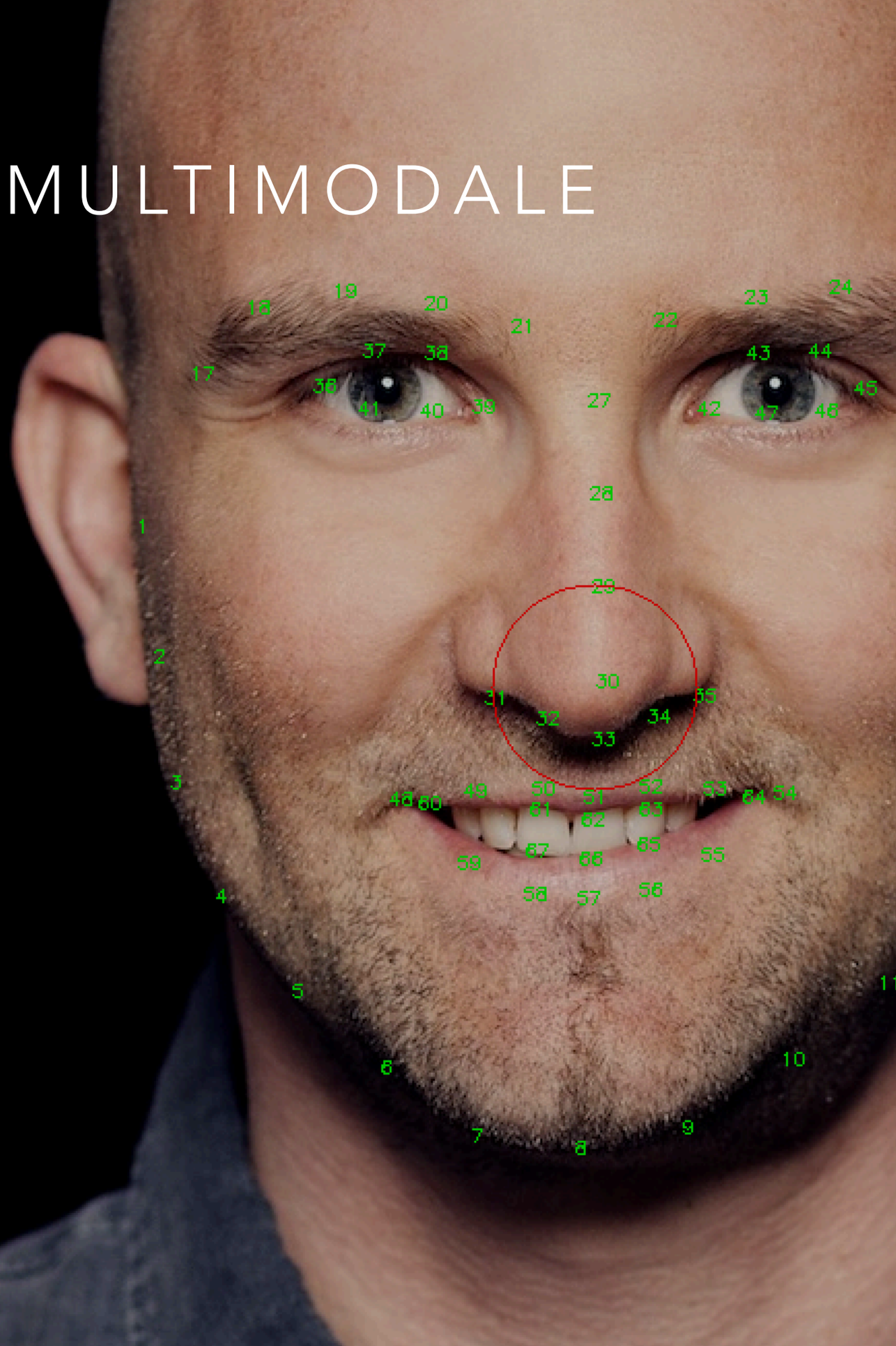
TELECOM PARISTECH - 2019

RECONNAISSANCE MULTIMODALE D'ÉMOTIONS



TELECOM ParisTech

X

pôle emploi

FIL ROUGE

CONTEXTE ET ENJEUX

CONTEXTE

- Informatique affective : *Etude et développement de systèmes ayant les capacités de reconnaître, d'exprimer, de synthétiser et modéliser les émotions humaines.*

- Usages pour le relation avec conseiller :
  - Assistance aux conseillers
  - Personnalisation et catégorisation des échanges

ENJEUX

- Tester et concevoir des solutions d'analyse des émotions :
  - Texte : Réponses à des questions en entretien
  - Audio : Enregistrements d'échanges sur une plateforme
  - Vidéo : Enregistrements d'échanges sur une plateforme
- Proposer une plateforme d'entraînement pour les candidats : réponse à des questions, entretien vidéo

DONNÉES

- Entrainement des modèles à partir d'Open Data :
  - *Texte : Pennebaker and King daily essays (1999)*
  - *Audio : The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)*
  - *Video : Facial Emotion Recognition 2013, Kaggle*

APPROCHE

- Mesure des traits BigFive pour le texte
- Mesure des émotions pour la vidéo et le son
- Création d'un modèle d'ensemble et développement de métriques

Table 1. Categorical target depending on the input data type.

Data Type	Categorical target
Textual	Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism
Sound	Happy, Sad, Angry, Fearful, Surprise, Neutral and Disgust
Video	Happy, Sad, Angry, Fearful, Surprise, Neutral and Disgust

FIL ROUGE

RESULTATS

QUELQUES CHIFFRES

1

WebApp

4

Contributeurs

8

Forks

261

Commits

3’864

Lignes de code Python

TECHNOLOGIES

TensorFlow

Keras

Flask

CSS

HTML

UN PROJET OPEN-SOURCE

GitHub

https://github.com/maelfabien/Multimodal-Emotion-Recognition

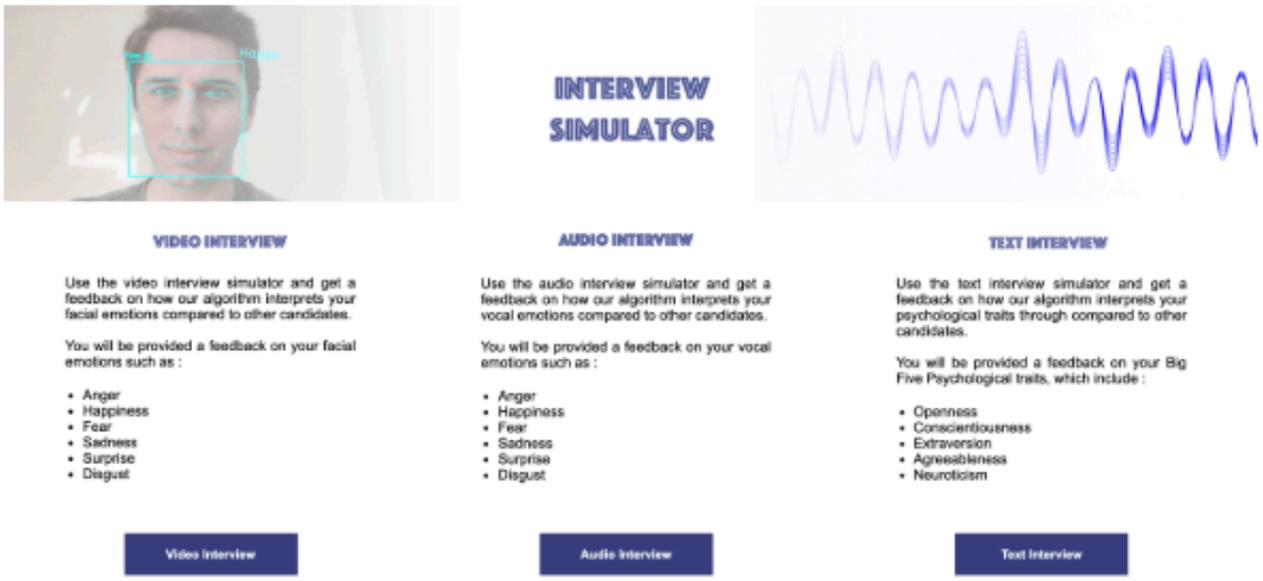
Real-Time Multimodal Emotion Recognition

Followers 23 | contributors 4 | commit activity invalid | python 3.2 | 3.3 | 3.4

Don't hesitate to ⭐ the repo if you enjoy our work !

In a nutshell

We developped a multimodal emotion recognition platform to analyze the emotions of job candidates, in partnership with the French Employment Agency. We analye facial, vocal and textual emotions, using mostly deep learning based approaches. We deployed a web app using Flask :



The tool can be accessed here : https://emotion.onrender.com/

We have also written a paper on our work : https://www.overleaf.com/read/xvtrrfpvzwhf

Table of Content :

- I. Context
- II. Data Sources
- III. Downloads
- IV. Methodology
  - a. Text Processing
  - b. Audio Processing
  - c. Video Processing
  - d. Ensemble Model
- V. How to use it ?
- VI. Research Paper
- VII. Deployment and WebApp
- VIII. Deployment

RÉSULTATS

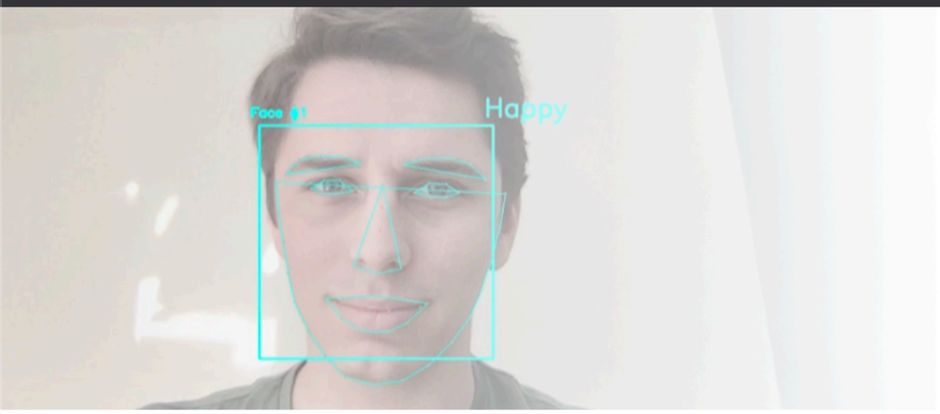
Texte	Audio	Video
45%	**75%**	**65%**

TEXTE

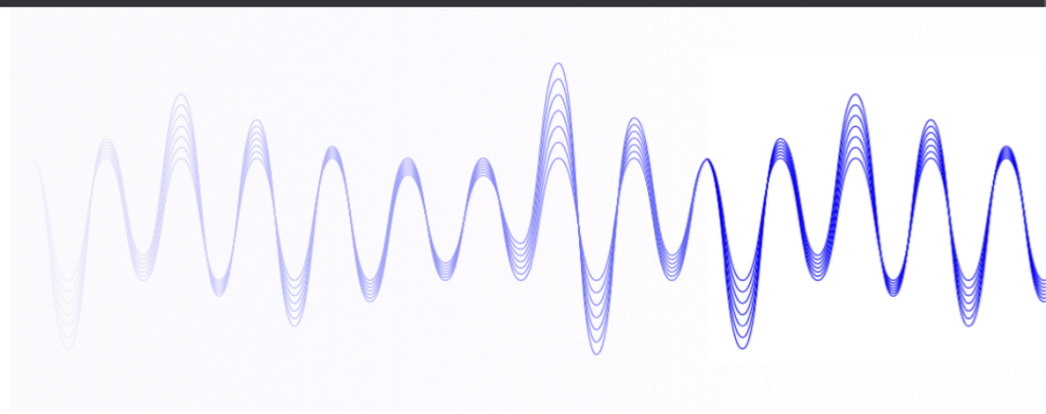
- Identifier des traits de personnalité (Big5) à partir de réponses à des questions spécifiques
- Natural Language Processing :
  - Pré-Traitement
  - Embedding
  - Classification

Chrome Fichier Modifier Afficher Historique Favoris Personnes Fenêtre Aide 100% Sam. 12:05

127.0.0.1:5000/text



TEXT INTERVIEW



Tell us about the last time you showed leadership.

