IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Survey On Human Facial Expression Recognition Techniques

1Sarita Sharma, 2Dr.Nirupama Tiwari

1Research Scholar, 2Associate Professor

1Sage University,

2Sage University

Abstract: Facial expression recognition is a crucial area of study in the field of computer vision. Research on nonverbal communication has shown that a significant amount of deliberate information is sent via facial expressions. Facial expression recognition is a crucial field in computer vision that deals with the significant impact of nonverbal communication. Expression recognition has lately been extensively used in the medical and advertising sectors. Facial emotion recognition is a technique that examines facial expressions in static images and videos to uncover information about an individual's emotional state. Facial expression recognition (FER) is a vital area of study within computer vision due to its significant impact on nonverbal communication. This review paper synthesizes recent advancements in FER, focusing on the integration of deep learning techniques and generative models to improve accuracy and robustness in diverse settings. Key contributions include enhancements in handling facial feature variability, occlusions, and lighting conditions. Notable methods such as Generative Adversarial Networks (GANs) and convolutional neural networks (CNNs) are explored for their effectiveness in creating and recognizing facial expressions. The research underscores the importance of both local and global facial features, demonstrating that a hybrid approach often yields superior results. Additionally, the paper discusses the application of FER in fields such as healthcare, human-computer interaction, and psychology, highlighting the broad implications of improved FER technology. By examining datasets, methodologies, and outcomes from various studies, this review identifies current trends and challenges, proposing directions for future research to enhance the understanding and interpretation of human emotions through facial expressions.

Keywords: Facial behavior analysis, Facial expression recognition, Datasets, Tools, Models, Complete facial recognition, incomplete facial recognition.

I INTRODUCTION

We can learn a lot about other people's intentions just by looking at their faces when they're upset or happy. A person's facial expressions and vocal intonation may often tell it a lot about their emotional state, including whether they're happy, angry, or sad.[1] while language modules make up one-third of human communication, non-verbal variables constitute two-thirds. Like other nonverbal components, facial expressions convey emotional meaning and are thus one of the most significant sources of information in interpersonal communication.[2-3] Numerous studies on automated facial expression analysis have been conducted because to its practical significance in socially adept robotics, medical treatment, fatigue tracking for drivers, and numerous other computer-human interaction programs. The domains of computer vision and machine learning explored many facial expression recognition (FER) algorithms to encode expression information from face representations.[4]Based on a cross-cultural investigation, Ekman and

Friesen[5] identified six main emotions in the early 20th century. This suggested that those core emotions are universally understood by humans, irrespective of cultural context.[6] Disgust, wrath, terror, joy, sadness, and surprise are the classic face expressions. Our regular adaptive displays, our ability to reflect their essence, and the fundamental emotions have always formed the core of the impact model. A wider range of emotions is reflected in models of emotional depiction like the stable model employing effective dimensions and the facial action coding system (FACS).Since facial expressions reveal the essence of people's emotions instantly, it is feasible to successfully communicate sentiments through them. Computer vision and AI researchers have shown that techniques developed to distinguish between emotional expressions on a person's face are very effective. The development of these techniques was motivated by the need to identify emotions from facial expressions [7] Visual inputs, rather than bodily connections, are better for emotion recognition. This remains true even if this function could be performed by means of wearable sensors. This preference for visual modalities of emotion detection is explained by the fact that these modalities offer improved relevance and variation [8].People produce a wide variety of facial expressions, each of which is shaped by the underlying emotions they are experiencing. Each facial expression is uniquely represented by a unique set of personality qualities and a unique distribution scale. as shown in the following:[4],The cranium One crucial aspect of human-computer interaction (HCI) is expression recognition, or FER for short. Computers' capacity to recognize and comprehend human emotions is enhanced by looking at images of people's faces [6]. Perhaps the evolution of computing over the years is to blame for the meteoric ascent of AI systems. Determining an individual's emotional state by observing their facial expressions is extremely important for HCI applications that include real-time interaction. Applications that utilize AI systems are one type of such application. in references [[9]. Alternative possibilities include apps for humanoid robots. Applications that recognize emotions have greatly aided the development and broad adoption of FER algorithms in various areas, including healthcare, security, safe driving, video gaming, and more. This method has been demonstrated to be essential in human-computer interaction and to produce intelligent results across several fields. Facial emotions are beneficial for investigating human behavior [**23**,**24**] as exhibited in **Figure 1**. Psychologically, it is proven that the facial emotion recognition process measures the eyes, nose, mouth and their locations.

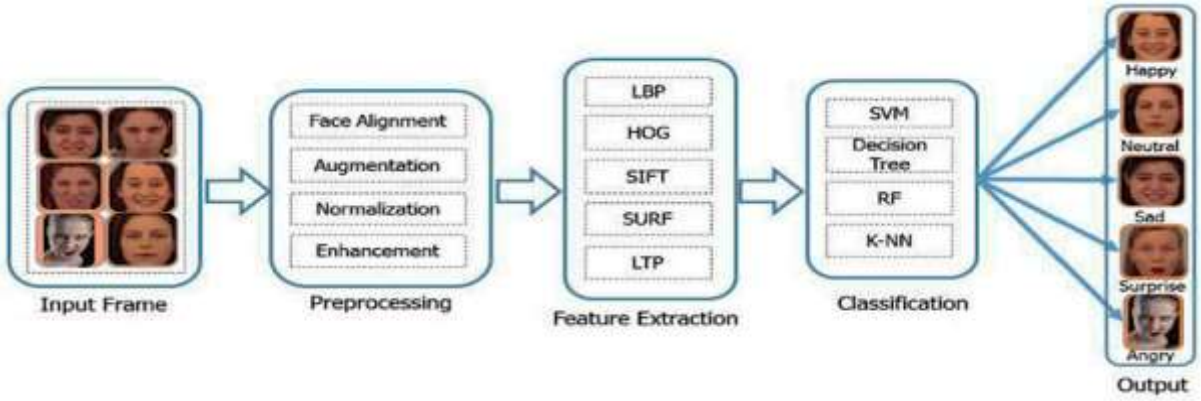

Figure 1. Facial emotion recognition (FER) process.

To facilitate the deployment of these applications, a number of methods have been proposed with the aim of automating face expression recognition. Particularly in controlled environments, these strategies produce outstanding outcomes [11]. Facial expressions of emotion are complex and dynamic, making it challenging to use traditional FER methods to analyze them. The advent of deep learning, most notably Convolutional Neural Networks (CNNs), has revolutionized the field by allowing automated algorithms to quickly learn accurate patterns from facial images [12]. A paradigm shift has occurred as a consequence of this. Recent work using CNNs for face detection tasks has demonstrated promising results, particularly with forward-facing, obstacle-free images CNNs have also shown remarkable results in several facial recognition tasks. However, when faces are obscured, whether by angles that obscures features or any other factor, the task becomes more challenging. It is possible to combine this with additional elements. Being able to recognize and understand expressions on faces that display a wide variety of sizes, angles, and positions is a valuable skill. However, in unregulated environments, the issue of developing accurate automated FER remains a

significant hurdle [13]But in everyday life, people frequently cover their faces with their hands or other accessories like scarves, hats, glasses, masks, or similar objects. The identifying process becomes more difficult as a result of this. Some examples of such accessories include eyeglasses, scarves, caps, and masks. The academic literature has offered a number of unique approaches to addressing these occlusions [14] below, they will go into greater depth about these tactics.

In addition to words, minor variations in facial expressions are a common way for humans to convey their feelings. It convey a great deal of emotion through our facial expressions, whether it's a temporary smile, a tightened jaw, or an arched eyebrow. By leveraging computer vision and artificial intelligence, FER enters this intriguing world and deciphers these emotional indicators from photographs and videos. Picture this: In the future, computers will be able to read our emotions just by looking at our faces, opening the door to more subtle and organic forms of human-computer connection... In order to increase engagement and comprehension, educational systems can adjust to students' emotional states and personalize their learning experiences. By responding to pedestrians' facial expressions, autonomous vehicles can increase safety and reduce the likelihood of accidents. The results of consumer surveys and other market research tools' analyses of product and ad reception might inform more effective advertising and promotion campaigns [15].

