

Rapport IA & Deep learning : TP7- reconnaissance faciale

Camille FAHY
IA et Deep Learning
ESIEE - 2024
France
camille.fahy@e
du.esiee.fr

Hugo KOTHE
IA et Deep Learning
ESIEE - 2023
France
hugo.kothe@e
du.esiee.fr

Abstract— La reconnaissance faciale est l'une des spécialités de recherches les plus étudiées dans le domaine de la vision par ordinateur. Elle permet d'authentifier ou d'identifier une personne grâce à son visage. Cette technologie peut s'appliquer à de nombreux domaines mais elle fait face à des défis tels que la quantité de données nécessaire pour obtenir un modèle performant, la complexité de calcul, l'overfitting...

Index Terms—Deep learning, face recognition, face detection, face encoding, accuracy, overfitting

I. INTRODUCTION

La reconnaissance faciale est l'une des spécialités de recherche les plus étudiées dans le domaine de la vision par ordinateur. Elle est très utilisée pour l'identification d'individus que cela soit pour des raisons de sécurité, sur les réseaux sociaux, ou encore en médecine.

La première étape pour la reconnaissance faciale est la détection de visage. Chaque visage est alors encadré. Cependant, la détection d'un visage doit être possible même si le visage est mal orienté, sous différentes positions, luminosités ou échelles. Il est donc nécessaire d'aligner les visages. De plus, l'ordinateur doit prendre les marques qui sont les plus pertinentes pour lui afin de mesurer les visages : c'est ce que l'on appelle l'encodage. Finalement, il suffit de rapprocher les mesures de l'image test à celles des images de notre base de données.

Grâce à cette méthode, nous avons pu construire et entraîner un modèle performant pour identifier des personnes grâce à leur image faciale et l'entraîner sur notre propre jeu d'images. Toutefois, cela a nécessité l'import d'une grande quantité de données, un modèle avec une complexité de calcul importante et l'ajustement des hyperparamètres.

II. RELATED WORKS.

Un domaine émergent de la reconnaissance faciale est le diagnostic de maladie. Ces technologies permettraient d'améliorer les performances de détection de certaines maladies. Un exemple, cité dans le papier, est les maladies endocriniennes qui se caractérisent par des modifications typiques des caractéristiques faciales des os, des muscles et des tissus mous. "Les symptômes généraux de ces maladies à un stade précoce se confondent facilement avec d'autres syndromes métaboliques. Le processus de diagnostic de référence est complexe, avec plusieurs cas de tests hormonaux et d'examens d'imagerie [26 , 27]. Les hautes performances de la reconnaissance faciale dans le diagnostic en ont fait une approche de dépistage rapide et accessible. L'exemple de l'acromégalie est présentée En raison de la libération accrue d'hormones de croissance, les personnes atteintes d'acromégalie ont souvent un visage rectangulaire, un élargissement du nez et des lèvres, un prognathisme et un front bombé [26]. Divers algorithmes ont été développés pour détecter le visage des patients atteints d'acromégalie. En 2006, Learned-Miller et al. [28] ont proposé un modèle 3D morphable pour classer la face frontale en différentes catégories de manière semi-automatique. Quarante-neuf

patients atteints d'acromégalie et témoins ont été identifiés avec un taux de précision de 85,7 %. "[1]

La reconnaissance faciale est aussi utilisée en IoT.[2] Cette étude présente un système de reconnaissance faciale pour le contrôle de l'ouverture d'une porte utilisant une raspberry pie. Ce projet utilise le deep learning. 5 images de 5 personnes différentes sont prises par un appareil photo pour réaliser le training set. Les images sont ensuite augmentées pour obtenir au total 2500 images, puis recadrées et réduites. Il utilise l'architecture AlexNet composée de 8 couches. La formation comprend d'abord 100 époques et est répétée avec 20 époques après la phase de test. Les performances de ce modèle sont bonnes mais nécessitent beaucoup de temps pour traiter l'image. Ce projet semble similaire à notre projet et propose un exemple d'application pour la sécurité.

