Recognizing Micro-Expressions Using Deep Neural Networks

Sergiu Cosmin Nistor [1], Gabriela Czibula [1]
[1] Department of Computer Science, Babeș-Bolyai University, Cluj-Napoca, Romania
sergiu.nistor@cs.ubbcluj.ro

Abstract

Micro-expressions are fast involuntary movements of the face that convey the emotions of people. They are hard to simulate or hide, so their recognition (when spotted) can be used as an indicative of true emotions. We propose in this paper a method for recognizing micro-expressions from high-speed video sequences. We use a deep neural network which we trained using a multi-stage approach. Our model is composed of a convolutional neural network which extracts representative features from individual frames of the sequence and a recurrent neural network which captures the evolution of the face during the video sequence. We use convolutional autoencoders for learning the most expressive facial features. We present our results on recognizing the emotion conveyed by the micro-expressions and the impact our multi-staged approach has on the performance of the network.

Keywords: micro-expression recognition; autoencoders; recurrent neural networks; convolutional neural networks
Domain: computer science
Section: Elaboration of the doctoral thesis

Motivation

Micro-expressions (MEs) are brief and subtle movements of the facial muscles. They usually last a fraction of a second. MEs are the result of deliberate or unconscious concealment. First discovered by Haggard and Isaacs [1] and independently by Paul Eckman [2], MEs are an accurate indicative of true emotion. We propose a deep learning method of classifying MEs by the emotion that they convey: positive, negative or surprise [3]. An F-score around 75% was obtained on a dataset consisting of publicly available images. Some methods for detecting and recognizing expressions take as input the entire face and try tracking the movement [4, 5], while others begin by isolating different regions of interest on the face, analyze them independently and combine the results for the final decision [6]. Both classical image processing methods like SVMs and optical flow [6] and deep learning [4, 5] were used to process expressions. For capturing spontaneous MEs, high stake situations must be created. The subjects who participated to the creation of ME datasets [7, 8, 9] were asked to look at highly emotional videos, conceal their emotions and were told they would be punished should their emotions be observed.

Methodology of Research

Our ME recognition solution uses a facial features extractor to obtain the most relevant characteristics of the face in a frame of the video. For this, we used a convolutional neural network (CNN). To enhance the ability of our extractor to observe the most relevant features, we also train it as part of a convolutional autoencoder (CAE). Autoencoders are specifically designed for feature extraction. One of their advantages is that they are self-supervised algorithms, as their output coincides with their input. They are composed of two parts: the encoder and the decoder. The encoder extracts the features and the decoder takes the features and tries to reconstruct the input from them. As the autoencoder is trained to minimize the differences between its inputs and its outputs, the features the encoder extracts become more representative. After the feature extractor processes each frame, we have a sequence of features from which to recognize the ME. For this task, we use a deep recurrent neural network (RNN).

We train our algorithm in three stages. In the first stage, we train our CAE on facial images reconstruction. Facial images are given as input and the exact same images should be obtained as output. In this way, the encoder learns to extract the most relevant features from each facial image. The second stage of our training is to specialize the encoder on facial expressions. We attach another fully-connected layer to the end of the encoder and we train it on macro-expression recognition. We consider four classes: positive, negative, surprise and neutral. The first three classes correspond to the classes considered for micro-expression recognition, while the last class allows the encoder to learn what is the difference between a facial expression that conveys an emotion and a neutral state of the face. The last stage of the training is recognizing MEs by processing the sequence of features using RNNs. The features corresponding to each frame are given to the RNN, in the order of the frames in the video. The output of the RNN after it processes the last frame features is processed by a feed-forward layer to produce the scores for each class and decide the final prediction.

Results and Comparison with State-of-the-art

As an efficient feature extractor is essential for our algorithm, we experimented with different CAE architectures and improved them until we designed a CAE that obtains a satisfactory performance. For ME recognition, we used two publicly available datasets [7, 8, 9]. We compared the results of our full algorithm with the results of three other options, each using a feature extractor that was differently trained: only on facial images reconstruction (FIR), only on macro-expression recognition (ER) or on both tasks in parallel (MIXT).

Conclusions

None of the alternative options that we considered surpassed our algorithm, demonstrating that each stage of our training is necessary for an efficient ME recognition system.

References

[1] Haggard, Ernest A., and Kenneth S. Isaacs. "Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy." In Methods of research in psy-chotherapy, pp. 154-165. Springer, Boston, MA, 1966.

[2] Ekman, Paul. Telling lies: Clues to deceit in the marketplace, politics, and mar-riage (revised edition). WW Norton & Company, 2009.

[3] Nistor, Sergiu Cosmin. "Multi-Staged Training of Deep Neural Networks for Mi-cro-Expression Recognition." In 2020 IEEE 14th International Symposium on Ap-plied Computational Intelligence and Informatics (SACI), pp. 000029-000034. IEEE, 2020.

[4] Borza, Diana, Razvan Itu, and Radu Danescu. "Micro Expression Detection and Recognition from High Speed Cameras using Convolutional Neural Networks." In VISIGRAPP (5: VISAPP), pp. 201-208. 2018.

[5] Breuer, Ran, and Ron Kimmel. "A deep learning perspective on the origin of faci-al expressions." arXiv preprint arXiv:1705.01842 (2017).

[6] Li, Xiaohong, Jun Yu, and Shu Zhan. "Spontaneous facial micro-expression de-tection based on deep learning." In 2016 IEEE 13th International Conference on Sig-nal Processing (ICSP), pp. 1130-1134. IEEE, 2016.

[7] Pfister, Tomas, Xiaobai Li, Guoying Zhao, and Matti Pietikäinen. "Recognising spontaneous facial micro-expressions." In 2011 international conference on computer vision, pp. 1449-1456. IEEE, 2011.

[8] Li, Xiaobai, Tomas Pfister, Xiaohua Huang, Guoying Zhao, and Matti Pie-tikäinen. "A spontaneous micro-expression database: Inducement, collection and baseline." In 2013 10th IEEE International Conference and Workshops on Automatic face and gesture recognition (fg), pp. 1-6. IEEE, 2013.

[9] Yan, Wen-Jing, Xiaobai Li, Su-Jing Wang, Guoying Zhao, Yong-Jin Liu, Yu-Hsin Chen, and Xiaolan Fu. "CASME II: An improved spontaneous micro-expression da-tabase and the baseline evaluation." PloS one 9, no. 1 (2014): e86041.