Background Theory

There are countless unspoken messages conveyed by the human face, a masterpiece of complex expressions and subtle shifts. For decades, academics have been engrossed with the development of intelligent systems that evaluate face features and infer emotional states through facial expression recognition (FER). The goal is to decipher this unspoken language. Datasets, models, and face-parts are the three main components that interact to determine the accuracy and performance of these systems [10]. Think of a little kid trying to figure out what animals are in a picture book using just photographs of puppies and kittens. As they encounter more diverse organisms, their initially distorted knowledge will gradually grow and improve. The performance of a FER system is also affected by the size and variety of the dataset that was utilized for training. Dimensions are important: To improve the model's generalizability and adaptability, bigger datasets expose it to more different facial expressions. Greater accuracy in recognition is achieved when there is an abundance of data that enables the model to understand subtleties both within and across expressions. Nevertheless, data alone won't solve the situation [4].

Diversity is key: The fabric of human faces should be reflected in the dataset used to train a FER system that is really robust. Ensuring gender and ethnic balance is just the beginning; other considerations that must be taken into consideration include age, cultural expression variances, and even environmental elements such as lighting and occlusion.

Misunderstandings and inaccurate results, especially for underrepresented groups, might arise from biases that exist within relatively homogeneous datasets. Combating the bias monster: Expressions on the face are not always expressive. If the training data is biased towards a given group or demographic, misinterpretations can occur due to cultural expressions, subtle micro-expressions, and even individual variances in muscle movement [19]. Careful data curation, diversified datasets, and methods to reduce imbalances, such as fairness-aware algorithms, are necessary to combat bias. The advent of deep learning has made Convolutional Neural Networks (CNNs) the go-to model for FER. Significant performance improvements have been achieved thanks to their capacity to automatically extract features from big datasets. Nevertheless, convolutional neural networks (CNNs) limit their usefulness in contexts with limited resources due to their data appetite and computational expense [16]. Besides convolutional neural networks, more conventional ML models such as Random Forests and Support Vector Machines (SVMs) provide interpretability and efficiency benefits. These methods are particularly useful for smaller datasets and can shed light on how the model makes its decisions. Another popular strategy is a hybrid one that uses both deep learning models and more conventional methods to get the best results possible [16].

Emerging possibilities: Outside of current paradigms, FER's future offers intriguing prospects. Generative models have the potential to generate new datasets in order to solve problems with data scarcity, and Recurrent Neural Networks (RNNs) demonstrate potential in comprehending the time-dependent behaviour of facial expressions.

The accuracy and robustness of FER systems can be further improved by exploration of these areas. Certain facial features are more expressive than others. Eyebrows, eyes, and the lips are a powerful trio that frequently communicate most of the emotional information. Missing other small clues, such as wrinkles on the forehead, the development of dimples, or even tightness in the neck, might result in incorrect judgments [17]. Looking only at certain things: Attention to prominent facial traits can boost efficiency, especially when resources are limited. Feature selection approaches can improve efficiency and decrease computing burden by identifying the most informative sections of the face. However, if the system is overly dependent on one or more traits, it may be susceptible to occlusions or differences in human face architecture. Complete harmony: An optimal strategy would combine feature selection with comprehensive analysis. To achieve the best possible accuracy and resilience, FER systems harvest data from critical areas while keeping track of the whole face landscape. One can gain a more complete picture of the expression by seeing how different facial features interact with one another and the overall face.

Our paper's notable contributions are summarized as follows:

Taxonomy: They provide a novel way to categorize FER and related datasets. Both classic methods and cutting-edge deep learning techniques are represented in this category. In the context of FER, it incorporates techniques like graph-based algorithms, transformers, and Generative Adversarial Networks (GANs). Furthermore, datasets are classified according to a separate taxonomy, with pictures and sequences being the two main types of datasets. There are two subcategories here: controlled and uncontrolled. Examples of the controlled group include film and laboratory settings, whilst examples of the uncontrolled group include "in the wild" environments. As far they are aware, no prior studies have tackled this all-encompassing categorization.

Comprehensive review:- Recent findings from research conducted by newcomers, who have not been covered in previous surveys, are presented in this work [16].

Highlighting top models: They offer evaluation findings from around 60 approaches, covering the most significant ones across various datasets.

Overview of popular datasets: the most commonly used datasets are introduced.

II RESEARCH METHODOLOGY

The research methodology for a comprehensive study on facial expression recognition involves several systematic steps to ensure a thorough and high-quality review of existing literature and the generation of new insights. The methodology includes the following components:

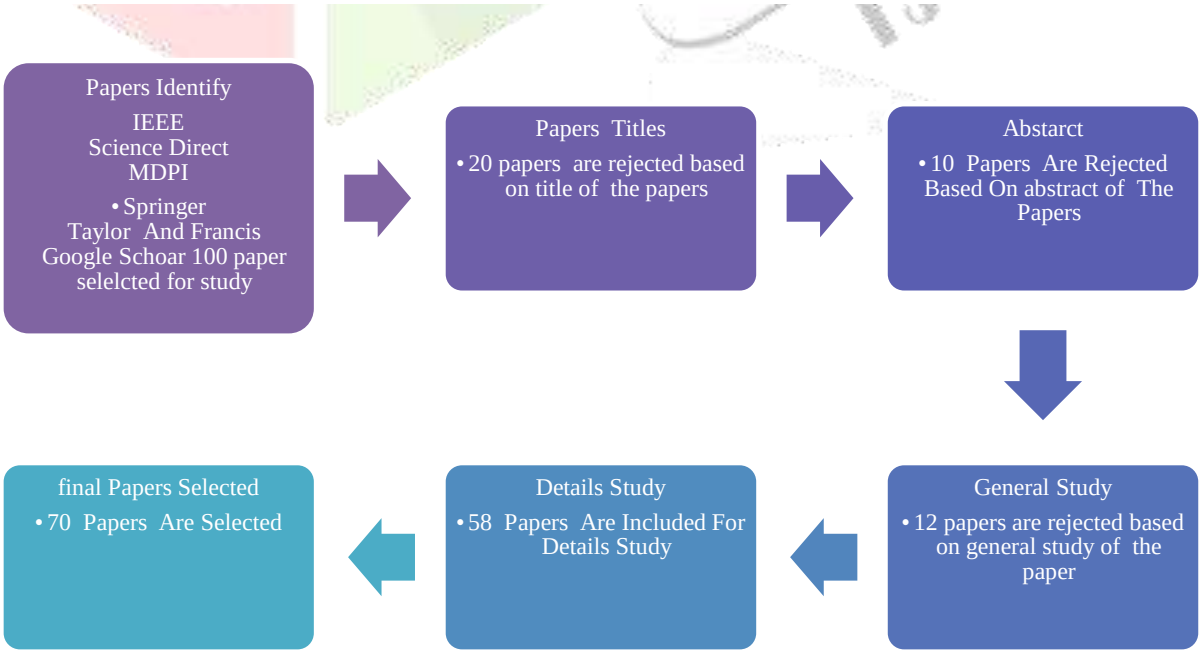

Figure. 3. Research methodology

Research Design

The study employs a systematic literature review (SLR) design, aiming to identify, evaluate, and synthesize existing research on facial expression recognition. This approach ensures a comprehensive understanding of the current state of knowledge and identifies gaps for future research.

Define Keywords and Phrases

To conduct a comprehensive literature review on facial expression recognition, we first need to define relevant keywords and phrases. These keywords should encompass the main themes and concepts of the research area:

Facial expression recognition, Deep learning algorithms, Convolutional Neural Networks (CNNs),Facial analysis, Datasets in facial recognition, Privacy concerns in facial recognition

Database Selection

Selecting appropriate databases ensures access to a wide range of relevant and high-quality academic resources. Key databases for this research include:

- IEEE Xplore
- PubMed
- Springer Link
- Science Direct
- ACM Digital Library
- Google Scholar

Initial Search

Using the defined keywords and selected databases perform an initial search to gather a broad range of articles, conference papers, and reviews related to facial expression recognition. Example search queries include:

- "Facial expression recognition using deep learning"
- "Emotion detection with CNNs"
- "Privacy concerns in facial recognition technology"
- "Whole face vs partial face recognition accuracy"
- "Applications of facial expression recognition in medicine"

Screening Titles and Abstracts

Review the titles and abstracts of the search results to determine the relevance of each study. Exclude any articles that do not directly pertain to the focus of the research or are clearly out of scope.

Full-Text Review

For articles that pass the title and abstract screening, conduct a full-text review to ensure they meet the inclusion criteria and contribute valuable insights to the research objectives. This step helps in refining the selection of studies for in-depth analysis.

