Commentary: The Dynamic Features of Lip Corners in Genuine and Posed Smiles

Yingqi Li[1], Zhongyong Shi[2,3], Honglei Zhang[4,5], Lishu Luo[4,5] and Guoxin Fan[5,6]*

[1] School of Humanity, Tongji University, Shanghai, China, [2] Psychiatry Department, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China, [3] Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States, [4] School of Management and Economics, Tianjin University, Tianjin, China, [5] Surgical Planing Lab, Radiology Department, Brigham and Women's Hospital, Boston, MA, United States, [6] School of Medicine, Tongji University, Shanghai, China

A Commentary on

The Dynamic Features of Lip Corners in Genuine and Posed Smiles
by Guo, H., Zhang, X.-H., Liang, J., and Yan, W.-J. Front. Psychol. 9:202. doi: 10.3389/fpsyg.2018.00202

For thousands of years of human history, we have learned how to fake or hide our genuine feelings and emotions to people around us intentionally or unconsciously. It is, indeed, an irony that this is what we view as emotional intelligence, and which we practice to win people over, display our politeness, tackle dilemmas, and deal with other complicated situations. Posed smiles are one of the most common faked expressions in our daily life. Indeed, it is a challenge for the computer vision system to recognize the genuine smile apart from posed smiles of an individual, and this may be difficult to interpret by humans too sometimes. Recently, an interesting work by Guo et al. (Guo et al., 2018) employed computer vision techniques to investigate the potential differences in the duration, intensity, speed, symmetry of the lip corners, and certain irregularities between genuine and posed smiles based on the UvA-NEMO Smile Database. The results are quite rewarding since they found that genuine smiles were correlated with higher onset, offset, apex, and total duration, as well as offset displacement and irregularity-b, compared with posed smiles. In addition, posed smiles were correlated with higher onset and offset speeds, irregularity-a, symmetry-a, and symmetry-d.

We cannot agree with the saying that only a handful of studies on the dynamic features of facial expressions have been conducted due to the lack of user-friendly analytic tools. On the contrary, in the past decades, hundreds of studies have focused on the dynamic features of facial expressions (Sandbach et al., 2012; Ko, 2018). Valstar et al. (2006) differentiated spontaneous brow actions from posed ones focusing on velocity, duration, and order of occurrence. Littlewort et al. (2009) distinguished fake pain from real pain by analyzing facial actions based on Gabor features. Dibeklioglu et al. (2012) analyzed the dynamics of eyelids, cheeks, and lip corners to tell genuine smiles from posed ones, and extracted 25 features, which were also cited by the author. Guo et al. (2018) said that not all these 25 features could be explained from a psychological perspective; hence, they extracted the duration, speed, intensity, symmetry, and irregularity aspects in their study. The question is why do all of the potential features need to be explained by psychological theory. It is possible that in this manner we may lose a lot of useful information to help distinguish genuine smiles from posed ones. Obviously, we still have great limited knowledge in psychology itself.

