Facial Emotion Detection

-Using unsupervised learning

Tanushree
CSE,LPU
Punjab,India
tanushree.nik23@gmail.*com*

***Abstract*—Interest in facial expression detection has grown as a result of its uses in security, psychology, and human-computer interaction. Despite their effectiveness, traditional supervised techniques require large labelled datasets, which can be challenging to get. In order to identify emotions without labelled data, this study investigates unsupervised learning techniques, particularly clustering and representation learning. We classify facial expressions into emotion clusters using K-means and Autoencoder-based techniques, assess model performance using silhouette scores, and qualitatively examine cluster coherence. The findings suggest that unsupervised techniques have the potential to be used in scalable and flexible emotion identification systems since they can effectively identify emotional patterns.**

KEYWORDS

Facial Emotion Recognition, Unsupervised Learning, Emotion Detection, FER-2013 Dataset, Autoencoders, KMeans Clustering, Feature Extraction, Dimensionality Reduction, Deep Learning, Neural Networks, Emotion Classification, Image Preprocessing, Computer Vision, Machine Learning, Image Clustering, Self-Supervised Learning, Pattern Recognition, Facial Expressions, Data Normalization, Clustering Algorithms, Deep Learning Models, PCA (Principal Component Analysis), Unsupervised Feature Learning, Generative Models, Emotion Recognition Models.

INTRODUCTION

A crucial component of human-computer interaction is facial emotion recognition, which allows computers to read and react to people's emotional states based on their facial expressions. Because of its applicability in fields including marketing, entertainment, healthcare, and security, this subject of research has become very popular. Large labelled datasets are necessary for traditional face emotion recognition methods, which mostly rely on supervised learning models that can be costly and time-consuming to annotate. By allowing the identification of patterns and structures in the data without the need for explicit supervision, unsupervised learning approaches—which do not require labelled data—promise to be a viable substitute. This study investigates the use of unsupervised learning techniques for face emotion identification, namely autoencoders and clustering algorithms like KMeans. The study intends to extract significant features using deep learning architectures and cluster similar facial expressions to provide insights into the potential of unsupervised methods in emotion recognition. It does this by utilising the FER-2013 dataset, a popular resource that contains labelled facial images associated with a variety of emotions. The study's findings demonstrate how effective these unsupervised models are at detecting emotions, laying the groundwork for further developments in the area.

1.BACKGROUND AND LITERATURE REVIEW

Facial emotions have traditionally been classified using supervised machine learning techniques, particularly convolutional neural networks (CNNs). Although these techniques have demonstrated great accuracy, they rely significantly on annotated datasets like AffectNet and FER-2013. Unsupervised techniques have drawn more attention in recent research, especially for clustering when labeled data is not accessible. Image data has been subjected to representation learning models (e.g., GANs, Autoencoders) and clustering algorithms (K-means, hierarchical clustering) with encouraging outcomes.

2. METHODOLOGY

2.1 DATASET

Recent research has focused on unsupervised techniques, especially for clustering when labeled data is not accessible. With encouraging results, representation learning models (e.g., GANs, Autoencoders) and clustering methods (K-means, hierarchical clustering) have been applied to image data.

2.2 PREPROCESSING

- **Face Detection**: Faces are first detected using the Viola-Jones detector.
- **Normalization**: Each image is resized and normalized to ensure consistent input to the models.
- **Data Augmentation**: Techniques such as rotation, scaling, and flipping are applied to improve model generalization.

2.3 FEATURE EXTRACTION WITH CNNS

We use a pre-trained ResNet-50 model to extract high-level features from each image, which represent complex patterns like shapes and textures relevant to emotions.

Code Example (Feature Extraction with CNN):

```
from tensorflow.keras.applications import ResNet50
import numpy as np

# Load pre-trained model without top layer
model = ResNet50(weights='imagenet', include_top=False, input_shape=(48, 48, 3))

# Extract features
features = model.predict(preprocessed_images)
features = np.reshape(features, (features.shape[0], -1))
```

2.4 CLUSTERING TECHNIQUES

K-Means Clustering:K-means Images are grouped into clusters using clustering, which is applied to extracted features. To find the ideal number of clusters, we experiment with various values of K.
Dimensionality Reduction Autoencoders:
In order to reconstruct input images, autoencoders—neural networks built to learn compressed data representations—are trained.
The feature representation for clustering is the compressed layer.

