Computer Vision Pipeline for Facial Expression Recognition

Mid Checkpoint

Kamil Akhmetov
B17-DS-01
Innopolis University
Innopolis, Republic of Tatarstan, Russia
k.ahmetov@innopolis.ru

Vadim Solovov
B17-DS-02
Innopolis University
Innopolis, Republic of Tatarstan, Russia
v.solovov@innopolis.ru

I. Introduction

Rapid development of neural network models has made deep learning frameworks standard approach for vast amount of machine learning tasks with applications in various branches of human life. Significant part of problems require solutions that are both effective and interpretable. This includes solutions specific for crucial fields as psychology and healthcare. Neural models while having exhaustive expressiveness are weakly interpreted [4], [5], meaning users may not rely on inferences of untrusted opaque algorithms. Traditional machine learning methods are more transparent, providing comprehensible results, however they require additional stages in the learning pipeline in comparison with neural networks that may facilitate end-to-end approaches. Capturing social signals from human beings is essential in order to empathize computer systems. Important information comes from facial expressions. In our project, we aim to build a Computer Vision pipeline for Facial Expression Recognition, which includes data preprocessing, feature extraction and classification stages. One of the main goals of the project is to find most representative features, and, as a consequence, improve the existing solutions, by both accurately determining facial emotions and saving interpretability so that we could be able to reason the features that contribute to the recognition of a specific expression. A solution we expect to achieve could open up opportunities for applications in critical areas of human life, for example, sociology, driver safety, virtual reality, cognitive sciences, etc.

II. Related Work

Deep Learning is a dominant approach for machine learning in general and computer vision in particular. Convolutional neural networks (CNN) and recurrent neural networks (RNN) are the deep learning based methods used for tasks as feature extraction, classification, recognition. The main advantage of CNN is that it completely (or highly) reduces the need in preprocessing steps and enables so called 'end-to-end' learning directly from the raw inputs [6]. In order to capture temporal changes CNN were applied in combination with RNNs or long short term memory (LSTM) networks. Commonly, deep learning approaches extract features with the desired characteristics directly from the raw data by the application of deep neural networks (mostly CNN), however, they require much higher computational power, comparing with traditional methods [7]–[12]. The other and often a more crucial drawback of deep learning approaches is the lack of interpretability still leaving most researches in such fields as physics or medicine using linear models. [4], [5]

Given the drawbacks of Deep Learning approaches, traditional methods remained in the spotlight of computer vision researches. There are several preprocessing, feature extracting and classifying techniques that are both powerful and comprehensible. For geometric features, relationship between facial components, information about positions and angles of different landmarks were used for constructing a training vector [13], [14]. Classifiers used to be either AdaBoost or SVM. The appearance features were usually extracted from the global face region, using, for example, local binary pattern (LBP) histogram for feature vectors and principal component analysis (PCA) for classifications [15]. But still this approach does not reflect changes in local mini-regions, that is why appearance features were obtained from different domains with different levels of information [16], [17]. Combining the benefits of previously described approaches [17], [18], hybrid features emerged in order to complement the weaknesses and provide even better results. Many systems [18], [19] were used to construct features in temporal domain extracting the geometric and appearance features tracked later frame-by-frame. Commonly speaking, classic FER approaches determine features and classifiers by experts.

III. Idea Explanation

In order to achieve interpretable results we rely on traditional models, thus the main focus is dedicated to feature extraction pipeline. The idea of the first iteration was to test several basic approaches and implement a framework for further experiments.

A. Preprocessing

Step of preprocessing converts input data points to a representation that is suitable for further feature extraction. This stage may include such operations as resizing, denoising, segmentation and so on.

B. Feature extraction

After the data is formatted and structurized, we are ready to extract apply algorithms for feature extraction and data vectorization. Finding and tuning these algorithms are the main part of our project. Currently we have implemented Histogram of Oriented Gradients [1] and Local Binary Pattern [2].

C. Classification

Feeding crafted feature vectors to a classifer is the final stage of the facial expression recognition pipeline. AdaBoost with Decision Trees and Support Vector Classifier models are well known to solve such kind of tasks, that is why they were our first choices as classifiers.

D. Code

Code for mid-project checkpoint is available on GitHub: https://github.com/kamilkoduo/cv-fer

IV. Experiments & Results

As we have stated above, for the first iteration we have tried several configurations constructed from the previously used methods. Our start point was to test LBP and HOG algorithms in combination with AdaBoost and SVC models.

A. Dataset

There are many datasets for human facial expressions such as CK+, JAFFE, TFEID, KFEID. For the sake of our experiments we have selected a relatively small facial expression dataset: Taiwanese Facial Expression Image Database. We have selected a subset of 336 samples uniformly distributed across 20 female and 20 male models and contains 8 facial expressions (including neutral).

