**REFERENCES**

[1] P. Ekman and W. V. Friesen, ``Constants across cultures in the face and emotion.,'' J. Personality Social Psychol., vol. 17, no. 2, pp. 124-129, 1971.

[2] I. J. Goodfellow et al., ``Challenges in representation learning: A report on three machine learning contests,'' Neural Netw., vol. 64, pp. 59-63, Apr. 2015.

[3] S. Lawrence, C. L. Giles, A. Chung Tsoi, and A. D. Back, ``Face recognition: A convolutional neural-network approach,'' IEEE Trans. Neural Netw., vol. 8, no. 1, pp. 98113, Jan. 1997.

[4] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, ``Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,'' IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 12851298, May 2016.

[5] T. Chang, G. Wen, Y. Hu, and J. Ma, ``Facial expression recognition based on complexity perception classication algorithm,'' 2018, arXiv:1803.00185. [Online]. Available: http://arxiv.org/abs/1803.00185

[6] M.-I. Georgescu, R. T. Ionescu, and M. Popescu, ``Local learning with deep and handcrafted features for facial expression recognition,'' IEEE Access, vol. 7, pp. 6482764836, 2019.

[7] C. Du and S. Gao, ``Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network,'' IEEE Access, vol. 5, pp. 1575015761, 2017.

[8] M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. Alrubaian, and G. Fortino, ``Facial expression recognition utilizing local direction-based robust features and deep belief network,'' IEEE Access, vol. 5, pp. 45254536, 2017.

[9] M. Z. Uddin, W. Khaksar, and J. Torresen, ``Facial expression recognition using salient features and convolutional neural network,'' IEEE Access, vol. 5, pp. 2614626161, 2017.

[10] D. Amodei et al., ``Deep speech 2: End-to-end speech recognition in english and mandarin,'' in Proc. Int. Conf. Mach. Learn., Jun. 2016, pp. 173182.

[11] C. Szegedy, S. Ioffe, and V. Vanhoucke, ``Inception-v4, inception-ResNet and the impact of residual connections on learning,'' in Proc. 31st AAAI Conf. Artif. Intell., 2016, p. 1.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ``ImageNet classification with deep convolutional neural networks,'' in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2012, pp. 10971105.

[13] M. R. Koujan, A. Akram, P. McCool, J. Westerfeld, D. Wilson, K. Dhaliwal, S. McLaughlin, and A. Perperidis, ``Multi-class classification of pulmonary endomicroscopic images,'' in Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI), Apr. 2018, pp. 15741577.

[14] O. Leonovych, M. R. Koujan, A. Akram, J. Westerfeld, D. Wilson, K. Dhaliwal, S. McLaughlin, and A. Perperidis, ``Texture descriptors for classifying sparse, irregularly sampled optical endomicroscopy images,'' in Proc. Annu. Conf. Med. Image Understand. Anal. Cham, Switzerland: Springer, 2018, pp. 165176.

[15] D. Hazarika, S. Gorantla, S. Poria, and R. Zimmermann, ``Self-attentive feature-level fusion for multimodal emotion detection,'' in Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR), Apr. 2018, pp. 196201.

[16] K.-Y. Huang, C.-H. Wu, Q.-B. Hong, M.-H. Su, and Y.-H. Chen, ``Speech emotion recognition using deep neural network considering verbal and nonverbal speech sounds,'' in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2019, pp. 58665870.

[17] M. E. Kret, K. Roelofs, J. J. Stekelenburg, and B. de Gelder, ``Emotional signals from faces, bodies and scenes inuence observers' face expressions, Fixations and pupil-size,'' Frontiers Human Neurosci., vol. 7, p. 810, Dec. 2013.

[18] D. Kollias, P. Tzirakis, M. A. Nicolaou, A. Papaioannou, G. Zhao, B. Schuller, I. Kotsia, and S. Zafeiriou, ``Deep affect prediction inthe-wild: Aff-wild database and challenge, deep architectures, and beyond,'' Int. J. Comput. Vis., vol. 127, pp. 123, Jun. 2019.

