COMP4423 – Assignment 1

LIU Qianli 20080796D

Feel free to visit the github repository to access the datasets used in the project!

Liu-KM/COMP4423-Computer-Vison: PolyU 2024 sem 2 COMP4423 (github.com)

1. Introduction

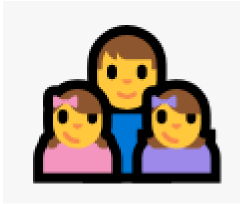
This report presents a concise overview of a project conducted for the COMP 4423 Computer Vision course, focusing on the construction and utilization of an extensive emoji database for classification purposes. By categorizing emojis into 26 unique groups based on various attributes and applying machine learning classifiers, the project achieved an impressive accuracy rate of over 97%. The classifier's robustness was further validated through testing with Emoji Kitchen images and a Kaggle facial expression dataset, revealing both strengths and limitations. Notably, a shift to deep learning methods demonstrated exceptional performance improvements. Key contributions include:

- The creation of a diverse emoji database, categorizing emojis into 26 distinct groups based on mood, type, number of people, and items, among other criteria.
- The development and application of a machine learning classifier that achieved an accuracy rate exceeding 97% on the emoji dataset.
- A comparative analysis of classifier performance on external datasets, highlighting the strengths and limitations of machine learning approaches.
- The successful transition to a deep learning methodology, which demonstrated superior efficacy in emoji classification and broader facial expression recognition tasks.

2. Method

2.1 Building the dataset

The initial phase of constructing the dataset encountered significant challenges, particularly with the encoding and rendering of complex emojis. An initial attempt to generate emoji images through Python, utilizing font files as both input and output, was hindered by the inherent limitations of the emoji encoding mechanism. Complex emojis, such as the family emoji (👨‍👩‍👧), were consistently rendered as separate, individual figures rather than a single, cohesive image. Despite extensive trials with numerous font files, this approach failed to produce the desired outcomes, leading to its eventual abandonment.



In search of a viable alternative, the project turned to an external resource, specifically the Noto Emoji fonts (github.com). This repository offered a collection of pre-rendered, colored emoji images, which provided a suitable foundation for the dataset. Selecting from this assortment, a curated set of emoji images was compiled to form the initial dataset. To further diversify and enrich the database, additional emojis were sourced from Emoji Kitchen, a platform known for its expansive range of emoji variations. This strategic integration resulted in a comprehensive dataset comprising 1616 emoji images (Appendix A), systematically categorized into 26 distinct groups based on criteria such as mood, type, number of people, and items.

This methodological pivot not only overcame the encoding challenges faced in the early stages but also established a robust and versatile emoji dataset. The diversity and breadth of the dataset played a pivotal role in the subsequent phases of the project, particularly in the training and evaluation of the machine learning classifier.

2.2 Image Processing

During the image processing stage, several crucial steps were undertaken to prepare the emoji images for classification. The first step involved standardizing all images to a uniform size of (72 \times 72) pixels. This standardization is vital for ensuring that all images are processed and analyzed under consistent conditions, eliminating size as a variable. Following the resizing, a suite of feature extraction methods was employed to capture various aspects of the images:

1. **Canny Edge Detection**: This algorithm identifies the edges in an image by finding areas with strong intensity gradients. The advantage of using Canny edge detection is its ability to capture the outline of emojis, providing a clear distinction between different shapes and forms, which is crucial for categorizing emojis based on their visual contours.
2. **Colour Histogram**: A colour histogram represents the distribution of colors in an image. By using this method, the model can understand the color composition of emojis, which is particularly useful for distinguishing emojis that differ primarily in color.
3. **Local Binary Patterns (LBP) Features**: LBP is a texture descriptor that compares each pixel with its surrounding pixels. It's beneficial for capturing the texture patterns of emojis, aiding in differentiating emojis with similar colors but different textures.
4. **Scale-Invariant Feature Transform (SIFT) Features**: SIFT extracts key points and their descriptors, which are invariant to image scale, rotation, and partially invariant to change in illumination. This robustness makes SIFT features highly effective for recognizing emojis that may appear in varied orientations and sizes.
5. **Histogram of Oriented Gradients (HOG) Features**: HOG focuses on the structure or shape of an object by capturing the distribution of direction gradients. It is particularly adept at distinguishing between emojis with similar colors but different forms or postures.

