Facial Landmark-Based Real-time Facial Emotion Recognition with Video Vision Transformer Model

sen.anuvab, halder.mritunjoy

July 19, 2023

1 Proposal

Real-time facial emotion analysis has numerous applications in various industries, and enhancing its performance for sentiment analysis in videos can unlock even more possibilities. In this paper we propose a Video Vision Transformer as a realtime facial recognition model. The work is ought be extended to aid in the detection of emotional fluctuations in videos. The detection of emotional dynamics within video content will help in facilitating applications in fields such as video analysis, emotion recognition, and sentiment analysis. [1].

Detecting emotional fluctuation is a challenging task and detecting emotional situation from that has great scope in human psychology. This detection may include several steps: They are as follows

- Face Detection: Use a face detection algorithm (such as Haar cascades, Viola-Jones, or deep learning-based methods like MTCNN or SSD) to identify and locate faces in each frame of the video. The best currently is YOLO algorithm.
- Facial Landmark Detection: Employ facial landmark detection algorithms (e.g., dlib, OpenCV's facial landmark detector, or deep learning-based methods like FaceNet or Dlib's CNN face detector) to identify specific points on each detected face, such as the corners of the mouth, eyes, and eyebrows. There is a very famous algorithm known as posenet. PoseNet has some great scope of modification. Main Research Area
- Facial Expression Analysis: Extract facial features from the landmarks, such as the distance between eyebrows, mouth curvature, or eye openness. Apply machine learning techniques (e.g., SVM, random forests, or deep learning models) to classify these features into different emotion categories, such as happiness, sadness, anger, surprise, etc.

  This step involves training a model on a labeled dataset of facial expressions. Main Research Area
- Correlation with body language: Human emotion has a great effect on body language. If a correlation function can be created with human body language and face expression it would be much more specific and robust model. New Research Area
- Temporal Analysis: To detect emotion fluctuations, analyze the sequence of classified emotions over time. Techniques like sliding windows or temporal smoothing can be applied to capture temporal patterns and identify changes in emotions.
- Optional: Incorporating acoustic analysis and create UI for real time analysis.

References

[1] F. Ma, B. Sun, and S. Li, "Facial expression recognition with visual transformers and attentional selective fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.