Face expressions understanding by geometrical characterization of deep human faces representation

Adrien Raison, Théo Biardeau, Pascal Bourdon, David Helbert

Face expressions understanding by geometrical characterization of deep human faces representation

Adrien Raison[a], Théo Biardeau[a], Pascal Bourdon[a], David Helbert[a]

[a]*XLIM-ASALI,CNRS U-7252, University of Poitiers, Poitiers, 86000, France*

Abstract

Face expressions understanding is a key to have a better understanding of the human nature. In this contribution we propose an end-to-end pipeline that takes color images as inputs and produces a semantic graph that encodes numerically what are facial emotions. This approach levrages low-level geometric details as face representation which are numerical representations of facial muscles activation patterns to build this emotional understanding. It shows that our method recovers social expectations of what characterize facial emotions.

1. Introduction

Face expressions are one of the main human communication channel. They are resultant of many complex human internal processes that themselves have been triggered by external stimulis. Understanding how human emotions work is a continuously investigated research topic and results favor both academic (e.g., human psychological understanding improvement) and industrial (e.g., advertising targeting) developements. In this study we propose a new framework allowing to understand visually what are facial emotions; in which manner they are recognizable from others; in other words, what are defining them intrinsically from a numerical point of view. Our approach relies on embedding human faces as graphs where nodes represent both semantic and positional landmarks. The spatial distribution of those

Email addresses: adrien.raison@univ-poitiers.fr (Adrien Raison), theo.biardeau@etu.univ-poitiers.fr (Théo Biardeau), pascal.bourdon@univ-poitiers.fr (Pascal Bourdon), david.helbert@univ-poitiers.fr (David Helbert)

landmarks is highly related to facial emotion revealing procedure. We firstly classify those graphs with a semantic-aware Graph Neural Network (GNN) model. We then provide a qualitative study that explains accurate internal decision processes the classifier has designed for it, which in this context, is a proxy for emotions numerical definition framework.

2. Related work

Face expression characterization is foremost an emotion recognition dependent problem. Face expression recognition problems have been deeply investigated according to different angles such as natural language processing paradigm [1]. Other studies has been conducted with EEG signals [2] that allow to have a deep notion of what are emotion from a neurological point of view. Other studies had took interest in human emotions characterization namely [3, 4] and a graph-based study has been done by [5]. Understanding facial emotions from a visual viewpoint, on a side, deals with facial alignment. Facial alignment is at the core interest of computer vision community. It consists of specifying facial landmarks position regarding specific semantics. Some relevant solutions have been proposed by the computer vision community leading to a wide deployment in industrial applications although it remains an important research topic still investigated. There are non-linear correlated phenomenons between facial landmarks groundthruth distribution and facial emotion revealing process. In this study we propose an end-to-end framework that fisrt, encodes human faces present in color images as graphs; then explain deep face representations classification yielding numerical definition of what are human facial emotions. Many explaining methods suited for GNN has been proposed in the litterature. XGNN [6] is a model-level approach that generates iteratively explaining graphs through a reinforcement learning procedure. GNNExplainer [7] is a mask generator model based on mutual information optimization. It starts with randomly initialized node and node features mask jointly optimized, with mutual information, against the class label of the assessed graph. LRP-GNN [8] adopts a walk-based approach, introducing the node anteriority and apply the original Layer-wise Relevance Propagation (LRP) [9] propagation rule. EiX-GNN [10] is a model-agnostic approach that is ordering subgraphs by determining the asymtotic behavior of a random walk that is performs over the explained graphs. Nonetheless, explaining is often a context-dependant task. In order to get rid of any contextualized-expertise to assess these methods in term of

relevance, objectives metrics [11] has been designed and have helped us to choose our explaining methods. Despites all these considerations, it turns out that our pipeline allows to recover social expectations of what are facial emotions recognition.

3. Problem formulation

Emotion is a human multi-factorial response to environemental interaction. For human-to-human communication, emotions help to design exchanged messages. Expressing emotions can be made by different channels. One of them is to put face configuration in a specific spatial configuration that is socially recognizable. From this, revealing an emotion is a transfer process between an initial face shape to an other specific one. The resulting pose can be defined as being what is the expressed emotion. For instance in 2-persons conversation, expressing happiness is done most of the time by smiling and uprising cheeks of one of the involved person. Noticing this change helps to visually perceive hapiness for the conversation partner, the emotional message is thus received if both individuals share the same facial emotion revealing process. Under computer-aided approach, understanding and characterizing human face emotions can be address by computationally understand the underlying conditionned facial geometry and especially given these precise configurations. We present here our proposed approach providing relevant results even on a wide range of social contexts.