These methods were chosen for their combined ability to capture a comprehensive range of visual characteristics, from basic color and shape to complex textures and gradients, thus enhancing the classifier's accuracy.

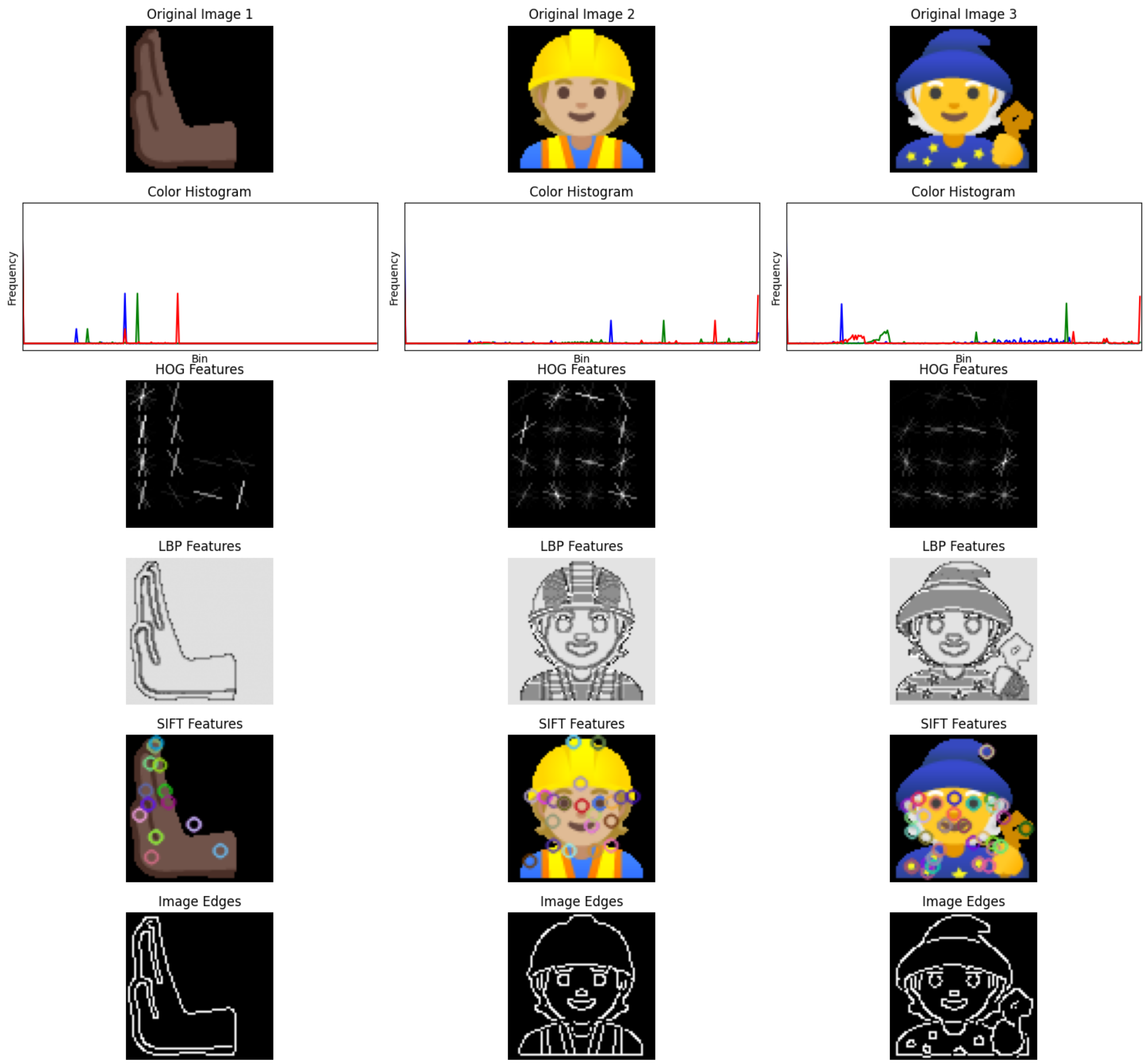
Following feature extraction, the dataset was divided into a training set and a test set using an 80:20 split. This division ensures that the model can be trained on a substantial portion of the data while reserving a separate subset for unbiased evaluation.

