IDS 705 Final Report

Enhancing Emotion Classification in Artworks through Genre Information

A Transfer Learning Approach

Team Members:
Osama Ahmed
Anastasiia Saenko
Haochong (Harry) Xia
Jiayi Zhou

Team Number: 4

Abstract

Art is a powerful medium that elicits a broad spectrum of emotions from viewers. However, translating these emotional nuances into computational terms remains a significant challenge. In this project, we investigate the relationship between art genres and the emotions they evoke to enhance emotion classification in artworks. Leveraging transfer learning techniques, we fine-tuned pre-trained ResNet50 models on a dataset combining WikiArt pieces and Artemis annotations to classify art pieces into positive or negative emotional categories across four genres. Results show that incorporating genre information improves emotion classification accuracy. Furthermore, our findings suggest that sequential training of convolutional blocks yields better performance than simultaneous block training. These insights pave the way for further exploration of model configurations aimed at optimizing emotion classification in art by incorporating genre information.

1. Introduction

Emotions elicited by art are deeply personal and can vary significantly from one person to another. Throughout history, it has been observed that specific art genres elicit particular emotional responses; for instance, the vivid colors and dynamic forms of Expressionism often evoke feelings of agitation and unease, while the gentle brushstrokes and light color palettes of Impressionism are typically associated with positive emotions like calmness and joy *(Damiano et al., 2023)*.

Recognizing and classifying the emotions that art evokes plays a crucial role in enhancing art discovery and personalizing selections on digital platforms. However, current search mechanisms struggle to capture and translate the subtle emotional nuances that users seek into specific thematic preferences. For example, searching for '*happy painting*' or '*sad painting*' frequently yields unsatisfactory results due to the complexity of mapping these broad emotional descriptors to particular artworks computationally *(Zhao et al., 2021)*. This discrepancy between user needs and technological capabilities highlights the need for more sophisticated methods in emotion classification, leveraging genre as a key contextual factor.

This project investigates the relationship between the genre of an artwork and the emotions it evokes to assess whether this can lead to more accurate predictions of emotional responses. Previous studies in this field have primarily focused on direct emotional categorization, often overlooking the role that artistic genre plays in shaping emotional perceptions. Works like ArtGraph *(Sinem et al., 2022)* leverage visual features and knowledge graph embeddings for emotion classification research and end up implicitly doing genre classification in the process, citing an underlying relationship between emotion evoked and genre. It is upon such research that our work explores emotion prediction by utilizing insights gained about genre. This approach not only enriches our predictions but also enables us to conduct experiments that quantify how genre inclusion improves emotion classification. We primarily focus on categorizing artworks into two emotional categories—positive and negative—across four genres: Renaissance, Post-Renaissance, Modern, and Contemporary.

2. Background

2.1 *The Project Background: Art, Emotions & Genres*

Art and Emotions

Art as a medium of expression predates as early as 45,000 years ago when early humans used cave paintings to document aspects of their daily lives—from hunting scenes to homesteads and imaginary creatures *(Brumm et al., 2021)*. These initial representations, ranging from simple doodle-like strokes to more complex abstract forms, have established the foundation for what has evolved into a multi-billion-dollar art industry spanning across thousands of stylistic categories, evoking human emotions, and influencing experiences globally *(Serrao et al., 2024)*.

Genres and Emotions

Art is organized into categories commonly known as genres, which differ by objective factors like stylistic features and historical context, as well as more subjective aspects such as social impact. Some genres are deeply rooted in specific historical periods, while others span decades or even centuries. Each genre is defined not only by its stylistic elements but also by the emotional responses it typically evokes giving each genre a unique yet universally impactful character *(DiMaggio, 1987)*. For example, the grandeur of Baroque art, originating around the 17th century primarily in Europe, popularized by artists like Caravaggio is characterized by its dramatic use of light and intricate detail, often evoking a sense of awe and emotional intensity, typically depicting themes of power and religion *(Dahiya, 2022)*. On the other hand, Impressionism, which emerged in the late 19th century in France, popularized by artists like Claude Monet is characterized by its focus on capturing everyday moments through the use of light and color conveying the beauty of fleeting moments in everyday life, often eliciting warm and positive emotions *(Clark, 2023)*.

