

A DEEP LEARNING FACIAL EXPRESSION RECOGNITION BASED SCORING SYSTEM FOR RESTAURANTS

¹O.RAMYA TEJA, ²M.SOUMYA, ³M.VANI, ⁴MANUPATI PAVANI

¹Assistant Professor, Department of Information Technology, **MALLA REDDY ENGINEERING COLLEGE FOR WOMEN**, Maisammaguda, Dhulapally Kompally, Medchal Rd, M, Secunderabad, Telangana.

^{2, 3, 4} Student, Department of Information Technology, **MALLA REDDY ENGINEERING COLLEGE FOR WOMEN**, Maisammaguda, Dhulapally Kompally, Medchal Rd, M, Secunderabad, Telangana.

ABSTRACT

This paper presents a novel approach to evaluating customer satisfaction in restaurants through a deep learning-based facial expression recognition system. Leveraging state-of-the-art convolutional neural networks (CNNs), our system analyzes facial expressions captured by in-house cameras to generate real-time satisfaction scores. This method aims to provide restaurant management with actionable insights into customer experiences, enhance service quality, and improve operational efficiency. We discuss the system's architecture, data collection, implementation, and evaluation results, demonstrating its effectiveness and potential applications. Recently, the popularity of automated and unmanned restaurants has increased. Due to the absence of staff, there is no direct perception of the customers' impressions in order to find out what their experiences with the restaurant concept are like. For this purpose, this paper presents a rating system based on facial expression recognition with pre-trained convolutional neural network (CNN) models. It is composed of an Android mobile application, a web server, and a pre-trained AIserver. Both the food and the environment are supposed to be rated. Currently, three expressions (satisfied, neutral and disappointed) are provided by the scoring system.

INTRODUCTION

Facial expression is one of the most powerful, natural and universal signals for human beings to convey their

emotional states and intentions [1], [2]. Numerous studies have been conducted on automatic facial expression analysis because of its practical importance in

sociable robotics, medical treatment, driver fatigue surveillance, and many other human-computer interaction systems. In the field of computer vision and machine learning, various facial expression recognition (FER) systems have been explored to encode expression information from facial representations. As early as the twentieth century, Ekman and Friesen [3] defined six basic emotions based on cross-culture study [4], which indicated that humans perceive certain basic emotions in the same way regardless of culture. These prototypical facial expressions are anger, disgust, fear, happiness, sadness, and surprise. Contempt was subsequently added as one of the basic emotions [5]. Recently, advanced research on neuroscience and psychology argued that the model of six basic emotions are culture-specific and not universal [6].

Although the affect model based on basic emotions is limited in the ability to represent the complexity and subtlety of our daily affective displays [7], [8], [9], and other emotion description models, such as the Facial Action Coding System (FACS) [10] and the continuous model using affect dimensions [11], are considered to represent a wider range of emotions, the categorical model that describes emotions in terms of discrete

basic emotions is still the most popular perspective for FER, due to its pioneering investigations along with the direct and intuitive definition of facial expressions. And in this survey, we will limit our discussion on FER based on the categorical model.

FER systems can be divided into two main categories according to the feature representations: static image FER and dynamic sequence FER. In static-based methods [12], [13], [14], the feature representation is encoded with only spatial information from the current single image, whereas dynamic-based methods [15], [16], [17] consider the temporal relation among contiguous frames in the input facial expression sequence. Based on these two visionbased methods, other modalities, such as audio and physiological channels, have also been used in multimodal systems [18] to assist the recognition of expression.

The majority of the traditional methods have used handcrafted features or shallow learning (e.g., local binary patterns (LBP) [12], LBP on three orthogonal planes (LBP-TOP) [15], non-negative matrix factorization (NMF) [19] and sparse learning [20]) for FER. However, since 2013, emotion recognition competitions such as