2.3 Constructing the Classifier

Given the assignment's stipulation to eschew deep learning models, the project explored various machine learning classifiers to determine the most effective approach for emoji classification. Each classifier offers distinct advantages and operates based on different principles:

1. **Support Vector Classifier (SVC)**: Utilizes hyperplanes to categorize data into classes. The strength of SVC lies in its effectiveness in high-dimensional spaces, making it suitable for complex feature sets derived from image processing.
2. **Decision Tree**: Models decisions and their possible consequences as a tree. Its simplicity and interpretability are advantageous, especially for datasets where relationships between features can be hierarchically structured.
3. **Random Forest**: An ensemble of decision trees, designed to improve classification accuracy through voting or averaging. Its robustness and ability to handle overfitting make it a strong candidate for complex classification tasks.
4. **K-Nearest Neighbors (KNN)**: Classifies data based on the closest training examples in the feature space. KNN is renowned for its simplicity and effectiveness in cases where the decision boundary is irregular.
5. **XGBoost**: An implementation of gradient boosted decision trees designed for speed and performance. XGBoost is favored for its efficiency and the capacity to handle sparse data.
6. **Logistic Regression**: Despite its name, it's a linear model for classification rather than regression. It is particularly useful for binary classification tasks and can be extended to multiclass classification.

These machine learning models were selected to provide a broad evaluation of different approaches, from ensemble methods and nearest neighbors to linear classifiers, thereby ensuring a comprehensive assessment of their suitability for emoji classification.



The table below presents a comprehensive summary of the classifiers' performance, delineated by both training and testing accuracy percentages across different feature sets:

Classifier	ImmageArray	Edge	HOG	LBP	ColorHist
Support Vector Classifier (SVC)	***98.39%***/90.43%	97.40%/94.44%	***98.39%***/94.75%	93.07%/93.52%	94.18%/87.96%
Decision Tree	93.07%/85.49%	91.09%/93.52%	91.34%/89.81%	90.84%/91.05%	88.74%/83.95%
Random Forest	97.03%/93.21%	96.41%/95.37%	97.40%/97.22%	96.41%/95.68%	95.42%/90.74%
K-Nearest Neighbors (KNN)	95.54%/92.90%	95.92%/94.75%	96.04%/93.21%	94.43%/95.37%	88.37%/81.17%
XGBoost	96.78%/93.21%	96.29%/94.44%	96.66%/95.06%	93.44%/94.14%	95.42%/89.20%
Logistic Regression	**98.64%**/84.88%	97.40%/91.98%	97.77%/92.90%	87.25%/88.58%	93.32%/84.57%

This table showcases a noteworthy variance in performance across classifiers and feature extraction methods. Notably, the Logistic Regression model achieved the highest training accuracy with the ImageArray feature set, while the Support Vector Classifier (SVC) and the same feature set tied for the best testing accuracy, indicating a robust ability to generalize from the training data to unseen images. The consistently high performance of the HOG feature set across multiple classifiers, particularly with the Random Forest classifier achieving a testing accuracy of 97.22%, underscores the effectiveness of texture and shape-oriented features in emoji classification. In contrast, the Color Histogram feature set, despite its intuitive appeal for emoji classification given their vibrant and distinctive color schemes, generally resulted in lower accuracy rates.