La reconnaissance faciale peut être utilisée pour la détection d'émotion. [3]

Dans cette étude, le prétraitement consiste à recadrer les yeux et la bouche en utilisant la méthode Haar Cascades. La base de données pour l'entraînement contient 20 images de 6 classes d'émotion. Pour tester les images, cet article adopte la base de données JAFFE. La base de données contient 213 images (chaque image : 256*256 pixels) de visages de femmes japonaises, chaque image étant définie dans l'expression originale. Il y a 10 personnes dans la bibliothèque d'expressions. Les performances de ce modèle sur les images tests donnent des résultats entre 85 et 95% selon les différentes émotions.

III. PROPOSITION

Here you will need to write about each of the topics that you worked on TP7.

A. Face detection

Write a short first paragraph on what is face detection (try to cite 1 work that you got this information from) Write a paragraph discussing the topics of:

- Talk about the data you received for face detection (what are the dimensions, how many images are there, etc.);
- Why do you need to do face detection for this data for your work ?
- What were the data pre-processing techniques you used ?
- How did you divide training and test ?
- Discuss about the small convnet you use, give a brief explanation of why you created it this way;
- Discuss about your small convnet performance on the test set;

- Does the result improve when you change : the number of layers, number of hidden units, learning rate initial value ?
- If possible include images (face before detection - after detection, performance graphs of your model, comparison between models, etc)



Fig. 1. Image caption.

La détection des visages est une étape essentielle de tous les algorithmes de reconnaissance faciale. Elle inclut l'alignement du visage, la reconnaissance du visage, l'analyse du visage. Le but de la détection des visages est de déterminer la présence de visages dans l'image et de renvoyer l'emplacement de l'image. Bien que cela semble être une tâche facile pour l'homme, cela peut s'avérer difficile pour les ordinateurs notamment à cause du positionnement du visage, de l'échelle, l'expression faciale ou la luminosité. [4]

Pour entraîner notre modèle de reconnaissance faciale, nous avons comme données d'entrée 218 images de dimension 2, de tailles variables, de 6 célébrités différentes.

D'abord, nous avons listé l'ensemble des chemins d'accès des images de notre ensemble. Puis, nous avons lu l'ensemble des images correspondantes, convertit l'image OpenCV de l'espace de couleur BGR (bleu, vert, rouge) en espace de couleur RGB (rouge, vert, bleu), extrait seulement le visage de la photo à l'aide de la fonction `extract_face(image, model="hog")`. L'outil "hog_detector" nous permet de détecter les visages et provient de la bibliothèque dlib. Nous avons noté que pour certaines images aucun visage n'était détecté. Cela concerne majoritairement les personnes ayant des lunettes ou le visage coupé.

Nous avons ensuite stocké ces visages extraits ainsi que le nom de la personne dans la liste "dataset" puis, dans un fichier "faces_dataset.pkl". Enfin, nous avons normalisé les images (en divisant les pixels par 255), encodé les labels et

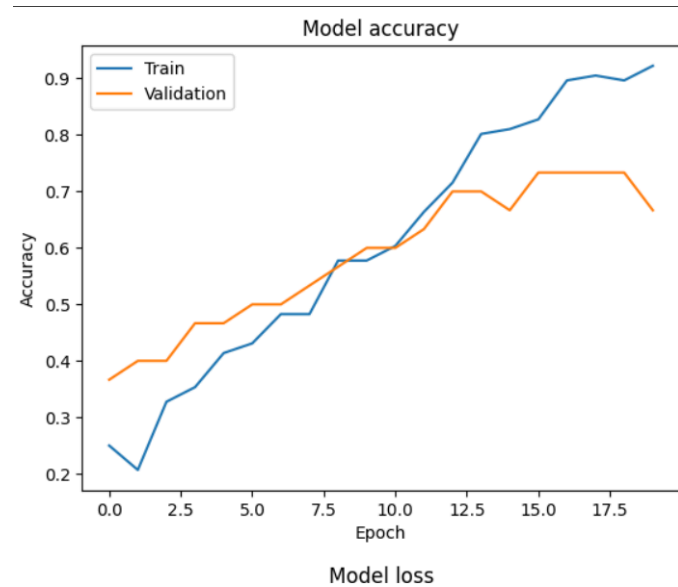
divisé le dataset en un ensemble d'entraînement et un ensemble de test avec la fonction `train_test_split(normalized_faces, integer_labels, test_size=0.3, random_state=42)`. Les données sont donc divisées en deux parties : 70% pour l'entraînement et 30% pour le test.

