Test Report

Zhizhen Li

This is the task project for GSoC 2025 application for the ArtExtract Project.

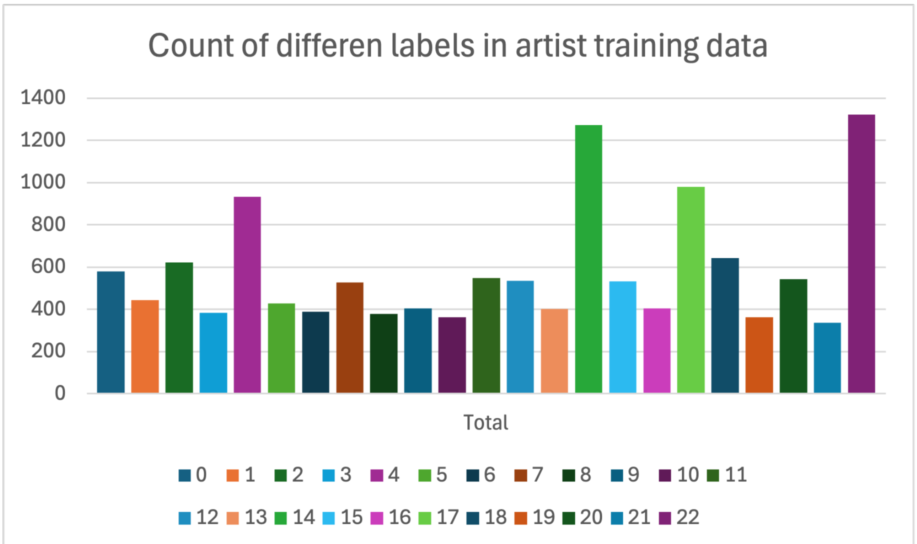
Introduction

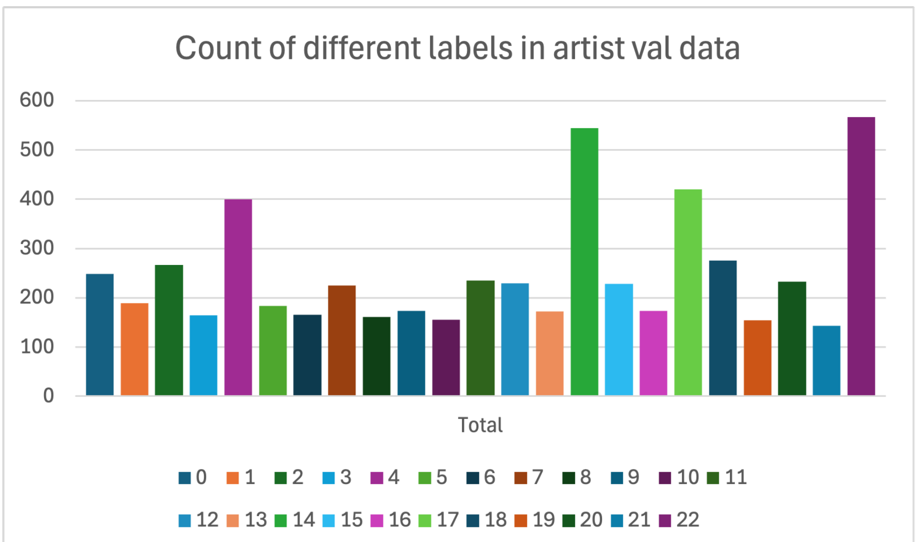
Nowadays, many edge-cutting techniques are applied in the art classification filed. Among them, the deep learning skill is considerably a effective way to finish the job. CNN model and its morphs, which can take pixel data, are used as key tools and perform well. (Zhao et al., 2021) This project is trying to raise a new methond to classify the artworks with different labels such as artists, genres and styles.

