

Review Article

Facial Expression Recognition from a Single Face Image Based on Deep Learning and Broad Learning

Mei Bie ^{1,2} Huan Xu ¹ Yan Gao ¹ and Xiangjiu Che ¹

¹College of Computer Science and Technology, Jilin University, Changchun 130012, China

²Institute of Education, Changchun Normal University, Changchun 130032, China

Correspondence should be addressed to Xiangjiu Che; chexj@jlu.edu.cn

Received 6 June 2022; Revised 23 June 2022; Accepted 28 June 2022; Published 20 July 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Mei Bie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With advances in computer vision and artificial intelligence technology, facial expression recognition research has become a prominent topic. Current research is grappling with how to enable computers to fully understand expression features and improve recognition rates. Most single face image datasets are based on the psychological classification of the six basic human expressions used for network training. By outlining the problem of facial recognition by comparing traditional methods, deep learning, and broad learning techniques, this review highlights the remaining challenges and future directions of deep learning and broad learning research. The deep learning method has made it easier and more effective to extract expression features and improve facial expression recognition accuracy by end-to-end feature learning, but there are still many difficulties in robustness and real-time performance. The broad learning system (BLS) is a broad network structure that is expanded by increasing the number of feature nodes and enhancement nodes appropriately to reinforce the structure and is also effective in facial expression recognition. However, outliers and noises in unbalanced datasets need BLS to solve in the future. Finally, we present several problems that still need to be addressed in facial expression recognition.

1. Introduction

Facial expression is an important way for humans to present their emotions. They can affect our daily lives by changing our attention, perception, and memory, helping us understand the intentions of others. As an information carrier, facial expressions can accurately express the true emotions of human beings. Humans can also learn the inner thoughts of others through facial expressions [1, 2] [3]. According to a study by psychologists Mehrabian et al., facial expressions are prominent in daily human communication, with a transmission rate of 55%, far more significant than speech (38%), language (7%), and so on [4]. Face expressions, on the other hand, follow the rules of facial muscle movement and are unaffected by gender, age, race, or cultural background [5]. As a result, facial expressions are a powerful tool for detecting emotions. Facial expression recognition research is crucial for advancing artificial intelligence and other domains. The research on facial expression recognition based on computer technology can enable intelligent devices such as

robots to better understand and identify human emotions, so as to actively judge human emotions, better serve humans, and achieve barrier-free interaction between humans and machines [1, 6, 7].

The performance of deep convolutional learning technology in natural language processing, computer vision, and other information technology sectors has grown increasingly significant due to ongoing in-depth research on deep convolutional networks. People's perceptions of deep learning convolutional neural networks have shifted as a result [8]. The application and performance of deep learning technology in image classification and recognition are better than the application and performance of previous algorithms in terms of recognition accuracy, and recognition speed as more universities and scientific research institutions conduct more in-depth research. As a result, more researchers are becoming aware of deep learning technologies. The deep learning technique eliminates numerous artificial technological ways to extract features and instead performs end-to-end feature learning directly on the picture.

As a result, we use deep learning-based technology in facial expression recognition to help it extract more effective and more prosperous expression characteristics, considerably enhancing facial expression identification accuracy. Face expression recognition is prone to complex issues such as difficulties extracting facial characteristics, low recognition accuracy, and slow recognition speed [9].

Although the facial expression recognition system has a wide range of applications and a bright future, several technical issues with the actual product landing still have to be resolved. In the real world, numerous adverse external effects alter the picture of the face, failing face frame extraction or poor recognition degree, and facial expression recognition in the video image must have a quick recognition rate to be successful system assistance. More facial expression recognition applications will gain traction due to their high accuracy and speed.

Deep learning research in facial expression has increasingly risen as deep learning technology in the field of image categorization has advanced. As a result, we focus on deep learning approaches in the context of expression recognition research. First, we will go through the history of face recognition technology and introduce some of the most popular expression datasets. The impacts of traditional and deep learning technology on different data sets are compared, and facial expression recognition technology based on deep learning is thoroughly examined from the perspectives of facial recognition challenges and expression recognition algorithms. Furthermore, the challenges and future directions in deep learning-based expression recognition are discussed.

2. Facial Expression Recognition Technology

2.1. Overview of Facial Expression Recognition Technology. As indicated in Figure 1, surprise (Su), fear (Fe), disgust (Di), happy (Ha), sadness (Sa), and anger (An) are the basic human emotions. In the past, automatic face expression interpretation from picture sequences was primarily a research area for psychologists. However, Suwa [10] did a preliminary investigation in 1978. Even though many academics feel the six basic emotions are culturally unique and not universal, the categorization approach based on the six fundamental expressions has been widely accepted by researchers and has aided the growth of expression recognition [11].

Facial expression recognition (FER) is a method that consists of three steps: preprocessing of facial images, extraction and representation of facial expression features, and recognition of facial expressions [13]. Figure 2 depicts the fundamental procedure. Facial image preprocessing is a stage of video processing that automatically recognizes faces in an input picture or sequence, detects faces for each frame, just detects faces in the first frame, and then tracks faces for the rest of the video sequence. Face analysis is complicated by variations in facial appearance caused by changes in position and illumination. As a result, preprocessing is necessary before analysis to align and normalize the visual-semantic information communicated by faces. Following the prepro-

cessing of face images, the next step is to extract usable data or information from the images, and the feature extraction procedure is critical for FER. This technique allows for a smaller and richer collection of characteristics, such as face edges, corners, diagonals, and other critical information (such as the distance between lips and eyes and the distance between two eyes), allowing the model to learn from the training data more rapidly. The FER pictures are categorized after face detection, preprocessing, and feature extraction phases.