The relationship between art genres and the emotions they evoke is essential for appreciating art but remains underexplored due to its subjective nature and the laborious process of interpretation and compilation *(Paasschen et al., 2015)*.

Emotion Classification

In the art field, modern data analysis and computational technologies have equipped researchers with tools not only for objective tasks such as genre classification and artistic reconstruction but also for more complex and subjective tasks like inferring the emotions evoked by artworks. Historically, the process of identifying emotions evoked by artworks has relied on human annotation, which involves examining thousands of reviews and commentaries. However, with the advent of deep learning technologies capable of transforming images into low-level representations and analyzing them pixel by pixel, new avenues for enhancing such tasks are being explored. In the field of image classification, current technologies *(Zhao et al., 2021)* struggle to capture the nuanced emotional content of artworks because they overlook the deep emotional connection and historical contexts that genres, epochs, and stylistic elements provide that is otherwise available to humans.

2.2 Related Works

WikiArt Emotions: An Annotated Dataset of Emotions Evoked by Art

The WikiArt Emotions project *(Mohammad & Kiritchenko, 2018)* embarked on an in-depth analysis to discern the elements that make art evocative. This extensive initiative annotated over 4,000 artworks from the WikiArt dataset *(WikiArt, 2010)*, predominantly paintings, to identify the emotions they stir in observers. These annotations were gathered through crowdsourcing, allowing a diverse range of human annotators to contribute their emotional interpretations. The project specifically examined how various elements, such as the artwork's title and visual content, influence emotional responses. The findings revealed that emotions like fear, happiness, love, and sadness not only predominate the emotional responses to these artworks but also elicit remarkably consistent reactions among different annotators. The culmination of this research led to the creation of the WikiArt Emotions Dataset. This valuable resource encompasses emotion annotations for a broad array of art spanning four major Western styles—Modern Art, Post-Renaissance Art, Renaissance Art, and Contemporary Art—across 22 style categories, each annotated for one or more of twenty distinct emotion categories.

Recognizing the Emotions Evoked by Artworks Through Visual Features and Knowledge Graph-Embeddings

This research explores the intersection of visual features and knowledge graph-embeddings in emotion classification from artworks. Utilizing data from Artemis and Artgraph, the study *(Sinem et al., 2022)* employed various models including ResNet50, citing its effectiveness in image classification tasks to enhance emotion-based information retrieval and knowledge discovery. Single-task emotion recognition achieved 61.32% accuracy while the multiclass combination of emotion, style, and genre recognition achieved 63.27%. From this approach, it was learned that emotion classification is largely dependent on visual features and contextual features like genre and suggested that synergy would allow for a more nuanced and effective classification process, suggesting that these elements can assist each other to improve accuracy.

3. Data

The datasets used in this study were sourced from WikiArt *(WikiArt, 2010),* which includes 80,020 unique images covering 27 artistic styles from 1,119 artists, and the Artemis dataset *(Achlioptas et al., 2021)*, comprising 455,000 human annotated emotion attributions from a collection of 80,000 artworks. We merged images from WikiArt with annotations from Artemis to create a dataset categorized by genre and corresponding emotion evoked. Emotions were further classified into binary categories: positive and negative to harmonize the translation from human annotations to standardized emotion categories. Positive emotions *(figure 1)* include feelings such as contentment and excitement, whereas negative emotions *(figure 2)* include sadness and fear. Due to imbalances, the original twenty genres were consolidated into four main genres: Renaissance, Post-Renaissance, Modern, and Contemporary *(figure 3)*. The resulting dataset used in this project comprises 10,000 images; 3,897 images associated with negative emotions and 6,191 with positive emotions. Genre data included 2,899 images each from Renaissance and Post-Renaissance, 2,900 from Modern, and 1,333 images from Contemporary.