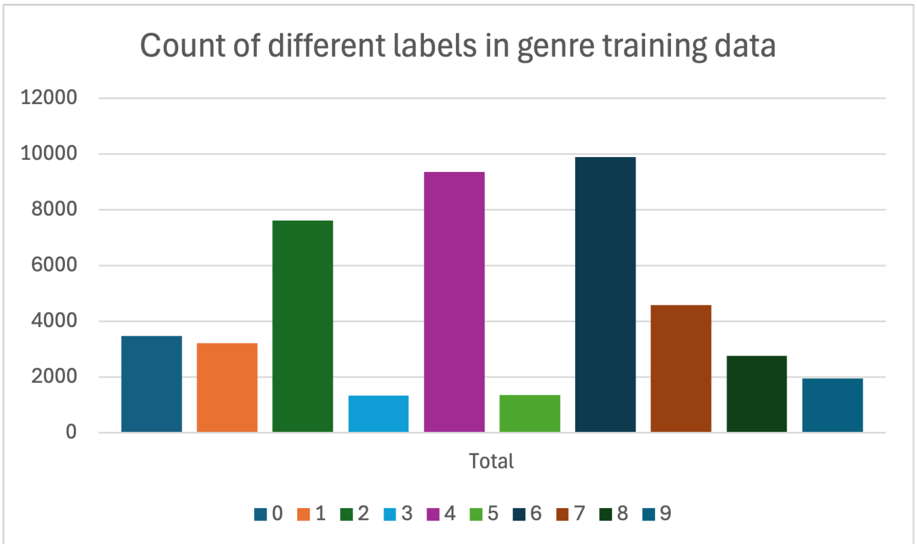
The aim of the project is to predict the artist, genre and style labels based on the pixel input. Moreover, the project tried to tell the outliers from the artworks data which may help people understand the variaty and stylistic chages in the given group of works.

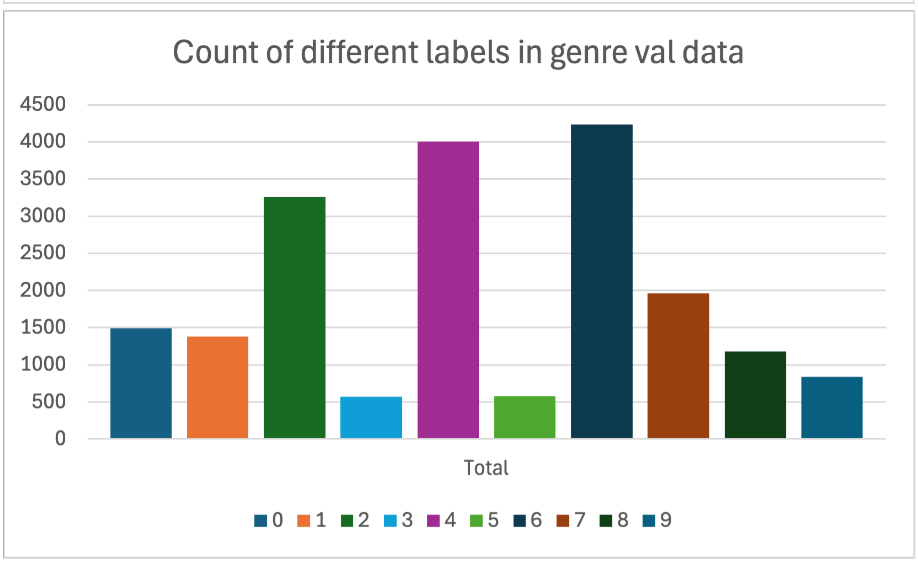
Data Overview

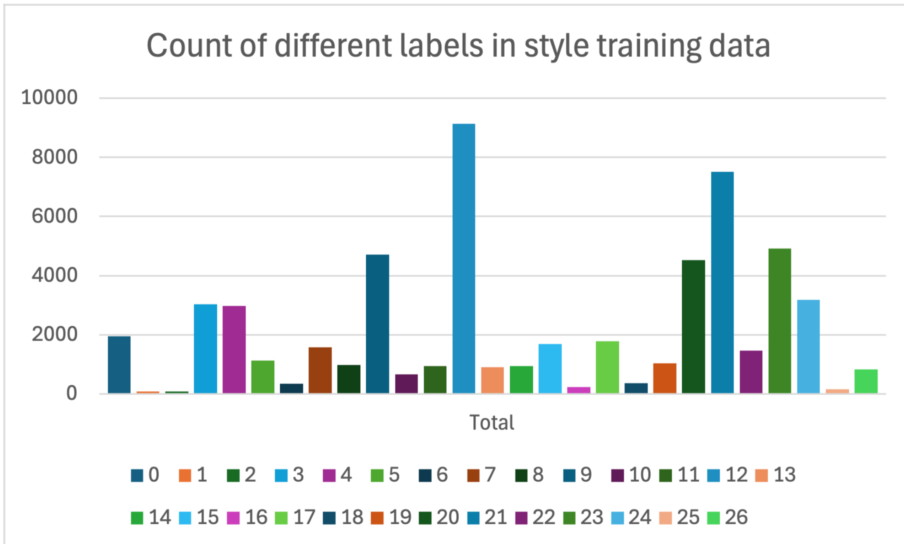
Wikiart dataset is uesd as the original data in the project. There are 13346 marks of artist data, 45503 marks of genre data and 57025 marks of style data for training and 5706 artist marks, 19492 genre marks and 24421 style marks for testing. The struture of them are shown below.

