

Advancements in Facial Expression Recognition: A Comparative Study of Traditional Machine Learning and Deep Learning Approaches

Yujie Jin^a

Artificial Intelligence, Soochow University, Suzhou, Jiangsu Province, 215000, China

Keywords: Facial Expression Recognition, Traditional Machine Learning, Deep Learning.

Abstract: In recent years, the use of facial expression recognition technology has become widespread with the development of artificial intelligence. However, limitations in face expression recognition still exist due to various factors such as environment and angle. This paper explores the processing methods of machine learning and deep learning models for face recognition and compares the differences between the two methods. In traditional machine learning models, this paper analyses the methodology of combining the results of Coded Hidden Markov Model (CHMM) and Fundamental Hidden Markov Model (FHMM) firstly, and the aim is to devise a comprehensive framework of criteria for classifying samples. The paper then analyses the key role of k-value in KNN classifiers by varying the k-value. This reveals that the choice of the k-value significantly affects the accuracy of emotion classification. The text appears to already meet the desired characteristics. No changes have been made. In deep learning, various CNN configurations such as region-based CNN (R-CNN), faster R-CNN, and 3D CNN have been analyzed for their precision on different datasets. Additionally, the study explores the extraction of face information using FPN target detection methods and the integration of LSTM networks with CNNs to efficiently capture sequential information from facial images and extract a more comprehensive representation of features. It is concluded that deep learning model is more effective and face emotion recognition is transitioning from traditional machine learning to deep learning.


1 INTRODUCTION

Facial expression recognition technology is a sophisticated approach that analyzes facial features in images to accurately assess an individual's emotional state. This technology primarily relies on discrete labels, categorizing expressions into six fundamental emotions: happiness, sadness, surprise, fear, disgust, and anger. Currently, facial expression recognition technology finds extensive applications across various domains. For instance, it plays a crucial role in medicine by enabling machines to identify specific emotional states and determine the psychological well-being of individuals for effective prevention and treatment of depression.

In the field of sales, it is feasible to identify customers' expressions and gather their product preferences in order to more accurately tailor the offerings to their liking. In education, students'

engagement and classroom participation can be assessed through their various expressions, providing valuable feedback for teachers and parents to assist students in adjusting their status and enhancing efficiency.

In the current era of digital information, artificial intelligence has made significant strides in its development. Algorithms such as Artificial Neural Network, K-Nearest Neighbor, Convolutional Neural Network (CNN), Long Short Term Memory have permeated various industries including medical care, transportation, and biology to support the advancement of diverse fields of study for humanity (Li, 2024; Qiu, 2019; Sun, 2020; Wang, 2024; Wu, 2024; Zhou, 2023). Moreover, it is crucial not to overlook their invaluable contributions to the field of education. In 2019, Gupta trained two CNN models to analyze the emotional state of a single student in a single image frame and analyze multiple students using a single image frame, respectively (Gupta, 2019).

^a <https://orcid.org/0009-0006-5521-5529>

In 2021, Pabba trained CNN models on BAUM-1, DAiSEE, and YawDD datasets to help teachers monitor student engagement and maintain appropriate levels of interaction (Pabba, 2021). In 2022, David Dukić put the test set in Inception-v3 and Training on ResNet-34 to know how these emotions are related to the tasks solved by the students to help teachers improve the lecture (David Dukić, 2022). The scholarly research on face recognition has sparked a profound interest in me regarding its application within the realm of education. However, it is worth noting that the impact of environmental factors and teacher performance on students remains unexplored through facial expression recognition. Hence, the paper is motivated to delve into this aspect for further investigation.

The remainder of the paper is organized as follows. Firstly, In Section 2, the paper will overview the detailed methods used by various researchers in the experimental process. Then, In Section 3, the paper will discuss the current shortcomings and puts forward the future direction of this field. Finally, the Section 4 will summarize the whole paper and present conclusions drawn from researches discussed.

2 METHODS

Framework of machine learning-based facial expression recognition. The current research on facial expression recognition for machine learning typically involves the following steps: Firstly, data collection is conducted either through public datasets or self-collection to acquire face images. Subsequently, the collected data undergoes preprocessing to extract crucial facial features or perform data augmentation. Following this, machine learning or deep learning models such as random forests, decision trees, neural networks are constructed and model parameters are configured for training purposes.

