A Comprehensive Analyze of Deep Learning Techniques and Applications

Shuo Song
Fakulti Teknologi dan Sains Maklumat, Universitiy Kebangsaan Malaysia, Malaysia

Keywords: Facial Emotion Recognition (FER), Deep Learning, Human-Computer Interaction (HCI).

Abstract: Facial emotion recognition (FER) is a key research direction in the field of artificial intelligence and human-computer interaction (HCI). This technology is of great significance for improving the emotion understanding ability of virtual reality, intelligent customer service, intelligent driving and other systems. Currently, deep learning models can significantly improve the recognition accuracy of FER, which makes up for the limitation that traditional methods are difficult to cope with complex and changing real-world scenarios. This paper systematically reviews the development history of FER, and analyzes its typical applications in human-computer interaction from the perspectives of both traditional methods and deep learning methods. In summary, it is found that deep learning methods have made significant progress in multimodal emotion recognition, real-time performance optimization, and personalized modeling, but still face challenges such as data imbalance, occlusion interference and cross-domain adaptability. This paper provides an outlook on the future development trend of FER, aiming to provide reference and inspiration for subsequent research.

1 INTRODUCTION

Facial expression is one of the most significant human nonverbal behaviors, which is widely used in human–human social communication. Facial expression can not only reflect subject's emotional feeling, but also be used in group communication and other social behaviors. In the past few years, due to the rapid development of artificial intelligence and computer vision, Facial Emotion Recognition (FER) has gradually become one of the key research focuses of automated affective computing. The purpose of Facial Emotion Recognition is to recognize the emotion of a subject from the face image or video stream automatically and classify it into several categories, so as to improve the intelligence level of Human-Computer Interaction (HCI).

In the HCI field, FER is widely used in intelligent customer service, education technology, mental health monitoring, intelligent driving, and Virtual Reality (VR) and Augmented Reality (AR) scenarios. For example, in intelligent customer service systems, FER can be used to analyze the user's emotional fluctuations, so as to optimize the interaction of customer service robots or voice assistants and improve the user experience; in intelligent driving systems, FER can be used to monitor the driver's attention and fatigue in real time, so as to improve the safety of driving; in mental health monitoring, FER can be used to combine physiological signals and voice analysis to assist in the diagnosis of emotional disorders. In addition, in the VR/AR environment, FER can enhance the realism of virtual characters' expressions, making the interaction more natural and immersive.

Despite the deep learning application greatly advances the performance of facial expression recognition (FER) recognition, there are still many challenges for deep learning to be applied in practical FER applications, such as imbalanced data categories, poor cross-domain generalization, sensitivity to occlusion and illumination variations, and high real-time performance requirements. Mollahosseini, Chan, and Mahoor (2016) noted that, in the FER2013 data set, there were only 547 samples in the aversion category, so the model prediction was unfairly influenced by this data category. Li et al. (2018) enhanced the stability of CNN models in complex environments by introducing an occlusion simulation mechanism. In addition, Mollahosseini et al. (2016) validated the superior performance of their model in cross-domain recognition through cross-validation

Song, S.
A Comprehensive Analyze of Deep Learning Techniques and Applications.
DOI: 10.5220/0014366900004718

In *Proceedings of the 2nd International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2025)*, pages 589-593
ISBN: 978-989-758-792-4

experiments conducted on seven publicly available databases, while Lopes et al. (2017) found that data enhancement and preprocessing techniques (e.g., rotating, cropping, and luminance normalization) significantly improved the model performance on multiple databases (Mellouk & Handouzi, 2020). For real-time application requirements, Pranav, Kamal, Chandran, and Supriya (2020) proposed a lightweight convolutional network with a simple structure for emotion recognition on mobile. Overall, FER research is evolving towards model lightweighting, multimodal fusion and high robustness.

So, in this paper, the aim is to present the development process of FERFacial Emotion Recognition (FER) technology, and it is achieved by analyzing its applications and challenges in HCI.Firstly, introduce the traditional method and its application environment; Secondly, discuss the application effect of deep learning based FER model in practice; Finally, make summary of the problems and look forward to the future development direction.

