Emotion recognition

Calvin GALAGAIN
ENS Paris Saclay - MVA
Calvin.galagain@ens-paris-saclay.fr

Abstract

Computer vision is now able to analyze a lot of information, especially that related to facial detection. The human face has more than forty muscles that generate a wealth of expressions that can be communicated. For a human being, this understanding is innate and trying to make algorithms that reproduce it is a real challenge.

1. Introduction

In the quest to learn tasks that humans know how to do naturally, this subject has many implications. Today to answer questionnaires which aim to know our feelings during a video, we must complete everything manually and, it is very long. One use could be to incorporate a sentiment recognition system on computer cameras to track changes in facial expressions live. This could also be useful to help people with illnesses. Autistic or blind people find it difficult to read other people's facial expressions, they would need technologies like the one we are going to seek to develop. Finally, the robotics world could also be very fond of a system that can read emotions because it could interact more intelligently with its users. To go further, we could try to read the expressions of a crowd or a driver to detect dangerous situations that could be avoided with prevention. Finally, in a more recent context, many people wear masks and the face is partially hidden. It is therefore more difficult to read the expressions that the person is feeling. It is interesting to see if algorithms could overcome this barrier.

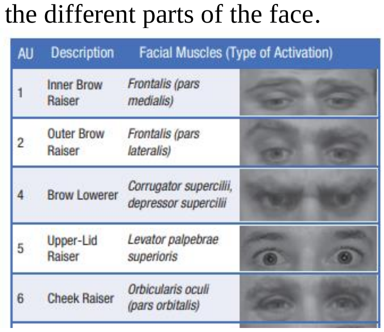
The problem is therefore very interesting because it has questioned many scientists since the 90s. But with the appearance of many databases, the problems have changed. Indeed, it is possible to have almost perfect results in controlled situations. If the subject is in optimal lighting, with an optimal background and conditions exactly the same as the other people in the database, the results are more than excellent. The real problem appears when we try to analyze faces in real situations “in the wild”. Sentiment detection promises to bring many applications that would be able to analyze our emotional state and improve our health or safety if these applications do not infringe on our freedoms.

2. Methodology

This work will aim to compare different Machine Learning methods in the service of image analysis. For this, the first step will be to find databases that provide data that are in real situations. We will first analyze RAF Face [2], EmotioNet [8] and AffectNet [1] which allow us to explore the problem very broadly. By creating a usable and processed database for our use, we can then compare different methods that exist in the state of the art. Some papers make it possible to realize the old methods that work or not. It would then be interesting to try new algorithms such as transformers [5] as we can this on recent paper [9].

A first objective may be to visualize the areas of attention on a face to detect emotions. We will use Transformers like a *vit_h_14* and pretrained weights to see if results can be usable.

And as shown in the article [10], one idea would be to implement a tailor-made attention mechanism where we could extract the different parts of the face.



AU	Description	Facial Muscles (Type of Activation)
1	Inner Brow Raiser	*Frontalis (pars medialis)*
2	Outer Brow Raiser	*Frontalis (pars lateralis)*
4	Brow Lowerer	*Corrugator supercilii, depressor supercilii*
5	Upper-Lid Raiser	*Levator palpebrae superioris*
6	Cheek Raiser	*Orbicularis oculi (pars orbitalis)*

Thus, with algorithms such as One vs All SVM, Fine-tuning from VGG, Fine-tuning from ResNet-50, personal algorithms from Neural Network or Transformers, this work will be able to compare new methods with previous work [3]. By comparing many databases, this work will be rich in proposals in the processing of data, in particular to seek to treat the biases that change the expression of feelings [6][7]. For example, works explain that the origin of people shows expressions that do not show the same on the faces [4]. We can try to compare the results by dividing the database according to the origin of the people.

It is possible to improve state-of-the-art models such as ResNet-18, VGG-Face, MS-Celeb-1M, MA-Net, ESR on AffectNet, FER+ and RAF-DB [11]. be used with unsupervised learning techniques to improve performance [12].

3. Evaluation

In view of the different databases that we will use, we will be led to classify emotions. The most common are these: Neutral, Happy, Sad, Surprise, Fear, Disgust, Anger (Contempt). By comparing the results with databases on faces in controlled situations, we can see that in real situations the problem is different.

By comparing all these methods, we can then create a cross-validation that will analyze the overall performance. The goal will be to understand the predictions of our models according to the labels present in our databases.

By processing our database, we will simplify the problem so that our algorithms generalize more easily. We will use landmark models to align the faces and make all faces have the same properties.

These results can be compared with the implications of the problem on what is likely to be feasible or not.

References

[1] A. Mollahosseini, B. Hasani, M. H. Mahoor, "**AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild"**, Arxiv 2017

[2] Shan Li, Weihong Deng, JunPing Du, "**Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild"**, IEEE 2017

[3] Vasavi Gajarla, Aditi Gupta, "**Emotion Detection and Sentiment Analysis of Images"**

[4] José Angel Soto, Robert W. Levenson, "**Emotion Recognition across Cultures: The Influence of Ethnicity on Empathic Accuracy and Physiological Linkage"**, Emotion. 2009

[5] Saptashwa Bhattacharyya, "**Understand and Implement Vision Transformer with TensorFlow 2.0"**, towardsdatascience 2022

[6] Jiabei Zeng, Shiguang Shan, Xilin Chen, **"Facial Expression Recognition with Inconsistently Annotated Datasets"**, ECCV 2018

[7] Yunliang Chen, Jungseock Joo, **"Understanding and Mitigating Annotation Bias in Facial Expression Recognition"**, ICCV 2021

[8] Fabian Benitez-Quiroz, C., Ramprakash Srinivasan, and Aleix M. Martinez. **"EmotioNet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild"**, IEEE 2016

[9] Fanglei Xue, Qiangchang Wang, Guodong Guo, **"TransFER: Learning Relation-aware Facial Expression Representations with Transformers",** Arxiv 2021

[10] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak, "**Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements"**, aps 2019

[11] Negar Heidari, Alexandros Iosifidis, **"Learning Diversified Feature Representations for Facial Expression Recognition in the Wild"**, Arxiv 2022

[12] Shuvendu Roy, Ali Etemad, **"Analysis of Semi-Supervised Methods for Facial Expression Recognition"**, Arxiv 2022