2.1 Traditional Machine Learning

2.1.1 HMM

X. Jiang utilizes the Active Appearance Model (AAM) to extract feature series from training sample images (Jiang, 2011). The hidden Markov expression models of six basic expressions are derived through the application of the Welch Baum algorithm and training with six types of feature sequences.

Subsequently, the facial expression sequences of each training object were fed into six training models,

and the output probabilities under each expression model were computed. This process enabled the determination of the alignment between a specific training expression sequence and the six expressions, resulting in a normalized matching degree vector across six dimensions. By evaluating the facial expression sequences of each training object in relation to the pattern matching output vector, a cumulative standardization and quantification process was undertaken to ascertain the similarity among the six expressions within the representation result of the training object.

Moreover, by averaging the similarity of opposing expressions for each training object, an overall reflection of the similarity between expression patterns was derived, providing a foundational rationale for subsequent classification decisions.

Drawing from the classification group, two three-state HMM classifiers were trained using the Baum-Welch algorithm to formulate a coded hidden Markov model. The essence of these classifiers lies in categorizing the six expressions into two classes, amalgamating three minor-class expression samples into a unified entity, and consolidating the six expression types into two categories featuring three expressions each. In the realm of expression recognition, in instances where the classification verdict of the encoded hidden Markov model falters—signifying an empty intersection of triple decision outputs—the study advocates for resorting to the utilization of pre-trained hidden Markov models to enhance classification accuracy.

To encapsulate, through a systematic amalgamation of steps and methodologies, this research amalgamates the outcomes of the coded hidden Markov model and the fundamental hidden Markov model to devise a comprehensive framework of sample classification criteria. This framework furnishes a holistic solution for the realms of facial expression recognition and classification.

2.1.2 KNN

The K-Nearest Neighbors (KNN) classifier stands as a prevalent non-linear classification method, well-suited for a wide array of applications, including the domain of emotion recognition. This algorithm operates by determining the K training data points closest to the test data, thus enabling the classification of unknown data. In the realm of KNN, various distance metrics can be leveraged to gauge the similarity between the test data and the training data. These metrics encompass the likes of the Manhattan, Euclidean, Minkowski, and Chebyshev distances,

offering a diverse toolkit for effective classification tasks.

In the realm of emotional expression recognition, diverse distance measures were employed in the M. Murugappan's study to effectively classify facial emotions, with the average accuracy of each measure being meticulously documented (Murugappan, 2020). Furthermore, experimental investigations delved into the impact of varying the k value, revealing that the selection of k value significantly influences the accuracy of emotion classification. The study underscored the pivotal role of k value in the KNN classifier, indicating that a lower k value correlates with a heightened emotion recognition rate.

2.1.3 RF and PCA

In the realm of image processing, the traditional Principal Component Analysis (PCA) serves as a widely employed technique for extracting expression features. Initially, this method involves converting the image matrix into a one-dimensional image vector. However, the application of the K-L transformation leads to a significant increase in the dimension of the image vector space, posing challenges in accurately computing the covariance matrix. To address this issue, the emergence of 2DPCA offers a direct approach to handling two-dimensional image matrices, thereby facilitating the computation of the covariance matrix with greater ease. Given its lower dimensionality, 2DPCA simplifies the calculation of eigenvalues and eigenvectors of the covariance matrix.

The fundamental concept behind 2DPCA lies in treating each image as an undefined control sequence that undergoes linear transformation into an m -dimensional column vector through matrix multiplication. Here, x denotes the n -dimensional projected column vector, while Y represents the mapping of an eigenvector of the matrix in the x -direction. Following the completion of expression feature extraction, the selection of a suitable classification method becomes paramount. Ju Jia purposed the random forest classifier is adopted for its attributes of rapid classification speed, robustness (Jia, 2016), and high recognition rates in high-dimensional scenarios. Random forest, constructed from multiple decision trees, proves effective in addressing multi-data classification challenges.

The random forest classifier comprises N decision trees (e.g., T_1, T_2, \dots, T_N), where each decision tree functions as a voting classifier. The ultimate outcome of random forest classification is the average of the voting results from all decision trees. In the

experiment conducted, two testing schemes are employed: one for testing trained individuals and another for testing untrained individuals. Post image preprocessing, the PCA and 2DPCA methods are utilized for extracting expression features, subsequently employed in training and classifying random forest and Support Vector Machine (SVM) classifiers.