2.4 Test on Unseen Data

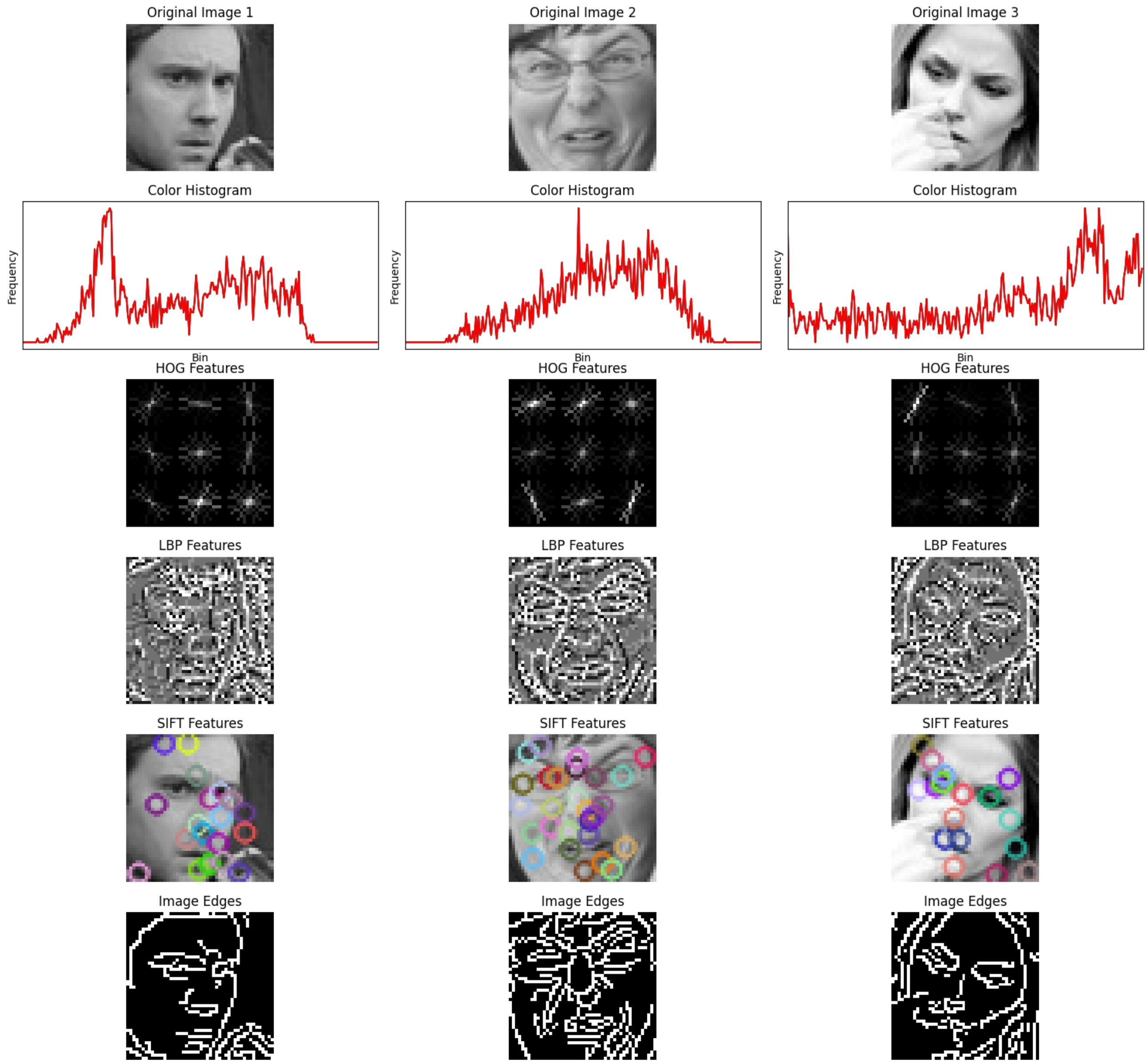
An evaluation on unseen emoji images revealed a notable trend among the classifiers: a propensity to classify diverse emojis into the "Crush" category. This observation underscores the challenges in achieving accurate predictions for emojis that significantly differ from those in the training dataset. The table below highlights this tendency across various classifiers:

