2022

Using Convolutional Neural Networks to Classify Art Genre

Jacob DuBois
John Carroll University, jdubois22@jcu.edu

Using Convolutional Neural Networks to Classify Art Genre

Jacob DuBois

Honors Program Capstone

Dr. Kilbride

May 6, 2022

Abstract

In 2006, Geoffry Hinton published a paper showing how to train a neural network capable of recognizing handwritten digits with greater than 98% accuracy, a technique he branded “deep learning.” Until this paper was published, training a deep neural network was widely considered impossible and most attempts were abandoned during the 1990’s (Hinton, 2007). As a result, this paper revived the interest of the scientific community and before long hundreds more papers proved that not only was deep learning possible, but capable of mind-blowing achievements that no other machine learning algorithm technique could even attempt. Since then, neural networks have become a part of everyday life and we have created models that drive cars, recognize our face and voice, read and write, and so much more that we take for granted. However, perhaps the most impressive branch of neural networks are convolutional neural networks or CNNs. These neural networks utilize a special component known as a conventional layer that they use to accomplish impressive feats of image recognition. These models are responsible for the unlocking function of our iPhones, the lane detection and autopilot features of modern cars, weather detection and prediction using satellite imagery, and so many other applications.

The increasing availability of these models and their impressive capabilities opens a new research perspective in the interaction of neural networks and practically any visual medium. There have been a plethora of models trained to watch and write movies, identify characters and actors on TV and movies, and a multitude of models

devoted to the creation and classification of artworks. Similarly, the goal of this project is to use these CNNs to classify and identify various art genres.

Introduction

Neural Networks

Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks (SNNs), are a subset of machine learning that works to mimic the computational processes of the human brain. The brain is made up of neurons who send signals to one another through connections known as synapses. Neurons transmit electrical signals to other neurons based on the signals they receive from other neurons. Similarly, artificial neurons within a neural network simulate this process by adding together the values (dot products) of the inputs it receives. If the data the neuron receives is above some threshold it sends its own signal to its output which is received by neurons on the next layer. This data is also weighted in such a way that some inputs are treated as more or less important compared to other inputs. These weights are adjusted throughout the training process in order to ensure that the final output of the network is accurate in what it is trying to achieve.

These artificial neural networks are at the very core of deep learning. They are versatile, powerful and scalable, making them ideal to tackle large and complex machine learning tasks such as classifying billions of images, speech recognition, and search engine optimization (Udyavar, 2017).

Convolutional Neural Networks

Convolutional neural networks (CNNs) are a specialized type of neural network model designed for working with visual data. The defining feature of this type of neural network is the use of a convolutional layer which performs an operation on the input data called a convolution. A convolution is a linear operation that involves the multiplication of weights with the input data, much like a traditional neural network. This multiplication is performed between an array of input data and a two dimensional matrix of weights known as a filter or a kernel. These filters are applied to the entire image through overlapping grids of pixels that shift over by a set number of pixels (defined as strides) until the entire image is covered. Each filter is specifically designed to identify important patterns within the image such as vertical or horizontal edges, circles and other shapes that are weighted through the neural network to identify a larger pattern or particular image. As a result, these models are particularly successful in image classification and other visual machine learning tasks (Saha, 2018).

Artworks

With the advent of the internet and network communication, art painting images can be easily found in everyday life. The availability of these artworks make it possible to easily explore and enjoy artworks scattered in galleries around the world without actually needing to visit them. This widespread availability has obviously been particularly useful for the study and research of art history, while also increasing the number of art fans and hobbyists. However, artwork in this new digital format has also opened new research opportunities in the intersections of art history and the interdisciplinary field of computer vision and machine learning. As a result, convolutional neural networks pose a myriad of interesting applications and research opportunities, previously thought impossible. For example, developers from around the world have been using these tools to create computer generated images and artworks, some of which have been put on display in galleries and museums. Furthermore, others have used these tools to dissect and describe artworks based on their subject, mediums, or artist using image recognition in the process.

However, another interesting subject of research in these fields, and the topic of this paper, is the use of these convolutional neural networks to classify and categorize artworks based on the genre. In these kinds of models a neural network is trained using a dataset of artwork images which it uses to recognize patterns in each image and genre. The model then makes a classifying prediction for each of these images and continues this process until a desired accuracy is reached. By the end of this process the model can then be given a completely new image and generate a prediction based on what it learned using the previous dataset.

