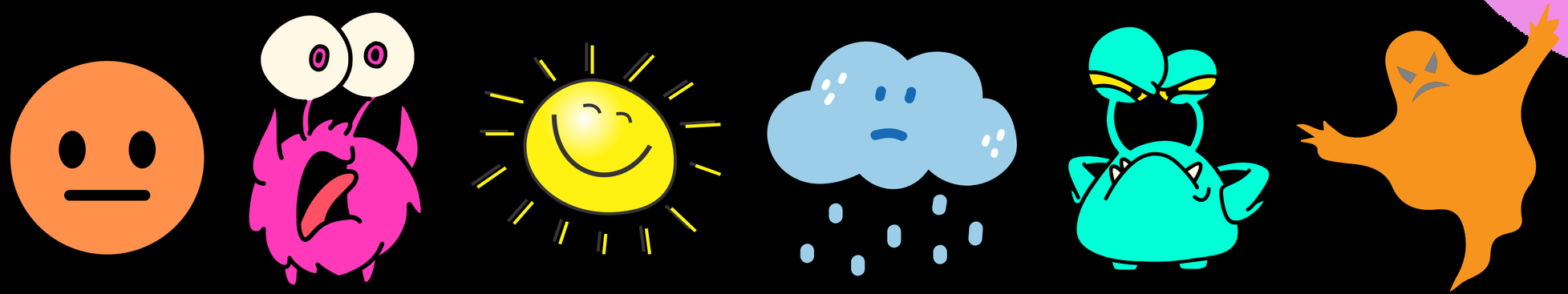
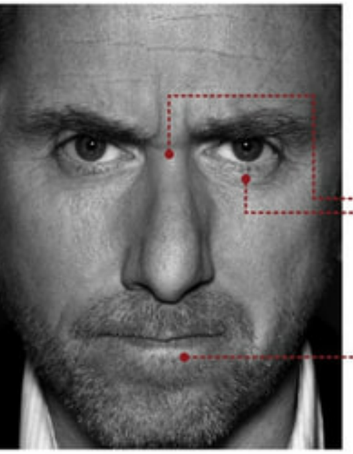
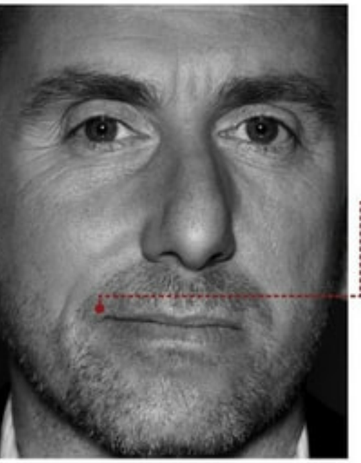
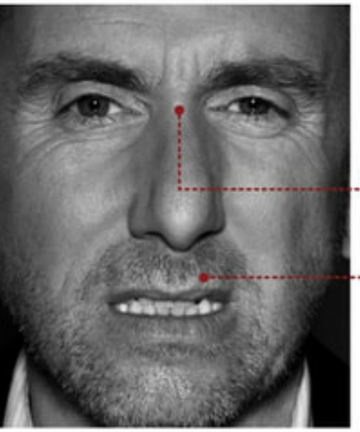
Emotion Recognition with Deep Learning

Camille Benoît
Alexandra Giraud









Why emotions? Applications?

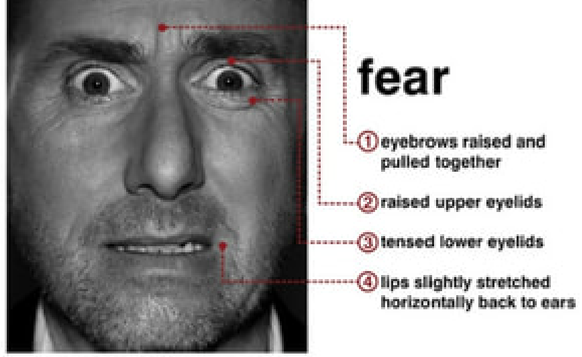
-Great challenge to tackle such an interesting issue