Figure 1*: Three art pieces that evoke positive emotions.*

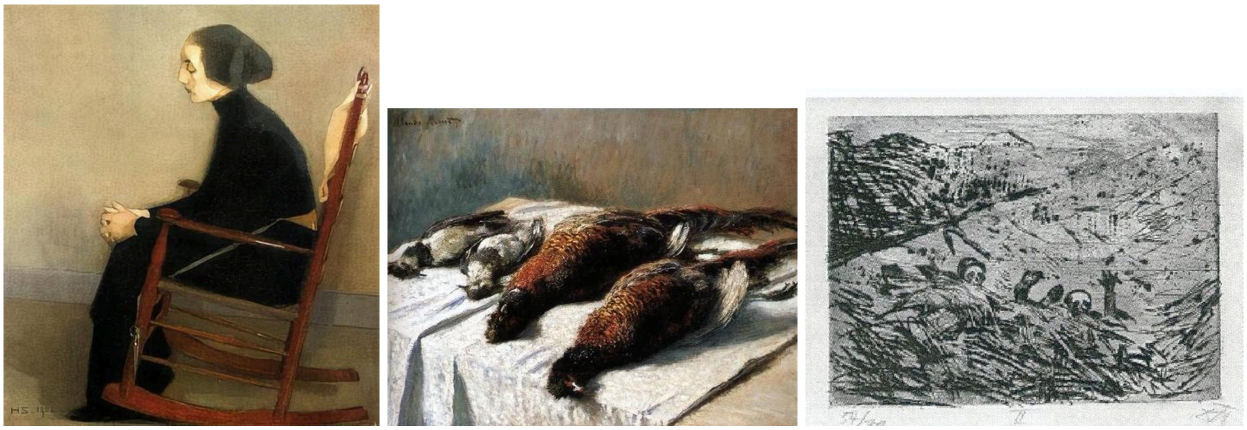

Figure 2: *Three art pieces that evoke negative emotions.*

4. Experiments

Our experimental design aimed to objectively assess the impact of integrating genre into the emotion classification process. We compared our transfer learning model, which incorporates genre, against a baseline model that does not. Building on prior research suggesting a nuanced relationship between genre and the emotions artworks evoke (Sinem et al., 2022), our hypothesis posits that genre significantly informs emotion classification. Both models share the same architecture, but while the baseline model classifies emotions directly, the transfer learning model includes an intermediate genre classification step. We then freeze the weights and fine-tune them for emotion classification, allowing us to evaluate the added value, if any, of incorporating genre into the emotion prediction process by analyzing differences in performance metrics between the two models.

4.1 Transfer Learning

Transfer learning is a machine learning technique that repurposes a model trained on one task for a different but related task, especially when labeled data for the target task is limited *(Oquab et al., 2014)*. In our project, we utilized transfer learning to enhance emotion classification in artworks by incorporating genre information as a contextual cue. We first trained a model to classify artworks by genre and then fine-tuned this model to identify the emotions evoked by these artworks. This sequential training approach leverages genre classification to improve emotion recognition in artworks, demonstrating how contextual cues can enhance the performance of machine learning tasks.

4.1.1 Preparing the Data

Before training the models, we preprocessed the images to ensure robust training and improved model generalization (Chollet, 2018). This preprocessing involved operations such as rotation, height shifting, width shifting, and rescaling *(figure 3)*. These steps enhanced the model's ability to learn invariant representations of the data, thereby improving its performance on unseen data.

Figure 3*: Data Preparation and Preprocessing Pipeline*

4.1.2 Selecting a Base Model

We used Convolutional Neural Networks (CNNs), a widely adopted and highly effective architecture for image classification tasks *(He et al., 2016)*. CNNs have demonstrated exceptional performance in various image-related tasks, including object recognition, scene understanding, and image classification. Their hierarchical structure allows them to automatically learn features from raw data, making them particularly well-suited for our task of classifying images based on their emotional content.

Furthermore, we employed ResNet (Residual Neural Network) architecture *(He et al., 2016)*. ResNet is renowned for its depth, which enables it to effectively capture intricate features from images which was crucial for our task, as the emotional content of images can be subtle and complex, requiring a model capable of capturing nuanced features. Additionally, the use of residual connections in ResNet mitigates the vanishing gradient problem, allowing for the successful training of very deep networks *(He et al., 2016)*.