Data Selection

Inclusion Criteria

- Studies published in peer-reviewed journals and reputable conference proceedings
- Research focusing on facial expression recognition
- Papers discussing the use of machine learning and deep learning algorithms, specifically CNNs
- Studies addressing both whole face and partial face recognition
- Research highlighting applications in various fields such as medicine and advertising

- Articles discussing privacy concerns related to facial recognition technology

Exclusion Criteria

- Studies not available in full-text format
- Articles published in non-English languages without an available translation
- Research focused on facial recognition for identity verification rather than emotion detection
- Studies with insufficient methodological detail or low-quality assessments

III LITERATURE STUDY

Face detection Algorithm

The process of determining the locations and sizes of human faces in arbitrary digital images is referred to as face detection. This approach applies to computers. Rather than focusing on other things, such as buildings, trees, or other bodies, it is only able to identify facial features. There are some systems that are able to detect and find faces simultaneously, while others first carry out a detection routine and then attempt to locate the object in question. An algorithm for tracking is required. Newer face detection algorithms try to handle the more general and challenging problem of multi-view face detection, whereas earlier face detection algorithms concentrated on the detection of first-person human faces. Variations in the image or video are taken into consideration by these algorithms. These variations include the appearance of the face, the lighting, and the stance. Face detection is an issue that needs to be solved by determining whether or not a picture contains a face. This is a two-class problem. A simplified version of the facial recognition problem can be seen in this approach. Face recognition should be able to classify a certain face, and there could be a wide variety of classes that are contenders. That being said, a great number of face detection approaches are extremely comparable to face recognition algorithms. In a different way of putting it, the methods that are utilized in face detection are frequently utilized. In the field of biometrics, face detection is utilized, frequently, as a component of a facial recognition system or in conjunction with it. Image database management, human computer interface, and video surveillance are some of the other applications for this technology. Recent digital cameras have begun to implement facial detection for the purpose of autofocus.

Viola-Jones Algorithm: When it comes to the detection framework, the features that are utilized always involve the sums of image pixels that are contained within rectangle areas. The Haar basis functions are the foundation for this. In figure 2, this is an illustration of the four distinct sorts of features that are utilized. It is possible to determine the value of any particular feature by subtracting the total number of pixels included within shaded rectangles from the total number of pixels contained within clear rectangles. In comparison to other options, such as steerable filters, rectangular features of this kind demonstrate a relatively low level of sophistication. Because of their sensitivity to both vertical and horizontal aspects, the input they provide is far coarser. With the use of an image representation known as an integral image, rectangular characteristics are able to be evaluated in a constant amount of time. This provides them with a significant speed advantage over their more complex relatives. The fact that every rectangular area in a feature is always adjacent to at least one other rectangle means that any feature with two rectangles can be computed using six array references, any feature with three rectangles can be computed using eight array references, and any feature with four rectangles can be computed using just nine array references.

Exhaustive Search Algorithm: One of the most essential technologies in face information processing is human face detection. The speed at which this technology operates is of utmost significance when it comes to real-time face detection for input images or input video sequences. When it comes to identifying faces in grayscale images, this technique introduces a revolutionary face window searching strategy that is based on evolutionary agent. By utilizing the evolutionary computing of dispersed agents, which each represent a different form of window, it is able to swiftly and the potential face windows. The findings of the experiments demonstrate that the evolutionary agent-based searching algorithm has the potential to boost the detection speed by a factor of five to seven times when compared to the conventional exhaustive searching method that is utilized in certain generic algorithms. The conventional method of searching for face windows begins in the upper left-hand corner of the image and proceeds to classify all of the possible sub windows until it identifies the windows that satisfy the classification requirements. Because there are typically a great number of areas in the input image that do not contain faces, the exhaustive searching

method is unable to make a speedy identification of the face like windows. It not only has a high false detection rate but also requires a significant amount of computational resources. It is common knowledge that humans are able to locate the face regions of an image in a short amount of time. It is relatively simple for a human being to determine the number of faces that are present in an image by counting them. If there are several faces in the image, individuals will concentrate on them simultaneously rather than searching the face windows that run from the top left corner to the bottom right corner of the image.

Branch and Bound algorithm: Recently, researchers have proposed many face recognition methods with the aim of improving the accuracy rate of face recognition. However, few face recognition methods only focus on computational cost. For reduction in the computational cost of face recognition, an effective face recognition method using Haar wavelet features and a branch and bound method is used. In this method extracts features of the Haar wavelet from a normalized face image, and recognizes the face by classifiers learned with the Ada-Boost M1 algorithm. To accelerate the recognition process we select features according to the accuracy of classification and apply a branch and bound method to the recognition tree into which the classifiers of an individual in the face database are merged.

Feature Extraction/Reduction Method

The feature extraction phase represents a key component of any pattern recognition system. Feature extraction involves detecting and isolating various desired features of patterns. It is the operation of extracting features for identifying interpreting meaningful information from the data. In facial expression recognition this is an essential pre-processing step as in pattern recognition. Following methods are used in feature extraction.

Fisher's Linear Discriminate: (FLD): The classification process is carried out in this space once the high-dimensional data is projected onto a line using this method. According to the projection, if there are two classes, the distance between the means is maximized, and the variation within each class is minimized. By projecting the data into a space that is low dimensional and maybe uncorrelated, it is able to limit the amount of variables that are introduced into the system. Input features are reduced to a level that is more manageable as a result of this effect. The projection process allows for the exclusion of certain variables that contain information that is not associated with face expression. The neural network is able to avoid learning unnecessary details from the input thanks to this advantage. Because of this, the performance and generalization of the network classifiers are optimized. It is at its most effective when applied to classification tasks, as it enhances both performance and the approach of feature reduction.

Principal Component Analysis (PCA): PCA is a linear transformation technique that transforms the data to a new coordinate system in such a way that the highest variance caused by projection of the data comes to lie on the first coordinate (which is referred to as the principal component), the second greatest variance on the second coordinate, and so on [14].By conserving lower-order principle components and disregarding high-order ones, principal component analysis (PCA) can be utilized to reduce the dimensionality of a data set while preserving the properties of the data set that are primarily responsible for its variation. It is common for such low-order components to contain the most significant parts of the data. Experience and to improve its performance on a job as it gains more prior experience [15]. **Deep Learning Era**

Deep learning, also known as DL [16], is currently regarded to be the fundamental technology that will be used in the Fourth Industrial Revolution. It has become one of the most popular research areas in the disciplines of machine learning (ML) and artificial intelligence (AI) this year. Image recognition, natural language processing (NLP), speech recognition software, genome engineering, and systems biology are the primary areas in which deep learning is utilized. [17] Deep learning also has applications in the field of systems biology. Deep learning models are becoming increasingly popular as an alternative to traditional ones for a number of reasons, including the following reason: Hardware developments, particularly graphics processing units (GPUs), speed up deep learning. (2) Deep learning has the ability to learn from raw data, which eliminates the need for manual feature engineering. As a result, the models are able to process information more quickly and effectively capture complicated patterns. (3) Additionally, deep learning models are able to build hierarchical representations of data, which allows them to capture detailed correlations and improve performance in tasks such as speech recognition, natural language processing, and identification of images. Traditional approaches frequently require assistance in order to successfully collect high-level properties in these areas. Deep learning models are able to tackle large-scale datasets and complicated challenges because of the intrinsic scalability that they possess. Improved performance can be achieved by training deep learning models in an effective manner on enormous amounts of data. Also, deep

learning models are able to learn end-to-end, which means that they can learn directly from input to output without relying on intermediate steps. This is a significant advantage.

The overall pipeline is simplified as a result of this. It has been demonstrated that deep learning models are exceptionally adaptable across a wide range of areas. In a variety of domains, including computer vision, they have accomplished achievements that are considered to be state-of-the-art.As a result of research and development efforts that have been ongoing over the years, deep learning has seen considerable breakthroughs. In the process of continuously introducing new architectures, algorithms, and optimization approaches, performance is being improved, and the frontiers of what is achievable in many applications are being pushed further and further. Before applying additional independent classifiers, such as a support vector machine (SVM) [18] or decision tree, it is possible to, in addition to end-to-end learning, make use of a CNN or another deep neural network as a strategy for feature extraction. This is an alternative to end-to-end learning. Using data augmentation approaches, as opposed to more traditional ways, is absolutely necessary in order to improve the performance of deep learning models.