Nous avons construit notre modèle en plusieurs couches :

- une couche de convolution 2D avec 32 filtres de taille (3,3), et une fonction d'activation ReLU
- une couche de max pooling avec une fenêtre de taille (2,2)
- une couche de convolution 2D avec 64 filtres de taille (3,3), et une fonction d'activation ReLU
- une couche de max pooling avec une fenêtre de taille (2,2)
- une couche de convolution 2D avec 128 filtres de taille (3,3), et une fonction d'activation ReLU
- une couche de max pooling avec une fenêtre de taille (2,2)
- une couche d'aplatissement
- une couche entièrement connectée avec ReLU comme fonction d'activation
- une couche de sortie avec softmax comme fonction d'activation qui permet d'effectuer la classification, ici en 6 classes.

Nous avons créé un petit réseau de convolution qui est souvent utilisé pour des tâches de classification avec un petit dataset.

Pour améliorer les performances de notre modèle et réduire l'overfitting, nous avons par la suite décidé d'ajouter deux couches de Dropout avec pour taux 0.2 après le max Pooling. Toutefois, les performances de notre modèle restent limitées malgré l'ajustement des hyperparamètres. Nous obtenons une précision de 70%.



B. Pose estimation

Write a short first paragraph what is Pose estimation (try to cite 1 work that you got this information from) Write paragraphs discussing the topics of:

- Talk about the data you are now using for pose estimation (what are the dimensions, etc.);
- Why do you need to do pose estimation for pose correction for this data for your work ?

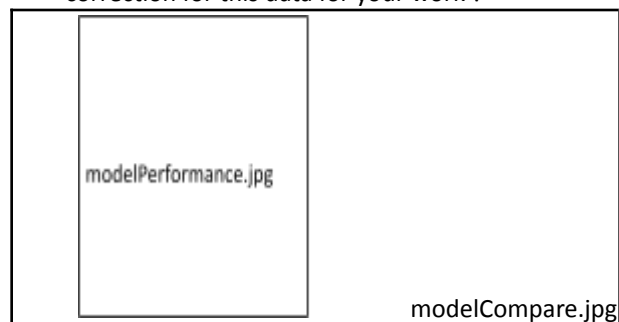


Fig. 2. Image caption.

- What were the techniques you used for pose estimation ?
- How did you divide training and test ?

- Discuss about how your previous small convnet models performs on this test set;
- Does the result improve ?
- If possible include images (New image sample of training image after pose correction- previous and after-, performance graphs, comparison between models, comparison between the performance of the model with this data and just the previous step, etc);
- Why do you think the results improved ?

