ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

DEEPRETAIL2020
2nd WORKSHOP ON DEEP UNDERSTANDING SHOPPER BEHAVIOURS AND INTERACTIONS IN INTELLIGENT RETAIL ENVIRONMENTS
Milan, Italy 11 January 2021

Who is in the crowd? Deep face analysis for crowd understanding

Simone Bianco, Luigi Celona, and Raimondo Schettini

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Imaging and Vision

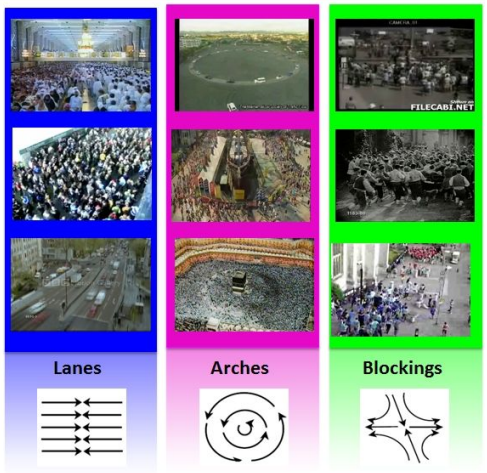
www.ivl.disco.unimib.it
luigi.celona@unimib.it
@luigi_celona

Crowd understanding

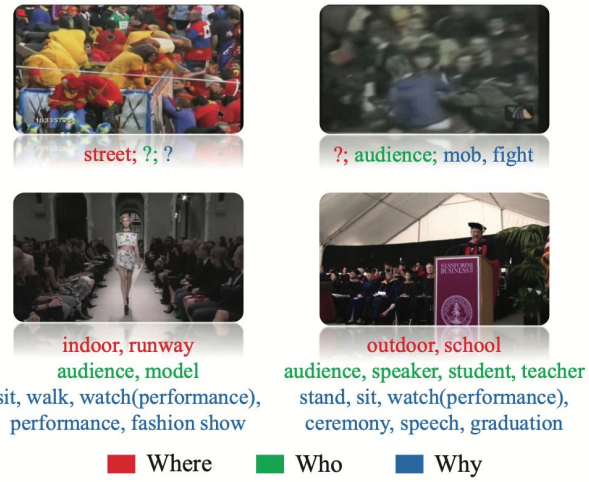
- Many applications such as surveillance, entertainment, marketing, and social sciences
- Most methods conduct a **coarse-grained analysis** of the crowd
- Recently, some studies have pointed the attention on the importance of providing an **attribute-based description of the crowd**

Crowd counting

Crowd behaviour analysis



Crowded scene understanding



Crowd understanding *based on face analysis*

- The human face is rich in information and conveys much of our age, emotions and lifestyle
- Knowing **who** is in the crowd, **what interests** them, **what they want**, **what they like or dislike** is relevant to the management and planning processes in various sectors of public and commercial services
  - Marketing: for retailers to know who the customers are, what motivates them, what they want, need, love, or hate
  - Ambient Intelligence: to send tailored messages, push notifications, and ads to improve the “customer experience”

