38-616 Final Project: Artwork Classification

Tianhan Ling, Jie Sing Yoo, Natalie Pham

April 2023

Abstract

The classification of fine art paintings based on their genres is a challenging problem due to the high degree of variation in style and composition. In this project, we developed a deep learning model using Convolutional Neural Networks (CNN) to classify paintings into six different genres, including Baroque, Impressionism, Post-Impressionism, Realism, Romanticism, and Symbolism. We experimented with several CNN models, evaluated their performance metrics including accuracy, weighted average F1-score, and confusion matrix, and explored various techniques to improve the model's performance. Our results indicate that the ResNet50 model with focal loss has the highest weighted average F1-score of 0.65, making it the best performing model for classifying fine art painting genres. However, there are some limitations, particularly in distinguishing between Impressionism and Post-Impressionism genres, which requires further improvement.

1 Introduction

A connoisseur is capable of identifying the genre and artist of an art piece by inspecting various properties of the art. However, human judgment is often subjected to errors, and visually inspecting all the small details in a fine art is a tedious process and can take a long time. Given the current development of machine learning and deep learning has matured, this project is an attempt to identify the genre of an artwork by feeding it into a neural network model.

The dataset for this project consists of 6669 images of fine art paintings by 38 influential artists. This project aims to classify the images into 20 different genres by differentiating the colors and geometric patterns in the images.

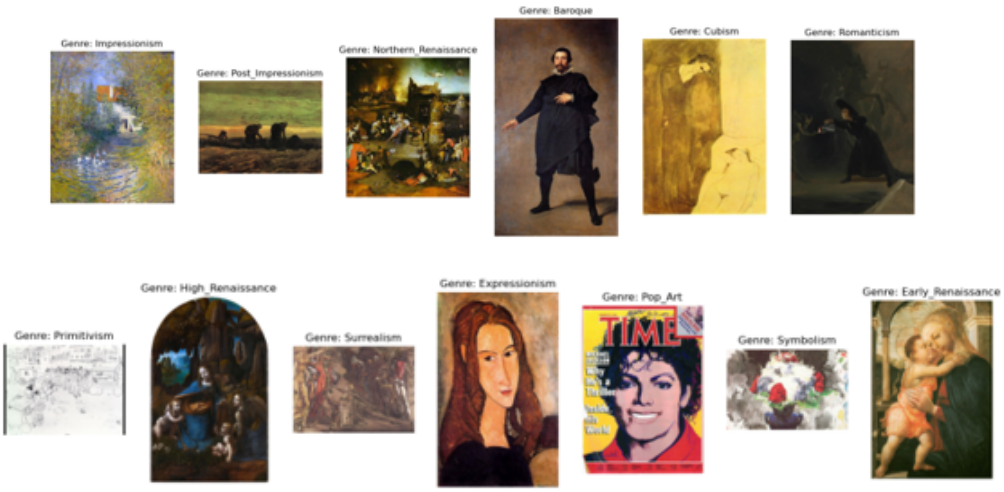

Figure 1: Sample data from each classes.

2 Methods

We utilized pre-trained models including VGG16 and ResNet50, in which the top layer was removed and two new fully connected layers were added at the end. These pre-trained models were previously trained on the ImageNet dataset, containing millions of images, allowing us to benefit from their learned features.

By discarding the top layer of the pre-trained models, we extracted the output from the remaining layers as input to our newly added fully connected layers. Our added layers included a 512-unit fully connected layer with a Rectified Linear Unit (ReLU) activation function, followed by batch normalization for input normalization, another ReLU activation function, and a fully connected layer with 16 units. Lastly, a dense layer with a softmax activation function was employed to classify the input image into one of the 13 fine art genres.

The diagrams below illustrate the structure of the pre-trained model in our project.

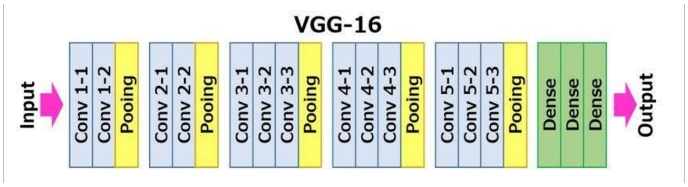

Figure 2: Structure of VGG16.