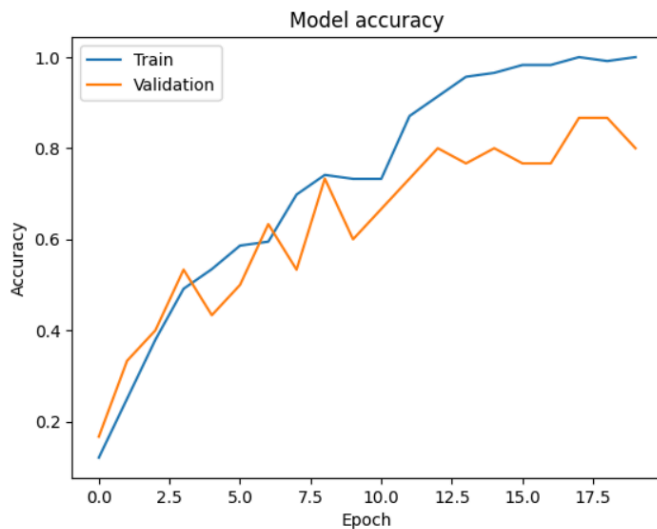
La variation de la pose du visage ou de l'éclairage font partie des principaux problèmes de la reconnaissance faciale. C'est pourquoi les outils d'estimation de la pose sont très utiles. Plusieurs techniques existent : celles basées sur les points de repères (Zhao et Gao 2006) qu'on utilisera lors de ce TP, les méthodes sous-spatiales basées sur l'apparence, qui traitent l'ensemble du visage comme un vecteur de caractéristiques dans certains sous-espaces de fonctionnalités (Gong, 1996), ou une combinaison des deux (Grundig et Hellwich, 2004). Cela permet d'aligner les visages de sorte qu'ils soient reconnaissables par l'ordinateur peu importe leur pose. [5]

Ainsi, en transformant nos visages préalablement extraits nous devrions obtenir de meilleures performances.

Nous avons appliqué la fonction landmarks() qui permet de calculer une liste de marques caractéristiques pour chaque visage. Ensuite, nous appliquons la fonction align_faces. Ainsi, nos visages sont alignés et centrés.

Comme précédemment, nous avons divisé nos données avec 30% pour le test et 70% pour l'entraînement. Nous obtenons de meilleurs résultats que dans la première partie avec une précision de 77.8%.

Test accuracy: 77.7777910232544



C. Face encoding

Write as the first paragraph what is Encoding - if you can be specific try to explain in the context of face encoding- (try to cite 1 work that you got this information from) Write paragraphs discussing the topics of:

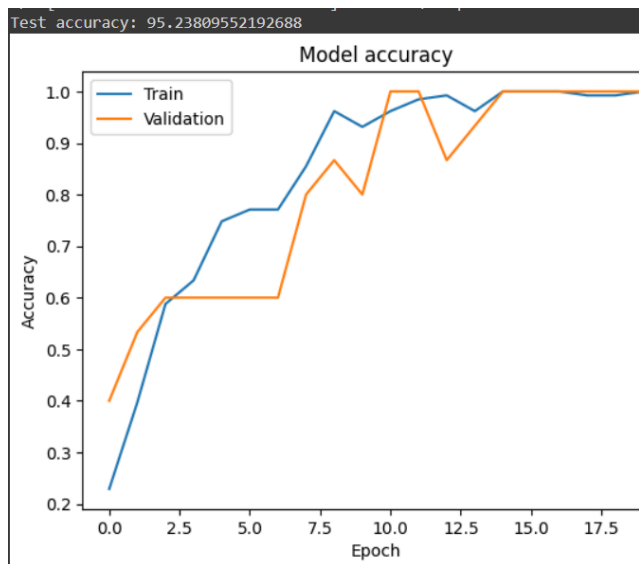
- What is the advantage of face encoding ?
- What is the data that you are going to be using during face encoding ?
- Discuss what the data becomes after face encoding; • Discuss about your training model for this new data;
- Does the result improve ?
- If possible include images (performance graphs, comparison between models, comparison between the performance of the model with this data and just the previous step, etc);
- Why do you think the results improved (or why not)?

L'encodage est une méthode utilisée en reconnaissance faciale pour déterminer les points importants à mesurer pour chaque visage, l'ensemble de ces mesures permettront d'identifier de manière unique chaque visage. Les 128 points à mesurer sont déterminés par des méthodes de deep learning. [6]

Contrairement aux modèles entraînés précédemment, nous avons comme données d'entrée des vecteurs de taille (128). En effet, nous ne nous concentrons pas sur tous les détails du visage mais seulement sur les 128 mesures nécessaires pour l'identifier.

De plus, notre modèle sera un réseau de neurones et non plus un réseau de convolution car l'image n'est plus une image à découper en plusieurs parties mais un simple vecteur.