4. Our approach

Face shape deformation is due to a joint facial muscle activation, some patterns occur and some of them are basis components for facial emotion revealing. Catching such patterns and characterizing them computationnally is the key to provide a computationnal understanding of facial emotions. Encoding such patterns require to define them physically and transcript it in computing machine. Faces are complex biological structure but we can represent facial muscle groups as semantic positional landmarks. To cover this issue, we use state-of-art face alignment method that has been especially designed for such semantic mapping. The landmark spatiallity variation is due to the adjacent facial muscles that squeeze or stretch conjointly in some face neighborhoods. Acquiring this relational knowledge, in particular when emotions are expressed, can be done by representing faces as a graph where

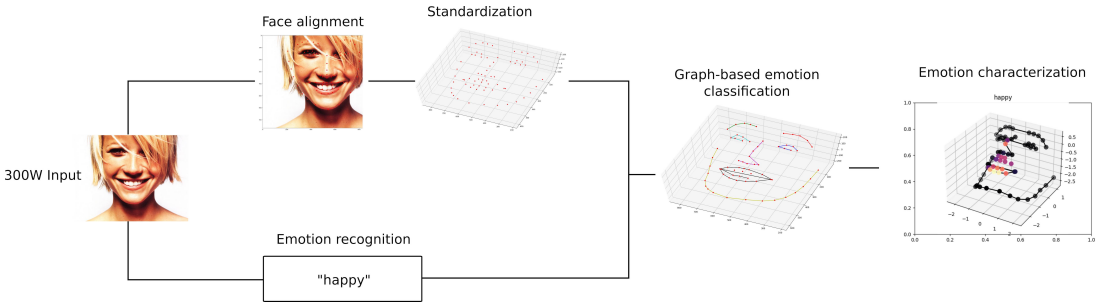


Figure 1: Our method relies on three compounds: a labeling module that is used to build our graph-fashion face emotional datasets; once this assembling is done, a deep model will classify our graph by parcellating in human-intractable high-dimensional space standardized graph vertices spatial positions; a conversion is then operated to put this complex representation into a human readable that provide an understanding of human facial emotions thank to the ScoreCAM GNN explainer.

each node is a facial landmark and where adjacency is represented by both intra-muscles (local) and inter muscles (global) scale interactions. We will describe our image-based facial graph-embedding process suited for various contexts; introducing our emotional supervised classification problem and its interpretability aspect. This last will let us highlight what are key insights to understand facial emotions at a visual scope.

4.1. Features conception and labeling

These raw materials are the inputs of our preprocessing step : image-based facial landmark detection and emotion recognition. For the sake of clarity, we denote D as being our images dataset. The dataset set size is denoted by $|D| \in \mathbb{N}$ and we see an image of size $k, l \in \mathbb{N}$ as a triplet $\mathbf{X} \in \mathbb{M}_{k,l}([0, 255] \cap \mathbb{N})^3$.

Facial landmark inference. From raw color images we determine graph features to learn thank to a face alignment method. The development of such method is at a core interest of computer vision community and detecting facial landmarks is a well-studied problem. Basically, assuming that we have an optimal image-based facial landmark detector $\texttt{fa}^\star$ that determined carefully $p \in \mathbb{N}$ three-dimensional facial landmarks, i.e. :

$$\begin{array}{rccc} \texttt{fa}^\star : & D & \to & \mathbb{R}^{3\times p} \\ & \mathbf{X} & \mapsto & (\mathbf{x}_i)_{i\in\{1,\cdots,p\}} \end{array}$$