Crowd understanding system

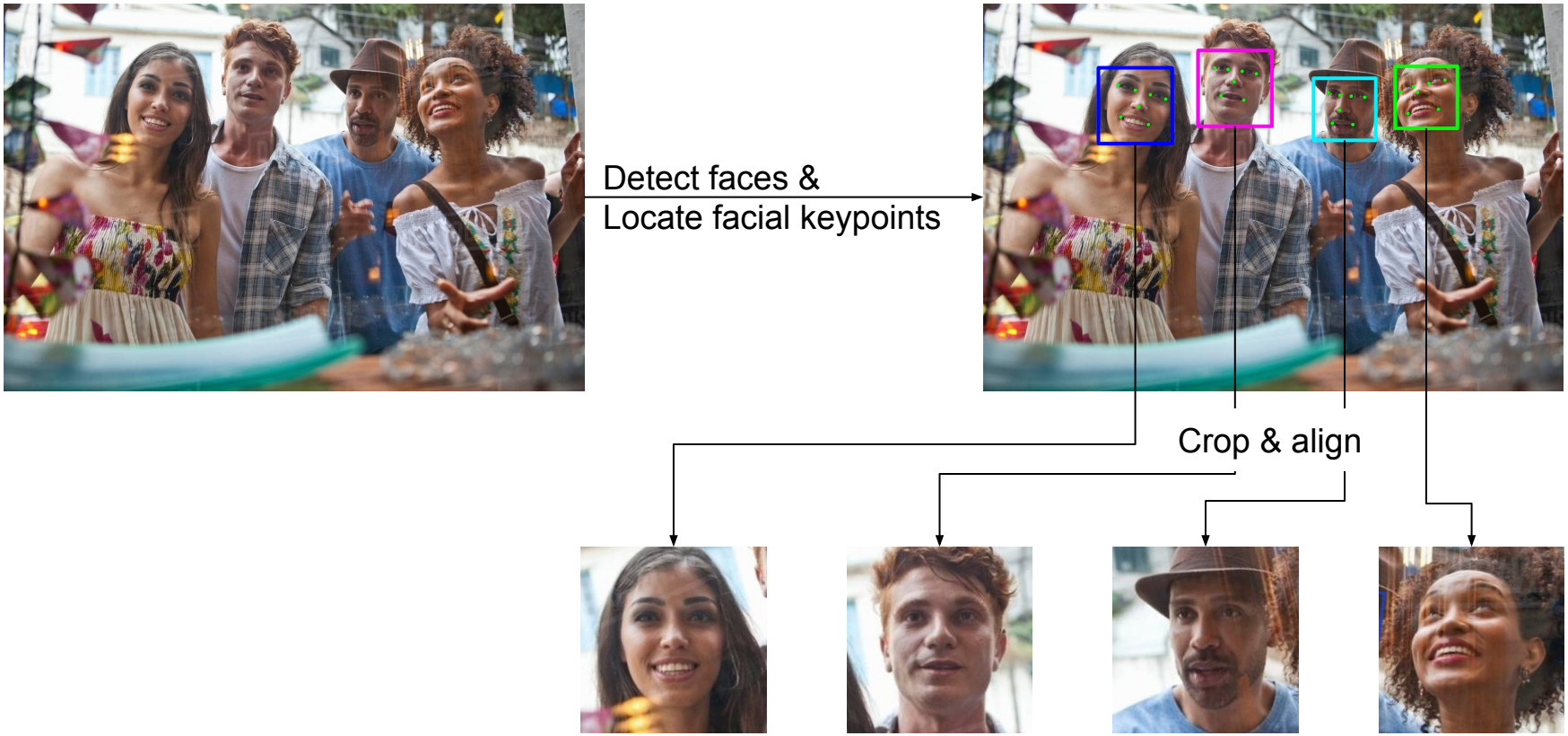
- A system capable of gathering information about people in the crowd basing on the **analysis of face**

- It conducts a fine-grained analysis of the crowd to gather information about
  - Demographic
  - Sentiment
  - Other attributes

- Challenging problem especially in dense crowds due to
  - The presence of cluttering, overlapping and occlusions
  - The low-resolution or low-quality of images/frames

Face detection

- The faces into a crowded image are detected
- For each face, 68 facial keypoints are located
- Each face is aligned by exploiting the position of the eyes, nose, and mouth
- We use the face detector and the landmark estimator of the DLib library [1]

