UNIVERSITÀ DI TRENTO

Facial expression

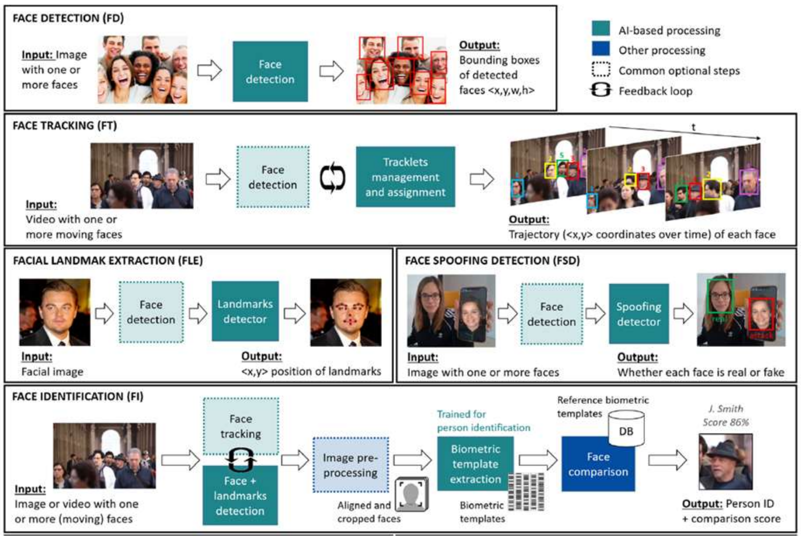
1

Facial processing

- Facial processing is the discipline that takes facial images / videos / data as input of a system and returns higher level information
- Most relevant facial processing tasks (Hupont et al., 2022):
  - Face detection
  - Face tracking
  - Facial landmark extraction
  - Face spoofing detection
  - Face identification
  - Face verification
  - Kinship verification
  - Facial expression recognition
  - Action Unit detection
  - Automatic lip reading
  - Facial attribute estimation
  - Facial attribute manipulation

2

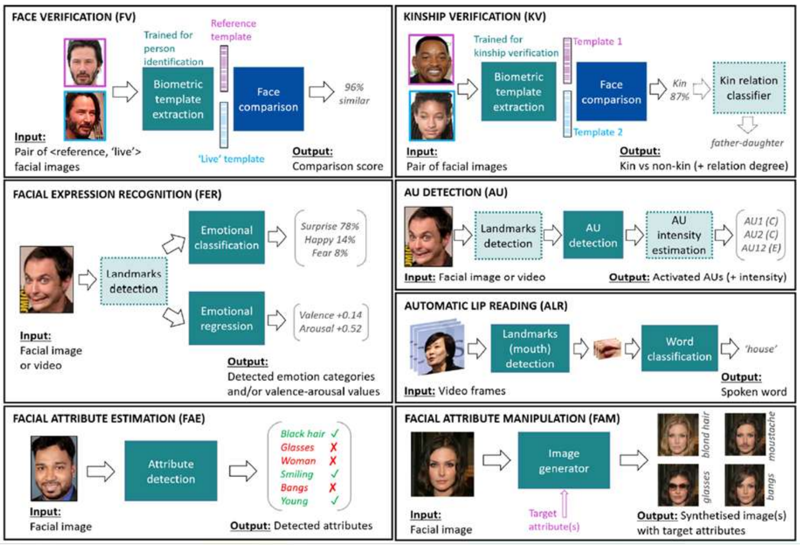
Computational pipelines for the facial processing tasks



Picture from (Hupont et al,, 2022)

3

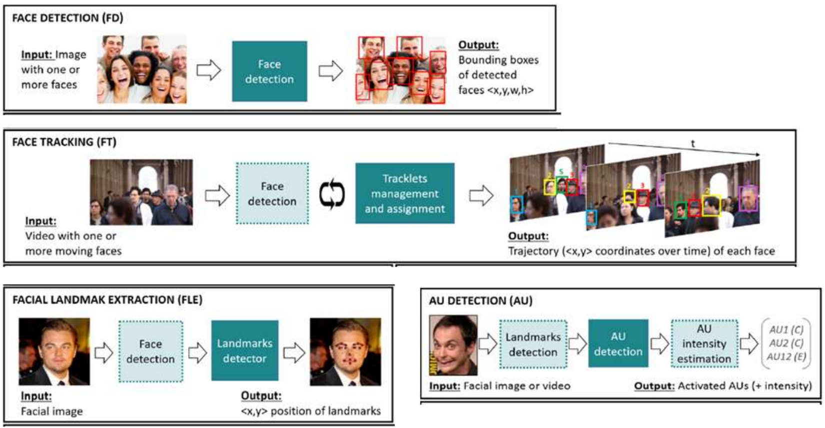
Computational pipelines for the facial processing tasks