It is worth noting that this has been accomplished before, that there are apps and services designed around this idea. One such application is called ‘Smartify.’ This application uses a similar deep convolutional neural network to allow users to scan and identify art work within a museum. However, this application works differently by using existing databases within museums. It uses image recognition not to make predictions about an art piece, but rather to predict which art piece an image is based on the art within the museum. However, the goal of this project is somewhat different. Rather than identifying the actual art piece we have created a neural network that can classify art work and disseminate information from that prediction. This requires much more data, and can be used on artwork regardless of where the piece is located.

I would like to create a neural network that can classify art work and disseminate information from that prediction. For example, if you used smartify in the Cleveland Museum of Art to identify a piece, say *Cupid and Psyche,* smartify would use its knowledge of the galleries at the museum to identify the piece. From there it would tell you information about that piece and perhaps its artist. My goal differs in that it would use a more robust knowledge base to identify the piece as a neoclassical artwork and then give you information about the neoclassical style and about the artists that contributed to its movement. The Smartify answer to this question requires more specific knowledge about a museum or place as well as the art that it contains, however my goal is more general, something that could be used virtually anywhere with almost any art piece.

Related work

Our project was inspired by a series of academic papers that sought to answer similar research questions. Primarily the paper "Fine-tuning Convolutional Neural Networks for Fine Art Classification," which used CNNs for a series of art related applications. This paper focuses more on identifying art by artists and artworks though its methodology and data sources are similar to our own. Similarly, another paper we found useful while developing our methodology was "Art Painting Identification using Convolutional Neural Networks," which focused more on the real word application of such a project. More specifically, the paper focuses on how CNNs are able to differentiate individual pictures from a variety of angles and perspectives. Although this is not entirely important to the foundation of this project, it does highlight an interesting problem we would face if we were to use our model on uncurated images.

Data collection

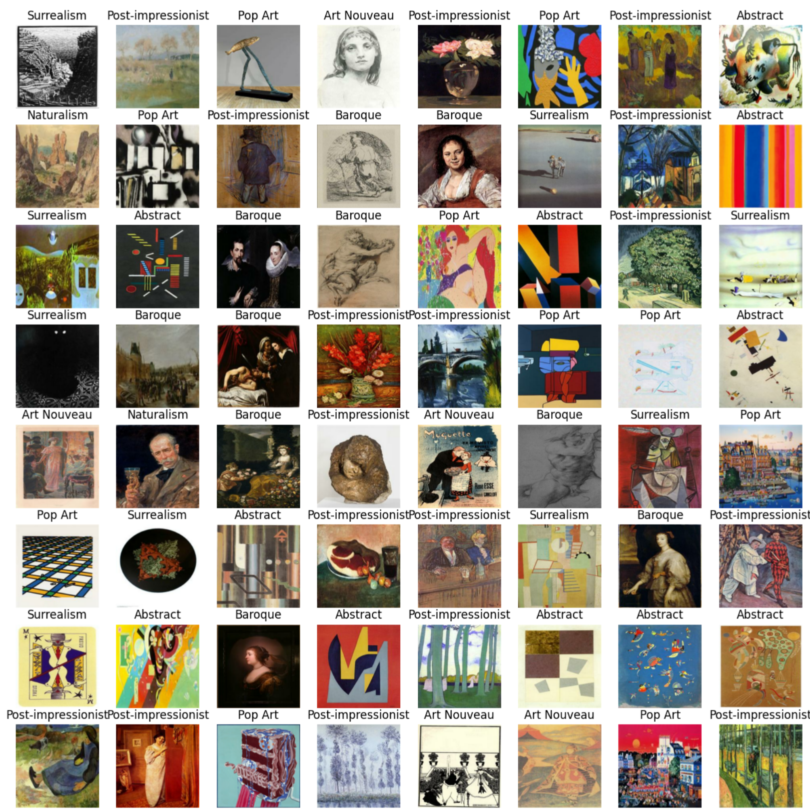
With the aim to include the largest possible number of paintings as well as a wide range of classifying genres, we needed a powerful and expansive source to create our datasets. In order to reach this goal, we used WikiArt, currently the largest online available collection of digitized paintings. WikiArt is a well organized collection and includes a plethora of information including artist, style, genre, nationality, etc for each piece in their collection. Furthermore, it includes artworks from a wide range of time periods, styles, and nationalities. As a result of its extensiveness, WikiArt is commonly used to create datasets used for these kinds of research applications.

In order to use this collection of artworks, we first had to create a script that cycled through the website's collection and downloaded artworks to our personal database. We accomplished this by using python and the Selenium, Beautiful Soup, and Urllib libraries. Furthermore, we also had to narrow our search parameters in order to create a dataset that was both robust and technologically feasible.