Since each instance $\mathbf{X}$ of D has its own recording context, the inference $\texttt{fa}^\star(\mathbf{X})$ lean on significant different spatially distributed subspace of $\mathbb{R}^3$ notably in terms of space-barycentric position or overall scaling ratio. Without a standardization procedure this will lead to an irrelevant statistical learning because intrinsecal emotion signal is only a in-between relative landmarks interaction. To gain statistical stability, we apply a standardization process relying on finding optimal rigid affine transformations. We will consider here the scale, rotation and translation transformation since this is these three affine transformations that overcome the wide-context image recording issue. This standardization puts $\texttt{fa}^\star(D)$ in a common, sustainable and statistically efficient dataset representation. This framework is defined regarding a mean face that stands as a regressed and standardized objective. This mean face is emotionless, synthetic, and has been designed by averaging spatially p landmarks of neutral aligned faces. It means also that neutral emotion will be our baseline to characterize other facial emotions. We denote this mean face has $(\mathbf{m}_i)_{i\in\{1,\cdots,p\}} \in \mathbb{R}^{3\times p}$.
The standardization process is formalized as follow; for all $(\mathbf{x}_i)_i \in \texttt{fa}^\star(D)$:

$$(s^\star_{\mathbf{X}}, \mathbf{R}^\star_{\mathbf{X}}, \mathbf{t}^\star_{\mathbf{X}}) = \underset{(s,\mathbf{R},\mathbf{t}^T)\in\Omega}{\arg\min} \sum_{i=1}^{p} \|\mathbf{m}_i^T - s\mathbf{R}\mathbf{x}_i^T - \mathbf{t}^T\|_2 \tag{1}$$

with $\Omega = \mathbb{R}^+ \times SO(3) \times \mathbb{R}^3$.
Note that those rigid transformations follow the data distribution prior. Indeed, due to the various contexts present in D, head poses are various as well. Considered geometric transformations must not alter the global biological structure of human faces which is, without abusive considerations, assumed to be constant across any individuals. And since affine transformations preserve notably colinearity, parallelism, ratio of length and barycenters, it conserves the global physical faces aspect and does not introduce any semantic alterations or non-linear effects. This problem (1) has a closed-form solution where proof is given in [12]. Due to face global face shape variation over individuals in D, non-linear variation (i.e individual singularity) is not taken into account but can be seen as noisy features that may not affect meaningfully the learning phase. We denote the standardized dataset as $D^\star$, it means:

$$D^\star = \{s^\star_{\mathbf{X}}\mathbf{R}^\star_{\mathbf{X}}\mathbf{X}^T - \mathbf{t}^{\star T}_{\mathbf{X}} | \mathbf{X} \in \texttt{fa}^\star(D)\} \tag{2}$$

Then we embed this new data standardized representation as a graph according the following scheme. From original definition, graphs are couple (V, E)

where V is the set of nodes and $E \subset V \times V$ is the set of edges. The set E is the adjacency representation of the considered graph. There is a matrix formulation of edges representation that is fully equivalent and allowing to perform easieer algebraic operation. The adjacency matrix $\mathbf{A} \in \mathbb{M}_{|V|}(\{0,1\})$ as its entry $a_{i,j}$ defined by $a_{i,j} = 1$ iff $(i,j) \in E$, 0 otherwise. In the context of deep representation of signals evolving over the graph, graphs are rather seen as $(\mathbf{L}, \mathbf{A})$ such that $\mathbf{L}$ is a column vector of size $|V|$ valued in $\mathcal{H}$, a d-dimensional Hilbert space ($d \in \mathbb{N}$). It thus means that each node in V as an $\mathcal{H}$-valued row vector of size d representing the signal evolving on this node. The adjacency matrix $\mathbf{A}$ is used to describe the domain structure.

We map bijectively $D^\star$ to $D^\star_G = \{G_{\hat{\mathbf{X}}} | \hat{\mathbf{X}} \in D^\star\}$ where for each $\hat{\mathbf{X}} \in D^\star$, $G_{\hat{\mathbf{X}}} = (\hat{\mathbf{X}}, \mathbf{A}) \in D^\star \times \{\mathbf{A}\}$. Note that $\mathbf{A}$ does not depend on $\hat{\mathbf{X}}$ because the topology of induced graphs (i.e., the face muscle interaction adjacency) is a common feature shared across all humans being, indeed as mentioned before giving a social context and an emotion, peoples express it according to the same joint muscle activation patterns. The design of $\mathbf{A}$ has been driven by the muscle momentum connexity assumption. The muscle momentum connexity assumption is based upon the fact that in every face localities, physically connected muscles act together as a group (i.e. given a muscle, when it is activating, adjacent muscles are more likely to be driven by this activation and acting in turn). And so, we assume that this asumption holds to reveal facial emotion.