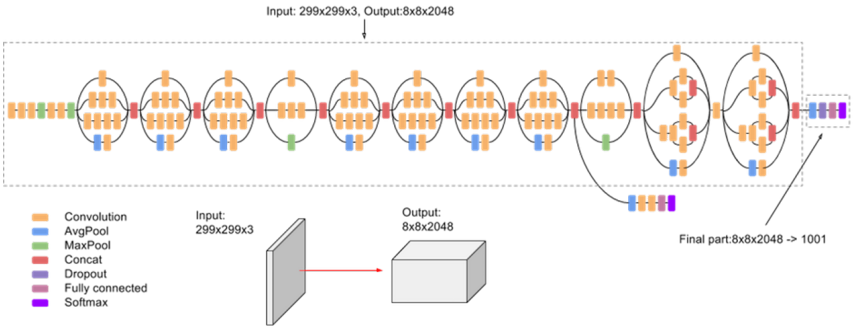

Figure 3: Structure of InceptionV3.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 4: Structure of ResNet 18/34/50/101/152.

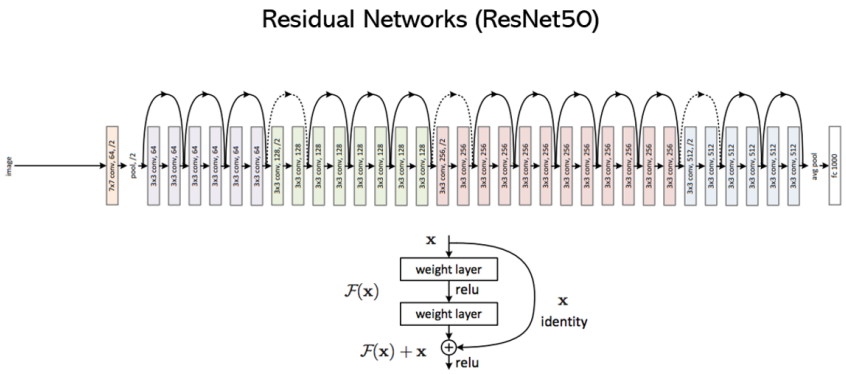

Figure 5: Structure of ResNet50.

3 Experiment and Results

3.1 Data Preprocessing

In the exploratory data analysis phase, we addressed the issue of imbalanced data by filtering out genres that contained fewer than 150 paintings. To enhance the robustness of our model, we implemented several data augmentation techniques such as rescaling the images to a size of (224 x 224 x 3), normalization by dividing the maximum pixel value, and applying horizontal and vertical flips. These measures were implemented to achieve the objective of accurately classifying images into 13 distinct genres.

Figure 6: Sample of data augmentation.

3.2 Model

In our model experiments, we separated the images into train and validation sets at a ratio of 80:20. We used the Categorical Cross-entropy Loss function and Adam optimizer to train our models. We set the number of training epochs to 50 and applied EarlyStopping to optimize training time. Specifically, we used a patience value of 20 epochs to stop training if the validation loss did not improve. Furthermore, we implemented a learning rate scheduler to reduce the learning rate by a factor of 0.1 if the validation loss did not improve after 5 epochs.

To achieve optimal performance, we experimented with several pre-trained models, including VGG16, InceptionV3, ResNet101, and ResNet50.

3.3 Results

To assess the effectiveness of each Convolutional Neural Network (CNN) model, we employed multiple metrics, such as the confusion matrix, accuracy, weight f1-score, and loss and accuracy plots. The accuracy metric helped