2 OVERVIEW OF METHODS FOR FACIAL EMOTION RECOGNITION

2.1 Traditional Methods

Early FER methods relied heavily on feature engineering techniques that used geometric and texture features to extract unique patterns in facial images for sentiment classification.

Geometric feature-based methods mainly analyze the spatial relationships between key points on the face, such as the positions of the eyes, eyebrows, nose, and mouth. The Facial Action Coding System (FACS), proposed by Ekman and Friesen, is a widely recognized approach in this category. It decomposes facial expressions into a set of Action Units (AUs) based on underlying muscle movements, allowing for the recognition of various emotional states. In HCI systems, FACS is commonly applied in intelligent customer service and remote sentiment analysis to enhance user experience—intelligent assistants, for instance, can adapt their interactions based on users' subtle facial cues to improve naturalness and engagement. In contrast, texture feature-based methods determine emotional states by analyzing local texture patterns in facial regions. Techniques such as Local Binary Pattern (LBP) and Gabor wavelets are commonly used in this context. LBP captures local pixel differences and demonstrates robustness to lighting changes, while Gabor wavelets extract features across multiple scales and orientations, thereby improving recognition accuracy. These texture-based approaches are extensively used in applications such as distance education and affective computing, where FER systems can monitor students' concentration and dynamically adapt teaching strategies to enhance learning outcomes.

2.2 Traditional Machine Learning Methods

Before the rise of deep learning, FER algorithms were predominantly based on traditional machine learning approaches like classifiers (e.g., SVMs), hidden Markov models (HMMs) or k-nearest neighbors (kNN). Typically, these approaches need you to extract so-called geometric features (distances between facial keypoints, angles) or texture features (local binary patterns, responses of filters like Gabor filters) from the image and use these vectors as input for the machine learning model. This "feature engineering" approach has the big disadvantage that it is knowledge- (domain expert) dependent and results in a bad generalization behavior for challenging scenarios.

Specifically, traditional classifiers are difficult to effectively establish robust classification boundaries when facing high-dimensional, nonlinearly distributed expression data. For example, SVM may be overfitted or underfitted when dealing with data of high dimensionality or unevenly distributed categories, and is highly sensitive to hyperparameter selection.Although HMM is suitable for sequence modeling, its assumption of state transfer is more simplified, and it is difficult to accurately capture the nonlinear and complex dynamic changes in facial expressions. In addition, distance-based methods such as kNN face "dimensional disaster" in high-dimensional space, their performance does not improve significantly with the increase of training data, and the computational cost is high.

In dynamic interaction environments, such as video surveillance, emotional robots, or intelligent customer service systems, the user's facial expression changes rapidly and is greatly affected by lighting, posture, occlusion, etc., and traditional methods are less robust to these disturbances. For example, in intelligent surveillance systems, SVM is usually used to analyze the emotional state in long-time static image sequences, but it is difficult to respond to instantaneous expression changes in real time, and it is not able to dynamically predict the emotional state based on the context. Therefore, although traditional machine learning methods played a key role in the early research of FER, their lack of adaptability to environmental changes, feature complexity, and real-time availability limits their wide application in real interactive systems. These limitations have provided an opportunity for the rise of deep learning methods, which have significantly improved the performance and application scope of emotion recognition systems through end-to-end learning and automatic feature extraction mechanisms.

2.3 Deep Learning Methods

In the past few years, with the rapid development of deep learning, the accuracy of FER has made a big breakthrough. Compared with traditional machine learning methods which need to manually extract features, deep learning models can automatically learn emotionally relevant higher-order semantic features from raw images in an end-to-end way, which greatly improves the accuracy and robustness of recognition. Among them, Convolutional Neural Networks（CNNs） are the most widely used models for the recognition of static facial images. By multiple convolution and pooling layers, local features are extracted from images and global representation are formed. This class of models have gained far better recognition level than traditional methods on the main stream datasets, FER2013, CK+, JAFFE, etc.

In addition, for video sequences or dynamic expression analysis, Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory Networks (LSTMs), can efficiently model emotional changes in time series and enhance the understanding of expression transition processes. For example, in intelligent customer service scenarios, LSTMs can continuously track subtle changes in user expressions to adjust interaction strategies and achieve more emotionally adaptive feedback.