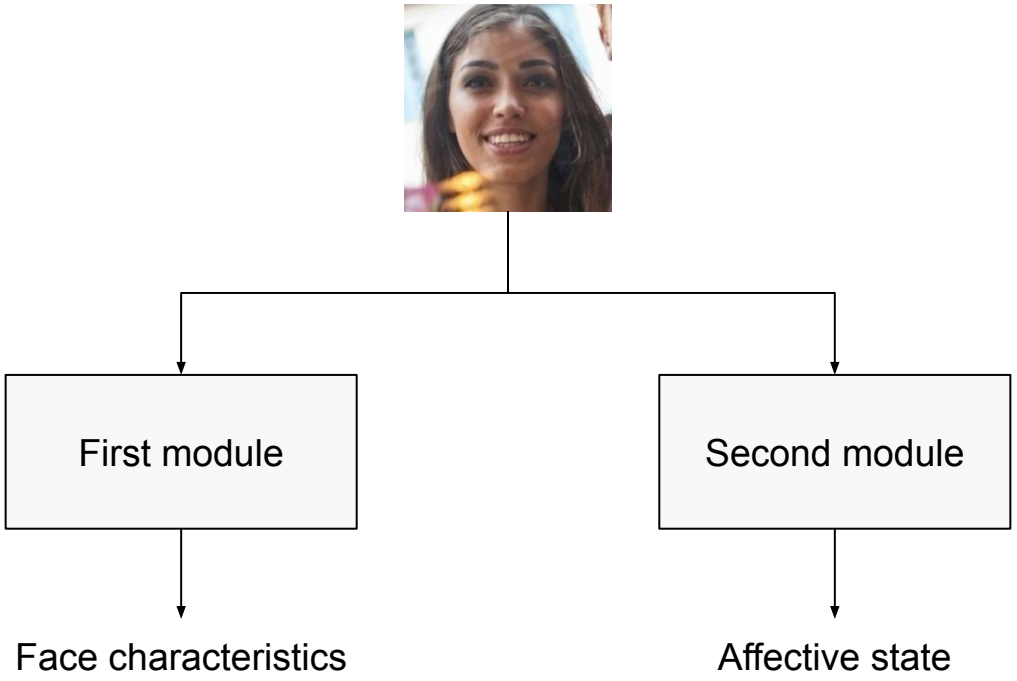


[1] King, D.E.: Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research 10, 1755–1758 (2009)

Face analysis

- Each detected face is processed by the face analysis module

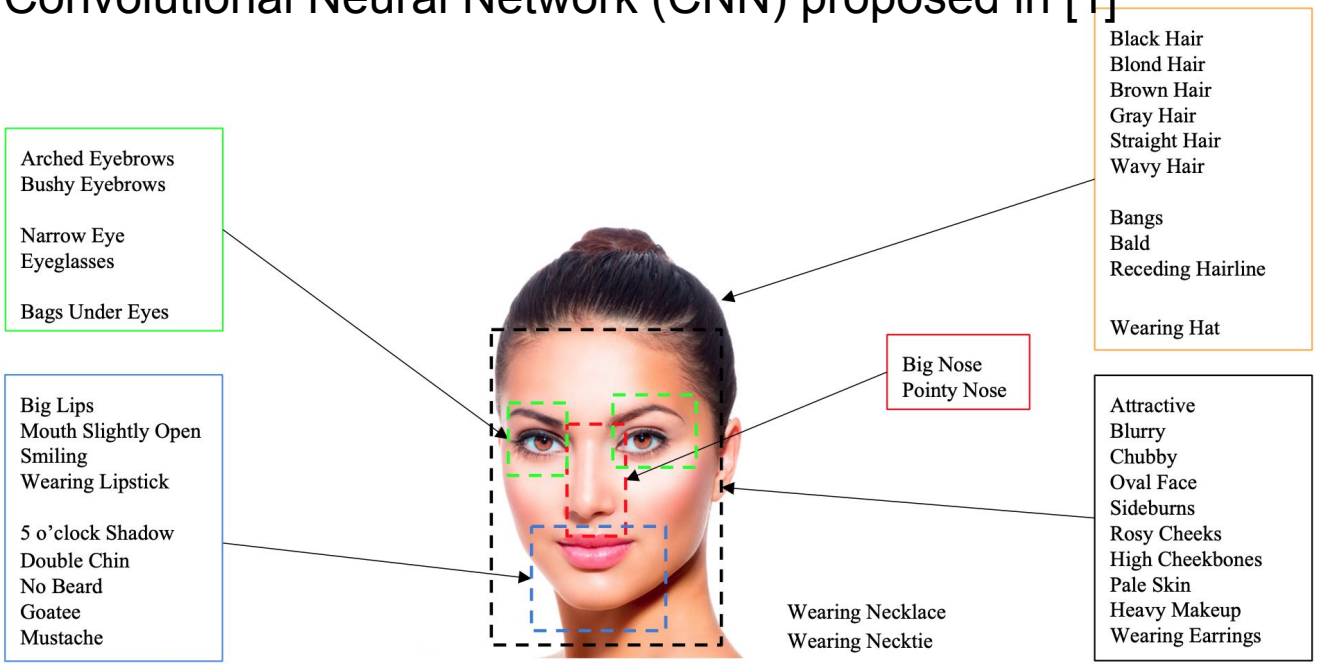
- The face analysis module consists of **two different modules**
  - The first estimates face characteristics
  - The second recognizes emotions