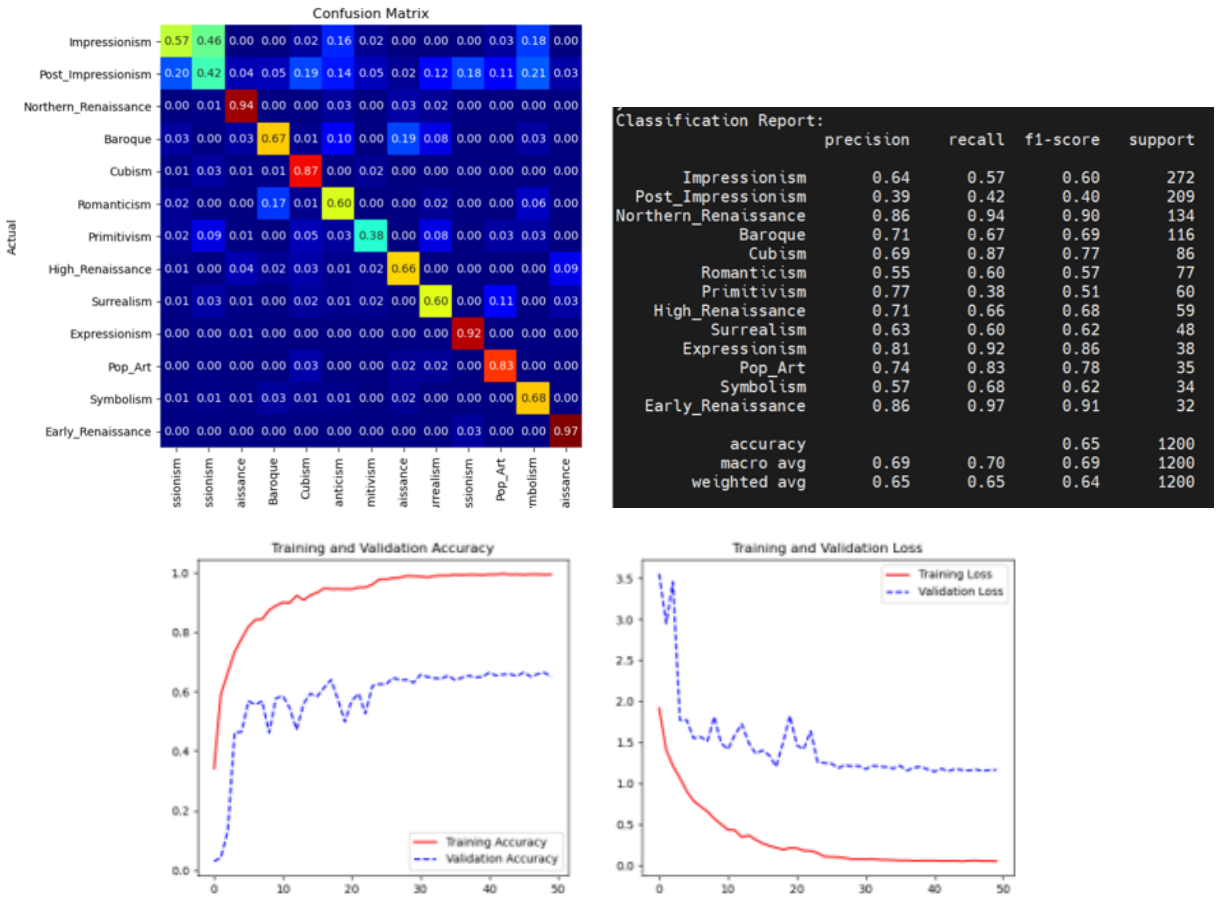

Figure 7: Metrics for ResNet50.

to measure the percentage of images that were correctly classified by the model. The loss and accuracy plots were used to track the performance of the model during training and validation phases.

Moreover, we utilized the confusion matrix to visualize the distribution of correctly and incorrectly classified genres and detect any misclassification patterns. We also used the weighted F1-score metric, which considers both precision and recall, to evaluate the performance of each model. These metrics provided us with a comprehensive evaluation of the performance of each CNN model and allowed us to compare their effectiveness in classifying the fine art painting genres.

Models	Accuracy	Macro average F1 score	Weighted average F1 score
VGG16	0.56	0.60	0.56
InceptionV3	0.42	0.39	0.41
ResNet101	0.65	0.66	0.60
ResNet50	0.65	0.69	0.64
ResNet50 with focal loss	0.66	0.69	0.65

Table 1: Result summary

Table 1 reveals that VGG16, InceptionV3, and ResNet101 exhibit inferior performance compared to ResNet50. This may be attributed to the fact that the validation loss and accuracy plots for the former models show greater fluctuations in comparison to ResNet50, likely due to the limited size of the dataset. Furthermore, all three models exhibit high misclassification rates for Impressionism and Post Impressionism genres, with a tendency to mistake Impressionism for Symbolism or Romanticism. Conversely, ResNet50 significantly decreases the likelihood of such misclassifications.