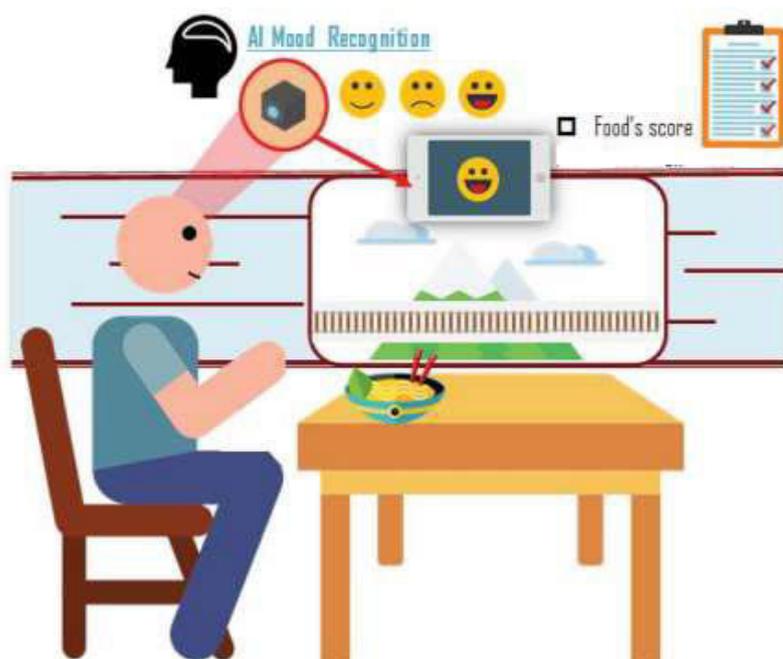
FER2013 [21] and Emotion Recognition in the Wild (EmotiW) [22], [23], [24] have collected relatively sufficient training data from challenging real-world scenarios, which implicitly promote the transition of FER from lab-controlled to in-the-wild settings. In the meanwhile, due to the dramatically increased chip processing abilities (e.g., GPU units) and well-designed network architecture, studies in various fields have begun to transfer to deep learning methods, which have achieved the state-of-the-art recognition accuracy and exceeded previous results by a large margin (e.g., [25], [26], [27], [28]). Likewise, given with more effective training data of facial expression, deep learning techniques have increasingly been implemented to handle the challenging factors for emotion recognition in the wild. Figure 1 illustrates this evolution on FER in the aspect of algorithms and datasets.

Exhaustive surveys on automatic expression analysis have been published in recent years [7], [8], [29], [30]. These surveys have established a set of standard algorithmic pipelines for FER. However, they focus on traditional methods, and deep learning has rarely been reviewed. Very recently, FER based on deep learning has been

surveyed in [31], which is a brief review without introductions on FER datasets and technical details on deep FER. Therefore, in this paper, we make a systematic research on deep learning for FER tasks based on both static images and videos (image sequences). We aim to give a newcomer to this field an overview of the systematic framework and prime skills for deep

Despite the powerful feature learning ability of deep learning, problems remain when applied to FER. First, deep neural networks require a large amount of training data to avoid overfitting. However, the existing facial expression databases are not sufficient to train the well-known neural network with deep architecture that achieved the most promising results in object recognition tasks. Additionally, high inter-subject variations exist due to different personal attributes, such as age, gender, ethnic backgrounds and level of expressiveness [32]. In addition to subject identity bias, variations in pose, illumination and occlusions are common in unconstrained facial expression scenarios. These factors are nonlinearly coupled with facial expressions and therefore strengthen the requirement of deep networks to address the large intra-class

variability and to learn effective expression-specific representations.



System Architecture

II.LITERATURE SERVERY

1. Near-field communication sensors and cloud-based smart restaurant management system, Hassain Saeed, Ali Shouman, Mais Elfar, Mostafa Shabka, Shikharesh Majumdar, In this paper, we introduce an efficient and user-friendly Smart Restaurant Management System. This system will solve key problems faced by restaurants today through the use of technologies such as Mobile and Web applications, Internet of Things (IoT), Near-Field Communications (NFC) sensors, and cloud computing. Restaurants have

many inefficiencies due to human limitations that can be resolved through

automation and device-to-device communication. This Smart Restaurant Management System accomplishes this by providing two interfaces for the two types of users in restaurants; an Android mobile application for customers and a web application for restaurant staff members. The Android mobile application allows customers to have a seamless dining experience with features such as finding available parking spaces easier through internet-connected infrared proximity sensors in the parking

lot, finding available tables at the restaurant easier through NFC sensors, ordering dishes through an interactive menu, and being able to pay the bill from their NFC equipped phones. The web application provides staff members benefits such as collecting data and statistics on the restaurant's performance in real time and automating the order placement system for waiters and cooks via IoT technology.

2. MobileNets: efficient convolutional neural networks for mobile vision applications, Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, We present a class of efficient models called MobileNets for mobile and embedded vision applications. MobileNets are based on a streamlined architecture that uses depth-wise separable convolutions to build light weight deep neural networks. We introduce two simple global hyper-parameters that efficiently trade off between latency and accuracy. These hyper-parameters allow the model builder to choose the right sized model for their application based on the constraints of the problem. We present extensive experiments on resource and accuracy tradeoffs and show strong

performance compared to other popular models on ImageNet classification. We then demonstrate the effectiveness of MobileNets across a wide range of applications and use cases including object detection, finegrain classification, face attributes and large scale geo-localization.