We designed facial muscles interactivity according an hand-crafted strategy built upon local and global semantic informations. Local adjacency is conceived according to the muscle momentum connexity assumption. At the local scope, landmarks, that represent the same positional semantic, are represented as a chained subgraph. Global adjacency is focused on gathering those local semantics in an prior-free representation, so as a complete adjacency structure according to the graph theory terminology. This strategy allows to embed efficiently both fine-details semantics (local scale) and higher-level semantics (global scale), including biological prior knowledge. The careful mixing of these connex components (according to the input graph) has a convinient algebraic representation. Indeed, the general adjacency matrix $\mathbf{A}$ is a block matrix designed respectively to respect the aforementioned $\{1, \cdots, p\}$ semantic parcellation. Landmarks semantic splitting mapping is described in Table 1.

Semantic features	Node index
Left eyebrow	17,18,19,20,21
Right eyebrow	22,23,24,25,26
Left eye	36,37,38,39,40,41
Right eye	42,43,44,45,46,47
Nose	27,28,29,30,31,32,33,34,35
Thin & cheeks	0,1,2,3,4,5,6,7,8,9,10,11,...,16
Upper mouth	48,49,50,51,52,53,54,64
Lower mouth	48,60,59,58,57,56,55,54,64,65,66,67

Table 1: Semantic face parts - Nodes index mapping

Face expression recognizer inference. As well as facial landmarks detection method, emotion recognition based on color images has been heavily investigated by the computer vision community. Many studies have been conducted and emotion recognition problems are often framed as supervised classification problems since nowadays many data are publicly available. Considering $C \in \mathbb{N}$ different emotions, an emotions recognizer is simply a optimal mapping $\texttt{er}^\star$ such that:

$$\begin{array}{rccc} \texttt{er}^\star: & D & \rightarrow & \{1, \cdots, C\} \\ & \mathbf{X} & \mapsto & c_{\mathbf{X}} \end{array}$$

We then infer $\texttt{er}^\star(D)$ to obtain labels that will be afterward used to classify emotions with respect to $D_G^\star \times \texttt{er}^\star(D)$, i.e., in a graph-based fashion rather than in image-based setting.

Note: Inferences made through $\texttt{fa}^\star$ and $\texttt{er}^\star$ may not be absolutely accurate in terms of, respectively, emotion and facial landmark groundthruths. The $D_G^\star \times \texttt{er}^\star(D)$ quality is highly dependent on $\texttt{fa}^\star$ and $\texttt{er}^\star$ accuracy. It has to be noticed that wrong assignment may lead to irrelevant results and may bias the conclusion of this study.

4.2. Feature classification

In order to acquire high understanding of facial emotions and to encode efficiently the joint landmarks relative position distribution given an emotion, we levrage the powerful data representation abilities that deep classifier

have. As a consequence, we used a deep model to classify emotions, especially a graph neural networks. Graph Neural Network is a general deep model introduced in [13] that is able to deal with data represented as graph. Contemporain GNN models has been proposed [14, 15]. We have used such models to perform our classification task. Now, we introduce $\theta \in \Theta$ such that Θ is an m-dimensional Euclidean space. We denote by f_θ our parametrized GNN classifier that determines for each instance the discrete conditional probability distribution given θ and the instanced graph $G_{\hat{\mathbf{X}}}$ over these C classes. The optimal parameter $\theta^\star \in \Theta$ exists and determines what is an optimal classifier that we denote $f_{\theta^\star}$.

4.3. Feature understanding

Once having determined $\theta^\star$, we can analyze how $f_{\theta^\star}$ internal decision processes have been designed. The classifier $f_{\theta_\star}$ combine non-linearly many different scales of input representations. Diving into such tortuous mixing is hopeless in order to have a human-understandable representation of this mixing. We rather use a method that seeks to highlight in a human-affordable and precise manner this representation combining. Explaining methods do not have a formalized formulation since what they are supplying are still under investigation from psychological, philosophical or even computional aspects.

5. Experiment

In this section we provide our actual framework from the dataset we have used to labelazing tools as well as our implementation setup that justify the results shown below.

5.1. Dataset

For a wide deployment purpose of our method, we use publicly and highly available RGB images coming from famous computer vision datasets.

300W. dataset [16, 17, 18] is widely used computer vision datasets for facial landmarks problems. This dataset is interesting since it supplies a very large contextual example including a wide range of emotions. This variety of data covers the majority of human emotions that a person using our tool can encounter in daily life. This dataset is composed of respectively 300 indoor-contextualized and 300 outdoor-contextualized instances. It is from this dataset that we have inferred three-dimensional landmarks and emotions.

