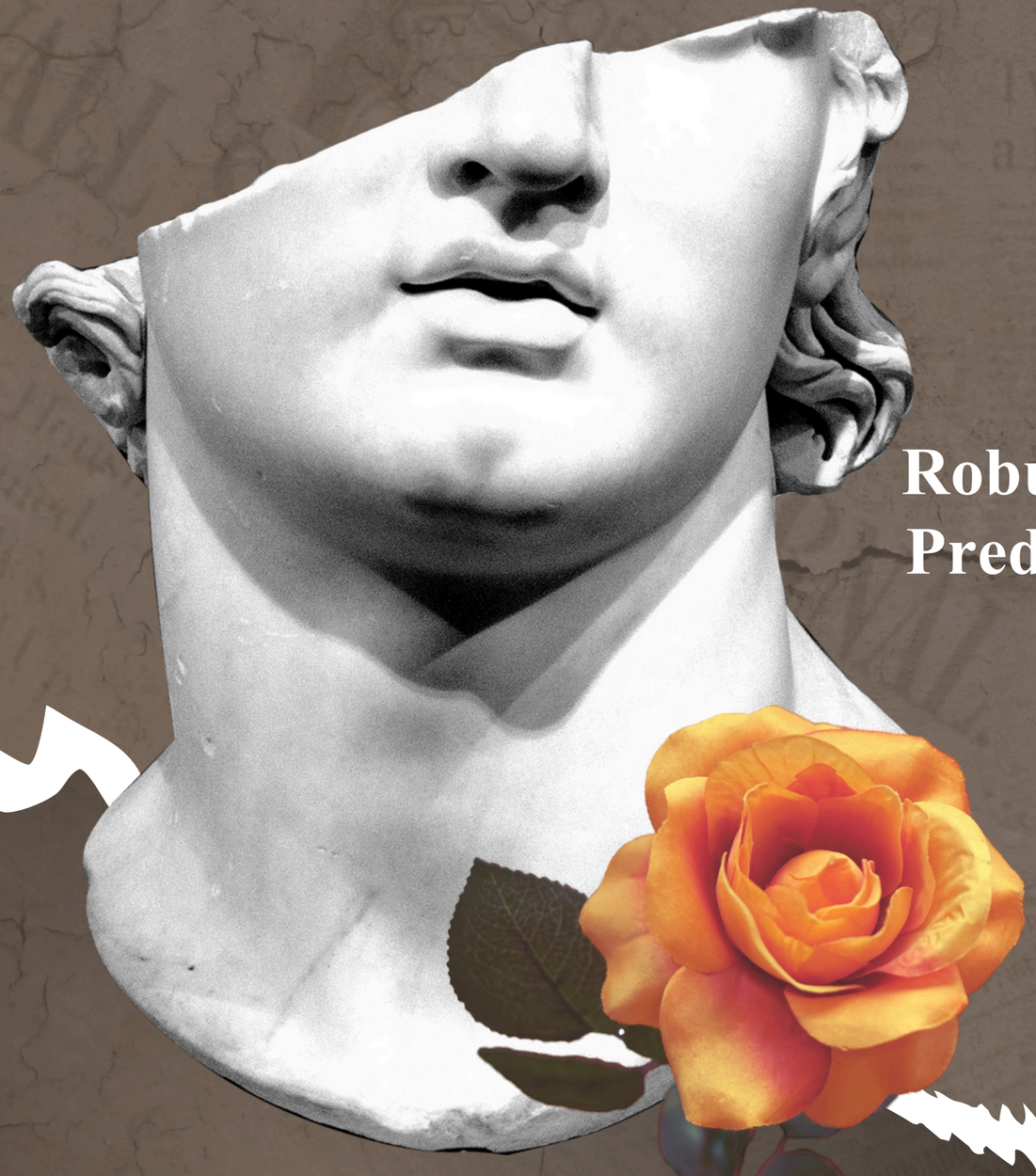
Jan. 2026

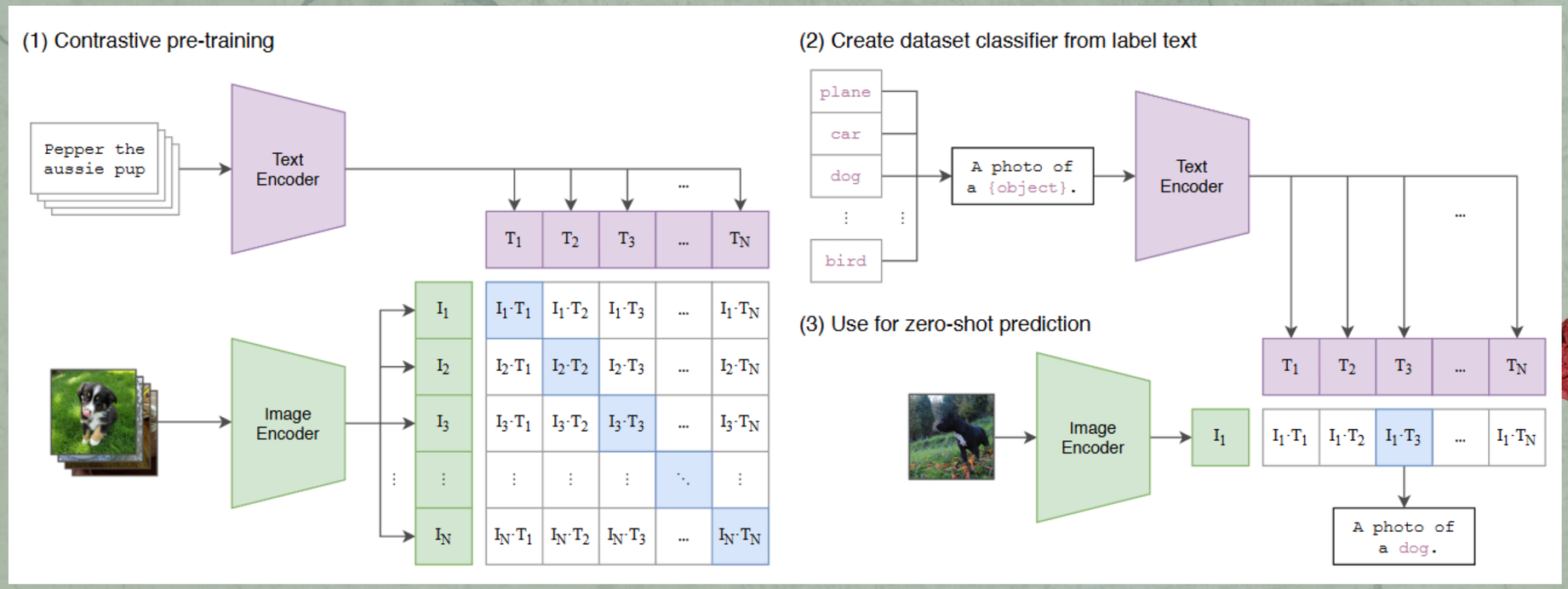


Robustness of Pre-Trained CLIP for Artist Predictions Under Visual Transformations

Presented by group 13

Introduction

- There is no established work that evaluates pre-trained CLIP for artist prediction.
- Robustness to visual transformations is a long-standing topic in CV. CLIP's robustness should be evaluated for specific tasks.
- SemArt-based studies have not explored pre-trained CLIP models.