Face analysis

First module: Facial characteristics

- The facial image is described in terms of
  - **Demographic**, *i.e.* gender and one among eight age groups
  - About 30 **visual attributes**
  - **Perceptual attributes** about image quality
  - The presence of **accessories**
- The Multi-task Convolutional Neural Network (CNN) proposed in [1]

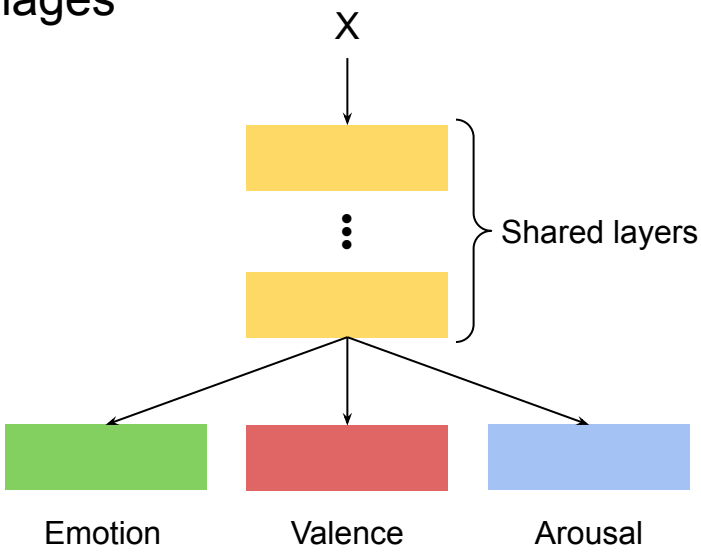


[1] Celona, L., Bianco, S., Schettini, R.: Fine-grained face annotation using deep multi-task CNN. MDPI Sensors **18**(8), 2666 (2018)

Face analysis

Second module: sentiment analysis of facial images

- A ResNet-50 architecture to simultaneously
  - Categorize emotion into **8 discrete classes** (Angry, Contempt, Disgusted, Happy, Neutral, Sad, Scared, Surprised)
  - Estimate emotion on a **continuous scale** (i.e. valence and arousal)



