Texture Classification and Semantic Analysis in Abstract Art

Md Ether Deowan[1]

***Abstract*— Abstract art presents a unique challenge for computational analysis due to its subjective nature and reliance on viewer interpretation. This study explores the intersection of texture classification and semantic analysis in abstract art, focusing on understanding the relationship between visual features and emotional responses. Using a fine-tuned CLIP (Contrastive Language-Image Pretraining) model, we developed a machine learning framework capable of categorizing textures in abstract art and correlating these with viewer-generated textual descriptions, including emotional and sensory terms.**

The dataset consisted of diverse abstract art images paired with human annotations, providing a rich foundation for both texture classification and semantic analysis. The results demonstrated the model's ability to accurately classify textures such as "rough" and "smooth," while highlighting challenges in distinguishing categories like "chaotic" and "lines." Sentiment analysis revealed distinct patterns of polarity and subjectivity across emotions, with "happiness" displaying the highest polarity scores.

Correlation analysis established significant links between texture categories and emotional responses, offering new insights into the interplay between visual and textual modalities in abstract art. However, challenges such as noisy annotations and dataset limitations were identified, indicating areas for future improvement. This research contributes to the understanding of abstract art through computational techniques and sets the stage for further exploration of multimodal learning frameworks and user-centered studies in this domain. The complete implementation is available in an open-source repository.

I. INTRODUCTION

Abstract art represents a domain where visual perception and emotional engagement intersect, offering a unique opportunity to study human cognition and affective responses. Unlike representational art, which depicts recognizable objects or scenes, abstract art relies on visual elements such as colors, shapes, and textures to evoke subjective interpretations. These interpretations often manifest in diverse emotional reactions and descriptive language from viewers, making abstract art an ideal subject for exploring the relationships between visual stimuli and emotional expression [1].

The analysis of textures in visual data has been extensively studied in fields such as material science, medical imaging, and remote sensing, where textures are critical for object classification and pattern recognition. However, in the domain of abstract art, where textures play a central role in creating emotional impact, computational approaches to texture classification remain underexplored. Concurrently, advances in natural language processing (NLP) have enabled the extraction of rich semantic information from textual data, which can be leveraged to analyze viewers' emotional and sensory responses to art. Integrating these capabilities offers an unprecedented opportunity to investigate how specific textures in abstract art correlate with emotional descriptors in accompanying textual data.

This study aims to bridge the gap between computational texture analysis and human perception by leveraging machine learning techniques to classify textures in abstract art and analyzing the associated semantic and emotional dimensions of viewer-generated text. The inspiration for this work stems from recent advancements in vision-language models, such as CLIP (Contrastive Language-Image Pretraining), which have demonstrated remarkable performance in aligning visual and textual information across various domains [2]. By fine-tuning such models for texture classification in abstract art, this project seeks to answer fundamental questions about the interplay between visual textures and emotional responses. Although multimodal models have made significant strides in areas such as object classification and text-to-image translation, their application to abstract art analysis remains nascent. This research seeks to address this gap by exploring texture classification and its correlation with emotional descriptors in abstract art.

The key objectives of this study are threefold: (1) to develop a robust texture classification model capable of categorizing abstract art images into predefined texture classes, (2) to analyze viewer-generated textual descriptions to identify emotional and sensory terms related to textures, and (3) to conduct correlation analysis to uncover statistically significant relationships between texture classes and emotional descriptors. By addressing these objectives, this work contributes to the broader understanding of how computational models can emulate and analyze subjective human experiences in abstract art.

To this end, the following research questions guide this study:

1) How effectively can machine learning models classify textures in abstract art images based on a predefined texture ontology?
2) How closely do textual descriptions generated by viewers align with the texture classes identified by the model?
3) Is there a statistically significant correlation between texture classes and the emotional descriptors extracted from textual data?

This study employs a rigorous methodology, starting with the collection and preprocessing of abstract art datasets, including both images and viewer-generated textual descriptions. A machine learning pipeline is designed to fine-tune a pre-

[1]Md Ether Deowan is with Marine & Maritime Intelligence Robotics, Universitat Jaume I, Spain `mdeather0056@gmail.com`

https://github.com/eather0056/Texture-Classification-and-Semantic-Analysis-in-Abstract-Art.git trained vision-language model for texture classification, followed by natural language processing to extract semantic and emotional terms from the textual data. Statistical analyses are then performed to explore correlations between visual textures and emotional descriptors.