Specifically, we utilized a pre-trained ResNet50 model, which had been trained on the ImageNet dataset *(Deng et al., 2009)* providing a strong foundation of general image features for us to leverage in this task.

Lastly, we established our baseline model, a resnet50 pre-trained on ImageNet fine-tuned exclusively for emotions. This baseline model served as the starting point for our comparative analysis.

4.1.3 Model Training

For comparison, we developed a version of the baseline model incorporating genre classification before emotion classification. We explored two architectural configurations to understand how this additional step may affect the model's performance depending on the choices of layers to unfreeze. This resulted in two baseline models and two models fine-tuned on genre before emotion for comparative analysis.

The experiments were conducted in two configurations, both fine-tuning the final layers of ResNet50 and freezing the rest of the model to leverage its feature extraction capabilities since it is pre-trained on ImageNet.

4.1.3.1 Configuration 1: Sequential Training

In the first configuration *(figure 4)*, we established our baseline model by training the fifth and final convolutional block of ResNet50 for emotion classification and replacing the final softmax layer of the pre-trained ResNet50 with a binary classification head for emotion classification (positive & negative).

In the second experiment, we started with a pre-trained ResNet50. However, this time, we first trained its fourth convolutional block on genre classification and subsequently trained the model's fifth convolutional block for emotion classification, utilizing the features learned from genre classification. This model configuration aimed to explore whether incorporating genre classification training improves emotion prediction. We hypothesized that the additional information from genre classification would enhance the model's ability to accurately classify the emotions evoked by artworks.