Picture from (Hupont et al,, 2022)

4

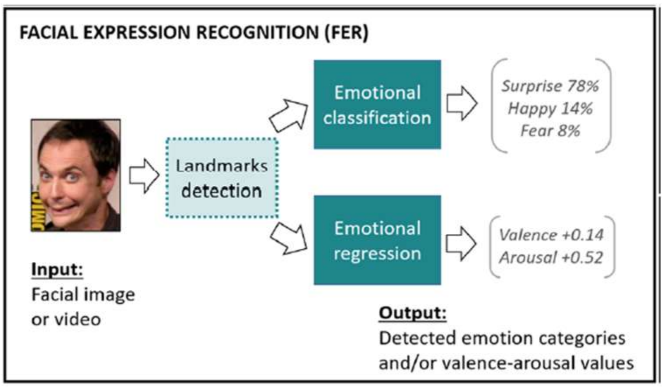
We will focus on …..



To do…

5

2D Facial Expression Recognition (FER)



FER is the technlogy that analyzes facial expressions from both static images and videos * in order to reveal information on one’s emotional state (EDS TechDispatch, issue 1, 2021)

* Other kinds of data can be used

6

Potential uses of FER

Personalized devices

- analyse emotions to display personalised messages in smart environments
- provide personalised recommendations e.g. on music selection /cultural material
- analyse facial expressions to predict individual reaction to movies

Customer behaviour analysis and advertising

- analyse customers' emotions while shopping focused on either goods or their arrangement within the shop
- advertising signage at a railway station using a system of recognition and facial tracking for marketing purposes

Education

- monitor students' attention
- design affective tutoring system
- detect engagement in online learning

Source: EDS TechDispatc, 1, 2021

7

Potential uses of FER

Employement

- help making of recruiters
- identify uninterested candidates in a job interview
- monitor moods and attention of employees

Healtcare

- detect autism or neurodegenerative diseases
- predict psychotic disorders or depression to identify users in need of assistance
- suicide prevention
- observe patients conditions during treatment

Public safety

- lie detectors and smart border control
- predictive screening of public spaces to identify emotions triggering potential terrorism threat

Source: EDS TechDispatc, 1, 2021

8

What is a facial expression?

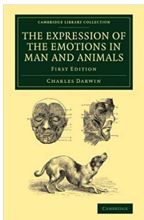
- A facial expression is one or more motions or positions of the muscles beneath the skin of the face
  (Freitas-Magalhães, 2011)
- Facial expressions serve as primary nonverbal means for human beings to regulate their interactions:
  - Communication of emotions
  - Emphasis on what is being said
  - Disagreement
  - …
- Facial expression duration [250ms, 5s]

9

At the beginning

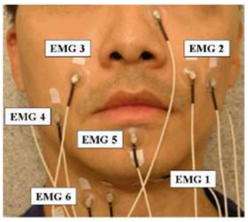
- Duchenne (1862) carried out the first study on facial expressions: he studied the electro-stimulation of individual facial muscles responsible for the production of facial expressions.

- Darwin (1872) published *“The expression of the emotions in man and animals”.* He explored the relevance of facial expressions for communication and described variations in facial expressions of emotions.

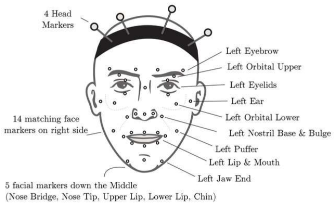


10

Devices for capturing faces



Picture from (Jou & Schultz, 2008)



Picture from (El Haddad et al., 2016)

11

Two consolidated approaches to facial expressions

- Message-based: aims to directly decode the meaning conveyed by a facial display (e.g. happy, sad and so on).