				
SVC	Crush	Happy	Crush	Happy
DecisionTree	Crush	Happy	Crush	Crush
RandomForest	Crush	Happy	Crush	Crush
KNN	Crush	Happy	Crush	Hand
XGBoost	Crush	Crush	Crush	Crush
LogisticRegression	Crush	Happy		Symbol

This succinct test illustrates the limitations and potential biases in the classifiers when faced with novel data, pointing to the importance of a diverse training set for improved generalization.

2.5 Extension to the FER Dataset

The **Facial Expression Recognition (FER) task** represents a significant and challenging area within the field of computer vision, focusing on the identification of human emotional states through facial expressions. This task not only tests the limits of classification models in understanding human emotions but also explores the applicability of these models beyond their original domains. For this project extension, the FER2013 dataset, provided by the "Challenges in Representation Learning: Facial Expression Recognition Challenge" competition on Kaggle, was utilized. The FER2013 dataset comprises grayscale images of facial expressions, making it an ideal benchmark for evaluating the adaptability and performance of the developed emoji classification models on a fundamentally different and more complex dataset.



The performance of the classifiers, when applied to the FER dataset, was evaluated using a range of feature extraction methods, as summarized in the table below. The entries represent accuracy scores, with the first percentage reflecting training accuracy and the second indicating testing accuracy.

Classifier	**ImageArray**	**ColorHistogram**	**HOG**	**LBP**	**ImageEdge**
Support Vector Classifier (SVC)	25.91%/24.19%	25.67%/24.19%	31.10%/30.55%	25.08%/24.19%	*24.21%/26.98%*
Decision Tree	23.62%/22.80%	20.62%/20.71%	24.45%/24.89%	19.82%/20.19%	*19.16%/21.67%*
Random Forest	34.45%/30.81%	26.82%/25.41%	37.86%/33.77%	24.76%/24.89%	*27.52%/28.63%*
K-Nearest Neighbors (KNN)	23.06%/24.37%	20.38%/20.71%	34.69%/26.54%	19.64%/19.32%	*14.56%/20.80%*

Classifier	ImageArray	ColorHistogram	HOG	LBP	ImageEdge
XGBoost	35.21%/27.68%	24.94%/24.98%	**40.93%**/30.20%	*23.16%/23.32%*	28.04%/26.11%
Logistic Regression	24.97%/27.07%	25.39%/25.07%	32.67%/31.51%	25.88%/23.24%	*23.89%/27.85%*

From the table, it is evident that traditional machine learning methods face challenges in accurately recognizing facial expressions from the FER dataset, with the highest accuracy achieved using Histogram of Oriented Gradients (HOG) features in conjunction with the XGBoost classifier. This outcome underscores the complexity of emotion recognition tasks, especially when compared to the relatively higher accuracy rates observed in emoji classification. The superior performance of HOG features highlights their effectiveness in capturing essential shape and contour information in facial expressions, a critical aspect of emotion recognition. Despite the modest success with HOG and XGBoost, the overall results signal the need for more sophisticated models or approaches, such as deep learning, which may offer improved capabilities for capturing the nuances of human emotional expressions.

