A Real time Face Emotion Detection System Based on YOLO11

Review Paper on AI-A Real time Face Emotion Detection System Based on YOLO11

Miss.Bhivsane.P.P[1], Mr.S.G.Shah[2]

[1]*(Computer Science & Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar— bhivsanepallavi@gmail.com)*

[2]*(Computer Science & Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar— sandeepshah@dietms.org)*

Abstract— Human emotion recognition and behavioral analysis have become important research areas in Artificial Intelligence, Computer Vision, and Human-Computer Interaction. Recent advancements in facial landmark detection, machine learning, and deep learning have enabled real-time analysis of emotions, attention levels, fatigue, and user engagement through webcams and mobile cameras. This review paper examines existing approaches used for facial emotion recognition, gaze tracking, blink detection, head pose estimation, and behavioral analytics. Various methodologies including Media Pipe Face Mesh, Convolution Neural Networks (CNNs), Facial Action Coding System (FACS), and machine learning-based inference models are analyzed. The paper highlights current limitations and identifies opportunities for developing lightweight, real-time behavioral tracking systems. This paper proposes a real-time face emotion detection and behavior analysis system based on YOLO11, Media Pipe Face Mesh, and Convolution Neural Networks (CNN). The proposed framework performs high speed face detection using YOLO11 and extracts detailed facial landmarks through Media Pipe Face Mesh. A CNN-based emotion classifier is trained to recognize seven primary human emotions: angry, disgust, fear, happy, neutral, sad, and surprise. In addition to emotion recognition, the proposed system performs behavioral analysis including blink detection, gaze tracking, eyebrow movement analysis, mouth state detection, head pose estimation, yawn detection, and attention score calculation

Index Terms— *YOLO11, Emotion Recognition, Computer Vision, CNN, Deep Learning, Media Pipe, Face Mesh, Human Behavior Analysis, Artificial Intelligence, Real-Time Detection*

I. INTRODUCTION

Behavioral tracking systems aim to analyze human facial expressions, eye movements, gaze direction, and head orientation to understand emotional and cognitive states. Such systems are widely used in online learning, healthcare monitoring, recruitment platforms, driver monitoring systems, and customer engagement analysis. Traditional emotion recognition methods relied heavily on handcrafted features.

Whereas modern approaches use deep learning and facial landmark analysis for improved Accuracy and real-time performance.

Human emotions are an essential component of communication and social interaction. Understanding human emotional states enables intelligent systems to interact naturally with users and improve decision making processes. Facial expressions are among the most significant indicators of human emotions and provide valuable behavioral information.

In recent years, Artificial Intelligence and Deep Learning technologies have revolutionized facial emotion recognition systems. Emotion recognition systems are now widely used in healthcare monitoring, driver fatigue detection, intelligent tutoring systems, mental health assessment, security surveillance, customer feedback analysis, and human-computer interaction systems.

Traditional facial emotion recognition methods relied on handcrafted features such as Haar Cascade, Histogram of Oriented Gradients (HOG), and Local Binary Patterns (LBP). However, these methods were limited in performance and often failed under varying environmental conditions.

Human emotions play a critical role in communication, decision-making, and behavioral responses. With the evolution of AI, machines are now capable of recognizing emotions using image processing, audio signals, text sentiment, and physiological patterns. Emotion recognition enhances usability in systems such as: Healthcare monitoring E-learning platforms Smart surveillance Customer support Human–robot interaction.

The research recognition from nonverbal behaviors. The role of positive emotions has been strongly emphasized over the recent two decades with the growth of positive psychology which focuses on the study and the promotion of positive aspects of human life, such as well-being, happiness, and personal strengths. Some psychologists different ate various positive states, like pride, gratitude, and sympathy by analyzing their appraisals, causes, circumstances in which they may appear, as well as the expressive patterns of these emotions. Specifically, emotion theorists propose that appraisals are the underlying process by which emotions arise.

This process involves the interpretation of events or situations based on their impact, which is determined by their congruence or incongruence with an individual's internal states, goals, and beliefs. Consequently, we focus on a variety of positive emotional states and review existing recognition or detection models by considering the modalities, the datasets used, the circumstances in which the data were collected, as well as the machine learning (ML) and deep learning (DL) techniques employed. To the best of the authors 'knowledge, although several surveys exist in the field of Affective Computing none specifically focus on the diversity of positive emotional states, making this work a unique contribution.

II. LITERATUREREVIEW

1. Positive Emotion Recognition—A Survey of Computational Models (2025) Summary: In this paper, we provide a systematic survey of works on positive emotion recognition and detection. We queried major research paper databases for 12 different emotional states: admiration, amusement, awe, compassion, contentment, elation, enthusiasm, excitement, gratitude, pride, relief, and sympathy. From an initial pool of more than 800 papers, we selected 81 that propose recognition models, and categorized them according to the modality and data type used for recognition. According to the results, the most frequently occurring labels are amusement and excitement, followed by compassion and contentment. In terms of data type, the highest number of solutions utilizes various physiological signals, with visual data being the second most common. As the first survey to focus on a large number of distinct positive emotions, our paper contributes to Affective Computing and, more broadly, to Human-Computer Interaction by demonstrating that: 1) the topic of positive emotion recognition remains largely unexplored, despite numerous potential applications; 2) there are significant shortcomings in the research on positive emotion recognition, including the lack of relevant datasets, contextual information, and adequate data collection procedures. Finally, we provide suggestions and guidelines to support future research in this area.