2.2. Facial Expression Database. Most expression recognition methods must be tested on a set of expression data, and the generation of expression data is thus a major driving factor in the advancement of expression recognition technology. Traditional expression recognition algorithms are sparse learning nonnegative matrix factorization (NMF) [15], local binary patterns (LBP) [16], LBP on two orthogonal planes (LBP-TOP) [17], and so on. Since 2013, however, general emotion recognition competitions such as FER2013 [18] and EmotiW [19–21] have gathered large-scale facial expression image data, primarily from the Internet or movie clips, that align with the real scene and effectively communicate promote the facial expression recognition transitions from laboratory controls to real-world scenarios. In image acquisition, facial expression images are usually acquired under laboratory conditions and subjected to specific preprocessing. At present, the well-known datasets include Japanese Female Facial Expression (JAFFE) and Fer2013. The main information of each dataset is shown in the following Table 1.

The CK+ (The Extended Cohn Kanade) expression database is a 593 video sequences laboratory-controlled database. The CK+ expression database video sequences were gathered from 123 participants, more than 60% of whom were women, all of whom were between 18 and 30 years old [22]. The following are the features of the video sequence. The lowest frame duration is ten frames, while the maximum frame duration is 60 frames. The face begins in a neutral state, with no expression, and gradually progresses to a peak of emotion.

The JAFFE (Japanese Female Facial Expression) database contains 219 images of 10 volunteers' facial expressions [23]. Six fundamental emotions were gathered from all volunteers, three to four photographs for each category and a neutral face image. The photos of facial expressions are stored in the database as black and white photographs that have been preprocessed.

There are 2,880 video sequences in the Oulu-CASIAPI1 expression database from 80 volunteers [24]. The video series begins with a neutral frame and progressively evolves to the peak emotion, similar to the CK+ facial expression database. FER experimental verification typically only employs 480 video sequences acquired by the Visual Identity System (VIS) in standard indoor lighting, with each sequence beginning with a frame (neutral face) and ending with three peaks (expressive face) as experimental data.

The expression database RAF-DB (Real-world Affective Faces Database) is based on real-world scenarios, and the

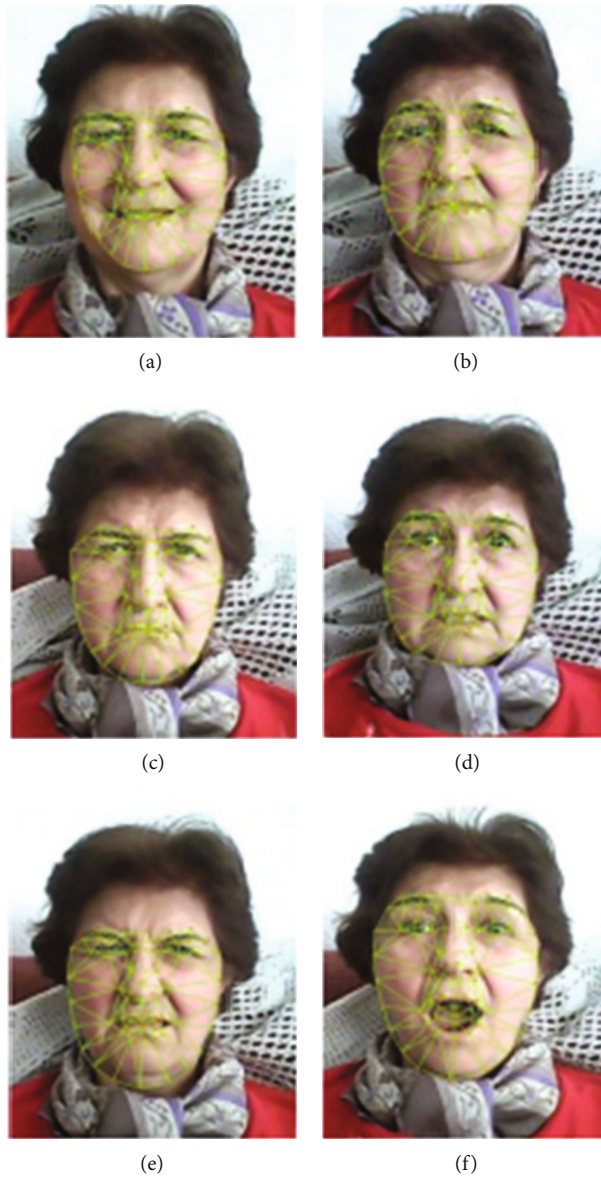


FIGURE 1: Humans have six basic expressions plus neutral: (a) happy, (b) sadness, (c) anger, (d) fear, (e) disgust, and (f) surprise [12].

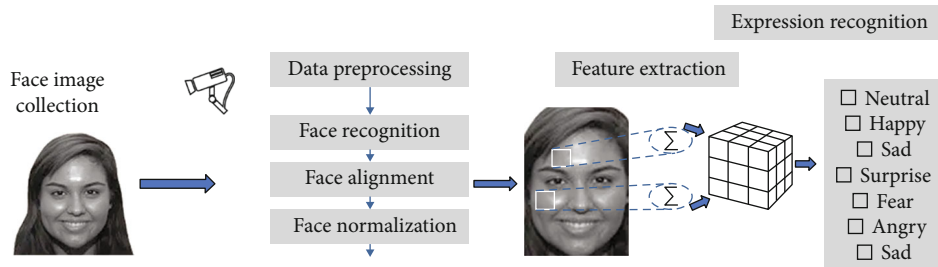


FIGURE 2: Flowchart of general facial expression recognition [14].

expression images in it are all retrieved from an Internet search engine in the form of emotion-related and age, race, and gender-related keyword combinations, totaling 29,672 pieces [25]. Each emoticon image was formed from roughly 40 persons, and the labels were annotated online by 315

annotators (students and university personnel) who underwent an hour-long online instruction on the psychological understanding of emotions.

The FERP plusM facial expression dataset grew out of the FER2013 facial expression dataset [26]. All pictures were

TABLE 1: Information table of facial expression database.