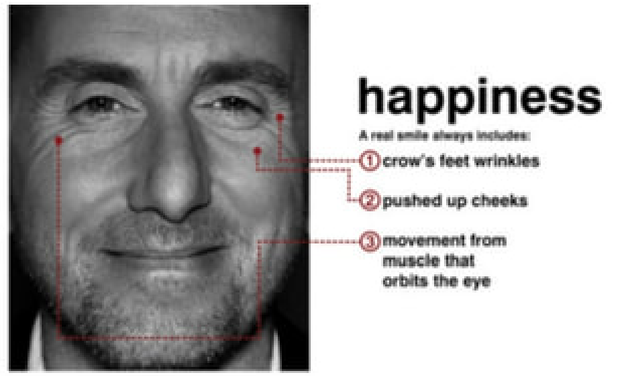
- Our initial idea was to apply it to **prevent drivers fatigue** and improve Road Security with AI

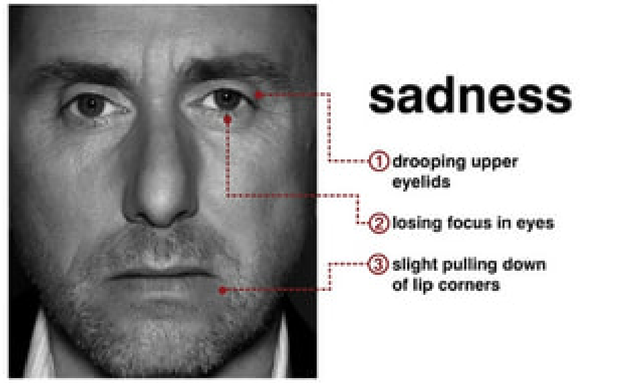
-Basis of 6 basic Human Emotions as described by Dr Ekman

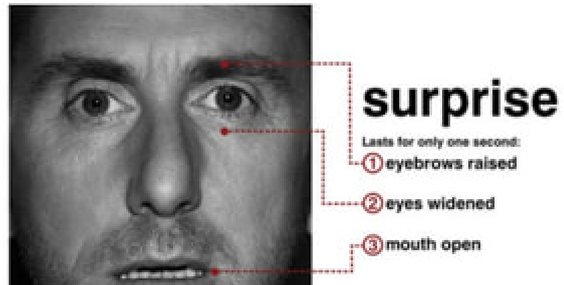
-Anger, Sadness, Suprise,Happiness,Fear and Disgust + Neutral face (absence of emotion)

We worked without the emotion disgust









OUR DATA : FER2013 DATASET

- FER2013 is one of the open rights datasets with the most images, around 30 000, with labels, training, and testing data
- Uniformisation of the data has been done (face-centered, same dimensions, lightning...)
- Ou can see some example in this slide

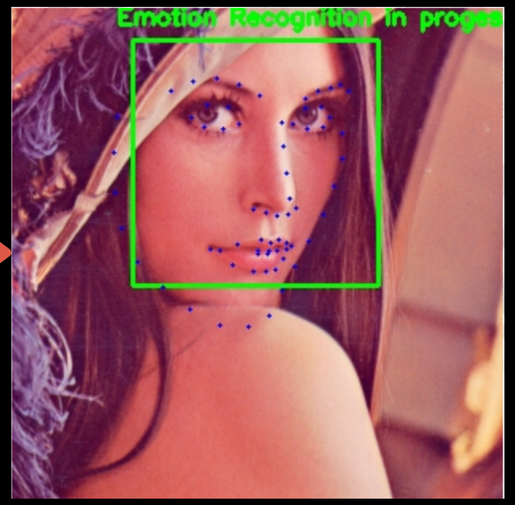
https://www.kaggle.com/deadskull7/fer2013

PROBLEMS: - still too small for Facial emotion recognition problem, state of the art accuracy is around 75 % (probably training accuracy)
- Some people have hands obscuring their faces