2.2 Deep Learning

Traditional methods exhibit lower reliance on hardware and data types in comparison to deep learning approaches. However, they require manual application of feature extraction and classification as independent steps. In contrast, deep learning methods are capable of simultaneously executing these two processes. Within the realm of deep learning techniques, facial expression recognition entails three primary stages: image preprocessing, deep feature learning, and deep feature classification. The image preprocessing phase is a critical step that typically encompasses the utilization of the Viola-Jones algorithm for face detection, face alignment, normalization, and enhancement to prepare the data. In the realms of deep feature learning and classification, numerous methodologies such as CNN, DBN, DAE, RNN and LSTM have been extensively researched and implemented.

2.2.1 CNN

In the realm of image preprocessing, CNN configurations continue to stand out as the most prevalent and cutting-edge, particularly in the realm of emotion recognition. Some of the popular CNN configurations include region-based CNN (R-CNN), faster R-CNN, and 3D CNN. These configurations demonstrate varying levels of accuracy across different datasets.

S. Begaj, A. O. Topal proposed the implementation details and results analysis of a CNN-based facial expression recognition model (Begaj, 2020). This CNN architecture comprises four convolutional layers, four max-pooling layers, one dropout layer, and two fully connected layers, totaling 899,718 parameters. The model's processing pipeline involves filtering images through Conv2D filters, applying ReLU activation functions, downsizing images with MaxPooling2D layers, and ultimately flattening and applying dropout layers.

Upon evaluating the model, it was observed that the training data outperformed the testing data, indicating signs of overfitting. Beyond the 25th epoch,

the training data accuracy surpassed that of the testing data, suggesting a bias towards the training data. Introducing data augmentation with the same model yielded improved results. In a third experiment, a pre-trained VGG16 model was applied to the dataset, with images sized at 250×250 pixels.

2.2.2 LSTM

Facial expression recognition, as a pivotal area of investigation within the domains of computer vision and human-computer interaction, has garnered considerable attention in recent years. L. Muyao and D. Weili provides an overview of the most recent advancements in facial expression recognition methodologies that are based on the integration of Convolutional Neural Network (CNN) and Long Short-Term Memory Network (LSTM) (Muyao, 2023).

LSTM, a type of recurrent neural network, is widely utilized for managing long-term dependencies in sequential data. In this investigation, the LSTM network is integrated with CNN to effectively capture sequential information from facial images and extract more comprehensive feature representations. The incorporation of the LSTM network enables the model to proficiently handle temporal sequence information, thereby enhancing the accuracy and robustness of facial expression recognition.

Meanwhile, the Feature Pyramid Network (FPN) has been introduced as an effective tool for multi-scale object detection, enabling the detection of objects at various scales through a top-down structure. In this research, the FPN object detection method is utilized to extract facial information, offering advantages such as enhanced detection accuracy and robust stability and reliability.

3 DISCUSSIONS

Traditional machine learning methods often require hand-designed feature extractors. These extractors need to be manually designed and selected for the task, and then provided to the classifiers (Random forests as mentioned above) for training and prediction. However, these models are relatively simple and only suitable for small-scale data and simple tasks. As a result, they may lack generalization ability when dealing with complex data and tasks. In contrast, deep learning can automatically learn more abstract and advanced features from the data, improving performance and generalization ability, while completing more complex face recognition tasks by

training on more data. Therefore, face emotion recognition is transitioning from traditional machine learning to deep learning.

However, there are still some limitations and challenges, there are more similarities among anger, disgust and sadness on the static images, and if it is possible to use the multi-frame image sequences to extract the dynamic information for PCA analysis, better recognition results could be obtained to distinguish more expressions. In addition, future research can further explore how to improve the real-time performance of the model and reduce the computational load of the model, so as to promote its use in practical applications. In addition, FPN and LSTM-based models can be considered to be fused with other models to further improve the expressiveness and robustness of the models.