Name	Picture quantity	Picture size	Expression type
JAFFE	219	256*256	7
Fer2013	35886	48*48	7
Yale emoji database	165	100*100	6
CK+ (video)	593	—	7
Oulu-CASIAPI (video)	2,880	—	7
RAF-DB	29,672	—	7
FERP plusM	35,887	48*48	7
Affected	450,000	—	7
AFEW (video)	957	—	7

scaled to 48x48 pixels after being preprocessed. The dataset is commonly separated into 28,709 photos for training, 3,589 images for validation, and 3,589 images for testing when used for experimental validation.

Affected is the world's most extensive library of real-world face expressions, with over 1 million photographs gathered from the web by searching emotion-related categories using various search engines. It offers facial expressions in two emotion models (category and dimensional), with 450,000 manually annotated with seven fundamental expressions (contempt, sorrow, fear, surprise, pleasure, disgust, and rage) and neutral expressions [27].

The Acted Facial Expressions in the Wild (AFEW) database is used to create the SFEW (Static Facial Expressions in the Wild) dataset. The expression video clips from movies that are designated as six fundamental expressions (anger, contempt, fear, sorrow, pleasure, and surprise), as well as neutral expressions and facial movements, comprise the material of AFEW. Units (Action Units, AUs) are labeled to mimic real-world face emotions as closely as possible [19].

2.3. Difficulties in Facial Expression Recognition Research. There are still numerous obstacles and problems to be solved when the scientific focus of expression recognition work switches to challenging real-world unconstrained environmental settings.

- (1) *Lighting Problem.* It is possible to lose vital information in specific face areas if there is insufficient or no light and the light is too bright. There is a loss of crucial knowledge. As a result, lighting conditions are an issue for expression recognition, and they can significantly impact the algorithm's robustness
- (2) *Lack of High-Quality Public Datasets.* Data bias and inconsistent labeling are common concerns when utilizing diverse datasets due to the diversity and subjectivity of the distribution of expression categories
- (3) *Occlusion and Pose Changes.* The FER model cannot get adequate facial information because the face is obstructed or the posture rotation is too great, and when the obtained facial information is less than 60%, the expression recognition cannot be conducted typically

- (4) *Multimodal Recognition.* Although facial expression recognition may produce good results from visible face photographs on its own, merging it with additional models into an integrated system, such as adding a voice model, adds more information and helps to enhance model dependability
- (5) *Big Data Problem.* Data is measured in terabytes in many scientific and commercial applications, posing issues for data storage, transport, and processing in FER systems
- (6) *Privacy Protection.* Many expression recognition algorithms rely on high-resolution face photos of users, yet many researchers pay little or no attention to ensuring that users' visual privacy is protected

3. Expression Recognition Algorithm

The science of expression recognition has advanced significantly over the years. Existing approaches may be split into standard machine learning algorithms and deep learning algorithms. Figure 3 depicts the framework for the two sorts of approaches. Face preprocessing, feature extraction, and classifier prediction are three discrete manual stages in traditional expression recognition methods, as depicted in Figure 3(a). As demonstrated in Figure 3(b), deep learning networks combine feature learning and classification into a single framework, eliminating the need for complicated manual methods.

3.1. Traditional Expression Recognition Algorithms. Preprocessing is the initial stage in classic expression recognition systems since low-quality photographs influence the system's accuracy. Preprocessing can help you eliminate as much task-irrelevant data as feasible while boosting task-relevant data. Face detection, histogram equalization, picture normalization, and face alignment are typical image preprocessing techniques. To extract the critical information that best recognizes expressions based on the digital signal of the image, well-designed feature extractors are required. Traditional approaches may be classified into the following kinds based on the extracted characteristics.

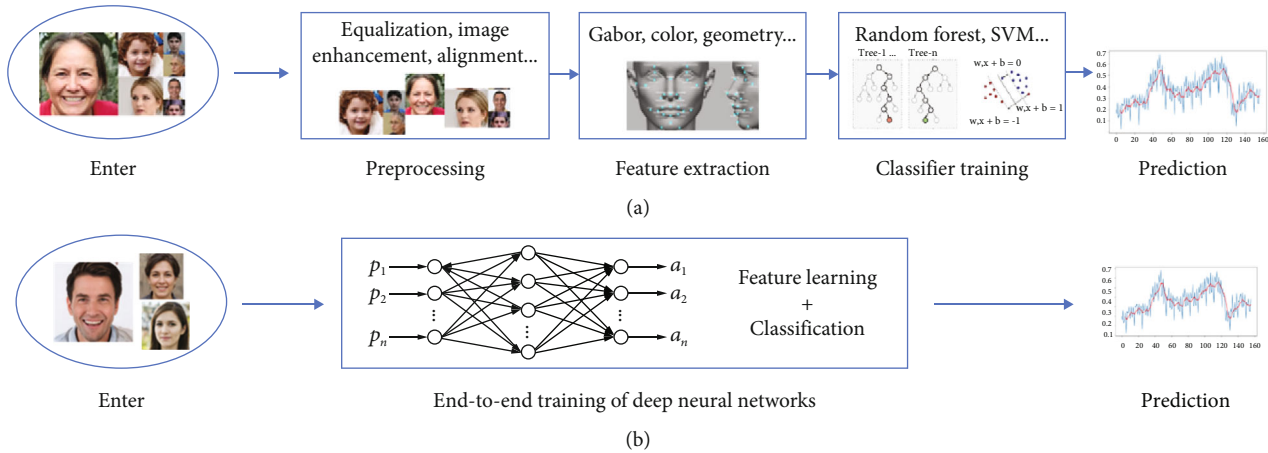


FIGURE 3: (a) Traditional facial expression recognition framework [28] and (b) deep learning-based empathy recognition framework [29].

3.1.1. Appearance-Based Methods. The texture, color, edge, and other aspects of face areas connected to expressions are apparent features. Histogram and grayscale characteristics are the simplest and most influential among them. The histogram is a statistical descriptor that specifies the number of gray levels in a picture. The grayscale feature, the image's grayscale, is more intuitive. Gabor and LBP features are also commonly employed in emotion recognition and picture texture extraction studies. Gabor features are retrieved using the Gabor filter, which emulates the visual stimulus-response of simple cells in the human visual system. Features at various scales and orientations can be retrieved by altering the Gabor filter's frequency domain settings.

