Facial Expression Recognition Using Local Sliding Window Attention

Subjects: Computer Science, Artificial Intelligence

Contributor: Shuang Qiu , Guangzhe Zhao , Xiao Li , Xueping Wang

There are problems associated with facial expression recognition (FER), such as facial occlusion and head pose variations. These two problems lead to incomplete facial information in images, making feature extraction extremely difficult.

facial expression recognition sliding window local feature enhancement

1. Introduction

Facial expressions are the most intuitive and effective body language symbols to convey emotions. Facial expression recognition (FER) enables machines to recognize human facial expressions automatically. FER has made a massive contribution to the fields of human–computer interactions [1], vehicle-assisted driving [2], medical services [3], and social robots [4]. FER is becoming an increasingly active field, and great progress has been made in this area in the past few decades.

Excellent recognition results [5][6] have been achieved on datasets collected in controlled laboratory environments, such as CK+ [7], MMI [8], and JAFFE [9]. However, in the wild, FER is still far from satisfactory. Occlusions and head pose variations are the common, intractable problems experienced in unconstrained settings. Therefore, researchers have proposed diverse real-world facial expression datasets such as FER2013 [10], RAF-DB [11], AffectNet [12] to advance the development of FER in the wild. The authors of [13][14][15][16] have made great contributions based on datasets of natural scenes. Li et al. [15] used the face as a whole to obtain various global features to produce a competitive performance, but they ignored the importance of local features. Occlusion or the appearance of poses in the wild can cause the effective information about the face to be incomplete. Therefore, it is difficult to mine significant expression features, resulting in the model having a poor discrimination ability. Karnati et al. [5] and Ruan et al. [6] showed that the learning of fine-grained features plays a huge role in expression recognition, thus considering both global features and local features is a better choice.

Most current methods for obtaining local features crop essential areas of the face through prior knowledge [16][17][18] or directly crop the face into patches of fixed size [14][19]. However, these methods have some problems: (1) The method of cropping faces through prior knowledge is in the data preprocessing stage, and most research methods conduct cropping based on landmarks. Although this method is intuitive, facial keypoint detection is prone to the effects of occlusion and pose. As shown in **Figure 1**a, landmarks are not accurately located. Therefore, the cropping method based on prior knowledge itself contains certain noise. (2) The cropping of images into fixed-scale patches does not require prior knowledge but may divide useful features into different patches. As shown in **Figure 1**b, eyes are an essential feature for judging expressions, and direct segmentation may divide the same eye into several different patches, causing the integrity of essential features to be destroyed. Thus, the sliding window-based cropping strategy is introduced to obtain the complete feature, as shown in **Figure 1**c. It requires no prior knowledge and ensures that essential features are not segmented, which is beneficial for FER.

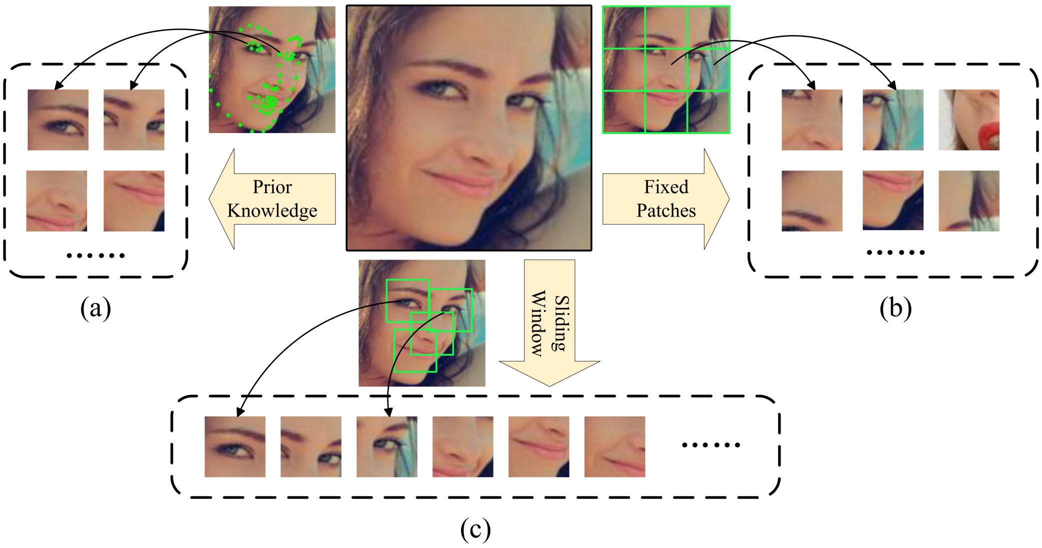

Figure 1. Different cropping strategies for facial expression images. (**a**) Cropping based on prior knowledge. Green dots represent facial landmarks. Positioning and cropping can be performed according to point coordinates. (**b**) Cropping based on fixed-size patches. The face image is equally divided into 9 regions. (**c**) Cropping based on sliding window. The window scans the entire face image and accurately locates the key regions of the face.