Picture from (Vinciarelli, Pantic, Bourlard, Pentland, 2008)

Concretely, labeling the meaning into categories. Example 6 basic emotions

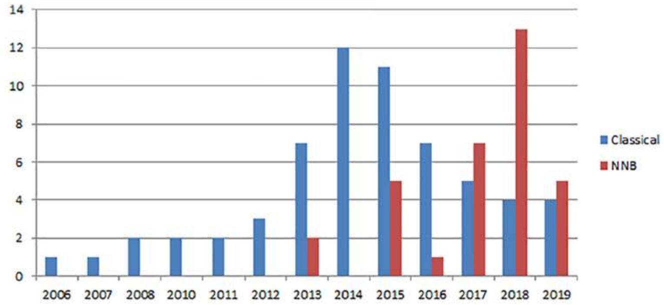
- Sign-based: aims to study the physical signal used to transmit the message (e.g. raised cheeks or depressed lips)

Picture from (Tian, Kanade, Cohn, 2004)

12

Concretely, describing the changes in the face configuration

Classical vs Deep Learning approaches

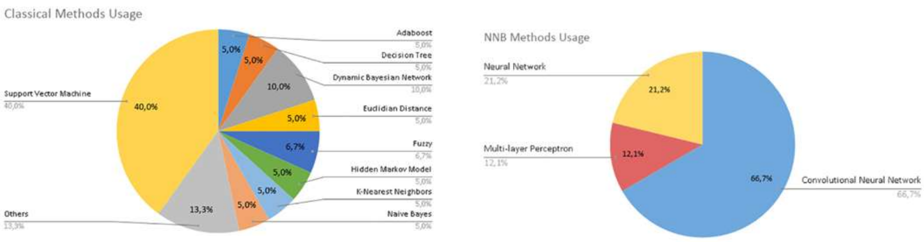


Picture from (Zago Canal, et al., 2022)

- Search done on:
  - ACM Digital Library, IEEE Xplore, Science Direct, Scopus (2006-2019)
  - 51 papers reporting 94 different methods were analyzed
  - Faces from images / videos

13

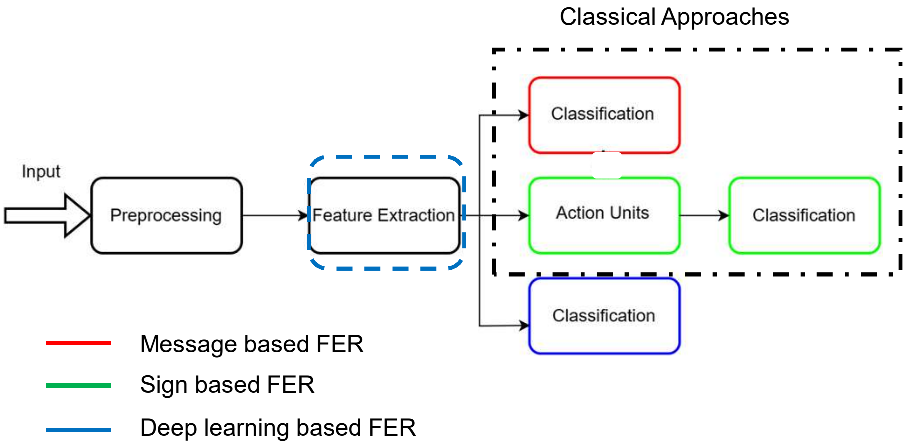
Classical vs Deep Learning approaches



Picture from (Zago Canal, et al., 2022)

14

Computational pipelines



15

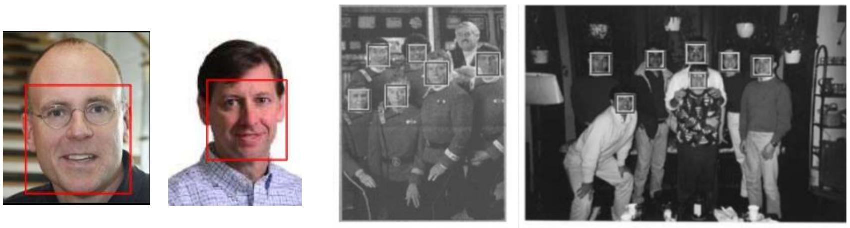
Common steps to FER

- Input:
  - Images (static FER)
  - Video (dynamic FER)
  - MoCap
  - EMG signals
- Preprocessing
  - Face detection (possibly tracking)
  - Facial landmarks localisation
  - Facial registration / normalization

16

Face detection

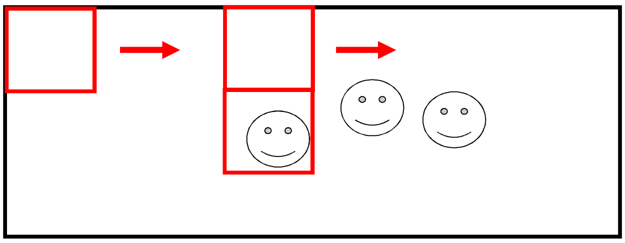
- Viola-Jones detector (2001, 2004) was the first algorithm making face detection practically feasible in real-world applications
- It is the reference face detection algorithm
- Until today, it was / is (?) widely applied in digital cameras and photo organization software



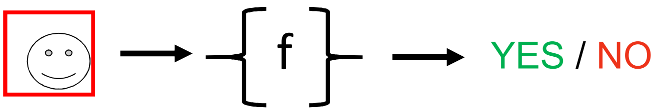
17

The Vola-Jones' idea

- Sliding a window across an image…



- …and evaluate a face model at each move



- Question: many faces generally?