3.1.2. Methods Based on Geometric Features. Facial muscle movements produce expressions, while geometric characteristics extract information about changes such as location, distance, and shape. Geometric feature-based approaches usually start by locating face important points or areas, then extracting features.

3.1.3. Methods Based on Subspace Learning. Face images have high dimensionality, include much information, have a lot of redundancy, and much irrelevant noise, all of which make it challenging to recognize expressions. A more abstract technique is feature extraction based on subspace learning. This technique uses an improved model to learn the mapping function, turns the original data into a more succinct and effective representation, and makes correct expression recognition easier. Principal component analysis, linear discriminant analysis, and manifold learning and refinements are examples of such approaches.

Traditional facial expression recognition systems mainly require an artificial design to extract facial expression elements in real-life applications. Artificially constructed features frequently fail to adapt to complicated and changing expressions, resulting in an inadequate final expression categorization outcome.

3.2. Expression Recognition Algorithm Based on Deep Learning. Hand-crafted features are time-consuming and inaccurate and are constrained mainly by the specified algorithms, which can only represent information under certain formal constraints. It has progressed from small sample jobs in typical laboratory situations to large-scale sample tasks in real-life scenarios in facial expression recognition. Deep learning has been applied in numerous computer vision tasks in recent years. Autoencoder and generative adversarial networks (GANs) are also used in several expression recognition methods [5, 30].

3.2.1. Expression Recognition Based on Convolutional Neural Network. CNN-based expression recognition research content is vast, covering architectural design and loss function design [31]. Yu and Zhang [32] built an eight-layer CNN architecture with five convolutional layers and three fully linked layers. Yu et al. preprocess pictures using a well-designed multilevel face detection framework (containing three current state-of-the-art detectors) before using a CNN architecture for end-to-end feature learning and classification, as illustrated in Figure 4.

3.2.2. Expression Recognition Based on Recurrent Neural Network. Recurrent neural network (RNN) is a recurrent neural network with a memory function and can learn time series data's dynamic evolution information. The RNN's nodes are linked together in a chain, and the input sequence data is processed recursively in the time evolution direction. RNNs are typically employed for temporal logic applications, such as natural language processing and video comprehension. Many studies use RNN or its variations, such as long short-term memory (LSTM) and bidirectional LSTM, for facial emotion recognition. The feature vector of the picture is commonly obtained as input data using a manual extractor or another deep learning model in these approaches.

As demonstrated in Figure 5, Yu et al. [33] introduced an end-to-end trainable deep learning model called STC-NLSTM for identifying medium expressions in picture sequences. SP3DCNN and NLSTM are two critical components of STC-NLSTM. 3DCNN is at the front end of the

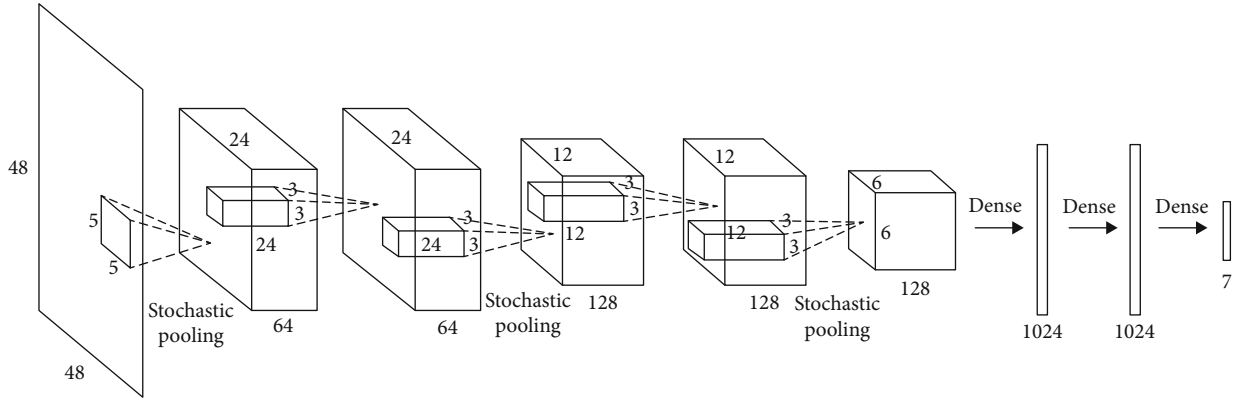


FIGURE 4: Network architecture based on convolutional neural network [32].

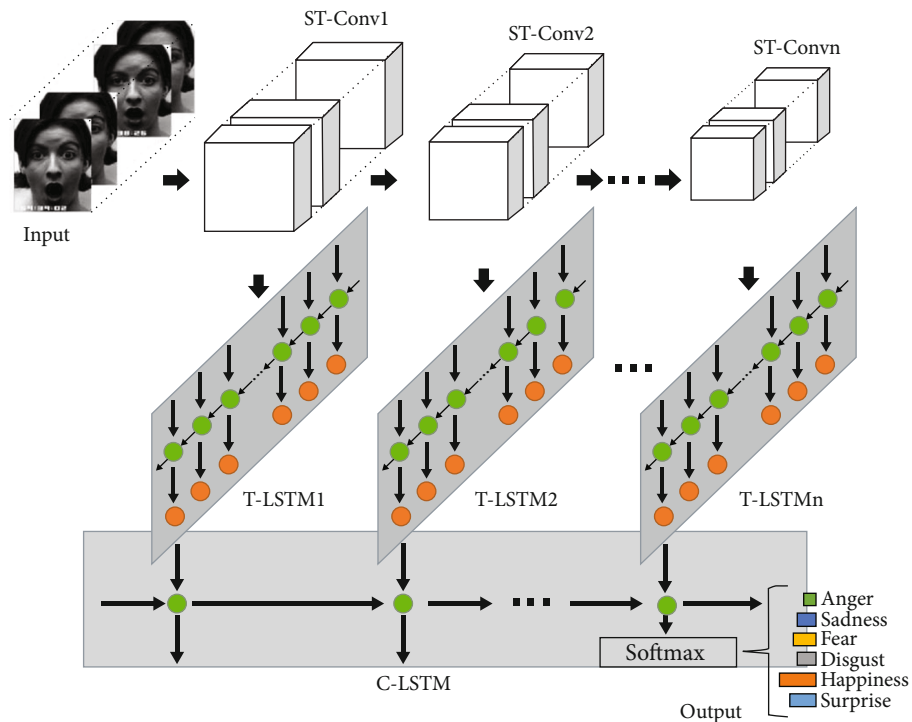


FIGURE 5: STC-NLSTM: end-to-end trained deep learning model [33].