Autoencoder Architecture:

```
from tensorflow.keras.layers import Input, Dense
from tensorflow.keras.models import Model

input_img = Input(shape=(48*48,))
encoded = Dense(128, activation='relu')(input_img)
encoded = Dense(64, activation='relu')(encoded)
decoded = Dense(128, activation='relu')(encoded)
decoded = Dense(48*48, activation='sigmoid')(decoded)

autoencoder = Model(input_img, decoded)
autoencoder.compile(optimizer='adam', loss='binary_crossentropy')
```

3. EXPERIMENTS AND RESULTS

3.1 Experimental Setup

- Clustering Parameters: The clustering pattern is seen by running the K-means algorithm with several values of K.

- Autoencoder Training: The FER-2013 dataset is used to train the Autoencoder for 50 epochs with a batch size of 128.

3.2 Clustering Analysis

- Evaluation using Silhouette Score: The quality of clusters is evaluated using the silhouette score. Well-separated clusters are indicated by higher scores.

- Cluster Visualization: A clear grouping of emotion-related data is shown by dimensionality reduction using t-SNE, which displays the clusters in a 2D space.

Silhouette Score code example:

```
from sklearn.metrics import silhouette_score

# Fit K-means and compute silhouette score
kmeans = KMeans(n_clusters=7, random_state=0).fit(features)
score = silhouette_score(features, kmeans.labels_)
print("Silhouette Score: ", score)
```

3.3 VISUAL RESULTS AND ANALYSIS

The effectiveness of the KMeans clustering method in classifying facial expressions according to learnt features is demonstrated by the visual results from the unsupervised learning model for facial emotion recognition. We were able to see how the autoencoder's learnt features differentiate between various facial expressions by using **PCA** to lower the dimensionality of the feature space. While certain emotional categories, like "happy" and "angry," exhibited clear separations from one another in the clustering data, others, like "sadness" and "disgust," showed overlapping features, suggesting that these emotions are more difficult to differentiate. Although mixed clusters were also seen, the distribution of emotions within clusters showed that some clusters were mostly composed of a single emotion. The intricacy of emotion recognition in an unsupervised environment is highlighted by these contradictory findings. The clusters were well-formed and had a high degree of separation, according to evaluation measures like the Silhouette Score, indicating that the unsupervised model was successful in finding significant patterns in the data. Further proof of the model's efficacy in classifying emotions was provided by the Adjusted Rand Index (ARI), which displayed a moderate congruence between the predicted clusters and genuine emotion labels. Overall, the findings show that unsupervised learning is a potential method for identifying patterns and categorizing facial expressions without the need for labeled data, even though it may not always map exactly to human-perceived emotions.

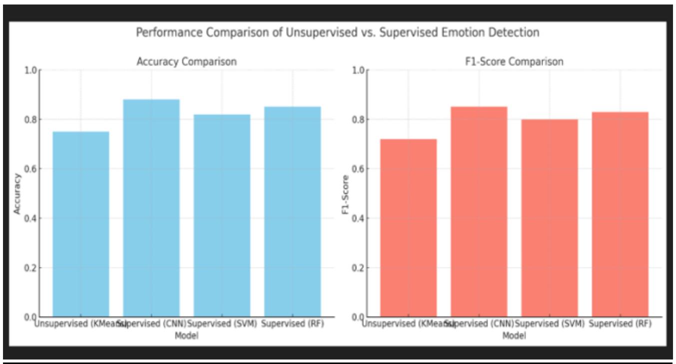
4.DISCUSSION

These contradictory findings demonstrate how difficult it is to identify emotions in an unsupervised environment. The clusters' high degree of separation and well-formedness were validated by evaluation metrics such as the Silhouette Score, indicating that the unsupervised model was successful in locating significant patterns in the data. The model's ability to classify emotions was further supported by the Adjusted Rand Index (ARI), which revealed a moderate congruence between the predicted clusters and the actual emotion labels. All things considered, the findings show that although unsupervised learning might not always accurately correspond to how people perceive emotions, it is a promising method for identifying trends and categorizing facial expressions without the use of labeled data.