Start Analysis

Or upload your Cover Letter :

Choisir un fichier Aucun fichier choisi

Start Analysis

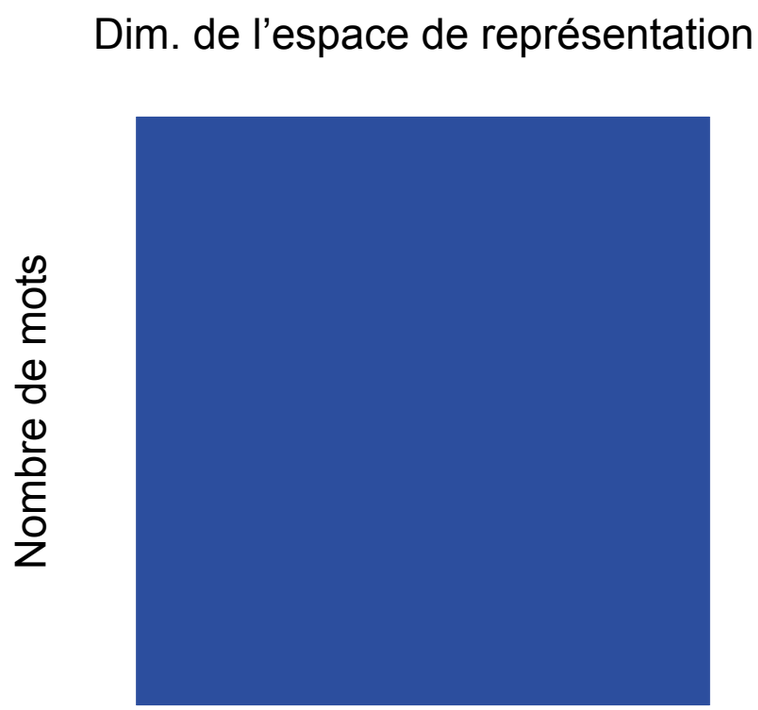
How does it work ?

[« Je me suis préparé pendant plusieurs mois à courir un marathon. C’était un rêve que … » **]**

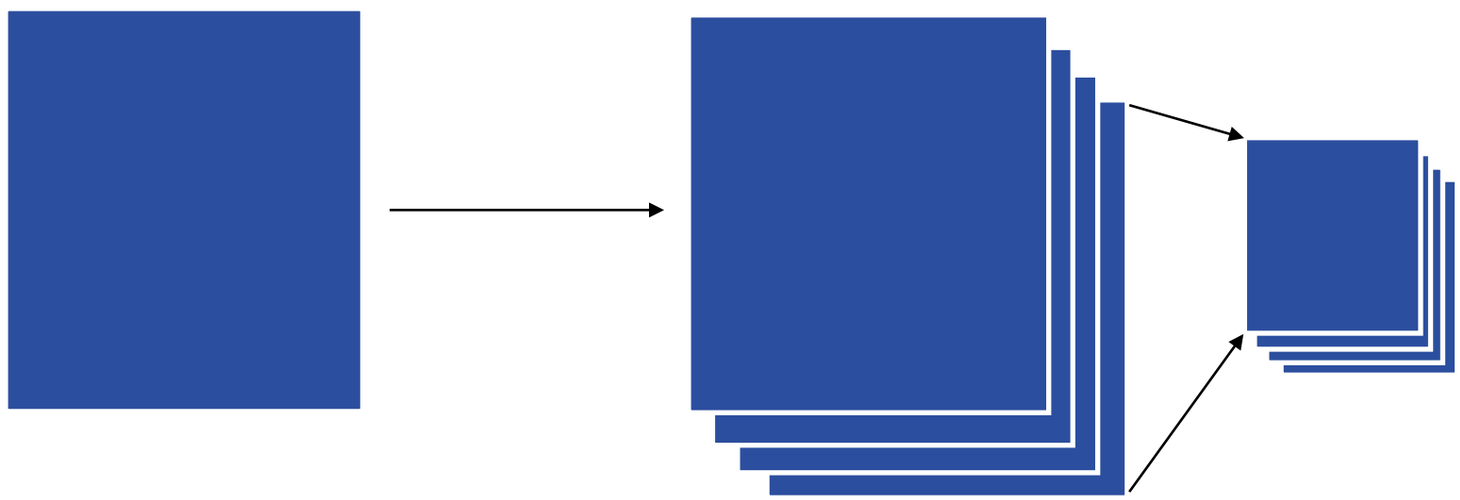
1. Extraire le texte

[« je », « être », « préparer », « plusieurs », « mois », « courir », « marathon » … **]**

