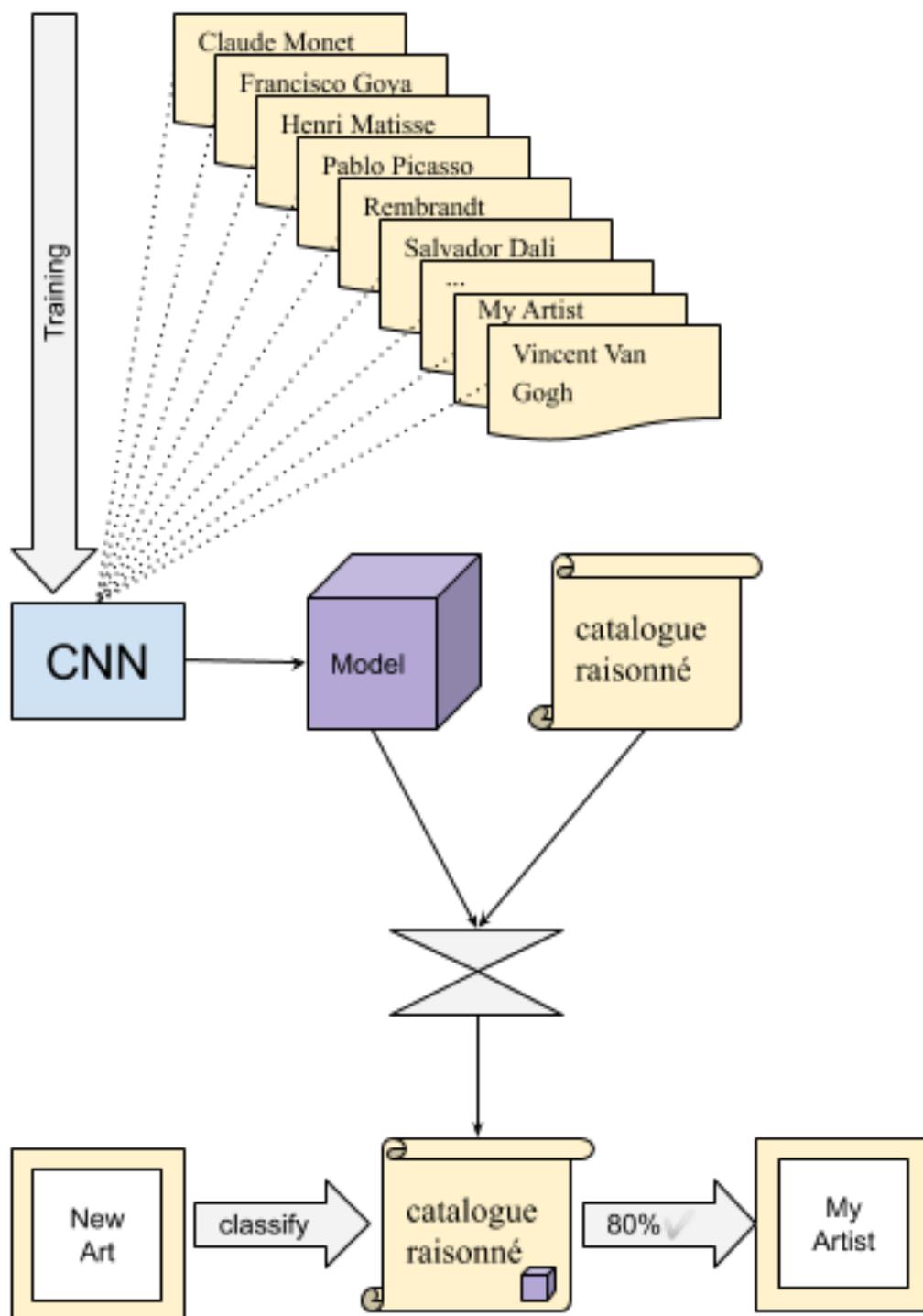
Figure 13: A pictorial of creating an artist's model from a CNN and associating it with a catalogue raisonné)