total network architecture for learning spatiotemporal properties of input picture sequences. T-LSTM and C-LSTM are the two sections of NLSTM. T-LSTM is primarily used to learn timing information, and many T-LSTMs are coupled to different layers of 3DCNN to learn more about its properties. C-LSTM connects the outputs of numerous T-LSTMs.

3.2.3. Expression Recognition Based on Deep Belief Network.

Deep belief network (DBN) is a generative model with a restricted Boltzmann machine in the layer structure next to it. By stacking greedy training, data satisfying a given distribution may be generated. DBNs are frequently connected to classifiers or provide superior initialization parameters for deep neural networks that are turned into discriminative models when dealing with classification challenges. DBNs have been used to solve problems in various fields, including facial expression recognition.

Lv et al. [34] used histogram of oriented gradients (HOG) features to train DBN as a facial key region detector. They separated face image blocks into different sizes and extracted different local areas from each block. The best detection findings are employed in deep neural networks to recognize expressions. As illustrated in Figure 6, Kurup et al. [35] created a semisupervised DBN network that comprises two parts: (a) for unsupervised feature learning, a DBN comprised of constrained Boltzmann machines; (b) for supervised feature learning and classification, softmax activations coupled at the top of the DBN. Face identification, picture local patch extraction, manual feature extraction, and feature dimensionality reduction were among the preprocessing steps Kurup et al. took. The expression image is reduced to a low-dimensional representation, which is then fed into the proposed DBN network for feature learning and classification.

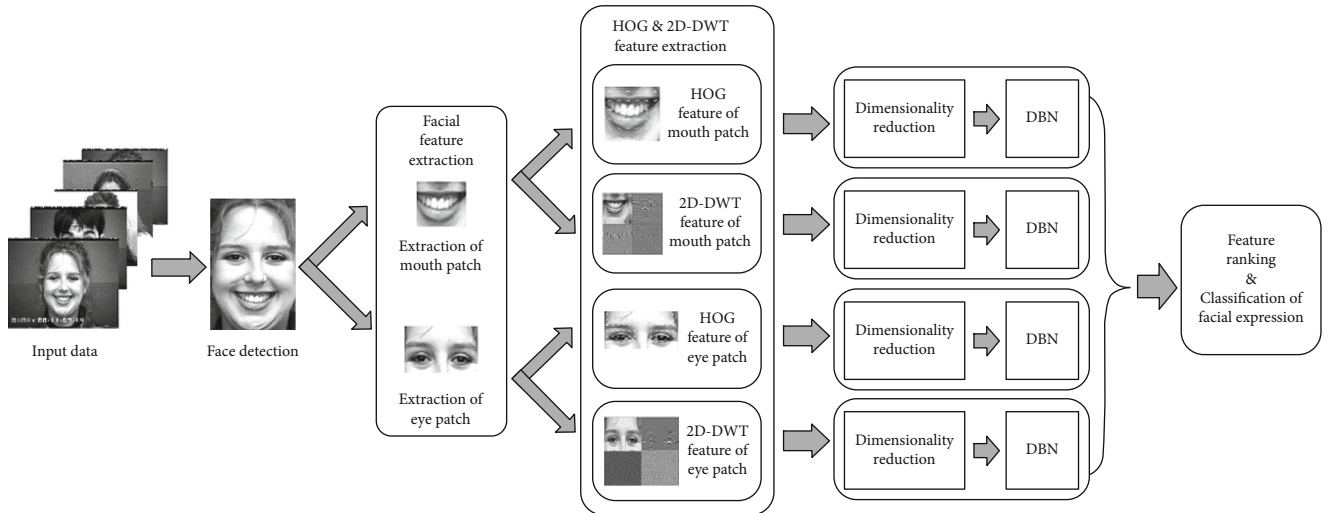


FIGURE 6: Semisupervised DBN network [35].

Deep learning successfully avoids human interference by performing end-to-end feature learning directly. The use of deep learning technology in facial expression recognition makes it easier and more effective to gather expression characteristics, boosting facial expression recognition accuracy. In 2016, Jeon et al. [36] presented a histogram of oriented gradients (HOG) for face identification, followed by the application of convolutional neural networks for feature extraction and classification tasks, with excellent results in various datasets. In 2017, Al-Shabi et al. [37] introduced a technique for integrating scale invariant feature transform (SIFT) with convolutional neural networks and created a hybrid classifier CNN-SIFT, a collection of two feature extractions with a very excellent effect on small sample detection and identification. The CK+ expression dataset is relatively small. However, the accuracy of this method on CK+ after training achieved 99.4%. According to the findings, deep learning and convolutional neural networks are superior to classical machine learning techniques in image identification, recognition, and classification [38].

3.3. Expression Recognition Algorithm Based on Broad Learning. The broad learning system (BLS) network is different from the general deep learning network [39]. It builds a horizontal network structure, and its network structure is not fixed during the training process. The widening network has reached the optimal classification effect.