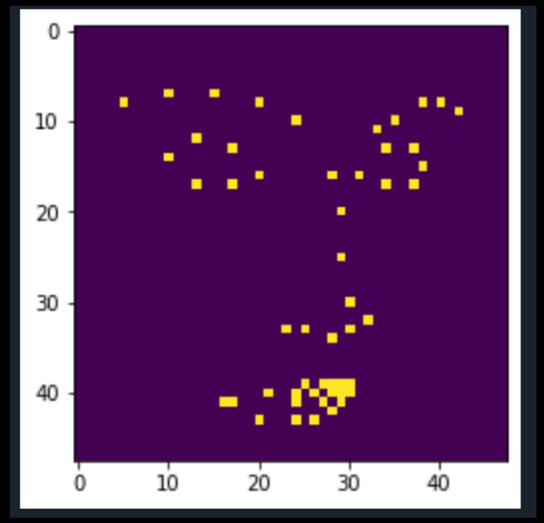
Preprocessing of the data

- CHOOSEN EMOTIONS: REMOVE DISGUST FROM DATASET
- FACE/**LANDMARK DETECTION** WITH OPENCV AND DLIB
- GENERATING **BINARY IMAGES** FOR INPUT

1. INPUT IMAGE (RANDOM SIZE)



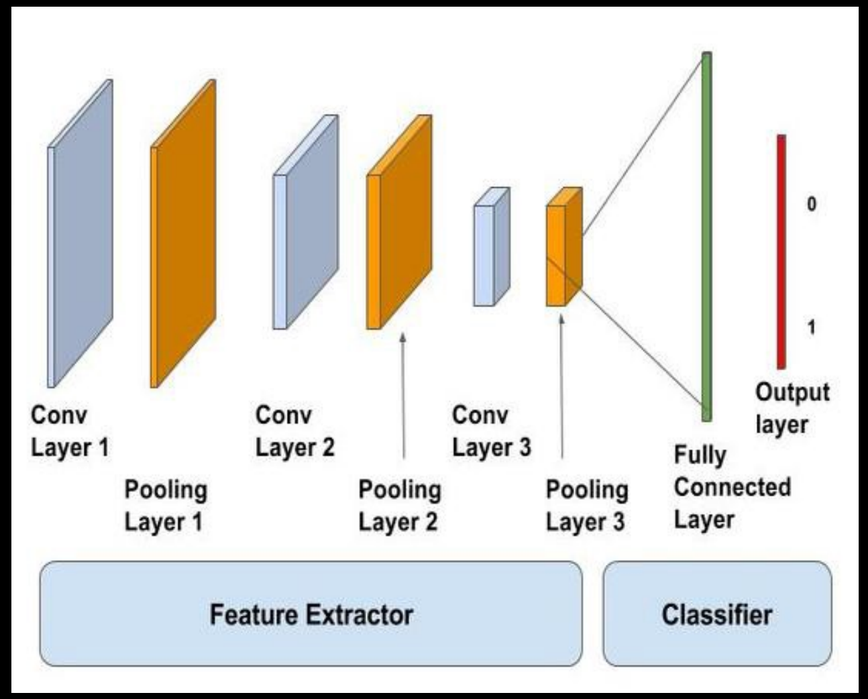
2. FACE DETECTION
3. CROP AROUND THE FACE
4. RESHAPE INTO (48,48)
4. LANDMARK DETECTION