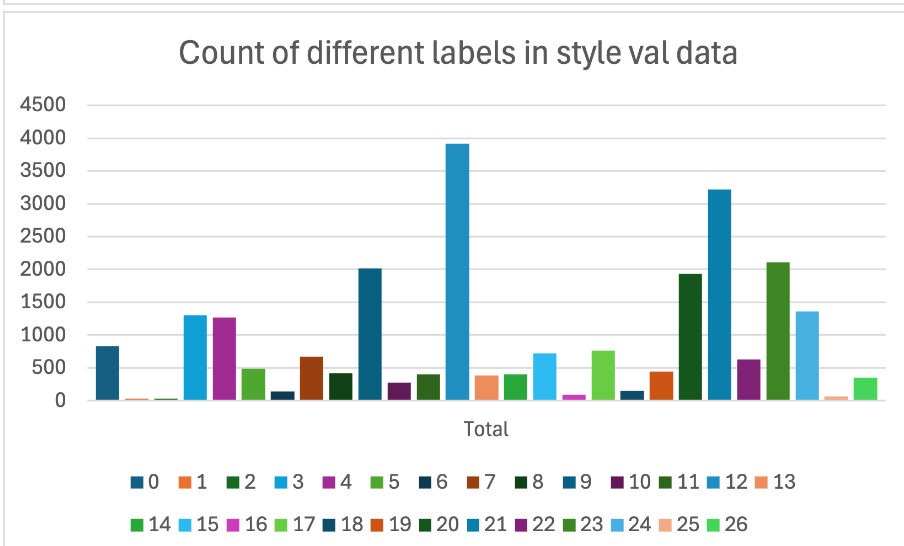

Count of differen labels in artist training data
1400
1200
1000
800
600
400
200
0
Total
0 1 2 3 4 5 6 7 8 9 10 11
12 13 14 15 16 17 18 19 20 21 22


Count of different labels in artist val data
600
500
400
300
200
100
0
Total
0 1 2 3 4 5 6 7 8 9 10 11
12 13 14 15 16 17 18 19 20 21 22


Count of different labels in genre training data
12000
10000
8000
6000
4000
2000
0
Total
0 1 2 3 4 5 6 7 8 9


Count of different labels in genre val data
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Total
0 1 2 3 4 5 6 7 8 9





It is clear that the distribution of the data is not even. For example, the label 3 and 5 in genre dataset own far less data than other labels. Things get worse in style data. 1, 2, 16, 25 labels owns too little data to be fully modeled.

Methods and Models

Lecoutre et al. (2017) pointed out that Alexnet and pre-trained Resnet model perform well in artword classification.

First, I numericalize the labels of all categories by **LabelEncoder** from the **scikit-learn** library to adapt the input requirements of the neural network model. In addition, in order to improve the generalization ability of the model, the following Data Augmentation (DA) methods are used. This process includes **Resize**, **Random Horizontal Flip**, **Random Rotation**, **Random Resized Crop** and **Normalization**.

AlexNet uses **five Convolutional Layers** followed by Fully Connected Layers, with the following structure:

- **Convolutional Layer 1 (Conv1)**: (11 ×11), step size 4, ReLU activation;
- **Pooling Layer 1 (Max Pooling 1)**: (3 ×3), step size 2;
- **Convolution Layer 2 (Conv2)**: (5 ×5), ReLU activation;
- **Pooling Layer 2 (Max Pooling 2)**: (3 ×3), Step 2;
- **Convolutional Layers 3-5 (Conv3-Conv5)**: (3 ×3), ReLU activation;
- **Pooling Layer 3 (Max Pooling 3)**: (3 ×3), step size 2;
- **Fully Connected Layer**: two 4096-dimensional hidden layers + output layer.