Data Augmentation

In general, the majority of FER datasets that are accessible to the public do not contain enough pictures for training. The major objective of the data augmentation stage in deep FER is to expand the size of a training dataset. As a consequence, this stage is an essential part of the process. The term "data augmentation" refers to the process of producing additional data points from the existing data in order to artificially enhance the amount of data. In addition, data augmentation is yet another key preprocessing feature that is utilized excessively in deep learning. Online data augmentation and offline data augmentation are the two categories of technologies that fall under the category of data augmentation. Increasing the amount of variety in the dataset can be accomplished quite effectively through the use of online augmentation in the training data loader. The enhanced data, on the other hand, are generated in a random fashion using a variety of methods, such as the GAN model, and the data loader follows the same pattern when sampling the data. In order to acquire a high level of accuracy, it is possible that a model will require extensive training. This can be avoided by the use of offline augmentation, which also produces a dataset that contains the necessary augmentations. When it is either impossible or prohibitively expensive to gather and classify data, offline augmentation can be used to increase the amount of the dataset.

The most frequent methods of augmentation, also known as offline augmentation, are rotation, flipping, saturation, translation, scaling, cropping, brightness, color augmentation, and contrast, among other techniques. It is always necessary to have a large amount of training data in order for deep learning to deliver reliable classification results when utilizing CNNs. In the field of machine learning, the term "overfitting" refers to an undesirable behavior in which a model is able to make correct predictions based on existing training data, but fails to perform well when confronted with fresh data that it has not before encountered. It is possible for the model to become susceptible to overfitting when the training procedure is carried out over a large number of epochs and the capacity of the network is high. This translates to the model performing well on the training set but failing to generalize successfully, which results in a modest training error but a big validation or test error. This becomes a significant problem due to the fact that the data base is so limited. There is a significant reduction in overfitting that occurs most frequently when the database is expanded using artificial label-preserving alterations.

Therefore, before training the CNN model, we add to the database by utilizing a variety of alterations to make a large number of very minor changes in appearance and orientation. Through the utilization of easy data augmentation approaches, the network was made more resistant to a variety of circumstances. According to observations, deep learning has achieved performance levels that are at the cutting edge in a variety of applications [19]. This section is a quick introduction to certain challenges that have been encountered in the application of FER. Following this, a little history on the deep learning methods that have been utilized in FER is presented, and then the deep learning approaches are discussed.

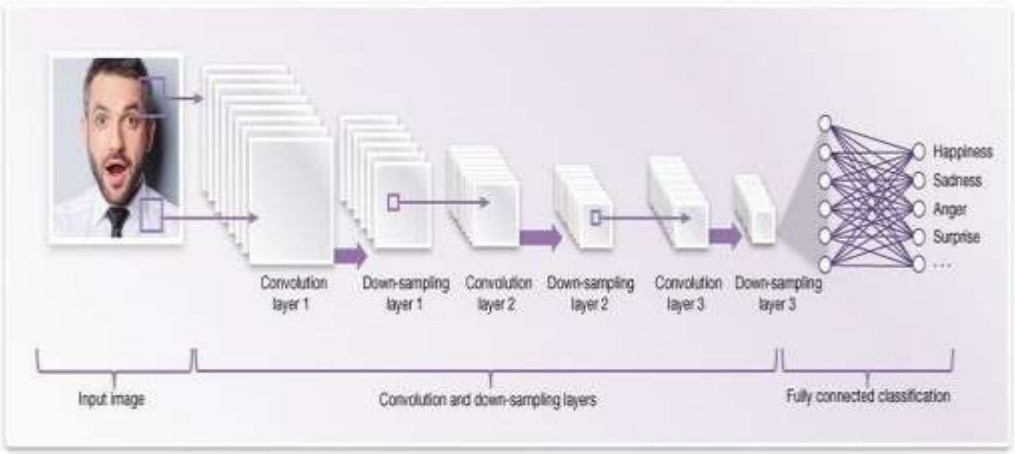

Figure 4. CNN architecture (https://github.com/somillko/Facial-Expression-Recognition accessed on 1 February 2023).

Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNN) [20] are specialized neural networks that are designed for processing data with a grid-like topology. This includes time-series data that is arranged in a one-dimensional grid, picture data that is grid-like in two dimensions, and volumetric data that is grid-like in three dimensions [21].Convolutional neural networks (CNNs) have become the dominant method in deep learning to accomplish a variety of tasks. Due to the fact that they develop hierarchical representations of images, CNNs are powerful tools for image identification tasks. They have been successfully applied in a multitude of computer vision applications, including face detection, facial expression recognition, object detection, self-driving or autonomous cars, auto-translation, text prediction, handwritten character recognition, climate analysis, X-ray image analysis, cancer detection, visual question answering, image captioning, biometric authentication, document classification, and 3D medical image segmentation [22]. CNNs have shown to be both versatile and effective in tackling a variety of visual recognition challenges, as demonstrated by these uses of CNNs.What makes them so effective is the fact that they are able to withstand a wide range of alterations, such as modifications in scale and face placement [23]. The shift-invariance, convolutional operations, and pooling layers that they possess allow them to perform better than multilayer perceptron (MLP) models when it comes to handling variances. CNNs are able to collect and extract features independent of factors such as facial location or scale changes because of their resilience. The Convolutional Layer, the Pooling Layer, and the Fully Connected Layer are the three primary types of layers that are typically found in a general CNN design. In most cases, the first layer of a CNN is a convolutional layer. This layer applies a collection of learnable filters to the input image, resulting in the generation of activation maps that include specific characteristics. Once the data has been processed through the many layers of the CNN, the network is able to recognize increasingly intricate patterns, such as face characteristics, contours, and ultimately the complete item. Numerous convolutional layers can be present in CNNs, which enables the network to construct hierarchical representations of the picture that is being fed into it. The number of neurons in a layer that are connected to the same input area is determined by the depth hyperparameter. Deeper networks are more complicated than networks with fewer neurons. In each stage of the convolution process, the stride hyperparameter helps to determine the number of pixels that are moved across the input matrix. Controlling the size of the output volumes is accomplished by the use of zero-padding, which involves adding zeros around the margins of the input. An example of a CNN architecture may be found included in Figure 8.

A pooling layer is typically added after the convolutional layers, which is sometimes referred to as down-sampling. The purpose of this layer is to minimize the size of the convolved features, which in turn reduces the amount of computing work that the network needs to complete [24].By pooling, complexity is reduced, efficiency is improved, and the risk of overfitting is reduced more significantly. When a CNN is constructed, the last layer is often a fully connected layer. This layer is responsible for performing the classification task based on the features that were retrieved by the layers that came before it. The fully connected layer (FC) [210] is made up of the weights and biases in addition to the neurons, and every node

in the output layer is connected directly to a node in the layer that came before it. This layer makes it possible for the 2D feature maps to carry out the classification work. It also makes it possible for the 2D feature maps to be based on the features that were extracted through the layers that came before them and the various filters that they included. When it comes to tasks like facial expression detection, CNN models that have been pre-trained, such as AlexNet [25], VGG [26], VGG-face [27], GoogleNet [28], Inception [28], and ResNet [29], can be especially helpful. The usefulness of these models in extracting significant features from images has been demonstrated by their training on large-scale image datasets such as ImageNet. Figure 5 demonstrates that both AlexNet and VGG contain a very large number of parameters, the majority of which are a result of their completely connected layers. GoogLeNet and ResNet, on the other hand, have a lower number of parameters, yet they nonetheless manage to achieve an accuracy rate of roughly 70 percent. The most recent iterations of Inception and ResNet have flattened the steep straight line that previous architectures have been following, which indicates that an inflection point is getting closer where costs begin to have a greater impact than the gains in accuracy. According to the plot, Inception V4, which is a mix of ResNet and Inception, has an amazing accuracy rate of 80% and may be the best option.