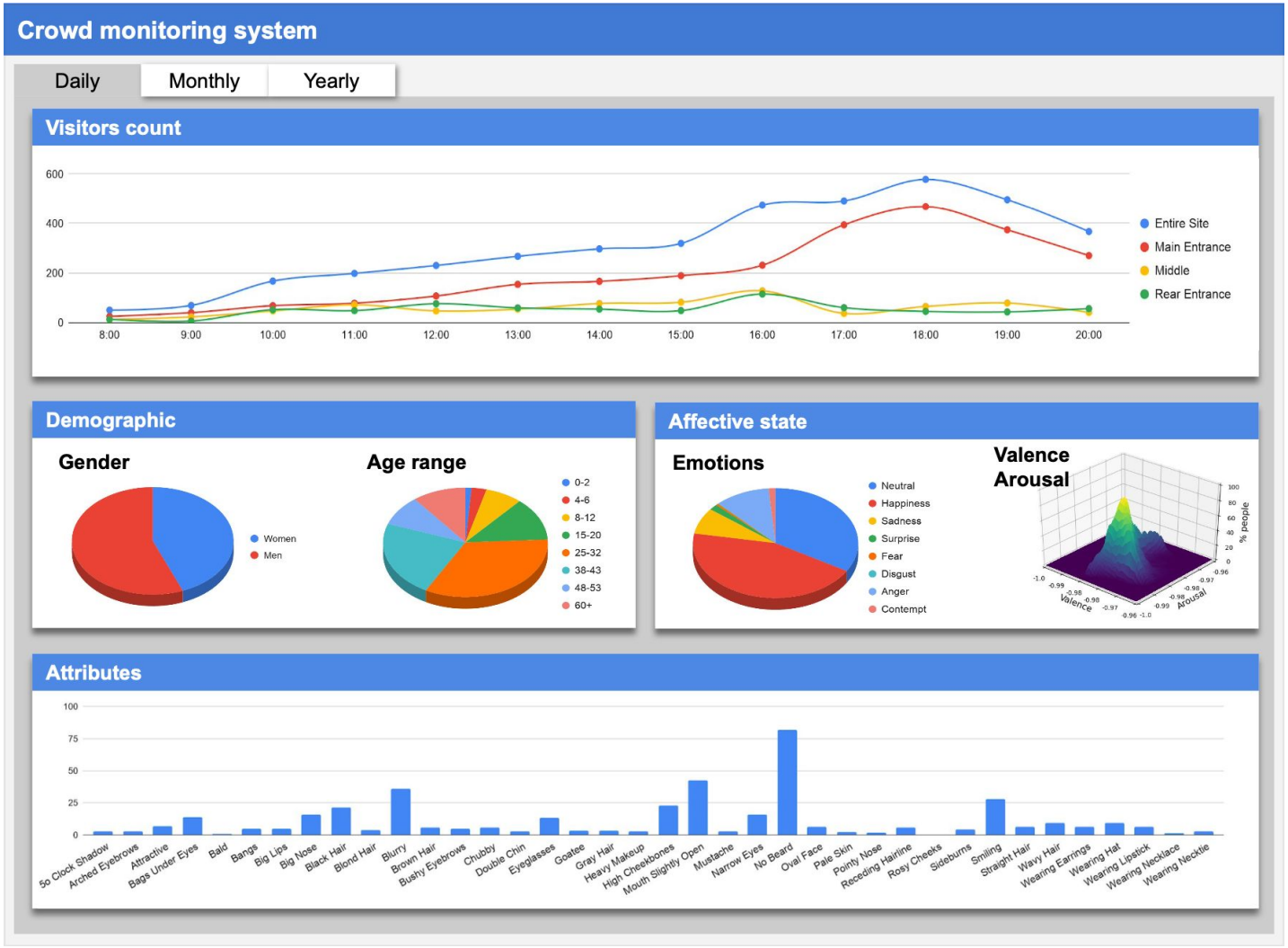
- Model trained on the AffectNet [1] database
  - Faces annotated by only one human coder in terms of discrete emotion categories and valence-arousal scores
  - 400,000 training images
  - 5,000 validation images

[1] Mollahosseini, A., Hasani, B., Mahoor, M.H.: Affectnet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing 10(1), 18–31 (2017)

Face analysis

Facial details	Parameters
Demographic	Gender, Age range (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53. 60+)
Affective state	8 emotions (Angry, Contempt, Disgusted, Happy, Neutral, Sad, Scared, Surprised) Valence Arousal
Other attributes	5 o'Clock Shadow, Arched Eyebrows, Bags Under Eyes, Bald, Bangs, Big Lips, Big Nose, Black Hair, Blond Hair, Brown Hair, Bushy Eyebrows, Chubby, Double Chin, Goatee, Gray Hair, Heavy Makeup, High Cheekbones, Mouth Slightly Open, Mustache, Narrow Eyes, No Beard, Oval Face, Pale Skin, Pointy Nose, Receding Hairline, Rosy Cheeks, Sideburns, Straight Hair, Wavy Hair
Perceptual quality	Attractive, Blurry
Accessories	Earring, Eyeglasses, Hat, Lipstick, Necklace, Necktie