5.2. Face alignment tools

As a three-dimensional landmark detector `fa`* we used [19]. This model is one of state-of-the-art 2D-to-3D landmark detector and it detects $p = 68$ facial landmarks. It relies on the Face Alignment Network. Several methods (e.g [20]) have been designed to directly find three-dimensional landmarks from raw images. . But for accuracy concerns, regressing models in two-dimensional space is easier from optimization point of view than in three-dimensional ones because the searching space is in the lower dimension. So a 2D-to-3D models that just extend the last dimension based on an already accurate 2D position is preferable to a 3D regression model that is trained from scratch.

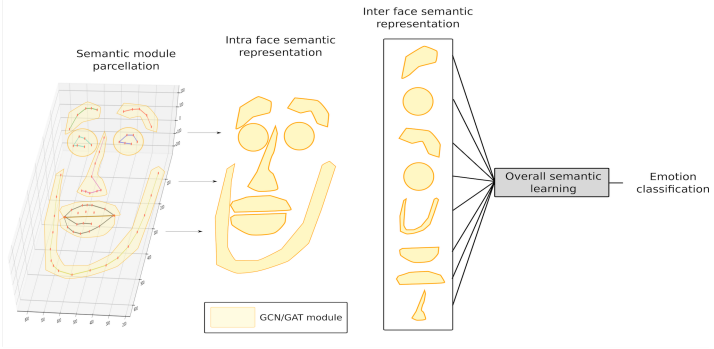
5.3. Face expression recognizer tools

As an emotion recognizer tool `er`* we have chosen the state-of-the-art model [4]. It achieves 76.82% of accuracy on FER2013 dataset which is the state-of-art dataset for such classification task. It has been trained to detect 7 emotion types : angry, fear, disgust, happy, sad, surprise and neutral. This method is using a segmentation network to refine feature maps that are then plugged into a Deep Residual Network and a U-Net based architecture.

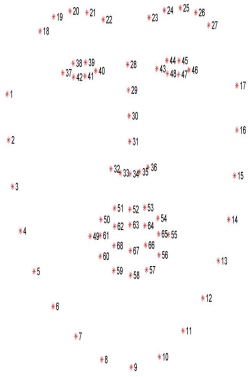
As far as we know, both tools achieved respectfully state-of-the-art results in their own domain. That is why we have chosen them to lead this study. We now provide our own classifier architecture that helps us to classify our graph representation regarding those $C = 7$ emotion types.

5.4. Classifier architecture and implementation setup

The architecture of f_θ (Figure 2a) is designed upon a local encoding that focuses on understanding each connex components and a global setting that combines, without any prior, these local information flows that afterward feed a linear layer allowing the weighting of each preprocessed semantics (with the concatenation of global average pooling and global maximum pooling layers) for the emotion classification. Our experimental setup is defined as follows : for our deep classifier we have used GCN modules for each semantic component encoding and a GCN for the subsequent global encoding. A linear module large of 512 neurons is used for the classification phase. We have used the Adam optimizer with a learning rate of 7×10^{-4}. From a hardware point, we have used Nvidia A100 40 Gb GPUs.



(a) Graph-based classifier structure: We designed the signal propaging scheme by concentrate semantic localities between themselves. Each semantic part is seen as a connex component of the assessed graph. As long as each component encodes a partial emotion information, we gather each of them in a semantic graph with eight components that are connected in an unconstrained manner, it means that it is a complete graph. The obtained embedded data distribution is then fed to a GCN module and a linear part that helps to classifying.



(b) The initial semantic node indexation that we used for our experimentation. This indexation is widely used for tackling face landmark assignment problems.

	Infidelity	Sparsity
GNNExplainer	1.67	0.23
LRP-GNN	0.79	0.14
ScoreCAM GNN	**0.21**	**0.89**

Table 2: Statistical comparative study of explaining methods: left column is the averaged infidelity over the dataset 300W given each assessed explaining methods. The infidelity [11] measures the unfaithfulness of a explaining method regarding an instance and a deep model. The lowest the infidelity, the better. We have measure as well the average sparsity that provided explanations actually are regarding each explaining method. Sparsity is another objective metrics that measuring the conciseness of each provided explanation. The higher the sparsity, the better. According both objective metrics, ScoreCAM GNN has shown empirically better results.