Deep learning methods also support multimodal emotion fusion, such as combining facial images, speech signals, physiological features, and other information for comprehensive discrimination, further improving the accuracy and application breadth of emotion recognition. Compared with traditional methods, deep models are more fault-tolerant to disturbing factors such as lighting changes, occlusion, and individual differences in expressions. However, deep learning methods also face some challenges, such as the dependence on large-scale high-quality data, high consumption of computational resources, and poor interpretability.

Nevertheless, with the improvement of computing power and optimization of training strategies, deep learning-based FER systems have gradually matured and have been widely used in many HCI application scenarios, such as smart home, virtual reality, and driving assistance, becoming one of the key technologies to promote the intelligence of human-computer interaction.

3 DEEP LEARNING-BASED FACIAL EMOTION RECOGNITION IN HUMAN-COMPUTER INTERACTION

3.1 Emotional Perception System for Students in Virtual Educational Environments

Students' facial expressions can reflect their level of understanding, learning status and psychological changes. However, in virtual teaching and distance education scenarios, it is often difficult for teachers to capture students' emotional feedback through traditional methods. Therefore, real-time recognition of student emotions has become a key tool to enhance the online education experience. Sachith Hans and Rao (2021) designed a deep learning architecture combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory Networks (LSTMs) for analyzing sequences of facial images in videos and extracting spatio-temporal features to recognize continuous emotional changes. It is shown that the model achieves 78.52% and 63.35% recognition accuracy on the training set CREMA-D and test set RAVDEES, respectively. The method enhances the model's focus on facial dynamics by preprocessing video images with the OpenFace tool, extracting facial regions and masking background interference. This deep learning-based student emotion perception system can assist teachers to dynamically understand students' attention and emotion changes, and provide reference for personalized teaching content pushing and interaction strategy adjustment.

3.2 Non-Contact Emotion Monitoring in Intelligent Healthcare Systems

In smart healthcare and telecare scenarios, patients may not be able to actively express their emotional states due to language barriers, ambiguous consciousness or physiological factors. Compared to traditional emotion monitoring methods that rely on physiological signals such as skin electrical response, heart rate, or body temperature, facial expression recognition (FER) systems have the advantages of being non-contact, real-time, and easier to deploy, which is especially suitable for healthcare environments where it is inconvenient to wear sensing devices. Carmen Bisogni et al. (2022) propose a multi-resolution deep convolutional network model for medical scenarios, which combines a migration learning strategy with image data enhancement techniques to improve the generalization ability of the model under different lighting, angle and resolution conditions. The system is tested on mainstream datasets such as Cohn-Kanade (CK+), Karolinska Directed Emotional Faces (KDEF), and Static Facial Expressions in the Wild, and outperforms the traditional model in terms of recognition accuracy and environmental adaptability. The system architecture includes three core modules: facial feature point detection, multi-resolution image processing and convolutional feature extraction. Its final recognition results can be fed back to medical terminals via an Internet of Things (IoT) platform to help doctors determine whether a patient is in an emotional state such as anxiety, pain, or depression, enabling more detailed and personalized remote care. The method is especially suitable for scenarios such as geriatric care, psychiatric treatment and post-operative rehabilitation, reflecting the unique value of FER in humanized healthcare.

3.3 Abnormal Emotion Recognition and Early Warning in Public Security Scenarios

When people in crowded public scenes like stations, airports and shopping malls exhibit mood changes in their faces, it may be a significant clue to identify abnormal behaviors. Akhand et al. designed a deep convolutional neural network based on transfer learning. Pre-trained models (VGG-16, ResNet-50, DenseNet-161) are used as feature extractor in their architecture, and KDEF is combined with JAFFE for fine-tuned training. The experimental results show that the model obtains 99.52% recognition accuracy on JAFFE data set, and has good cross-view robustness on KDEF data set of multi-angle face images(Saleem, 2021). The system focuses on side face and head deviated view in recognition, which makes the system more applicable in real security scene. In actual application, the system can access the face video stream collected by the camera and give the real-time marking and warning for the person with abnormal expression such as anger, fear, surprise, etc, to assist the security man to make rapid handling and improve the security management level of the public place.