How robust is a pre-trained CLIP model for top 100 most prolific artists predictions when visual properties of paintings are changed?

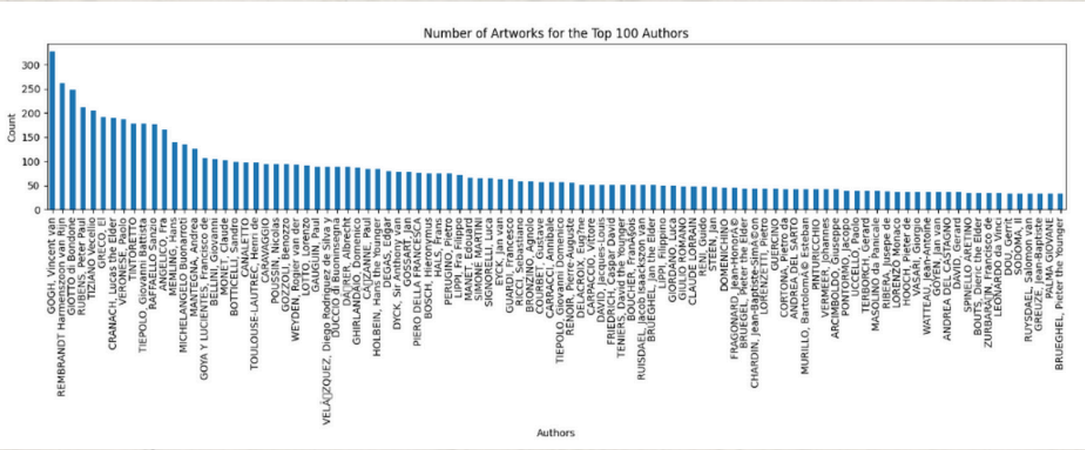
1. How well does pre-trained CLIP perform at top 100 most prolific artists predictions?
2. How does removing color information affect CLIP's artist predictions?
3. How does geometric distortion affect CLIP's artist predictions?
4. How does Elastic Transform affect CLIP's artist predictions?