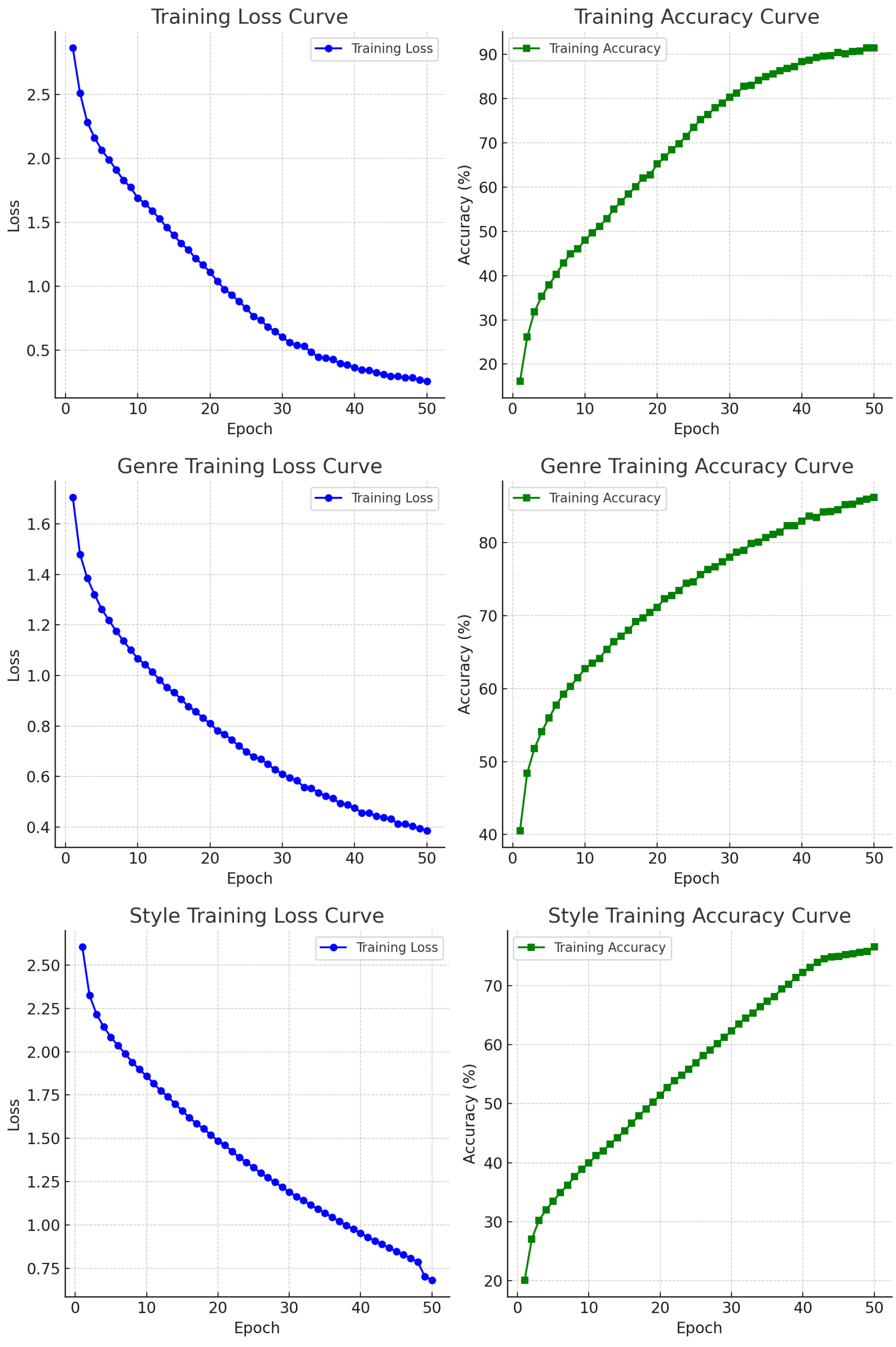
ResNet-50 adopts **Residual Learning**, the core idea of which is to retain gradient information through **Skip Connections**, so as to effectively train deep neural networks. In this study, we use **ResNet-50 pre-trained model** and re-train the **last 20 layers** (Fine-tuning).

Finally, this study uses **Soft Voting** for model integration. Specifically, AlexNet and ResNet-50 generate separate predictive probability distributions on the test set, and the final classification result is taken as the average** of the predictive probabilities of the **two models, thus improving the overall classification performance.