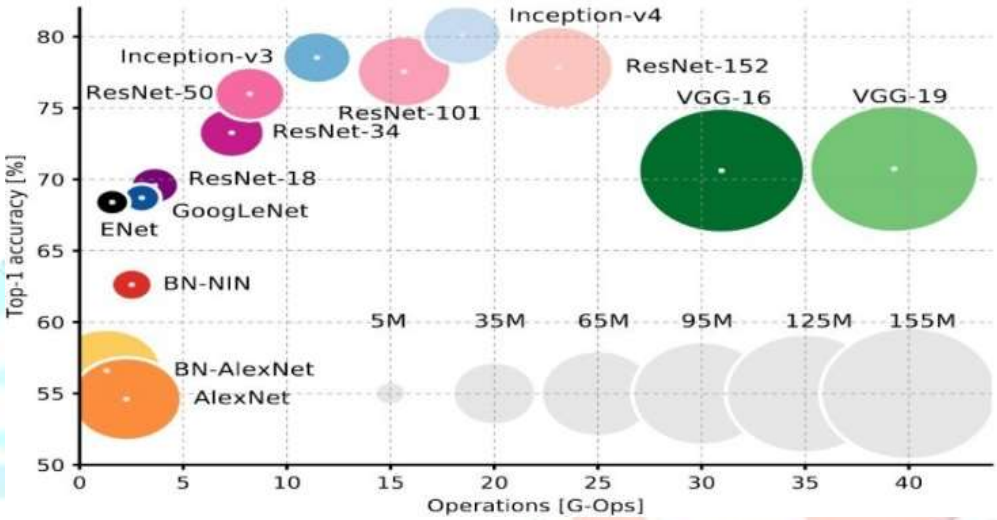

Figure 5 Accuracy vs. Operations, Size Parameters.

A proposal was made by [30] to enhance the dataset in order to construct high-capacity models without overfitting, which would ultimately result in improved facial expression recognition (FER) performance. Another interesting model for FER is Xception [31], which has demonstrated small gains over Inception V3 on the ImageNet dataset. Xception is a promising development. For the purpose of efficiently capturing spatial data, Xception employs depthwise separable convolutions technology. It is also recommended to use DenseNet [32] since it improves feature propagation, stimulates feature reuse, minimizes the number of parameters, and helps to mitigate the problem of disappearing gradients. In addition to this, EfficientNet [33] has established itself as one of the most successful models for understanding image recognition. From EfficientNet-B0 all the way up to EfficientNet-B7, its model versions are available. In order to attain state-of-the-art precision while minimizing the amount of computer resources required, these models are designed. Nevertheless, it is essential to keep in mind that training a Convolutional Neural Network can be a time-consuming process, particularly when applied to big datasets, and may necessitate the utilization of specialist hardware such as GPUs. Conventional convolutional neural networks (CNNs) are fantastic at extracting spatial features from input images; but, they struggle to capture the temporal interactions that occur in video sequences for some reason. The existence of this constraint underscores the necessity of utilizing specialist architectures, like as recurrent neural networks (RNNs) or spatiotemporal models, in order to effectively simulate the temporal dynamics of video videos. FER and other image recognition tasks can be considerably advanced by the use of tactics such as data augmentation, the utilization of sophisticated models such as Xception, DenseNet, and EfficientNet, and the exploration of specific architectures. For the purpose of capturing the spatial and temporal features of video clips, Tran et al. [34] proposed the use of 3D-CNNs. The most significant disadvantage has to do with the rise in the total number of training parameters.

It has been previously established that preprocessing is an extremely important factor in the overall performance of deep learning models. The techniques of data augmentation, cropping, downsampling, and normalization are considered to be conventional methods for enhancing the robustness and accuracy of various models. Through the application of a variety of changes to the initial data, the process of data

augmentation entails the generation of extra training samples. According to the findings of [24], the combination of various preprocessing methods has the potential to considerably increase the accuracy of CNNs. A zero-bias model was presented by [25] for the fully connected layer in CNNs. This model offers an additional avenue for optimization. According to the findings of certain researchers, increasing the level of sophistication and depth of CNN architectures has the potential to enhance the accuracy and performance of computer vision algorithms. FaceNet2ExpNet is a relatively new architecture that was proposed by 26]. Based on a face net that had been updated in the past, the author initially proposed a probabilistic distribution function as a means of describing the high-level neuron response. This results in feature-level regularization, which makes use of the extensive face information that the face net possesses. It was also suggested that label supervision be implemented during the second phase in order to improve the final discriminative capabilities.

Multitask Networks

As a comprehensive framework for resolving the inherent challenges of facial expression recognition (FER), multitask learning has gained gaining recognition as a possible solution. Subject identification, lighting conditions, and head position are only some of the parameters that can have an effect on FER in real-world applications. However, these are not the only aspects that can have an effect. This delicate interaction between the many latent components cannot be captured by traditional FER models, which typically concentrate on learning a single job. Multitask learning paradigms, on the other hand, provide a more thorough approach: they incorporate extra relevant tasks, which in turn improve the sensitivity of features to expression-specific cues and reduces the impact of confounding variables [27]. In addition, multitask networks are capable of performing two distinct functions: categorization and data enhancement or enhancement. Although multitask learning has demonstrated its potential in a variety of applications, one of the most significant challenges is figuring out how to assign the appropriate amount of importance to each separate job. Furthermore, the effectiveness of the multitask learning model is greatly impacted by these weights.[29] Presented a multilingual convolutional neural network (MSCNN) that was trained under dual supervision for recognition and verification tasks. Every job has its own loss function, which is designed to improve the ability to differentiate between different facial expressions while also reducing the amount of variation that occurs within the same class. By taking this strategy, the model's attention is efficiently directed toward the subtleties of facial expressions.[30] have introduced a dynamic multitask learning framework that is capable of adaptively updating task weights based on their relevance throughout the training process. This work is noteworthy because it was done by Ming et al. Furthermore, this dynamic technique achieved outstanding performance metrics, such as an accuracy of 99.5% on the CK+ dataset and an accuracy of 89.6% on the Oulu-CASIA dataset which demonstrates its practical utility in real-world applications. [31] An novel solution was developed by combining the well-known LightFace face recognition library with a facial feature analysis that takes into account many levels of complexity. In order to obtain a high performance in face recognition, this hybrid framework makes use of the most advanced face recognition architectures available today. These architectures include VGG-Face, Google FaceNet, OpenFace, Facebook DeepFace, DeepID, ArcFace, Dlib, and SFace. Notably, the framework makes use of the same preprocessing techniques as are utilized by the aforementioned recognition models. These techniques include face detection and alignment. For the purpose of expanding its scope of application, the system also conducts assessments based on factors such as age, gender, sentiment, and race. A basic VGG-Face model is developed with pre-trained weights, which serve as the cornerstones for the whole solution. Both robustness and simplicity are taken into consideration when developing this model.

Savchenko has made significant contributions to multi-task learning networks, and some of his work has achieved accuracy metrics that have never been seen before [39,40]. In an important study, Savchenko presented a multi-task learning model that simultaneously addresses face identification, gender, ethnicity, and age classification on the UTKface dataset and emotion recognition using the AffectNet dataset. This model was developed by Savchenko. MobileNet, EfficientNet, and ResNet are examples of lightweight backbone architectures that are utilized by the model in order to achieve high levels of computational efficiency. Among the preprocessing processes is the application of the MTCNN algorithm for face detection, followed by cropping.

Furthermore, Savchenko and his team released Multi-task EfficientNet-B2 and its variation, Multi-task EfficientNet-B0 [32]. Both of these networks were developed in following work. These architectural designs are intended to carry out a variety of functions, such as the detection of faces, the identification of

faces, and the recognition of facial expressions. In the beginning, face sequences for each individual participant are extracted by utilizing a combination of face detection, tracking, and clustering approaches. After been pre-trained on facial recognition tasks, a single neural network is then fine-tuned to recognize emotional characteristics in each frame. This is accomplished by utilizing a powerful optimization approach that has been specifically designed for static images taken from the AffectNet dataset.

Notably, the extracted facial features can be utilized to quickly infer collective emotional states, individual emotional expressions (such as happiness and sadness), and varied levels of student engagement, ranging from disengagement to high involvement. This is a significant advantage. This model has established the state-of-the-art benchmark for distinguishing eight distinct emotions, with a 63.03% accuracy rate on the large-scale AffectNet dataset. It is important to highlight this fact because it has already established the benchmark. In addition to this, it has achieved a level of accuracy of 66.29%, which places it in third place for the seven-emotion categorization. [33] The Discrimi-native Deep Multi-Tasking Learning (DDMTL) framework, which includes a Siamese-based loss function, was proposed as a complex multi-task learning system. Improved facial expression recognition capabilities are the result of this novel approach, which includes data distribution information and expression labels. [34] A multi-task learning system that is particularly adept at recognizing facial emotions "in the wild" was proposed, which contributed to the advancement of the subject. Their method makes use of Graph Convolutional Networks (GCN) in order to take advantage of the complex relationships that exist between categorical and dimensional feelings.