Dataset Overview

- SemArt dataset - Aston University
- European paintings - 13th to 19th centuries
- Metadata:
  - Title
  - Technique
  - Date
  - Type
  - School
  - Timeframe
  - Author

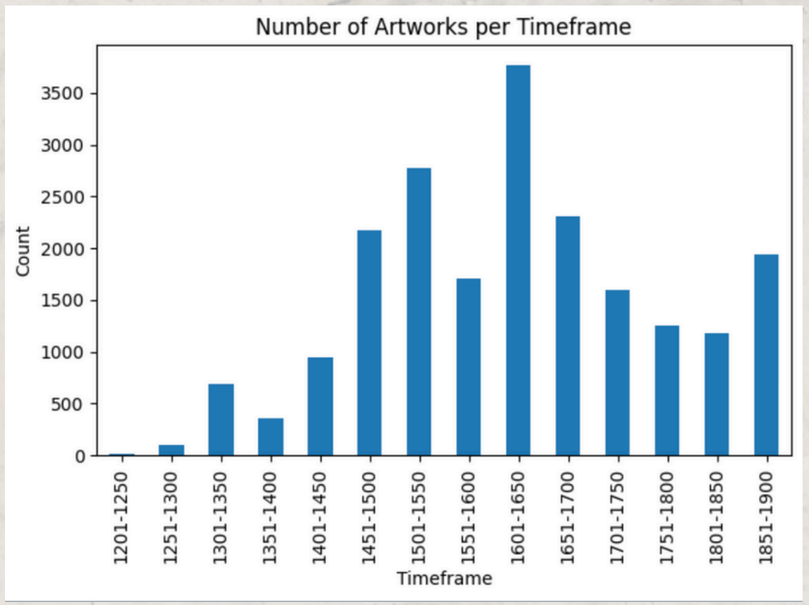
Artists

- Total number of distinct artists: 3,253
- Most prolific artists:
  - Vincent van Gogh
  - Rembrandt van Rijn
  - Giotto di Bondone