B. Performance evaluation

The data was split in following manner: randomly chosen 80% of data points of each expression are used for training and the remaining 20% used for testing, thus classes are uniformly distributed across both splits. Based on these data we have trained 4 configurations.

	LBP	**HOG**
SVC	0.34	0.37
AdaBoost	0.43	0.46

TABLE I: Accuracy Score

As we can see from the Table 1, while HOG feature extractor in combination with Ada-Boost classifier shows better accuracy score, it is still insufficient and alternatives need to be researched.

V. Conclusion & Future plans

To conclude, we have prepared a framework for our experiments and tested some configurations. We have seen, that improvement is needed. Future plans include finding other techniques in order to achieve better results.

References

[1] McConnell, R K. Tue . "Method of and apparatus for pattern recognition". United States.

[2] T. Ojala, M. Pietikainen and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," Proceedings of 12th International Conference on Pattern Recognition, Jerusalem, Israel, 1994, pp. 582-585 vol.1, doi: 10.1109/ICPR.1994.576366.

[3] L. F. Chen and Y. S. Yen, "Taiwanese facial expression image database," Ph.D. dissertation, Brain Mapping Lab., Inst. Brain Sci., Nat. Yang-Ming Univ., Taipei, Taiwan, 2007

[4] Irene Sturm, Sebastian Lapuschkin, Wojciech Samek, Klaus-Robert Mddotuller, Interpretable Deep Neural Networks for Single-Trial EEG Classification, ¡![CDATA[Journal of Neuroscience Methods]]¿ (2016), http://dx.doi.org/10.1016/j.jneumeth.2016.10.008

[5] Dong, Yinpeng Bao, Fan Su, Hang Zhu, Jun. (2019). Towards Interpretable Deep Neural Networks by Leveraging Adversarial Examples.

[6] Walecki, R.; Rudovic, O.; Pavlovic, V.; Schuller, B.; Pantic, M. Deep structured learning for facial action unit intensity estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, HI, USA, [2017]; pp. 3405–3414

[7] Kahou, S.E.; Michalski, V.; Konda, K. Recurrent neural networks for emotion recognition in video.In Proceedings of the ACM on International Conference on Multimodal Interaction, Seattle, WD, USA,9–13 November 2015; pp. 467–474.

[8] Kim, D.H.; Baddar, W.; Jang, J.; Ro, Y.M. Multi-objective based Spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. IEEE Trans. Affect. Comput. [2017],PP.

[9] Chu, W.S.; Torre, F.D.; Cohn, J.F. Learning spatial and temporal cues for multi-label facial action unit detection. In Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, [2017]; pp. 1–8.54.

[10] Hasani, B.; Mahoor, M.H. Facial expression recognition using enhanced deep 3D convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Hawaii, HI, USA, [2017]; pp. 1–11.55.

[11] Graves, A.; Mayer, C.; Wimmer, M.; Schmidhuber, J.; Radig, B. Facial expression recognition with recurrent neural networks. In Proceedings of the International Workshop on Cognition for Technical Systems,Santorini, Greece, 6–7 October 2008; pp. 1–6.56.

[12] Jain, D.K.; Zhang, Z.; Huang, K. Multi angle optimal pattern-based deep learning for automatic facial expression recognition. Pattern Recognit. Lett. 2017,1, 1–9. [CrossRef]57.

[13] Suk, M.; Prabhakaran, B. Real-time mobile facial expression recognition system—A case study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, [2014]; pp. 132–137.23.

[14] Ghimire, D.; Lee, J. Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines. Sensors 2013,13, 7714–7734

[15] Happy, S.L.; George, A.; Routray, A. A real time facial expression classification system using local binary patterns. In Proceedings of the 4th International Conference on Intelligent Human Computer Interaction,

[16] Khan, R.A.; Meyer, A.; Konik, H.; Bouakaz, S. Framework for reliable, real-time facial expression recognition for low resolution images. Pattern Recognit. Lett. [2013], 34, 1159–1168.

[17] Ghimire, D.; Jeong, S.; Lee, J.; Park, S.H. Facial expression recognition based on local region specific features and support vector machines. Multimed. Tools Appl. [2017], 76, 7803–7821.

[18] Benitez-Quiroz, C.F.; Srinivasan, R.; Martinez, A.M. EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, [2016]; pp. 5562–5570.

[19] Torre, F.D.; Chu, W.-S.; Xiong, X.; Vicente, F.; Ding, X.; Cohn, J. IntraFace. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Ljubljana, Slovenia, [2015]; pp. 1–8