2. Uncertain Facial Correction (2024) Summary: Expression Recognition via Multi-Task Assisted In this article, the MTAC framework to allevi ate the uncertainty in facial expression images. The target FER branch measured uncertainty to calculate the confidence score and strengthen valid samples during model training. The auxiliary VA branch executed category balancing and joint feature learning with the support of continuous emotion labels. The auxiliary AU branch constructed the data-driven AU graph to generate semantic representations. The relabeling strategy corrected extremely uncertain samples under the feature-level similarity constraint based on the updated memory templates. Our MTAC‘s modular design allows for adding and removing branches based on what is needed during training and inference. Extensive experiments on five large-scale datasets showed that MTAC was robust to uncertain samples and achieved superior results in the FER task. Although the MTAC performs competitive FER with full or weak supervision, the requirement of neighboring VA and AU annotations might limit the practical deployment. Alternative auxiliary tasks like face recognition and landmark detection could be introduced. Conversely, MTAC can be expanded to produce labels for incremental learning, pre-train universal en coders of facial expressions, and address the uncertain problem in other data modalities.

3. Multi Task Learning Of Emotion Recognition And Facial Action Unit Detection With Adaptively Weights Sharing Network Summary: In this paper, presented a multi-task learning method by adaptively sharing the weights in FER .The emotion recognition task and AU detection task are both fundamental and crucial tasks in FER. Our work can learn optimal combinations of task-specific and task shared representations and saving the cost of fine-tuning parameters manually. We performed extensive experiments that show our approach is able to equal or even out performed dedicated approaches on both emotion recognition and AU detection.

4. Multistep Deep System for Multimodal Emotion Detection with Invalid Data Summary: A multistep deep system to reliably detect multimodal emotion using collecting records containing invalid data is proposed. The proposed system includes a feature extraction and emotion detection method using peripheral physiological signals and video modalities via deep neural networks. The invalid data are filtered out in the discriminative module using the semantic compatibility and continuity. The experiments are conducted using a public database containing different proportions of invalid data.

The results verify the effectiveness of the discriminative module. Besides, the performance of the MSD is compared with the state-of-the-art approach in two conditions (the records contain invalid data and do not contain invalid data), and the proposed system based on peripheral physiological signals and video significantly improves the detection performance. The promising results imply that the proposed system can be deployed in many IoT scenarios, even without the complex network structure and brilliant data acquisition facilities.

5. Evaluating the Effect of Emotion Models on the Generalizability of Text Emotion Detection Systems Summary: The arbitrary selection of emotion models while labeling such datasets poses significant challenges in the performance and generalizability of the produced machine learning predictors, primarily when evaluated against unseen data, as it effectively introduces bias to the process. This study investigates the impact of emotion model selection on the efficacy of machine learning systems for text emotion detection. Eight labeled datasets were employed to train linear regression, feed forward neural network, and BERT-based deep learning models. Results demonstrated a notable decrease in accuracy when models trained on one dataset were tested on others, underscoring the inherent incompatibilities in labeling across datasets.

To prove that the emotion model significantly impacts predictors ‘performance, we propose a standardized emotion label mapping utilizing James Russell‘s circumplex model of affect that turns the emotion model into a parameter rather than a fixed element. Cross-dataset testing with this shared emotion mapping yielded significant, non-negligible changes in accuracy (both improvement and degradation). This fact highlights the impact of the emotion model (traditionally arbitrarily selected) during machine learning training and performance, arguing that improvements in accuracy reported in related research literature might be due to differences in the used emotion model rather than the new algorithms introduced.

6. Deep Fusion of Neurophysiologic and Facial Features for Enhanced Emotion Detection(2025) Summary: This research presents a novel multimodal emotion recognition system integrating neurophysiologic signals and facial data to enhance accuracy and reliability. The system processes face image sequences alongside EEG, EOG, GSR, BVP, RSP, EMG, SKT, and pulse wave signals to predict valence-arousal, liking, and dominance labels. Leveraging transformers and deep neural networks, it captures complex temporal and spatial patterns in the data. Experiments on the DEAP dataset, conducted on both per-subject and inter subject bases, demonstrate superior performance compared to single-modality methods and parity with state-of-the-art multimodal approaches.

