A Graph-Structured Representation with BRNN for Static-based Facial Expression Recognition

Lei Zhong[1,*], Changmin Bai[1,*], Jianfeng Li[1,*], Tong Chen[1], Shigang Li[1,2], Yiguang Liu[3]

[1]School of Electronic and information engineering, Chongqing key laboratory of nonlinear circuit and intelligent information processing, Chongqing, China

[2]Graduate School of Information sciences, Hiroshima City University, Hiroshima, Japan

[3]College of Computer, Sichuan University, Chengdu, China

[*]Lei Zhong and Changmin Bai have contributed equally; Jianfeng Li is the corresponding author.

***Abstract*—Facial expression is controlled by facial muscle and can be considered as appearance and geometric variation of key parts. One key challenging issue of static-based facial expression recognition is to capture effective information from a single facial image. In this paper, we propose a graph representation with Bidirectional RNN (BRNN) for static-based facial expression recognition. Each node on the graph represents appearance information around the facial landmarks. Edges represent the geometric information encoded by the distance between two nodes. A bidirectional recurrent neural network utilized to process the graph extracts the appearance and geometric representation. The final representation from BRNN is fed into a fully connected layer and a Softmax layer to infer expressions. Experimental results show that this method achieves significant improvements over the state-of-art methods on three widely used facial databases (Oulu-CASIA, CK+, and MMI), and our method reduces the error rates of the previous best methods by 42.2%, 35.9% and 18.7%, respectively.**

Keywords-graph structured representation; facial expression recognition; BRNN

I. INTRODUCTION

Facial expression recognition has received increasing attention in the field of computer vision in recent years, and it plays an important role in many applications such as health care and human-computer interaction. Early research on facial expression recognition mainly focuses on feature learning, feature selection and classifier construction. First, features related to facial geometry or facial appearance changes are extracted from still frames or video, such as LBP-TOP [1], HOG 3D [2], and STM-ExpLet [3]. Then, a subset of the extracted features, which can be effective in distinguishing one expression from others, is selected to promote an efficient classification and to enhance the generalization capability [4]. Finally, according to the extracted features, an effective classifier is constructed to recognize facial expressions. However, desirable results are difficult to achieve with the traditional classification method.

In recent years, due to the great improvement in computer performance, deep learning has achieved remarkable results in many computer vision fields. In the field of facial expression recognition, many deep learning methods were proposed [5, 6, 7]. Different from the early method, in which images are directly input to the neural network, researchers have recently focused on how to optimize the neural network input. *Jung et al.* [8] propose to use a small CNN to capture temporal appearance features from an image sequence and use the other network to extract temporal geometry features from temporal facial landmark points. *Zhang et al.* [9] propose to use a Part-based Hierarchical Recurrent Neural Network (PHRNN) to capture dynamic features from consecutive frames and use a Multi-Signal Convolutional Neural Network (MSCNN) to extract spatial features from still images. Although these methods boost the performance of facial expression recognition, the capability of these methods is limited as they do not explicitly exploit the spatial relationship among the facial landmark, which are crucial for understanding facial expression. As we all know, facial expression is controlled by facial muscle and can be considered as dynamic variation of key parts (e.g. eyes, nose and mouth) [9]. When a facial muscle group deforms the skin of the face locally, the reflectance properties of the skin change [10]. In common, they use 66 facial landmarks to annotate the face. Among 66 facial landmarks, 49 facial landmarks (without facial contour) can well represent the key parts of the face [9]. In this paper, Similar with [11], they proposed to build graphs over the scene objects and over the question words, and they describe a deep neural network that exploits the structure in these representations.

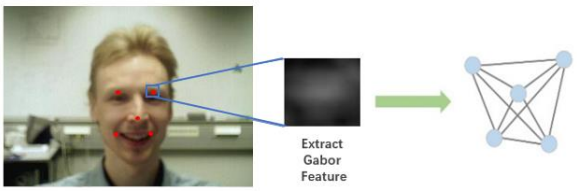

Fig. 1. We encode the facial expression image as a graph on which a BRNN operates.