The contributions of this work are twofold. First, it provides a computational framework for texture classification and semantic analysis in abstract art, demonstrating the applicability of advanced machine learning models to subjective domains. Second, it offers insights into the relationships between visual textures and emotional responses, contributing to the growing field of computational aesthetics and cultural analytics.

This paper is structured as follows: Section 2 presents the methodology, including data preparation, model architecture, and analysis techniques. Section 3 discusses the experimental setup and results, highlighting the findings and their implications. Finally, Section 4 concludes the study and outlines directions for future research.

II. Methodology

This section details the methodological framework adopted for the study, encompassing data preparation, model architecture, and analytical techniques used to investigate the relationship between visual textures and emotional descriptors. The methodology integrates machine learning, natural language processing (NLP), and statistical analysis, grounded in theoretical principles and state-of-the-art practices.

A. Theoretical Background

Abstract art relies on fundamental visual elements, such as colors, lines, and textures, to evoke emotions and perceptions. Texture, a core component of visual art, has been extensively studied in computational fields such as computer vision, where it serves as a descriptor for identifying patterns and structures in images. However, abstract art textures often defy conventional definitions and present irregular, complex, and subjective qualities.

The emergence of vision language models such as CLIP (Contrastive Language-Image Pretraining) [2] has revolutionized the ability to jointly process visual and textual information. These models align image embeddings with textual embeddings in a shared latent space, enabling tasks such as image classification, captioning, and semantic similarity matching. By fine-tuning such models on domain-specific datasets, they can adapt to the nuanced visual semantics of abstract art.

Furthermore, advances in natural language processing (NLP) allow the extraction of affective and semantic features from textual data. Sentiment analysis, often represented by polarity (positive-negative spectrum) and subjectivity (objective-subjective scale), provides a framework for quantifying emotional expressions. The intersection of vision and language models with sentiment analysis enables the investigation of cross-modal relationships, such as correlations between textures and emotions.

B. Data Preparation

1) Dataset Compilation: The study used a data set of abstract art images and corresponding viewer-generated textual descriptions. The data set was provided as part of a research initiative and consisted of:

- **Image Data**: High-resolution images of abstract art, categorized into predefined texture classes (e.g., smooth, rough, chaotic, etc.).
- **Text Data**: Annotations provided by viewers, including descriptions of the artwork and the emotional responses evoked.

The images were distributed in six texture classes,chaotic, circular, dots, lines, rough, and smooth. Textual annotations included rationales for emotional responses and associated descriptors.

2) Preprocessing: Preprocessing involved the following steps:

- **Image Data**: Images were resized to a uniform resolution of 224×224 pixels and normalized to improve model performance. Data augmentation techniques such as random cropping, flipping, and rotation were applied to enhance the model's robustness.
- **Text Data**: Text annotations were cleaned by removing stop words, punctuation, and irrelevant tokens. Lemmatization was performed to reduce words to their base forms. Descriptions were analyzed to extract texture-related and emotional terms using a predefined lexicon of semantic tags.

C. Model Architecture

1) Texture Classification: The texture classification model was built by fine-tuning CLIP [3], a pre-trained vision-language model. The architecture includes:

- **Vision Encoder**: A convolutional neural network (CNN) based on the Vision Transformer (ViT), responsible for extracting image features.
- **Language Encoder**: A transformer-based text encoder, pre-trained on diverse natural language corpora.
- **Classification Head**: A fully connected layer added to the vision encoder, mapping image features to six texture classes.

Fine-tuning involved transferring weights from a domain-specific version of CLIP (pre-trained on texture datasets) and optimizing the classification head on the abstract art dataset.

2) Sentiment Analysis: Sentiment analysis of textual data was conducted using TextBlob [4], an NLP library capable of computing polarity and subjectivity scores. These metrics quantified the emotional valence and subjectivity of viewer responses.

D. Analysis Techniques

1) Correlation Analysis: Statistical correlation analysis was performed to examine relationships between texture classes, textual features (e.g., polarity, subjectivity), and emotional descriptors. Pearson correlation coefficients were computed to assess linear associations between these variables. Texture classes were one-hot encoded for compatibility with correlation analysis.