painting

64 - 7x7 conv

max pooling

64 - 3x3 conv

128 - 1x1 conv

128 - 3x3 conv

256 - 1x1 conv

256 - 3x3 conv

512 - 1x1 conv

512 - 1x1 conv

global average pooling

fully connected

softmax

classification output
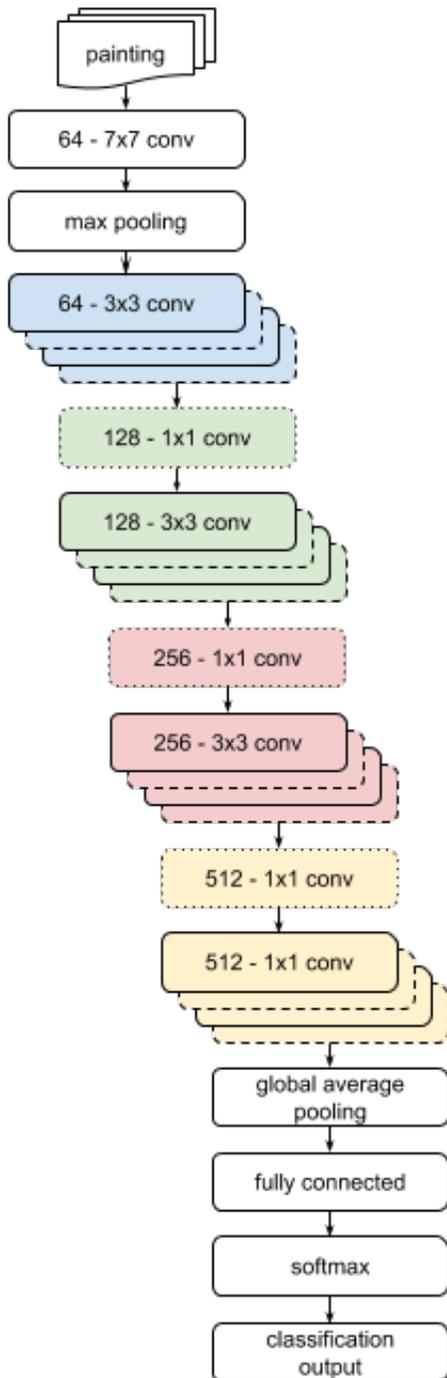
Figure 14: ResNet 18 CNN Architecture with a focus on convolutions
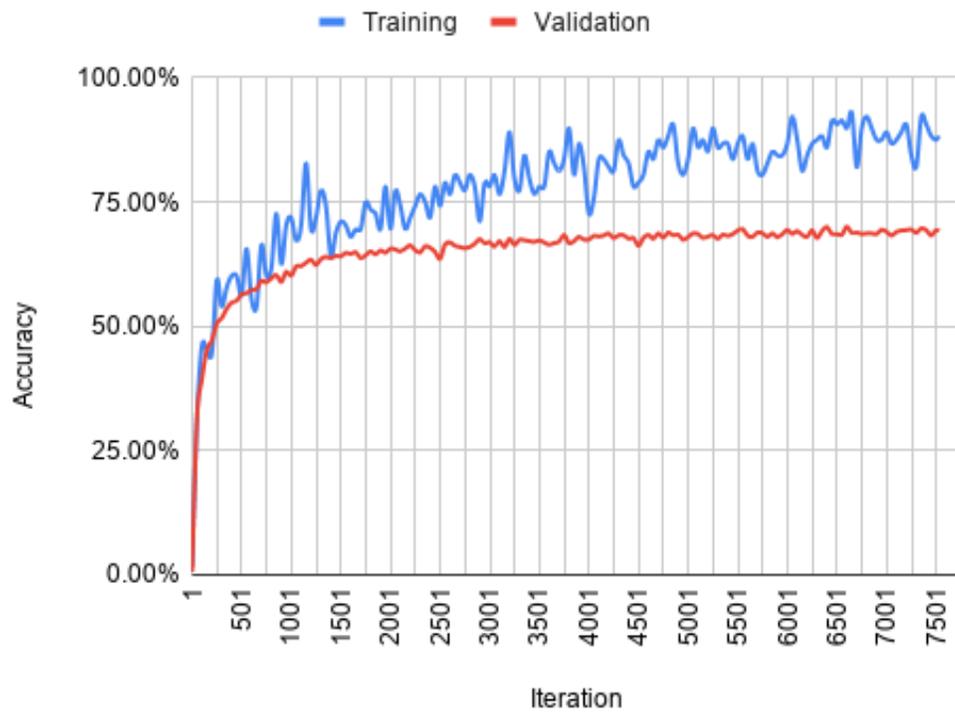
Figure 15: Progress of ResNet 18 model plotting the training and validation curves for accuracy by iteration
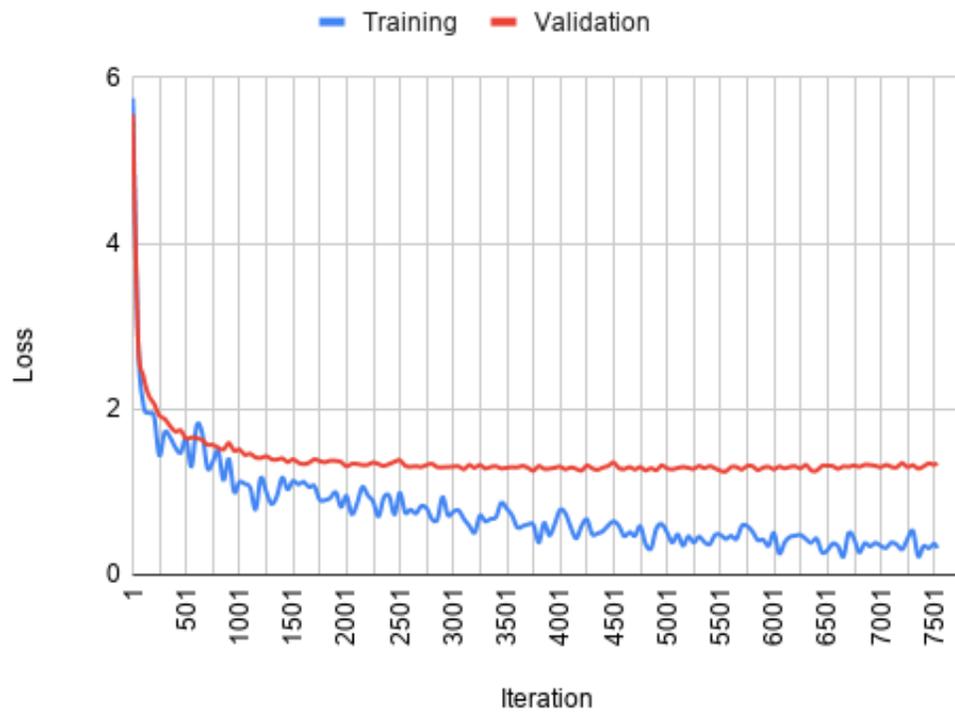
Figure 16: Loss of ResNet 18 model plotting the training and validation curves for loss by iteration
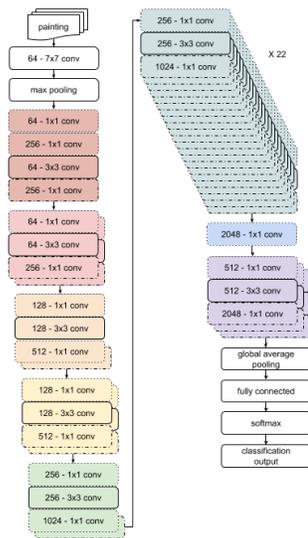
Figure 17: ResNet 101 CNN Architecture with a focus on convolutions
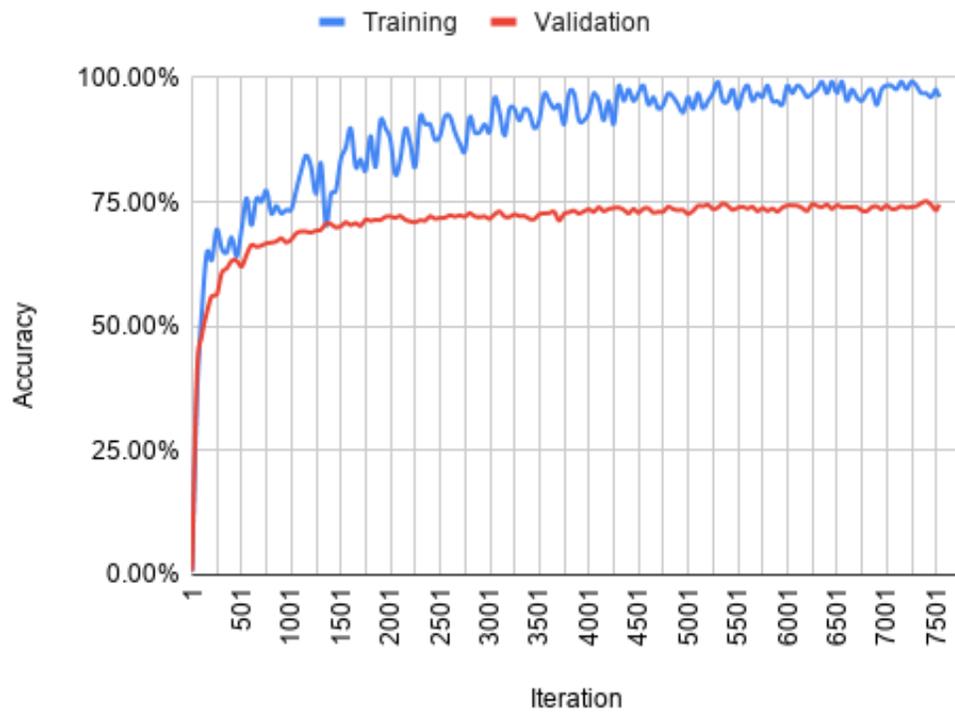
Figure 18: Progress of ResNet 101 model plotting the training and validation curves for accuracy by iteration
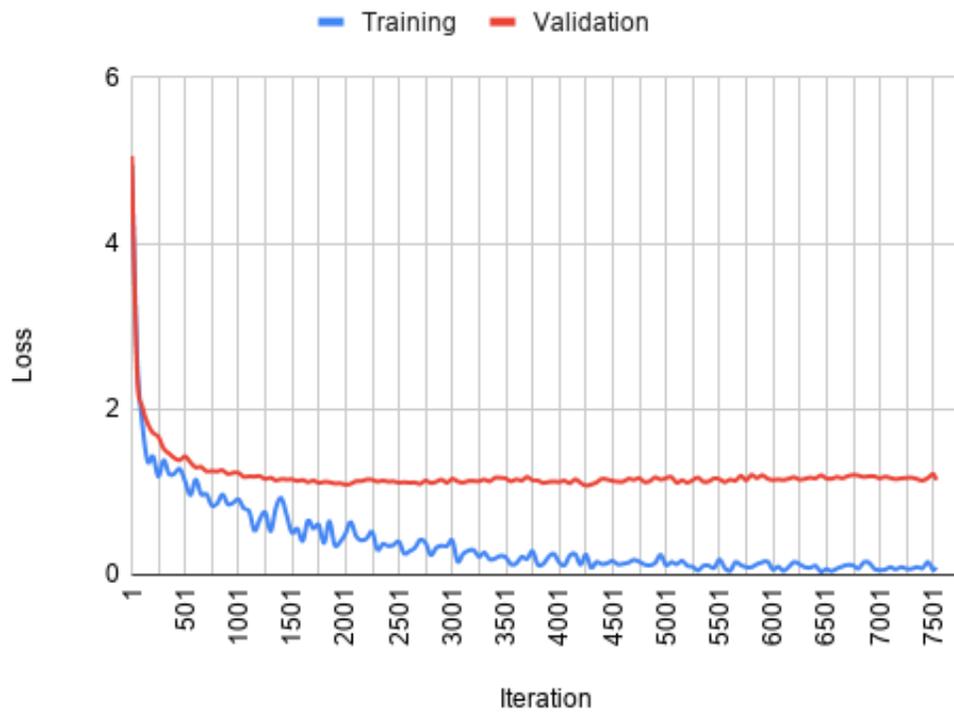
Figure 19: Loss of ResNet 101 model plotting the training and validation curves for loss by iteration
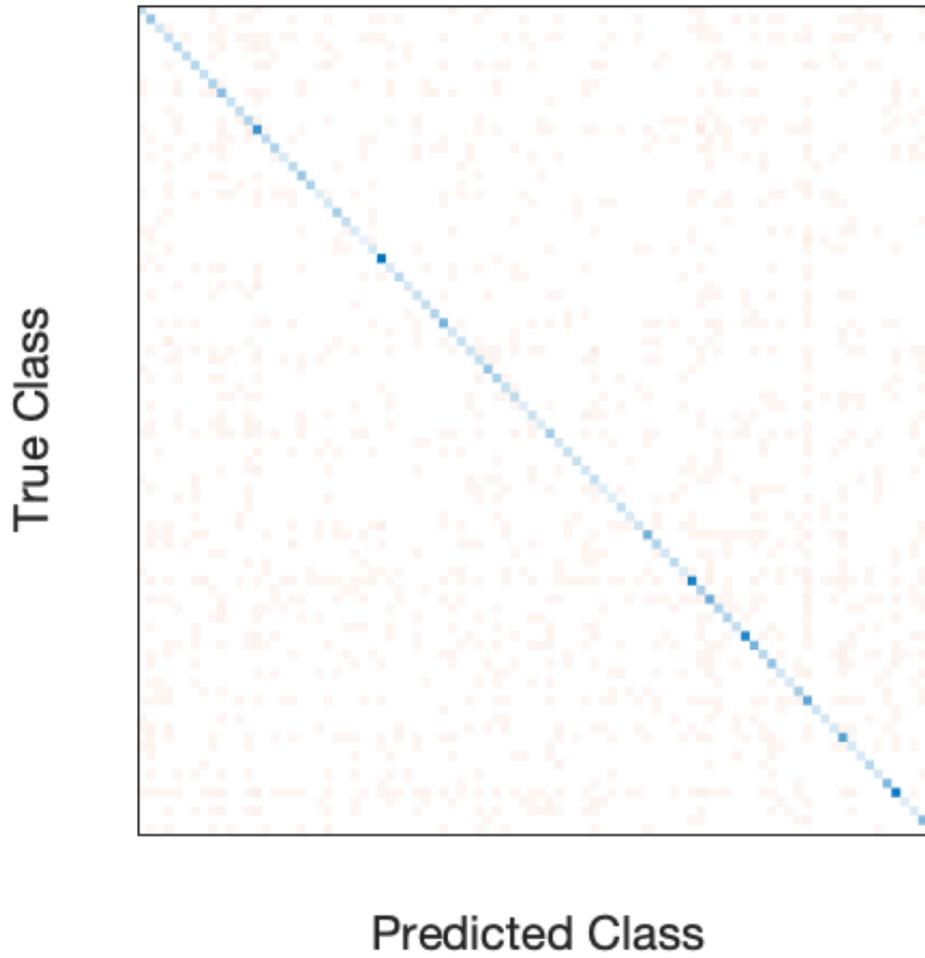
Figure 20: Confusion matrix for baseline experiment showing the blue diagonal of true positive predictions and red points of false negatives and false positives
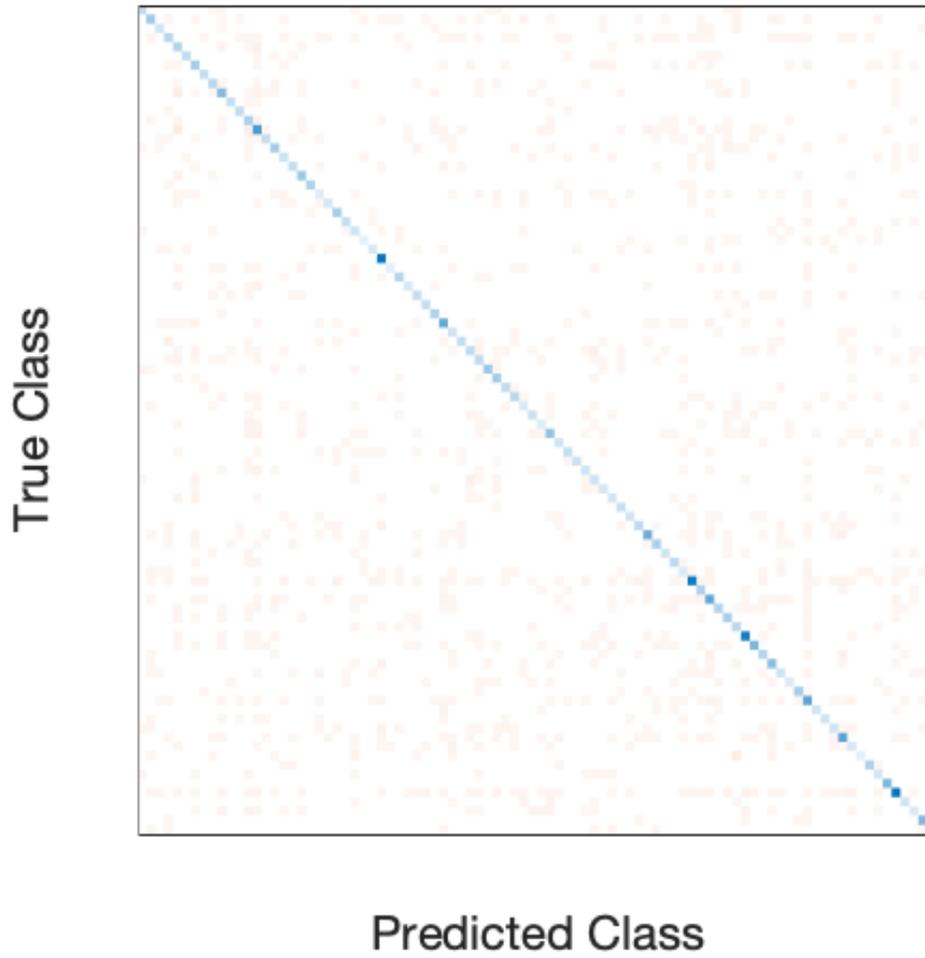
Figure 21: Confusion matrix for proposed experiment showing the blue diagonal of true positive predictions and red points of false negatives and false positives
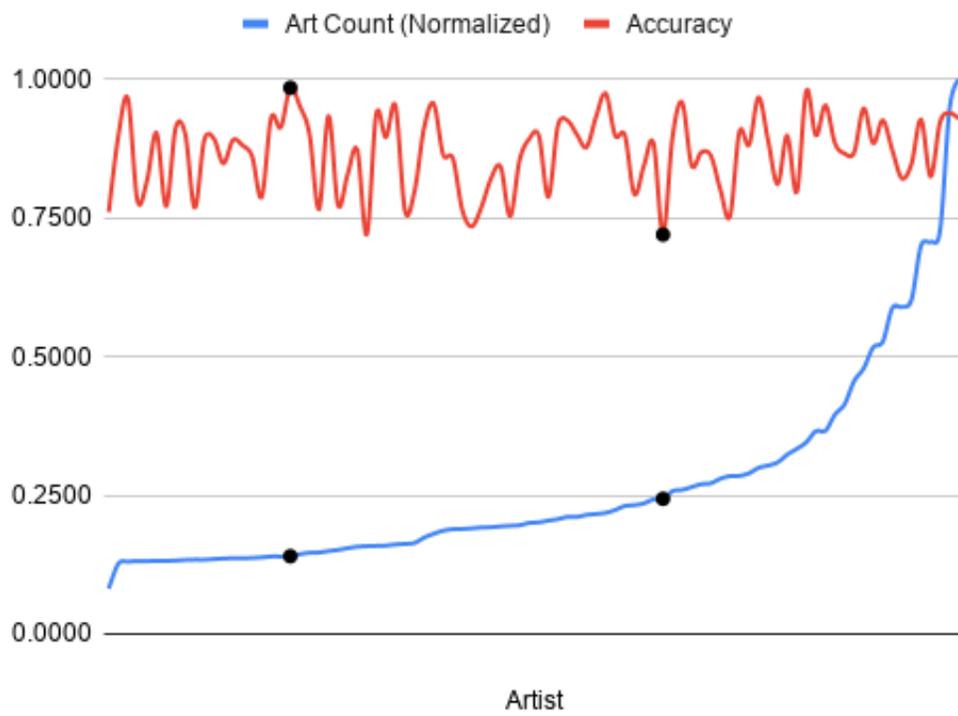
Figure 22: Artwork Count vs. Accuracy
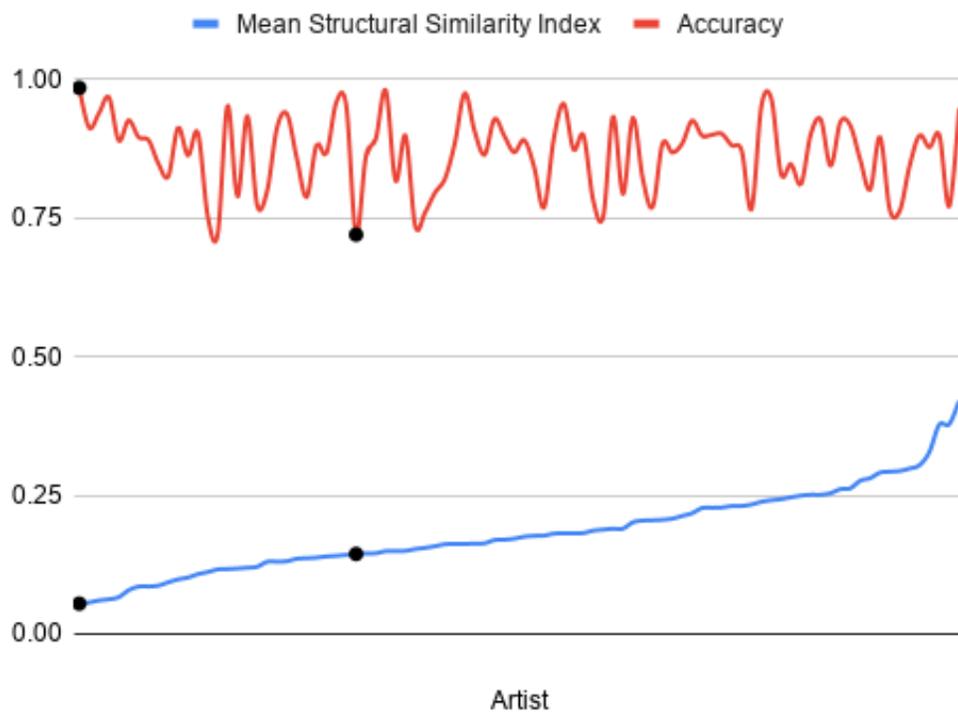
Figure 23: Mean SSIM vs. Accuracy
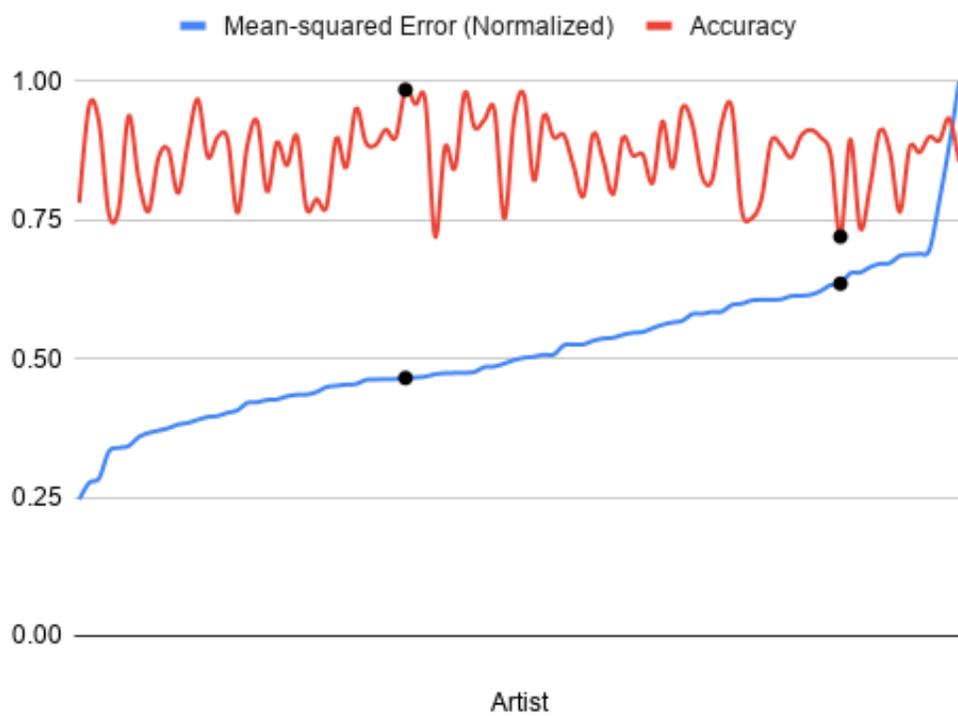
Figure 24: Mean MSE vs. Accuracy

CHAPTER 3: RIJKSMUSEUM AND ANNEALING

3.1   The Rijksmuseum Challenge

The popularity of machine learning algorithms produced numerous applications in the past ten years. One application is that of art authentication which assures that a piece of art is created by an artist. A certificate of authenticity created from proper art authentication significantly increases the value of a piece of art which impacts all parties in an art transaction. The models produced by machine learning algorithms provide an objective measure to authenticate an artist to their artwork collection. In the past ten years numerous machine learning algorithms have been used to address art authentication on a variety of datasets. This work extends art authentication with residual neural networks and the Rijksmuseum data set. Our results show contributions is four key areas: A performance increase of 11.35% over the baseline for 34 artists; A new baseline for 1,199 artists; A standard methods for recreating the Rijksmuseum data set; and A standard method for measuring results from imbalanced data for the Rijksmuseum data set.

When buying or selling a piece of art, it is common to require proof of the artwork's authenticity. Proof of authenticity is normally accomplished through artifact provenance which consist of documentation such as certificate of authenticity, past ownership, artist signature, and other physical attributes such as dimension, medium, and title.

The value of artwork is directly proportional to proper authentication. Therefore, proper art authentication impacts all parties involved with a piece of art such as artist, buyer, seller, curator, appraiser, and insurance adjuster.

Conversely, there are issues with artwork authentication when artifact provenance is fraudulent or missing. For 15 years, Ann Freedman, the president of Knoedler & Company, unknowingly owned $80 million of fraudulent art. Glafira Rosales commissioned fraudulent reproductions of Rothko, Motherwell, and Pollock masterpieces from a local artist and sold them to Freedman. Rosales walked away with $20 million before FBI forensics on the masterpieces revealed historically inconsistent chemicals [77]. The German army stole numerous amounts of art between 1938 and 1945 during their invasion of Europe [41]. Paris and Vienna were areas of interest for the German army due to the lavish collections held by private collectors and galleries in the area [31]. Wissbroecker et al.[114] discuss the litigation attempts of recovering art during this time period. Some of this art that was not destroyed still exists by holders aware and unaware of the art asset. When one of these missing pieces of art surfaces, provenance may be missing.