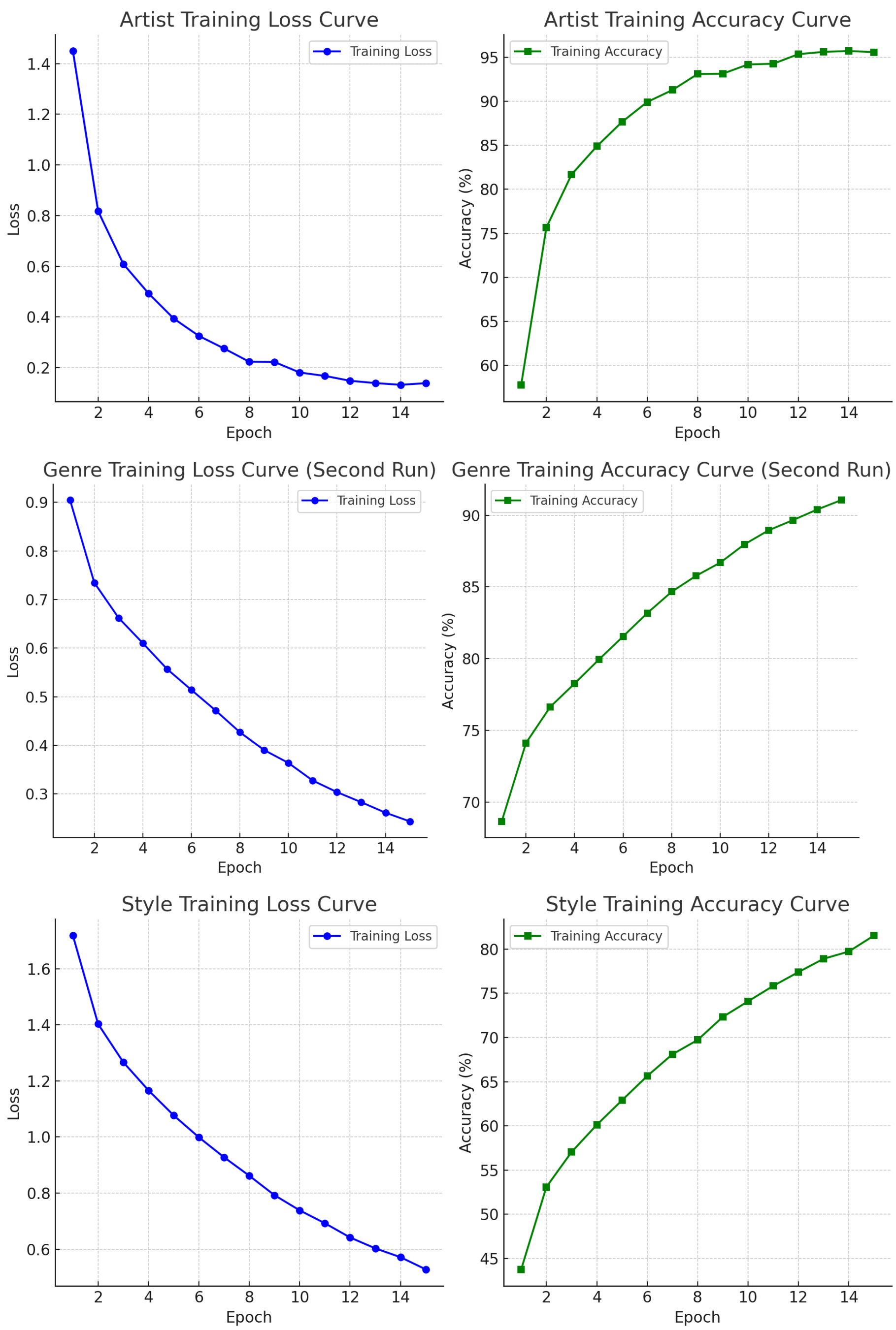
Model Training and Evaluation

As shown in the figures below, the Resnet models converges more quickly. However, Alexnet model has better generalization ability, as Alexnet model got 89.28%, 86.03% and 79.83% on the test data while Resnet model got 94.94%, 75.31% and 56.56%.