2. Facial Expression Recognition in the Wild

Some traditional methods use feature engineering, including SIFT [20], HOG [21], and Gabor [22], to focus on global facial information. This has achieved good results. Convolutional Neural Networks (CNN) have been employed with success in the image field, and Fasel et al. [23] found that they have strong robustness to facial pose and scale in face recognition tasks. Tang et al. [24] and Kahou et al. [25] designed deep CNN models to win the championships of FER2013 and Emotiw2013. With the development of GPU hardware, more and more models using deep CNN are being developed. FER based on deep learning has gradually gained the upper hand and become a mainstream research method. Researchers rely on powerful deep learning to quickly overcome the challenge associated with FER in a controlled environment.

More and more researchers are turning their attention to challenging wild environment conditions and working on solving facial recognition problems under natural conditions [26] including illumination changes, occlusions, head pose variations, and blur. Karnati et al. [5] proposed the use of multiscale convolution to obtain more fine-grained features and reduce intraclass differences, aiming to solve the illumination problem. Zhang et al. [27] proposed the use of “uncertainty” learning to quantify the degree of “uncertainty” for various noise problems in facial expressions, mixing “uncertainty” features from different faces to separate noise and expression features. Zou et al. [28] regarded expressions as a weighted sum of different types of expressions and learned basic expression features through a sequential decomposition mechanism. Fan et al. [29] designed a two-stage training program to further recognize expressions using identity information. Zhang et al. [13] proposed Erasing Attention Consistency, which mines the features of the expression itself, rather than the label-related features, to suppress the learning of noise information during training. Ruan et al. [6] learned intraclass features and interclass features by decomposing and reconstructing. Jiang et al. [30] proposed an identity and pose disentangled method, which separates expression features from the identity and pose.

3. Facial Expression Recognition Based on Local Features

Due to the enormous difficulty in recognizing expression images in the natural environment, guiding models to mine local significant features has become the choice of more researchers. Local-based FER uses some strategies to crop the face image into several local regions and solves the noise problem associated with facial images in the real world by obtaining local information. Li et al. [18] cut out 24 small patches according to the landmark coordinates, generated corresponding weights according to the degree of occlusion, and then predicted the result with the global features. Wang et al. [17] considered multiple cropping strategies and proposed the RB-Loss method to assign different weights to different regions. Zhao et al. [19] reduced the occlusion and pose interference through the use of multiscale features and local attention modules. Liu et al. [16] proposed adaptive local cropping, and particularly cropped the eye and mouth parts, guiding the model to find more distinguishable parts. This method is robust to occlusion and pose changes. Krithika et al. [31] segmented the face and background information and then cut out the eyes, nose, and mouth and proposed a Minimal Angular Feature-Oriented Network to obtain specific expression features. Xue et al. [14] guided the model to learn diverse information within patches and identified rich relationships between patches. Although these methods focus on local features through facial key point cropping or fixed-size cropping, the former will cause inaccurate key point positioning due to facial occlusion and head poses, and the latter may divide essential features into different patches, thereby destroying the integrity of features. Even if this effect is mitigated by mixing the two cropping strategies, the model will still ignore finer expression details, such as brows and muscle lines.

References

1. Chowdary, M.K.; Nguyen, T.N.; Hemanth, D.J. Deep learning-based facial emotion recognition for human–computer interaction applications. Neural Comput. Appl. 2021, 1–18.
2. Zhang, Y.; Hua, C. Driver fatigue recognition based on facial expression analysis using local binary patterns. Optik 2015, 126, 4501–4505.
3. Bisogni, C.; Castiglione, A.; Hossain, S.; Narducci, F.; Umer, S. Impact of deep learning approaches on facial expression recognition in healthcare industries. IEEE Trans. Ind. Inform.

2022, 18, 5619–5627.