Blockchain and digital rights management (DRM) are new ways to address art authentication. Wang et al.[112] develop a system that leverages the provenance capability of blockchain to protect a unique identifier assigned to a digital art asset. Zhaofeng et al.[119] develop a digital watermarking algorithm based on discrete cosine transfer (DCT), Arnold transform, human vision system (HVS) model, and Watson model to protect digital assets. Both of these methods address art authentication of digital assets and are easily applied to contemporary art or art with existing

authentication and digital representation. However, these methods cannot be used for physical art that is fraudulent, has missing provenance, or is produced by an artist unwilling to use a supervised technical method for art authentication. The need for an unsupervised method to authenticate digital art assets derived from physical art still exists.

With the popularity of digital image processing and unsupervised machine learning, Johnson et al.[45] provides objective measures for determining Van Gogh's artistic style. This work branched off into numerous research efforts analyzing artistic style thus providing a basis to mitigate the issues of missing provenance with a digital signature of an artist's work. Soon after the van Gogh experiments by Johnson et al.[45] and the WikiArt dataset creation by Pirrone et al.[79], related research continued for multiple artists in 2010 and 2011. Blessing et al.[9] ran experiments on seven artists and achieved 85.13% using histogram of oriented gradients (HOG) for feature extraction and SVM for classification. The data for this experiment sources from Google image search [9]. We consider this source of data closely tied to WikiArt because all artists were publicly available through WikiArt at the time of this experiment. Influenced by Blessing et al.[9] work, Jou et al.[52] conducted similar experiments using histogram of oriented gradients (HOG) for feature extraction and Naïve Bayes for classification. This approach leads to a reduced accuracy of 65% with less artists. It is important to note that the data for this experiment sources from specific websites for each artist, and two artists are not part of the Blessing et al.[9] experiments [52].

The success of Russakovsky et al.[87] winning the ImageNet challenge pushed image

classification to new performance standards. The number of artists in experiments using data from WikiArt greatly increases with the use of convolutional neural networks (CNN) after the ImageNet challenge starts in 2015 [87]. From an artist classification perspective, Viswanathan et al.[110] and Dobbs et al.[27] build on the success of ImageNet winners by applying residual neural networks to increase the performance of artist classification using the WikiArt data set. Likewise, Mensink et al.[70] and Van Noord et al.[106] apply machine learning algorithms to increase the performance of artist classification using the Rijksmuseum data set. Van Noord et al.[106] use a machine learning algorithm called PigeoNet which is a Convolutional Neural Network (CNN) derived from CaffeNet and AlexNet. PigeoNet is the state of the art algorithm for artwork classification on the Rijksmuseum data set. The Rijksmuseum is the national museum of the Netherlands. They tell the story of 800 years of Dutch history, from 1200 to now. In addition, they organize several exhibitions per year from their own collection and with (inter) national loans [72].

Our objective is to make four contributions towards art authentication using the Rijksmuseum data set. First, we use a performance annealing residual neural network (ResNet) to increase artist classification accuracy of artwork from 78.30% to 87.19%, a 11.35% increase, for 34 artists from baseline experiments using the Rijksmuseum data set. The ResNet series of machine learning algorithms performed better in the ImageNet challenge compared to CaffeNet and AlexNet which are the primary algorithms used in PigeoNet baseline. Second, we increase multi class count for ResNet 101 art authentication experiments from 958 to 1,199 and provide a baseline for an artist class count of 1,199 for the Rijksmuseum data set. Third, we provide

standard methods for recreating the Rijksmuseum data set. Fourth, we provide standard methods for measuring results from imbalanced data for the Rijksmuseum data set.

3.2 Art Authentication and the Rijksmuseum

To satisfy our objectives, we need a source of data, a residual neural network method that can be customized, a custom method of annealing, a high performance cluster to run experiments, and a method for measuring the performance of our experiments. These materials and methods are documented and shared on GitHub for reproduction by an independent party. To review the instructions for reproducing our experiments for this research, please see this GitHub link. Finally, we support our methods with a theoretical basis.

3.2.1 Data Source

The data source for our experiments is publicly available for research by the Rijksmuseum and consists of 112,039 artworks from 6,629 artists. Each artwork has a corresponding image and xml metadata file. The high quality images are stored as 300 dpi compressed jpeg and were taken in a controlled environment [70]. Special organization and translation scripts developed in Matlab prepare the data for our core experiments. For our experiments, we use images from all types of artworks for artists with more than ten artworks. Artwork types include but are not limited to images of paintings, prints, photographs, ceramics, furniture, silverware, doll's houses, and miniatures. We include artworks from anonymous and unknown artists in our experiments even though these two categories are not relevant to art authentication.

Since both anonymous and unknown classes contain multiple artists, they provide a group for which an artist should not identify.

3.2.2   Residual Neural Network

Similar to Kim et al.[56], we use Matlab's implementation of Residual Neural Networks [40] to train, validate, and test our models. Specifically, we use a ResNet 101 implementation because research from Dobbs et al.[27] shows ResNet 101 performs better than ResNet 18 for art authentication experiments. Matlab provides an extensible scripting method which facilitates our implementation. We resize images to 224x224x3 to match network input size and use 70% of artwork images for training, 10% for validation, and 20% for test. We use the same training, validation, and test proportions as the baseline experiment. Images also undergo random rotation, scaling, and reflection to prevent overfitting. Training makes a pass through all images in batches of 128 for up to 30 times or epochs and validation occurs after 50 iterations within each epoch. After each epoch, training data shuffles paintings to handle the situation where the mini batch size does not equally partition the data. Once a training model is generated, we make predictions using the model and generate a confusion matrix from the ground truth and output predictions.

We use additional parameters from Matlab. We use transfer learning by utilizing Matlab's pretrained networks which are based on models from ImageNet [26]. These models provide an optimal starting point for our experiments. Without this starting point, we would need to specify initial weights for training such as the weight guidelines established by Cao et al.[14] which propose using Gaussian, Gamma or Uniform

Table 13: Annealing Parameters from Van Noord et al.[106] Baseline Experiment

Artist Count	Artist Loss
958	0
197	761
97	100
34	63

distribution for weight initialization. We also use several recommended hyper parameters. The solver used is stochastic gradient descent with momentum (SGDM) with a learning rate of 0.01 and a momentum of 0.9. To reduce overfitting, we add a weight decay regularization term with a value of .0001 to the loss function.

3.2.3   Annealing

We use Matlab to apply a custom annealing process to harden results based on performance. Annealing takes place after we generate a model with the ResNet 101 algorithm. Class counts from baseline experiments determine how many times annealing takes place. Table 13 and Table 14 show the annealing parameters for our two experiments. In both tables, the first row represents our initial experiment with the artist count of the respective baseline experiment and an artists loss of zero because the annealing process is not applied on the first step. Subsequent rows represent the next experiment and artist count from the respective baseline. The annealing process orders the performance of each artist and drops the worst performing artists such that the artist count is in line with the baseline experiment artist count. The artist loss attribute represents the number of artists dropped this way. The annealing process continues until the last baseline artists experiment completes. Algorithm 1 demonstrates the process.

Algorithm 1 Annealing ResNet Algorithm

$m \leftarrow m_i$ ▷ Initialize model from ImageNet (transfer learning)
$\mathbf{A} \leftarrow \mathbf{A_i}$ ▷ Initialize annealing list
$\mathbf{I} \leftarrow \mathbf{I_i}$ ▷ Initialize image list
while $\mathbf{A} \neq \langle \rangle$ **do**
    $\mathbf{CM} \leftarrow ResNet(m, \mathbf{I})$ ▷ Create confusion matrix
    $\mathbf{I} \leftarrow Anneal(\mathbf{CM}, \mathbf{I}, head(\mathbf{A}))$ ▷ Anneal images
    $\mathbf{A} \leftarrow tail(\mathbf{A})$
end while

Table 14: Annealing Parameters from Mensink et al.[70] Baseline Experiment

Artist Count	Artist Loss
374	0
300	74
200	100
100	100

3.2.4   High Performance Cluster

The time to train models from over 112,039 images for 1,199 artist classes is prohibitive for a personal computer. Non-iterative approaches such as those discussed by Wang et al.[111] may not need HPC resources, but further research is needed before we can select a non-iterative approach over the proven results of ResNet 101. If we used a personal computer, it could take weeks to get results and prohibit us from using the personal computer on a day to day basis. Moreover, any type of interruption such as coding bugs and power outage would interrupt the process, and we would need to start over. Therefore, we use our institution's high performance cluster for our experiments. We target a node with 128 gigabytes (GB) of memory and four graphics processing units (GPUs) for our experiments.

3.2.5   Performance Measurement

Once our experiments are complete, we use Matlab to calculate the performance of our models based on the output confusion matrix of each experiment. We calculate mean class average (MCA) which is the same performance metric used in the baseline [70, 106].

3.2.6   Theory

The theoretical foundations of this work are based off of transfer learning and deep residual neural networks. Traditionally, these theories are applied to practical applications which classify images or objects within an image [26]. Two practical developments come from our theoretical basis. First, higher classification performance amongst many classes will extend support for art authentication in situations of good and bad art provenance [27]. Second, improving performance extends the labeling and querying organizational applications of the Rijksmuseum challenge [70]. We aim to leverage these theories to classify artists given images of their paintings.

3.2.6.1   Transfer Learning

Torrey et al.[104] describes transfer learning as a technique used to transfer knowledge from a source task to improve the learning rate in a target task by allowing the training process to start with higher start, slope, and asymptotic characteristics (Figure 25). We propose to use transfer learning from ImageNet models which are used for classifying whole images and objects within images based on a large lexical database of English called WordNet [71].
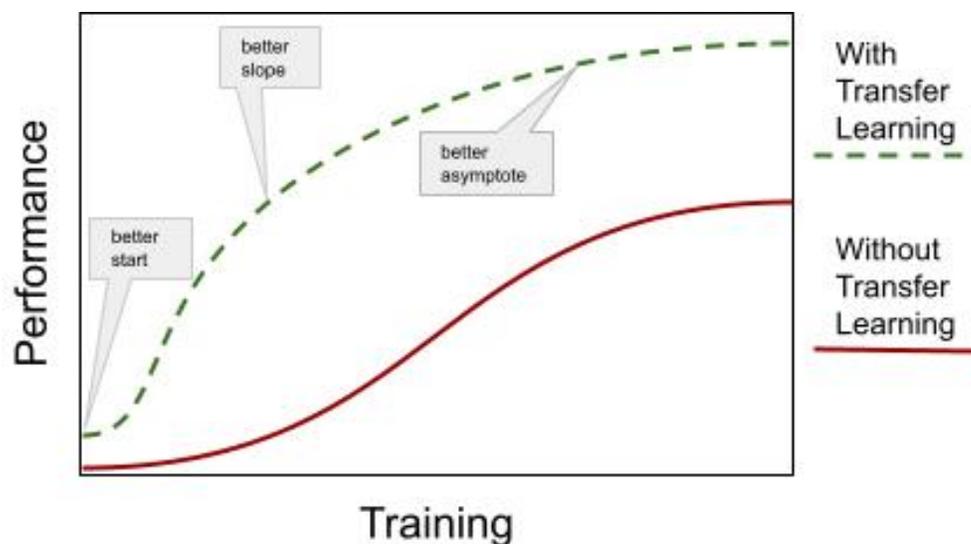
Figure 25: Three ways for transfer learning improvement

3.2.6.2   Deep Residual Neural Network

He et al.[40] solve the exploding and vanishing gradient problem of deep neural networks with a deep residual learning framework which allows for much deeper networks using the concept of skip connections. Philipp et al.[78] provide mathematical proof demonstrating how skip connections can largely circumvent the exploding and vanishing gradient problem. With admissible, deeper neural networks, we expect to extend the performance of previous work. Specifically, we plan to use the ResNet 101 algorithm for our experiments (Figure 26).

3.3   Experiments on Rijksmuseum

Two HPC jobs execute ten experiments that take a combined time of 2.65 days to produce ten models for artists with ten or more artwork images. Our first job run produces seven results for class count 34 to 368 and the second job run picks up three results with larger class counts of 1199, 958, and 374. Each of our ten experiments

Layer Name	Output Size	Layer #	Multiplier	Block
convolution 1	112 x 112	1	1 x	7 x 7 64
convolution 2	56 x 56	2-10	3 x	1 x 1 64 / 3 x 3 64 / 1 x 1 256
convolution 3	28 x 28	11-22	4 x	1 x 1 128 / 3 x 3 128 / 1 x 1 512
convolution 4	14 x 14	23-91	23 x	1 x 1 256 / 3 x 3 256 / 1 x 1 1024
convolution 5	7 x 7	92-100	3 x	1 x 1 512 / 3 x 3 512 / 1 x 1 2048
avg pool, fc, softmax	1 x 1	101	1 x	W
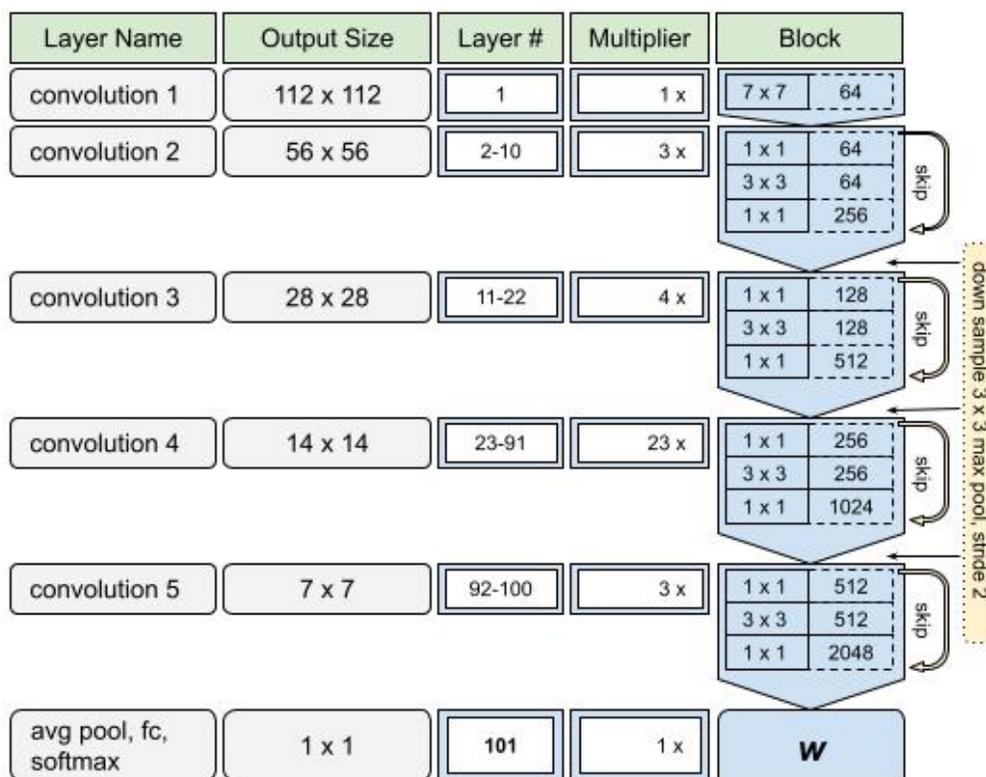
Figure 26: This figure demonstrates the structure of ResNet 101. Each of the three layer building blocks support an identity skip connection which allows for a deeper network that is admissible.

produces four artifact results: confusion matrix, training information, classification labels, and classification model. First, we share our core result measures which we calculate from the confusion matrix artifact. Second, we share training information and explain related detail and visualization. Third, we share classification labels and their importance. Fourth, we share the classification model and its importance. ResNet 101 hyper parameters for these results are shown in table 15.