Display statistics

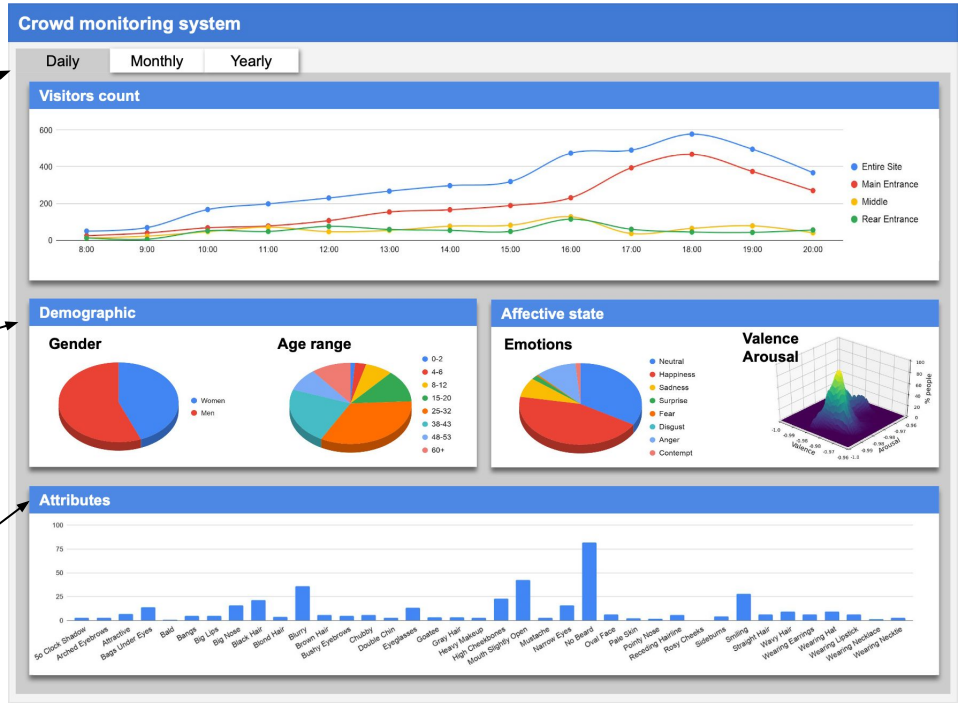


Display statistics

Statistics for several time periods

- Percentage of people per gender
- Percentage of people per age range

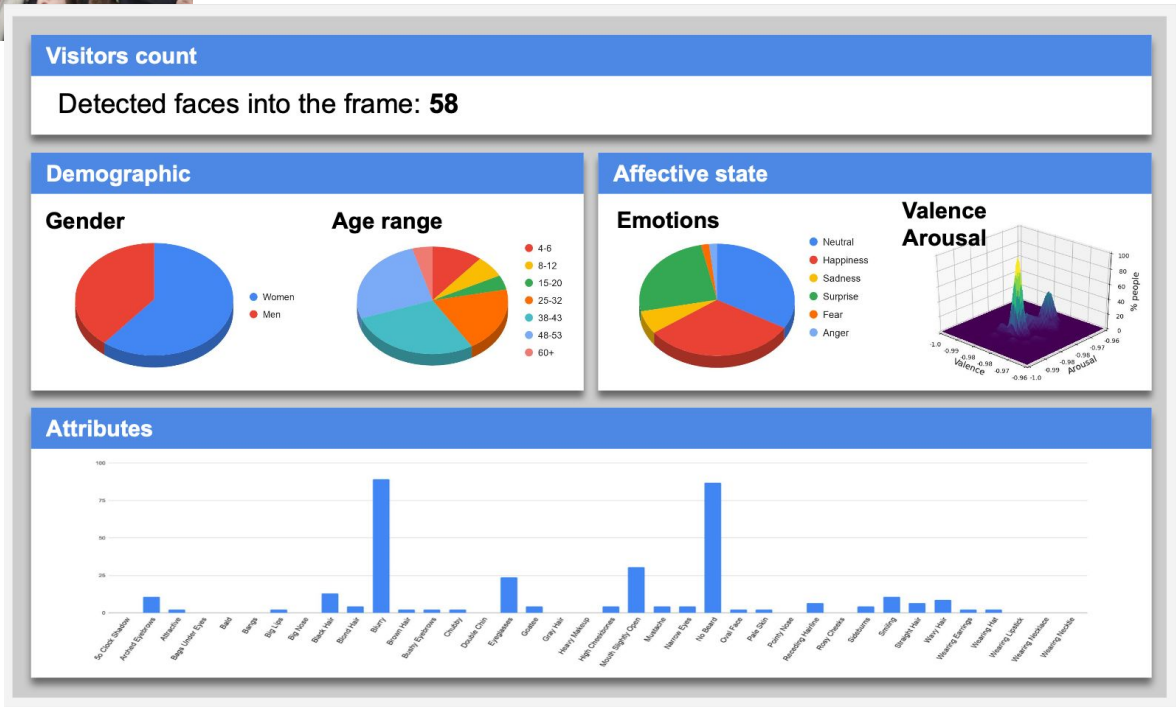
Percentage of individuals per attributes