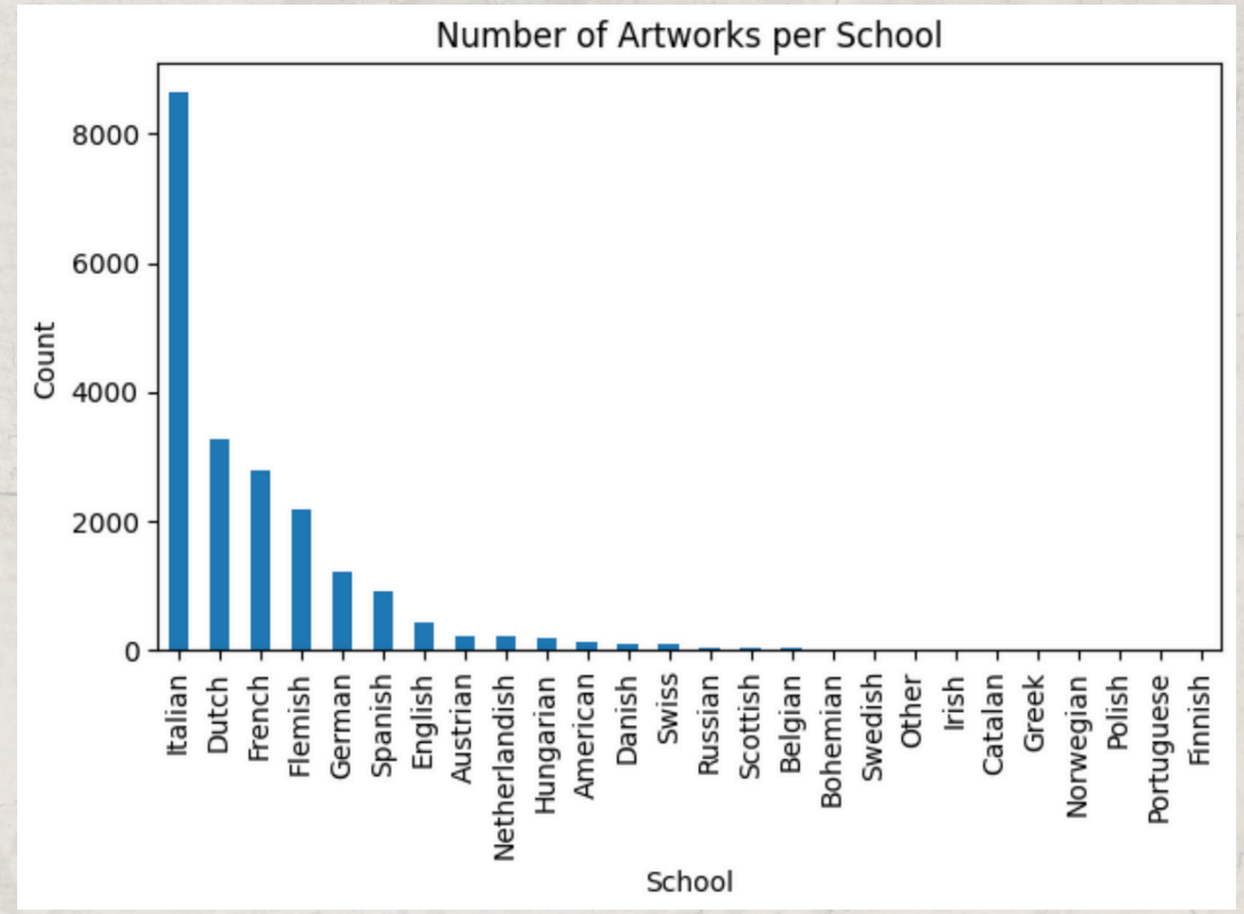
Timeframe

- Paintings span from the 13th to the 19th century
- Oldest period: 106 paintings
- Gradual increase over time
- Peak in 17th century: 3,770 paintings
- Last half of the 19th century: 1,936 paintings