3.3.1 Confusion Matrix

In this section, we discuss the results calculated from the confusion matrix artifact. We also explain the related MCA calculation in more detail.

We list primary MCA performance measure results for this research in table 16 and table 17. We split results into two tables to align to the baseline experiment results from Mensink et al.[70] and Van Noord et al.[106] respectively. The results table displays class or artists count, balanced baseline performance from the respective research paper, our performance, and the increase from the previous state of the art baseline performance to our performance.

Regarding the MCA result calculations, Dobbs et al. [27] determine the optimal method for calculating art authentication MCA for ResNet 18 and ResNet 101 is using a balanced macro method. Micro calculations (subscript of μ) aggregate measures before the final class measure calculation. Macro calculations (subscript of M) aggregate after each individual class calculation [93]. We use the macro version for its propensity to balance the low and high extreme calculations. Undefined calculations can occur when a class has a true positive and false negative value of zero which causes a division by zero situation. We handle these real situations which cause issues in the macro calculation by adding very small values which prevent zero and do not compromise the calculation. Balanced accuracy (subscript of β) addresses class imbalance as described in Grandini et al.[37]. Our data classes are imbalanced, so we utilized the same technique. Both techniques are combined and represented with equation 1 where l is class count, fp is false positive, fn is false negative, tp is true positive, and tn is true negative.

$$MCA_{\beta}M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i+fn_i} + \sum_{i=1}^{l} \frac{tn_i}{tn_i+fp_i}}{l} \tag{1}$$

We omit results of error rate and the macro and micro versions of precision, recall, and F1 score. The performance of these calculations is not applicable to this research because they are not significant, and the baseline research does not report these measures.

Neither baseline addresses art authentication for all artists with ten or more artworks. We run this experiment to initialize our Van Noord et al.[106] baseline and obtain a MCA of 61.49% for 1,199 artists.

3.3.2   Training Information

Training information consists of training/validation accuracy and loss along with information for base learn rate and aggregate validation accuracy and loss. We found that validation accuracy is not useful to visualize due to undefined data, and aggregate measures are not useful for reporting performance due to the fact that the aggregated data does not consider previous learning. The base learn rate is not useful because it stays constant. We focus on the training accuracy because it provides a good visualization for the time and rate for which each experiment completes. It is important to point out that the training measures do not take data balance into account and over fitting often occurs when generating the model. We account for these issues by using test data to produce final performance on the training models. The results of the test measures consider data imbalance and over fitting.

Table 18 displays the iterations taken for each experiment. Note, the iteration gap between 368 and 374. The gap is a direct result from breaking up or experiments into two jobs. We explain the reason for this in the discussion section and note that the

counts are in proportion with the learning rate of their respective jobs.

In table 19 and table 20 we show the top performers with corresponding artwork count for each classification experiment. Due to space, we omit the top performers with less than 20 artworks in table 20. Note, there are not as many top performers in table 19 because the artwork cutoff is 50. Whereas the artwork cutoff in table 20 is 10.

We show training performance results for each of our ten models in figure 27. As expected, the slope and asymptote of the training curve reduces as the number of classes increase. The time also increases with the number of classes. This is seen best when comparing the 200 class orange curve with the 1199 class green curve. For the most part, each step of the curve is in proportion. The only exception being class count 374. This is a special situation that benefits from the learning of class 1,199 and 958 models which puts it in line with class 100 learning rate. This is likely the cause for the small increase between 374 and 300 in table 17

3.3.3   Classification Labels

We collect classification labels for each experiment to assist with reproducing and extending results in future research. As expected, the number of classification labels reduce with each experiment by an expected amount due to our annealing process. The only exception to this rule occurs between the 374 and 368 class experiments where only 84 classes are in common. We explain the reason for this in our discussion section.
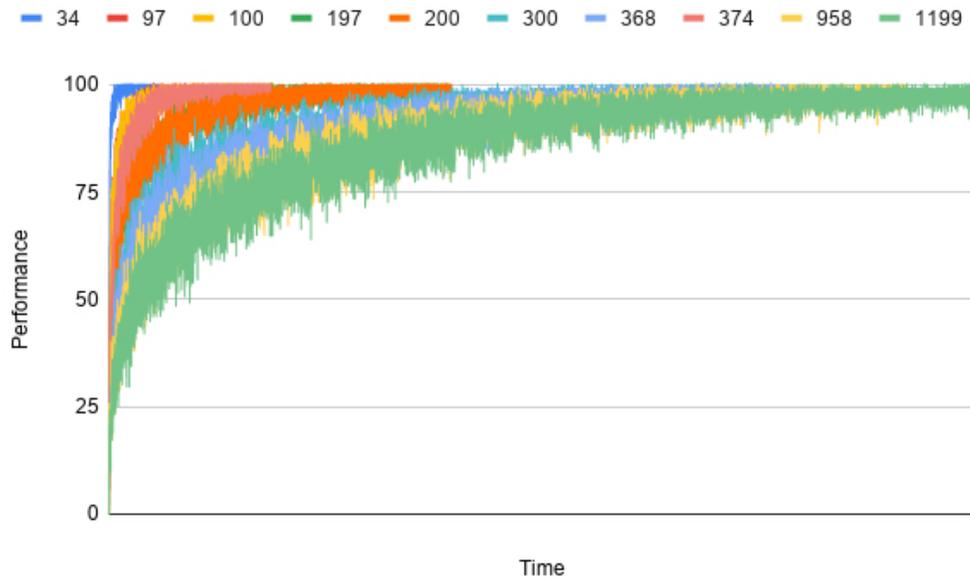
Figure 27: Training performance for all ten experiments which are represented by the numbers at the top of the figure

3.3.4   Classification Model

We collect a classification model for each experiment. Each experiment leverages its model to produce a confusion matrix based on the ground truths and applying the classification model to each test case. These models are not shared because they are too large for the GitHub repository and sharing the models on GitHub defeats the purpose of required result duplication of this research.

3.3.5   Result Discussion

In this section, we discuss our experiments and results in more detail. First, we discuss the significance of this work and contribution. Second, we discuss the results of one or our experiments that facilitate visualization. While result measures are extensible up to the larger 1,199 class experiment, the resultant confusion matrix is

difficult to visualize. Third, we discuss and put to rest potential data congruence concerns with our experiments.

3.3.5.1   Contribution

The significance of this work is self-evident from a performance perspective. Experiments listed in table 16 and table 17 produce performance gains over or close to the baseline. Given the data and metric calculation congruence gap between our experiments and the baseline, the contribution stands due to our data and calculation reproduction potential. While there are no specific baseline comparisons for 1,199 artists, we note that this research of related work did not find any artist classification baselines for any artwork data set that performs better than 32.40% MCA for 1,199 artist classes. These results have a strong correlation to similar ResNet 101 experiments on different art collections. For example, Dobbs et al. [27] report 72.96% MCA on 90 artist classes from WikiArt which is in line with the 72.69% and 72.34% MCA that we report for 97 and 100 artist classes, respectively. Our experiments extend the solution to the art authentication problem by increasing performance and reproducibility. An increase in performance has a direct correlation to the value of art which benefits all

It is noteworthy to point out how the ResNet 101 model outperforms the baseline. Both of the baseline approaches use traditional algorithms to extract features and perform classification. As seen in Figure 26, the ResNet 101 approach uses a deep convolutional neural network which efficiently uses a general feature extraction approach over multiple convolutions of an image to build a model for classification. Efficiency gains are realized by reducing the number of fully connected layers and identity or

skip blocks.

3.3.5.2   Artist Confusion

As expected, performance is inversely related to the number of artists as seen in Figure 28 and Figure 29. This relationship is apparent due to the fact that the probability of an artist's style will be confused with another artist's style naturally increases as the number of artists classified increases. We also know from the research of Dobbs et al.[27] that similar ResNet 101 performance experiments show no correlation to similarity (SSIM), estimator quality (MSE), or artwork count which gives us further confidence of the validity of artist confusion through learning.

We demonstrate artist confusion via the confusion matrix in Figure 30 for 34 artist classes. The saturation of blue on the diagonal stands for the number of a true positive predictions. The saturation of red outside of the diagonal stands for the number of a false negative predictions on the horizontal axis and false positive predictions on the vertical axis. The red gradient values to the right and bottom of the confusion matrix represent aggregate false negative and false positive respectively. The blue gradient values to the right and bottom of the confusion matrix represent aggregate true positive values. These confusion matrices supply a high-level visual that supports the fact that our results produce more true positive results versus false negative and false positive results. Visualization for artists classes greater than 34 is prohibitive due to presentation space. Table 21 lists the most confused artists for each experiment. We define the most confused artists by extracting the two artists that correspond to the item with the maximum value in a confusion matrix excluding diagonal values.

Duplicate values are included. We exclude diagonal values because these represent true positives or an artist that is not being confused with another.

In table 21, we report on the two artists confused for each artist count experiment. We also show the corresponding HPC job to support the fact that intermingling the HPC job runs with baseline experiments does not have an impact on our results in a meaningful way. For example, we show that both jobs have experiments with similar confusion between anoniem/Scherm Laurens and GordonRobert-Jacob/Meissener-Porzellan-Manufa Our 200 class experiment confuses Picart Bernard with Houbraken Jacob and Tanjé Pieter to the same degree. We also see that the initial runs of both HPC jobs have an artist being confused with anoniem or anonymous. After each annealing step completes, the two artists most confused tighten up to two unique artists. Given the confusion matrix in figure 30, we can cross reference to table 21 to see 34 instances where artist 26, Jacob Houbraken, is predicted, but artist 19, Rober Nanteuil, is the correct class. Likewise, Rober Nanteuil is predicted, but the true class is Jacob Houbraken, albeit to the lesser degree of one. The fact that these two artists are confused with one another is no surprise as they are both portrait artists.

Further inspection of figure 30 reveals that artist 13, Arnoud van Halen has the most false negatives, and artist 26, Robert Nanteuil has the most false positives. Both of these artists are portrait artists. Artist 23, Meissen porcelain manufacturer, has most true positives. This art consists of images of porcelain pieces. The features from these pieces are visually distinctive which contributes to the art providing the best performance. These results support the Van Noord et al.[106] experiments. Specifically, Meissen porcelain manufacturer accuracy in the baseline is 97.5% and we

increase this accuracy to 99.8%. We display top performing true positives for each experiment in table 22.
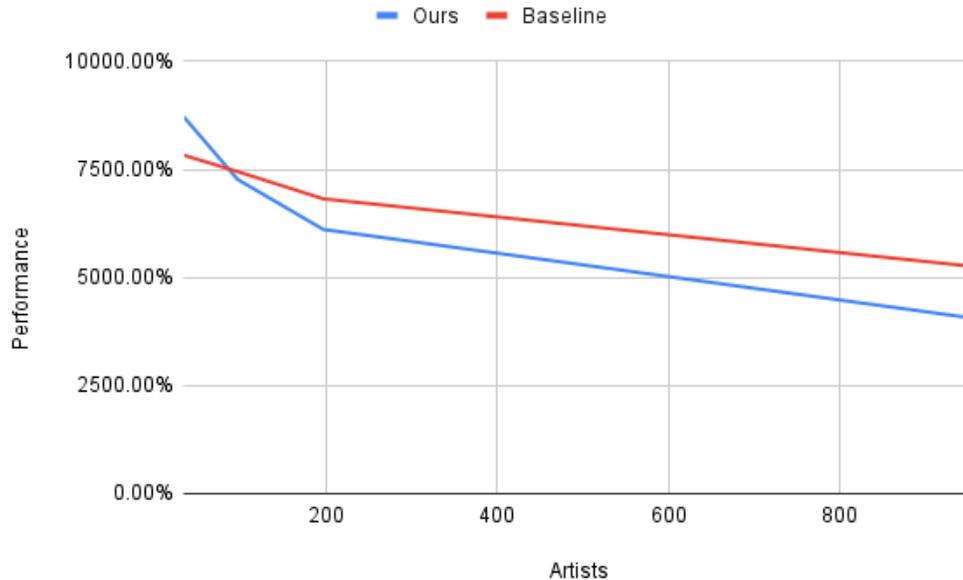
Figure 28: Performance of ours versus the Van Noord et al.[106] related experiments

3.3.5.3   Data Congruence

It is important to make several observations on potential artwork gaps in the data that feeds our MCA results and baseline MCA results. We verify that the complete domain of artwork that we use to produce our experiments is the same as the baseline. We also start our experiments with artists having the same number of artworks. This equates to 50 artworks for the Mensink et al.[70] experiment which ensures at least ten artworks per artist for testing which implies a 35/5/10 split for the first experiment. This equates to 10 artworks for the Van Noord et al.[106] experiment which is the same starting point. However, there may be a margin of error with the artwork domain selection due to the interpretation of which artists make initial
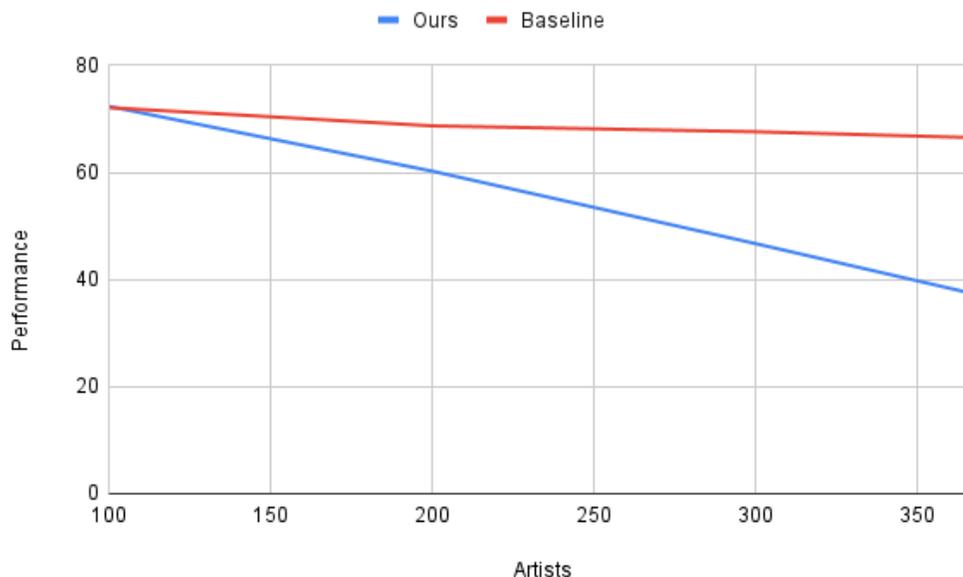
Figure 29: Performance of ours versus the Mensink et al.[70] related experiments