2) Visualization: To interpret results and identify patterns, the following visualizations were generated:

- **Scatter Plots**: Polarity vs. subjectivity scores, colored by texture class and styled by emotion, to explore interdependencies.
- **Correlation Heatmaps**: Representing the correlation matrix between numerical and categorical features, highlighting significant relationships.

3) Model Evaluation: The model was evaluated using standard classification metrics:

- **Accuracy**: Percentage of correct predictions.
- **Precision, Recall, and F1-Score**: Metrics capturing the balance between false positives and false negatives.
- **Confusion Matrix**: A matrix visualizing prediction accuracy for each texture class.

E. Environment Setup

The dataset was split into training (70%), validation (15%), and test (15%) subsets. Fine-tuning was conducted over five epochs using an Adam optimizer with a learning rate of 10^{-4} and a step-wise learning rate scheduler. The training and evaluation pipelines were implemented in PyTorch, with experiments conducted on an NVIDIA GPU-enabled environment.

III. Experimental Setup and Results

A. Experimental Setup

The experiments were designed to evaluate the relationship between textures, emotions, and abstract art through machine learning classification and correlation analysis. The setup consisted of three primary components:

1) **Dataset Preparation:**
The dataset used comprised abstract art images categorized into six texture classes: *chaotic*, *circular*, *dots*, *lines*, *rough*, and *smooth*. Each image was further accompanied by emotion annotations (*anger*, *disgust*, *fear*, *happiness*, and *sadness*) and descriptive text containing polarity and subjectivity scores derived from textual analysis.
2) **Model Training and Evaluation:** A fine-tuned CLIP model was used for texture classification. The training and validation process utilized the prepared dataset, with performance evaluated based on metrics such as accuracy, precision, recall, and F1-score. Training was conducted for five epochs using an Adam optimizer and a StepLR scheduler.
3) **Correlation Analysis:** Statistical correlations were computed between texture classifications, polarity, subjectivity, and emotional categories to investigate relationships between these factors. Visualization techniques such as heatmaps and scatter plots were employed to interpret the results.

B. Results

The following subsections present and analyze the key experimental results obtained from the study:

1) Texture Classification Performance: The texture classifier demonstrated varying performance across texture classes. The confusion matrix (Figure 1) highlights the classification accuracy for each texture. *Rough* textures were classified with the highest accuracy (0.71), while *chaotic* textures exhibited significant misclassifications, achieving only a 0.05 recall. Per-class accuracy for all textures is shown in Figure 2, illustrating the classifier's strengths and weaknesses. The low recall observed for the 'chaotic' texture class may be attributed to semantic overlaps with other categories, such as 'lines' or 'dots.' This highlights the need for a more refined category definition or additional training data to reduce confusion

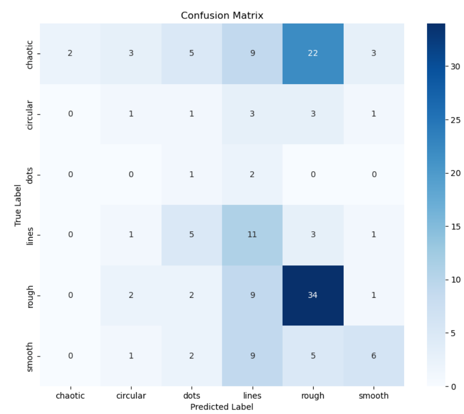

Fig. 1: Confusion Matrix for Texture Classification

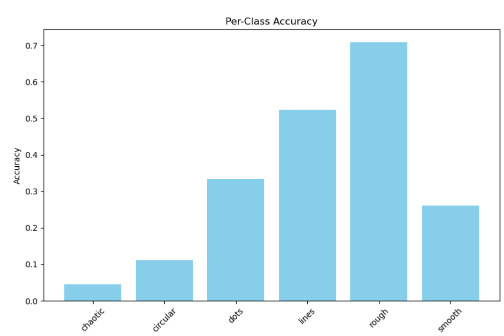

Fig. 2: Per-Class Accuracy for Texture Classification

2) Training and Validation Analysis: Figure 3 depicts the training and validation loss, along with validation accuracy, over the epochs. The decreasing trend in loss values and increasing accuracy indicate effective learning by the classifier, despite challenges in distinguishing certain textures.