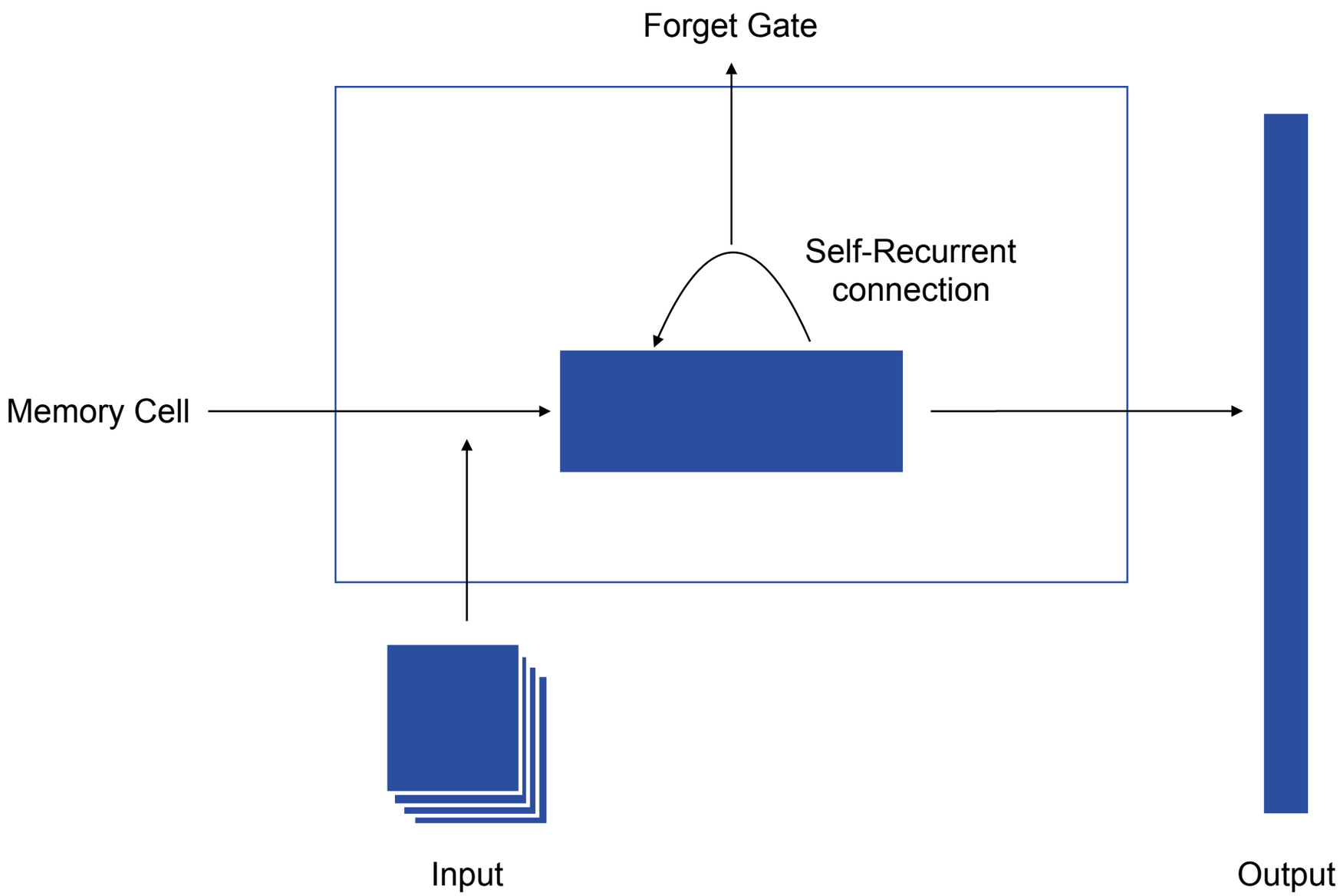
2. Pré-traitement



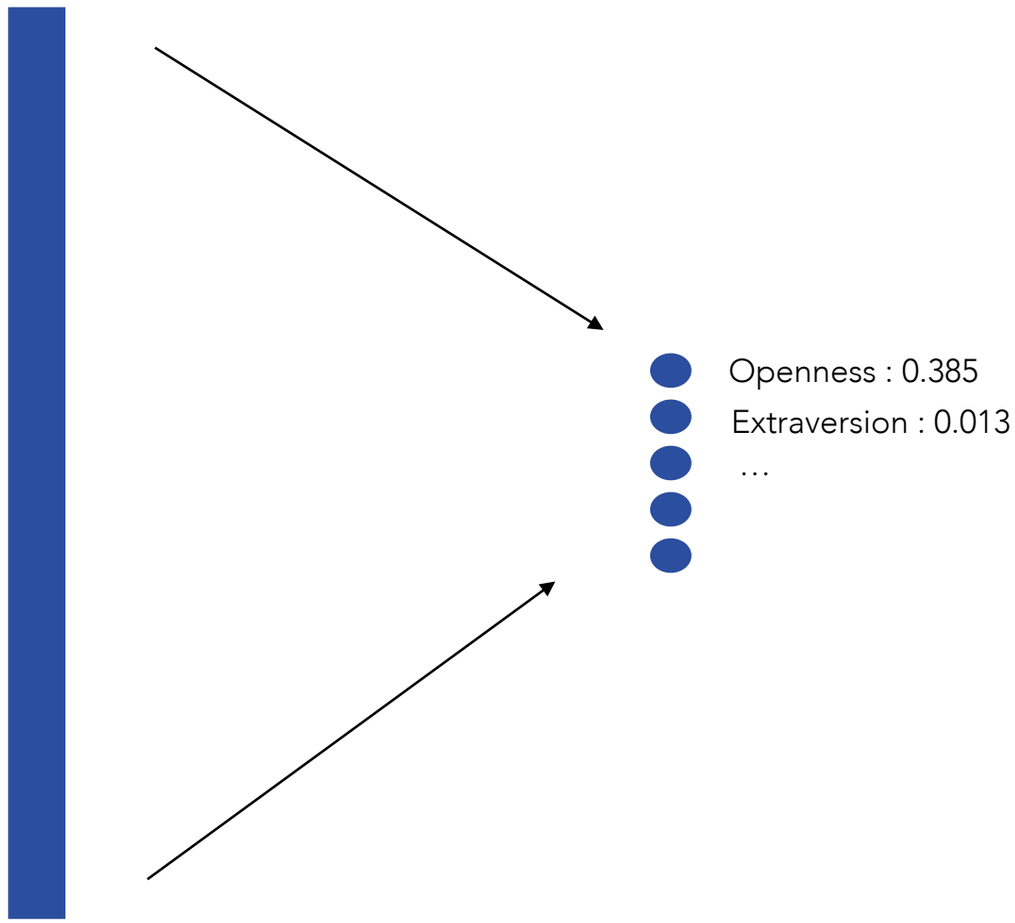
3. Word2Vec



4. Convolution, Pooling


Forget Gate
Self-Recurrent
connection
Memory Cell
Input
Output

5. LSTM

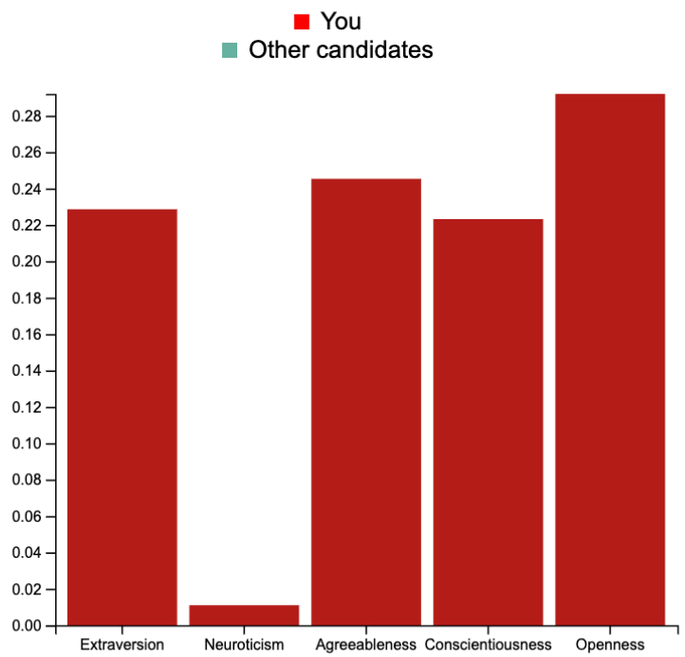


6. Dense network

TEXTE

- Difficultés:
  - Très nombreuses options de **pré-traitement des données textuelles**
  - Corpus d’entrainement de relativement petite taille
- Ajustements:
  - Agrégation de différente bases de données

Perceived Psychological Traits



Your most visible trait is :

Openness

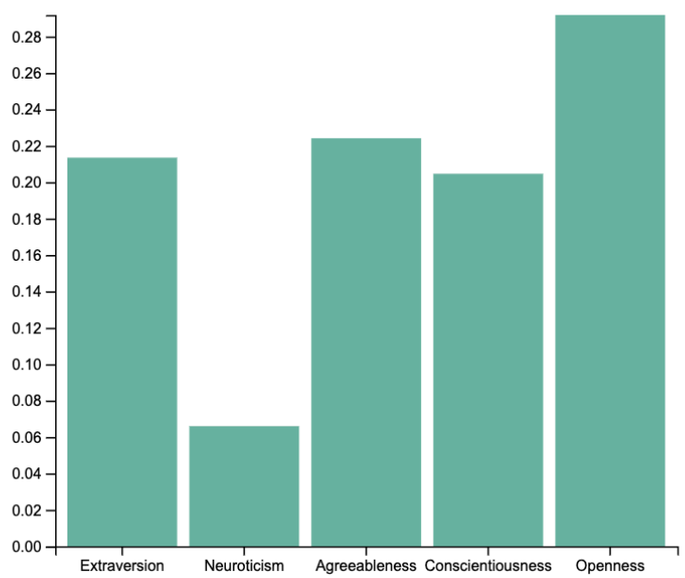
Psychological Traits :

- Extraversion : 22%
- Neuroticism : 1%
- Agreeableness : 24%
- Conscientiousness : 22%
- Openness : 29%

Most common words :

- to
- ;
- ’
- t
- you
- your
- easy
- The
- doesn
- turn
- don
- So
- come
- worth
- having

Other candidates



Their most visible trait is :

Openness

Psychological Traits :

- Extraversion : 21%
- Neuroticism : 6%
- Agreeableness : 22%
- Conscientiousness : 20%
- Openness : 29%

Most common words :

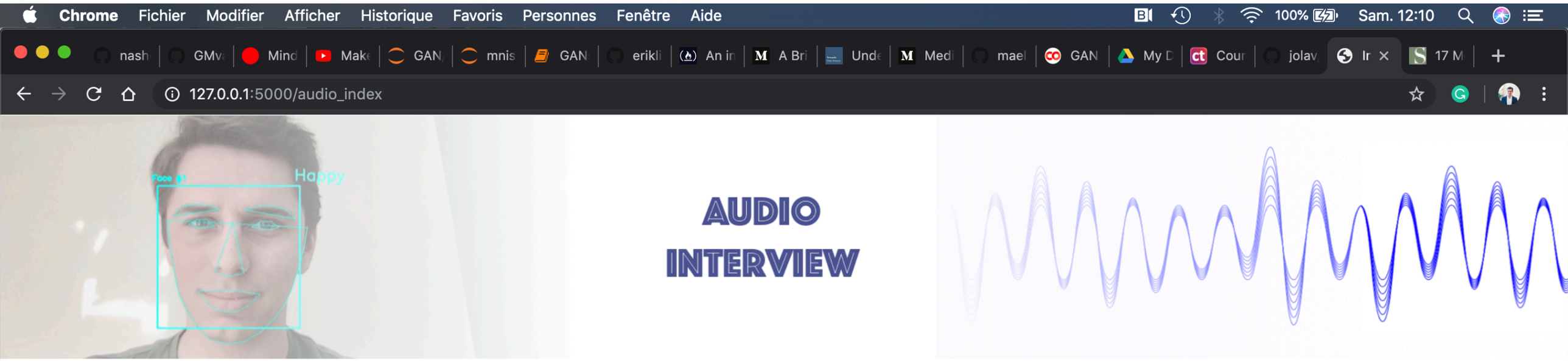
- .
- I
- to
- and
- a
- in
- the
- of
- for
- have
- Hi
- my
- ’
- :
- Hello

AUDIO

- Idée: identifier des émotions dans la voix.

- Méthode:
  - **Extraction** et **discrétisation** du signal audio
  - **Pré-traitement** du signal
  - **Classification** des émotions

- Modèle:
  - Variable: *Spectrogram*
  - Classifieur: *CNN+LSTM*



Tell us about the last time you showed leadership.

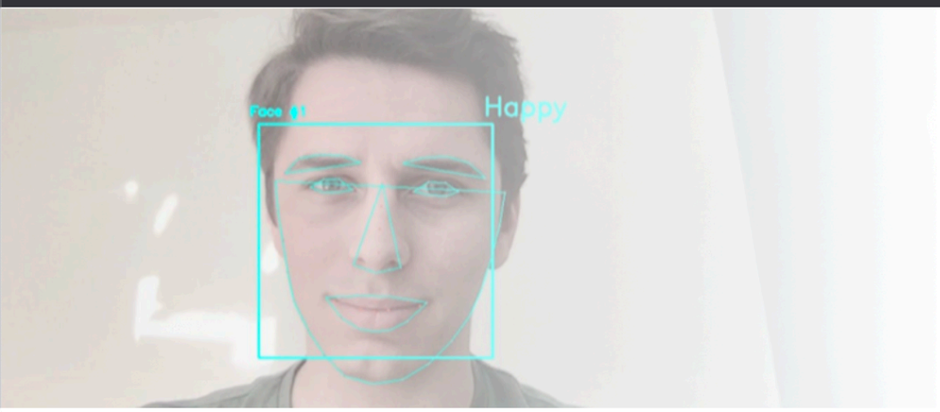
Start Recording

How does it work ?

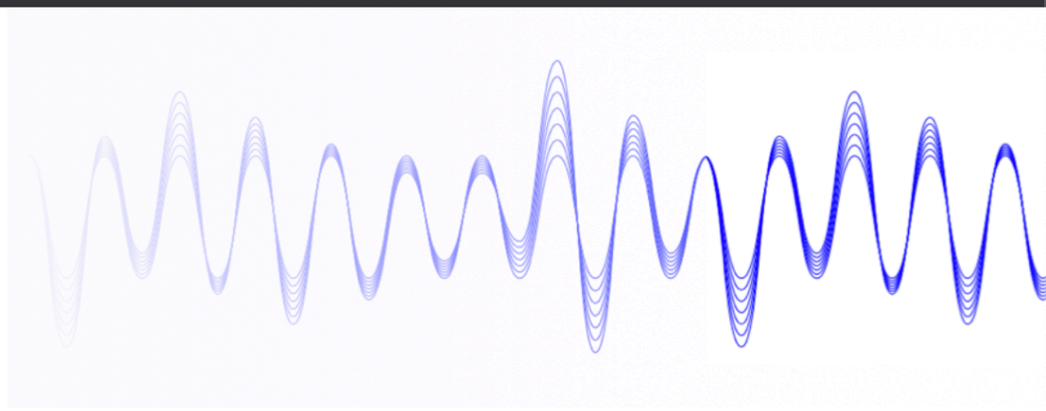
Back

Chrome Fichier Modifier Afficher Historique Favoris Personnes Fenêtre Aide 100% Sam. 12:11

127.0.0.1:5000/audio_recording



AUDIO INTERVIEW



Tell us about the last time you showed leadership.