Classifications

This was accomplished by selecting only a handful of genres that were both different enough from each other as well as contained enough data to run the model. Originally we were able to scrape data from 12 different genres: Abstract, Naturalism, Surrealism, Post-impressionism, photo-realism, Superflat, Baroque, Art Nouveau, Pop Art, Renaissance, Impressionism, and Neoclassicism. After these genres were decided and sufficient data was collected, we hand filtered the data in order to remove images. This included images that were not indicative of the genre as well as images, mainly website banners, that were not of actual artwork. With the data collected and filtered we were then able to run multiple models in order to find which genres the computer could best differentiate. These test models allowed us to conclude on a dataset using only seven of the original twelve genres: Pop art, Abstract Art, Baroque, Naturalism, Post-Impressionism, Surrealism, and Art Nouveau. We found that the model had a particularly difficult time differentiating between Baroque, Renaissance, and

Neoclassical as well as Post-Impressionist and Impressionist artworks. As a result, we decided to limit the dataset by keeping the largest of the genres (Baroque and Post-Impressionist) and removing the Impressionist, Renaissance, and Neoclassical datasets. Furthermore, we saw that there were an extremely limited number of artworks included in both the superflat and photo realism genres so we removed those as well. Each of the remaining datasets contains 3545 artworks, with the exception of Pop art which contains 2690 pieces, and naturalism which contains only 1237 pieces.



A Sample of the Training Dataset

CNN Architecture

The main CNN architecture used in our final model was created using the Tensorflow library and contains seven layers including the input and output layers. Each is a convolutional layer with 64 nodes and a rectified linear unit, or relu, activation function. Max-pooling layers also follow after each of the layers and the last fully connected layer is connected to a softmax layer that determines the probability of each class. The input of the network is also a 256x256 pixel crop of the resized RGB image.

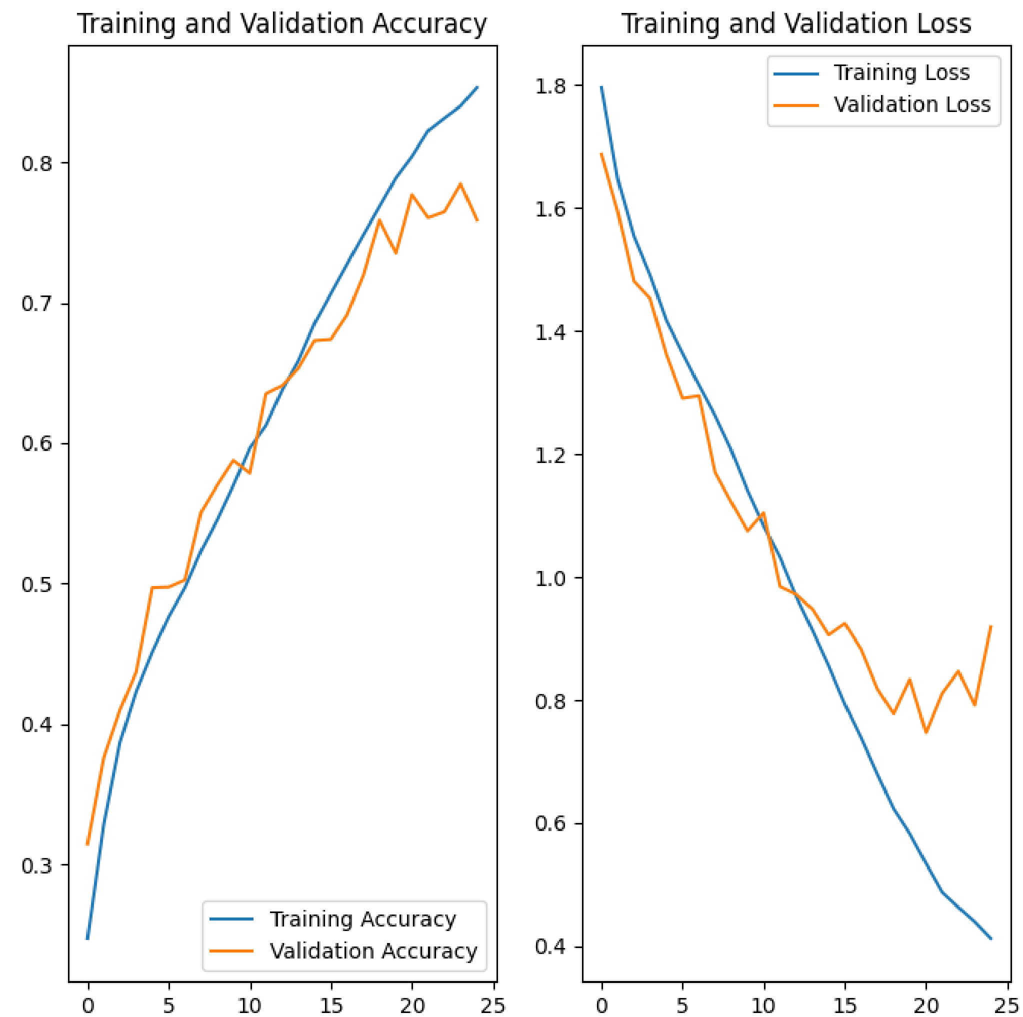
In order to train this model we partitioned the complete dataset into smaller testing, training, and validation datasets. The training dataset contains 80% of the data while the testing and validation sets each contain 10% of the data. Using the training data we trained the model over 25 epochs, each with a batch size of 62.

```
model = models.Sequential([
    resize_and_rescale,
    layers.Conv2D(32, kernel_size = (3,3), activation='relu', input_shape=input_shape),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(64,  kernel_size = (3,3), activation='relu'),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(64,  kernel_size = (3,3), activation='relu'),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(64, (3, 3), activation='relu'),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(64, (3, 3), activation='relu'),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(64, (3, 3), activation='relu'),
    layers.MaxPooling2D((2, 2)),
    layers.Flatten(),
    layers.Dense(64, activation='relu'),
    # Output layer
    layers.Dense(n_classes, activation='softmax'),
])
```