III.EXISTING SYSTEM

As there is no staff available in unmanned restaurants, it is difficult for the restaurant management to estimate how the concept and the food is experienced by the customers. Existing rating systems, such as Google and TripAdvisor, only partially solve this problem, as they only cover a part of the customer's opinions. These rating systems are only used by a subset of the customers who rate the restaurant on independent rating platforms on their own initiative. This applies mainly to customers who experience their visit as very positive or negative.

IV.PROPOSED SYSTEM

In order to solve the above problem, all customers must be motivated to give a rating. This paper introduces an approach for a restaurant rating system that asks every customer for a rating after their visit to increase the number of ratings as much as possible. This system

can be used unmanned restaurants; the scoring system is based on facial expression detection using pretrained convolutional neural network (CNN) models. It allows the customer to rate the food by taking or capturing a picture of his face that reflects the corresponding feelings. Compared to text-based rating system, there is much less information and no individual experience reports collected. However, this simple fast and playful rating system should give a wider range of opinions about the experiences of the customers with the restaurant concept.

V. MODULES

1. Face Detection:

Face detection or localization is an important step for image classification since only the principal component of face such as nose, eyes, mouth are needed for classification. Face detection algorithms can be broadly classified into feature, knowledge, template and appearance based methods. Our proposed system uses Viola Jones object detection algorithm for face localization which comes under feature based classification. Viola Jones object detection algorithm uses Haar featurebased cascade classifiers. The Haar Cascade classifier is extremely important element of the face detection.

The presence of the features in any of the input image is determined by the Haar features.

2. Facial Expression Recognition classification:

After learning the deep features, the final step of FER (Facial Expression Recognition) is to classify the given face into one of the basic emotion categories. Unlike the traditional methods, where the feature extraction step and the feature classification step are independent, deep networks can perform FER in an end-to-end way. Specifically, a loss layer is added to the end of the network to regulate the back-propagation error; then, the prediction probability of each sample can be directly output by the network. In CNN, softmax loss is the most common used function that minimizes the cross-entropy between the estimated class probabilities and the ground truth distribution.

3. Convolutional neural network (CNN):

CNN has been extensively used in diverse computer vision applications, including FER. At the beginning of the 21st century, several studies in the FER literature found that the CNN is robust to face location changes and scale variations and behaves better than the

multilayer perceptron (MLP) in the case of previously unseen face pose variations, employed the CNN to address the problems of subject independence as well as translation, rotation, and scale invariance in the recognition of facial expressions.

VI. REFERENCES

1. Deng, J., Guo, J., & Xu, J. (2019). A survey of deep learning for facial expression recognition. *Journal of Visual Communication and Image Representation*, 62, 228-244. <https://doi.org/10.1016/j.jvcir.2019.05.002>
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. <https://www.deeplearningbook.org/>
3. Khan, S. H., Hayat, M., & Shah, M. (2020). Facial expression recognition using deep convolutional neural networks: A comprehensive review. *Computer Vision and Image Understanding*, 198, 102001. <https://doi.org/10.1016/j.cviu.2020.102001>
4. Li, S., Deng, W., & Du, J. (2019). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, 10(3), 366-382. <https://doi.org/10.1109/TAFFC.2017.2747156>
5. Mollahosseini, A., Chan, D., & Mahoor, M. H. (2016). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1), 18-31. <https://doi.org/10.1109/TAFFC.2017.2753614>
6. Zhang, K., Zhang, Z., & Li, Z. (2018). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10), 1499-1503. <https://doi.org/10.1109/LSP.2016.2587778>
7. Liu, X., & Zhang, Y. (2017). A survey of facial expression recognition in the wild. *Journal of Computer Science and Technology*, 32(4), 697-718. <https://doi.org/10.1007/s11390-017-1760-6>
8. Jiang, H., Zhang, J., & Zhang, S. (2021). Facial expression recognition with deep learning: A review. *ACM Computing Surveys*, 54(3), 1-32. <https://doi.org/10.1145/3446348>
9. Hassain Saeed, Ali Shouman, Mais Elfar, Mostafa Shabka, Shikharesh Majumdar, and Chung Horng-Lung, "Near-field communication sensors and cloud-based smart restaurant management system," in Proceedings of

the 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), pp. 686-691, 2016.

10. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, "MobileNets: efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.