Joint Prediction of Style and Genre in the WikiArt Dataset

and Bob's Bluffing Gallery

Wassim Mecheri

April 2025

WikiArt Dataset & Classification Task

- 81,444 artworks from WikiArt.org
- Labels: 129 artists, 11 genres, 27 styles
- Task: Predict both **genre** and **style**
- Multi-label classification with imbalanced labels

Preprocessing

- WikiArt: 81,444 paintings labeled with style, genre, and artist.
- Significant class imbalance in both genre and style.
- Balanced subset created using a network flow approach:
  - Modeled genre-style pairs as a bipartite graph
  - Applied network flow to find valid matchings
  - Uniform sampling across selected pairs
  - Removed rare styles
  - Remapped style label indices
- Final dataset: 17,061 images, 11 genres, 11 styles

Training

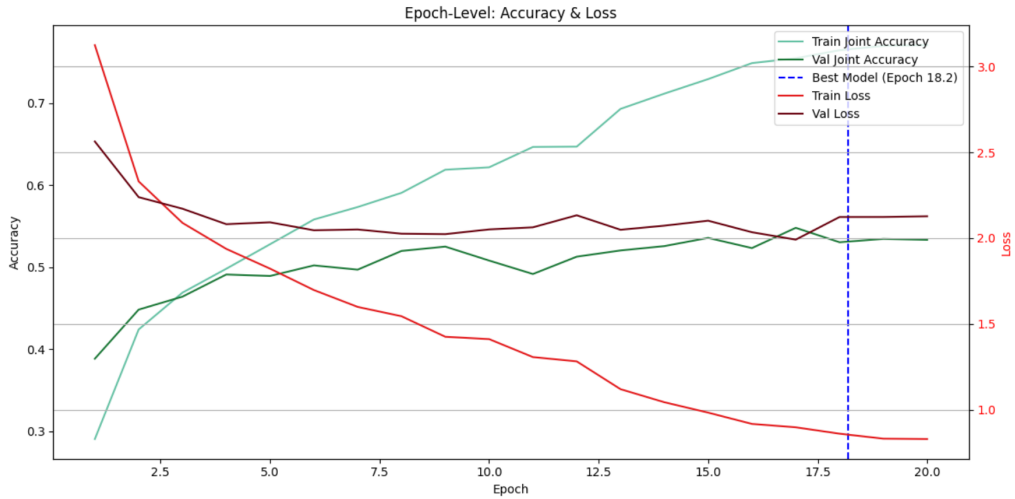
- Dataset split: 67.5% training, 12.5% validation, 20% test
- Models trained:
  - Simple CNN (baseline)
  - ResNet-50 frozen
  - Fine-tuned ResNet-50
- Optimization:
  - Adam optimizer
  - ReduceLROnPlateau scheduler
- Data augmentation used:
  - Random cropping
  - Horizontal flipping
  - RandAugment

Model Architectures

- **Baseline CNN**: 3 convolutional layers
- **ResNet-50 (frozen)**: used as a feature extractor
- **Fine-tuned ResNet-50**: last 4 blocks unfrozen
- Multi-head architecture: separate heads for genre and style