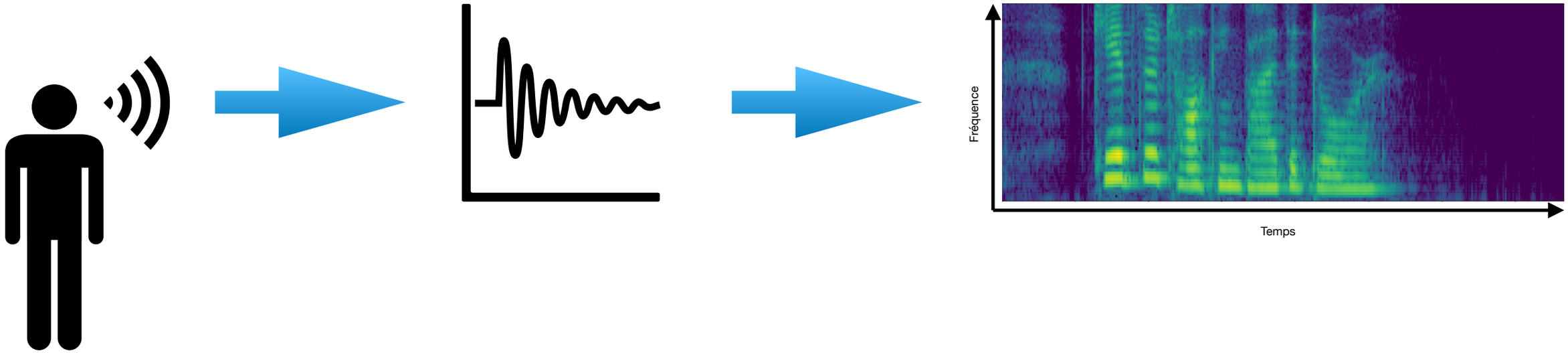
Start Recording

The recording is over! You now have the opportunity to do an analysis of your emotions. If you wish, you can also choose to record yourself again. ✖

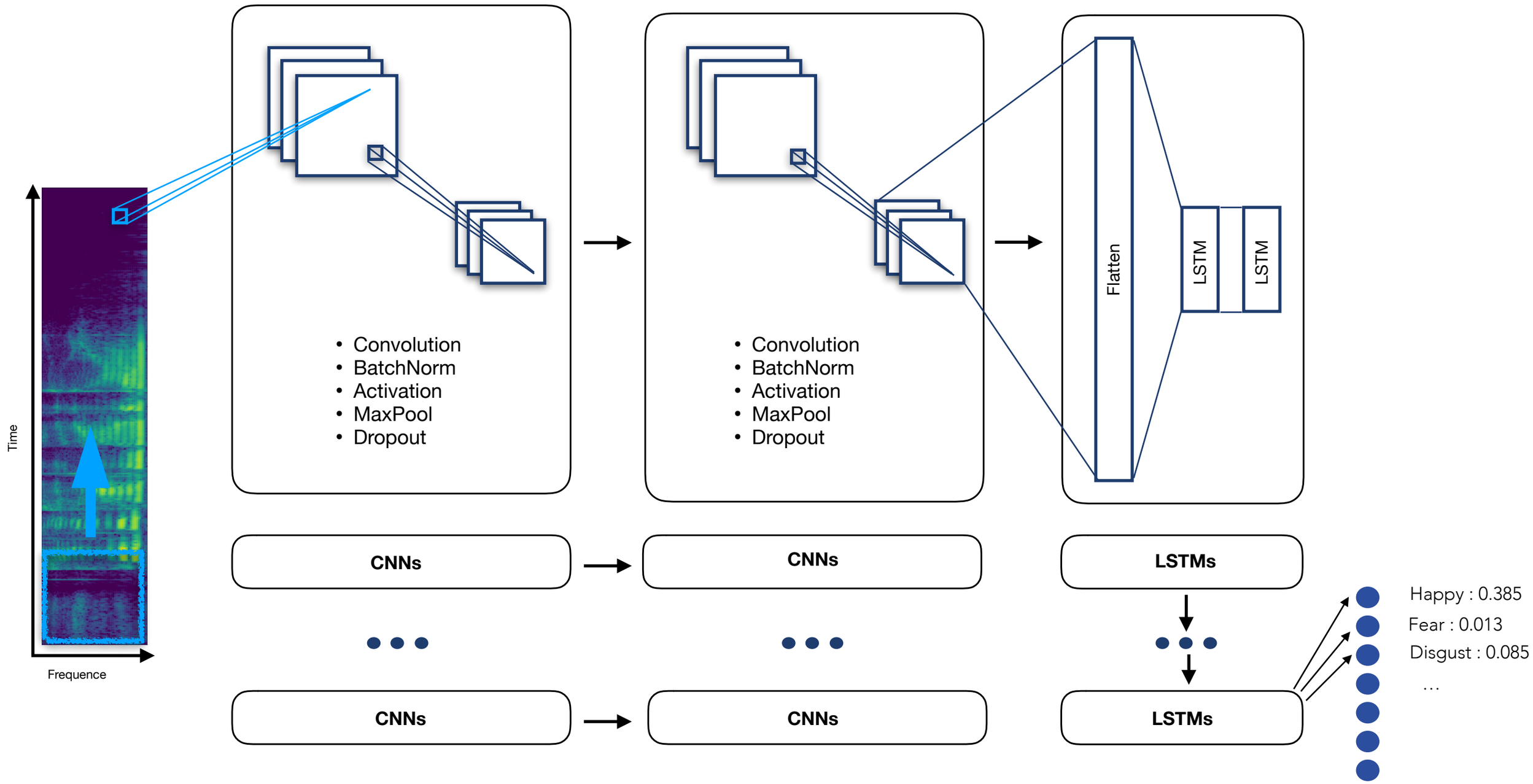
Get Emotion Analysis

How does it work ?

Back



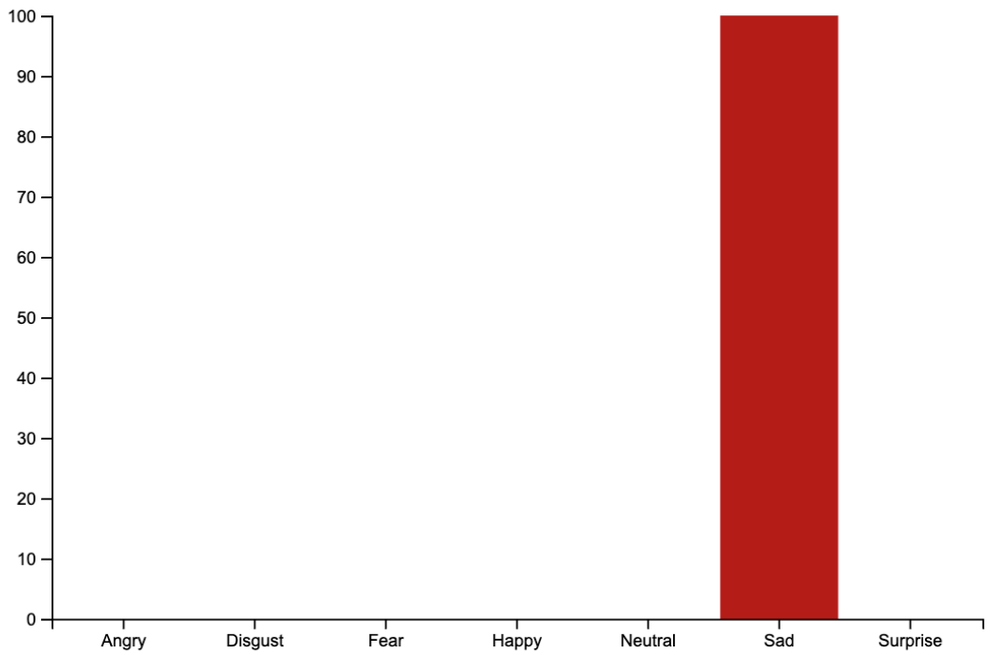
1. Extraction du signal et du Spectrogram


Time
Frequence
• Convolution
• BatchNorm
• Activation
• MaxPool
• Dropout
• Convolution
• BatchNorm
• Activation
• MaxPool
• Dropout
Flatten
LSTM
LSTM
CNNs
CNNs
LSTMs
CNNs
CNNs
LSTMs
Happy : 0.385
Fear : 0.013
Disgust : 0.085
...

2. CNNs + LSTMs

AUDIO

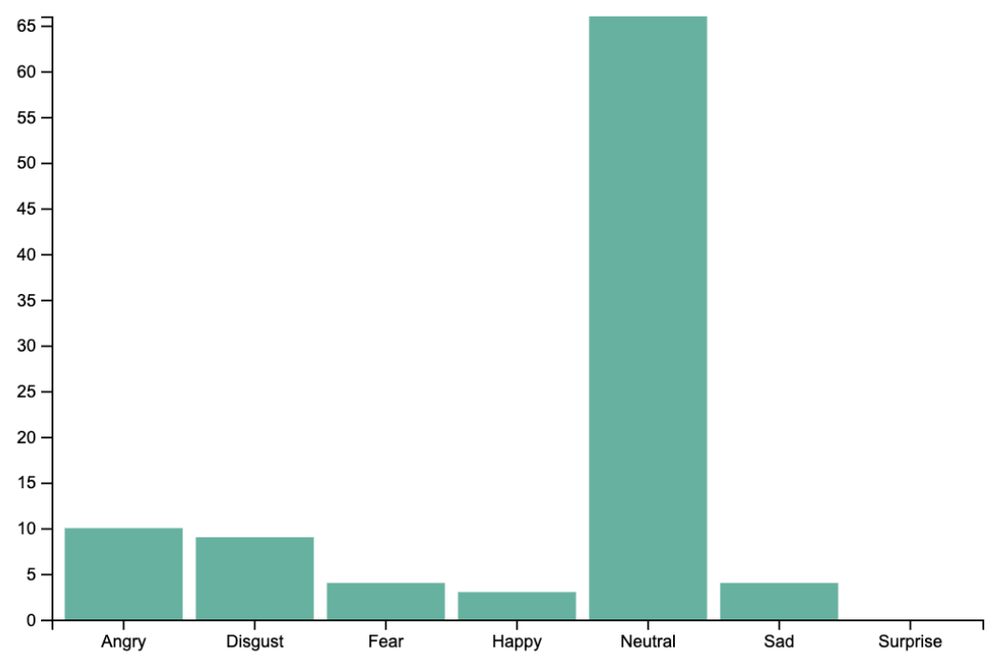
- Difficultés:
  - **Spectrogram** le plus « propre » possible
  - **Architecture** la plus performante
- Ajustements:
  - Suppression des fréquences **> 4k Hz**
  - **Augmentation** des données d'apprentissage
  - Distinction de **genre**