Compared with the general deep learning network, the BLS has a simple structure and fewer parameters with horizontal expansion. The basis of the BLS network is the random vector functional-link neural network (RVFLNN) [40]. The network structure of RVFLNN is shown in Figure 7(a). For a given input X , a normal neural network will multiply the input X by a weight and add a bias to the next hidden layer. Nevertheless, the RVFLNN network is more than that. It also multiplies the input X by a set of random weights and adds a bias to get an enhancement layer

through nonlinear mapping of the activation function. Finally, the input X is connected to the output layer together with the enhancement layer data. Based on the RVFLNN network, the BLS network has been improved.

On the one hand, BLS does not directly use the original input data as the network input but does a linear transformation, which is equivalent to feature extraction and uses the transformed feature data as the feature layer of the network. The feature layer is nonlinearly mapped to generate an enhancement layer, and finally, the feature layer and the enhancement layer are combined as the network's input layer. BLS has a core incremental learning part, and its network structure is not fixed. During the training process, incremental learning can be achieved by increasing the number of feature nodes or enhancement nodes to improve the network model. A basic BLS network structure is shown in Figure 7(b). The BLS system was tested using the extended Cohn-Kanade dataset (CK+), and the BLS method was very effective in facial expression recognition compared to convolutional neural networks [41].

The general neural network has many network parameters due to deep layers and many neural nodes, and the training is very time-consuming. The BLS network model has a simple structure and few adjustable parameters and is suitable for data with low feature dimensions. When the network training fit is insufficient, the number of nodes in the feature layer or enhancement layer can be appropriately increased to improve the recognition effect, and the training process is simple and fast.

4. Summary and Outlook

Expression characteristics are complicated and variable, and they account for a significant percentage of emotional computing. Face recognition research using computer technology might help intelligent objects like robots better comprehend and recognize human emotions, allowing for barrier-free contact between humans and machines.

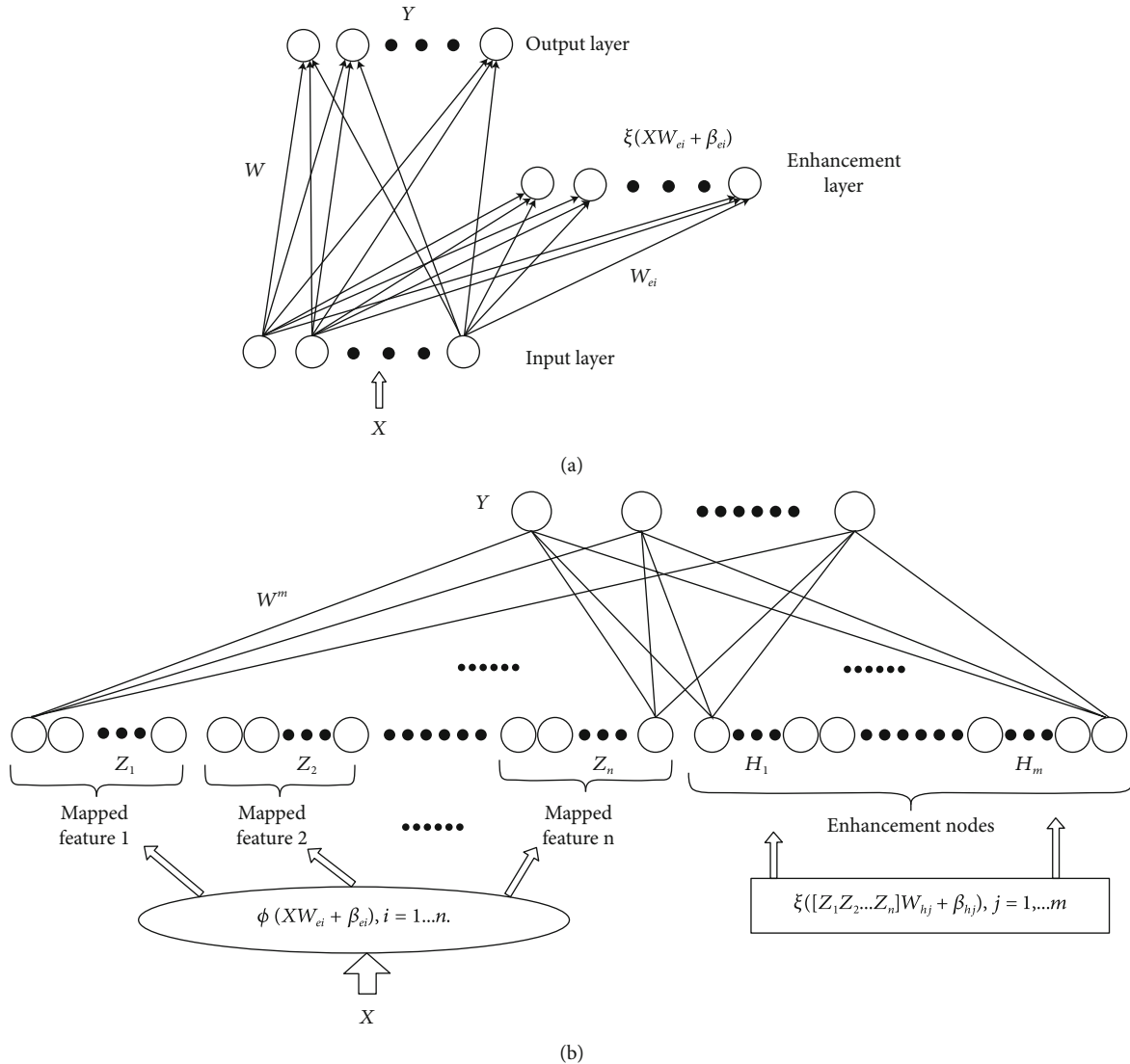


FIGURE 7: (a) RVFLNN network structure [40]. (b) BLS network structure [42].

Although deep learning and broad learning technology have made significant progress in real-world facial expression identification, there are still specific issues to be resolved in the current field.