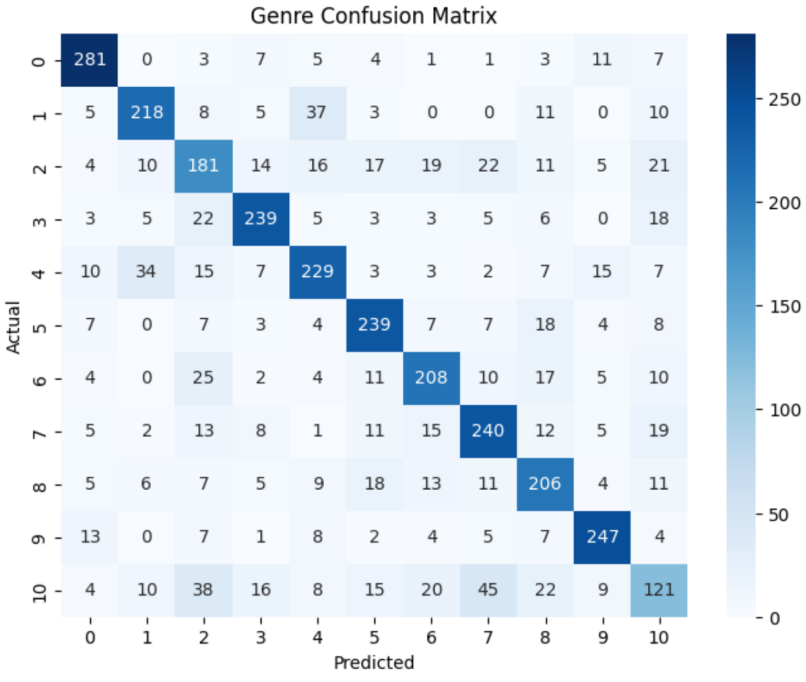
Model Results

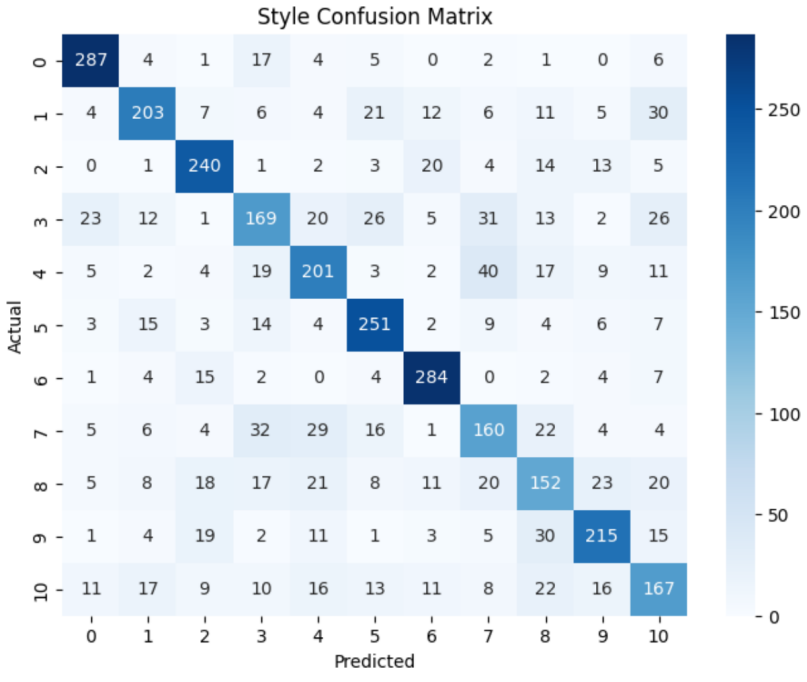
- Fine-tuned ResNet-50 performance:
  - **Genre Accuracy:** 70.6%
  - **Style Accuracy:** 68.2%
  - **Joint Accuracy:** 53.8%



Test Results

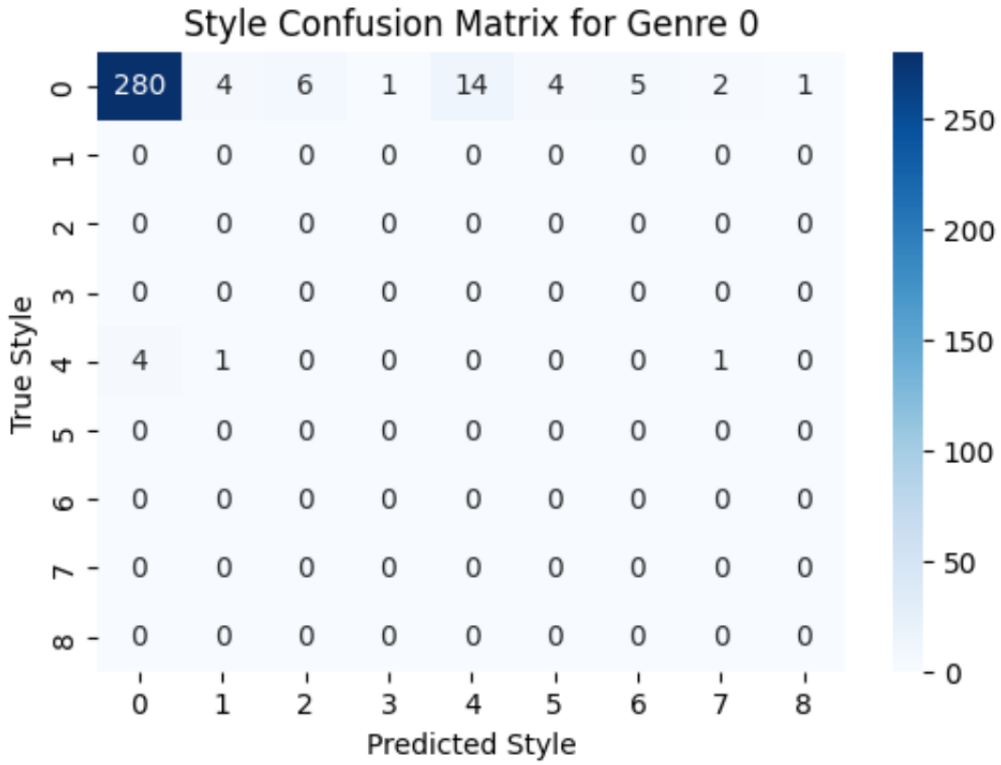
- Persistent bias in underrepresented genre-style combinations
- Confusion remains high for overlapping stylistic features





Bias Analysis

- Joint distribution of genre-style pairs is highly imbalanced despite balanced individual labels
- Model overfits to frequent genre-style combinations, underperforms on rare ones
- Underrepresented styles are often misclassified or ignored within certain genres



Application: Bob's Bluffing Gallery

- Bob = Novice AI tour guide
- Describes art using model predictions only
- Built with Flask + HTML/CSS/JS
- Descriptions generated by TinyLlama (LLM)

TinyLlama Integration

- TinyLlama-1.1B-Chat-v1.0 used for speed (~3s)
- Larger models (Mistral, Phi-4) were too slow
- Custom prompts and rules to guide responses
- Still imperfect, but adds charm to Bob

Conclusion & Future Work

- Fine-tuned ResNet-50 works well for joint prediction
- Turned model flaws into an engaging game
- Future:
  - Improve prompt handling and LLM quality
  - Image-to-text models for richer context
  - Balance joint label pairs more precisely