6. Results

In this section, we firstly provide the performances we have reached for classifying emotions within our graph-based approach. Then, we will illustrate we are our results regarding the emotion singularization obtained through ScoreCAM GNN which has been proven to be the most relevant regarding objective assessment metrics.

6.1. Classification performances

For the classification task label distribution needs to be balanced for an efficient learning. It turns out that emotions were not uniformly distributed over instances of 300W. Consequently, we had a focus only on classes which occur at least at 10% among 300W instances. Those emotions are happy, neutral, sad and fear, it thus means that $C = 4$. We also note some emotion misattributions by $\texttt{er}^{\star}$. Under our experimental setup, we reach 66% of accuracy on the graph-variant of 300W dataset. This result remains acceptable considering the misattribution problem and general classification model accuracy that surround 80% of accuracy. Based on these results, we now provide some qualitative emotion characterization measurements.

6.2. Relative explaining methods assessment

In order to supply relevant results as possible, we have leaded a statistical comparative study among state-of-the-art explaining methods. In this study we retain GNNExplainer, the state-of-the-art method, LRP-GNN and ScoreCAM GNN that provide explanations under realistic time amount with respect to XGNN or SubgraphX. As shown in Table 2, we have measured some

explaining methods objective assessment metrics (faithfulness and sparsity) and ScoreCAM GNN seems to be the most relevant explaining method. As a consequence, we will only considering ScoreCAM GNN explanations to lead our emotional understanding study. ScoreCAM GNN [21] is an extension to non-Euclidean domain of ScoreCAM, initially introduced by [22]. ScoreCAM linearly combine the highest level of data representation and weight each of them by its own contribution weight relatively to the classification task. The outcome of ScoreCAM GNN is, given a classifier and an instance, a normalized distribution of these contributions for classifying the instance, of each element of the domain the instance leans on (i.e. nodes). In our context, it provided the impact of each node regarding the classification of the considered graph according to $f_{\theta^\star}$. In other terms, it provides the importance of each landmark which has an almost one-to-one physical mapping (facial muscle neighborhood) involved in emotion description, so a deep facial emotion understanding.

6.3. Face expression understanding

Understanding deep model behaviors may be dependent on model accuracy since model accuracy reflects the numerical understanding of the model learning tasks. Under the above considerations, some results may be a bit irrelevant but it appears experimentally that provided explanation of $f_{\theta^\star}$ behaviors are in accordance with the social baseline of what an emotion is with respect to other emotions thus what is characterizing facial emotions visually. As shown in Figure 3, we have displayed the four emotions characterization with their initial image representation. For classification task, single instance explaining method always provided their explanation with including what is the relevant information (current class understanding) and what is not relevant (remaining classes) since classification are seamlessly a space partition problem. We firstly described what is the neutral emotion because it can be seen as a baseline for understanding and characterizing other emotions. Indeed we can see neutral emotion as a non-emotional emotion. Other emotions characterization will be then supplied in a constrastive fashion, with as a baseline, the neutral emotion that is geometrically unambiguous. For happy emotion, what we noticed under what ScoreCAM GNN revealed is that upper, lower mouth and nose are conjointly involved in the raw description of what happiness is over a human face. Analogously, sadness emotion can be seen as the symmetric of happiness and we found out that recognizing sadness from the geometrical point of view is also done by a specific mouth