18

Key steps of the Viola-Jones' detector

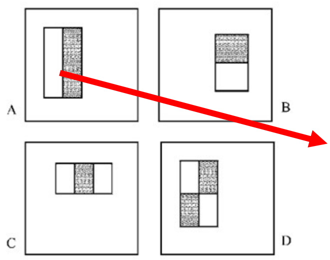
- Extracting Haar-like features
- Computing intergral images for fast feature extraction
- Boosting for feature selection
- Attentional cascade for having a fast rejection of no-faces (optimization step!)

19

Haar feature extraction

- The input image is scanned using a 24 x 24 window
- We want to compute very efficient features
- Haar features are computed on each window as the difference between the normalized sum of the pixel values in the dark area and the normalized sum of the pixel values in the white area:

if such difference is close to 1, we have an something



Haar filters

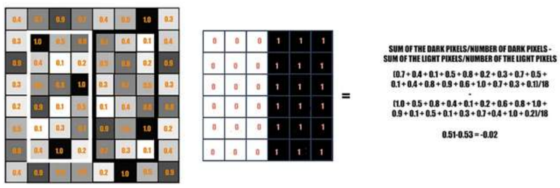


Fig. The rectangle on the left is a sample representation of an image with pixel values 0.0 to 1.0. The rectangle at the center is a haar kernel which has all the light pixels on the left and all the dark pixels on the right. The haar calculation is done by finding out the difference of the average of the pixel values at the darker region and the average of the pixel values at the lighter region. If the difference is close to 1, then there is an edge detected by the haar feature.

20

https://towardsdatascience.com/face-detection-with-haar-cascade-727f68dafd08

Integral image

- Computation cost of Haar's features over an image:

(N x M -1) per pixel per filter per size

- N, M are the Haar's filter size
- Complexity: O(N x M) per filter

- We can do defintely better with the *trick of the integral image*

- Given an image, its integral image ii at location (*x*, *y*) is a new image that contains the sum of the pixel values above and to the left of *x* and *y*, inclusive:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y').$$

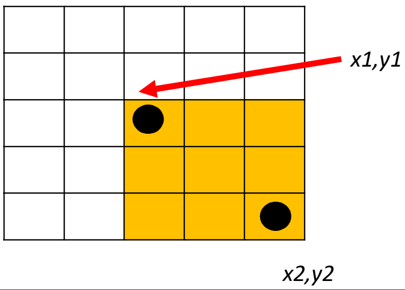
21

Let's see how it works

- Integral image allows us to compute the sum of the values in a rectanglular filter in one step only (O(1) per filter)
- How it works:

For computing the sum S inside a generic rectangle:

$$S = II(x_2, y_2) - II(x_1 - 1, y_2) - II(x_2, y_1 - 1) + II(x_1 - 1, y_1 - 1)$$



22

And now try it with a quiz!

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

0	1
0	1

1	2
5	6

Compute the Haar features on the submatrix:
- using the direct computation;
- using the integral image

23

Viola- Jones Haar Cascade

- Considering all possible filter parameters (i.e. position, size, and type) there are 160K features for each 24 x 24 window!
- The idea is to look for the most informative features and discard the other ones
- AdaBoost is adopted: it consists in the use of many *weak classifiers* for building a *strong classifier* which is a weighted combination of the weak classifiers
- Each weak classifier uses just one feature

24

Viola- Jones Haar Cascade

- Each weak classifier uses just one feature

polarity feature threshold

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) < p_j \theta_j \\ 0 & \text{otherwise} \end{cases}$$

Weak classifier

24 x 24 window

Number of weak classifiers

$$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

Strong classifier

- Optimal T is around 200 to have 95% of accuracy, but 1fps!

25

Viola-Jones Haar Algorithm

Given:
N images labeled + (face) or – (no face).
Images are given weights w.
Initially, all weights w are set equally.