and subsequent performance cuts. For example, we use the same criteria to select 374 artists as Mesink et al.[70], but our selection produces 368 artists. To get our 374 artists performance number, we included a special annealing step after the 958 experiment when running experiments with respect to Van Noord et al.[106]. This is a necessary step to produce the measure. The exception step results in an intersection of 84 classes between class experiment 374 and 368 which would normally result in a six class difference. A similar discrepancy occurs with Van Noord et al.[106] where initial selection of artists with ten or more artworks starts our experiments with 1,199 artists instead of 958 artists. Artist domain shear may also be present with each annealing step. Without the exact artist and art selection for each point of analysis and the code to reproduce these relative states, it's impossible to know if the exact same images are used between baseline and current experiments. Moreover, the annealing process only takes performance into account when generating the artists classes for
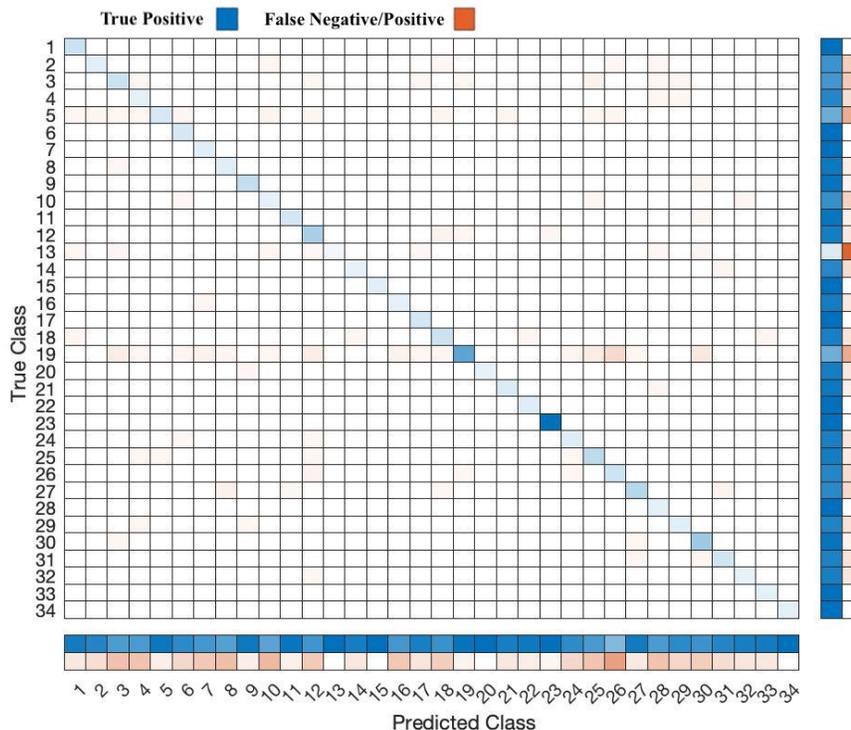
Figure 30: Confusion matrix from our 34 artist experiment

each experiment. Whereas the baseline takes both performance and artwork count into consideration. This is clearly seen in class 34 and 97 measures of table 19 where we show artwork counts in the 60s where the baseline has counts between 128 and 256. Despite these data discrepancies, we believe that our performance increase supports an advancement due its significance over the baseline and curve as seen in figure 28 and 29. To assist with research that would extend our results in the future, we share artist selection from annealing, training progress, and confusion matrices with our code on GitHub.

3.3.5.4   Conclusions

We contribute toward art authentication research using the Rijksmuseum data set by applying a performance annealing residual neural network to baseline experiments.

We also improve the artist classification aspect of the Rijksmuseum challenge for all experiments. Our best increase is 11.35% for 34 artists. Lastly, we provide standard performance calculations and a data source for reproducing our experiments.

3.3.5.5   Future Work

In future work, we want to extend this research in five areas. First, the investigation of adversarial attacks on art authentication will be useful to understand. Second, we want to investigate the saliency maps that support art authentication. Third, we want to investigate increased performance estimates with multi-class classification using pairwise coupling techniques. Fourth, we want to investigate the use of recommender systems with art authentication. Fifth, we believe there may be performance opportunities by tweaking the algorithm, hyper parameters, and transfer learning models of our current experiment. For example, we could use ResNet 152, leverage transfer learning to take advantage of the models produced during the annealing process, or use a different neural network approach such as the feed-forward neural networks with random weights (NNRW) approached discussed by Cao et al.[15].

Table 15: ResNet Hyper Parameters Used for All Experiment Results.

Parameter	Value	Purpose
Image Size	224 × 224 × 3	Resize to match network input
Training	70%	Baseline value
Validation	10%	Baseline value
Test	20%	Baseline value; performance measure source
Image Rotation	random	prevent overfitting
Image Scaling	random	prevent overfitting
Image Reflection	random	prevent overfitting
Image Batch	128	Based on total image count and available resources
Maximum Epochs	30	Training Governor
Validation	every 50 iterations in epoch	Training Governor
Image shuffle	each epoch	Handles indivisible image partition
Initial Input Weight	ImageNet Transfer Learning	Initial weights for neural network
Solver	SGDM	Algorithm that updates weights and biases to minimize the loss function
Learning Rate	0.01	Tuned to ensure training doesn't take too long or results do not diverge
Momentum	0.9	Parameter contribution of the previous iteration to the current iteration
Weight Decay Regularization	0.0001	Reduces overfitting

Table 16: Performance of ResNet 101 with Annealing Compared to Van Noord et al.[106] Baseline

Artist Count	Baseline Performance	Performance	Increase Amount	Increase Percent
958	52.50%	40.51%	-11.99	-22.84%
197	68.20%	61.12%	-7.08	-10.39%
97	74.50%	72.69%	-1.81	-2.43%
34	78.30%	87.19%	8.89	11.35%

Table 17: Performance of ResNet 101 with Annealing Compared to Mensink et al.[70] Baseline

Artist Count	Baseline Performance	Performance	Increase Amount	Increase Percent
374	66.50%	58.60%	-7.90	-11.88%
300	68.70%	46.70%	-22.00	-32.02%
200	72.10%	60.24%	-11.86	-16.45%
100	76.30%	72.34%	-3.96	-5.18%

Table 18: Iteration Count for Each Experiment

Artist Count	Iteration Count
34	1,020
97	3,090
100	3,120
197	6,450
200	6,480
300	12,150
368	13,440
374	3,060
958	15,120
1,199	16,380

Table 19: Top Performers for Each Experiment in First HPC Job

Artist Count	Top Performing Artist(s)	Artwork Count
34	Hausdorff	66
	Voet-430	61
97	Hausdorff	66
	Noé_Michel	54
	Voet-430	61
100	Hausdorff	66
	Noé_Michel	54
197	Breen_Adam-van	84
	Hausdorff	66
	Voet-430	61
200	Hausdorff	66
	Voet-430	61
300	Den-Haag_Porseleinfabriek	66
	Hausdorff	66
368	Hausdorff	66

Table 20: Top Performers for Each Experiment in Second HPC Job

Artist Count	Top Performing Artist(s)	Artwork Count
374	Corvinus_Johann-August	34
	Fuchs_Adam	25
	Hausdorff	66
	Meester-van-Antwerpen-(I)	29
	Montano_Giovanni-Battista	37
	Ravesteyn_Jan-Antonisz-van	28
958	Adam_Richard	20
	Crespi_Giuseppe-Maria	23
	Fuchs_Adam	25
	Groenning_Gerard-P	31
	Hausdorff	66
	Ikku_Jippensha	26
	Kunimasa_Utagawa	23
	Le-Gouaz_Yves-Marie	20
	Matteini_Teodoro	25
	Montano_Giovanni-Battista	37
	Naiwincx_Herman	21
	Rabel_Daniel	25
1199	Adam_Richard	20
	Den-Haag_Porseleinfabriek	66
	Fuchs_Adam	25
	Ikku_Jippensha	26
	Kunimasa_Utagawa	23
	Naiwincx_Herman	21
	Noé_Michel	54

Table 21: Artist Most Confused with Each Experiment

Artist Count	Artist 1	Artist 2	HPC Job
34	Houbraken_Jacob	Nanteuil_Robert	1
97	Galle_Philips	Lepautre_Jean	1
100	Gordon_Robert-Jacob	Meissener-Porzellan-Manufaktur	1
197	Coornhert_Dirck-Volckertsz	Cort_Cornelis	1
200	Houbraken_Jacob	Picart_Bernard	1
	Tanjé_Pieter	Picart_Bernard	1
300	anoniem	Harrewijn_Jacobus	1
368	anoniem	Scherm_Laurens	1
374	Gordon_Robert-Jacob	Meissener-Porzellan-Manufaktur	2
958	anoniem	Scherm_Laurens	2
1199	anoniem	Scherm_Laurens	2

Table 22: Artist with Most True Positives for Each Experiment

Artists Count	Artist Name	True Positive Count
34	Meissener-Porzellan-Manufaktur	201
97	Rembrandt-Harmensz-van-Rijn	249
100	Rembrandt-Harmensz-van-Rijn	231
197	Rembrandt-Harmensz-van-Rijn	201
200	Rembrandt-Harmensz-van-Rijn	215
300	anoniem	1718
368	anoniem	1494
374	Rembrandt-Harmensz-van-Rijn	213
958	anoniem	1087
1199	anoniem	819

CHAPTER 4: ARTFINDER AND LARGE CLASSIFICATION

4.1   Contemporary Art

The identification of the artist of a contemporary painting answers the question who painted the artwork. This is also known as art authentication, and the answer to this question is manifest through art gallery exhibition and is reinforced through financial transaction. Art authentication has visual influence via the uniqueness of the artist's style in contrast to the style of another artist. The significance of this contrast is proportional to the number of artists involved and the degree of uniqueness of an artist's collection. This visual uniqueness of style can be captured in a mathematical model produced by a Machine Learning (ML) algorithm on painting images. However, art authentication is not always possible for contemporary art since art can be anonymous, forged, gifted, or stolen. Here, we show an image only art authentication attribute marker of contemporary art for a very large number of artists. We found that it is possible to authenticate contemporary art for 2,368 artists with an accuracy of 48.97%. These results come from a model generated from a contemporary art database of 170,056 paintings and tested on 42,514 paintings from the same artists but unbeknownst by the model. Our results demonstrate the largest effort for image only art authentication to date with respect to the number of artists involved and the accuracy of authentication. Art authentication is paramount for

the value of an artwork. We anticipate this research will contribute as an additional attribute marker to support art authentication where traditional art authentication methods are inadequate or missing. This attribute marker is used for any artist in the model as a binary attribute. For the prediction of a piece of art for an artist in question, a successful prediction provides a favorable outcome for one artist and an unfavorable outcome to the remaining 2,367 artists. Both the accuracy of the prediction and the number of artists being considered matters.

In the past five years, art authentication has received increased attention due to artificial intelligence, digital image processing, forensic techniques, and legal cases. From an artificial intelligence perspective, supervised deep learning algorithms on painting images achieved an accuracy of 67.78% authenticating art for 90 artists using the WikiArt dataset [27] and an accuracy of 32.40% authenticating art for 1,199 artists using the Rijksmuseum dataset.[28] On the digital image processing front, an accuracy of 91.7% was achieved authenticating art for two artists using Principal Component Analysis (PCA) and a custom van Gogh and Raphael dataset. These results involve fewer artists with the advantage of reduced resource cost.[60] An accuracy of 88% was achieved for authenticating art on an undisclosed number of artists using a decision tree on attribution markers and a custom dataset consisting of 43 authentic paintings and 12 forged paintings. It's important to note that the attribution markers consist of typical forensic metrics that are currently used by art historians for art authentication purposes in addition to markers from the painting image.[64] A similar concept to attribution markers is the forensic technique of optical coherence tomography (OCT) which provides analysis on the cracks in paintings. Both the nature of

paintings cracks and the map of painting cracks for an authenticated artwork provide a quick method for determining art forgeries.[57] From a legal perspective, an art expert is used to authenticate art using methods of connoisseurship, provenance, and scientific analysis. Art experts are not legally regulated, and the methods are subject to human error. A look into the future indicates companies like Art Recognition and academic institutions Rutgers University have proprietary capability to detect intentional forgeries with 80% accuracy with respect to an undisclosed number of artists which represents a step forward in eliminating human error.[67]

Figure 31 demonstrates our process for the creation of an image only art authentication attribute marker to model 2,368 artists. When the process begins, images are partitioned into a training, validation, and test sets. To learn the model, the process trains for up to 30 epochs. An epoch is a learning event that includes all paintings in the training set. In each epoch, training paintings are shuffled and mutated to prevent over fitting the model and the artist's style is gradually learned in batches. At regular intervals the model is validated using validation paintings and the results of learning validation makes changes to the model which are used in the next iteration. Once validation results meet a threshold or 30 epochs pass, the process stops with the current state of the model. This model is used on test paintings to determine the artist. The results of this test produce a confusion matrix which is a table showing true negatives/positives, false negatives, and false positives. True negatives/positives indicate that the model made a correct negative or positive prediction with respect to the artist and painting in question. False negatives indicate that the model predicted another artist instead of the actual artist. False positives indicate that the model predicted the actual artist,

but it should have predicted another artist instead. There are a variety of metrics that can be calculated from this confusion matrix. We use equation 4 for our primary metric.
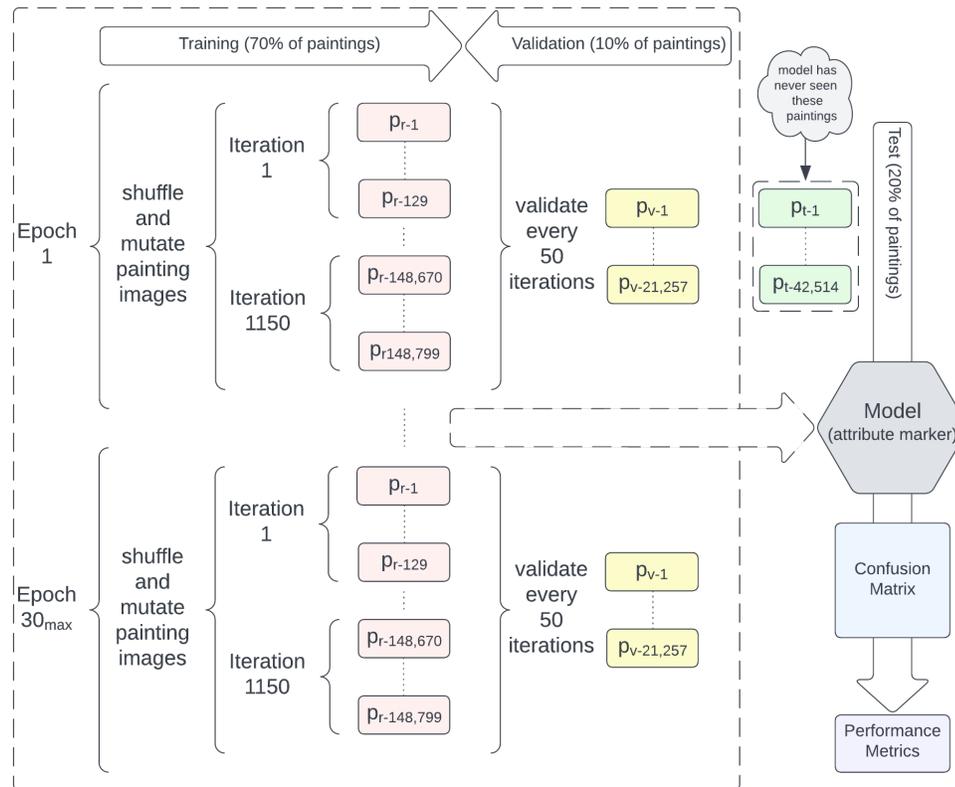
Figure 31: Process for image only art authentication model attribute marker. p_{ri} are paintings in the training set. p_{vi} are paintings in the validation set. p_{ti} are paintings in the test set. Paintings in the training, validation, and test set are mutually exclusive.