[35] Developed a model that produced amazing performance metrics in a landmark study that was published in 2021. The model attained total accuracy of one hundred percent on the CK+ dataset and ninety-four percent on the FER+ dataset. The architecture comprises three critical components: a primary deep learning model, which can be instantiated as either AlexNet or Inception; dual attention mechanisms, specifically, Grid Wise attention for low-level feature extraction and Visual Transformer Attention for high-level feature discernment; and a FER module capable of differentiating between simple and complex facial expressions, utilizing the C-F labelling technique as a reference point. To add insult to injury, they also presented a large-scale model variant that combines the first two components in order to give an improved performance. In addition, they conceived up a revolutionary Emotional Education Mechanism (EEM) with the purpose of facilitating the effective deployment and optimization of lightweight financial education models. [36], saw the unveiling of a CNN model with twelve tasks, which provides a novel method to multitask learning by merging three independent paradigms of parameter sharing across tasks. This model was recently introduced. MobileNet, ResNet, and other backbone designs are utilized in this approach. To perform a range of tasks, including gender, age, facial expression recognition, and ethnicity classification.In Figure 6, different Mutli-tasking Networks, (a)EmotionGCN and (b) Multi- task EfficientNet-B0 are displayed.

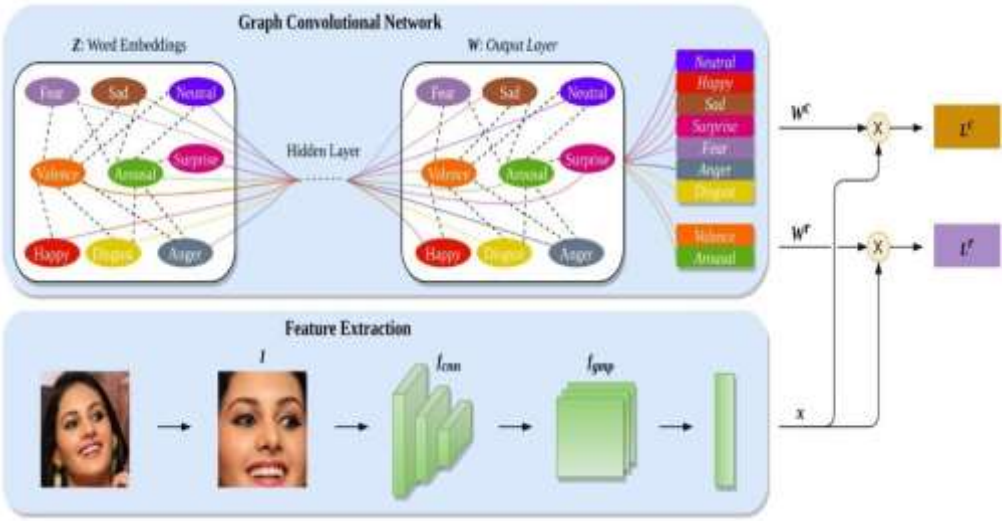

Figure 6. Multi-tasking Networks

The GCN of Emotion [37] A graph that links seven expression labels with two valence-arousal dimensions is incorporated into the Emotion-GCN model, which was developed for the purpose of Facial Expression Recognition (FER) in natural situations. This model is designed to be used in natural environments. The processing of word embeddings into classifiers and regressors for the purpose of mapping facial expressions is accomplished through the utilization of Graph Convolutional Networks (GCNs). A CNN that is based on DenseNet is used to extract image representations, and then global max-pooling is used to enhance them. This allows the model to perform expression classification as well as valence-arousal regression through end-to-end training [38].

In [39], The Affective Behavior Analysis in the Wild (ABAW) Competition includes a challenge that requires participants to learn many tasks simultaneously. A unified algorithm that is capable of completing many affective analysis tasks simultaneously is the task that the participants are tasked with building. These tasks include valence–arousal estimation, expression classification, and action unit recognition. In order to facilitate the creation of highly effective algorithms for emotion analysis in ecologically realistic situations, this effort aims to promote developments in affective computing, which will ultimately lead to the development of algorithms.

Facial expression recognition is crucial for understanding human emotions and nonverbal communication. With the growing prevalence of facial recognition technology and its various applications, accurate and efficient facial expression recognition has become a significant research area. However, most previous methods have focused on designing unique deep-learning architectures while overlooking the loss function. This study presents a new loss function that allows simultaneous consideration of inter- and intra-class variations to be applied to CNN architecture for facial expression recognition. More concretely, this loss function reduces the intra-class variations by minimizing the distances between the deep features and their corresponding class centers. It also increases the inter-class variations by maximizing the distances between deep features and their non-corresponding class centers, and the distances between different class centers. Numerical results from several benchmark facial expression databases, such as Cohn-Kanade Plus, Oulu-Casia, MMI, and FER2013, are provided to prove the capability of the proposed loss function compared with existing ones.

Cascaded Networks

A challenge that demands participants to learn multiple tasks at the same time is included in the Affective Behavior Analysis in the Wild (ABAW) Competition contest. One of the tasks that the participants are tasked with completing is the construction of a unified algorithm that is capable of executing a number of emotional analysis tasks simultaneously. Calculating valence–arousal, classifying expressions, and recognizing action units are some of the tasks that fall under this category. This endeavor intends to foster breakthroughs in affective computing, which will ultimately lead to the development of algorithms, with the goal of facilitating the production of highly effective algorithms for emotion analysis in ecologically realistic circumstances using these algorithms.

Databases Used for Facial Emotion

Recognition Facial recognition is getting better and more prevalent each year, and consequently, facial databases have expanded tremendously [79]. When modeling of recognition requires visual or audio examples, you have to give, model enhancement or training of the model requires a database of those kinds, and this, as well as class labels for them, which gets progressively larger and larger as the number of examples increases, expand is required [80]. For example, there are various possible applications for emotional recognition, ranging from simple human-robot collaboration [81] to being used to identify people suffering from depression to serving as a depression detector [82]. out into the real world to realize how light conditions and occlusions work in the context of real life in order to comprehend the situations [84]. Using this method [85], the subjects were easily rotated and the effects of different lighting on their appearance were studied in the M2VTS database which features the faces of 37 subjects in a variety of rotated and lit positions.)e emotions included in a database define its function. Many databases, like CK, MMI, eNTERFACE, and NVIE, opt to record the six basic emotion categories proposed by Ekman. Many databases, like the SMO, AAI, and ISL meeting corpus, try to classify or contain general positive and negative emotions. Some try to evaluate deception and honesty, such as the CSC corpus database.)e most well-known 3D datasets are BU-3DFE, BU-4DFE, Bosphorus, and BP4D. BU-3DFE and BU-4DFE both have six-expression posed datasets, with the latter having a higher resolution. Bosphorus tries to address the issue of having a wider range of facial expressions, while BP4D is the only one of the four that uses

induced rather than posed emotions.)e main benefit of deep learning is that it opens the neural networks to various databases, allowing them to grow with the addition of a wide range of new inputs, examples, facial expressions, and constant changes in expressions. CK database also has seven expressions but it contains 132 subjects that are posed with natural and smile. It contains totally 486 image sequences with 640 490 pixel resolution of gray images. Some of the sample images of the CK database are shown in Figure 7

Figure.7 Sample images from CK database.

Controlled datasets (or lab): The creation of controlled datasets or labs takes place in settings such as studios or research facilities. These datasets comprise the capture of facial expressions under conditions that are carefully controlled, including lighting, camera angles, and backgrounds. The JAFFE and CK+ datasets are two examples of controlled datasets. There are numerous other examples.

