Chapter 9
Real-Time Facial Emotion Analysis for Adaptive Teaching Strategies Using Deep Learning

V. Suganthi
Vels Institute of Science, Technology, and Advanced Studies, India

M. Yogeshwari
https://orcid.org/0009-0001-2627-4814
Vels Institute of Science, Technology, and Advanced Studies, India

ABSTRACT

Facial emotion extraction is a process of identifying and extracting emotional information from human facial expressions. Due to its potential applications in a variety of fields, including psychology, marketing, and human-computer interaction, this technology has been gaining popularity recently. Technology for detecting facial expressions can be applied to smart classrooms to improve students' learning. By analyzing the emotions of students, teachers can gain insights into how engaged and attentive students are during the lesson and adjust their teaching style accordingly. This can help to improve the learning outcomes of students and create a more dynamic and engaging classroom environment. Facial emotion detection technology can be integrated into existing classroom tools, such as video conferencing software or smart boards. Students' facial expressions can be analyzed in real-time to identify emotions such as happiness, sadness, confusion, or boredom. This data can then be used to provide feedback to teachers about the effectiveness of their lesson and the engagement level of students. All papers found during the search will also sentence to review the current situation and pinpoint any potential gaps.

INTRODUCTION

Digital Image Processing (DIP) is a field of study that focuses on the processing of digital images using various techniques and algorithms. It involves the manipulation of digital images to improve their quality, extract information, or transform them into other forms that can be more easily interpreted by humans

DOI: 10.4018/979-8-3693-1355-8.ch009

or other machines (Abdullah & Sai, 2023). The digital images can be captured using various imaging devices such as cameras, scanners, and satellites (Alzubi et al., 2023). They can be in various formats like grayscale or color, two-dimensional or three-dimensional, static or dynamic (Anand et al., 2023). DIP techniques can be applied to these images to enhance or modify their characteristics for various applications such as medical imaging, remote sensing, robotics, and computer vision (Bin Sulaiman et al., 2023). Filtering, segmentation, extracting features, and detection of patterns are a few of the primary methods utilised in DIP (Biswaranjan Senapati & Rawal, 2023). These methods can be applied to digital photos to improve their visual quality, extract relevant information from them, find and classify items or patterns, and carry out a number of other tasks (Calo et al., 2023). DIP has become an essential component of many fields such as medicine, engineering, and science, and has wide-ranging applications in fields such as video surveillance, satellite imagery, and even social media (Chakrabarti & Goswami, 2008). As the field of digital image processing continues to evolve, new algorithms and techniques are being developed to meet the growing demands of various applications (Cirillo et al., 2023).

The Vision Transformer (ViT) and Convolutional Neural Networks (CNNs) are both powerful deep learning models used in computer vision tasks. While there are some similarities between the two, there are also big differences (Devi & Rajasekaran, 2023). The way ViT and CNNs analyze visual data is one of their key distinctions. CNNs use convolutional layers to extract spatial features from an image, while ViT processes the entire image as a sequence of tokens using self-attention mechanisms (Jasper et al., 2023). This allows ViT to capture global relationships between different parts of the image, while CNNs are better suited for extracting local features (Jeba et al., 2023). Another difference between the two models is their computational requirements. ViT requires significantly more memory and computational power than CNNs due to its larger number of parameters and the use of self-attention mechanisms (Jeba et al., 2023). This can make ViT less practical for some applications, especially those with limited resources (Lodha et al., 2023).

In terms of performance, ViT has shown promising results in image classification tasks, even Inseveral benchmarks, it performs better than the most advanced CNNs. While CNNs may be better suited for smaller datasets with more homogeneous images, ViT is particularly effective at identifying patterns in huge datasets with different images (Kanyimama, 2023). A form of machine learning algorithm known as an artificial neural network (ANN) is based on the structure and operation of biological neural networks found in the human brain (Priyadarshi et al., 2020). ANNs are layers of linked nodes, often known as "artificial neurons," that are trained on input data to make predictions and learn from it (Magare et al., 2020). Each synthetic neuron takes in one or more inputs and then uses a mathematical formula to generate an output. Until the output layer generates the final prediction or classification, the neurons in one layer's output become the inputs to the next layer. The Block Diagram for the Extraction of Facial Emotions is shown in Figure 1.

In order to reduce the discrepancy between its predictions and the actual outputs, the network modifies the strength of its connections between neurons throughout training (Minu et al., 2023). The backpropagation method enables the network to discover intricate linkages and patterns in the input data (Murugavel & Hernandez, 2023). ANNs come in a wide variety of forms, each with a unique architecture and set of uses. Convolutional neural networks, for instance, are utilised for image recognition and natural language processing, whereas feed forward neural networks, the most fundamental kind of ANN, are employed for straightforward classification tasks (Nagaraj & Subhashni, 2023). In general, ANNs are an effective machine learning technology that can be used to tackle a variety of issues, from forecasting stock prices to recognising objects in photos.