4.2   Art Authentication and Artfinder

4.2.1   Machine learning development and evaluation

We develop an ML model to predict an artist by training a model using the state-of-the-art ResNet algorithm to learn relationships between input painting images and corresponding artists which have been labelled manually by visual inspection. We

use recommended hyper parameters to continually validate the model being generated until a desired result is achieved or max validation steps occur. The resulting model is evaluated on an unseen test dataset.

4.2.1.1   Training, Validation and Testing Datasets

An image dataset from ArtFinder of over 212,570 paintings with minimum size of 1200px $\times$ 1200px and sRGB color profile is used for training, validation, and test sets. Images are resized using bilinear interpolation into a $224 \times 224 \times 3$ tensor. The training set is randomly generated from 70% of the images and the validation set is randomly generated from 10% of the images. The remaining 20% of images is set aside for testing after the model is trained. An epoch is defined as a pass through the images in batches of 130-140 images for up to 1,050 times. Up to 30 epochs of training and validation occur in a cross folded fashion every 50 iterations. To mitigate over-fitting concerns, images are shuffled, rotated between -90 and 90 degrees, randomly scaled between 1 and 2 times, and undergo random reflection on the x axis with each epoch. We use transfer learning from ImageNet models for additional initialization of training parameters.

4.2.1.2   Model Selection

We leverage existing work from our literature review to select a model for training. Features extracted from images using SIFT, HOG, and other digital image processing algorithms consumed by basic ML algorithms such as support-vector machines (SVM), decision trees, and k-nearest neighbors algorithm (k-NN) train models quickly, but accuracy starts to suffer quickly as approximately fifty artist classes are approached.

Deep neural networks remedy this limitation at the cost of training time and the need for high performance cluster to generate the model. Of these networks, ResNet 101 outperforms earlier versions of ResNet as well as AlexNet, VGG, GoogLeNet, PigeoNet, and CaffeNet.[110, 18, 17, 70, 106, 99, 58] Moreover, there may be some performance improvements with SENet and deeper versions of ResNet. However, the scope of this work is not to perform a detailed model comparison or improve upon a model that is working well. Therefore, we continue using the ResNet 101 with an annealing process to produce our models.

4.2.1.3   ResNet Architecture

The ResNet algorithm solves the exploding and vanishing gradient problem of deep neural networks with a deep residual learning framework which allows for much deeper networks using the concept of skip connections.[39] A mathematical proof demonstrates how skip connections can largely circumvent the exploding and vanishing gradient problem [78].   ResNet works well with classifying art because the deep network enables multiple passes on an artist's body of work at varying filter sizes in a generic manner. This process produces a model which does a very good job at learning an artist's style.

4.2.1.4   Artist Selection

We perform 24 experiments to understand how classification metrics and artist style evolves as the number of artists is reduced. Our first experiment is seeded with artists having 10 or more artworks. This provides 2,368 artists for our first experiment. Our next experiment consists of 2,300 artists and continues to 100 artists reducing

the artist count by 100 with each annealing iteration. The artists selection criteria to determine which artists will be dropped is based on macro balanced accuracy. We use macro balanced accuracy over micro balanced accuracy because the metric provides a more granular selection for the fitness of an artist which results in less ties [28].

4.2.1.5   Evaluation Using the Testing Set

In a final effort to determine the fitness of our model, we test our model on 20% of the paintings split out before training so we know the model has never seen these paintings. Since the number of paintings produced by artists is naturally unbalanced, and we want a true representation of the artist model without over and under sampling, we use a macro balanced accuracy calculation from the confusion matrix produced by our test.[37, 92] The accuracy ranges from 48.97% for our largest experiment to 91.23% for out smallest experiment. Test accuracy is approximately equal to validation accuracy in all our experiments. This indicates our model isn't subject to over fitting concerns. Note, validation accuracy is calculated by Matlab with each validation iteration which takes into consideration ROC analysis.

4.2.1.6   Limitations with Image Based Art Authentication

Several limitations exist with performing art authentication with painting images alone. First and foremost, it is difficult to acquire data. Both physical and online art galleries protect image data because the image is the primary proprietary asset for sale. Access to the complete collection of an art gallery for research purposes requires a trusted relationship with the gallery or legitimate method of crawling the galleries online website for image data. Second, there is a varying number of

paintings produced by artists which naturally leads to imbalanced data. The task of gathering more data samples is difficult because the time it takes for an artist to produce new works is nondeterministic. From a sampling perspective, we do not want to under sample because our model does not get the opportunity to learn more about an artist's paintings, and we do not want to over sample because we do not get a true representation of the artist's body of work. Therefore, we are left with using techniques to acquire meaningful multi class metrics from our tests that assumes input classes are not in balance.[37, 92] Third, there are many other attribution markers other than a digital representation of the painting when authenticating a painting. These markers have traditionally been used by art historians for art authentication. Over 30 attribution markers are discussed in state-of-the-art research dealing with art authentication. For example, there are markers corresponding to the UV, IR, and X-ray physical analysis of a painting. Markers characterizing the pigments and medium characteristic of the artist and time period are considered. Moreover, there are markers having nothing to do with the actual image such as signature and ownership documents and history.[64] Fourth, there are no paintings representing true negatives on purpose in our experiments.

4.2.1.7 Data source

The data for our experiments comes from the online art marketplace, ArtFinder. Data from this website was collected over several years via automated web crawling technology [81]. We were given permission to use this data in this research and report aggregate results only. We are unable to report on specific artists and paintings in

this research hence the omission of artist name and painting images. Many hours were spent observing raw data images to ensure sound data is being used for our experiments.

4.2.2   Data and Code Availability

The data and code for the training, validation, and test classes are made anonymous and are made available at our GitHub repository `https://github.com/btdobbs/` `Contemporary-Art-Authentication-with-Large-Scale-Classification`.

4.3   Experiments on Artfinder

High level results for our experiments are listed in Table 23. This table represents all 24 experiments starting with 2,368 artists and ending with 100 artists. Validation accuracy (Val Acc) is the accuracy obtained during training. Test accuracy (Test Acc (M)) is the primary metric of interest and is the calculated macro balanced accuracy of the test paintings which were not observed during training. Note, test accuracy is a bit higher than validation accuracy which indicates our model did not encounter any over fitting issues during training. The number of paintings observed during the 70/10/20 split is represented by Train/Val/Test Cnt respectively. The batch size of images used during each iteration of training is represented in the Batch column and the total number of Iterations per epoch is represented by the Iterations header. With each experiment, we also calculated the average number of artworks per artist. We verify that this number always increases except for the very last experiment with 100 artists. It's important that this number is increasing to ensure our model is not influenced most by artists with fewer artworks. Specifically, we start with an average

of 18 paintings per artist for 2,368 artists and end with 41 paintings per artist for 200 artists. The average dips down to 28 for our last experiment of 100 artists. This metric is not listed in table 23 due to limited space.

Table 23: Experiments Results

Artists	Val Acc (μ)	Test Acc (M)	Test Acc Cnt	Train Cnt	Val Cnt	Test Cnt
2,368	67.62%	65.33%	48.97%	148,799	21,257	42,514
2,300	68.09%	66.02%	50.93%	146,913	20,988	41,975
2,200	68.67%	67.20%	52.88%	145,100	20,729	41,457
2,100	69.15%	67.63%	54.84%	143,759	20,537	41,074
2,000	69.71%	68.37%	57.35%	142,464	20,352	40,704
1,900	70.49%	68.95%	59.35%	140,021	20,003	40,006
1,800	71.42%	70.23%	61.05%	137,081	19,583	39,166
1,700	72.49%	71.47%	63.66%	133,795	19,114	38,227
1,600	73.29%	72.76%	65.34%	130,239	18,606	37,211
1,500	74.29%	73.41%	66.80%	126,900	18,129	36,257
1,400	75.76%	74.41%	68.34%	123,260	17,609	35,217
1,300	76.66%	75.93%	70.51%	116,050	16,579	33,157
1,200	77.81%	77.43%	71.77%	110,663	15,809	31,618
1,100	78.83%	78.46%	74.01%	107,279	15,326	30,651
1,000	79.59%	79.57%	75.40%	101,945	14,564	29,127
900	81.34%	81.57%	77.20%	96,138	13,734	27,468
800	82.49%	82.35%	78.36%	89,667	12,810	25,619
700	83.75%	83.35%	80.33%	81,379	11,626	23,251
600	85.59%	85.71%	82.66%	74,582	10,655	21,309
500	86.46%	86.85%	83.60%	67,130	9,590	19,180
400	88.15%	88.51%	85.47%	55,542	7,935	15,869
300	91.11%	91.30%	88.88%	45,567	6,510	13,019
200	93.17%	93.36%	91.15%	28,609	4,087	8,174
100	96.20%	96.29%	91.23%	9,625	1,375	2,750

4.3.1   Confusion Matrix

The confusion matrix in Figure 32 represents our largest experiment. Due to the large number of artist classes, we use a pixel-based confusion matrix where the intensity color of the pixel represents the strength of the metric. The diagonal from

Table 24: Experiments Results Continued

Batch[1]	Iterations[2]
129	1,150
134	1,100
132	1,100
131	1,100
130	1,100
133	1,050
131	1,050
134	1,000
130	1,000
134	950
130	950
129	900
130	850
134	800
136	750
128	750
128	700
136	600
136	550
134	500
139	400
130	350
143	200
193	50

the upper left-hand corner to the lower right-hand corner in the confusion matrix represent correct predictions in the form of true negative and positive predictions. A distinct, visible diagonal is a favorable condition for confusion matrix as this will likely indicate a favorable accuracy metric. Horizontal pixels represent false positives and vertical pixels represent false negatives. The confusion matrix is also partitioned by the primary art style of the artists represented. This provides a method to determine which styles are confused. The primary art style for an artist is determined by the largest count of paintings of a given style for the artist. The

first alphabetical style is used for ties. For example, "artist 1004" has the following painting styles by count: Impressionistic(13), Expressive and gestural(3), Urban and Pop(3), Abstract(2), Geometric(1), Organic(1), and Photorealistic(1). Therefore, we would attribute Impressionistic to "artist 1004". We use the name "artist 1004" because our agreement with ArtFinder, the provider of data, is to keep artist and painting names anonymous.
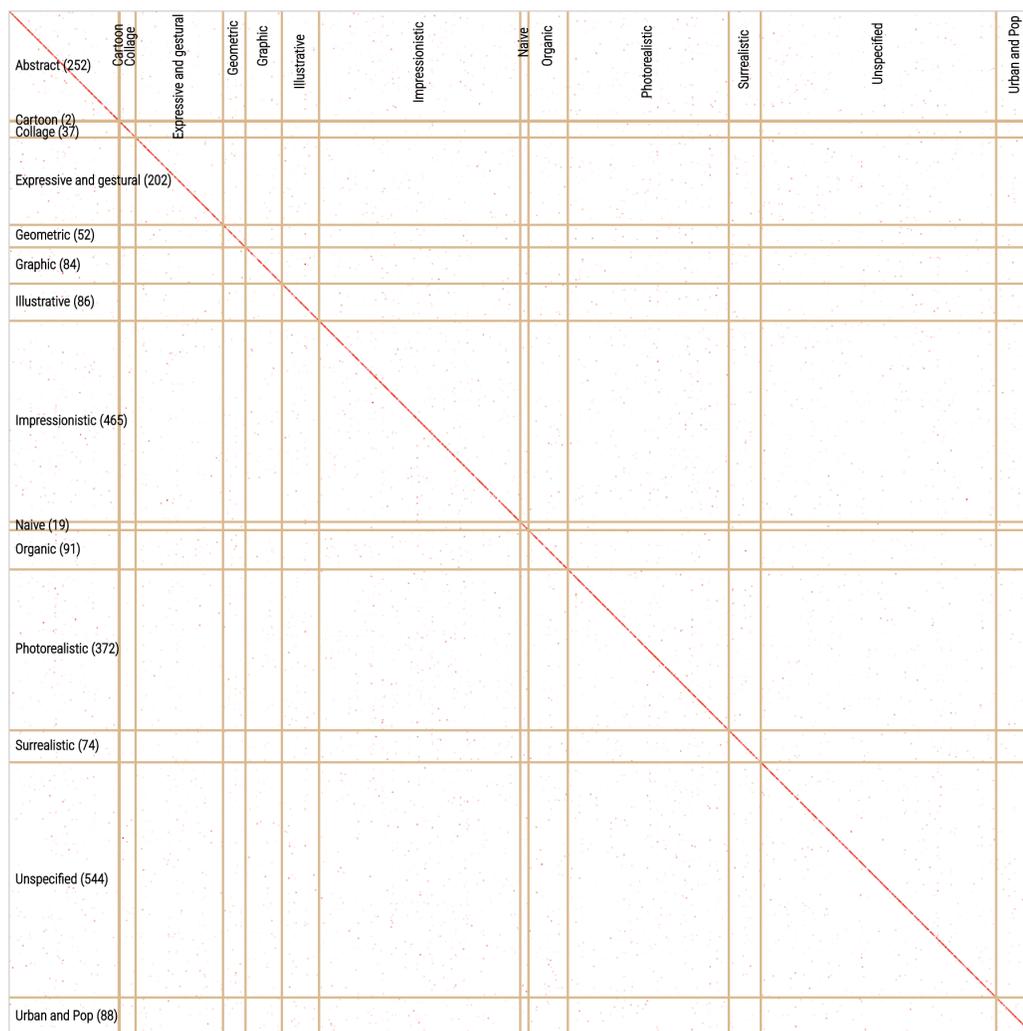
Figure 32: Pixel Based Confusion Matrix of Largest Experiment

4.3.2 Accuracy

The typical average accuracy calculation from our confusion matrix shown in equation 2 cannot be used because it applies to binary classification.[92] We combine several techniques from state-of-the-art multi classification performance measure research to arrive at equation 4. This equation represents macro balanced accuracy which provides a reasonable accuracy because it prevents unbalanced majority and minority classes from influencing the overall accuracy.[37, 92] The corresponding micro balanced accuracy show in equation 3 is also available. This equation reduces to the average multi classification recall calculation over all artists thus it is not used even though it provides a better number for reporting. If our data was balanced on the front end of our experiments, this metric would be legitimate and would converge with macro balanced accuracy.[37] We still report on micro balanced accuracy to demonstrate that it coincides with validation accuracy which is unbalanced. This demonstrates that our model is not over fitting.

$$\text{AvgAcc} = \frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + tn_i + fn_i + fp_i}}{l} \tag{2}$$

$$\text{BalAcc}_\mu = \frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l} tp_i + fn_i} = \frac{\sum_{i=1}^{l} tp_i}{\text{Total Predictions}} \tag{3}$$

$$\text{BalAcc}_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l} \tag{4}$$

4.3.3   Result Discussion

4.3.4   Multiclass Classifier as Binary Classifier

In Table 23, the artist count is inversely proportional to validation and test accuracy. Given state-of-the-art research using multi classification for image only art authentication, this behavior is expected.[27, 28] With our experiments, we hope to reproduce the counter-intuitive phenomenon that a large number of classes can improve multi classification metrics as the number of classes grows.[2] While we did not observe this phenomenon, multi classification for binary classification art authentication problem is important because we don't need to training multiple binary classifiers for the artists of interest.[43] Moreover, training a model on more than one artist produces a model of an artist's paintings in addition to what is not considered a painting by the artist in question. Overall model accuracy is reduced in these situations, but a true positive provides more information regarding for which artists the painting in question does not belong. To show this concept in our experiments, consider artist1051 which exists in all experiments. In 16 of the experiments including the experiment with the most and least artists, the model predicts the artist with 100% accuracy with the test painting data. In three of the experiments, the model predicts the artist with 87.50% accuracy with the test painting data. In five of the experiments, the model predicts the artist with 85.71% accuracy with the test painting data. Given that the accuracy is high in all experiments, the test with the most artists is more meaningful because there are many other potential artists modeled to be confused with in the test.