Ainsi, en encodant nos images, la précision de notre modèle a augmenté à 95,2% contre 77% sans encodage. Contrairement à précédemment, nous observons peu d'overfitting. La courbe de validation suit la courbe d'entraînement. Nous aurions pu augmenter le nombre d'epochs pour voir jusqu'où peut être augmentée la précision.



D. Face recognition

Write as the first paragraph what is face recognition (try to cite 1 work that you got this information from). What is the difference between face recognition and face detection ? Write paragraphs discussing the topics of:

- What are the classifiers you chose to do this task ? Why ?
- Discuss about the performance of each classifier in this task;
- What is your best classifier ? Write about the performance;
- If possible include images (performance graphs, comparison between models, etc);

La reconnaissance faciale est une méthode d'identification des visages basée sur leur caractéristiques biométriques. [7]

La différence entre la reconnaissance faciale et la détection de visage est que la reconnaissance faciale associe un visage à un nom alors que la détection de visage détecte seulement la présence ou non d'un visage.

Nous avons testé 4 classificateurs :

- logistic regression
- KNN
- SVM
- Neural Network

```
Logistic Regression: Accuracy = 0.9682539682539683, Training Time = 0.04177451133728
SVM: Accuracy = 0.9841269841269841, Training Time = 0.006528615951538086
kNN: Accuracy = 0.9841269841269841, Training Time = 0.016952209854125977
Neural Network: Accuracy = 0.9841269841269841, Training Time = 0.7190883159637451
```

La logistic regression a la moins bonne précision. Malgré un temps d'entraînement rapide, le modèle n'est pas le plus adapté.

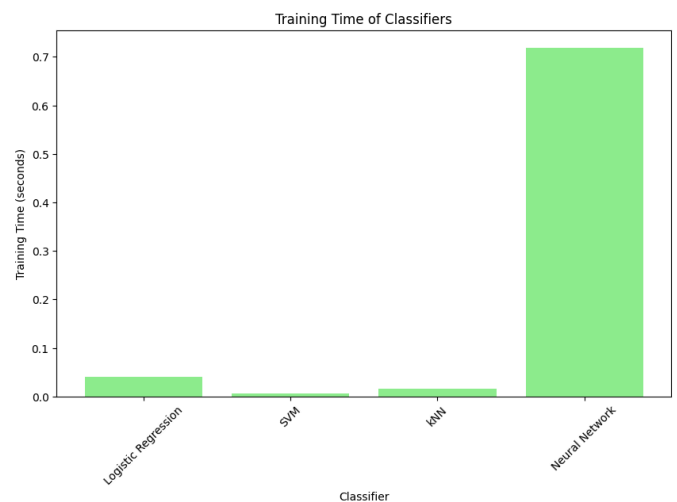
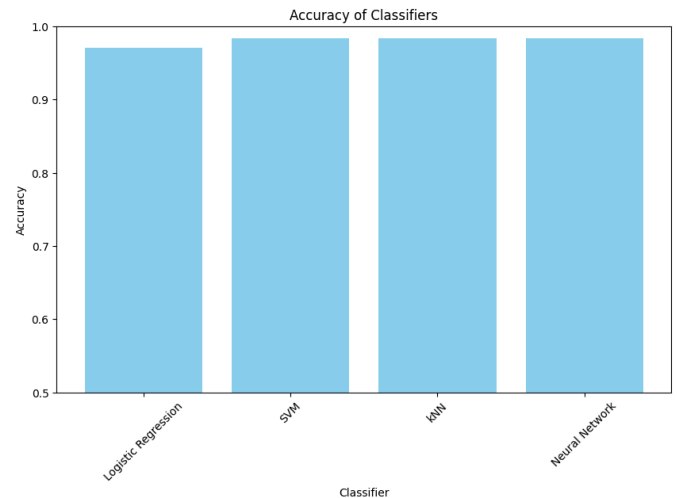
Le Support-Vector Machine à une très bonne précision (0.98) et un temps d'entraînement très court (0.0065), ce qui en fait un très bon classificateur.