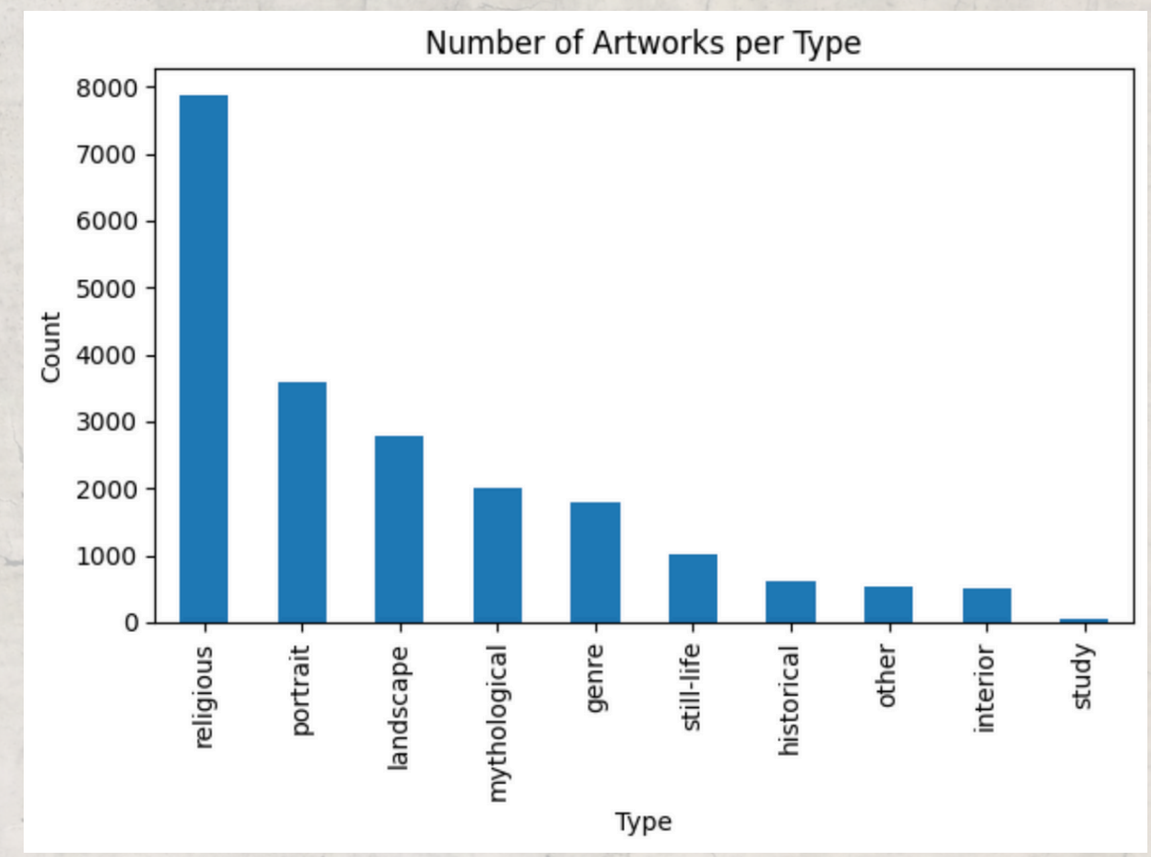
Schools

- Italian schools:
  - 8,652 paintings
  - 42% of the dataset
- Dutch, French, and Flemish schools:
  - Combined 40% of the dataset
- Least represented schools:
  - Greek
  - Norwegian
  - Polish
  - Portuguese
  - Finnish



Artwork Types

- 10 artwork categories
- Most common:
  - Religious paintings: 7,872 (37%)
- Other common types:
  - Portrait: 17%
  - Landscape: 13%
- Least common:
  - Study
  - Interior
  - Other

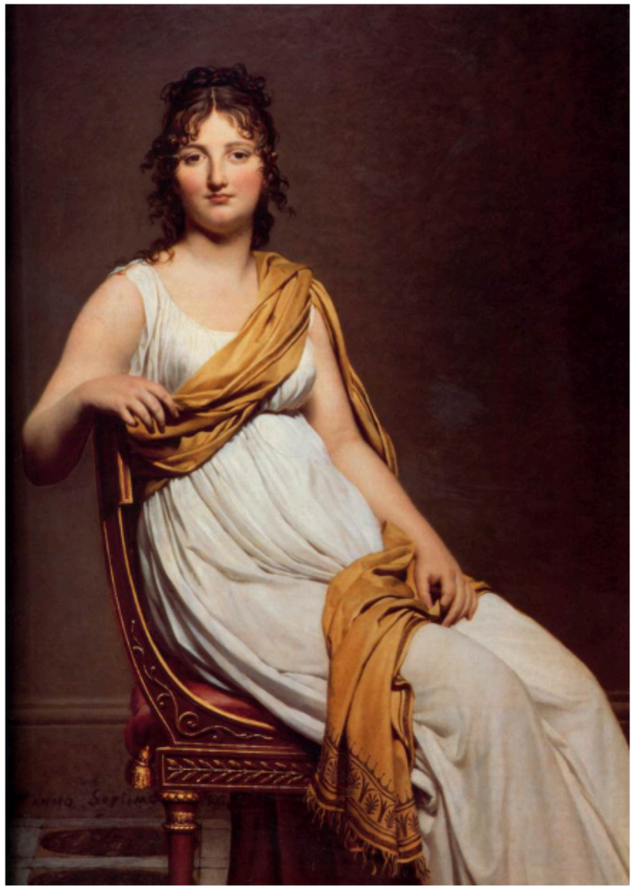


CLIP Model

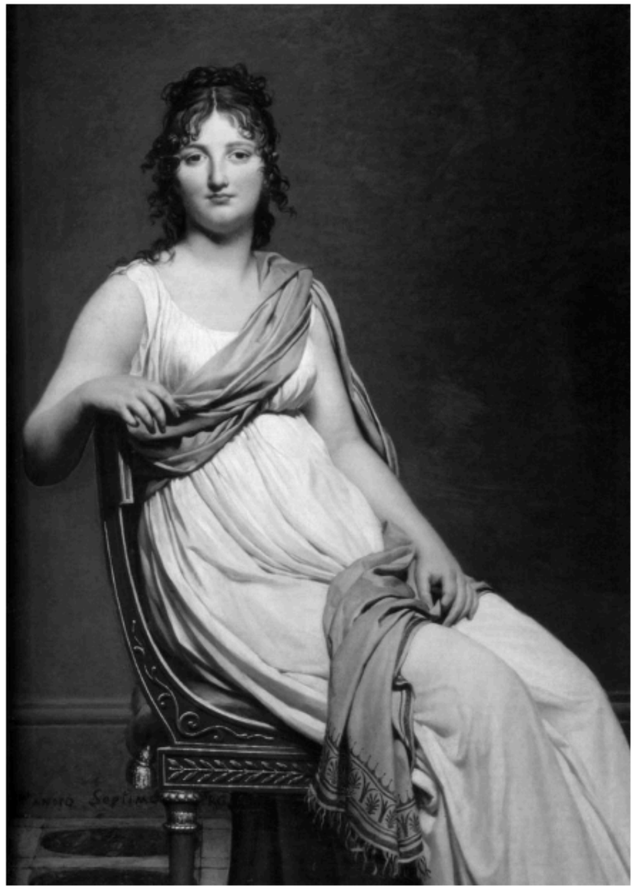
- ViT-B/32 backbone in a zero-shot classification
- Top 100 most prolific artists
- Painter with the highest Cosine Similarity score