FIGURE 1 | Hybrid model combining CNN and hand-crafted dynamic features (the smile picture belongs to the first author, and informed consent was obtained from the first author).
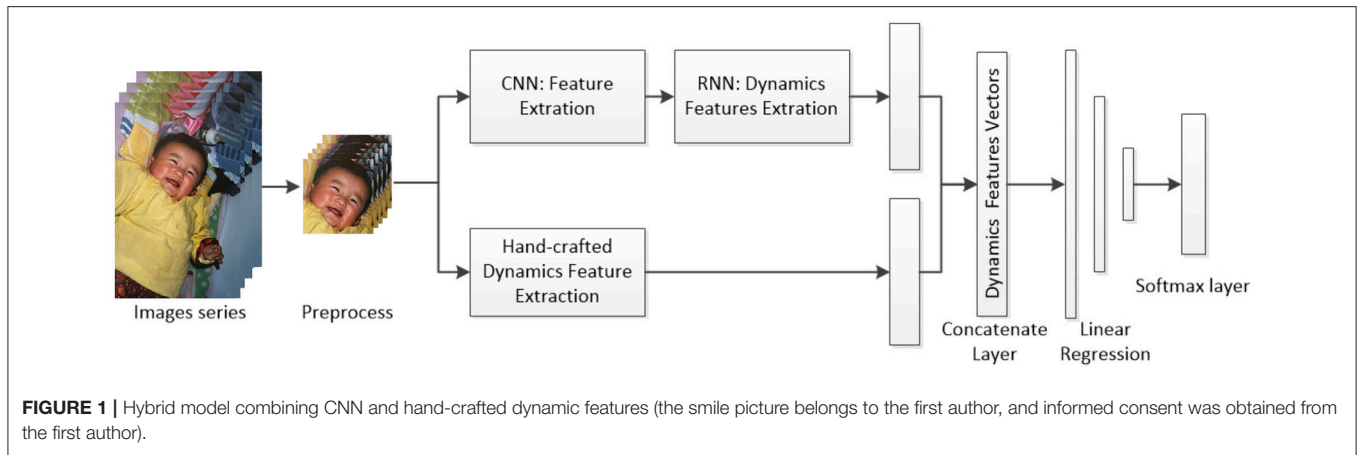
Indeed, the value of all the above-mentioned pioneering works should be appreciated, as they helped improve the recognition of posed smiles from spontaneous expressions over time. However, the hand-crafted features built by rules may lead to inadequate abstraction and representations. We are wondering whether the 25 features encompass the whole story to tell genuine smiles from posed ones, and how many of these extracted features would help the computer vision system to recognize posed smiles from genuine facial expressions. Obviously, there is still a lot of work left for us to consider and all of the features identified by different studies and extracted from different datasets need to be analyzed to help conduct the recognition performance. We cannot tell how much the dynamic features of the lip corners would help to differentiate genuine smiles from posed smiles from the diagnostic data presented in the current study.

Recently, deep learning has led to overwhelming performances in image or video processing over conventional methods such as facial recognition and classification (Peng et al., 2017; Rodriguez et al., 2017; Majumder et al., 2018; Yu et al., 2018). Many start-up companies have already built their businesses displaying outstanding performance in the field of facial recognition in security. It is not surprising that researchers have already adopted convolutional neural networking (CNN) to differentiate genuine smiles from posed ones, and the recognition performances have been promising (Kumar et al., 2017; Mandal et al., 2017). However, another question arises as to whether deep learning will take over this area and wipe out the necessity of studying hand-crafted features.

In reality the recognition performances, as consequences of deep learning in classifying the genuine smiles and posed smiles may rely heavily on the size of the training data. Unfortunately, datasets containing labeled genuine smiles and posed smiles are limited (Xu et al., 2017). However, the good news is that hand-crafted features combined with deep learning may have the potential to improve the recognition performances compared with deep learning alone supported by limited data (Pesteie

et al., 2018). It is possible to build a hybrid model by inputting features from deep learning along with well-known features obtained from conventional methods into a classifier (**Figure 1**). We admit that deep learning has also been criticized for its level of interpretability, known as the black box. However, many researchers have realized the importance of solving the problem of the black box associated with deep learning, and solutions have been proposed to tackle the same (Gunning, 2017; Samek et al., 2017; Shwartz-Ziv and Tishby, 2017).

Considering outstanding recognition performance, we do believe that deep learning will dominate the area of image recognition and classification, including discriminating genuine smiles from posed ones. As for the black box, we should regard it as an accompanying aspect of deep learning, instead of being a mere limitation. It would be better if we can solve the problem of the black box similar to how Newton figured out why apples always fell to the ground. When that day comes, deep learning will have a greater impact than it has today, though we admit that more efforts are needed to solve the problem of the black box associated with deep learning.

AUTHOR CONTRIBUTIONS

YL designed and wrote the manuscript. ZS revised the manuscript. HZ and LL gave critical comments. GF reviewed and approved the manuscript.

