Milestone Report: Artistic Style Recognition using Deep and Shallow Neural Networks

Zhu Fangyu
2023011410
Class 32
xx@mails.tsinghua.edu.cn

Lu Yingxi
2023011435
Class 31
lu-yx23@mails.tsinghua.edu.cn

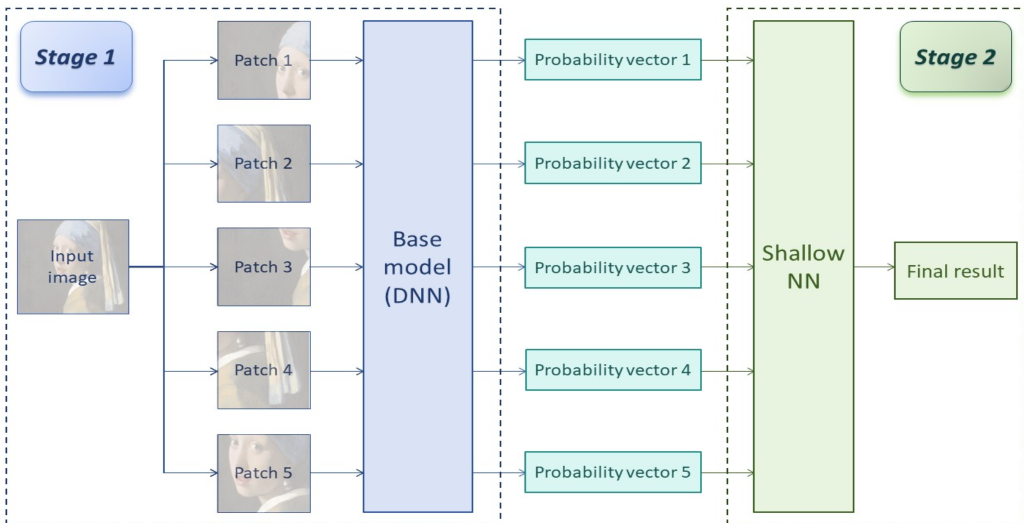

Figure 1. Main pipeline

1. Introduction

In this project, we aim to develop a model capable of classifying paintings according to their artistic style. Following the approach proposed in [2], our model consists of two components: a deep neural network (DNN) classifier and a shallow neural network (SNN).

At the first stage, each input image is divided into five distinct patches: top-left, top-right, bottom-left, bottom-right, and center (as illustrated in Figure 2). The DNN independently classifies each of these five patches. In the second stage, the SNN functions as a decision-maker: it takes the probability distributions generated by the DNN for the five patches as input and produces the final classification result. According to [2], this two-stage architecture helps reduce error accumulation by training the two stages separately.

Our project consists of two main parts:

- **Architecture implementation and validation:** Implement the architecture mentioned in [2] (since there is no open-source code repo), proof its effectiveness by ablation study.
- **Adaptive refinement and optimization:** To further explore and gain insights, we investigate how certain design choices affect model accuracy, including:
  - **Selection of the fifth patch**: Instead of always choosing the center region as the fifth patch, we experiment with alternative strategies such as using a downsampled version of the entire image or extracting a region that contains the main semantic content.
  - **Architectural variations**: Beyond the simple five-patch individual classification scheme, we explore more complex hierarchical architectures. We evaluate their performance and analyze the results to understand how these modifications influence classification accuracy.

2. Problem Statement

In this project, we address the problem of classifying paintings according to their artistic styles. Given the vast number

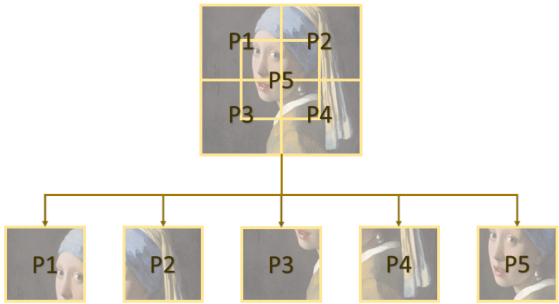

Figure 2. Patch division

of possible artistic styles, we focus on five major modern styles: **Cubism**, **Impressionism**, **Expressionism**, **Realism**, and **Abstract**.

Dataset. We utilize the Painter by Numbers dataset [3] as the primary source of paintings for our study. (TODO: Describe the preprocessing steps, including how multiple style labels are consolidated into the five selected categories; explain the preparation of input data for the shallow neural network.)