Grayscale

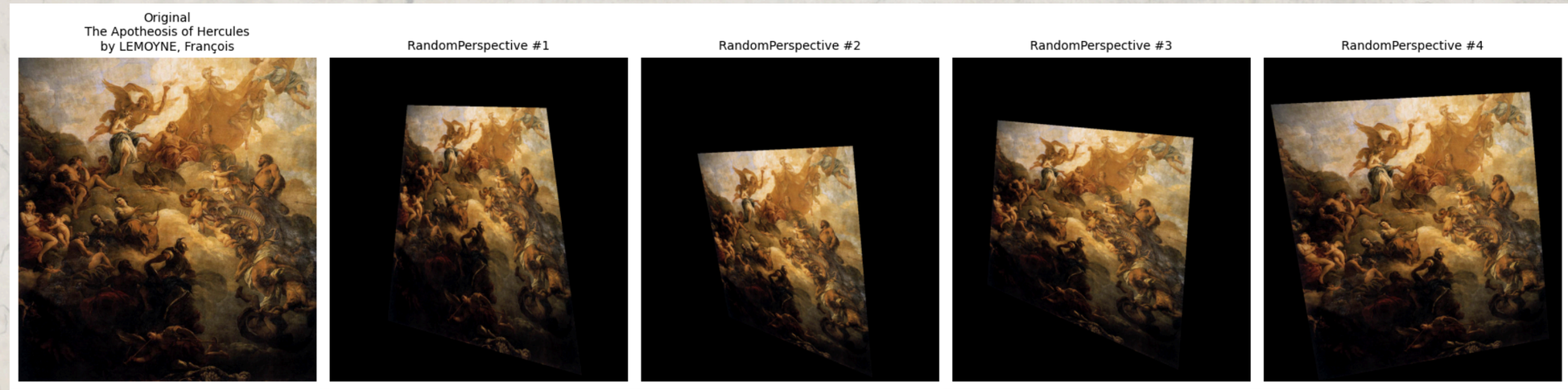
Original
Madame Raymond de Verninac
by DAVID, Jacques-Louis



Grayscale (PyTorch)
Madame Raymond de Verninac
by DAVID, Jacques-Louis



Random Perpective



Elastic

Original
Peter Zrinyi and Ferenc Frangepán in the Wiener-Neustadt Prison
by MADARÁSZ, Viktor

Elastic Transform (alpha=250.0, sigma=5.0)
Peter Zrinyi and Ferenc Frangepán in the Wiener-Neustadt Prison
by MADARÁSZ, Viktor

Results

Table 1: Pre-trained CLIP performance for top-100 artist prediction

Metric	Value
Accuracy	29.56%
Macro Precision	0.25
Macro Recall	0.27
Macro F1 Score	0.23
Weighted F1 Score	0.28