3. Discussion and Conclusion

3.1 Brief project summary

This project embarked on an ambitious journey to navigate the complexities of emoji classification and extend its methodologies to facial expression recognition, culminating in a series of insightful findings and confronting inevitable limitations. The experimental outcomes—characterized by high accuracy in emoji dataset classification and notably lower accuracy in the FER dataset—align with the initial hypotheses. The simplicity of emoji features, contrasted with the richer, more nuanced features of human faces, inherently favored the emoji classification task. Yet, the facial expression recognition results fell short of expectations, with only a singular classifier surpassing the 40% accuracy mark, underlining the challenges posed by simple machine learning models in capturing the subtleties of human emotions.

3.1 Key Findings

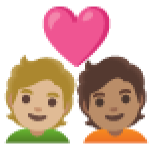
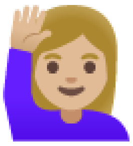
1. **High Emoji Classification Accuracy**: The project achieved significant success in emoji classification, attaining high accuracy rates. However, discrepancies in unseen data performance, particularly the misclassification issues surrounding the 'crush' category—despite its minimal representation in the dataset—highlight the complexities of predictive modeling and the influence of dataset composition on classifier performance.
2. **Variable Accuracy in Facial Expression Recognition**: The face emotion monitoring task demonstrated considerable variation in accuracy across different features, with HOG features emerging as the most indicative of emotional states. This underscores the critical role of feature selection in effectively capturing the essence of facial expressions.
3. **PCA Dimensionality Reduction**: The application of PCA for dimensionality reduction occasionally led to improvements in model accuracy. This suggests the presence of redundant information within the feature sets, pointing to the potential benefits of simplifying data representations to enhance model efficacy.

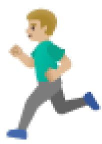
3.2 Limitations and Future Directions

1. **Complexity of Emoji Encoding**: The project faced challenges in processing and rendering complex emoji codes, limiting the scope for expanding the training dataset with more diverse emoji representations. Future work could explore advanced encoding and rendering techniques to overcome these hurdles.
2. **Abstract Nature of Emojis**: The inherent abstraction and variability among emojis posed challenges in creating a comprehensive and representative dataset. Manual dataset creation was successful but highlighted the need for more sophisticated methods to capture the broad spectrum of emoji types and categories.
3. **Potential of Deep Learning**: Given the limitations imposed by the assignment's scope, deep learning methodologies were not explored. However, the results suggest that deep learning could offer substantial advantages for both emoji classification and facial expression recognition. Future explorations could leverage deep learning to address the noted limitations, potentially harnessing convolutional neural networks (CNNs) and other advanced models to capture the complex patterns underlying both emojis and human facial expressions.