Alexnet model's performance

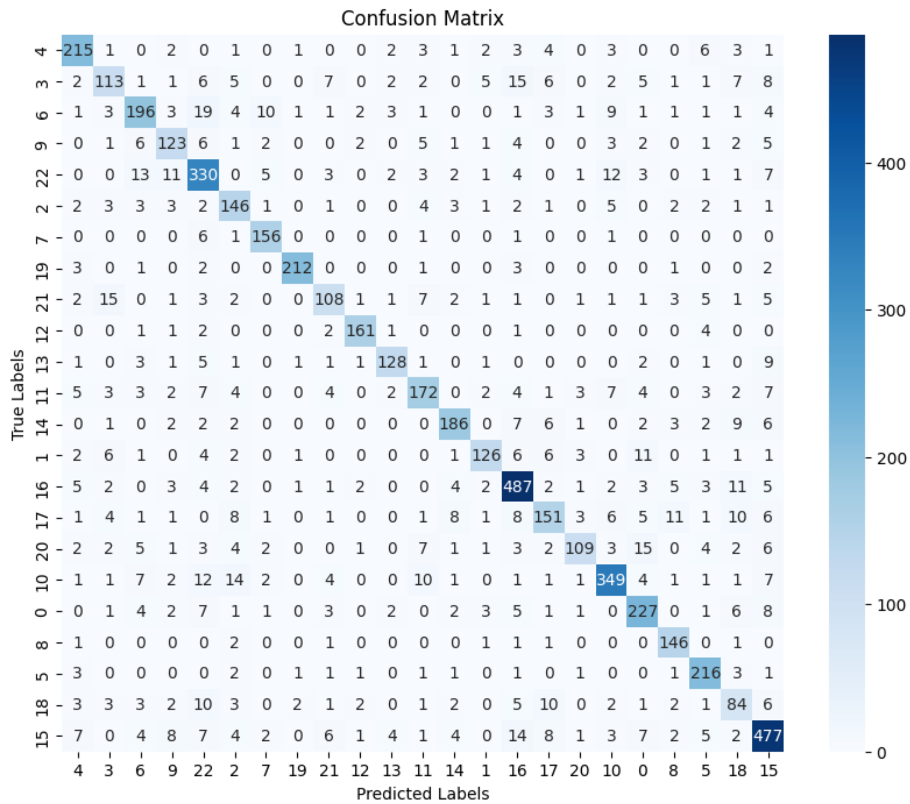


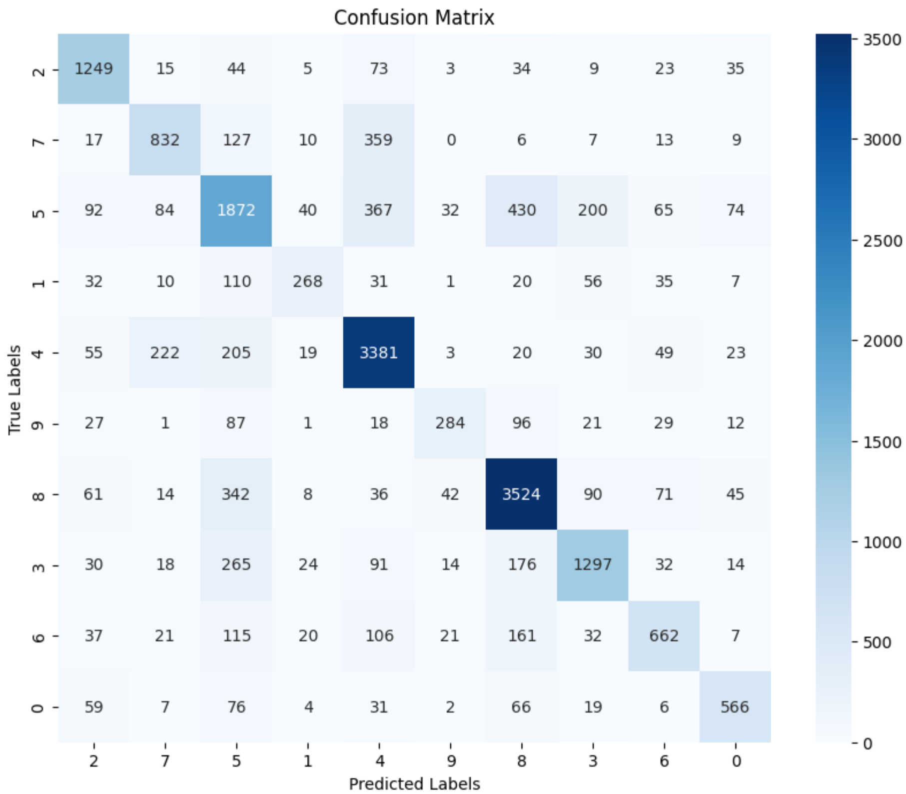
Resnet model's performance



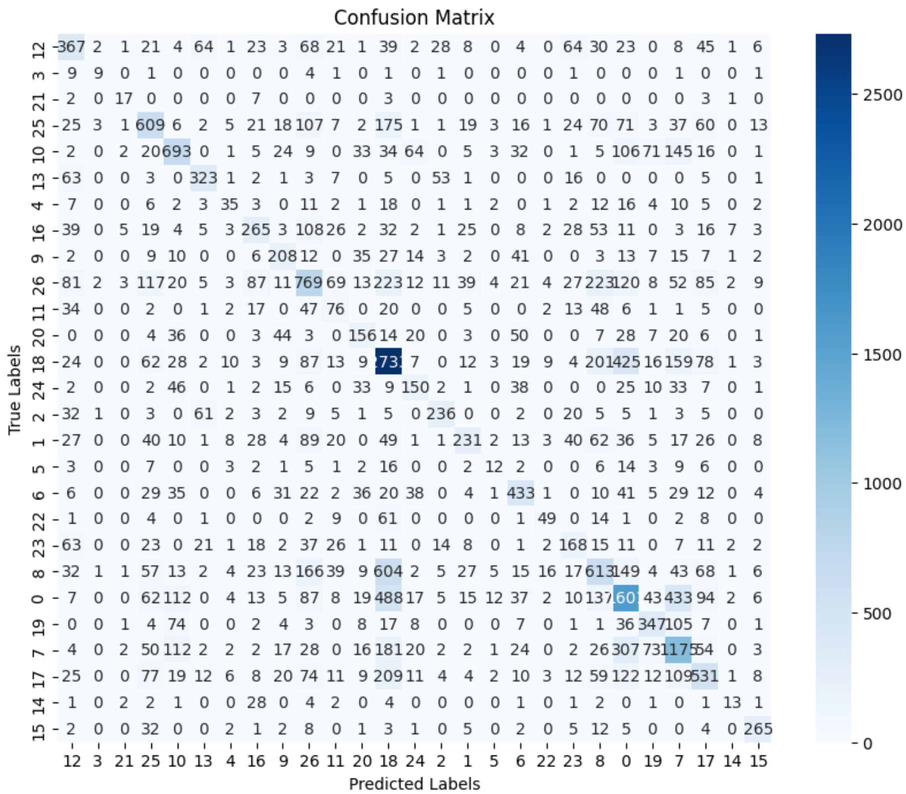
In order to taking the advantages of both models, soft voting strategy has been applied. Based on the perfomance of the models on the test data, they got different weight when voting. After stacking the two model, the accuracy rate increased considerably.

The stacked model's confusion matrixs are shown below.





	2	7	5	1	4	9	8	3	6	0
2	1249	15	44	5	73	3	34	9	23	35
7	17	832	127	10	359	0	6	7	13	9
5	92	84	1872	40	367	32	430	200	65	74
1	32	10	110	268	31	1	20	56	35	7
4	55	222	205	19	3381	3	20	30	49	23
9	27	1	87	1	18	284	96	21	29	12
8	61	14	342	8	36	42	3524	90	71	45
3	30	18	265	24	91	14	176	1297	32	14
6	37	21	115	20	106	21	161	32	662	7
0	59	7	76	4	31	2	66	19	6	566