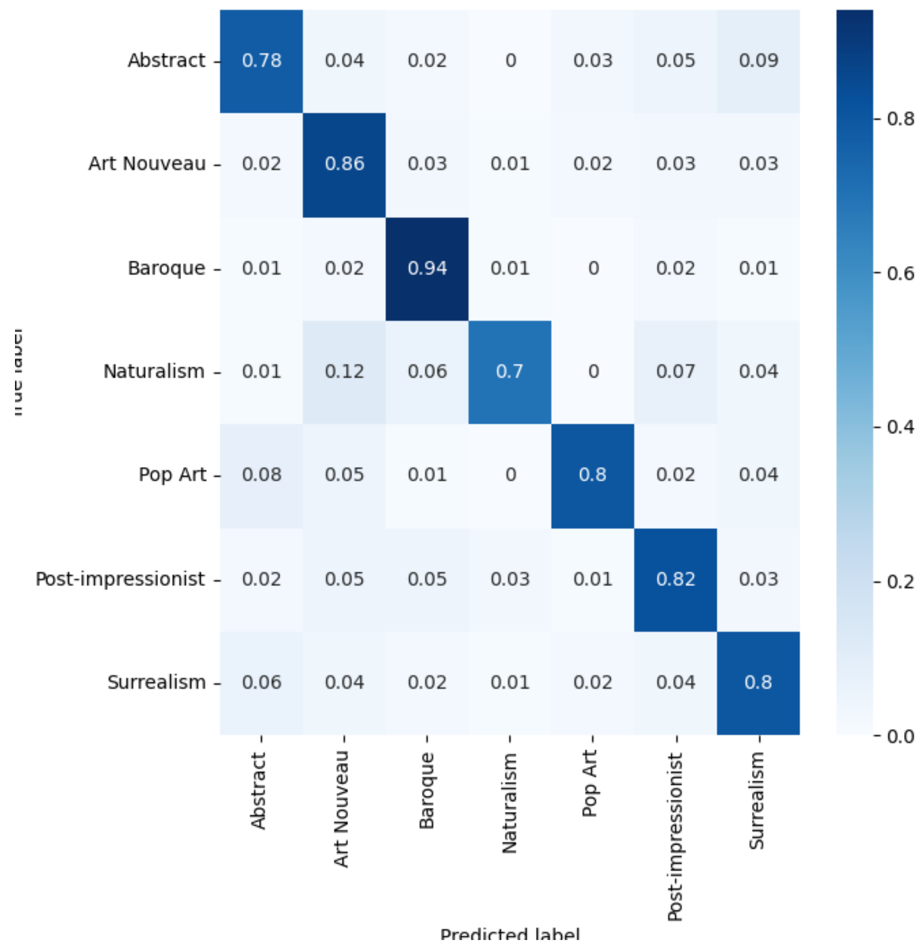
The Actual code creating the Neural Network Architecture

Results

After training and testing the model we generated a training accuracy of 86.12% and a testing accuracy of 81.77%. Multiple tests using the same architecture generated similar results with test accuracy ranging from 80-84% and training accuracy surpassing 90%. The variation between the test and training accuracy alongside the validation accuracy graph suggest that the model's accuracy peaked around the 20th epoch, though the model is not overfit. We can also tell, through the use of a confusion matrix, that the model was particularly accurate when identifying specific genres of art. In particular, the model had 94% accuracy when identifying Baroque artwork and it rarely confused baroque artworks for other genres. Unsurprisingly, the model had the most trouble with identifying Naturalist artworks, most likely because of its small dataset compared to the other genres.