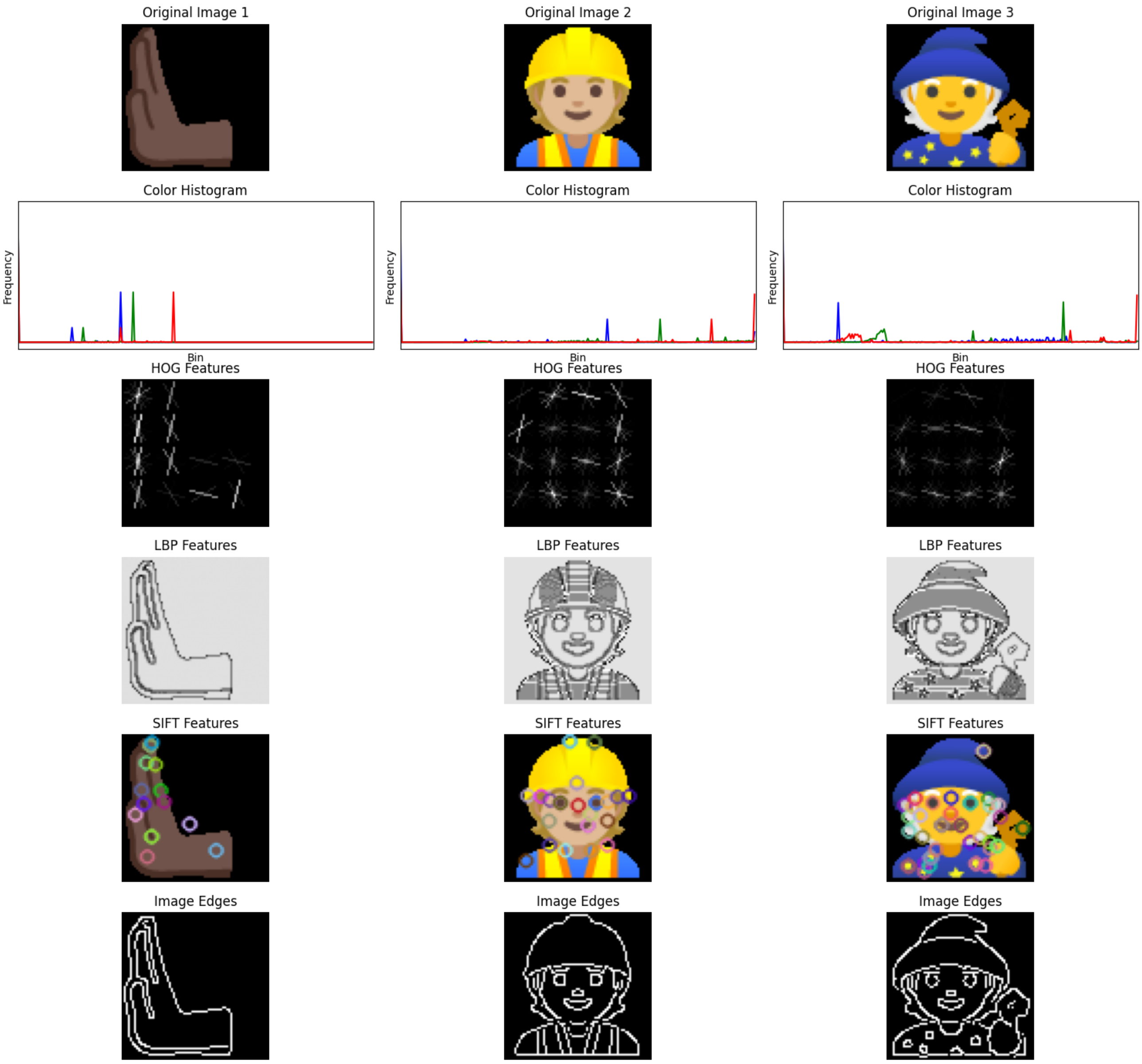
In conclusion, this project has laid a foundational understanding of the challenges and opportunities in emoji classification and facial expression recognition. While achieving commendable success in emoji classification, the venture into facial expression recognition has illuminated the path for future research, particularly highlighting the untapped potential of deep learning in transcending the observed limitations. The journey underscores the intricate balance between feature selection, model choice, and the nuanced characteristics of the data, setting the stage for further exploration in the rich and evolving domain of computer vision.