4.3.5   True Negatives

Adding purposeful true negatives to our experiments would be an interesting addition. This could be accomplished by adding a true negative in the form of a contemporary art forgery. Producing the art forgeries is difficult because forgery paintings are difficult to acquire due to obfuscation of the forgery and rareness of the forgery event. It is also cost prohibitive to commission forgeries due to constraints on time and money. True negatives could also be accomplished by keeping a random sample of paintings from the artists which were removed after each experiment's annealing process. While changing the process to include true negatives from the previous experiments is a straightforward task, it is a time prohibitive task to retrain the models.

4.3.6   Contemporary Art Performance

Performance with the ArtFinder contemporary art dataset outperforms previous experiments with historical art datasets from WikiArt and Rijksmuseum. In table 25, we compare accuracy results using the same macro balanced accuracy metric. In all cases, our ArtFinder experiments outperform WikiArt and Rijksmuseum experiments 10+%.[27, 28] This increase in accuracy for contemporary art may be because that we start with over twice as many artists, and the annealing process can select the best artists for classification once reaching a comparable artist count in previous experiments. It could also be because contemporary art has progressed from historical art in such a way that provides more opportunities to learn artistic style.

Table 25: Dataset Accuracy Comparison

Artists	ArtFinder Acc	WikiArt Acc	Rijksmuseum Acc
1,200	71.77%	n/a	32.40%
1,000	75.40%	n/a	40.51%
400	85.47%	n/a	58.60%
300	88.88%	n/a	46.70%
200	91.15%	n/a	61.12% & 81.66%
100	91.23%	72.96%	72.34% & 72.69%

4.3.7 Artist Style

For all experiments, we show in Figure 33 the percentage of artists styles represented. This is important to show because it shows a variety of art styles are represented from our first experiment with 2,368 artists to our last experiment with 100 artists. The evolution of painting style representations demonstrates that our models do not favor a specific art style.

4.3.8 Uniqueness

This research aims to maximize the number of artists and the accuracy in our experiments. Maximizing both numbers yield the best accuracy with respect to as many artists as possible. It's helpful to have one metric heuristic that considers both numbers. Therefore, we define a uniqueness heuristic as a metric to judge our experiments in terms of a ratio between the accuracy and artist count and number of paintings. We consider the number of artists per painting as artist density. We choose the artist painting density as the numerator and accuracy as the denominator because we are interested in how artist painting density is partitioned by accuracy. This way, we approach artist painting density as accuracy approaches 100% in which case our
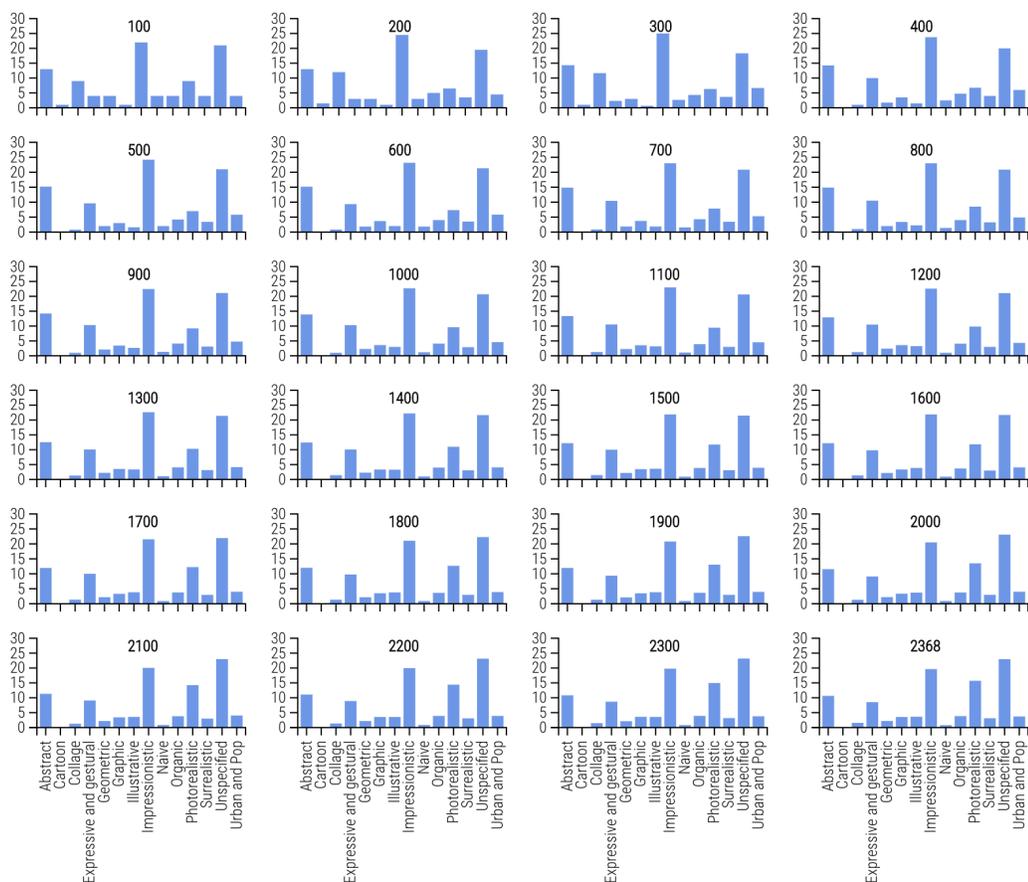
Figure 33: Artist Style Evolution

paintings uniquely define our artists. We calculated uniqueness as the ratio of artist painting density and accuracy seen in equation 5. We show that our largest class experiment yields the largest uniqueness score of 11.37%. As hoped, this uniqueness heuristic is proportional to the artist count in our experiments.

$$\text{Uniqueness} = \frac{\frac{\text{Artist Count}}{\text{Painting Count}}}{\text{Accuracy}} \tag{5}$$

CHAPTER 5: CONCLUSION

5.1   Dissertation Overview

In this dissertation, we examine the process of art authentication from a non-tagged art database via images of paintings from the artists of three galleries. Moreover, the following answers to art authentication questions were discovered from this research.

- ResNet 101 with ImageNet transfer learning performed best when classifying artists given images of their artworks.

- Datasources in the form of artist and artwork were documented and made publicly available for the ArtFinder, Rijksmuseum, and WikiArt such that experiments can be reproduced and improved.

- The multi-class macro balanced accuracy measure performed best for measuring the fitness of an algorithm given the unbalanced nature of artwork data.

- Management of the ResNet 101 algorithm with annealing increased performance.

- A uniqueness calculation was developed and this calculation is the best way to measure the performance of ResNet 101 with annealing from a multi-classification accuracy and cardinality perspective.

- The multi-classification cardinality limits were pushed to 2,400 artists which is the largest experiment of this kind.

- We demonstrated how the models developed from our classification results can be applied to an artist's catalogue raisonné and as an art authentication attribution marker.

In Chapter 2, we discussed the challenge of impermanence with the catalogue raisonné and how a digital catalogue raisonné with a classification model would solve the impermanence issue. We outperformed Stanford's state-of-the-art research for the WikiArt database from a balanced accuracy and cardinality perspective by creating a model with 90 artists and achieved an accuracy of 72.96% using the ResNet 101 algorithm.

In Chapter 3, we applied what we learned in chapter 2 and extend it with an annealing process on the Rijksmuseum gallery for the purpose of art authentication. We demonstrated a performance increase of 11.35% over the baseline for 34 artists from a previous Rijksmuseum experiment and established a new baseline for 1,199 artists.

In Chapter 4, we applied what we learned in chapter 2 and chapter 3 to push cardinality limits to provide a new art authentication marker with the ArtFinder gallery. We established 24 new baselines from 2,368 artists with an accuracy of 48.97% to 100 artists with an accuracy of 91.23%. Our results demonstrate the largest effort for image only art authentication to date with respect to the number of artists involved and the accuracy of authentication.

5.2 Future Directions

The continuation of art authentication research is ideal for students attending a liberal arts college who would like to learn more about advanced computer science and mathematics subjects dealing with digital image processing and machine learning. For example, the following topics can be expanded on or unpacked in several different ways.

- New algorithms core and ancillary to existing algorithms can be applied to existing datasets using the same measures. The algorithms can be implemented with various languages and APIs.

- New art datasets can be used with existing algorithms.

- New measurements for accuracy can be developed and compared with existing measurements.

- New methods for handling data imbalance can be developed and compared with existing methods.

- Core metrics of class cardinality and accuracy can be improved or verified.

- New experiments with transfer learning can be applied.

- A study of the evolution of an image as it is processed by a machine learning algorithm would be a superb learning experience and would produce some interesting visual results.

- Expanding outside of art authentication, art analytics can provide opportunity to study any form of art and the unstructured nature of the images and texts that are characterized by it.

5.3   Acknowledgments

5.4   Data and Materials Availability

All data, code, and materials used in the analysis are available via the link documented in the Data and Code Availability section of each chapter.

REFERENCES

[1] ABC News Internet Ventures. One art lover's crusade to catalog the world. `https://fivethirtyeight.com/features/one-art-lovers-crusade-to-catalog-the-world/`, February 2019. accessed on 2019-02-10.

[2] F. Abramovich and M. Pensky. Classification with many classes: challenges and pluses. *Journal of Multivariate Analysis*, 174:104536, 2019.

[3] Artnet. Artnet price database. `https://www.artnet.com/`, February 2020. accessed on 2020-02-16.

[4] Artprice. Artprice auction database. `https://www.artprice.com`, February 2020. accessed on 2020-02-16.

[5] J. Bailey. Artnome. `https://www.artnome.com/art-analytics/`, January 2019. accessed on 2019-01-27.

[6] A. M. Ben-Ahmeida and A. Y. B. Sasi. Improved image retrieval based on fuzzy colour feature vector. In *International Conference on Graphic and Image Processing (ICGIP 2012)*, volume 8768, page 87686C. International Society for Optics and Photonics, 2013.

[7] S. Bianco, R. Cadene, L. Celona, and P. Napoletano. Benchmark analysis of representative deep neural network architectures. *IEEE Access*, 6:64270–64277, 2018.

[8] A. Blessing and K. Wen. Using machine learning for identification of art paintings. *Technical report*, 2010.

[9] A. Blessing and K. Wen. Using machine learning for identification of art paintings. *Technical report*, page 5, 2010.

[10] Blouin. Blouin art sales index. `https://www.blouinartsalesindex.com`, February 2020. accessed on 2020-02-16.

[11] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 401–408, 2007.

[12] M. Buda, A. Maki, and M. A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249–259, 2018.

[13] P. Cannon-Brookes. Impermanence: A curator's viewpoint. *International Journal of Museum Management and Curatorship*, 2(3):283–285, 1983.

[14] W. Cao, M. J. Patwary, P. Yang, X. Wang, and Z. Ming. An initial study on the relationship between meta features of dataset and the initialization of nnrw. In *2019 international joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2019.

[15] W. Cao, X. Wang, Z. Ming, and J. Gao. A review on neural networks with random weights. *Neurocomputing*, 275:278–287, 2018.

[16] E. Cetinic, T. Lipic, and S. Grgic. Fine-tuning convolutional neural networks for fine art classification. *Expert Systems with Applications*, 114:107–118, 2018.

[17] E. Cetinic, T. Lipic, and S. Grgic. Fine-tuning convolutional neural networks for fine art classification. *Expert Systems with Applications*, 114:107–118, 2018.

[18] J. Chen. Comparison of machine learning techniques for artist identification.

[19] J. Chen and A. Deng. Comparison of machine learning techniques for artist identification. *StandfordCS229 Report*, 2018.

[20] X. Chen, H. Fang, T.-Y. Lin, R. Vedantam, S. Gupta, P. Dollár, and C. L. Zitnick. Microsoft coco captions: Data collection and evaluation server. *arXiv preprint arXiv:1504.00325*, 2015.

[21] X. Chen and C. Lawrence Zitnick. Mind's eye: A recurrent visual representation for image caption generation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2422–2431, 2015.

[22] L. Cieplinski. Mpeg-7 color descriptors and their applications. In *International Conference on Computer Analysis of Images and Patterns*, pages 11–20. Springer, 2001.

[23] L. A. Contributors. Linked art. `https://linked.art`, June 2017. accessed on 2019-04-05.

[24] W. Corp. Wikiart visual art encyclopedia. `https://www.wikiart.org`, February 2019. accessed on 2019-02-02.

[25] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

[26] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE, 2009.

[27] T. Dobbs, A. Benedict, and Z. Ras. Jumping into the artistic deep end: building the catalogue raisonné. *AI & SOCIETY*, Jan. 2022.

[28] T. Dobbs and Z. Ras. On art authentication and the rijksmuseum challenge: A residual neural network approach. *Expert Systems with Applications*, 200:116933, 2022.

[29] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2625–2634, 2015.

[30] Z. Falomir, L. Museros, I. Sanz, and L. Gonzalez-Abril. Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (qart-learn). *Expert Systems with Applications*, 97:83–94, 2018.

[31] H. Feliciano and H. Felliciano. *The lost museum: the Nazi conspiracy to steal the world's greatest works of art.* Basic Books New York, 1997.

[32] A. Friedenthal. John smith's rembrandt research project: An art dealer establishes the first catalogue raisonné of the paintings (1836). *Netherlands Yearbook for History of Art / Nederlands Kunsthistorisch Jaarboek Online*, 69(1):212–247, 2020.

[33] A. Friedman. Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of experimental psychology: General*, 108(3):316, 1979.

[34] George Hopper. Nighthawks. `https://www.artic.edu/artworks/111628/nighthawks`, January 1942. accessed on 2019-03-07.

[35] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[36] Google. Google arts and culture. `https://artsandculture.google.com`, February 2020. accessed on 2020-02-16.

[37] M. Grandini, E. Bagli, and G. Visani. Metrics for multi-class classification: an overview. *arXiv:2008.05756 [cs, stat]*, 2008.

[38] J. Han and K.-K. Ma. Fuzzy color histogram and its use in color image retrieval. *IEEE transactions on image processing*, 11(8):944–952, 2002.

[39] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[40] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[41] E. J. Henson. The last prisoners of war: Returning world war ii art to its rightful owners-can moral obligations be translated into legal duties. *DePaul L. Rev.*, 51:1103, 2001.

[42] Hiscox. Hiscox online art trade report 2021. `https://www.hiscox.co.uk/sites/default/files/documents/2022-04/21674b-Hiscox_online_art_trade_report_2021-part_two_1.pdf`, March 2021. accessed on 2022-06-01.

[43] P. Honeine, Z. Noumir, and C. Richard. Multiclass classification machines with the complexity of a single binary classifier. *Signal Processing*, 93(5):1013–1026, 2013.

[44] A. Hosny. The green canvas. `https://ahmedhosny.com/greenCanvas`, January 2019. accessed on 2019-01-27.

[45] C. Johnson, E. Hendriks, I. Berezhnoy, E. Brevdo, S. Hughes, I. Daubechies, J. Li, E. Postma, and J. Wang. Image processing for artist identification. *IEEE Signal Processing Magazine*, 25(4):37–48, 2008.

[46] C. R. Johnson, E. Hendriks, I. J. Berezhnoy, E. Brevdo, S. M. Hughes, I. Daubechies, J. Li, E. Postma, and J. Z. Wang. Image processing for artist identification. *IEEE Signal Processing Magazine*, 25(4):37–48, 2008.