Neural networks on graph structures have recently received significant attention. The discussed neural network architecture includes both recurrent neural networks [12,13] and convolutional neural networks (CNNS) [14,15]. The approach most similar to ours is the Gated Graph Sequence Neural Networks [16], which associate gated recurrent units (GRU) to each other by iteratively passing messages between neighbors. Additionally, in a related work, *Damien Teney et al.* [11] build graphs over the scene objects and over the question words and use the GRU unit to iterate each node on a graph. Finally, the features of all objects and all words are

978-1-7281-0089-0/19/$31.00 ©2019 IEEE

combined (concatenated) pairwise to predict scores over a fixed set of candidate answers. This method achieved an excellent result in visual question answering (VQA).

Inspired by the above methods, we propose a graph representation with BRNN for Static-based Facial Expression Recognition. As shown in Figure 1, we make a fully connected graph by connecting facial landmarks to each other. Different expressions will produce different texture changes around each facial landmark and different geometric changes between each node. We use the Gabor filter to extract texture features around facial landmarks, which represent nodes on the graph, and then, we use the Euclidean distance to represent the edges between these nodes. There are two main benefits to using the graph representation for facial expressions: (1) Each node represents a texture feature near the facial landmarks and can play an effective role in feature selection while removing useless information such as features near the cheeks. (2) Each edge represents the distance between two nodes and can well represent the geometric changes caused by different facial expressions. Finally, we use BRNN to iterate each node in the graph to extract features and then input the extracted features to a classifier to obtain the result of the facial expression.

The main contributions of this paper are three-fold.

(1) We propose a graph representation for static-based facial expression recognition and describe how to use a neural network capable of processing these representations to infer expressions. (2) The use of a graph structure to represent facial expression images can reduce useless information and save the time of training the neural network for facial expression recognition. (3) On three public datasets for facial expression recognition, the proposed model achieves a superior performance compared with the previous methods.

See Fig.2 for an overview of our method.

II. Our Method

A. Graphic representation of facial expression

First, we use the DRMF method [17] to calibrate 66 facial landmarks from a human face, and we remove 17 facial landmarks of the external outline of the face, for these 17 facial landmarks have small effect on the facial expressions. Then, we adopt the Gabor filter to extract the texture information near the facial landmarks. The formula that we used to generate the Gabor kernel function is as follows:

$$g(x,y;\lambda,\theta,\phi,\sigma,\gamma) = e^{-\frac{1}{2}\left[\left(\frac{x'}{\sigma}\right)^2+\left(\frac{\gamma y'}{\sigma}\right)^2\right]}\cos\left(\left(\frac{2\pi * x'}{\lambda}+\phi\right)\right) \quad (1)$$

A feature of the Gabor filter is that it contains two parameters, the dimension λ and the angle θ. Different parameter settings and combinations will produce different results. When setting the parameters, we let θ = {0, π/4, π/2, 3π/4, π, 5π/4, 3π/2, 2π}, λ = {4, $4\sqrt{2}$, 8, $8\sqrt{2}$, 16}, which will generate a group of 5×8 = 40 sets of Gabor vectors. To select the best parameter settings in the fusion of the Gabor vector, we choose to cascade and average two methods for the test. Considering the size of the Gabor core, we choose 3×3, 5×5 and 7×7 three-scales to test. The specific results are shown in the section about the experiment. Now, we have finished processing the node of the Graph, which contains the texture information for expression changes. We use x_i $(i = 1,2,\ldots,49)$ to represent the feature vector of the 49 nodes in the graph. Different expressions have different displacements of the facial landmarks. Therefore, different expressions will be distinct in the geometric distribution of facial landmarks. To magnify the distinction between different expressions and to uncover the weight relationships between the nodes, we introduce the geometric information of the facial landmarks as the edges of the graph. We calculate the Euclidean distance between any two facial landmarks and generate a 49×49 matrix. This matrix represents the geometric relationship between one landmark and the others. We use e_{ij} $(i,j = 1,2,\ldots,49)$ to represent the distance of each edge in the graph.