Your most frequent emotion is : **Sad**

- Angry : 0%
- Disgust : 0%
- Fear : 0%
- Happy : 0%
- Neutral : 0%
- Sad : 100%
- Surprise : 0%

Other candidates



Other candidates most frequent emotion is : **Neutral**

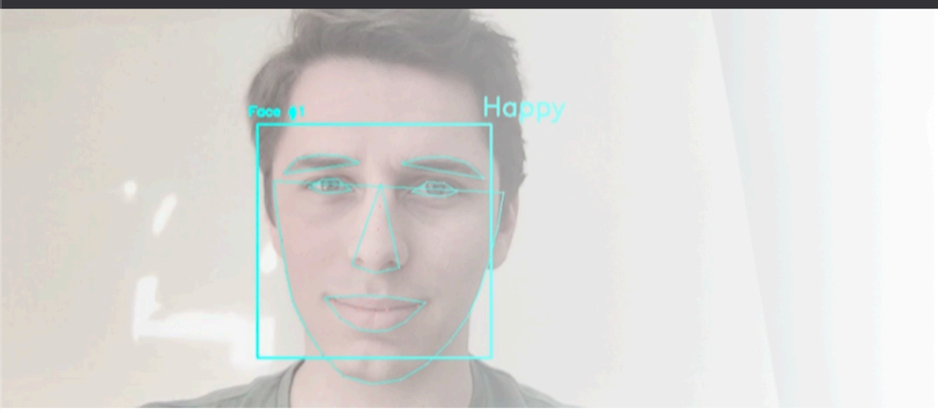
- Angry : 10%
- Disgust : 9%
- Fear : 4%
- Happy : 3%
- Neutral : 66%
- Sad : 4%
- Surprise : 0%

VIDEO

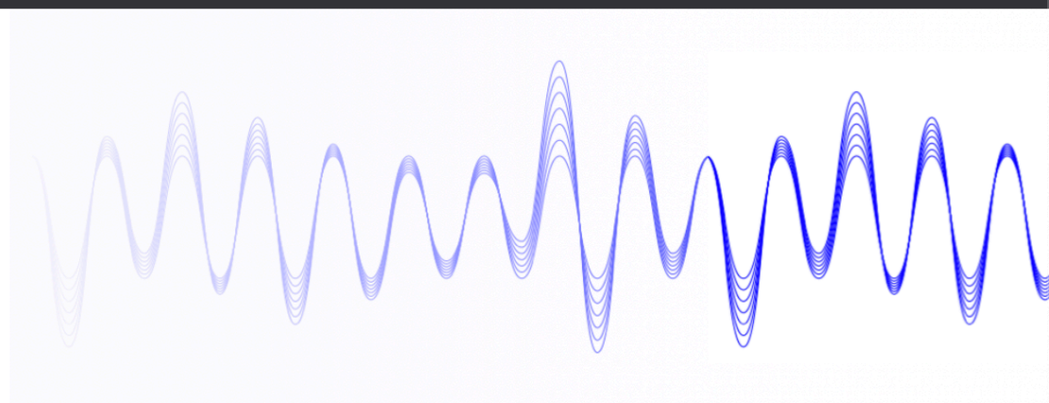
- Identifier des émotions à partir d'une vidéo
- Traitement de l'image :
  - Pré-Traitement de l'image : Identification, extraction du visage
  - Filtres manuels + SVM
  - CNN, XCeption

Chrome Fichier Modifier Afficher Historique Favoris Personnes Fenêtre Aide 100% Sam. 12:15

127.0.0.1:5000/video



VIDEO INTERVIEW



Tell us about the last time you showed leadership.

Start Recording

You will have 45 seconds to discuss the topic mentioned above. Due to restrictions, we are not able to redirect you once the video is over. Please move your URL to /dash instead of /video_1 once over. You will be able to see your results then. ✖

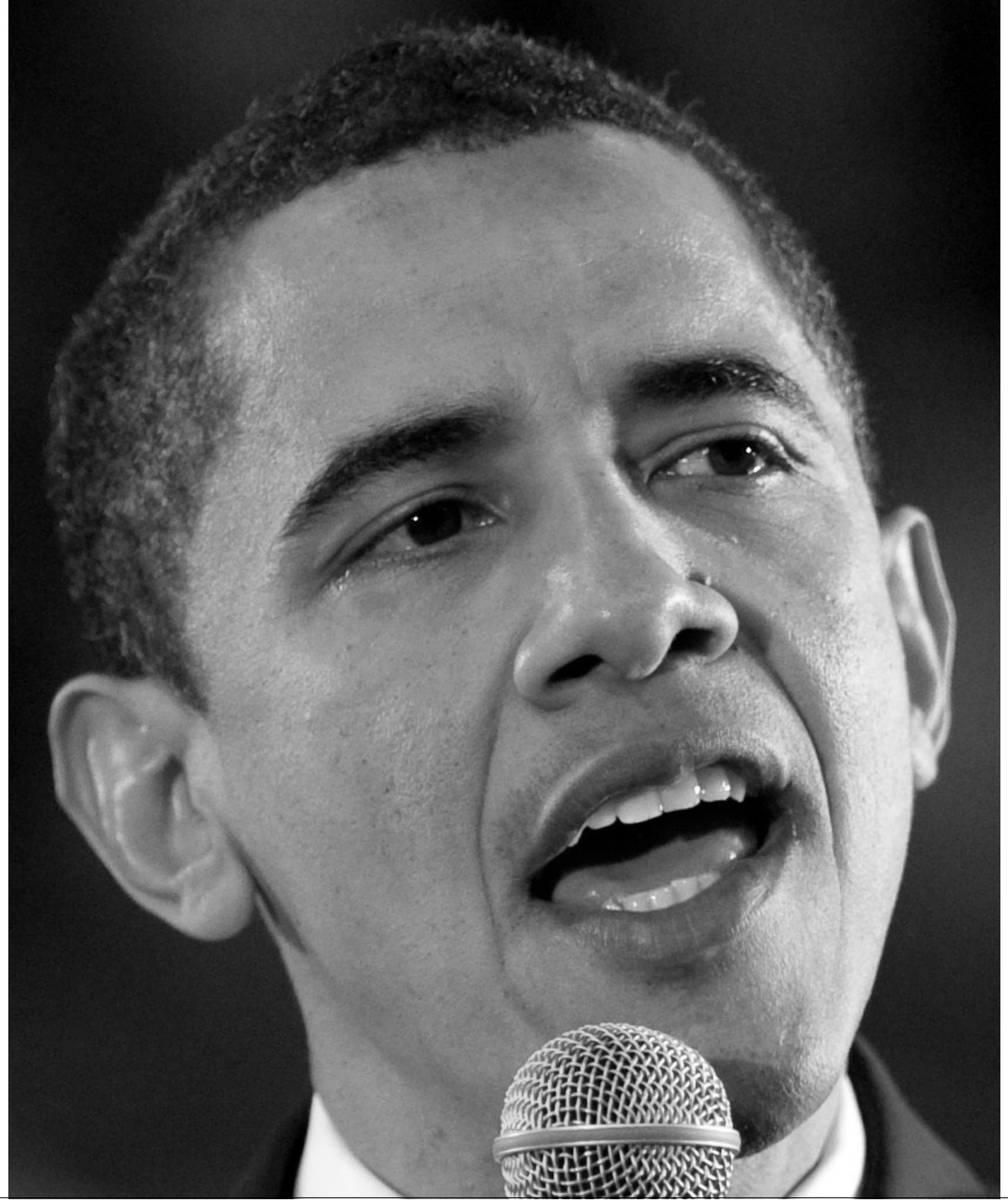
How does it work ?

Back

1. Extraire chaque image

2. Appliquer un filtre gris

3. Identifier un visage (Histogramme de Gradients Orientés)



4. Zoomer sur le visage

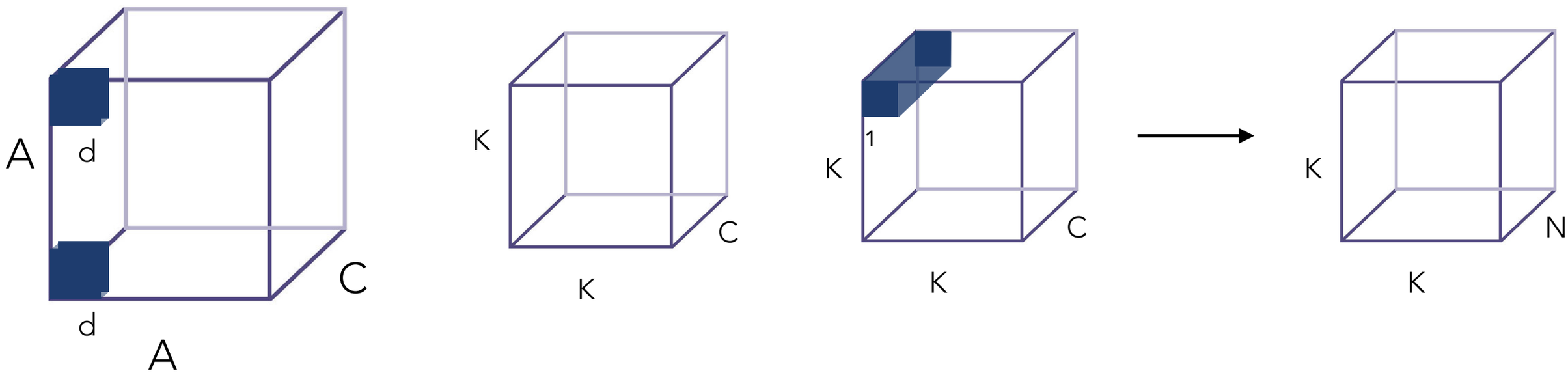
5. Réduire la dimension

DepthWise Separable Convolutions

Approche en 2 étapes :