In addition, current models for face emotion recognition are often less interpretable, often require additional computational and storage resources, and are susceptible to factors such as data quality and feature selection. This may increase model complexity and overhead. Furthermore, model adaptability needs improvement, and models trained on European and American faces may not be easily applicable to other races, such as Asians. In general, external factors such as light intensity, occlusions, and pose shifts can affect the results. To address these issues, cross-domain integration (e.g. computer vision, pattern recognition) can be promoted to break down disciplinary boundaries, while enhancing dataset diversity for broader and deeper applications.

4 CONCLUSIONS

This paper explores the use of traditional machine learning and deep learning in facial expression recognition. The article summarises the weaknesses of current machine learning and deep learning models in terms of interpretability, adaptability, and robustness. It discusses four machine learning methods (HMM, KNN, RF, and PCA) and two deep learning methods. Future research can further delve into knowledge migration and transformation between different domains to achieve broader cross-domain integration. For example, combining methods such as migration learning, domain adaptation and meta-learning to better handle the heterogeneity of data from different domains and improve the generalization ability of models. Additionally, with the continuous development of techniques such as Generative Adversarial Networks (GANs) and self-supervised learning, it is possible for future research

to focus more on increasing the diversity of datasets using synthetic data and unsupervised learning methods which may help to improve the adaptability of models when facing new domains and unknown data.

REFERENCES

- Begaj, S., Topal, A. O., & Ali, M. 2020 Dec. Emotion recognition based on facial expressions using convolutional neural network (CNN). In 2020 International conference on computing, networking, telecommunications & engineering sciences applications (CoNTESA) (pp. 58-63). IEEE.
- Dukić, D., & Sovic Krzic, A. 2022. Real-time facial expression recognition using deep learning with application in the active classroom environment. *Electronics*, 11(8), 1240.
- Gupta, S. K., Ashwin, T. S., & Guddeti, R. M. R. 2019. Students' affective content analysis in smart classroom environment using deep learning techniques. *Multimedia Tools and Applications*, 78, 25321-25348.
- Jia, J., Xu, Y., Zhang, S., & Xue, X. 2016 Aug. The facial expression recognition method of random forest based on improved PCA extracting feature. In 2016 IEEE International Conf. on Signal Processing, Communications and Computing (ICSPCC) (pp. 1-5). IEEE.
- Jiang, X. 2011 A facial expression recognition model based on HMM. In *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology* (Vol. 6, pp. 3054-3057). IEEE.
- Li, M., He, J., Jiang, G., & Wang, H. 2024. DDN-SLAM: Real-time Dense Dynamic Neural Implicit SLAM with Joint Semantic Encoding. arXiv preprint arXiv:2401.01545.
- Muyao, L., & Weili, D. 2023 July. Facial Expression Recognition Based on FPN and LSTM. In 2023 IEEE 5th International Conference on Power, Intelligent Computing and Systems (ICPICS) (pp. 762-767). IEEE.
- Murugappan, M., Mutawa, A. M., Sruthi, S., Hassouneh, A., Abdulsalam, A., Jerritta, S., & Ranjana, R. 2020 Sept. Facial expression classification using KNN and decision tree classifiers. In 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP) (pp. 1-6). IEEE.
- Pabba, C., & Kumar, P. 2022. An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition. *Expert Systems*, 39(1), e12839.
- Qiu, Y., Chang, C. S., Yan, J. L., Ko, L., & Chang, T. S. (2019, October). Semantic segmentation of intracranial hemorrhages in head CT scans. In 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) (pp. 112-115). IEEE.
- Sun, G., Zhan, T., Owusu, B.G., Daniel, A.M., Liu, G., & Jiang, W. 2020. Revised reinforcement learning based on anchor graph hashing for autonomous cell activation in cloud-RANs. *Future Generation Computer Systems*, 104, 60-73.
- Wang, H., Zhou, Y., Perez, E., & Roemer, F. 2024. Jointly Learning Selection Matrices For Transmitters, Receivers And Fourier Coefficients In Multichannel Imaging. arXiv preprint arXiv:2402.19023.
- Wu, Y., Jin, Z., Shi, C., Liang, P., & Zhan, T. 2024. Research on the Application of Deep Learning-based BERT Model in Sentiment Analysis. arXiv preprint arXiv:2403.08217.
- Zhou, Y., Osman, A., Willms, M., Kunz, A., Philipp, S., Blatt, J., & Eul, S. 2023. Semantic Wireframe Detection. publica.fraunhofer.de.