[19] D. Kollias, A. Schulc, E. Hajiyev, and S. Zafeiriou, ``Analysing affective behavior in the rst ABAW 2020 competition,'' 2020, arXiv:2001.11409. [Online]. Available: http://arxiv.org/abs/2001.11409

[20] W. Y. Choi, K. Y. Song, and C.W. Lee, ``Convolutional attention networks for multimodal emotion recognition from speech and text data,'' in Proc. Grand Challenge Workshop Hum. Multimodal Lang. (Challenge-HML), 2018, pp.

2834.

[21] E. Marinoiu, M. Zanr, V. Olaru, and C. Sminchisescu, ``3D human sensing, action and emotion recognition in robot assisted therapy of children with autism,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 21582167.

[22] Z. Du, S. Wu, D. Huang, W. Li, and Y. Wang, ``Spatio-temporal encoder- decoder fully convolutional network for video-based dimensional emotion recognition,'' IEEE Trans. Affect. Comput., early access, Sep. 10, 2019, doi: 10.1109/TAFFC.2019.2940224.

[23] J. Yang, K. Wang, X. Peng, and Y. Qiao, ``Deep recurrent multi-instance learning with spatio-temporal features for engagement intensity prediction,'' in Proc. 20th ACM Int. Conf. Multimodal Interact., Oct. 2018, pp. 594598.

[24] P. Barros, N. Churamani, and A. Sciutti, ``The FaceChannel:Alight-weight deep neural network for facial expression recognition,'' in Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), Buenos Aires, AR, USA, Apr. 2020 pp. 449453.

[25] M. Koujan, L. Alharbawee, G. Giannakakis, N. Pugeault, and A. Roussos, ``Real-time facial expression recognition `in the wild' by disentangling 3D expression from identity,'' in Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), Buenos Aires, AR, USA, May 2020 pp. 539546.

[26] S. Xiao, P. Ting, and R. Fu-Ji, ``Facial expression recognition using ROI-KNN deep convolutional neural networks,'' Acta Automatica Sinica, vol. 42, no. 6, pp. 883891, 2016.

[27] M. Liu, S. Li, S. Shan, R. Wang, and X. Chen, ``Deeply learning deformable facial action parts model for dynamic expression analysis,'' in Proc. ACCV. Cham, Switzerland: Springer, 2014, pp. 143157.

[28] G. Zhao, H. Yang, and M. Yu, ``Expression recognition method based on a lightweight convolutional neural network,'' IEEE Access, vol. 8, pp. 3852838537, 2020.

[29] A. F. Abate, P. Barra, S. Barra, C. Molinari, M. Nappi, and F. Narducci, ``Clustering facial attributes: Narrowing the path from soft to hard biometrics,'' IEEE Access, vol. 8, pp. 90379045, 2020, doi: 10.1109/ACCESS.2019.2962010.

[30] K. Simonyan and A. Zisserman, ``Very deep convolutional networks for large-scale image recognition,'' 2014, arXiv:1409.1556. [Online]. Available: https://arxiv.org/abs/1409.1556

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, ``Rethinking the inception architecture for computer vision,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 28182826.

[32] F. Chollet, ``Xception: Deep learning with depthwise separable convolutions,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2016, pp. 12511258.

[33] K. He, X. Zhang, S. Ren, and J. Sun, ``Deep residual learning for image recognition,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770778.

[34] G. A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, ``Mobilenets: Efcient convolutional neural networks for mobile vision applications,'' CoRR, vol. abs/1704.04861, pp. 19, Apr. 2017.

[35] K. Zhang, Z. Zhang, Z. Li, andY. Qiao, ``Joint face detection and alignment using multitask cascaded convolutional networks,'' IEEE Signal Process. Lett., vol. 23, no. 10, pp. 14991503, Oct. 2016.