CONCLUSION

The project Emotion Detection Using Deep Learning successfully integrates advanced artificial intelligence techniques to recognize human emotional states and generate appropriate system responses. Through the use of convolution neural networks (CNN), LSTM units, and transformer-based architectures, the system demonstrates reliable accuracy in classifying emotions such as happiness, sadness, anger, fear, surprise, and neutrality. This not only improves human–computer interaction but also provides valuable support in domains such as mental health monitoring, smart education, security surveillance, and personalized user experiences. The experiments show that deep learning–based models outperform traditional machine learning approaches and are capable of functioning effectively in real-time environments with proper preprocessing and dataset balancing. The project establishes a solid foundation for future work involving multimodal emotion recognition, context-aware actions, IoT integration, and deployment on mobile and embedded platforms. In this work, we survey computational models for the recognition of the following positive emotions: admiration, amusement, awe, compassion, contentment, elation, enthusiasm, excitement, gratitude, pride, relief, and sympathy. The choice is inspired by the previous works exploring the differences between positive emotions, especially. Further, we conduct preliminary queries to main databases of research papers to gain a better understanding of the variety of papers addressing classification, detection, and recognition of these emotions. More specifically, the labels are selected based on how reflective and inclusive they are of the current literature, while also taking practical concerns into account. After preliminary queries, some labels are excluded due to their ambiguous nature both in theory and in the contexts in which the research was conducted. We intentionally excluded generic descriptions of positive state, such as joy, happiness, and their synonyms such as enjoyment. This decision is made since our aim is to survey specific, well-defined emotional states, as well as the contexts in which they appear.

REFERENCES

[1] R.W.Picard, *AffectiveComputing*. Cambridge, MA, USA: MITPress, 1997.

[2] P. Ekman and W. V. Friesen, *Facial Action Coding System: A Technique for the MeasurementofFacialMovement*.PaloAlto, CA, USA: ConsultingPsychologistsPress, 1978.

[3] J.Redmon,S.Divvala,R.Girshick,andA.Farhadi,"YouOnlyLook Once:Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788.

[4] Ultralytics, "YOLO11 Architecture and Real-Time Object Detection Engine OptimizationDocumentation,"*UltralyticsDeepLearningReposito ry*, 2024.[Online]. Available: https://github.com/ultralytics/ultralytics

[5] P. Viola and M. Jones, "Rapid Object Detection using Boosted Cascade of Simple Features,"in*Proceedingsofthe2001IEEEComputerSocie tyConferenceonComputer Vision and Pattern Recognition (CVPR)*, Kauai, HI, USA, 2001, pp. I-511–I-518.

[6] M.Abadi*etal.*"TensorFlow:e-ScaleMachineLearningonHeterogeneous Distributed Systems,"*arXiv preprint arXiv:1603.04467*, 2016.

[7] D. P. Kingman and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proceedingsofthe3rdInternationalConferenceonLearningRepres entations (ICLR)*, San Diego, CA, USA, 2015.

[8] C.Lugaresi*etal.*"MediaPipe: AFrameworkforBuildingPerceptionPipelines,"*arXiv preprint arXiv: 1906.08172*, 2019.

[9] T. Soukupova and J. Cech, "Real-Time Eye Blink Detection Using Facial Landmarks,"in*Proceedingsofthe21stComputerVisionWinterWor kshop (CVWW)*, Rimske Toplice, Slovenia, 2016, pp. 1-8.

[10] G.Bradski,"TheOpenCV Library,"*Dr.Dobb'sJournalofSoftwareTools*,vol.25, Pp.120-123,2000.

[11] I.J.Goodfellow, Y.Bengio, andA.Courville, DeepLearning.Cambridge, MA, USA: MIT Press, 2016.

[12] Z.Zheng,P.Wang,W.Liu,J.Li,R.Ye,andD.Ren,"Distance-IoULoss:Faster and Better Learning for Bounding Box Regression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 12565-12572, 2020.

[13] J. Redmon et al., "YOLO: Real-Time Object Detection," IEEE Conference on Computer Vision and Pattern Recognition.

[14] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, and L. S. Mattos, __Toward emotion recognition from physiological signals in the wild: Approaching the methodological issues in real-life data collection,‘‘ Frontiers Psychol., vol. 11, p. 1111, Jul. 2020.

[15] Dahl R.E., Harvey A.G. Sleep in children and adolescents with behavioral and emotional disorders https://www.sciencedirect.com/science/article/pii/S1556407X07000513

[16] Wilson G.F., Russell C.A. Real-time assessment of mental workload using psychophysiological measures and artificial neural networks Hum. Factors, 45 (4) (2003), pp. 635-644 Google Scholar

[17] Kang S., Park C.Y., Kim A., Cha N., Lee U.Understandin g emotion changes in mobile experience CHI ’22, Association for Computing Machinery, New York, NY, USA (2022), 10.1145/3491102.3501944

[18] V.Lepetit,F.Moreno-Noguer,andP.Fua,"EPnP:AnAccurateO(n)Solutiontothe PnP Problem, *“International Journal of Computer Vision*, vol. 81, no. 2, pp. 155-166, 20