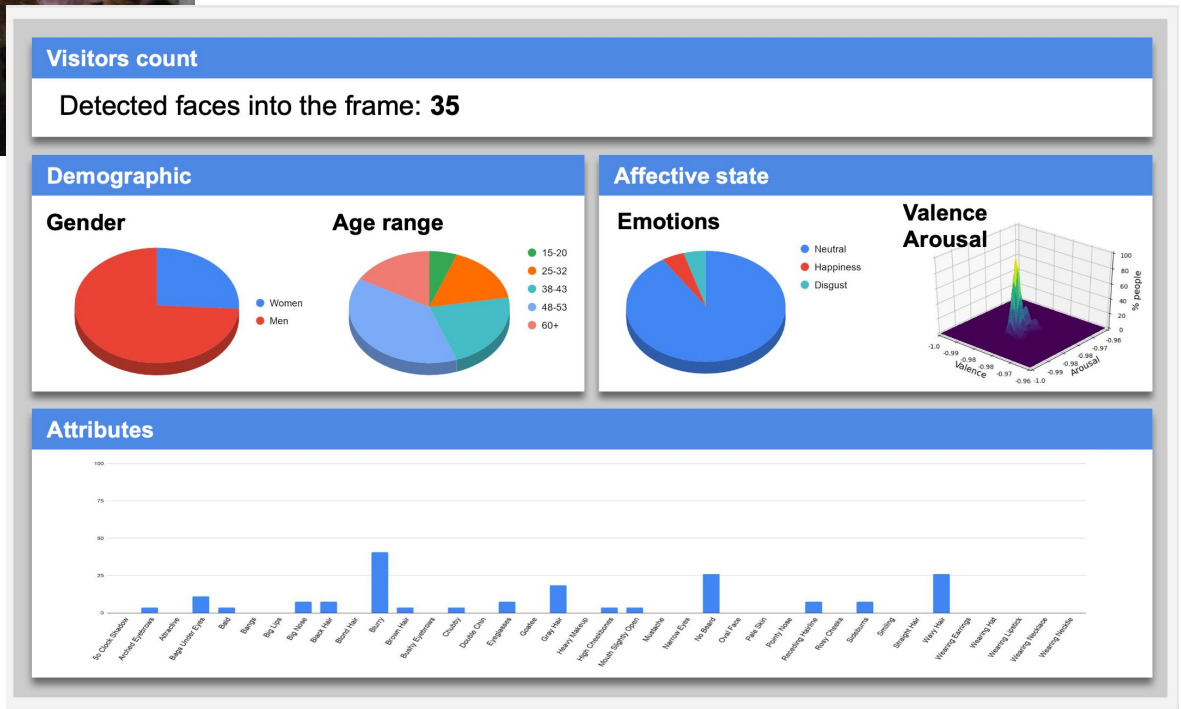
Number of people per hour

- Percentage of people per emotion
- Distribution of people for the valence-arousal dimensions