Etape 1 : Depthwise convolution

Etape 2 : Pointwise convolution

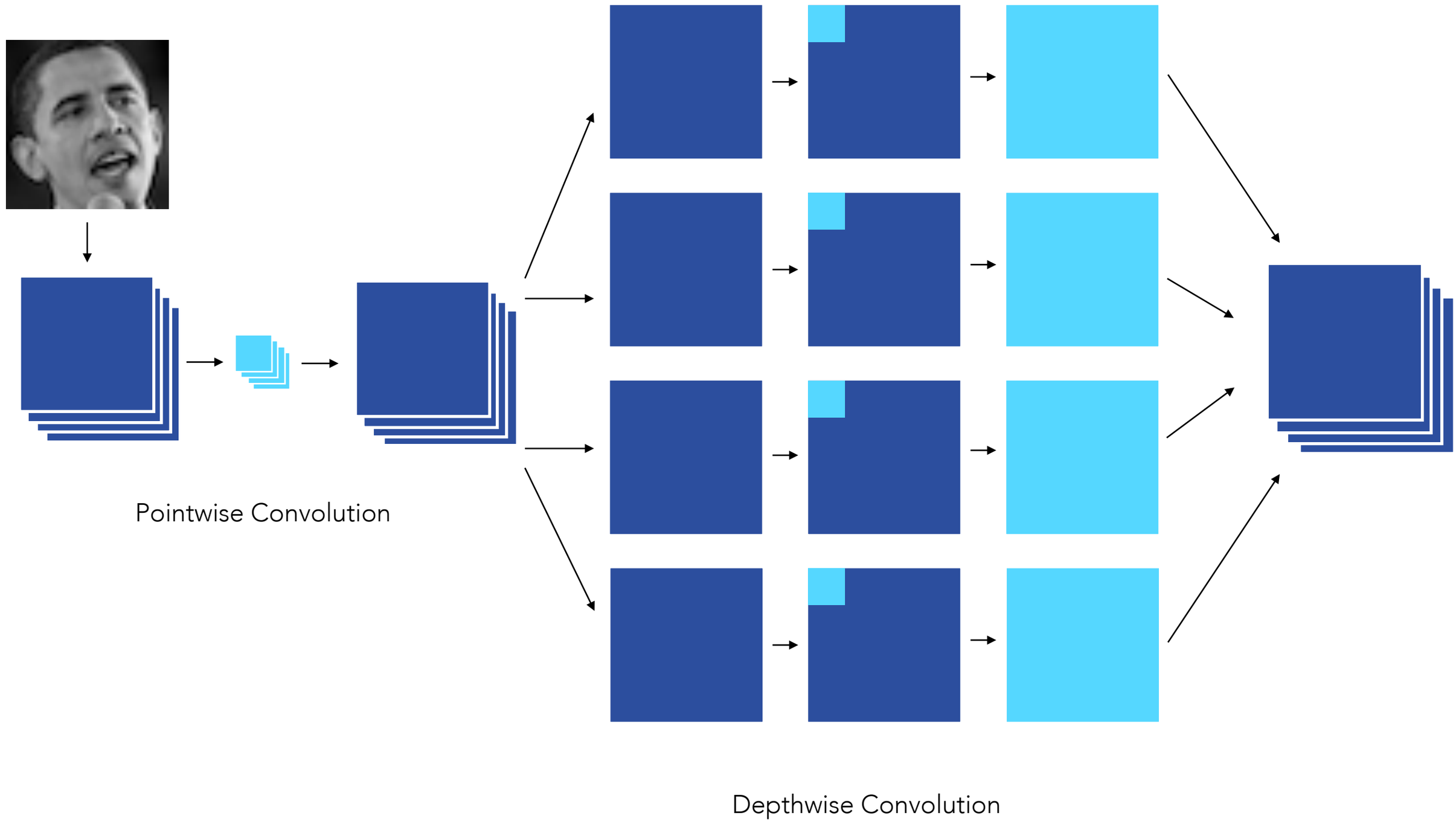


Permet de réduire le nombre de paramètres et accélérer le temps d'entrainement

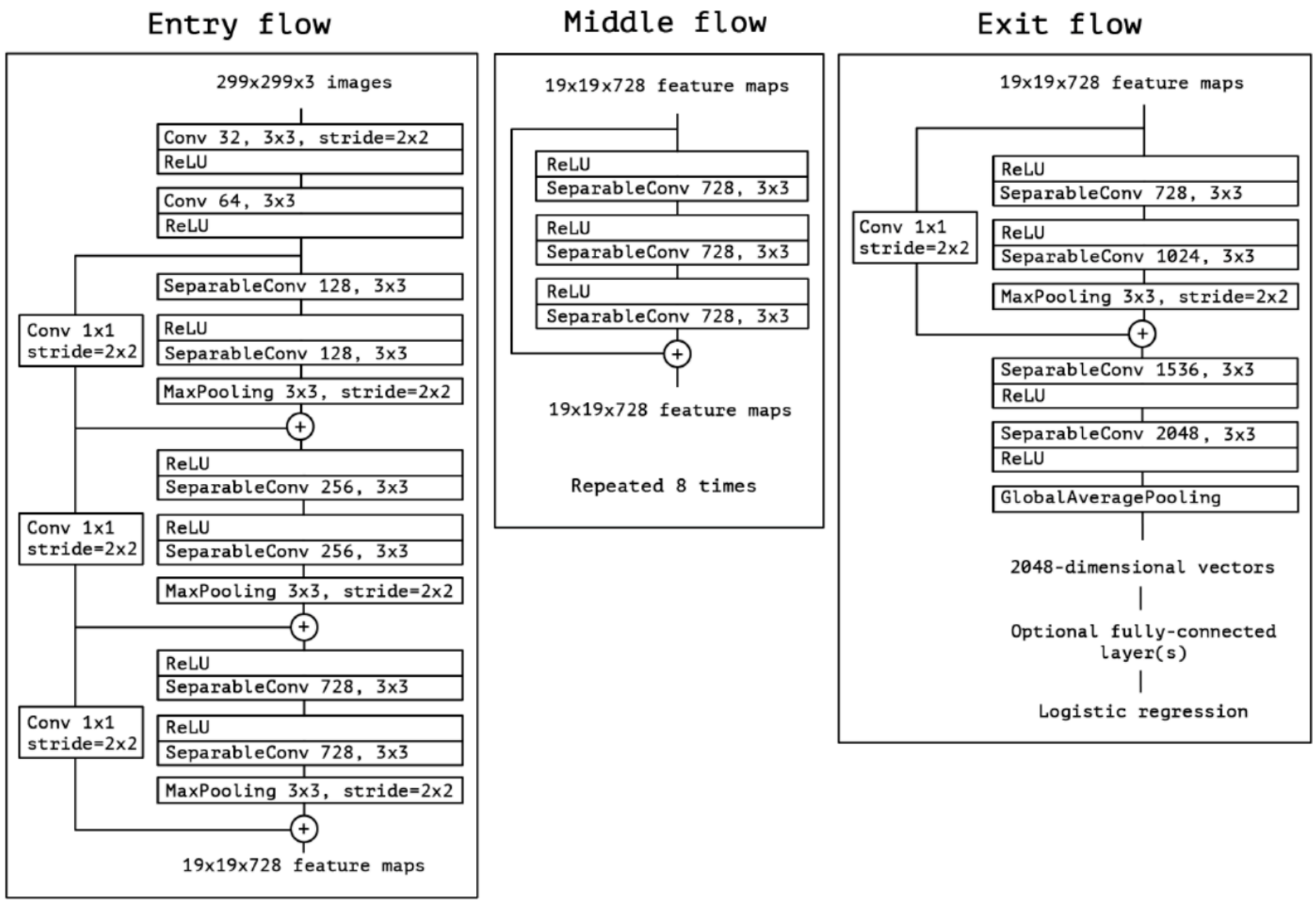
± 2h sur GPU

---

6. Appliquer des filtres convolutifs (XCeption)

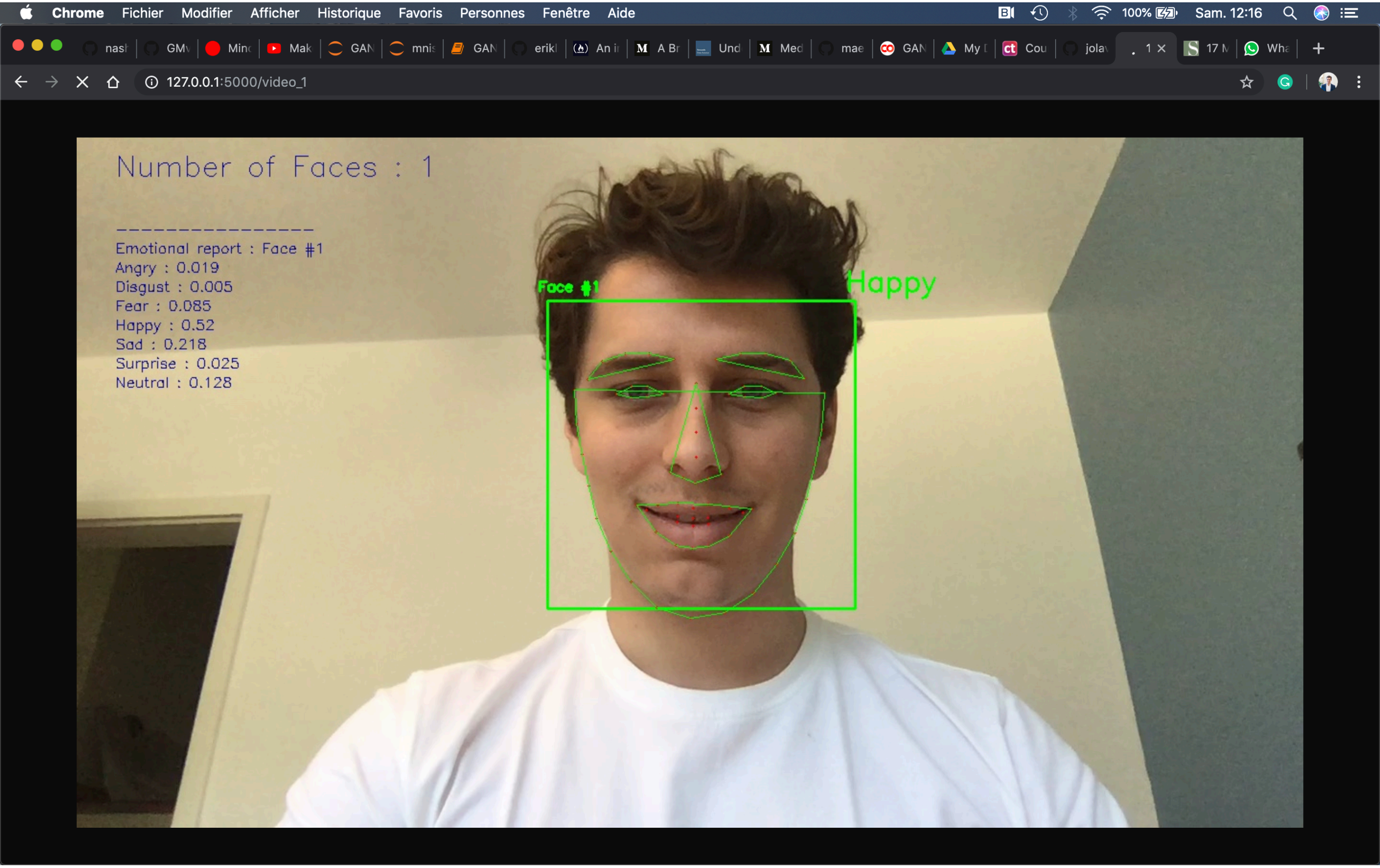


6. Appliquer des filtres convolutifs (DenseNet, Inception, ResNet, XCeption…)

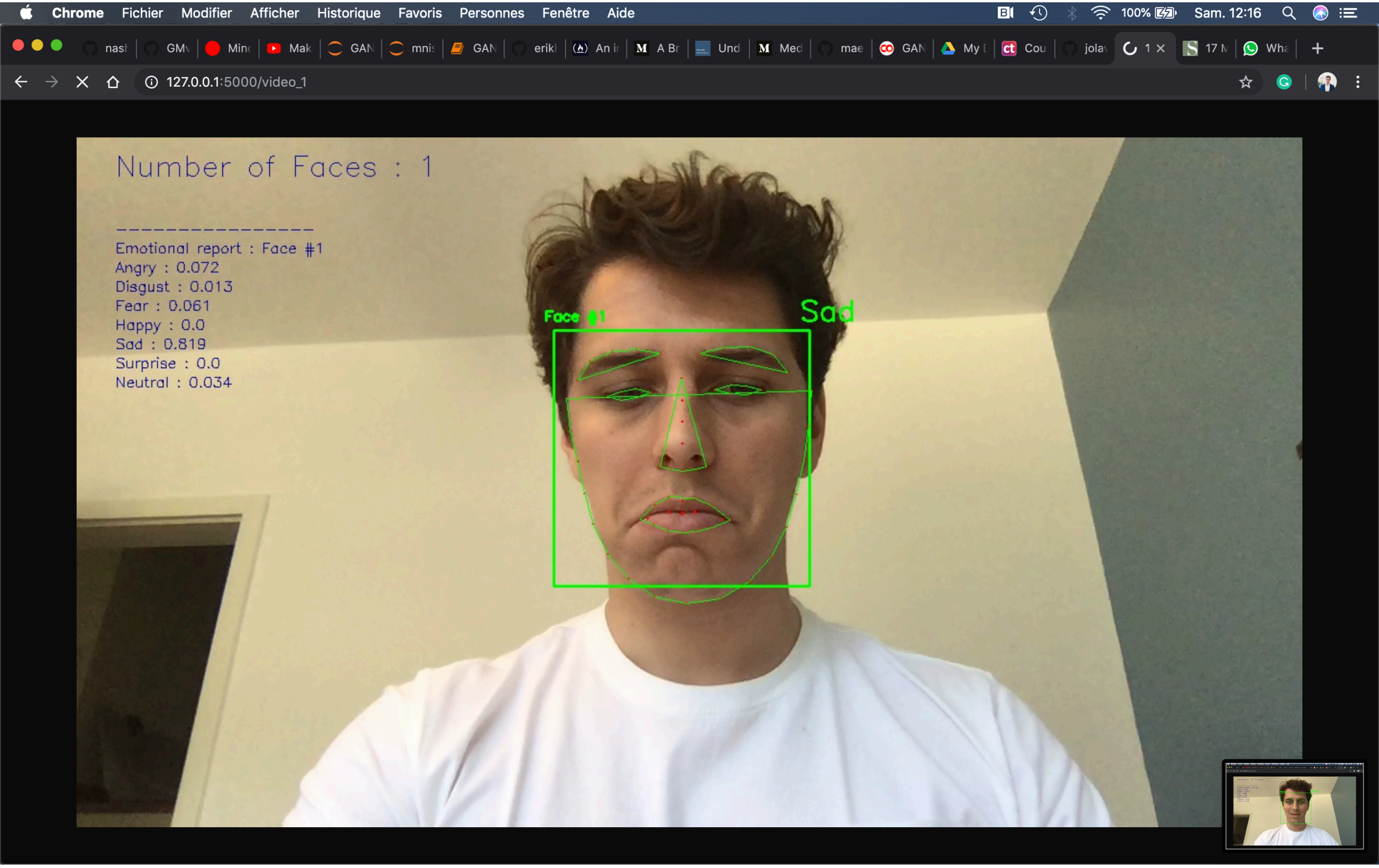


Overall Architecture of Xception (Entry Flow > Middle Flow > Exit Flow)

7. Structure du modèle


Chrome Fichier Modifier Afficher Historique Favoris Personnes Fenêtre Aide
100% Sam. 12:16
127.0.0.1:5000/video_1
Number of Faces : 1

Emotional report : Face #1
Angry : 0.019
Disgust : 0.005
Fear : 0.085
Happy : 0.52
Sad : 0.218
Surprise : 0.025