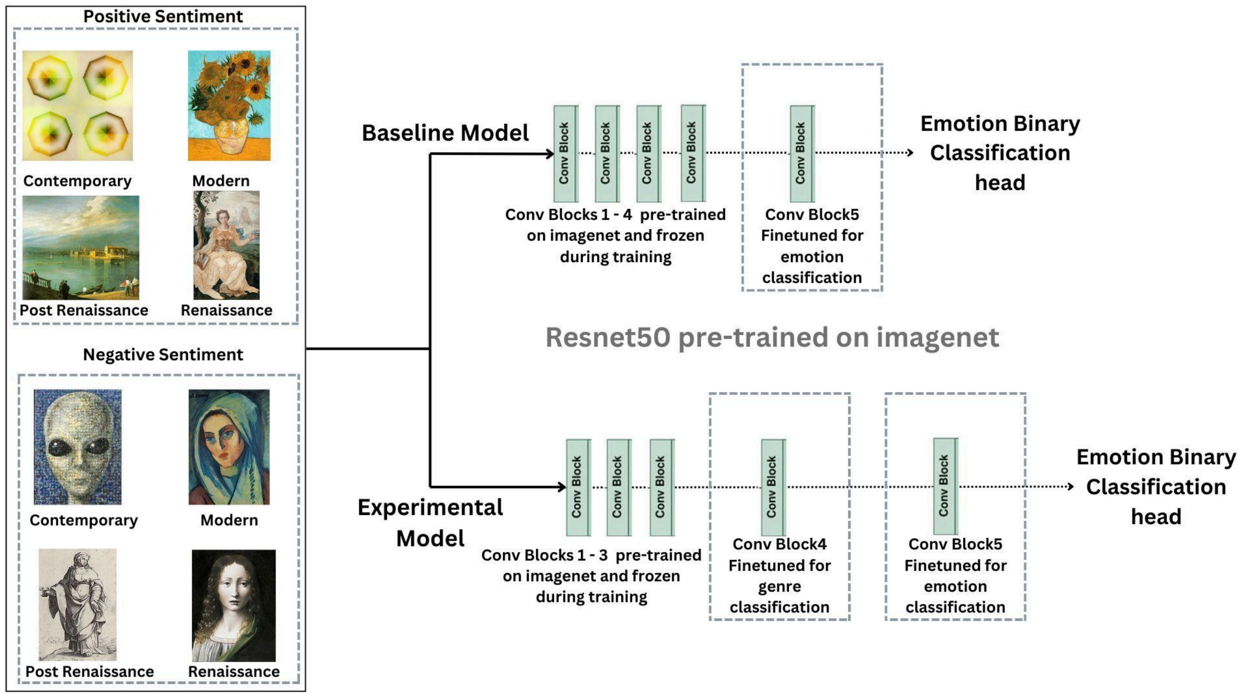

Figure 4*: Experimental Configurations for Sequential Model Training*

4.1.3.2 Configuration 2: Simultaneous Training

In the second configuration *(figure 5)*, we sought to eliminate any effects introduced by the sequential training of the fourth and fifth convolutional blocks of the model in our transfer learning model. To achieve this, we repeated the same experiments, but this time, we trained both the fourth and fifth convolutional blocks simultaneously. For the baseline model, we trained both convolutional blocks for emotion classification. In the experimental model, we first trained both the fourth and fifth convolutional blocks simultaneously for genre classification as part 1, then followed by emotion classification as part 2 *(figure 5)*. By keeping all other factors constant, except for the genre training step, we aimed to directly compare the effects of genre classification on emotion prediction.

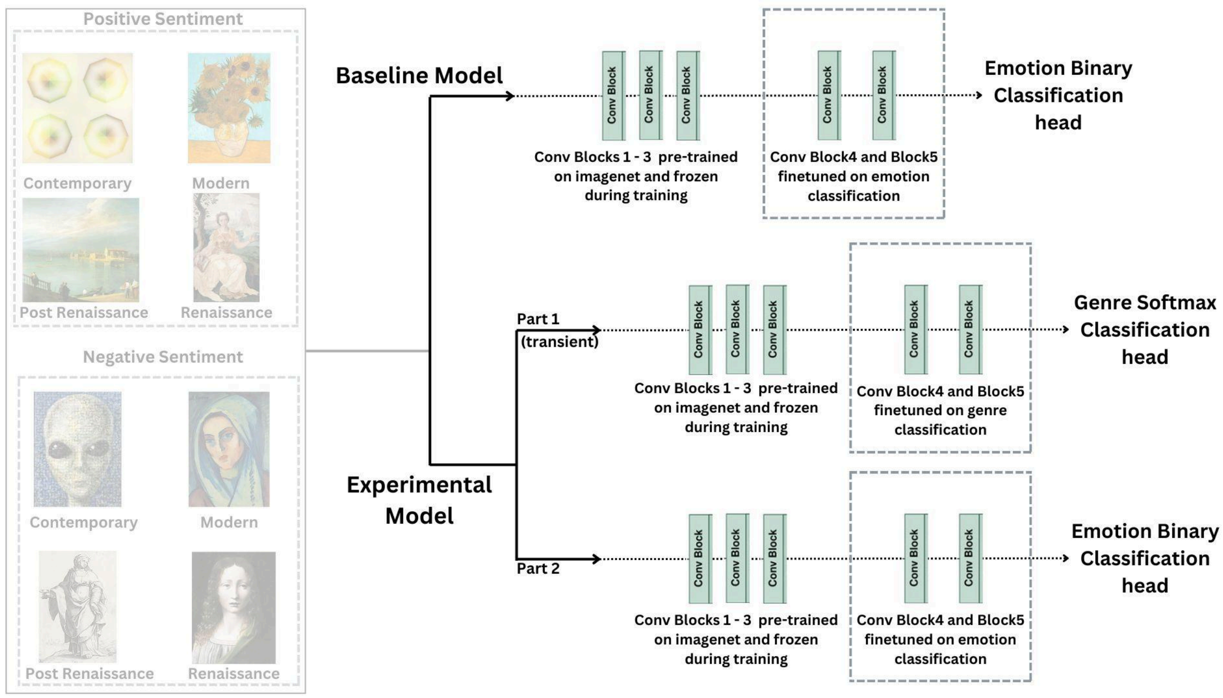

Figure 5*: Experimental Configurations for Simultaneous Model Training*

4.1.4 Evaluation and Comparison

We evaluated the performance of the transfer learning models on our two configurations on a set of test data reserved for emotion classification. Accuracy and F1 score were chosen as the metrics for assessing how well the model identified the emotions evoked by art pieces, considering the slight imbalance in the data categories. This performance was compared against the baseline models, which were trained directly on the emotion classification task without the transient genre training step. The transfer learning models had better performance confirming the hypothesis that knowledge of art genres enhances emotion classification.