B. Processing graphs with neural networks

In this section, we will describe a deep neural network suitable for processing the graph from the previous section. To utilize the past contextual information and the future contextual information between nodes, we adopt the bidirectional recurrent neural network (BRNN) [18] to iterate each node on our graph. A bidirectional recurrent neural network is illustrated in Fig. 3. We replace the nonlinear units in Fig. 3 with GRU blocks [19]. Before using a neural network to process the graph, we combine the information from each node x_i with the information from the connected edges e_{ij} to form new nodes n_i $(i = 1,2,\ldots,49)$. By comparing different combination methods, we find that the best performance is achieved by averaging the connected edges and then multiplying it by the node (Eq. 2).

$$n_i = x_i \cdot \frac{\sum_{j=0}^{N} e_{ij}}{N} \quad (2)$$

Next, we input each n_i to a GRU unit in order. The forward layer and backward layer $\overrightarrow{h_i^t}$, $\overleftarrow{h_i^t}$ in the BRNN are defined as follows, where T is the number of iterations:

$$\overrightarrow{h_i^t} = GRU\left(\overrightarrow{h_i^{t-1}}, n_i\right) \quad t \in [1,T] \quad (3)$$

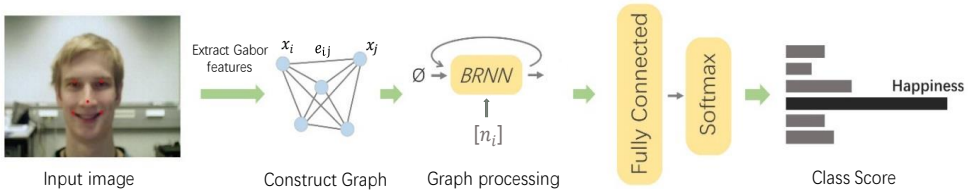

Fig. 2. Architecture of the proposed model.

$$\overleftarrow{h_i^t} = GRU\left(\overleftarrow{h_i^{t+1}}, n_i\right) \quad t \in [1, T] \tag{4}$$

The initial value is:

$$h_i^0 = 0 \tag{5}$$

Finally, we combine the forward output and the backward output as the input to the fully connected layer (Eq. 6) and a SoftMax layer (Eq. 7).

$$y_i = f(W_1\overrightarrow{h_i^T} + W_2\overleftarrow{h_i^T}) \tag{6}$$

$$y' = f'(W_3\sum_{i=1}^{N} y_i + b_1) \tag{7}$$

W1, W2, b1, and b2 are learned weights and biases, f is a ReLU, and f' a SoftMax function. The final output vector y' contains scores for the possible expression.

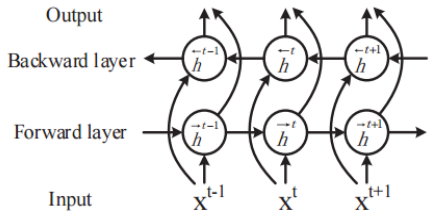

Fig. 3. The architecture of a bidirectional recurrent neural network [18].

III. Evaluation

We evaluate the performance of our model on three widely used facial expression databases, which are the CK+, Oulu-CASIA, and MMI databases. We set up multiple groups of comparison experiments on the Oulu-CASIA database to select the best parameters in a preprocessing experiment and the best neural network structure. Then, we compare our model with the state-of-art method on three facial databases.

A. Databases and Protocols

1) The Oulu-CASIA Database [20]: The Oulu-CASIA database consists of six basic expressions (anger, disgust, fear, happiness, sadness, and surprise). All expression sequences begin at the neutral emotion and end with the peak of the emotion. We select the last three frames from the expression sequence as our new dataset. Meanwhile, we employ the most popular 10-fold cross-validation protocol.

2) The CK+ database [21]: The CK+ database is the most extensively used laboratory-controlled database for expression recognition. CK+ contains 593 video sequences from 123 subjects. Similar to the Oulu database, we select the last three frames from the sequences as our database and adopt the 10-fold cross-validation strategy.

3) The MMI database [22]: The MMI database is laboratory-controlled and includes 326 sequences from 32 subjects. In this paper, we conduct our experiments on all of the 205 sequences and select the middle three frames from the sequences as our database.