Appendix

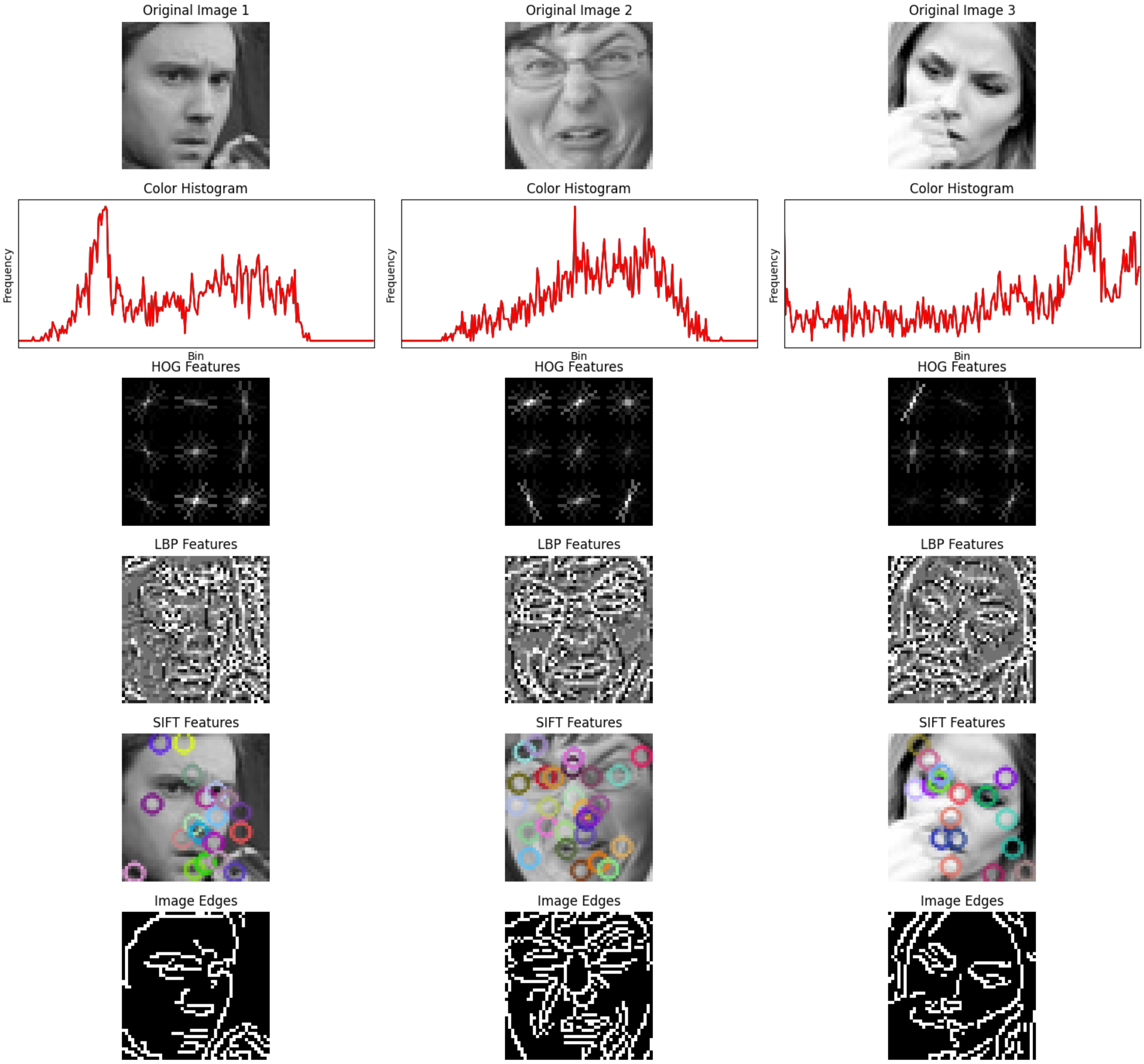
Class	Count	Image
hand	266	
two_people	210	
halfbody	168	
whole_person	144	

Class	Count	Image
symbol	120	
happy	96	
crush	76	
water_sport	72	
two_people_full	70	
head	68	
bicycle	36	
run	36	
clock	25	
heart	21	
police	18	

Class	Count	Image
vampire	18	
graduate_student	18	
fireman	18	
sapceman	18	
chef	18	
planet	18	
farmer	18	
worker	18	
wizard	18	
fruit	16	
Santa	12	



Original Image 1
Original Image 2
Original Image 3
Color Histogram
Frequency
Bin
HOG Features
LBP Features
SIFT Features
Image Edges


Original Image 1
Original Image 2
Original Image 3
Color Histogram
Frequency
Bin
HOG Features
LBP Features
SIFT Features
Image Edges