5. Results

Overall, we observe an improvement in accuracy in both configurations when the genre component is introduced. This confirms our hypothesis that incorporating knowledge of art genres enhances emotion classification performance. In configuration one, sequential training of the fourth convolutional block on genre followed by training the fifth convolutional block on genre, resulted in a 22.3% and 4.2% improvement in accuracy and F1 Score respectively from the 56.3% and 0.71 baseline values *(Table 1)*. Similarly, in configuration two, where the fourth and fifth blocks were trained simultaneously, incorporating genre information led to a 6.3% and 3% improvement in accuracy and F1 Score respectively from the 50.6% and the 0.66 baseline values (*Table 2)*. These results highlight the importance of considering multiple aspects, such as genre, in enhancing emotion classification performance in artworks.

Fine-tuning Process	Output Observed	Layers Fine Tuned	Accuracy	F1 Score
Trained on Genre (transient)	Genre	Conv4	63.8%	0.6
Trained on Emotions	Emotions Evoked	Conv5	56.3%	0.71
Trained on Genre then Emotions	Emotions Evoked	Conv4 then Conv5	**68.9%**	**0.74**

Table 1. *Experimental Conditions and Performance Comparison of the first configuration; sequential training of the fourth and fifth layers of Resnet50 pre-trained on Imagenet*

Fine-tuning Process	Output Observed	Layers Fine Tuned	Accuracy	F1 Score
Trained on Genre (transient)	Genre	Conv4 & Conv5	62.9%	0.61
Trained on Emotions	Emotions Evoked	Conv4 & Conv5	50.6%	0.66
Trained on Genre then Emotions	Emotions Evoked	Conv4 & Conv5	**53.8%**	**0.68**

Table 2*. Experimental Conditions and Performance Comparison of the second configuration; simultaneous training of the fourth and fifth layers of Resnet50 pre-trained on Imagenet*

Training the fourth and fifth layers sequentially in the first configuration, with the addition of genre information, outperformed simultaneous training in the second configuration. This observation led us to investigate the sequential training approach further. In an extra experiment, we trained the fourth and fifth layers sequentially, but this time focusing solely on emotion classification. The results *(table 3)* provided valuable insights into the impact of training order on model performance.

Fine-tuning Process	Output Observed	Layers Fine Tuned	Accuracy	F1 Score
Trained on Emotions	Emotions Evoked	Conv4	67.8%	**0.76**
Trained on Emotions	Emotions Evoked	Conv4 then Conv5	**69.2%**	0.75

Table 3. *Sequential Training of the fourth convolutional block of pre-trained resnet50 followed by the fifth convolutional block on emotions only*

In this setting, we can observe that the highest F1 score is achieved by training only the fourth convolutional block on emotion without training the last convolutional block. This indicates that there are ways to optimize the model further to enhance its learning of emotion classification before including the information introduced by the genre contextual cues.

This result does not conflict with our previous findings that confirmed our hypothesis regarding the positive impact of genre on emotion classification. We consistently observed improvement when introducing genre information. However, this result suggests that including emotion in the fourth convolutional block is more beneficial than including genre in the same block in the sequential setting, which suggests that we can explore further configurations that maximize generalization performance using just emotion-evoked labels before we move to the step of including genre labels.

6. Conclusions

Our study confirms the hypothesis that incorporating genre information improves emotion classification accuracy in artworks. By leveraging transfer learning techniques and pre-trained ResNet50 models, we explored the relationship between art genres and the emotions they evoke. Our findings also suggest that sequential training of convolutional blocks yields better performance than simultaneous training. These insights pave the way for more accurate emotion classification in art, highlighting the importance of considering multiple experiment configurations to enhance emotion classification performance.