Neutral : 0.128
Face #1
Happy

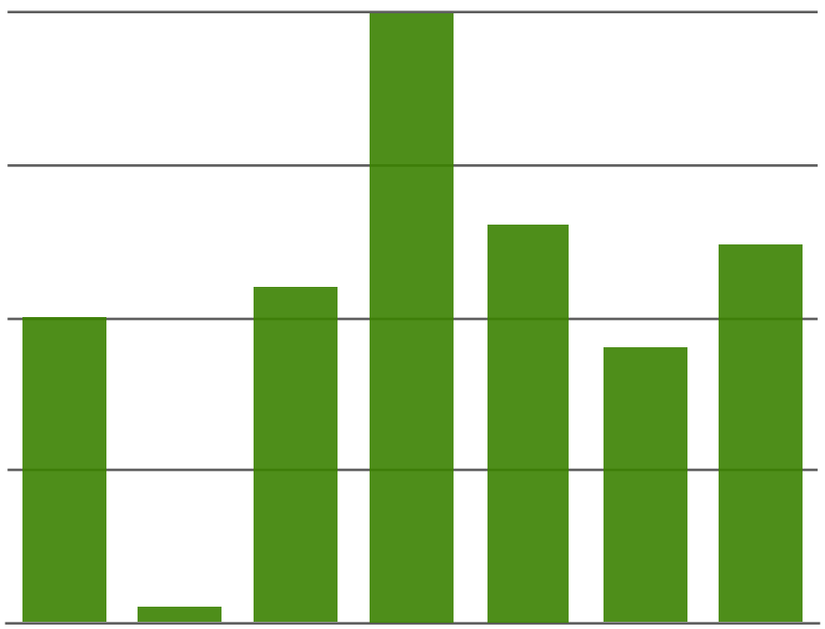

Chrome Fichier Modifier Afficher Historique Favoris Personnes Fenêtre Aide
100% Sam. 12:16
127.0.0.1:5000/video_1
Number of Faces : 1

Emotional report : Face #1
Angry : 0.072
Disgust : 0.013
Fear : 0.061
Happy : 0.0
Sad : 0.819
Surprise : 0.0
Neutral : 0.034
Face #1
Sad

VIDEO

- Difficultés:
  - **Label des données** d'entrainement
  - **Passage à l'échelle** (images/secondes)

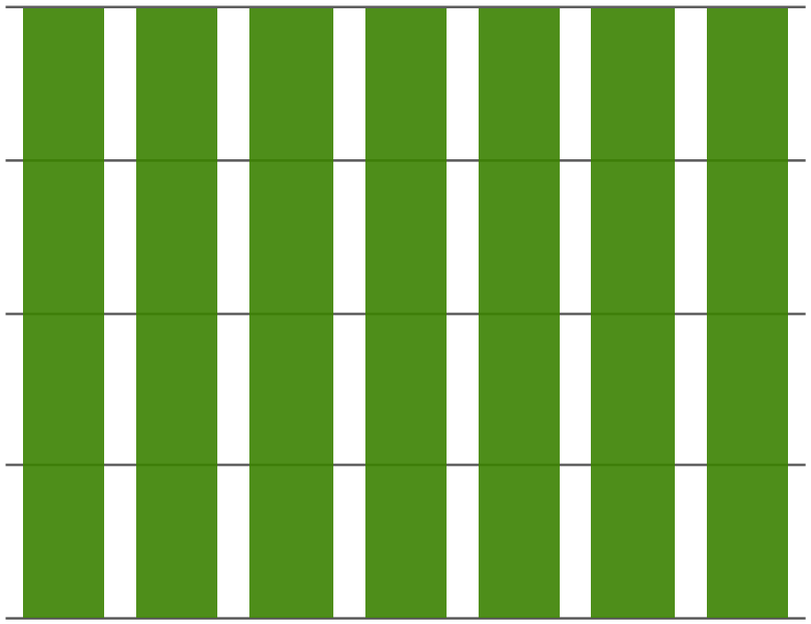
- Ajustements:
  - Data Augmentation
  - Early Stopping
  - Réduction du « Learning Rate » sur plateau
  - L2-Regularization
  - Class Weight



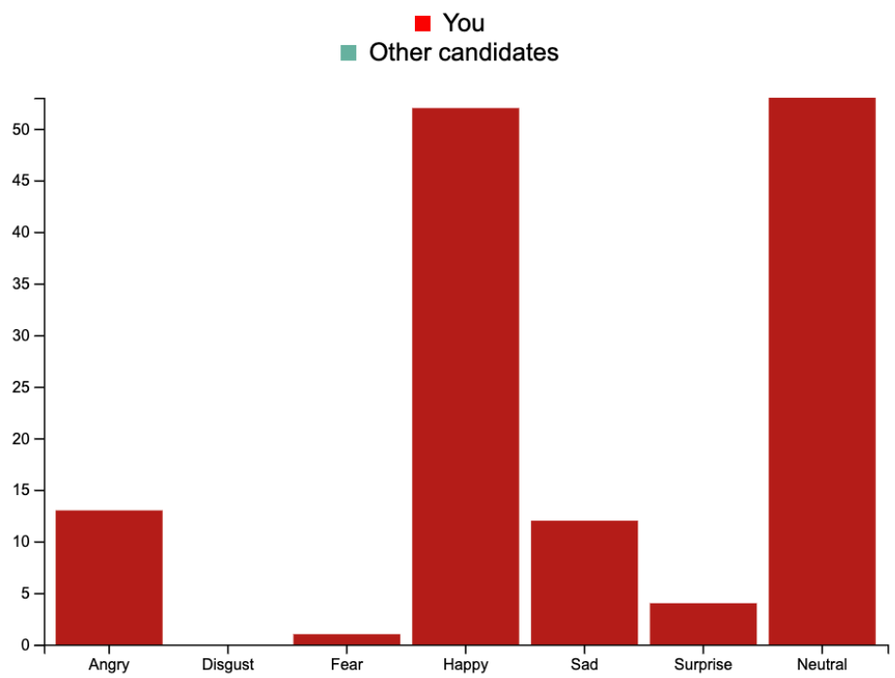
VIDEO

- Difficultés:
  - **Label des données** d'entrainement
  - **Passage à l'échelle** (images/secondes)