ResNet50 has generally performed well in predicting the genres, however, there are still some misclassifications

observed between Impressionism and Post Impressionism. This issue can be attributed to the high similarity between these two genres, which sometimes makes it difficult even for humans to distinguish between them.

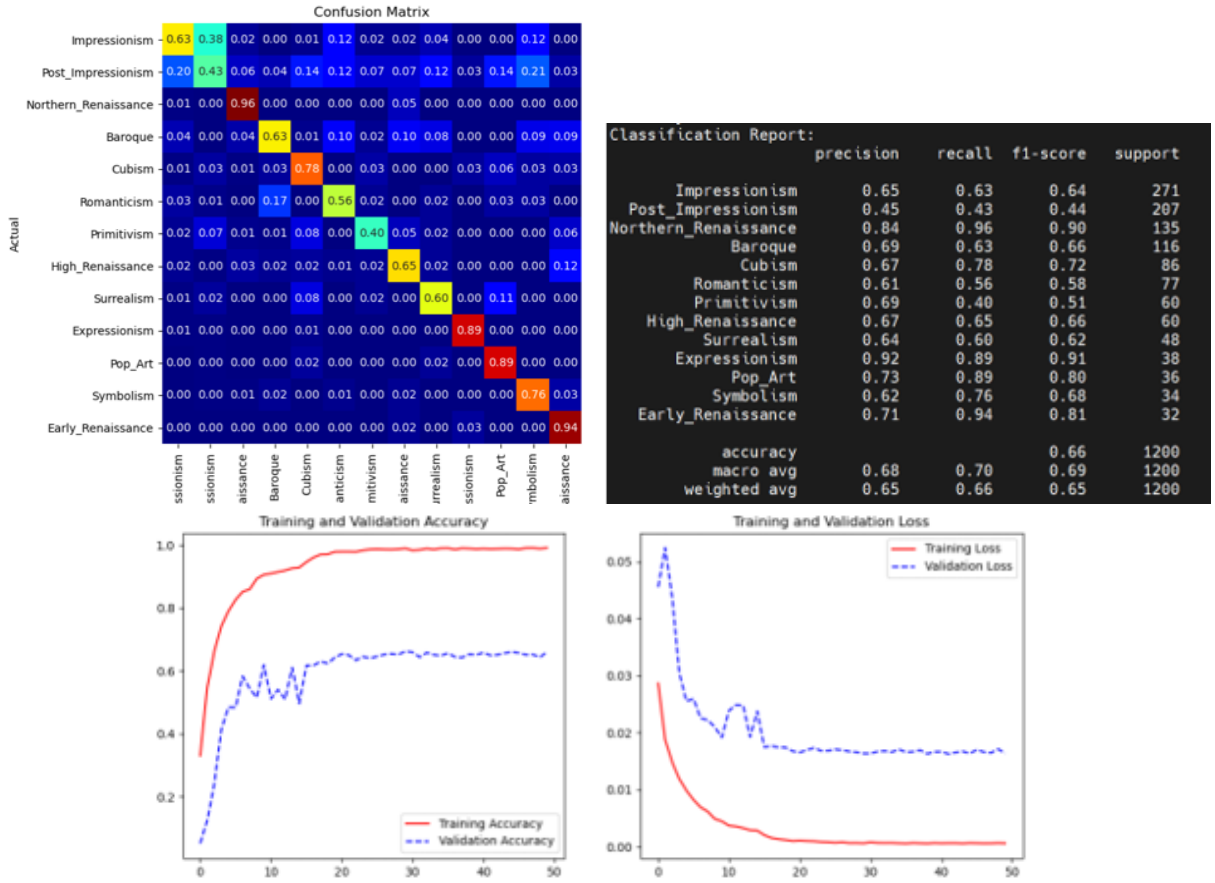

Figure 8: Metrics for ResNet50 with focal loss.