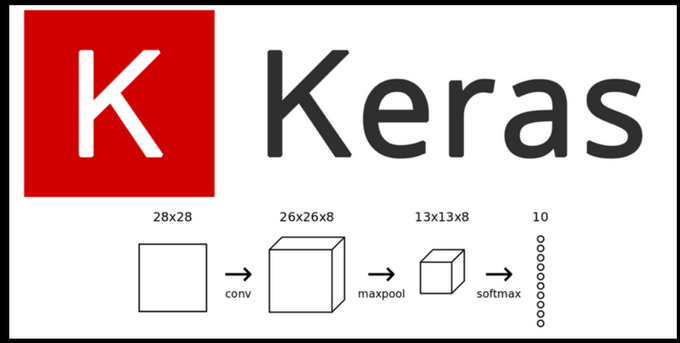
5. DELETE FACE SHAPE LANDMARK (NOT ENOUGH INFORMATION
6.CREATION OF A BINARY IMAGE (NORMALIZED)
7. RESHAPE THE INPUT IN (48,48,1) FOR OUR MODEL

OUR MODEL : CNN

–TYPE OF NEURAL NETWORK INSPIRED BY ANIMAL VISION ADAPTED TO IMAGES AND VIDEOS
– CAPABLE OF EXTRACTING FEATURES, LOW LEVEL AND HIGH LEVEL



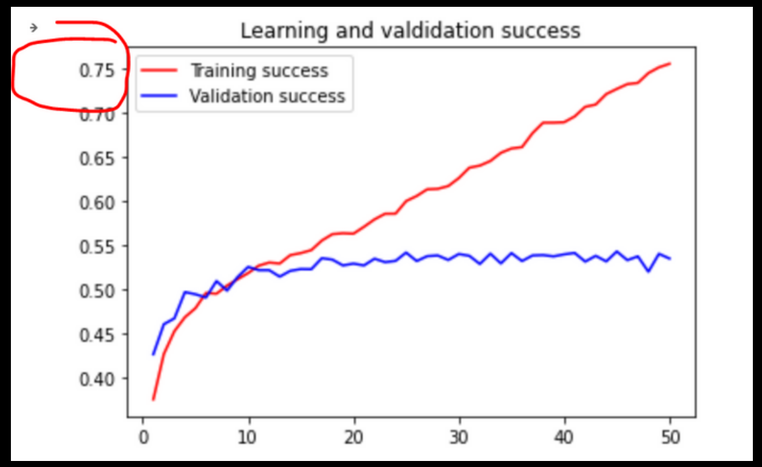
1. INPUT : SHAPE (48,48,1) (BINARY IMAGES)
1. 3 CONVOLUTIONAL LAYERS WITH MAX POOLING
2. 2 FC LAYERS WITH DROPOUT 0.2
3. SOFTMAX ACTIVATION FUNCTION (GIVES PROBABILITY DISTRIBUTION)

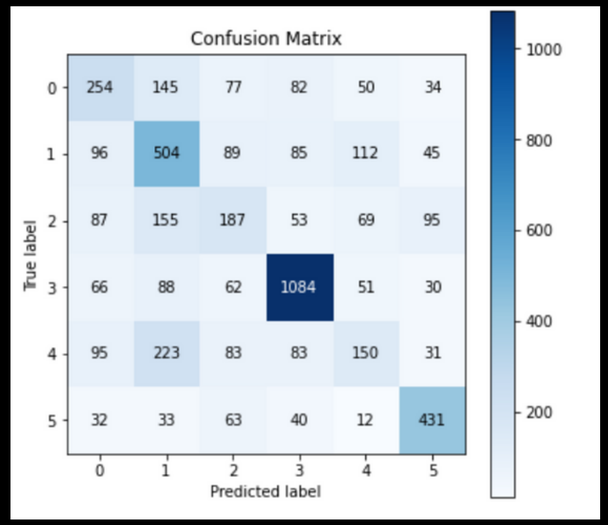



```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 46, 46, 64)        640
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 44, 44, 64)        36928
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 22, 22, 64)        0
_________________________________________________________________
dropout (Dropout)            (None, 22, 22, 64)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 20, 20, 64)        36928
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 18, 18, 64)        36928
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 9, 9, 64)          0
_________________________________________________________________
dropout_1 (Dropout)          (None, 9, 9, 64)          0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 7, 7, 128)         73856
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 5, 5, 128)         147584
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 2, 2, 128)         0
_________________________________________________________________
flatten (Flatten)            (None, 512)               0
_________________________________________________________________
dense (Dense)                (None, 1024)              525312
_________________________________________________________________
dropout_2 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_1 (Dense)              (None, 1024)              1049600
_________________________________________________________________
dropout_3 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_2 (Dense)              (None, 6)                 6150
=================================================================
Total params: 1,913,926
```