- (1) *Head Posture Is Causing Self-Occlusion* [39]. Occlusion is a prevalent phenomenon in unconstrained natural landscapes and a tough challenge in computer vision. Head posture changes are common in expression photos, causing extreme self-occlusion, which has two immediate effects: losing essential facial information and altering the look of expressions. Local parts such as the eyes, lips, and others are not visible; the same emotion will seem different in different head situations.
- (2) *Overfitting Problem in Deep Expression Recognition*. The model in deep learning generally has a very deep hierarchical structure, and the represented function

is quite complicated, which makes overfitting very easy. The overfitting problem is especially obvious when used to expression recognition. Because human emotions are so varied, face pictures are frequently a mixture of numerous expressions rather than a single one. As a result, expression photographs have a natural quasiblur. On the other hand, hard labels can only reflect one category and cannot explain real-life expressions. Current algorithms usually use hard labels, which can easily aggravate model overfitting and severely reduce the generalization ability [40]

- (3) *The Challenge of Extracting Discriminative Expression Features*. The key to expression recognition in machine learning is to learn discriminative features from pictures and then do class matching [41]. A high degree of interaction between parameters such as backdrop and light intensity generates the picture

in an unconstrained natural scene, while numerous irrelevant qualities are jumbled on the surface. As a result, the image's complexity is tremendous, and there is much noise, making learning expression characteristics incredibly tough.

- (4) *The Selection and Extraction of Feature Information Is the Basis for Building a Wide Neural Network Model*. The video emotional features extracted by the convolutional network can hardly describe the facial expression changes in the video, so it needs to be combined with other texture features to improve the recognition rate of various emotions.

The rapid development of deep learning in recent years is evident to all, and new structures are constantly emerging, which also drives the development of facial expression recognition technology, and the accuracy on many public datasets keeps rising [42]. However, with the gradual deepening of the number of layers, the number of parameters becomes too large, requiring more resources to optimize [43]. On the other hand, broad learning has the advantages of fast computation, incremental online learning, and ease of scaling [35]. However, it is still necessary for BLS to solve the outliers and noises problem [44]. The future will focus more on facial expression recognition in-the-wild conditions [5], such as education, medical, traffic, criminal investigation, and business. Some factors such as lower image resolution, occlusion, and pose changes will make the expression discrimination and understanding more challenging [45]. Facial expression recognition that incorporates multimodal information is helpful to realizing more detailed emotional understanding, and it is the future development direction [46, 47]. A question worthy of in-depth study is whether the two algorithms can always maintain good computing power and generalization performance.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no potential conflicts of interest for the research, authorship, and/or publication of this article.

Acknowledgments

This work was financially supported by National Natural Science Foundation of China under Grant 62172184, Science and Technology Development Plan of Jilin Province of China under Grant 20200401077GX and 20200201292JC, Social Science Research of the Education Department of Jilin Province (JJKH20210901SK), Jilin Educational Scientific Research Leading Group (ZD21003), and Humanities and Social Science Foundation of Changchun Normal University (2020[011]).

References

- [1] F. An and Z. Liu, "Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM," *The Visual Computer*, vol. 36, no. 3, pp. 483–498, 2020.
- [2] J. Serrano-Puche, "Affect and the expression of emotions on the internet: an overview of current research," *Second International Handbook of Internet Research*, pp. 529–547, 2020.
- [3] R. W. Picard, *Affective Computing*, MIT press, 2000.
- [4] A. Mehrabian and A. James, *RUSSELL; JA An Approach to Environmental Psychology*, Cambridge MAUSThe MIT Press, 1974.
- [5] H. Yang, U. Ciftci, and L. Yin, "In facial expression recognition by de-expression residue learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2168–2177, Salt Lake City, UT, USA, 2018.
- [6] J.-H. Kim, B.-G. Kim, P. P. Roy, and D.-M. Jeong, "Efficient facial expression recognition algorithm based on hierarchical deep neural network structure," *IEEE Access*, vol. 7, pp. 41273–41285, 2019.
- [7] F. Kong, "Facial expression recognition method based on deep convolutional neural network combined with improved LBP features," *Personal and Ubiquitous Computing*, vol. 23, no. 3-4, pp. 531–539, 2019.
- [8] O. Ghorbanzadeh, T. Blaschke, K. Gholamnia, S. R. Meena, D. Tiede, and J. Aryal, "Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection," *Remote Sensing*, vol. 11, no. 2, p. 196, 2019.
- [9] S. Li and W. Deng, "Deep facial expression recognition: a survey," *IEEE Transactions on Affective Computing*, 2020.
- [10] M. Sown, "A preliminary note on pattern recognition of facial emotional expression," in *The 4th International Joint cConferences on Pattern Recognition*, Kyoto, Japan, 1978.
- [11] B. Yang, J.-M. Cao, D.-P. Jiang, and J.-D. Lv, "Facial expression recognition based on dual-feature fusion and improved random forest classifier," *Multimedia Tools and Applications*, vol. 77, no. 16, pp. 20477–20499, 2018.
- [12] E. Lozano-Monator, M. T. López, A. Fernández-Caballero, and F. Vigo-Bustos, "In facial expression recognition from webcam based on active shape models and support vector machines," in *International Workshop on Ambient Assisted Living*, pp. 147–154, Springer, 2014.
- [13] D. Sen, S. Datta, and R. Balasubramanian, "Facial emotion classification using concatenated geometric and textural features," *Multimedia Tools and Applications*, vol. 78, no. 8, pp. 10287–10323, 2019.
- [14] P. Barros, N. Churamani, and A. Sciutti, "The facechannel: a fast and furious deep neural network for facial expression recognition," *SN Computer Science*, vol. 1, no. 6, pp. 1–10, 2020.
- [15] R. Zhi, M. Flierl, Q. Ruan, and W. B. Kleijn, "Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 1, pp. 38–52, 2011.
- [16] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas, "Learning active facial patches for expression analysis," in *2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE*, pp. 2562–2569, Providence, RI, USA, 2012.
- [17] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