Uncontrolled datasets (or In-the-Wild): In-the-wild datasets are made up of photographs or videos that were gathered from actual situations that occur in the real world. Variations in lighting, occlusions, and posture variations, as well as issues due to elements such as changes in lighting, occlusions, various camera quality, and pose variations, are some of the extra challenges that these datasets provide. Many datasets, such as AffectNet [41] and Emotic [42], are examples of datasets that are collected in the wild.In order to assess the effectiveness of facial expression recognition (FER) systems, researchers make use of a wide range of datasets. These datasets include a wide range of contexts, from controlled laboratory environments to demanding situations that occur in the real world. The utilization of a wide variety of datasets guarantees the development of models that are reliable and accurate, and that function effectively in a variety of settings and fields of application. In the beginning, datasets are classified according to the type of data, which may be images or videos. Following that, they are separated into two categories: regulated (in the laboratory) and uncontrolled (in the wild). The last sub-category is determined by the year in which the datasets were initially created. Lists of datasets that are appropriate for the image category:

Contribution in facial expression fields

Shintaro Kondo et al. (2022) the use of video inputs and an increase in frame rates were investigated as potential means of improving the quality of face expression models applied by dialogue bots. In order to generate facial emotions in movies of emotional speech, this method makes use of advancements made in past research.[46]

Wenbo Zheng et al. (2022) in order to synthesize facial expressions, the Local and Global Perception Generative Adversarial Network (LGP-GAN) was established. This network focuses on key facial regions such as the lips and the eyes. This approach makes use of a two-stage process in order to enhance the quality and level of detail of the expressions that are created.[47]

Takanori Shiomi et al. (2022) the classification of various facial expressions and the evaluation of their intensity are the recommended approaches for estimating the intensity of facial expressions. They determined that both explicit and implicit methods were viable for recognizing facial expressions after conducting a comparison between the two.[48]

Dejian Li et al. (2021) a comprehensive adversarial network that generates new facial expressions from landmarks and expression labels was presented. This network was designed to preserve identification information while simultaneously producing expressions that are realistic.[49]

Lingzhao Ju et al. (2022) MAP Net, a mask-based focus parallel network, was created in order to address problems with posture and occlusion in facial emotion identification. For the purpose of enhancing the precision of expression prediction, this technique makes use of a binary mask in conjunction with parallel network layers.[50]

Win Shwe Sin Khine et al. (2022) applied transfer learning on deep learning models for facial expression recognition, achieving high accuracy with Efficient Net B0 on the CK+ dataset[51]

Rajesh Kumar et al. (2023) A real-time facial expression recognition system was developed by utilizing CNN and Haar cascade classifiers. The system was trained on the FER2013 dataset and is intended for use in a variety of fields.[52]

Jianing Teng et al. (2021) A combination of two-dimensional and three-dimensional convolutional neural networks (CNNs) was used to develop the Typical Facial Expression Network (TFEN), which was designed to enhance facial expression recognition (FER) by disentangling facial information from emotional features.[53]

Yun Zhang et al. (2021) An inversion-based GAN-based Attentive Expression Embedding Network (GI-AEE) was proposed as a means of editing facial expressions with high quality, while also resolving concerns of artifacts and blurring.[54]

Jie Cai et al. (2021) the Identity-Free conditional GAN (IF-GAN) was proposed in order to obtain state-of-the-art results by generating images that have a consistent synthetic identity. This was done in order to eliminate identity-related variability in FER.[55]

Young Eun An et al. (2021) utilized landmark-based deep learning techniques in order to predict mental health statuses based on facial signals in order to investigate the impact that mental health disorders have on facial expressions.[56]

Jingjie Yan et al. (2023) This study demonstrated the usefulness of the Facial Expression of Neonatal Pain (FENP) database in the field of research on neonatal pain and facial expression identification. The database was created by utilizing cutting-edge recognition techniques to assess pain emotions in neonates.[57]

Yifan Xia et al. (2022) Particular emphasis was placed on the extraction and reproduction of information in crucial facial regions, such as the eyes and mouth, in order to produce high-quality facial expressions. This was accomplished by the use of LGP-GAN to synthesize facial expressions. [58]

IV CONCLUSION

The important future enhancements described from recent papers are FER for side view faces using the subjective information of facial sub-regions and use different parameters to represent the pose of the face for real-time applications. FER is used in real-time applications such as driver sate surveillance, medical, robotics interaction, forensic section, detecting deceptions. This survey paper is useful for software developers to develop algorithms based on their accuracy and complexity. Also, it is helpful for hardware implementation to implement with low cost depends on their need. This survey compares algorithms based on preprocessing, feature extraction, classification and major contributions. The performance analysis is done based on the database, complexity rate, recognition accuracy and major contributions. This survey discusses the properties such as availability of preprocessing and feature extraction and expression count. The power of algorithms, advantages is discussed elaborately to reach the aim of this survey. ROI segmentation method is used for preprocessing and it. While FER is an important source of information about an individual's emotional state, it is always limited by learning only the six basic emotions plus neutral. It is in conflict with what is present in everyday life, which contains more complex emotions.)is will encourage researchers to expand their databases and develop powerful deep learning architectures capable of recognizing all basic and secondary emotions in the future. Additionally, emotion recognition has evolved from a unimodal analysis to a complex system multimodal analysis in the modern era. Leon et al. in [123] demonstrate that multimodality is a necessary condition for optimal emotion detection. Researchers are now focusing their efforts on developing and commercializing powerful multimodal deep learning architectures and databases.

References

1. S. LokeshNaik, A. Punitha, P. Vijayakarthik, A. Kiran, A. N. Dhangar, B. J. Reddy, et al., "Real Time Facial Emotion Recognition using Deep Learning and CNN," in 2023 International Conference on Computer Communication and Informatics (ICCCI), 2023, pp. 1-5.
2. D. Duncan, G. Shine, and C. English, "Facial emotion recognition in real time," Computer Science, pp. 1-7, 2016.
3. A. Patwal, M. Diwakar, A. Joshi, and P. Singh, "Facial expression recognition using DenseNet," in 2022 OITS International Conference on Information Technology (OCIT), 2022, pp. 548-552.
4. H. Dino, M. B. Abdulrazzaq, S. Zeebaree, A. B. Sallow, R. R. Zebari, H. M. Shukur, et al., "Facial expression recognition based on hybrid feature extraction techniques with different classifiers," TEST Engineering & Management, vol. 83, pp. 22319-22329, 2020.
5. S. M. Saleem, S. R. Zeebaree, and M. B. Abdulrazzaq, "Real-life dynamic facial expression recognition: a review," in Journal of Physics: Conference Series, 2021, p. 012010.
6. S. Meriem, A. Moussaoui, and A. Hadid, "Automated facial expression recognition using deep learning techniques: an overview," International Journal of Informatics and Applied Mathematics, vol. 3, pp. 39-53, 2020.
7. I. Talegaonkar, K. Joshi, S. Valunj, R. Kohok, and Kulkarni, "Real time facial expression recognition using deep learning," in Proceedings of international conference on communication and information processing (ICCIP), 2019.
8. D. Dukić and A. Sovic Krzic, "Real-time facial expression recognition using deep learning with application in the active classroom environment," Electronics, vol. 11, p. 1240, 2022.
9. M. Sajjad, F. U. M. Ullah, M. Ullah, G. Christodoulou, F. A. Cheikh, M. Hijji, et al., "A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines," Alexandria Engineering Journal, vol. 68, pp. 817-840, 2023.
10. Zhang, Z.; Song, Y.; Qi, H. Age progression/regression by conditional adversarial autoencoder. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5810–5818.
11. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
12. Kollias, D. ABAW: Valence-Arousal Estimation, Expression Recognition, Action Unit Detection & Multi-Task Learning Challenges. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, New Orleans, LA, USA, 18–24 June 2022; pp. 2328–2336.
13. Huang, Y.; Khan, S.M. Dyadgan: Generating facial expressions in dyadic interactions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 June 2017; pp. 11–18.
14. Yang, H.; Ciftci, U.; Yin, L. Facial expression recognition by de-expression residue learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2168–2177.
15. Wu, R.; Zhang, G.; Lu, S.; Chen, T. Cascade EF-GAN: Progressive Facial Expression Editing with Local Focuses. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
16. Liu, Y.; Zhang, X.; Li, Y.; Zhou, J.; Li, X.; Zhao, G. Graph-based facial affect analysis: A review. IEEE Trans. Affect. Comput. **2022**,14, 2657–2677.
17. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Philip, S.Y. A comprehensive survey on graph neural networks. IEEE Trans. Neural Netw. Learn. Syst. **2020**, 32, 4–24.
18. Liao, L.; Zhu, Y.; Zheng, B.; Jiang, X.; Lin, J. FERGCN: Facial expression recognition based on graph convolution network. Mach. Vis. Appl. **2022**, 33, 40.

