PROJET DATA

Data Challenge Reminiz : Reconnaissance de célébrités

2017–2018

Daviet Mathieu
Hélénon François
Yan Sen

Superviseurs
Pennerath Frédéric

CentraleSupélec
2018

Table des matières

Chapitre 1

Présentation Challenge Reminiz

1.1 Contexte

Reminiz cherche à identifier en temps réels des personnages publiques à la télévisons (acteurs, chanteurs, présentateurs tv...). Un utilisateur pourrait ainsi filtrer le contenu en fonction d ses acteurs préférés à des fins ludiques ou de recherches documentaires. Dans cette optique il cherche des algorithmes qui puissent être à la fois précis et permettre un traitement rapide afin de pouvoir traiter la quantité considérable de contenu vidéo existante.

1.2 Objectifs spécifiques du challenge

Le but du challenge est de reconnaître un certain nombre de célébrités apparaissant dans une ou plusieurs sources vidéos différentes.

La mesure d'évaluation est tout simplement le taux de classification correcte.

1.2.1 Base de données

On dispose d'un set d'entraînement composé de :

— 9888 images téléchargées sur internet correspondant à 992 acteurs différents ainsi qu'une classe anonyme. Il y a environ une dizaine de d'image par classes (1 acteur = 1 classe)

— 97442 images groupées en 2461 tracks. Un track est une séquence de quelques images extraites d'une scène d'un film ou d'un épisode de série. Les images sont donc très proches les unes des autres. Un track contient en général un seul acteur (voir figure 1.1). Chaque track contient un identifiant unique renseignant sur le film dont le track est issu.

Les images internets et les celles extraites dans les tracks sont annotées avec un identifiant unique par acteur.

On dispose également d'un set de tests composé de :

— 25989 images groupées en 677 tracks. Les identifiants des track suivent la même nomenclature que pour le set d'entraînement ce qui renseigne sur le film dont sont issus les tracks.

— Pas d'images de type "internet"

Les images sont de tailles et de qualités variables que ce soit pour les tracks ou les images internets.

FIGURE 1.1 – Exemple de 2 images issues d'un même track

Chapitre 2

Classification par réseau neuronal profond

2.1 Exploitation base de données

Une première étape importante est de pouvoir manipuler les bases de tests et d'entraînement. Charger et travailler sur toutes la base de données en une fois est impossible car la mémoire vive est très vite saturée. Il a fallu charger intelligemment les images par batchs de données et stocker des résultats intermédiaires sur disque afin de pouvoir faire tous les traitements décrits dans la suite du rapport.

Par ailleurs afin d'accélérer les traitements, on utilise les fonctionnalités de multiprocessing de python afin de réaliser en parallèle le chargement en mémoire et le prétraitement des images (voir figure 2.1).

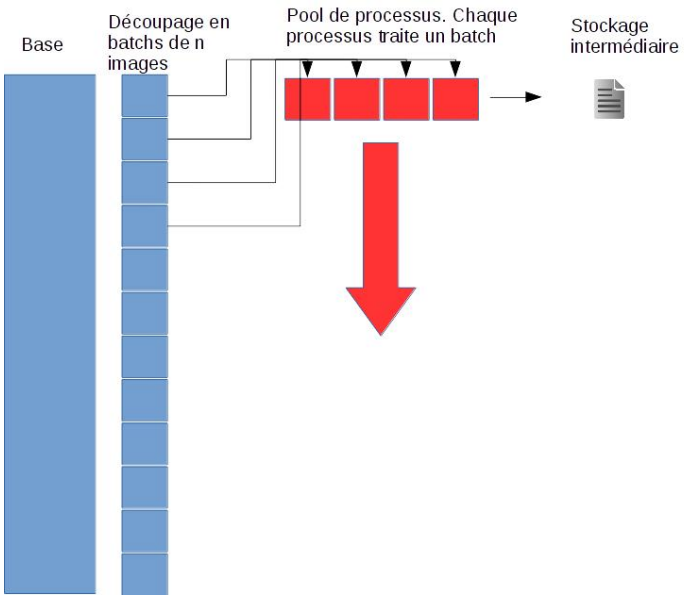

FIGURE 2.1 – Procédure de traitement des images

2.2 Un réseau pré-entrainé le Vgg-Face

Les réseaux convolutions profonds ont montré de très bonnes performances ces dernières années dans la reconnaissance d'objets et de personnes. L'apprentissage de tel réseaux pouvant demander des jours voire semaines de calcul sur des plateformes performantes, plutôt que de chercher à créer notre propre réseau, nous avons utilisé un réseau pré-entrainé. Une équipe d'Oxford [1] a notamment entraîné un réseau pour la classification de visages sur des personnalités et des bases de données différentes de celles utilisées par Reminiz.

Les couches "cachées" d'un réseau profond extraient des caractéristiques complexes des signaux d'entrées tandis que les dernières couches se chargent de la classification.

L'idée est alors que les couches cachées ont pu apprendre à extraire des caractéristiques suffisamment robustes et généralisables pour l'appliquer à notre cas.

On a alors téléchargé les poids du réseau et on les a utilisés après avoir recréer l'architecture du vgg-face sous le framework Keras. On a supprimé les dernières couches du réseau responsables de la classification.

Finalement on obtient en sortie du réseau un vecteur de 2622 features supposées représentatifs d'un visage.

2.3 Problèmatiques spécifiques à Reminiz

On peut cependant soulever certains problèmes liés à l'utilisation d'un réseau pré-entrainé. L'apprentissage a été réalisée sur des sets d'images particuliers et il faut donc adapter notre base de données pour que les images correspondent au même "type" d'image utilisée pour l'apprentissage du réseau. On peut notamment penser aux différences suivantes entre les images issues des bases du Vgg et celles des séquences de films :

- Images de taille 224*224*3 vs taille variable
- Personnalités de face, visage vertical vs orientation diverses des visages
- Images ne contenant que le visage (pas le corps) vs images contenant le haut du corps
- Bonne condition d'éclairage en générale vs éclairage variable

Afin de maximiser la reconnaissance il est donc nécessaire de pré-traiter les images.

FIGURE 2.2 – Comparaison d'une image issue des bases du vgg à gauche vs celle issues d'une séquence de film à droite

Bibliographie

[1] A. Zissermanr O. M. Parkhi, A. Vedaldi. Deep face recognition.