Repeat T times:

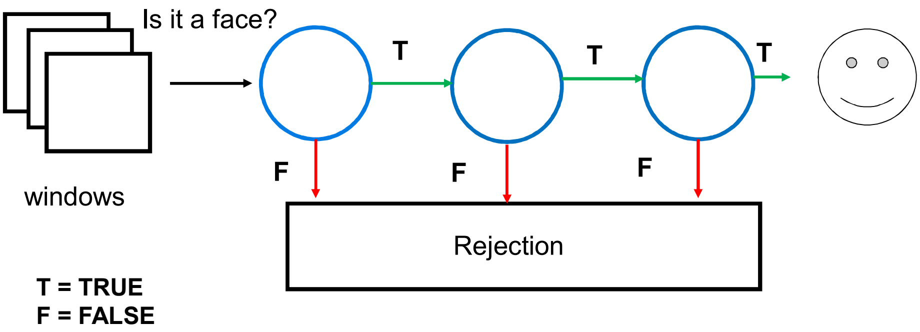
Step 1:
Choose the most efficient weak classifier. Threshold θ is estimated to maximize accuracy. Assign the classifier a weight α proportional to its accuracy.

Step 2:
Update the weights w to emphasize the examples which were incorrectly classified. This makes the next weak classifier to focus on "harder" samples.

Result:
The final (strong) classifier is a weighted combination of the T "weak" classifiers, weighted according to their accuracy.

26

Viola- Jones Haar Cascade



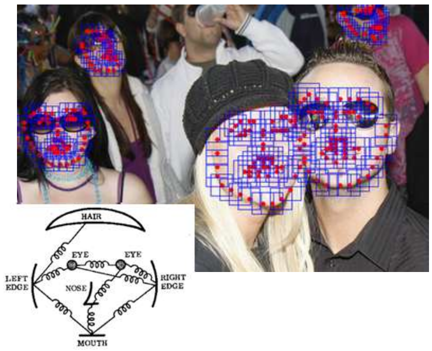
Cascade: speed up the algorithm; reduce the false positive rate

78% of accuracy, but 15fps!

27

Face detection

- A recent classification (Zafeiriou, 2015):
  - Algorithms based on rigid templates
  - Algorithms applying Deformable Parts-based Models (DPM)



Picture from (Fischler and Elschlager,1973)

28

Demo Time!

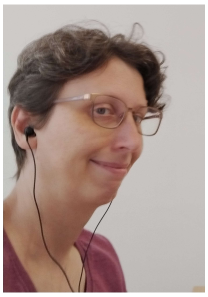
List of the issues:

30

Detecting profile face?

- How can we solve this issue?
  - Collect (many!) profile faces
  - Train a new detector on those





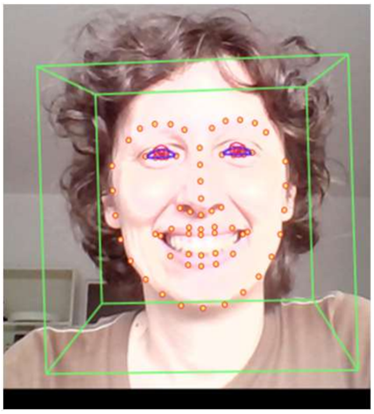
31

Demo Time: MediaPipe

Face detection

Facial landmarks localization

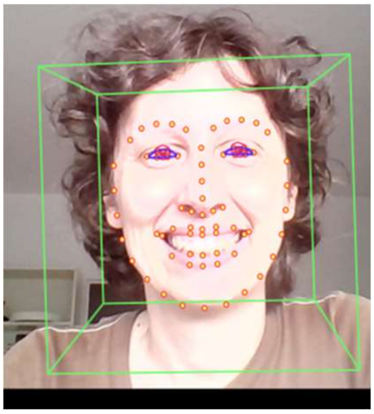
- Facial landmarks are defined as distinctive face locations, such as corners of the eyes, center of the bottom lip, tip of the nose, and so on
- Taken in sufficient numbers they define the face shape.
- Localization and tracking of facial landmarks can improve accuracy of analysis of facial expression.



33

Facial landmarks localization

- Two classes of algorithms:
  - Generative Models
    - Active Appearance Models (AAM)
  - Discriminative Models
    - Response map fitting
    - Regression-based approach
    - Deformable Parts Model



34

Supervised Descent Method (Xiao, de la Torre 2013)

- Regression-based approach
- SDM is a supervised method that learns to optimize non-linear least squares problems
- SDM learns generic descent maps in a supervised manner. It is able to overcome many draw-backs of second order optimization schemes, such as non-differentiability and expensive computation of the Jacobians and Hessians
- Very popular method able to work in real-time settings

35

IntraFace (de la Torre et al, 2013)

36