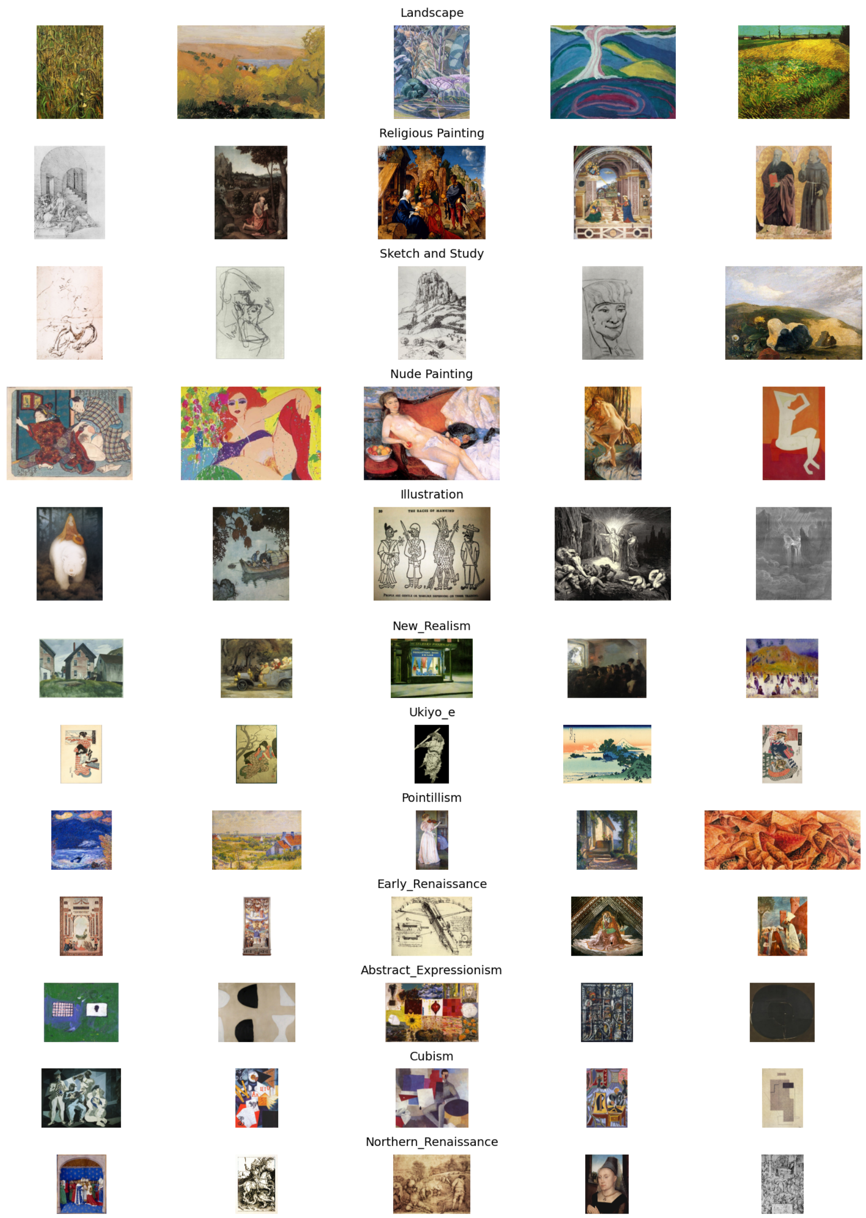


Results and Discussion

The final accuracy of the stacked model is 80.79% in artist data, 71.41% in genre data and 49.50% in style data. It is clear that the style classification is not ideal enough. There are two main reasons of the result. Firstly, the style of a given picture is debatable. It is fear to take a picture as different styles for different reasons. Thus, we may adjudt the evaluation methodology into if the style label in the top 3 propably styles. After adjusting the metho, the accuracy will increase to 80.42%.

It can be telled clearly that beside the lack of training data, there exist some labels which can be confused by other labels, causing the limitation on the accuracy. To find out why, we may analyse them in this part.

It can be noticed that several of the often confused classifications (e.g., 3-5-8) show a large number of paintings dominated by figures, which in a way creates an obstacle for the computer's recognition. In addition, we can also find that the existing classifiers are less effective in learning LANDSCAPE as a LABEL. By observing the given training samples, it can be found that there is a great variability within the LANDSCAPE samples, and the classification criteria are rather rough. Modernist and abstract paintings have large differences in visual characteristics from classical paintings, which leads to difficulties in computer recognition.



It can be noticed that my model is good at distinguishing between figure paintings and non-figure paintings. Since both ukiyo-e and neorealism contain a lot of portraits, the computer tends to classify them into one category. It is worth noting, however, that this confusion is not meaningless. For example, the computer's confusion of Impressionism with ukiyo-e reflects the profound influence of ukiyo-e art on Impressionist art. (Yonemura, 1996) Neorealism was greatly influenced by the paintings of Impressionism, but there is a lack of exploration of the relationship between Neorealism and classical Japanese aesthetics in both the literary and artistic worlds. An examination of the philosophical ideas of neorealist painters such as Edward Hopper reveals that their emphasis on the isolation and loneliness of the individual in modern life is deeply compatible with traditional Japanese aesthetic thought. (Koob, 2004) This may provide ideas for future art studies.

Reference

Zhao, W., Zhou, D., Qiu, X., & Jiang, W. (2021). Compare the performance of the models in art classification. Plos one, 16(3), e0248414.

Lecoutre, A., Negrevergne, B., & Yger, F. (2017, November). Recognizing art style automatically in painting with deep learning. In Asian conference on machine learning (pp. 327-342). PMLR.

Yonemura, J. K. (1996). The Influence of Ukiyo-e on Impressionism and Post-impressionism. California State University, Dominguez Hills.

Koob, P. N. (2004). States of being: Edward Hopper and symbolist aesthetics. American Art, 18(3), 52-77.