4. Ruiz-Garcia, A.; Webb, N.; Palade, V.; Eastwood, M.; Elshaw, M. Deep learning for real time facial expression recognition in social robots. In Proceedings of the International Conference on Neural Information Processing (ICONIP); Springer: Cham, Switzerland, 2018; pp. 392–402.
5. Karnati, M.; Seal, A.; Yazidi, A.; Krejcar, O. FLEPNet: Feature Level Ensemble Parallel Network for Facial Expression Recognition. IEEE Trans. Affect. Comput. 2022, 13, 2058–2070.
6. Ruan, D.; Yan, Y.; Lai, S.; Chai, Z.; Shen, C.; Wang, H. Feature Decomposition and Reconstruction Learning for Effective Facial Expression Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021; pp. 7656–7665.
7. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.
8. Pantic, M.; Valstar, M.; Rademaker, R.; Maat, L. Web-based database for facial expression analysis. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Amsterdam, The Netherlands, 6–9 July 2005; p. 5.
9. Lyons, M.; Akamatsu, S.; Kamachi, M.; Gyoba, J. Coding facial expressions with Gabor wavelets. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG), Nara, Japan, 14–16 April 1998; pp. 200–205.
10. Goodfellow, I.J.; Erhan, D.; Carrier, P.L.; Courville, A.; Mirza, M.; Hamner, B.; Cukierski, W.; Tang, Y.; Thaler, D.; Lee, D.H.; et al. Challenges in representation learning: A report on three machine learning contests. In Proceedings of the International Conference on Neural Information Processing (ICONIP), Daegu, Republic of Korea, 3–7 November 2013; pp. 117–124.
11. Li, S.; Deng, W.; Du, J. 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 (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2584–2593.
12. Mollahosseini, A.; Hasani, B.; Mahoor, M.H. AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. IEEE Trans. Affect. Comput. 2019, 10, 18–31.
13. Zhang, Y.; Wang, C.; Ling, X.; Deng, W. Learn from all: Erasing attention consistency for noisy label facial expression recognition. In Proceedings of the European Conference on Computer Vision (ECCV), Tel Aviv, Israel, 23–27 October 2022; pp. 418–434.
14. Xue, F.; Wang, Q.; Guo, G. TransFER: Learning Relation-aware Facial Expression Representations with Transformers. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Virtual Event, 11–17 October 2021; pp. 3581–3590.

15. Li, Y.; Gao, Y.; Chen, B.; Zhang, Z.; Lu, G.; Zhang, D. Self-Supervised Exclusive-Inclusive Interactive Learning for Multi-Label Facial Expression Recognition in the Wild. IEEE Trans. Circuits Syst. Video Technol. 2022, 32, 3190–3202.
16. Liu, H.; Cai, H.; Lin, Q.; Li, X.; Xiao, H. Adaptive Multilayer Perceptual Attention Network for Facial Expression Recognition. IEEE Trans. Circuits Syst. Video Technol. 2022, 32, 6253–6266.
17. Wang, K.; Peng, X.; Yang, J.; Meng, D.; Qiao, Y. Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition. IEEE Trans. Image Process. 2020, 29, 4057–4069.
18. Li, Y.; Zeng, J.; Shan, S.; Chen, X. Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism. IEEE Trans. Image Process. 2019, 28, 2439–2450.
19. Zhao, Z.; Liu, Q.; Wang, S. Learning Deep Global Multi-Scale and Local Attention Features for Facial Expression Recognition in the Wild. IEEE Trans. Image Process. 2021, 30, 6544–6556.
20. Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003, 31, 3812–3814.
21. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 886–893.
22. Liu, C.; Wechsler, H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans. Image Process. 2002, 11, 467–476.
23. Fasel, B. Robust face analysis using convolutional neural networks. In Proceedings of the 2002 International Conference on Pattern Recognition, Quebec City, QC, Canada, 11–15 August 2002; Volume 2, pp. 40–43.
24. Tang, Y. Deep learning using linear support vector machines. arXiv 2013, arXiv:1306.0239.
25. Kahou, S.E.; Pal, C.; Bouthillier, X.; Froumenty, P.; Gülçehre, Ç.; Memisevic, R.; Vincent, P.; Courville, A.; Bengio, Y.; Ferrari, R.C.; et al. Combining modality specific deep neural networks for emotion recognition in video. In Proceedings of the 15th ACM International Conference on Multimodal Interaction (ICMI), Sydney, Australia, 9–13 December 2013; pp. 543–550.
26. Li, S.; Deng, W. Deep facial expression recognition: A survey. IEEE Trans. Affect. Comput. 2020, 13, 1195–1215.
27. Zhang, Y.; Wang, C.; Deng, W. Relative Uncertainty Learning for Facial Expression Recognition. Adv. Neural Inf. Process. Syst. 2021, 34, 17616–17627.
28. Zou, X.; Yan, Y.; Xue, J.H.; Chen, S.; Wang, H. Learn-to-Decompose: Cascaded Decomposition Network for Cross-Domain Few-Shot Facial Expression Recognition. In Proceedings of the

European Conference on Computer Vision (ECCV), Tel Aviv, Israel, 23–27 October 2022; pp. 683–700.

29. Fan, Y.; Li, V.O.; Lam, J.C. Facial expression recognition with deeply-supervised attention network. IEEE Trans. Affect. Comput. 2020, 13, 1057–1071.
30. Jiang, J.; Deng, W. Disentangling Identity and Pose for Facial Expression Recognition. IEEE Trans. Affect. Comput. 2022, 13, 1868–1878.
31. Krithika, L.; Priya, G.L. MAFONN-EP: A minimal angular feature oriented neural network based emotion prediction system in image processing. J. King Saud-Univ.-Comput. Inf. Sci. 2022, 34, 1320–1329.

Retrieved from https://encyclopedia.pub/entry/history/show/119897