Evaluation. Classification accuracy is adopted as the primary evaluation metric to assess model performance.

Baseline. In this study, we adopt the ArtNet model [1] as our baseline for comparative analysis. ArtNet is specifically designed to classify paintings into five distinct modern art styles: Cubism, Impressionism, Expressionism, Realism, and Abstract. A key feature of ArtNet is its preprocessing strategy, where each input image is first padded to a uniform size and then divided into five patches. Both the full image and its patches are included in the training set, enabling the model to learn from global composition as well as local texture details such as brush strokes and color tones.
This dual-level learning approach aligns well with our patch-based decision-making framework, making ArtNet an appropriate choice for our baseline model.

Expected Results. We expect the following improvements at different parts of the project:

- **Implementation part:** By introducing a shallow neural network as a decision-making module that aggregates the predictions from individual patches, we expect the overall classification accuracy to surpass that of the baseline model.
- **Optimization part:**
  - By adopting a more strategic or adaptive approach for selecting the fifth patch—such as using a downsampled version of the entire image or extracting a patch containing the main semantic content—we anticipate achieving higher classification accuracy compared to simply selecting the center patch.
  - By employing a more sophisticated hierarchical structure for patch extraction, rather than independently classifying five uniformly sized patches, we aim to further improve accuracy and provide insights into the spatial importance of different regions within a painting.

3. Technical Approach

Data Preprocessing

Architecture Implementation and Validation The overall architecture of our model is illustrated in Figure 1. In this design, the base model utilized for patch-level prediction is the ArtNet model.
During inference, the input image is first divided into five patches—top left, top right, bottom left, bottom right, and center. Each patch is then independently processed by the baseline model to produce a probability distribution over the five artistic style classes. These five probability vectors are subsequently concatenated and fed into a shallow neural network, which produces the final style classification.
The primary objective of this stage is to validate that integrating a shallow decision-making network into the prediction pipeline leads to higher classification accuracy compared to directly using the baseline model for whole-image prediction.

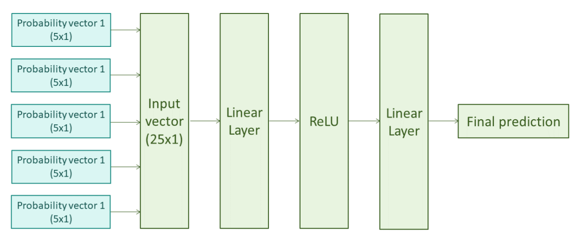

Figure 3. Shallow NN architecture

The architecture of the shallow neural network, consisting of a single hidden layer, is shown above. By utilizing the prepared training data—comprising five probability vectors as inputs and the corresponding ground truth labels as outputs—the shallow network can be trained to learn the mapping parameters. Once trained, the shallow network is directly employed during inference to generate the final style prediction.

Adaptive Refinement and Optimization

- **Selection of the Fifth Patch:** We propose two strategies for selecting the fifth patch:

- *Global Context Patch:* The original image is resized to match the patch size and directly used as the fifth patch. This approach aims to incorporate global contextual information, which is expected to enhance classification accuracy.
- *Semantic-Driven Patch:* A patch containing the most salient semantic content is detected and selected as the fifth patch. This method focuses on capturing the most informative region of the image to assist the decision-making network in improving prediction performance.

- **Architectural Variations:** We also explore more advanced hierarchical architectures for patch selection and decision aggregation. In particular, we investigate the use of architectures such as the Swin Transformer [4], and analyze their impact on the overall classification accuracy.

4. Intermediate/Preliminary Results

Since now, we have completed the architecture implementation and validation part, and received some preliminary results.

References

[1] Rendi Chevi. Artnet - painting style classification and feature visualization. https://github.com/rendchevi/artnet-app. Accessed: 2025-04-24. 2

[2] Saqib Imran, Rizwan Ali Naqvi, Muhammad Sajid, Tauqeer Safdar Malik, Saif Ullah, Syed Atif Moqurrab, and Dong Keon Yon. Artistic style recognition: combining deep and shallow neural networks for painting classification. *Mathematics*, 11(22):4564, 2023. 1

[3] Kaggle. Painter by numbers. https://www.kaggle.com/c/painter-by-numbers/data. Accessed: 2025-04-24. 2

[4] Zhenghua Zhang, Zhangjie Gong, Qingqing Hong, and Ling Jiang. Swin-transformer based classification for rice diseases recognition. In *2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI)*, pages 153–156, 2021. 3