- [18] I. J. Goodfellow, D. Erhan, P. L. Carrier et al., "Challenges in representation learning: a report on three machine learning contests," in *International conference on neural information processing*, Springer, 2013.
- [19] A. Dhall, O. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon, "In video and image based emotion recognition challenges in the wild," in *EmotiW 2015, Proceedings of the 2015 ACM on international conference on multimodal interaction*, pp. 423–426, Seattle, WA, USA, 2015.
- [20] A. Dhall, R. Goecke, J. Joshi, J. Hoey, and T. Gedeon, "EmotiW 2016: video and group-level emotion recognition challenges," in *Proceedings of the 18th ACM international conference on multimodal interaction*, pp. 427–432, Tokyo, Japan, 2016.
- [21] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon, "From individual to group-level emotion recognition: EmotiW 5.0," in *Proceedings of the 19th ACM international conference on multimodal interaction*, pp. 524–528, Glasgow, Scotland, 2017.
- [22] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference On Computer Vision And Pattern Recognition-Workshops*, pp. 94–101, San Francisco, CA, 2010.
- [23] P. Borisagar, S. Jani, Y. Agrawal, and R. Parekh, "An efficient and compact review of face recognition techniques," in *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), IEEE*, pp. 1–5, Bhopal, India, 2020.
- [24] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen, "Facial expression recognition from near-infrared videos," *Image and Vision Computing*, vol. 29, no. 9, pp. 607–619, 2011.
- [25] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2852–2861, Honolulu, HI, USA, 2017.
- [26] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "In training deep networks for facial expression recognition with crowd-sourced label distribution," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 279–283, Tokyo, Japan, 2016.
- [27] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: a database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2017.
- [28] J. Shao and Y. Qian, "Three convolutional neural network models for facial expression recognition in the wild," *Neurocomputing*, vol. 355, pp. 82–92, 2019.
- [29] W. Zhou, J. Cheng, X. Lei, B. Benes, and N. Adamo, "Deep learning-based emotion recognition from real-time videos," in *Human-Computer Interaction. Multimodal and Natural Interaction. HCII 2020. Lecture Notes in Computer Science, vol 12182*, M. Kurosu, Ed., Springer, Cham, 2020.
- [30] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, and A. M. Dobaie, "Facial expression recognition via learning deep sparse auto-encoders," *Neurocomputing*, vol. 273, pp. 643–649, 2018.
- [31] R. Lu, Y. Li, P. Yang, and W. Zhang, "Facial expression recognition based on convolutional neural network," *Journal of Physics: Conference Series, IOP Publishing*, vol. 1757, no. 1, 2021.
- [32] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pp. 435–442, Motif Hotel, Seattle, 2015.
- [33] Z. Yu, G. Liu, Q. Liu, and J. Deng, "Spatio-temporal convolutional features with nested LSTM for facial expression recognition," *Neurocomputing*, vol. 317, pp. 50–57, 2018.
- [34] Y. Lv, Z. Feng, and C. Xu, "Facial expression recognition via deep learning," in *2014 International conference on smart computing*, pp. 303–308, Hong Kong, China, 2014.
- [35] A. R. Kurup, M. Ajith, and M. M. Ramón, "Semi-supervised facial expression recognition using reduced spatial features and deep belief networks," *Neurocomputing*, vol. 367, pp. 188–197, 2019.
- [36] J. Jeon, J.-C. Park, Y. Jo et al., "A real-time facial expression recognizer using deep neural network," in *Proceedings of the 10th international conference on ubiquitous information management and communication*, pp. 1–4, Danang, Viet Nam, 2016.
- [37] T. Connie, M. Al-Shabi, W. P. Cheah, and M. Goh, "Facial expression recognition using a hybrid CNN-SIFT aggregator," in *International Workshop on Multi-disciplinary Trends in Artificial Intelligence*, pp. 139–149, Springer, 2017.
- [38] K. Shan, J. Guo, W. You, D. Lu, and R. Bie, "Automatic facial expression recognition based on a deep convolutional-neural-network structure," in *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), IEEE*, pp. 123–128, London, United Kingdom, 2017.
- [39] X. Gong, T. Zhang, C. P. Chen, and Z. Liu, "Research review for broad learning system: algorithms, theory, and applications," *IEEE Transactions on Cybernetics*, pp. 1–29, 2021.
- [40] M. A. Boucher, J. Quilty, and J. Adamowski, "Data assimilation for streamflow forecasting using extreme learning machines and multilayer perceptrons," *Water Resources Research*, vol. 56, no. 6, article e2019WR026226, 2020.
- [41] T. Zhang, Z.-L. Liu, X.-H. Wang, X.-F. Xing, C. P. Chen, and E. Chen, "Facial expression recognition via broad learning system," in *2018 IEEE international conference on systems, man, and cybernetics (SMC)*, pp. 1898–1902, Miyazaki, Japan, 2018.
- [42] H. Zhao, J. Zheng, W. Deng, and Y. Song, "Semi-supervised broad learning system based on manifold regularization and broad network," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 3, pp. 983–994, 2020.
- [43] P. Zhen, H. B. Chen, Y. Cheng, Z. Ji, B. Liu, and H. Yu, "Fast Video Facial Expression Recognition by a Deeply Tensor-Compressed LSTM Neural Network for Mobile Devices," *ACM Transactions on Internet of Things*, vol. 2, no. 4, pp. 1–26, 2021.
- [44] K. Yang, Z. Yu, C. P. Chen, W. Cao, J. J. You, and H. San Wong, "Incremental weighted ensemble broad learning system for imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [45] M. Hassaballah and S. Aly, "Face recognition: challenges, achievements and future directions," *IET Computer Vision*, vol. 9, no. 4, pp. 614–626, 2015.
- [46] J. Zhang, Z. Yin, P. Chen, and S. Nichele, "Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review," *Information Fusion*, vol. 59, pp. 103–126, 2020.
- [47] S. Bhaskar and T. M. Thasleema, "LSTM model for visual speech recognition through facial expressions," *Multimedia Tools and Applications*, pp. 1–18, 2022.