[36] D. Kingma and J. Ba, ``Adam: A method for stochastic optimization,'' 2014, arXiv:1412.6980. [Online]. Available: https://arxiv.org/abs/1412.6980

[37] M. Lin, Q. Chen, and S. Yan, ``Network in network,'' 2019, arXiv:1312.4400. [Online]. Available: https://arxiv.org/abs/1312.4400

[38] J. Z. Ji, Application Research onWeakly-Supervised Learning in Computer Vision. Chengdu, China: Univ. of Electronic Science and Technology of China, 2011.

[39] J. Tobias Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, ``Striving for simplicity: The all convolutional net,'' 2014,

arXiv:1412.6806. [Online]. Available: http://arxiv.org/abs/1412.6806

[40] X. Glorot, A. Bordes, and Y. Bengio, ``Deep sparse rectifier neural networks,'' in Proc. 14th Int. Conf. Artif. Intell. Statist., 2011, pp. 315323.

[41] G. Zeng, J. Zhou, X. Jia, W. Xie, and L. Shen, ``Hand-crafted feature guided deep learning for facial expression recognition,'' in Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), May 2018, pp. 423430.

[42] H. Wang and S. Hou, ``Facial expression recognition based on the fusion of CNN and SIFT features,'' in Proc. IEEE 10th Int. Conf. Electron. Inf. Emergency Commun. (ICEIEC), Jul. 2020, pp. 190194, doi: 10.1109/ICEIEC49280.2020.9152361.

[43] A. Mollahosseini, D. Chan, and M. H. Mahoor, ``Going deeper in facial expression recognition using deep neural networks,'' in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Lake Placid, NY, USA, Mar. 2016, pp. 110, doi: 10.1109/WACV.2016.7477450.

[44] S. Miao, H. Xu, Z. Han, and Y. Zhu, ``Recognizing facial expressions using a shallow convolutional neural network,'' IEEE Access, vol. 7, pp. 7800078011, 2019, doi: 10.1109/ACCESS.2019.2921220.

[45] K. Liu, M. Zhang, and Z. Pan, ``Facial expression recognition with CNN ensemble,'' in Proc. Int. Conf. Cyberworlds (CW), Chongqing, China, Sep. 2016, pp. 163166, doi: 10.1109/CW.2016.34.

[46] Z. Yi-Kui and L. Jian, ``Facial expression recognition based on transferring convolutional neural network,'' J. Signal Process., vol. 34, no. 6, pp. 729738, 2018.

[47] Y. Fang, ``Research of facial expression recognition based on convolutional neural network,'' M.S. thesis, Dept. Comput. Softw. Comput. Appl., Xidian Univ., Xi'an, China, 2017.

[48] L. L. Xu, S. M. Zhang, and J. L. Zhao, ``Expression recognition algorithm for parallel convolutional neural networks,'' J. Image Graph., vol. 24, no. 2, pp. 02270236, 2019.

[49] H. Ma and T. Celik, ``FER-Net: Facial expression recognition using densely connected convolutional network,'' Electron. Lett., vol. 55, no. 4, pp. 184186, Feb. 2019, doi: 10.1049/el.2018.7871.

[50] W. Hua, F. Dai, L. Huang, J. Xiong, and G. Gui, ``HERO: Human emotions recognition for realizing intelligent Internet of Things,'' IEEE Access, vol. 7, pp. 2432124332, 2019, doi: 10.1109/ACCESS.2019.2900231.

[51] F. Chang, A. Tuan Tran, T. Hassner, I. Masi, R. Nevatia, and G. Medioni, ``ExpNet: Landmark-free, deep, 3D facial expressions,'' in Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), Xi'an, China, 2018, pp. 122129, doi: 10.1109/FG.2018.00027.

[52] M. V. Zavarez, R. F. Berriel, and T. Oliveira-Santos, ``Cross-database facial expression recognition based on fine-tuned deep convolutional network,'' in Proc. 30th SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI), Niteroi, Brazil, Oct. 2017, pp. 405412, doi: 10.1109/SIBGRAPI.2017.60.

[53] A. M. Ali, H. Zhuang, and A. K. Ibrahim, ``An approach for facial expression classification,'' IJBM, vol. 9, no. 2, pp. 96112, 2017.