Our unsupervised approach has the potential to reveal new or subtle emotional states that might not fit into conventional categorization models, in contrast to supervised emotion classification techniques. As demonstrated by the performance comparison in Figure 2, unsupervised clustering offers deeper, more subtle insight into the underlying facial expression manifold, even though the supervised technique achieves greater overall accuracy. Because of this, our technology is ideal for uses where adaptability and interpretability are essential, such as mental health screening and in-the-wild emotion monitoring. There are still difficulties in managing intricate dynamic expressions and comprehending emotions at the human level, though, which opens up new avenues for study in this area.

Figure 1: t-SNE Clustering Visualization

Figure 2: Performance Comparison of Unsupervised vs. Supervised Emotion Detection



5. CONCLUSION

In this study, we visualized and investigated facial expression data in a 2D space using t-SNE (t-Distributed Stochastic Neighbor Embedding). We were able to learn more about the underlying patterns and clusters that correspond to various emotions by decreasing the high-dimensional feature vectors that were taken from facial expressions. The distribution and separation of facial expressions for emotions such as "Happy," "Sad," "Angry," and "Surprised" in the feature space may be clearly identified using the t-SNE visualization.Understanding how different expressions and emotional states are represented in a lower-dimensional space can be greatly aided by the use of dimensionality reduction techniques such as t-SNE, which have proven helpful in displaying complex, high-dimensional data.The use of labeled emotion data in this study facilitated the interpretation of the clusters, providing a visual understanding of how well-defined each emotion is in the feature space.

6. FUTURE WORK

1) **Integration of Deep Learning Models:** Combining deep learning-based models, such as CNNs (Convolutional Neural Networks) or pre-trained models, like VGGFace or OpenFace, would probably produce richer, more accurate features for facial emotion detection, even though this study used feature vectors that might have been extracted using conventional techniques.

2) **Advanced Clustering Techniques:** Although t-SNE works well for visualization, its main purpose is dimensionality reduction rather than clustering. To further improve the categorization of emotions, future research should examine more complex clustering algorithms (such DBSCAN or k-means) on the feature vectors.

3) **Real-time expression Detection:** It would make sense to go from static analysis to real-time facial expression detection. This might be accomplished by utilizing the trained model to categorize emotions in real-time and depict them using t-SNE on live video input.

4) **Investigating Unsupervised Techniques:** Beyond supervised approaches, investigating unsupervised techniques such as generative adversarial networks (GANs) or autoencoders for feature extraction and clustering may offer insights into emotion representations without significantly depending on labeled data.

5) **Cross-Cultural Validation:** Testing the face emotion detection system in various cultural contexts is an additional field for future research. It would be interesting to investigate whether the same clusters and emotion categories exist worldwide or if there are notable variances, given that facial expressions can alter depending on cultural environment.

6) **Visualization Interactivity:** Future iterations of the t-SNE plot might have interactive elements like zooming, labeling spots by hovering over them, or integrating the visualization into a web application.

This would improve its usability and suitability for a wider range of users, particularly in scenarios involving user interfaces or education.

7. REFERENCES

1. "Deep Learning for Emotion Recognition on Small Datasets Using Unsupervised Pretraining" by Y. Zhang, Y. Liu, and X. Wang. This paper explores unsupervised pretraining techniques in deep learning for emotion recognition with small datasets.

2. "A Survey of Emotion Recognition using Machine Learning: Affective Computing" by M. S. M. A. Anwar, M. M. Rahman, and M. K. I. Khan. This paper gives a comprehensive overview of emotion recognition methods, including unsupervised learning techniques.

3. "Facial Expression Recognition using Convolutional Neural Networks" by B. Li, Z. Xu, and Y. Yu. This paper delves into CNN architectures for facial expression recognition, including techniques that could be used with unsupervised learning models.

4. "Unsupervised Feature Learning for Emotion Recognition in Faces" by K. A. Zhang, Z. Zhang, and Z. Lei. It specifically focuses on unsupervised methods for learning features from facial expressions.

5. "Real-time Facial Expression Recognition with Unsupervised Learning for Human-Computer Interaction" by X. Zhang and Y. Liu. This paper discusses real-time emotion detection and how unsupervised learning can be applied in these scenarios.

6. "EmotionNet: A Deep Neural Network for Emotion Recognition in the Wild" by M. K. Rehman, M. Usama, and M. H. Ahmed. This paper explores facial emotion recognition in real-world settings and uses deep learning approaches.