Graph Depicting Training/Validation Loss and Accuracy Over Epochs



Confusion Matrix of the Model and Test Data

Conclusion

In this paper, we proposed the use of convolutional neural networks for artwork identification. We first collected thousands of digital images of artworks using public databases and python scripts. Then we fed the images into a CNN and narrowed our classifying datasets as we tuned the model. As a result, we were able to create a fairly reliable model that could accurately differentiate between seven artistic genres. Though the accuracy (81.77%) of this model is not perfect, it does show potential for future research applications. With access to better technology, more data, and time, I am certain that this model could be improved enough to surpass 90% accuracy with far more classifying genres.

Annotated bibliography

Brownlee, Jason. “How Do Convolutional Layers Work in Deep Learning Neural Networks?” *Machine Learning Mastery*, 16 Apr. 2020, https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/.

This article mainly talks about convolutional neural networks and in particular how layers interact with nodes. Gives detailed and easy to understand visual representations of convoluted networks and in particular the filters and feature maps that separate them from other types of neural networks.

Cetinic, Eva, et al. “Fine-Tuning Convolutional Neural Networks for Fine Art Classification.” *Expert Systems with Applications*, Pergamon, 12 July 2018, https://www.sciencedirect.com/science/article/pii/S0957417418304421?casa_token=14NnxXb5BnEAAAAA%3A4bDRyjm5Su3buD9YZHk_QFtOA4tQJhRGDxu0J074t948J1fA1OOCTekyTsgr0fUtv61ihMZQfQ.

This paper seeks to answer similar questions to our research question. They use convolutional neural networks for a series of applications including using them to identify art subject, artist, and the individual artwork.

Geron, Aurelien. *Hands-on Machine Learning with Scikit-Learn, Keras and Tensorflow Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly, 2019.

This book is my main source of knowledge when it comes to machine learning and neural networks in particular. It gives informative explanations of virtually every type of neural network as well as code and real world applications. This book is vital in explaining these topics as well as creating the convolutional neural network at the heart of my capstone project.

Hinton, Geoffrey E. "To Recognize Shapes, First Learn to Generate Images." *Progress in Brain Research*, 2007, pp. 535–547., https://doi.org/10.1016/s0079-6123(06)65034-6.

Highlighted in the introduction of this paper, this publication marks the beginning of deep learning as a whole. Up until this publication it was widely considered to be impossible to implement such an accurate and robust model and it marks the beginning of data science and the concept of deep learning. Not only is this paper pivotal in the field, but the methodology and code is made easy to understand and is extremely informative for those just starting on the path to deep learning.

Hong, Yiyu, and Jongweon Kim. *Research India Publications Was Established in 1998 and Now Today We ...Yiyu Hong1*. International Journal of Applied Engineering Research, 4 Nov. 2017, https://www.ripublication.com/ijaer17/ijaerv12n4_17.pdf

Rather than research the ability for CNNs to identify art genres or artists, this paper seeks to use Convolutional Neural networks to identify the actual artwork. This is done by simulating a myriad of transformations and perspectives.

Saha, Sumit. “A Comprehensive Guide to Convolutional Neural Networks - the eli5 Way.” *Medium*, Towards Data Science, 17 Dec. 2018, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

This article focuses on convolutional neural networks and gives a very easy to understand description of nodes and how they function. In particular, this article focuses heavily on the convolutional layer aspect of the model and gives very easy to understand visuals that explain concepts like feature maps, stride, dot products, and pooling.

Udyavar, Nehal. “A Beginner's Guide to Neural Networks: Part One.” *Medium*, Towards Data Science, 10 Apr. 2017, https://towardsdatascience.com/a-beginners-guide-to-neural-networks-b6be0d442fa4.

This article, as well as its second part gives an easy to understand overview of neural networks and how they function. It also has informative visuals that explain the actual math that goes behind neural networks that is often overlooked when discussing the topic.

Editor. "Image Recognition with Deep Neural Networks and Its Use Cases." *AltexSoft*, AltexSoft, 13 Jan. 2020, https://www.altexsoft.com/blog/image-recognition-neural-networks-use-cases/.

This article focuses on Smartify and the science behind its model. It explains neural networks and in particular the deep neural networks used within the application that allows it to identify and classify artwork.