B. Database classification

To select appropriate parameters, we conduct several groups of comparative experiments on the Oulu-CASIA database. First, we conduct three groups of experiments with the Gabor kernel size of 3×3, 5×5 and 7×7. The Gabor feature vectors are averaged, and the network structure is BRNN. Table 1 show the results of the different sizes of Gabor kernels on the Oulu-CASIA database. The Gabor kernel size of 3×3 achieves the best performance, and the results of the 7×7 are the worst (Table 1). Then, we compare the effect of the Gabor feature vector cascade and average in eight directions. The size of the Gabor kernel is 3×3. In our implementation (Table 2), we found that the better performance occurred with the average function, taking care to average over the connected neighbors. Therefore, we choose to average the Gabor feature vector in eight directions in the following experiment. Meanwhile, we remove the edge features by setting e_{ij} = 1. The result is shown in Table 3. The result confirms that the model makes use of the spatial relations between facial landmarks encoded by the edges of the graph. Finally, we compare the performance of BRNN and RNN on the Oulu-CASIA database. Both BRNN and RNN use GRU as the basic unit. As shown in Table 4, BRNN performs better than RNN. It confirms that BRNN integrates the direction context information from connected neighbors into each node's own representation. From the above experiments, our model can achieve the best performance when the Gabor kernel size is 3×3, the Gabor feature vector is averaged in eight directions and the structure of RNN is BRNN. As a result, we will use these settings to compare with the state-of-art method. Table 5, Table 6, and Table 7 compare the performance of our models with the current state-of-the art methods on three databases. For the Oulu-CASIA database, the best traditional algorithm for facial expression recognition is STM-ExpLet, which achieves a 74.59% accuracy. Recently, *Zhang et al.* [9] proposed the PHRNN-MSCNN model to capture the dynamic variation of facial physical structures from videos and used two signals to increase the variations of different expressions. This method obtained an 86.25% accuracy. Table 5 shows that our Graph-BRNN method achieved a satisfactory performance that outperformed the state-of-the-art method. For the CK+ and MMI databases, our proposed models also significantly outperform the previous best methods. Figure 4 shows the confusion matrix from the three databases. We can see that our methods perform well on happiness, disgust, and surprise on the Oulu-CASIA and MMI databases but have poor performance on fear, sadness and anger. The reasons are that the changes in the area around the facial landmarks for fear and sadness are relatively slight, which makes it difficult for the Gabor filter to capture the texture changes. In addition, the appearances of fear, sadness and anger are similar, which is an impediment for a neural network to distinguish.

All experiments were conducted on the TensorFlow deep learning framework. To prevent overfitting, we set the dropout of the input of GRU as 0.5, and the optimizer method is the Adam optimizer.

TABLE I
Compare different size of Gabor kernel on Oulu-CASIA database

The size of Gabor kernel	Accuracy

3×3	93.6807%
5×5	90.2392%
7×7	87.9831%

TABLE II
Different Gabor vector fusion method on Oulu-CASIA database

Gabor vector fusion method	Accuracy
Average (the size of Gabor kernel is 3×3)	93.6807%
Cascade (the size of Gabor kernel is 3×3)	89.4040%

TABLE III
With edge features and without edge features on Oulu-CASIA database

Method	Accuracy
With edge features	93.6807%
Without edge features	82.6302%

TABLE IV
Compare Basic RNN with BRNN on Oulu-CASIA database

Structure of RNN	Accuracy
RNN	86.2548%
BRNN	93.6807%

TABLE V
Compare of Different method on the Oulu-CASIA Database

Method	Descriptor	Accuracy
Liu et al [3].	STM-ExpLet	6 classes:74.59%
Guo et al [23].	Atlases	6 classes:75.52%
Jung et al [8]	DNN	6 classes:74.17%
Jung et al [8]	CNN-DNN	6 classes: 81.46%
Zhao et al. [24]	PPDN	6 classes: 84.59%
Yu et al. [25]	DPCN	6 classes: 86.23%
Zhang et al. [9]	PHRNN-MSCNN	6 classes: 86.25%
Yang et al. [35]	DeRL	6 classes: 88.00%
Our method	**Graph-BRNN**	**6 classes: 93.06%**

TABLE VI
Comparison of Different method on the CK+ Database

Method	Descriptor	Accuracy
Cai et al [26]	Island loss	7 classes: 94.35%
Zeng et al [27]	DSAE	7 classes: 95.79%
Meng et al [28]	multitask network	7 classes: 95.37%
Liu et al [29]	(N+M)-tuplet clusters loss	7 classes: 97.10%
Yang et al. [35]	DeRF	7 classes: 97.30%
Our method	**Graph-BRNN**	**7 classes: 98.27%**