19. Wu, C.; Chai, L.; Yang, J.; Sheng, Y. Facial expression recognition using convolutional neural network on graphs. In Proceedings of the 2019 Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019; pp. 7572–7576.
20. Wasi, A.T.; Šerbetar, K.; Islam, R.; Rafi, T.H.; Chae, D.K. ARBEx: Attentive Feature Extraction with Reliability Balancing for Robust Facial Expression Learning. arXiv **2023**, arXiv:2305.01486.
21. Perveen, N.; Gupta, S.; Verma, K. Facial expression recognition system using statistical feature and neural network. Int. J. Comput. Appl. **2012**, 48, 17–23.
22. Meng, D.; Peng, X.; Wang, K.; Qiao, Y. Frame Attention Networks for Facial Expression Recognition in Videos. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 3866–3870.
23. Hasani, B.; Mahoor, M.H. Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks. In Proceed- ings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Honolulu, HI, USA, 21–26 July 2017.
24. Xie, S.; Girshick, R.; Dollar, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
25. Kervadec, C.; Vielzeuf, V.; Pateux, S.; Lechervy, A.; Jurie, F. Cake: Compact and Kuo, C.M.; Lai, S.H.; Sarkis, M. A Compact Deep Learning Model for Robust Facial Expression Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–23 June 2018.
26. Kollias, D.; Cheng, S.; Ververas, E.; Kotsia, I.; Zafeiriou, S. Deep Neural Network Augmentation: Generating Faces for Affect Analysis. Int. J. Comput. Vis. **2020**, 128, 1455–1484.
27. Vo, T.H.; Lee, G.S.; Yang, H.J.; Kim, S.H. Pyramid with super resolution for in-the-wild facial expression recognition. IEEE Access **2020**, 8, 131988–132001.
28. Psaroudakis, A.; Kollias, D. MixAugment & Mixup: Augmentation Methods for Facial Expression Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2367–2375.
29. Zhang, Y.; Wang, C.; Deng, W. Relative Uncertainty Learning for Facial Expression Recognition. Adv. Neural Inf. Process. Syst.**2**021, 34, 17616–17627.
30. Zhou, H.; Meng, D.; Zhang, Y.; Peng, X.; Du, J.; Wang, K.; Qiao, Y. Exploring emotion features and fusion strategies for audio-video emotion recognition. In Proceedings of the 2019 International Conference on Multimodal Interaction, Suzhou, China, 14–18 October 2019; pp. 562–566.
31. Kumar, V.; Rao, S.; Yu, L. Noisy student training using body language dataset improves facial expression recognition. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 756–773.
32. Adrian, R. Deep Learning for Computer Vision with Python Volume 1; Pyimage-Search. 2017.
33. Cheng, S.; Kotsia, I.; Pantic, M.; Zafeiriou, S. 4dfab: A large scale 4d database for facial expression analysis and biometric applications. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5117–5126.
34. Sun, N.; Tao, J.; Liu, J.; Sun, H.; Han, G. 3-D Facial Feature Reconstruction and Learning Network for Facial Expression Recognition in the Wild. IEEE Trans. Cogn. Dev. Syst. **2023**, 15, 298–309.
35. Wu, Z.; Wang, X.; Jiang, Y.G.; Ye, H.; Xue, X. Modeling Spatial-Temporal Clues in a Hybrid Deep Learning Framework for Video Classification. In Proceedings of the 23rd ACM International Conference on Multimedia, New York, NY, USA, 26–30 October 2015; pp. 461–470.
36. Dang, C.N.; Moreno-García, M.N.; De la Prieta, F. Hybrid deep learning models for sentiment analysis. Complexity **2021**, 2021, 9986920.

37. Baltrušaitis, T.; Ahuja, C. Morency, L.P. Multimodal Machine Learning: A Survey and Taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. **2019**, 41, 423–443.
38. M. Daneshmand, A. Abels, and G. Anbarjafari, "Real-time, automatic digi-tailor mannequin robot adjustment based on human body classification through supervised learning," International Journal of Advanced Robotic Systems, vol. 14, no. 3, Article ID 1729881417707169, 2017.
39. A. Bolotnikova, H. Demirel, and G. Anbarjafari, "Real-time ensemble based face recognition system for nao humanoids using local binary pattern," Analog Integrated Circuits and Signal Processing, vol. 92, no. 3, pp. 467–475, 2017.
40. M. Valstar, B. Schuller, K. Smith et al., "Avec 2013: the continuous audio/visual emotion and depression recognition challenge," in Proceedings of the 3rd ACM International Workshop on Audio/visual Emotion challenge, pp. 3–10, Barcelona Spain, October 2013.
41. Trong-Dong PhamMinh-Thien DuongQuoc-Thien Ho,Seongsoo Lee,Min-Cheol Hong **(2023)** CNN-Based Facial Expression Recognition with Simultaneous Consideration of Inter-Class and Intra-Class Variations (2) 24 10.3390/s2324965
42. Wu, B.F.; Lin, C.H. Adaptive feature mapping for customizing deep learning based facial expression recognition model. *IEEE Access* **2018**, *6*, 12451–12461.
43. hand, M.; Roy, S.; Siddique, N.; Kamal, M.A.S.; Shimamura, T. Facial emotion recognition using transfer learning in the deep CNN. *Electronics* 2021, *10*, 1036.
44. C. Watson and P. Flanagan, "Nist special database 18 mugshot identification database,," 2016
45. Khan, M.M.; Ward, R.D.Ingleby, M. Automated classification and recognition of facial expressions using infrared thermal imaging. In Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems, Singapore, 1–3 December; 2004; Volume 1, pp. 202–206.
46. Shintaro Kondo; Seiichi Harata; Takuto Sakuma; Shohei Kato acial Expressions Generating Model Reflecting Agent's Emotion Response Using Facial Landmark Residual Networks2022 IEEE 11th Global Conference on Consumer Electronics (GCCE) 2022
47. Yifan Xia; Wenbo Zheng; Yiming Wang; Hui Yu; Junyu Dong; Fei-Yue Local and Global Perception Generative Adversarial Network for Facial Expression Synthesis Wang IEEE Transactions on Circuits and Systems for Video Technology ,2022
48. Takanori Shiomi; Hiroki Nomiya; Teruhisa Hochin Facial Expression Intensity Estimation Considering Change Characteristic of Facial Feature Values for Each Facial Expression 2022 23rd ACIS International Summer Virtual Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Summer) ,2022
49. Dejian Li; Wenqian Qi; Shouqian Sun Facial Landmarks and Expression Label Guided Photorealistic Facial Expression Synthesis IEEE Access ,2021
50. Lingzhao Ju; Xu Zhao Mask-Based Attention Parallel Network for in-the-Wild Facial Expression Recognition ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) ,2022
51. Win Shwe Sin Khine; Prarinya Siritanawan; Kazunori Kotani Facial Expression Features Analysis With Transfer Learning 2022 14th International Conference on Knowledge and Systems Engineering (KSE), 2022
52. Rajesh Kumar A Deep Learning Approach To Recognizing Emotions Through Facial Expressions 2023 Global Conference on Wireless and Optical Technologies (GCWOT), 2023
53. Jianing Teng,Dong Zhang; Wei Zou,Ming Li,Dah-Jye Lee Typical Facial Expression Network Using a Facial Feature Decoupler and Spatial-Temporal Learning IEEE Transactions on Affective Computing,2021
54. Yun Zhang;Ruixin Liu, Yifan Pan, Dehao Wu,Yuesheng Zh,Zhiqiang Bai GI-AEE: GAN Inversion Based Attentive Expression Embedding Network For Facial Expression Editing 2021 IEEE International Conference on Image Processing (ICIP) Year: 2021
55. Jie Cai; Zibo Meng; Ahmed Shehab Khan; James O'Reilly; Zhiyuan Li; Shizhong Han; Yan Tong Identity-Free Facial Expression Recognition Using Conditional Generative Adversarial Network 2021 IEEE International Conference on Image Processing (ICIP) Year: 2021
56. Young Eun An,Ji Min Lee,Min Gu Kim;Sung Bum Pan Categorization of facial expressions using rearranged landmarks 2021 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) 2021

57. Jingjie Yan; Guanming Lu; Xiaonan Li; Wenming Zheng; Chengwei Huang; Zhen Cui;Yuan Zong;Mengying Chen;Qiang Hao;Yi Liu;Jindu Zhu;Haibo FENP: A Database of Neonatal Facial Expression for Pain Analysis LiIEEE Transactions on Affective Computing,2023

58. Yifan Xia;Wenbo Zheng;Yiming Wang;Hui Yu;Junyu Dong;Fei-Yue Wang Local and Global Perception Generative Adversarial Network for Facial Expression Synthesis IEEE Transactions on Circuits and Systems for Video Technology,2022