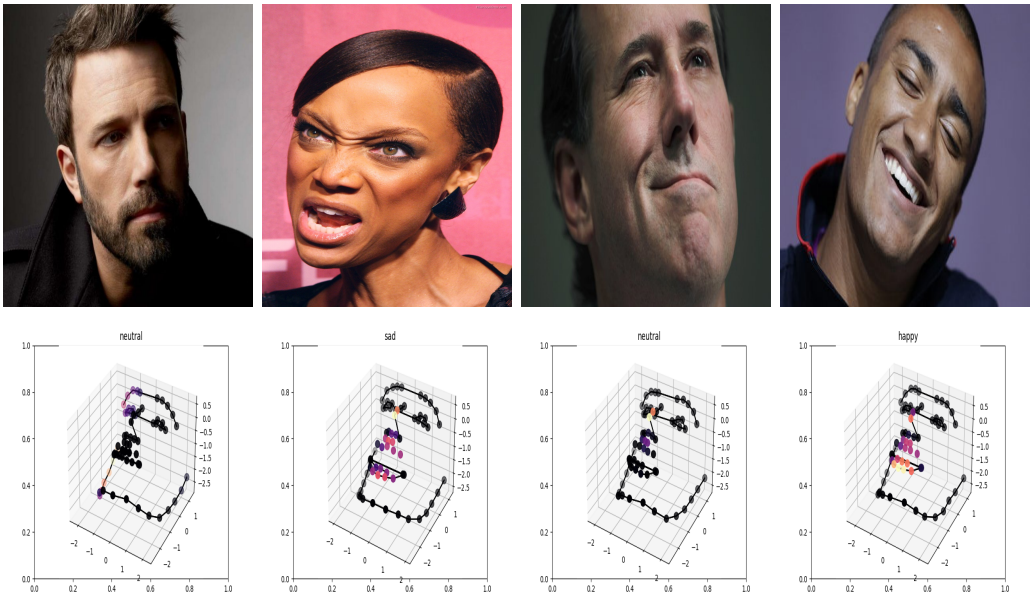


Figure 3: Face expression characterization through ScoreCAM GNN: In these four dual representation of human faces we highlight where is the numerical definition of facial emotions. Neutral emotion is the baseline emotion to characterize other ones. It shows that happiness and sadness are both highly embedded in mouth and nose motions which are socially relevant.

and nose muscle configuration that is far different from the happiness configuration. These results are in social accordance of what we can expect from the response of what characterizing facial emotions.

7. Conclusion

In this study we propose an original approach to characterize visually what are human face emotions. By using, state-of-the-art methods to recover high-level information, such as image-based emotion recognition or semantic facial landmarks positioning, we encode human faces as a graph to enforce and leverage the relational aspects of muscle activation patters involved in human emotions revealing processes. Even with uncomplete information due mainly to self-weakness of labeling methods, that stands out as being state-of-the-art method in their own field of application, we recover social expected results of what are emotions and which muscle group are involved to visually described them. Further works may include more accurate labeling methods in order to increase the overall accuracy of the graph based classifier. Other explaining methods may be used to provide emotional understandings.

References

[1] Z. Lian, J. Tao, B. Liu, J. Huang, Z. Yang, R. Li, Context-Dependent Domain Adversarial Neural Network for Multimodal Emotion Recognition, in: Interspeech 2020, ISCA, 2020, pp. 394–398. doi:10.21437/Interspeech.2020-1705.
URL https://www.isca-speech.org/archive/interspeech_2020/lian20b_interspeech.html

[2] D. O. Bos, EEG-based Emotion Recognition 18.

[3] M. A. Neerincx, J. van der Waa, F. Kaptein, J. van Diggelen, Using Perceptual and Cognitive Explanations for Enhanced Human-Agent Team Performance, in: Engineering Psychology and Cognitive Ergonomics: 15th International Conference, EPCE 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Springer-Verlag, Berlin, Heidelberg, 2018, pp. 204–214. doi:10.1007/978-3-319-91122-9_18.
URL https://doi.org/10.1007/978-3-319-91122-9_18

[4] L. Pham, T. H. Vu, T. A. Tran, Facial Expression Recognition Using Residual Masking Network, in: 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 4513–4519, iSSN: 1051-4651. doi:10.1109/ICPR48806.2021.9411919.

[5] L. Guerdan, A. Raymond, H. Gunes, Toward Affective XAI: Facial Affect Analysis for Understanding Explainable Human-AI Interactions, in: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), IEEE, Montreal, BC, Canada, 2021, pp. 3789–3798. doi:10.1109/ICCVW54120.2021.00423.
URL https://ieeexplore.ieee.org/document/9607734/

[6] H. Yuan, J. Tang, X. Hu, S. Ji, XGNN: Towards Model-Level Explanations of Graph Neural Networks, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 430–438. doi:10.1145/3394486.3403085.
URL https://doi.org/10.1145/3394486.3403085

[7] Z. Ying, D. Bourgeois, J. You, M. Zitnik, J. Leskovec, GNNExplainer: Generating Explanations for Graph Neural Networks, in: Advances in Neural Information Processing Systems, Vol. 32, Curran Associates, Inc., 2019.
URL https://proceedings.neurips.cc/paper/2019/hash/d80b7040b773199015de6d3b4293c8ff-Abstract.html