TABLE VII
Comparison of Different method on the MMI Database

Method	Descriptor	Accuracy
Zhong et al [30]	CSPL	6 classes:73.53%
Liu et. al [31]	3DCNN-DAP	6 classes:63.4%
Jung et.al [8]	CNN-DNN	6 classes:70.24%
Hasani et al [32]	3DCNN-LSTM	6 classes: 77.50%
Kim et al [33]	CNN-LSTM	6 classes: 78.61%
Hasani et al [32]	CNN-CRF	6 classes: 78.68%
Zhang et al. [9]	PHRNN-MSCNN	6 classes: 81.18%
Sun et al [34]	Network ensemble	6 classes :91.46%
Our method	**Graph-BRNN**	**6 classes: 94.44%**

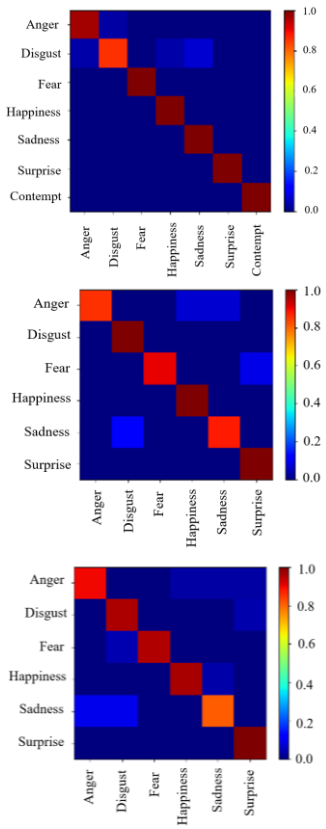

Fig. 4. Confusion matrix on the CK+, MMI and Oulu-CASIA databases

IV. Conclusion

In this paper, we first presented a deep neural network for facial expression recognition that processes graph-structured representations of facial expressions. In our opinion, the variation in the area around facial landmarks contains useful information to distinguish different expressions, while the entire facial image contains much useless information. Therefore, we utilize the Gabor filter to extract texture features of the facial landmarks, which construct the nodes in our graph. The distance between each landmark is taken as the edge of our graph representing the geometric information from the facial image. Finally, we adopt BRNN iterations on each node of our graph to predict the expression. Experimental results on three databases demonstrate that our model achieved state-of-the-art performance.

Acknowledgment

This work is partially supported by the Fundamental Research Funds for the Central Universities (SWU117024), NSFC under Grant 61860206007 and 61571313, National Training Program of Innovation and Entrepreneurship for Undergraduates (201810635078), and by funding from Sichuan Province under Grant 18GJHZ0138.

REFERENCES

[1] Shan, C., Gong, S., & McOwan, P. W. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing, 27(6), 803-816.

[2] Klaser, A., Marszałek, M., & Schmid, C. (2008, September). A spatio-temporal descriptor based on 3d-gradients. In BMVC 2008-19th British Machine Vision Conference (pp. 275-1). British Machine Vision Association.

[3] Liu, M., Shan, S., Wang, R., & Chen, X. (2014, June). Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on(pp. 1749-1756). IEEE.

[4] Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2005, June). Recognizing facial expression: machine learning and application to spontaneous behavior. InComputer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 2, pp. 568-573). IEEE.

[5] Graves, A., Mayer, C., Wimmer, M., Schmidhuber, J., & Radig, B. (2008). Facial expression recognition with recurrent neural networks. In Proceedings of the International Workshop on Cognition for Technical Systems.

[6] Kaya, H., Gürpınar, F., & Salah, A. A. (2017). Video-based emotion recognition in the wild using deep transfer learning and score fusion. Image and Vision Computing, 65, 66-75.

[7] Kim, Y., Yoo, B., Kwak, Y., Choi, C., & Kim, J. (2017). Deep generative-contrastive networks for facial expression recognition. arXiv preprint arXiv:1703.07140.

[8] Jung, H., Lee, S., Yim, J., Park, S., & Kim, J. (2015, December). Joint fine-tuning in deep neural networks for facial expression recognition. In Computer Vision (ICCV), 2015 IEEE International Conference on (pp. 2983-2991). IEEE.

[9] Zhang, K., Huang, Y., Du, Y., & Wang, L. (2017). Facial expression recognition based on deep evolutional spatial-temporal networks. IEEE Transactions on Image Processing,26(9), 4193-4203.

[10] Angelopoulo, E., Molana, R., & Daniilidis, K. (2001). Multispectral skin color modeling. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 2, pp. II-II). IEEE.