- Ajustements:
  - Data Augmentation
  - Early Stopping
  - Réduction du « Learning Rate » sur plateau
  - L2-Regularization
  - Class Weight



Perceived emotions



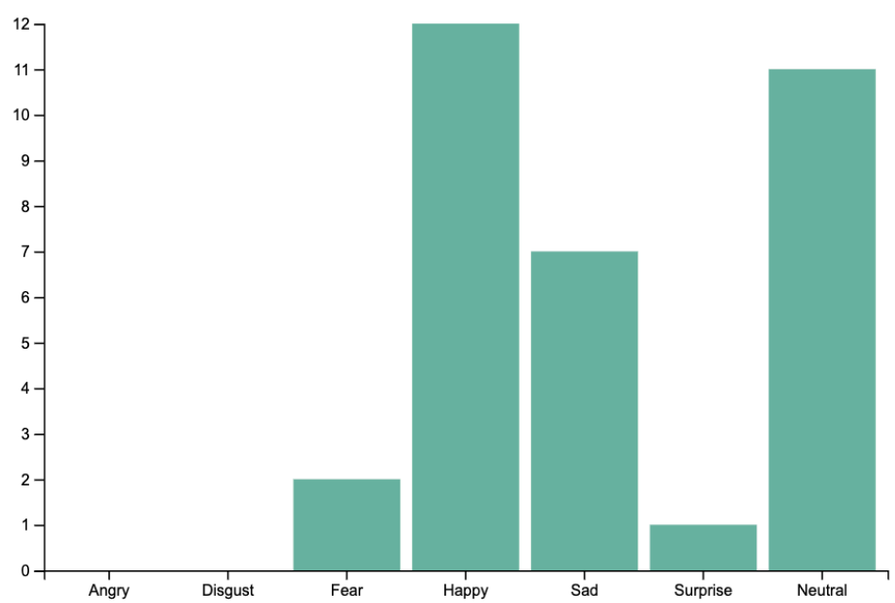
Facial Emotions

Your most frequent emotion is :

Neutral

- Anger : 0%
- Disgust : 0%
- Fear : 0%
- Happiness : 24%
- Sadness : 36%
- Surprise : 0%
- Neutrality : 39%

Other candidates



Other candidates most frequent emotion is :

Happy

- Angry : 7%
- Disgust : 0%
- Fear : 1%
- Happy : 38%
- Neutral : 11%
- Sad : 2%
- Surprise : 38%

FIL ROUGE

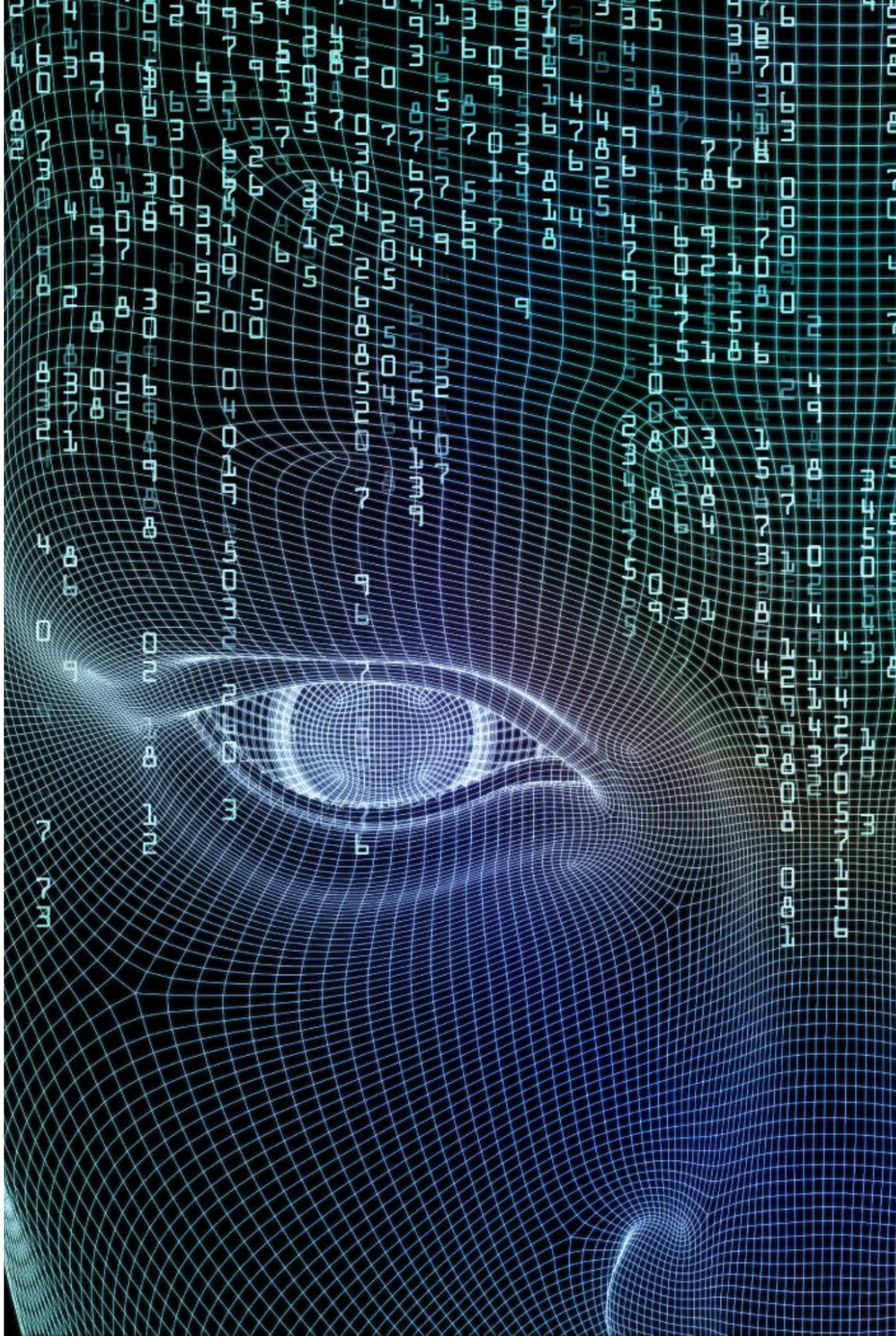
PERSPECTIVES

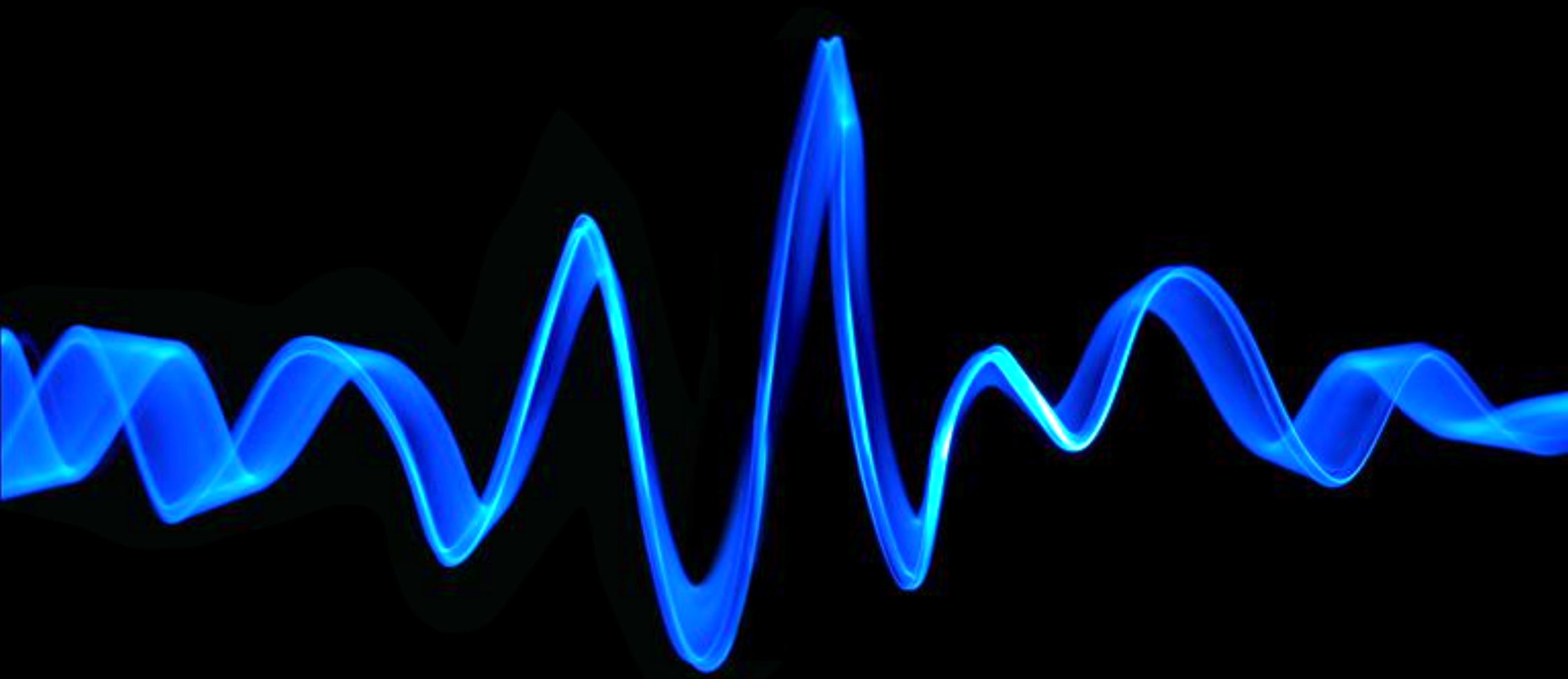
AMÉLIORATIONS

- Déploiement de la Web App en ligne
- Mise en place de bases de données SQL pour les scores des candidats
- Fluidifier l'app Flask
- Meilleure visualisation avec D3.js

FIL ROUGE

CONCLUSION



TELECOM PARISTECH - 2019

MULTIMODAL EMOTION RECOGNITION