Moving forward, we plan to explore further model configurations to optimize emotion classification using emotion labels only. Once we establish the optimal spot achievable using just emotion labels, we will incorporate genre contextual cues to further leverage the contribution of genre to this space. Additionally, we aim to incorporate an explainability component like Gradient-weighted Class Activation Mapping (Grad-CAM) *(Selvaraju et al., 2019)* to understand how decisions are made in the modeling process. This understanding will better inform the incorporation of the most important contextual cues from genre in the emotion classification problem.

Roles

- Data Collection/Cleaning: Jiayi Zhou, Anastasiia
- Data Exploration/Preprocessing: Haochong Xia, Jiayi Zhou
- Experiment Design and Modelling: Osama
- Result Analysis: Haochong Xia, Osama
- Report Organization: Anastasiia, Haochong Xia, Jiayi Zhou, Osama

References

1. Achlioptas, P., Ovsjanikov, M., Haydarov, K., Elhoseiny, M., & Guibas, L. (2021, Jan 19). ArtEmis: Affective Language for Visual Art. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.48550/arXiv.2101.07396
2. Brumm, A., Oktaviana, A. A., & Burhan, B. (2021, Jan 13). Oldest cave art found in Sulawesi. *Science advances*. https://doi.org/10.1126/sciadv.abd4648
3. Chollet, F. (2018). *Deep Learning with Python*. Manning.
4. Clark, A. (2023, September). Impressionism After Impressionism. *Oxford Art Journal*, *46*(2). https://doi.org/10.1093/oxartj/kcad013
5. Dahiya, B. (2022). Baroque Art History. *Journal of Research in Humanities and Social Science*, *10*. https://www.questjournals.org/jrhss/papers/vol10-issue4/Ser-3/I10045563.pdf
6. Damiano, C., Gayen, P., Rezanejad, M., & Banerjee, A. (2023, April). Anger is red, sadness is blue: Emotion depictions in abstract visual art by artists and non-artists. *Journal of Vision*, *23*(4). https://doi.org/10.1167/jov.23.4.1
7. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 10.1109/CVPR.2009.5206848
8. DiMaggio, P. (1987). Classification in Art*. *American Sociological Review*, *52*(4), 440-445. https://doi.org/10.2307/2095290

9. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 10.1109/CVPR.2016.90
10. Mohammad, S. M., & Kiritchenko, S. (2018). WikiArt Emotions: An Annotated Dataset of Emotions Evoked by Art. *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*. http://saifmohammad.com/WebPages/wikiartemotions.html
11. Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 10.1109/CVPR.2014.222
12. Paasschen, J. v., Bacci, F., & Melcher, D. P. (2015). The Influence of Art Expertise and Training on Emotion and Preference Ratings for Representational and Abstract Artworks. *PloS one*, *10*(8). https://doi.org/10.1371/journal.pone.0134241
13. Selvaraju, R. R., Cogswell, M., & Das, A. (2019). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision (IJCV) in 2019*. https://doi.org/10.1007/s11263-019-01228-7
14. Serrao, F., Chirico, A., Gabbiadini, A., Gallace, A., & Gaggioli, A. (2024). Enjoying art: an evolutionary perspective on the esthetic experience from emotion elicitors. *Frontiers in psychology*. https://doi.org/10.3389/fpsyg.2024.1341122
15. Sinem, A., Castellano, G., & Digeno, V. (2022, August 07). Recognizing the Emotions Evoked by Artworks Through Visual Features and Knowledge Graph-Embeddings. *Image Analysis and Processing. ICIAP 2022 Workshops*. https://link.springer.com/chapter/10.1007/978-3-031-13321-3_12

16. *WikiArt*. (2010). WikiArt.org - Visual Art Encyclopedia. Retrieved April 21, 2024, from https://www.wikiart.org/
17. Zhao, S., Huang, Q., & Tang, Y. (2021, March 19). Computational Emotion Analysis From Images. https://ar5iv.labs.arxiv.org/html/2103.10798