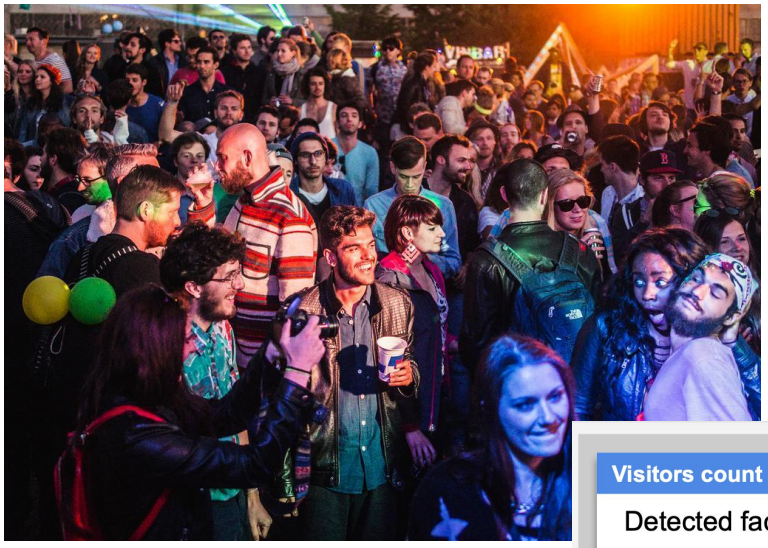
Qualitative results: Shopping center

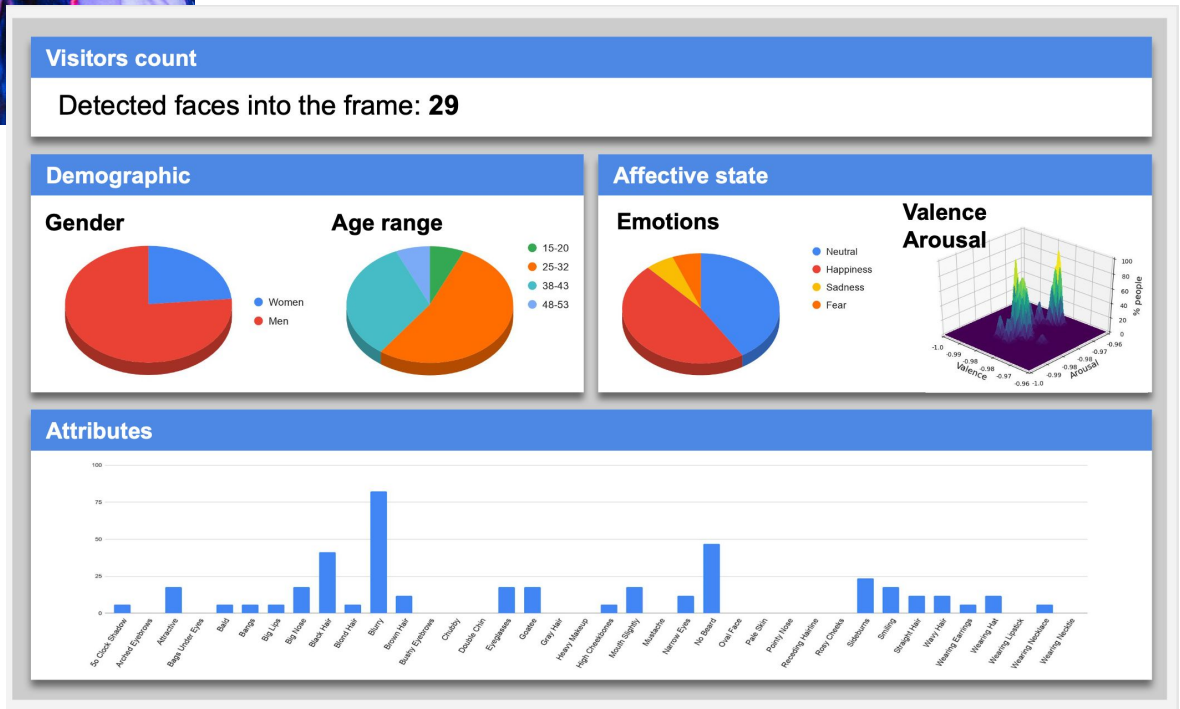


Qualitative results: Museum



Qualitative results: Concert





Conclusions

- We presented a **crowd understanding system based on face analysis** for collecting crowd statistics regarding
  - the demographic constitution
  - the affective state
  - other attributes
- The proposed system is applicable in various contexts where it is intended to **collect information on people in a crowd for statistical purposes** (e.g. marketing) and can be used as a **complement to existing crowd monitoring systems**

Future works

- ❖ Collect a database of crowded images for the development and validation of the proposed system
- ❖ Use a more effective face detector capable of detecting faces in more challenging poses and imaging conditions
- ❖ Include a face tracking algorithm to avoid recalculating statistics for the same individual in different frames
- ❖ Make the list of recognizable attributes customizable and also including for example the analysis of the upper body (e.g. outfit analysis)

Questions?

ivl.disco.unimib.it