ACKNOWLEDGMENTS

REFERENCES

Dibeklioglu, H., Salah, A. A., and Gevers, T. (2012). "Are you really smiling at me? Spontaneous versus posed enjoyment smiles," in *Computer Vision – ECCV 2012*, eds A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid (Berlin; Heidelberg: Springer Berlin Heidelberg), 525–538.

Gunning, D. (2017). *Explainable Artificial Intelligence (xai)*, Arlington, VA: Defense Advanced Research Projects Agency (DARPA), nd Web.

Guo, H., Zhang, X. H., Liang, J., and Yan, W. J. (2018). The dynamic features of lip corners in genuine and posed smiles, *Front. Psychol.* 9:202. doi: 10.3389/fpsyg.2018.00202

Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors (Basel)* 18:E401. doi: 10.3390/s18020401

Kumar, G. A. R., Kumar, R. K., and Sanyal, G. (2017). "Discriminating real from fake smile using convolution neural network," in *2017 International Conference on Computational Intelligence in Data Science (ICCIDS)* (Chennai: IEEE), 1–6.

Littlewort, G. C., Bartlett, M. S., and Lee, K. (2009). Automatic coding of facial expressions displayed during posed and genuine pain. *Image Vis. Comput.* 27, 1797–1803. doi: 10.1016/j.imavis.2008.12.010

Majumder, A., Behera, L., and Subramanian, V. K. (2018). Automatic facial expression recognition system using deep network-based data fusion. *IEEE Trans. Cybern.* 48, 103–114. doi: 10.1109/TCYB.2016.2625419

Mandal, B., Lee, D., and Ouarti, N. (2017). "Distinguishing posed and spontaneous smiles by facial dynamics," in *Computer Vision – ACCV 2016 Workshops*, eds C.-S. Chen, J. Lu, and K.-K. Ma (Cham: Springer International Publishing), 552–566.

Peng, M., Wang, C., Chen, T., Liu, G., and Fu, X. (2017). Dual temporal scale convolutional neural network for micro-expression recognition. *Front. Psychol.* 8:1745. doi: 10.3389/fpsyg.2017.01745

Pesteie, M., Lessoway, V., Abolmaesumi, P., and Rohling, R. N. (2018). Automatic localization of the needle target for ultrasound-guided epidural injections. *IEEE Trans. Med. Imaging* 37, 81–92. doi: 10.1109/TMI.2017.2739110

Rodriguez, P., Cucurull, G., Gonalez, J., Gonfaus, J. M., Nasrollahi, K., Moeslund, T. B., et al. (2017). Deep pain: exploiting long short-term memory networks for facial expression classification. *IEEE Trans Cybern.* 1–11. doi: 10.1109/TCYB.2017.2662199

Samek, W., Wiegand, T., and Müller, K.-R. (2017). *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models.* arXiv:1708.08296 [Preprint].

Sandbach, G., Zafeiriou, S., Pantic, M., and Yin, L. (2012). Static and dynamic 3D facial expression recognition: a comprehensive survey. *Image Vis. Comput.* 30, 683–697. doi: 10.1016/j.imavis.2012.06.005

Shwartz-Ziv, R., and Tishby, N. (2017). *Opening the Black Box of Deep Neural Networks via Information.* arXiv:1703.00810 [Preprint].

Valstar, M. F., Pantic, M., Ambadar, Z., and Cohn, J. F. (2006). "Spontaneous vs. posed facial behavior: automatic analysis of brow actions," in *Proceedings of the 8th international conference on Multimodal Interfaces*, ACM, Banff, Alberta, Canada, 162–170.

Xu, C., Qin, T., Bar, Y., Wang, G., and Liu, T. Y. (2017). "Convolutional neural networks for posed and spontaneous expression recognition," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 769–774.

Yu, Z., Tan, E. L., Ni, D., Qin, J., Chen, S., Li, S., et al. (2018). A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition. *IEEE J. Biomed. Health Inform.* 22, 874–885. doi: 10.1109/JBHI.2017.27 05031