
COSE474-2024F: Final Project Proposal

Art Style Classification using CLIP and WikiArt Dataset

Geonyeong Koh

1. Introduction

Classifying images through artificial intelligence is a core task of computer vision. And understanding what an image signifies is relatively simple task. However, understanding the stylistic features of art and categorizing different art styles are complex tasks because it require deep knowledge of visual patterns.

This project aims to use CLIP for classifying artworks into different art styles with the WikiArt dataset. The goal is to develop a system that can automatically classify artworks into styles such as Impressionism, Baroque, and Surrealism with some additional training.

2. Problem Definition & Challenges

The task is to automatically classify paintings into their respective art styles using CLIP. There are several challenges.

Inter/Intra Class Variability : Some art styles, like Impressionism and Post-Impressionism, share subtle characteristics, making classification difficult. And each art style can vary greatly in its visual expression, making generalization difficult for the model.

Dataset imbalance: Some art styles are overrepresented in the dataset, which can bias the model towards certain classes.

Zero-shot limitations: CLIP's zero-shot learning may struggle with fine-grained distinctions between similar art styles without specific training.

3. Related Works

Inception-ResNet Models: This CNN-based models have achieved competitive results in art classification tasks. The study fine-tuned pre-trained models through transfer learning and used Ensemble method.

EfficientNet: EfficientNet has shown strong performance with fewer parameters, making it a good fit for tasks like art style classification and transfer.

CLIP for Zero-shot Classification: CLIP introduced a powerful approach to linking visual and textual information.

Although the accuracy of CLIP is lower compared to models specifically developed for style classification but it can perform three tasks such as predicting art style, author and time period

4. Datasets

The WikiArt dataset will be used, containing over 80,000 images across 27 different art styles. This dataset is a standard benchmark for art classification tasks and provides rich metadata, including the artist, style, and creation date.

5. State-of-the-Art Methods and Baselines

Inception-ResNet v2: achieves around 68.55% accuracy on WikiArt, focusing on convolutional features for fine-grained visual classification.

EfficientNet: EfficientNet achieves high performance with less computational cost and achieves 79.23% accuracy with proper fine-tuning.

CLIP: As a zero-shot model, CLIP achieves 24.57% accuracy without any additional training on art-specific data.

The baseline of this project will be the zero-shot performance of CLIP on the WikiArt dataset, and the goal is to fine-tune the model to exceed the baseline performance and approach or surpass the results achieved by SOTA models.

6. Schedule

October 28 – November 3: Data preparation and preprocessing.

November 4 – November 10: Implement CLIP zero-shot classification on WikiArt.

November 11 – November 24: Fine-tune CLIP and experiment with different settings.

November 25 – December 1: Compare the fine-tuned CLIP performance with other models like Inception-ResNet v2 and EfficientNet.

December 2 – December 10: write final project paper.