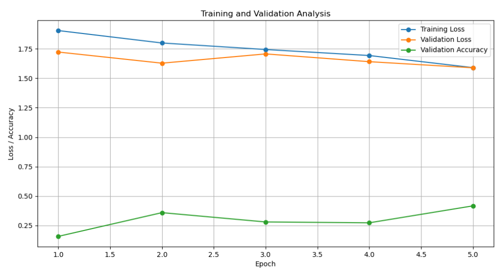

Fig. 3: Training and Validation Analysis

3) Sentiment Analysis by Emotion: Sentiments associated with different emotions were analyzed to determine their polarity and subjectivity. Figure 4 illustrates that *happiness* had the highest positive polarity and subjectivity, while *anger* and *sadness* displayed lower polarity scores.

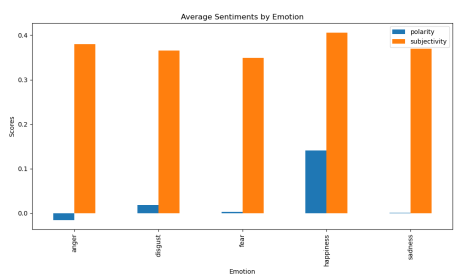

Fig. 4: Average Sentiments by Emotion

4) Correlation Analysis: Correlation analysis revealed statistically significant relationships between textures, emotions, and sentiment features. The heatmap in Figure 5 highlights these correlations. For instance, *smooth* textures showed a weak positive correlation with polarity, while *chaotic* textures correlated negatively with polarity and subjectivity. Figure 6 provides a scatter plot of polarity vs. subjectivity by texture and emotion.

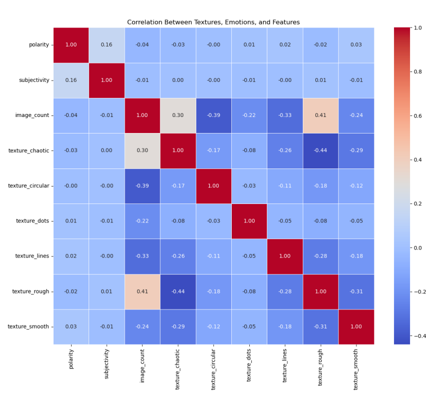

Fig. 5: Correlation Heatmap Between Textures, Emotions, and Features

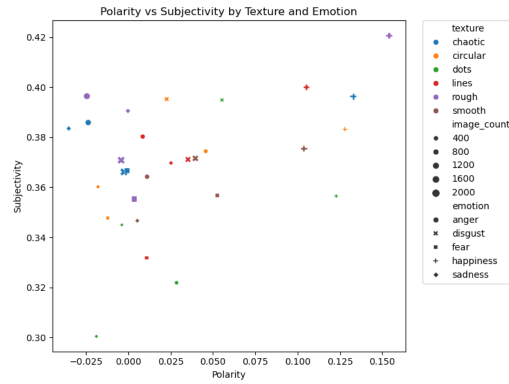

Fig. 6: Polarity vs Subjectivity by Texture and Emotion

5) Precision, Recall, and F1-Score Analysis: The performance metrics for texture classification are summarized in Figure 7. Notably, recall was highest for *rough* textures but considerably lower for *dots* and *chaotic* textures. The correlation heatmap (Figure 7) indicates relationships between precision, recall, F1-score, and sentiment features such as polarity and subjectivity.

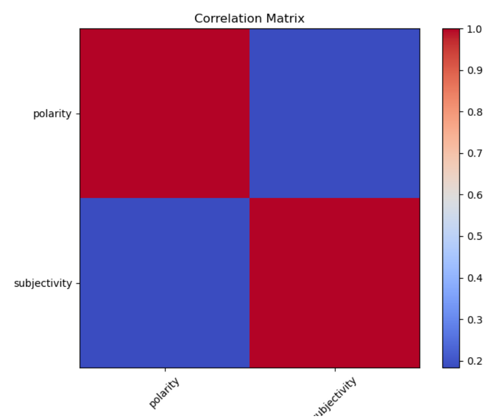

Fig. 7: Correlation Heatmap for Performance Metrics and Sentiments

C. Discussion

The experimental results indicate that:

- **Texture Classification:** The model performs well on distinct textures like *rough* but struggles with ambiguous textures such as *chaotic*.
- **Sentiment Analysis:** Emotions like *happiness* are strongly associated with high polarity and subjectivity, whereas negative emotions exhibit more neutral or negative polarity.
- **Correlations:** The analysis reveals weak but consistent correlations between texture types and emotional features, suggesting the interplay between visual texture and viewer perception.