Observed from the confusion matrix of the ResNet50 model, the model was more likely to improve the accuracy of easy-to-classify classes such as Early Renaissanse and Pop Art. Hence we used the down weighting technique to reduces the influence of easy examples on the loss function, resulting in more attention being paid to hard examples. This technique can be implemented by adding a modulating factor to the Cross-Entropy loss. The focal loss function defined as follow:

$$FL(p_t) = -\sum_{i=1}^{N} \alpha(1-p_i)^{\gamma} * log(p_i)$$

where N is the number of training samples, p_i is the predicted probability of the true class of sample i, α and γ are modulating factors.
Compared to cross-entropy loss, the focal loss allows the model to train with part of the loss focus on hard-to-classify classes (such as Impressionism, Post Impressionism and Primitivism). In our project, we trained the CNN with ResNet50 layers with the focal loss used $\alpha = 0.25$ and $\gamma = 2$. For improvement, we can tune α and γ to get a model with better accuracy.

4 Discussion

We attempted to improve the performance of our model by using data augmentation techniques such as zooming, rotation, and shear range, but the overall weighted average f1 score did not improve, particularly for the post-impressionism genre.

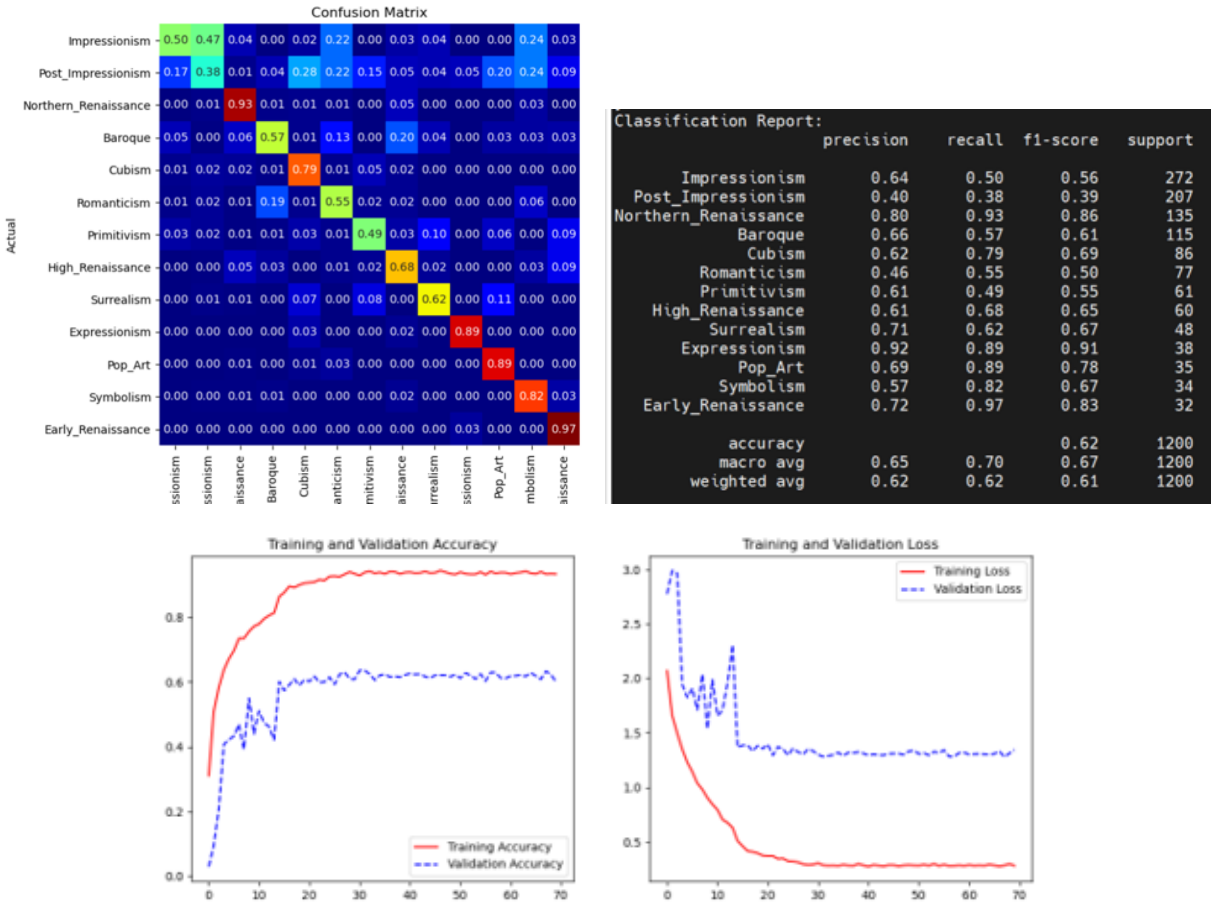

Figure 9: Metrics for ResNet50 with data augmentation.