[8] T. Schnake, O. Eberle, J. Lederer, S. Nakajima, K. T. Schütt, K.-R. Müller, G. Montavon, Higher-Order Explanations of Graph Neural Networks via Relevant Walks, IEEE Trans. Pattern Anal. Mach. Intell. (2021) 1–1ArXiv:2006.03589 [cs, stat]. doi:10.1109/TPAMI.2021.3115452.
URL http://arxiv.org/abs/2006.03589

[9] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, W. Samek, On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation, PLOS ONE 10 (7) (2015) e0130140, publisher: Public Library of Science. doi:10.1371/journal.pone.0130140.
URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0130140

[10] A. Raison, P. Bourdon, D. Helbert, EiX-GNN : Concept-level eigencentrality explainer for graph neural networks (Jun. 2022).
URL https://hal.archives-ouvertes.fr/hal-03686299

[11] C.-K. Yeh, C.-Y. Hsieh, A. Suggala, D. I. Inouye, P. K. Ravikumar, On the (In)fidelity and Sensitivity of Explanations, in: Advances in Neural Information Processing Systems, Vol. 32, Curran Associates, Inc., 2019.
URL https://proceedings.neurips.cc/paper/2019/hash/a7471fdc77b3435276507cc8f2dc2569-Abstract.html

[12] B. K. P. Horn, Closed-form solution of absolute orientation using unit quaternions, J. Opt. Soc. Am. A, JOSAA 4 (4) (1987) 629–642, publisher: Optica Publishing Group. doi:10.1364/JOSAA.4.000629.
URL https://opg.optica.org/josaa/abstract.cfm?uri=josaa-4-4-629

[13] F. Scarselli, M. Gori, Ah Chung Tsoi, M. Hagenbuchner, G. Monfardini, The Graph Neural Network Model, IEEE Trans. Neural Netw. 20 (1) (2009) 61–80. doi:10.1109/TNN.2008.2005605.
URL http://ieeexplore.ieee.org/document/4700287/

[14] T. N. Kipf, M. Welling, Semi-Supervised Classification with Graph Convolutional Networks, arXiv:1609.02907 [cs, stat]ArXiv: 1609.02907 (Feb. 2017).
URL http://arxiv.org/abs/1609.02907

[15] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph Attention Networks, arXiv:1710.10903 [cs, stat]ArXiv: 1710.10903 (Feb. 2018).
URL http://arxiv.org/abs/1710.10903

[16] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, M. Pantic, 300 Faces In-The-Wild Challenge: database and results, Image and Vision Computing 47 (2016) 3–18. doi:10.1016/j.imavis.2016.01.002.
URL https://linkinghub.elsevier.com/retrieve/pii/S0262885616000147

[17] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic, 300 Faces in-the-Wild Challenge: The First Facial Landmark Localization Challenge, in: 2013 IEEE International Conference on Computer Vision Workshops,

IEEE, Sydney, Australia, 2013, pp. 397–403. doi:10.1109/ICCVW.2013.59.
URL http://ieeexplore.ieee.org/document/6755925/

[18] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic, A Semi-automatic Methodology for Facial Landmark Annotation, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE, OR, USA, 2013, pp. 896–903. doi:10.1109/CVPRW.2013.132.
URL http://ieeexplore.ieee.org/document/6595977/

[19] A. Bulat, G. Tzimiropoulos, How Far are We from Solving the 2D & 3D Face Alignment Problem? (and a Dataset of 230,000 3D Facial Landmarks), in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, 2017, pp. 1021–1030. doi:10.1109/ICCV.2017.116.
URL http://ieeexplore.ieee.org/document/8237378/

[20] X. Zhu, Z. Lei, X. Liu, H. Shi, S. Z. Li, Face Alignment Across Large Poses: A 3D Solution 10.

[21] A. Raison, P. Bourdon, D. Helbert, ScoreCAM GNN : une explication optimale des réseaux profonds sur graphes, in: XXVIIIème Colloque Francophone de Traitement du Signal et des Images (GRETSI), Nancy, France, 2022.
URL https://hal.archives-ouvertes.fr/hal-03694349

[22] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, X. Hu, Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Seattle, WA, USA, 2020, pp. 111–119. doi:10.1109/CVPRW50498.2020.00020.
URL https://ieeexplore.ieee.org/document/9150840/