[47] J. Johnson. *Compositional Visual Intelligence*. PhD thesis, Stanford University, 2018.

[48] J. Johnson, A. Karpathy, and L. Fei-Fei. Densecap: Fully convolutional localization networks for dense captioning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4565–4574, 2016.

[49] J. Johnson, R. Krishna, M. Stark, L.-J. Li, D. Shamma, M. Bernstein, and L. Fei-Fei. Image retrieval using scene graphs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3668–3678, 2015.

[50] J. M. Johnson and T. M. Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):27, 2019.

[51] J. Jou and S. Agrawal. Artist identification for renaissance paintings, 2011.

[52] J. Jou and S. Agrawal. Artist identification for renaissance paintings, 2011.

[53] Kaggle. Painter by numbers. `https://www.kaggle.com/c/painter-by-numbers`, October 2016. accessed on 2020-03-20.

[54] A. Karpathy and L. Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3128–3137, 2015.

[55] A. Karpathy, A. Joulin, and L. F. Fei-Fei. Deep fragment embeddings for bidirectional image sentence mapping. In *Advances in neural information processing systems*, pages 1889–1897, 2014.

[56] P. Kim. *MATLAB deep learning: with machine learning, neural networks and artificial intelligence.* For professionals by professionals. Apress, 2017. OCLC: 1002109952.

[57] S. Kim, S. M. Park, S. Bak, G. H. Kim, C.-S. Kim, C. E. Kim, and K. Kim. Advanced art authentication in oil paintings using precise 3d morphological analysis of craquelure patterns. *unknown*, 2021.

[58] K. Kondo and T. Hasegawa. CNN-based criteria for classifying artists by illustration style. In *Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing*, pages 93–98. ACM, 2020.

[59] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[60] R. Leonarduzzi, H. Liu, and Y. Wang. Scattering transform and sparse linear classifiers for art authentication. *Signal Processing*, 150:11–19, 2018.

[61] C. Li and A. C. Bovik. Three-component weighted structural similarity index. In S. P. Farnand and F. Gaykema, editors, *unknown*, page 72420Q, 2009.

[62] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999.

[63] M. Lux and S. A. Chatzichristofis. Lire: lucene image retrieval: an extensible java cbir library. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 1085–1088, 2008.

[64] B. I. Łydżba-Kopczyńska and J. Szwabiński. Attribution markers and data mining in art authentication. *Molecules*, 27(1):70, 2021.

[65] H. Mao, M. Cheung, and J. She. Deepart: Learning joint representations of visual arts. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1183–1191. ACM, 2017.

[66] J. Mao, W. Xu, Y. Yang, J. Wang, Z. Huang, and A. Yuille. Deep captioning with multimodal recurrent neural networks (m-rnn). *arXiv preprint arXiv:1412.6632*, 2014.

[67] B. Mar. Experts' role in art authentication. *unknown*, 2021.

[68] MATLAB. Neural pattern recognition tool. `https://www.mathworks.com/help/deeplearning/gs/classify-patterns-with-a-neural-network.html`, March 2020. accessed on 2020-03-08.

[69] D. Matthew and R. Fergus. Visualizing and understanding convolutional neural networks. In *Proceedings of the 13th European Conference Computer Vision and Pattern Recognition, Zurich, Switzerland*, pages 6–12, 2014.

[70] T. Mensink and J. van Gemert. The rijksmuseum challenge: Museum-centered visual recognition. In *Proceedings of International Conference on Multimedia Retrieval*, ICMR '14. Association for Computing Machinery, 2014.

[71] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[72] R. Museum. Rijks museum. `https://www.rijksmuseum.nl`, November 1798. accessed on 2020-04-05.

[73] Museum of Modern Art. The museum of modern art (moma) collection. `https://github.com/MuseumofModernArt/collection`, May 2019. accessed on 2019-05-25.

[74] W. G. of Art. Web gallery of art. `https://www.wga.hu/`, February 2020. accessed on 2020-02-16.

[75] J.-R. Ohm, L. Cieplinski, H. J. Kim, S. Krishnamachari, B. Manjunath, D. S. Messing, and A. Yamada. Color descriptors. *Introduction to MPEG-7*, pages 187–212, 2002.

[76] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175, 2001.

[77] J. Panero. "i am the central victim": Art dealer ann freedman on selling $63 million in fake paintings. *New York Magazine*, 27, 2013.

[78] G. Philipp, D. Song, and J. G. Carbonell. Gradients explode-deep networks are shallow-resnet explained. *arXiv*, 2018.

[79] R. Pirrone, V. Cannella, O. Gambino, A. Pipitone, and G. Russo. WikiArt: An ontology-based information retrieval system for arts. In *2009 Ninth International Conference on Intelligent Systems Design and Applications*, pages 913–918. IEEE, 2009.

[80] H. Pishro-Nik. *Introduction to probability, statistics, and random processes.* Kappa Research, LLC, 2014.

[81] L. Powell, A. Gelich, and Z. W. Ras. How to raise artwork prices using action rules, personalization and artwork visual features. *Journal of Intelligent Information Systems*, 57(3):583–599, 2021.

[82] R. Quian Quiroga and C. Pedreira. How do we see art: an eye-tracker study. *Frontiers in human neuroscience*, 5:98, 2011.

[83] F. Rayar. Imagenet mpeg-7 visual descriptors-technical report. *arXiv preprint arXiv:1702.00187*, 2017.

[84] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[85] A. M. Research. All art index. `https://www.artmarketresearch.com`, February 2020. accessed on 2020-02-16.

[86] K. Rogers. Viewpoint: The catalogue raisonné scholars association (CRSA) - ProQuest.

[87] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet large scale visual recognition challenge. *arXiv:1409.0575 [cs]*, 2015.

[88] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[89] I. Sanz, L. Museros, Z. Falomir, and L. Gonzalez-Abril. Customising a qualitative colour description for adaptability and usability. *Pattern Recognition Letters*, 67:2–10, 2015.

[90] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

[91] L. Shamir, T. Macura, N. Orlov, D. M. Eckley, and I. G. Goldberg. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. *ACM Transactions on Applied Perception (TAP)*, 7(2):8, 2010.

[92] M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009.

[93] M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437, 2009.

[94] Sothebys. Sothebys. `https://www.sothebys.com/`, February 2020. accessed on 2020-02-16.

[95] J. Stallabrass. *Contemporary art: a very short introduction*, volume 146. Oxford University Press, 2006.

[96] Stanford. Imagenet. `http://image-net.org`, October 2016. accessed on 2020-03-20.

[97] G. Strezoski. Omniart the only artistic dataset you will ever need. `http://www.vistory-omniart.com`, October 2018. accessed on 2020-04-05.

[98] G. Strezoski, A. Shome, R. Bianchi, S. Rao, and M. Worring. Ace: Art, color and emotion. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1053–1055, 2019.

[99] G. Strezoski and M. Worring. OmniArt: Multi-task deep learning for artistic data analysis. *arXiv*, 2017.

[100] G. Strezoski and M. Worring. Omniart: a large-scale artistic benchmark. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(4):1–21, 2018.

[101] G. Strezoski and M. Worring. OmniArt: A large-scale artistic benchmark, 2018.

[102] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[103] C. Szegedy, S. Reed, D. Erhan, D. Anguelov, and S. Ioffe. Scalable, high-quality object detection. *arXiv preprint arXiv:1412.1441*, 2014.

[104] L. Torrey and J. Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.

[105] H. J. Van Miegroet, K. P. Alexander, and F. Leunissen. Imperfect data, art markets and internet research. In *Arts*, page 76. Multidisciplinary Digital Publishing Institute, 2019.

[106] N. van Noord, E. Hendriks, and E. Postma. Toward discovery of the artist's style: Learning to recognize artists by their artworks. *IEEE Signal Processing Magazine*, 32:46–54, 2015.

[107] E. A. Vessel, G. G. Starr, and N. Rubin. Art reaches within: aesthetic experience, the self and the default mode network. *Frontiers in Neuroscience*, 7:258, 2013.

[108] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164, 2015.

[109] N. Viswanathan. Artist identification with convolutional neural networks. *Standford193CS231N Report*, 2017.

[110] N. Viswanathan and Stanford. Artist identification with convolutional neural networks, 2017.

[111] X. Wang and W. Cao. Non-iterative approaches in training feed-forward neural networks and their applications, 2018.

[112] Z. Wang, L. Yang, Q. Wang, D. Liu, Z. Xu, and S. Liu. Artchain: blockchain-enabled platform for art marketplace. In *2019 IEEE International Conference on Blockchain (Blockchain)*, pages 447–454. IEEE, 2019.

[113] E. Winkleman. *Selling contemporary art: How to navigate the evolving market.* Simon and Schuster, 2015.

[114] D. Wissbroecker. Six klimts, a picasso, & (and) a schiele: Recent litigation attempts to recover nazi stolen art. *DePaul-LCA J. Art & Ent. L.*, 14:39, 2004.

[115] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057, 2015.

[116] H. Yang and K. Min. A multi-column deep framework for recognizing artistic media. *Electronics*, 8(11):1277, 2019.

[117] H. Yang and K. Min. Classification of basic artistic media based on a deep convolutional approach. *The Visual Computer*, 36(3):559–578, 2020.

[118] M. Yelizaveta, C. Tat-Seng, and A. Irina. Analysis and retrieval of paintings using artistic color concepts. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 1246–1249. IEEE, 2005.

[119] M. Zhaofeng, H. Weihua, and G. Hongmin. A new blockchain-based trusted drm scheme for built-in content protection. *EURASIP Journal on Image and Video Processing*, 2018(1):1–12, 2018.

APPENDIX A: COLOR FEATURES

5.5 Color Layout

Color Layout is a MPEG-7 color descriptor. Its primary use is for fast searching in an image database, and its representation is a compact, spatial distribution of colors [75, 22].

5.6 Color Structure

Color Structure is a MPEG-7 color descriptor. Its primary use is for multimedia retrieval, and its representation is a generalized color histogram that retains some spacial representation [75, 22].

5.7 Dominant Color

Dominant Color is a MPEG-7 color descriptor. Its primary use is for searching an image database via several colors, and its representation is a compact structure of colors in an image [75, 22].

5.8 Scalable Color

Scalable Color is a MPEG-7 color descriptor. Its primary use is for storage efficiency and image frame groups, and its representation is a series of Haar transforms over an initial 256 bin color histogram [75, 22].

5.9 Entropy

Entropy measures the aggregate expectation of color pixels occurring in an image. The formula below computes the overall, red, green, and blue components. $H(X) = -\sum p(X) \log p(X)$

APPENDIX B: TEXTURE FEATURES

5.10   Edge Histogram

The Edge Histogram is a MPEG-7 texture descriptor that consists of a histogram of five orientations. After a grid overlays an image, each grid element uses the pixels within to assign one of five orientations which feed the histogram through aggregation [83].

5.11   SIFT

SIFT or scale-invariant feature transform is a popular algorithm used in computer vision object detection. It decomposes feature vectors in an image in such a way to preserve the relative location of such vectors. The decomposition allows features of the image to be recognized such that image transformation does not affect detection performance [62].

5.12   GIST

The psychological phenomenon that "an abstract representation of the scene that spontaneously activates memory representations of scene categories" is known as GIST and is how humans recognize scenes [33]. This phenomenon is applied by perceptual mapping properties such as naturalness, openness, roughness, ruggedness, and expansion to second-order statistics and spatial arrangement of structures [76].

Table 26: All Artist artwork distribution along with the training, validation, and test counts used in experiments

Artist	Artwork	Training	Validation	Test

	Count	Count	Count	Count
Albert Bierstadt	336	235	51	50
Albrecht Durer	707	495	106	106
Alexander Roitburd	264	185	40	39
Alfred Freddy Krupa	598	419	90	89
Alfred Sisley	471	330	71	70
Amedeo Modigliani	349	244	53	52
Boris Kustodiev	645	452	97	96
Byzantine Mosaics	255	179	38	38
Camille Corot	498	349	75	74
Camille Pissarro	881	617	132	132
Charles M Russell	278	195	42	41
Chicote Cfc	307	215	46	46
Childe Hassam	550	385	83	82
Claude Monet	1,366	956	205	205
David Burliuk	400	280	60	60
Edgar Degas	625	438	94	93
Egon Schiele	299	209	45	45
Ernst Ludwig Kirchner	387	271	58	58
Eugene Boudin	560	392	84	84
Eyvind Earle	422	295	64	63
Felix Vallotton	314	220	47	47

Ferdinand Hodler	256	179	39	38
Fernand Leger	446	312	67	67
Francis Bacon	312	218	47	47
Francisco Goya	284	199	43	42
George Grosz	158	111	24	23
George Stefanescu	264	185	40	39
Giovanni Battista Piranesi	1,353	947	203	203
Gustave Courbet	270	189	41	40
Gustave Dore	389	272	59	58
Gustave Loiseau	258	181	39	38
Henri De Toulouse Lautrec	372	260	56	56
Henri Fantin Latour	289	202	44	43
Henri Martin	408	286	61	61
Henri Matisse	999	699	150	150
Honore Daumier	254	178	38	38
Hryhorii Havrylenko	408	286	61	61
Ilya Repin	541	379	81	81
Isaac Levitan	449	314	68	67
Ivan Aivazovsky	579	405	87	87
Ivan Shishkin	522	365	79	78
Jacek Yerka	308	216	46	46
James Tissot	432	302	65	65

Jean Auguste Dominique Ingres	259	181	39	39
Joaqu N Sorolla	365	256	55	54
John Henry Twachtman	255	179	38	38
John Singer Sargent	800	560	120	120
Katsushika Hokusai	265	186	40	39
Kazimir Malevich	360	252	54	54
Kenneth Noland	271	190	41	40
Konstantin Korovin	317	222	48	47
Konstantin Makovsky	366	256	55	55
Konstantin Somov	254	178	38	38
Konstantin Yuon	293	205	44	44
Louis Comfort Tiffany	261	183	39	39
Lucian Freud	283	198	43	42
M C Escher	469	328	71	70
Marc Chagall	1,018	713	153	152
Martiros Sarian	551	386	83	82
Mary Cassatt	304	213	46	45
Maurice Prendergast	379	265	57	57
Max Ernst	368	258	55	55
Nicholas Roerich	1,834	1,284	275	275
Odilon Redon	455	319	68	68
Pablo Picasso	1,139	797	171	171

Paul Cezanne	587	411	88	88
Paul Gauguin	512	358	77	77
Peter Paul Rubens	395	277	59	59
Pierre Auguste Renoir	1,409	986	212	211
Pyotr Konchalovsky	925	648	139	138
Raphael Kirchner	525	368	79	78
Rembrandt	765	536	115	114
Rene Magritte	371	260	56	55
Robert Henri	263	184	40	39
Roger Weik	502	351	76	75
Salvador Dali	1,164	815	175	174
Samuel Peploe	252	176	38	38
Stanley Spencer	270	189	41	40
Theodor Severin Kittelsen	375	263	56	56
Theophile Steinlen	1,136	795	171	170
Thomas Eakins	306	214	46	46
Titian	245	172	37	36
Utagawa Kuniyoshi	418	293	63	62
Vasily Surikov	267	187	40	40
Veletanlic Darmin	668	468	100	100
Vincent Van Gogh	1,931	1,352	290	289
William Adolphe Bouguereau	259	181	39	39

William Merritt Chase	377	264	57	56
Zdislav Beksinski	708	496	106	106
Zinaida Serebriakova	415	291	62	62