4 CHALLENGES AND FUTURE DIRECTIONS IN FACIAL EMOTION RECOGNITION

The variability of facial expressions among individuals presents significant challenges for classification models in facial expression recognition (FER). Accessories such as glasses and masks, along with changes in head pose, further reduce recognition accuracy. Moreover, FER datasets often suffer from emotional category imbalance—particularly with emotions like fear and disgust, which have fewer samples—leading to potential model bias. Another issue lies in the poor generalization ability across different datasets, resulting in inconsistent performance in varying environments (Ranganathan, 2016). Additionally, the high computational demands of deep learning models limit their deployment in real-time applications. To address these challenges, researchers are exploring several innovative strategies. Few-Shot Learning and Zero-Shot Learning are being investigated to enable FER systems to recognize emotions with limited annotated data. Multimodal FER, which integrates facial expressions, voice, body language, and physiological signals, is also gaining attention for its potential to enhance emotional recognition accuracy. Furthermore, techniques such as adversarial training and domain adaptation are being developed to improve robustness to unseen environments. Edge AI technologies are also emerging as a promising solution, aiming to design lightweight, real-time FER models that can operate efficiently on mobile and embedded devices.

5 CONCLUSION

This paper systematically reviews the development history of FER technology, focusing on the typical applications of deep learning-based methods in the field of HCI. By comparing traditional machine learning methods with current mainstream deep learning architectures, the advantages of deep models including CNN and RNN in feature extraction and sentiment classification are analyzed. Meanwhile, the performance performance, deployment conditions and challenging factors of current FER systems at the application level are sorted out based on several representative literatures in the context of actual HCI scenarios such as education, healthcare and security.

It is found that deep learning methods significantly improve the recognition accuracy and robustness of FER systems in complex environments, especially in dealing with non-frontal viewpoints, face occlusion, and lighting changes, etc., and substantial breakthroughs have been achieved. In addition, the combination of migration learning, data augmentation, multi-scale modeling and other techniques further enhances the model's generalization ability and practical deployment feasibility.

Although the current FER technology has been well applied in many fields, there are still problems such as insufficient real-time computing capability, weak privacy protection mechanism, and unstable cross-domain migration effect. Future research can start from the directions of multimodal emotion fusion, privacy computation under the federated learning framework, and lightweight design of models, to promote the in-depth development of FER systems in personalized emotion understanding and intelligent interaction, and to lay a solid foundation for the construction of a more natural and efficient human-computer collaboration system.

REFERENCES

Akhand, M.A.H., Roy, S., Siddique, N., Kamal, M.A.S., Shimamura, T., 2021. Facial emotion recognition using transfer learning in the deep CNN. Electronics, 10, 1036.

Bisogni, C., Castiglione, A., Hossain, S., Narducci, F., Umer, S., 2022. Impact of deep learning approaches on facial expression recognition in healthcare industries. IEEE Transactions on Industrial Informatics, 18(8), 5619–5627.

Hans, A.S.A., Rao, S., 2021. A CNN-LSTM based deep neural networks for facial emotion detection in videos. IJASIS, 7(1), 11–20.

Li, S., Deng, W., 2022. Deep facial expression recognition: a survey. IEEE Transactions on Affective Computing, 13(3), 1195–1215.

Mellouk, W., Handouzi, W., 2020. Facial emotion recognition using deep learning: review and insights. Procedia Computer Science, 175, 689–694.

Mollahosseini, A., Chan, D., Mahoor, M.H., 2016. Going deeper in facial expression recognition using deep neural networks. In IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 1–10.

OpenAI, 2025. Research on computer vision and AI applications.

Pranav, E., Kamal, S., Chandran, C.S., Supriya, M.H., 2020. Facial emotion recognition using deep convolutional neural network. In 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 317–320.

Ranganathan, H., Chakraborty, S., Panchanathan, S., 2016. Multimodal emotion recognition using deep learning architectures. In IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 1–9.

Saleem Abdullah, S.M., Abdulazeez, A.M., 2021. Facial expression recognition based on deep learning convolution neural network: a review. J. Smart Computing and Data Mining, 2(1), 53–65.