OUR RESULTS

EPOCH ==50; BATCH_SIZE == 500; TRAINING_TIME~4 HOURS



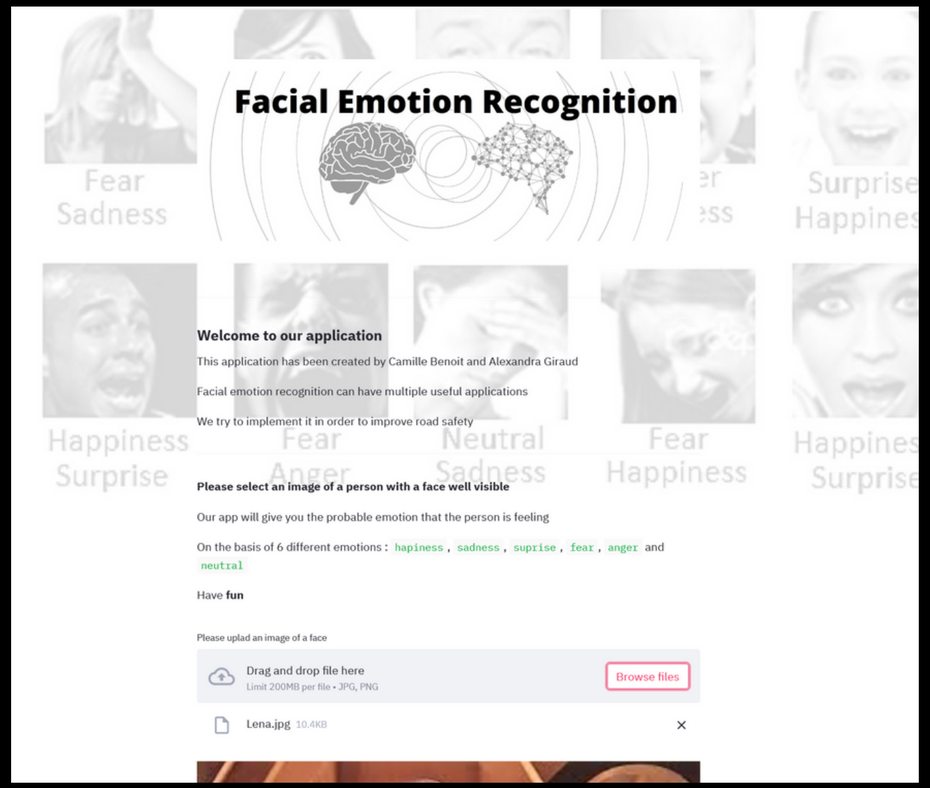


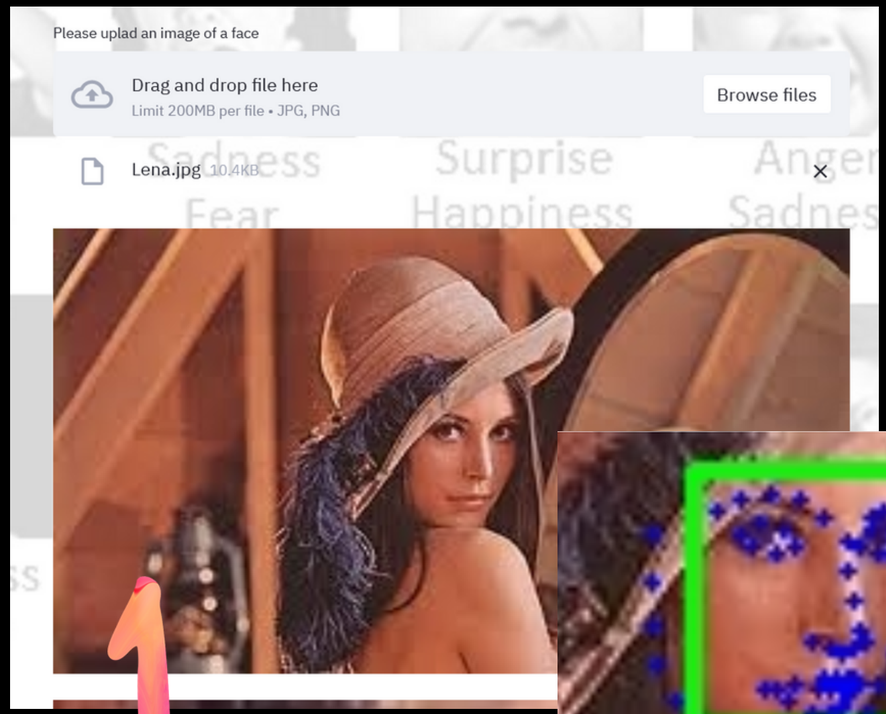
TRAINING_ACCURACY == 75% (QUITE GOOD FOR CNN AND DATASET)
TEST_ACCURACY == 55% (QUITE GOOD FOR CNN AND DATASET), BUT **OVERFIT**

OUR MODEL CONFUSES NEUTRAL, HAPPY AND SAD!
THE EMOTION WITH THE HIGHTEST RECOGNITION RATE IS HAPINESS

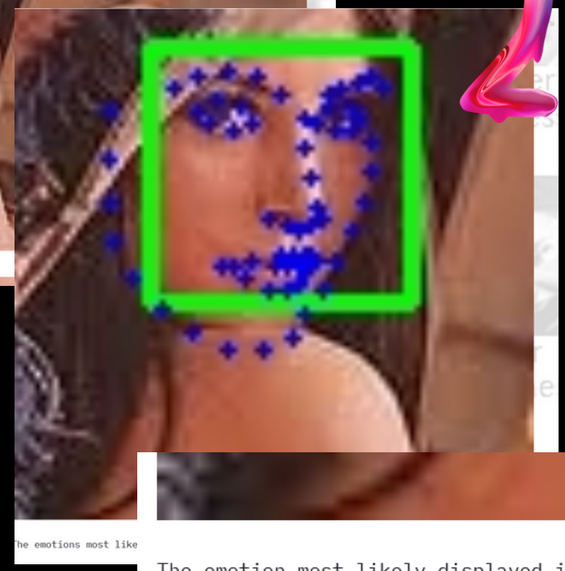
BUT CAN A HUMAN REALLY DO BETTER ?

Demonstration of the webApp

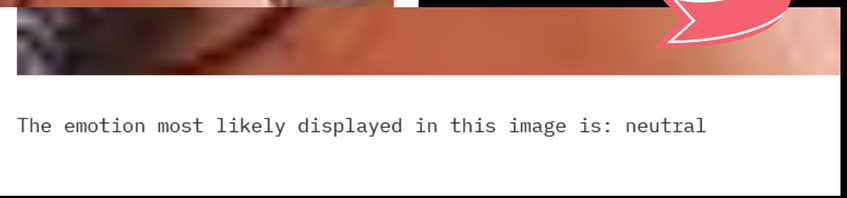




1



2



3

Streamlit

Problems & Further work

IMPROVE MODEL

Needs deeper and more complex architecture than basic CNN

Data augmentation / noise reduction

Transfert learning (VGG16 architecture and weights) (Started)

Landmarks with distance data (additional info)

Multiple People Analysis

Fine tuning the model

Sometimes there are fails in landmark detection

OTHER LIMITATIONS

Not any good dataset available in the word (very difficult to make, big enough for CNN)

"a model can only be as good as the data that it receives"

Results in the state of the art **don't exceed 78%** for 7 emotions

Ambiguity and **difficulty** related to the concept of emotions

THANK YOU !