[11] Teney, D., Liu, L., & van den Hengel, A. (2017, July). Graph-Structured Representations for Visual Question Answering. InComputer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on (pp. 3233-3241). IEEE.

[12] Tai, K. S., Socher, R., & Manning, C. D. (2015). Improved semantic representations from tree-structured long short-term memory networks. arXiv preprint arXiv:1503.00075.

[13] Oord, A. V. D., Kalchbrenner, N., & Kavukcuoglu, K. (2016). Pixel recurrent neural networks. arXiv preprint arXiv:1601.06759.

[14] Henaff, M., Bruna, J., & LeCun, Y. (2015). Deep convolutional networks on graph-structured data. arXiv preprint arXiv:1506.05163.

[15] Bruna, J., Zaremba, W., Szlam, A., & Lecun, Y. (2014). Spectral networks and locally connected networks on graphs. In International Conference on Learning Representations (ICLR2014), CBLS, April 2014.

[16] Li, Y., Tarlow, D., Brockschmidt, M., & Zemel, R. (2015). Gated graph sequence neural networks. arXiv preprint arXiv:1511.05493.

[17] Asthana, A., Zafeiriou, S., Cheng, S., & Pantic, M. (2013, June). Robust discriminative response map fitting with constrained local models. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on (pp. 3444-3451). IEEE.

[18] Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing,45(11), 2673-2681.

[19] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

[20] Taini, M., Zhao, G., Li, S. Z., & Pietikainen, M. (2008, December). Facial expression recognition from near-infrared video sequences. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on (pp. 1-4). IEEE.

[21] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010, June). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on (pp. 94-101). IEEE.

[22] Valstar, M., & Pantic, M. (2010, May). Induced disgust, happiness and surprise: an addition to the mmi facial expression database. In Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect (p. 65).

[23] Guo, Y., Zhao, G., & Pietikäinen, M. (2012). Dynamic facial expression recognition using longitudinal facial expression atlases. In Computer Vision–ECCV 2012 (pp. 631-644). Springer, Berlin, Heidelberg.

[24] Zhao, X., Liang, X., Liu, L., Li, T., Han, Y., Vasconcelos, N., & Yan, S. (2016, October). Peak-piloted deep network for facial expression recognition. In European conference on computer vision (pp. 425-442). Springer, Cham.

[25] Yu, Z., Liu, Q., & Liu, G. (2017). Deeper cascaded peak-piloted network for weak expression recognition. The Visual Computer, 1-9.

[26] Cai, J., Meng, Z., Khan, A. S., Li, Z., O'Reilly, J., & Tong, Y. (2018, May). Island Loss for Learning Discriminative Features in Facial Expression Recognition. In Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on (pp. 302-309). IEEE.

[27] Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y., & Dobaie, A. M. (2018). Facial expression recognition via learning deep sparse autoencoders. Neurocomputing, 273, 643-649.

[28] Meng, Z., Liu, P., Cai, J., Han, S., & Tong, Y. (2017, May). Identity-aware convolutional neural network for facial expression recognition. In Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on (pp. 558-565). IEEE.

[29] Liu, X., Kumar, B. V. K. V., You, J., & Jia, P. (2017). Adaptive Deep Metric Learning for Identity-Aware Facial Expression Recognition. IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp.522-531). IEEE Computer Society.

[30] Zhong, L., Liu, Q., Yang, P., Liu, B., Huang, J., & Metaxas, D. N. (2012, June). Learning active facial patches for expression analysis. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (pp. 2562-2569). IEEE.

[31] Liu, M., Li, S., Shan, S., Wang, R., & Chen, X. (2014, November). Deeply learning deformable facial action parts model for dynamic expression analysis. In Asian conference on computer vision (pp. 143-157). Springer, Cham.

[32] Hasani, B., & Mahoor, M. H. (2017, July). Facial expression recognition using enhanced deep 3D convolutional neural networks. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on (pp. 2278-2288). IEEE.

[33] Kim, D. H., Baddar, W., Jang, J., & Ro, Y. M. (2017). Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. IEEE Transactions on Affective Computing.

[34] Sun, N., Li, Q., Huan, R., Liu, J., & Han, G. (2017). Deep spatial-temporal feature fusion for facial expression recognition in static images. Pattern Recognition Letters.

[35] Yang, Huiyuan, Umur Ciftci, and Lijun Yin. "Facial Expression Recognition by De-Expression Residue Learning." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.