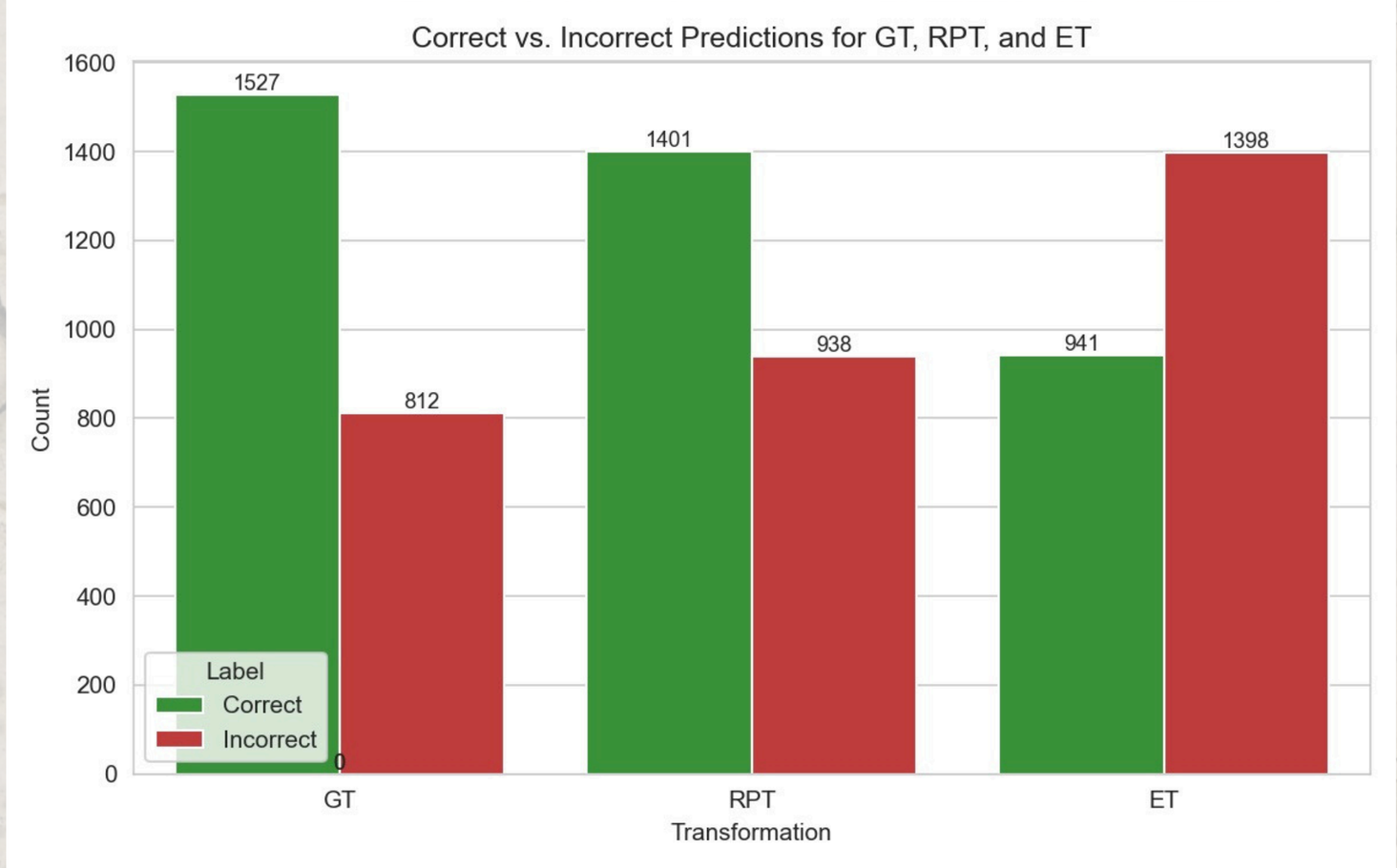
```
Top 10 Best Performing Authors:
  CANALETTO                          - 86.6% (84/97)
  CÉZANNE, Paul                      - 83.3% (70/84)
  VERMEER, Johannes                  - 83.3% (35/42)
  GAUGUIN, Paul                      - 83.1% (74/89)
  RENOIR, Pierre-Auguste             - 80.4% (45/56)
  CARAVAGGIO                         - 77.9% (74/95)
  FRAGONARD, Jean-Honoré             - 75.6% (34/45)
  ARCIMBOLDO, Giuseppe               - 73.8% (31/42)
  RUYSDAEL, Salomon van              - 69.7% (23/33)
  CHARDIN, Jean-Baptiste-Siméon      - 65.9% (29/44)

Bottom 10 Authors:
  BOUTS, Dieric the Elder            - 0.0% (0/35)
  GOSSART, Jan                       - 0.0% (0/79)
  PALMA GIOVANE                      - 0.0% (0/33)
  RAFFAELLO Sanzio                   - 0.0% (0/177)
  GIULIO ROMANO                      - 0.0% (0/49)
  GIORDANO, Luca                     - 0.0% (0/50)
  SPINELLO ARETINO                   - 0.0% (0/35)
  GHIRLANDAIO, Domenico              - 0.0% (0/87)
  GOYEN, Jan van                     - 0.0% (0/36)
  RICCI, Sebastiano                  - 0.0% (0/59)
```

Accuracy of 29.56%, correctly identifying the artist for 2339 out of 7913 paintings.