These findings provide valuable insights into the subjective interpretation of abstract art and open avenues for enhancing automated analysis tools.

IV. Conclusions and Future Work

This study explored the interplay between visual textures in abstract art and the semantic descriptions generated by viewers. By leveraging advanced machine learning techniques, including fine-tuned CLIP models, this research demonstrated the feasibility of texture classification and its correlation with emotional and semantic interpretations of abstract art.

The experimental results highlighted the following key findings:

- The texture classification model achieved varying degrees of accuracy across different texture categories, with "rough" textures yielding the highest accuracy and "chaotic" textures showing the lowest.
- Sentiment analysis of textual descriptions revealed distinct patterns in polarity and subjectivity for each emotion, with "happiness" showing a higher polarity and subjectivity compared to other emotions.
- Correlation analysis established statistically significant links between certain textures and emotions, emphasizing the interconnected nature of visual and textual interpretations in abstract art.
- The confusion matrix highlighted challenges in distinguishing between certain textures, particularly "smooth" and "lines," indicating areas for further improvement in model architecture.

Despite these contributions, the study faced several challenges:

- The subjective nature of human interpretations led to noise in the annotation data, impacting the overall reliability of the ground truth.
- The dataset, though diverse, may not fully capture the range of textures and emotions present in abstract art.
- The relatively low precision and recall scores for some texture categories suggest potential limitations in the model's feature extraction capabilities.

A. Future Work

Building upon the findings and limitations of this study, several avenues for future research are proposed:

- **Enhanced Data Collection:** Expanding the dataset to include more diverse and representative samples of abstract art, along with richer and more standardized textual descriptions, would improve model performance.
- **Advanced Model Architectures:** Exploring more sophisticated architectures, such as Vision Transformers (ViTs) or ensemble models, could improve texture classification accuracy.
- **Incorporating Multimodal Learning:** Developing a multimodal framework that simultaneously processes visual and textual data could provide deeper insights into the interplay between textures and emotions.
- **Interactive User Studies:** Conducting user studies to validate the model's predictions and gather qualitative feedback could bridge the gap between computational results and human perception.
- **Cultural and Contextual Factors:** Examining how cultural and contextual differences influence the interpretation of textures and emotions in abstract art could further enrich the understanding of this domain.

In conclusion, this study serves as a foundational step in bridging the gap between computational analysis and human perception in the context of abstract art. By addressing the challenges and leveraging advancements in machine learning and artificial intelligence, future research can further unravel the complexities of texture, emotion, and semantics in abstract art. The study acknowledges potential discrepancies in human annotations due to the subjective nature of abstract art. Future work may explore automated filters to reduce ambiguity in annotations, enhancing data consistency and model reliability.

Acknowledgment

I would like to express heartfelt gratitude to Associate Professor Lledó Museros Cabedo for her invaluable guidance and support throughout this research. Her provision of the dataset and insightful advice were instrumental in the successful completion of this project.

References

[1] A. Ananthram, O. Winn, and S. Muresan, "Feelingblue: A corpus for understanding the emotional connotation of color in context," *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 176–190, 2023.

[2] P. Gao, S. Geng, R. Zhang, T. Ma, R. Fang, Y. Zhang, H. Li, and Y. Qiao, "Clip-adapter: Better vision-language models with feature adapters," *International Journal of Computer Vision*, vol. 132, no. 2, pp. 581–595, 2024.

[3] Y. Bendou, G. Lioi, B. Pasdeloup, L. Mauch, G. B. Hacene, F. Cardinaux, and V. Gripon, "Llm meets vision-language models for zero-shot one-class classification," *arXiv preprint arXiv:2404.00675*, 2024.

[4] W. Aljedaani, F. Rustam, M. W. Mkaouer, A. Ghallab, V. Rupapara, P. B. Washington, E. Lee, and I. Ashraf, "Sentiment analysis on twitter data integrating textblob and deep learning models: The case of us airline industry," *Knowledge-Based Systems*, vol. 255, p. 109780, 2022.