We experimented with a binary model to distinguish between impressionism and post-impressionism, but the confusion matrix showed some misclassifications for post-impressionism.

We also tried to combine impressionism and post-impressionism as one genre, but this resulted in further imbalance due to the fact that these two genres comprise almost 30% of the dataset. Consequently, the performance of the model decreased after combining these two genres.

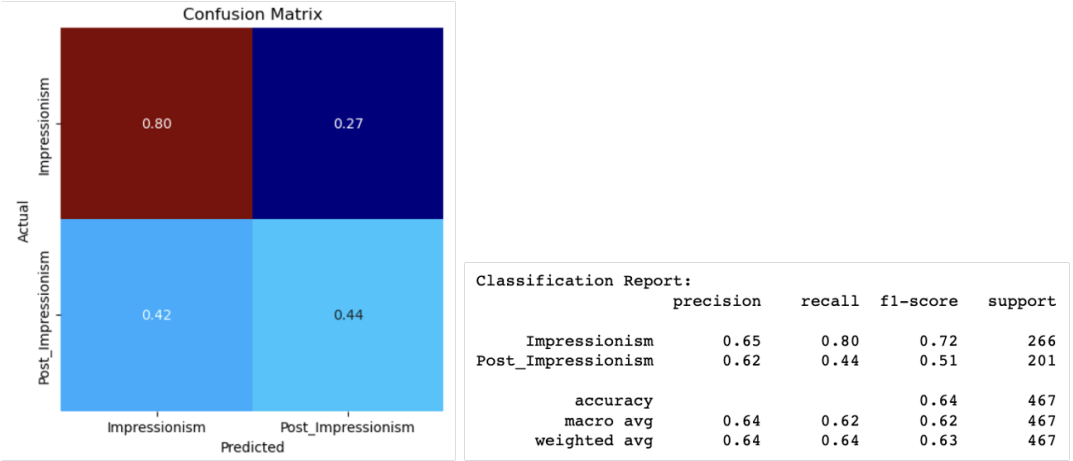

Figure 10: Metrics for ResNet50 classifier for Impressionism and Post Impressionism.

5 Conclusion

In a nutshell, we conducted an in-depth analysis of several convolutional neural network models for the classification of fine art painting genres. Through our evaluation, we found that ResNet50 outperformed VGG16, InceptionV3, and ResNet101 models in predicting the genres of fine art paintings. However, misclassification between Impressionism and Post Impressionism remained an issue due to the high similarity between these two genres. Despite our efforts to improve the performance of our model through data augmentation and a binary classification approach, we were unable to improve the overall weighted average F1 score significantly. In addition, combining Impressionism and Post Impressionism as one genre resulted in further imbalance and a decrease in model performance. Our analysis highlights the challenges of classifying similar fine art painting genres and the importance of continued research to improve classification accuracy in this field.

References

[1] Simonyan, K. and Zisserman, A. (2015) "Very Deep Convolutional Networks for Large-Scale Image Recognition", The 3rd International Conference on Learning Representations (ICLR2015). https://arxiv.org/abs/1409.1556

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition". https://arxiv.org/abs/1512.03385

[3] Popular networks, "ResNet (34, 50, 101): Residual CNNs for Image Classification Tasks". https://neurohive.io/en/popular-networks/resnet/

[4] Roshan Nayak, "Focal Loss: A better alternative for Cross-Entropy". https://towardsdatascience.com/focal-loss-a-better-alternative-for-cross-entropy-1d073d92d075

[5] Sean Benhur, "A friendly introduction to Siamese Networks". https://towardsdatascience.com/a-friendly-introduction-to-siamese-networks-85ab17522942