Performance varies across individual artists. Several artists achieve high prediction accuracy. A group of artists receive no correct predictions despite having a large number of available works.

Results


Correct vs. Incorrect Predictions for GT, RPT, and ET
Count
1600
1400
1200
1000
800
600
400
200
0
1527
812
0
1401
938
941
1398
Label
Correct
Incorrect
GT
RPT
ET
Transformation

Results

- Only 23.39% of all paintings were correctly predicted across all transformations
- 15.31% of all paintings were incorrectly predicted across all transformations

Grayscale	Perspective	Elastic	Count	%
1	1	1	547	23.39
1	1	0	478	20.44
1	0	1	167	7.14
0	1	1	149	6.37
1	0	0	335	14.32
0	1	0	227	9.71
0	0	1	78	3.33
0	0	0	358	15.31

Table 2: Prediction pattern across transformations (N=2339). Entries indicate whether the prediction is correct (1) or incorrect (0) after each transformation.

- The confidence of the model decreases after transformation

Transformation	$\bar{c}_{\text{base}}$	$\bar{c}_{\text{trans}}$	$\bar{c}_{\text{trans}}$ - $\bar{c}_{\text{base}}$	$\bar{c}_{\text{corr}}$	$\bar{c}_{\text{incorr}}$
GT	0.318	0.295	-0.024	0.300	0.285
RPT	0.318	0.298	-0.020	0.302	0.293
ET	0.318	0.308	-0.011	0.313	0.304

Table 3: Confidence statistics (N=2339).

Results

- Micro Recall = Micro F1 = Accuracy
- Overall performance decreases after transformation, and the decrease is much larger when averaged equally across artists, indicating that some artists are much harder to predict after transformation

Transformation	Micro Recall	Macro Recall	Weighted Recall	Micro F1	Macro F1	Weighted F1
GT	0.653	0.494	0.653	0.653	0.437	0.662
RPT	0.599	0.408	0.599	0.599	0.388	0.615
RT	0.402	0.292	0.402	0.402	0.256	0.419

Table 4: Recall and F1 scores after each transformation (N=2339 paintings, 76 artists).

Limitations

- Limited computation
- Selection Bias
- CLIP designed for text–image representation
- F1, Precision Recall overestimating overall performance

Conclusion

- Artist attribution by CLIP.
- Grayscale conversion had the smallest impact.
- Random perspective transformations caused moderate performance drop.
- Elastic transformations had the strongest negative effect
- CLIP shows limited robustness and favors structural cues over color.

Bibliography

1. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning Transferable Visual Models From Natural Language Supervision. In: Proceedings of the 38th International Conference on Machine Learning, pp. 8748–8763. PMLR (2021)

2. Alayrac, J.B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M., et al.: Flamingo: A Visual Language Model for Few-Shot Learning. In: Advances in Neural Information Processing Systems, pp.23716–23736 (2022)

3. Ghildyal, A., Wang, L.Y., Liu, F.: WP-CLIP: Leveraging CLIP to Predict Wölfflin's Principles in Visual Art. arXiv preprint arXiv:2508.12668 (2025)

4. Conde, M.V., Turgutlu, K.: CLIP-Art: Contrastive Pre-training for Fine-Grained Art Classification. arXiv preprint arXiv:2204.14244 (2022)

5. Garcia, N., Renoust, B., Nakashima, Y., Yanai, K.: SemArt: A Dataset for Semantic Art Understanding. In: Proceedings of the 26th ACM International Conference on Multimedia, pp. 259–267. ACM, New York (2018)

6. Shorten, C., Khoshgoftaar, T.M.: A Survey on Image Data Augmentation for Deep Learning. Journal of Big Data 6(1), 1–48 (2019)

7. Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., Brendel, W.: ImageNet-Trained CNNs Are Biased Towards Texture. In: Proceedings of the 7th International Conference on Learning Representations (2019)

8. Hendrycks, D., Dietterich, T.: Benchmarking Neural Network Robustness to Common Corruptions. In: Proceedings of the 7th International Conference on Learning Representations (2019)

9. Usama, M., Asim, S.A., Ali, S.B., Wasim, S.T., Mansoor, U.B.: Analysing the Robustness of Vision-Language Models to Common Corruptions. arXiv preprint arXiv:2504.13690 (2025)

10. Dahal, A., Murad, S.A., Rahimi, N.: Embedding Shift Dissection on CLIP: Effects of Augmentations on VLMs Representation Learning. arXiv preprint arXiv:2503.23495 (2025)


THANK
YOU!