La méthode de k plus proche voisins a aussi une bonne précision (0.98) et un temps court d'entraînement (0.016).

Le réseau de neurones a une bonne précision (0.98) mais un long temps d'entraînement (0.7).

Ainsi, le classificateur le plus adapté est le SVM (Support-Vector Machine).

Nous avons noté que cela est vrai en utilisant Google colab, mais que cela n'est pas forcément vrai sur toutes les machines.



E. Personal dataset

- Talk about how you created your dataset; • Does your model work on your personal set?
- How is your model's performance in your dataset compared to the previous dataset? Is it good/bad ? (What is your opinion on why)
- Compare your dataset to the original data;

Nous avons pris des photos de nous et de nos familles. Nous avons réduit la qualité des photos.

Les performances sur notre modèle sont plutôt bonnes.

F. Extra - Bias analysis

- What is bias ?
- Why is bias a problem in machine learning ?
- Can you comment on a situation where bias can be a problem ?
- Did you create a set trying to diversify your input examples ? (Is it all from the same person ?, Do you present different people ethnicities to verify if the model had bias towards some specific group (meaning better performed in one group)?)
- Can you think of some statistics to calculate and present based on the the original face detection data ? What about you data ?

[8] On peut penser que l'apprentissage automatique pourrait être une solution pour réduire les inégalités, les décisions injustes. Pourtant les décisions de l'apprentissage automatique peuvent, elles aussi, être biaisées. En effet, lorsque la machine apprend par elle-même à partir d'exemples, il se peut que certaines données soient biaisées, ainsi la machine va intégrer ces biais au modèle. Ces biais peuvent se produire lorsque que dans les données d'entraînement la population n'est pas représentative de la population globale ou lorsque le modèle n'est pas assez complexe pour représenter correctement la structure de données.

Imaginons qu'une entreprise utilise un algorithme d'IA pour trier les CV des candidats à un poste. Voici comment les biais pourraient poser problème dans ce contexte :

Si l'algorithme est entraîné sur des données qui sont basées sur des préjugés comme l'âge, l'origine ou le genre. L'algorithme pourra reproduire ces biais dans ces décisions.

Si l'ensemble des données utilisées pour l'entraînement ne représente pas correctement l'ensemble des candidats potentiels, l'algorithme peut avoir du mal à décider pour des profils peu représentés.

Si les critères de sélection sont basés sur des stéréotypes, cela se reflétera dans les décisions prises par l'algorithme.

Pour tenter de réduire les biais, on peut analyser nos données d'entraînement et calculer les répartitions en fonction de plusieurs caractéristiques. Par exemple, dans notre cas, on pourrait calculer la représentation de différentes ethnies. On pourrait aussi tester notre modèle sur différents sous-ensembles de personnes pour s'assurer que les performances sont égales.

Les biais sont donc à prendre en compte dès la création du modèle.

CONCLUSION

Here you need to summarize in 1-2 paragraphs this work and what you learned.

Pour conclure, ce TP nous a permis d'en savoir plus sur la reconnaissance faciale et les techniques de deep learning associées. Nous avons pris connaissance de l'état de l'art et des notions de détection de visages, du problème de la variation de pose, de l'encodage. Nous avons tenté d'augmenter les performances de notre modèle grâce à plusieurs techniques telles que le drop out, l'early stopping, le choix du bon classificateur... Nous avons ainsi appliqué toutes ses méthodes à notre jeu de données et obtenu un modèle performant.

Ce TP nous sera très utile dans nos futurs projets en intelligence artificielle et nous permettra de mieux comprendre les enjeux actuels de ce domaine.

REFERENCES

- [1] 'Review on facial recognition based application in diseases diagnosis' by Jiaqi Qiang, Danning Wu, Hanze Du, Huijuan Zhu, Shi Chen, and Hui Pan
- [2] 'IoT based facial recognition door access control home security system using raspberry pi' A. R. Syafeeza, M. K. Mohd Fitri Alif, Y. Nursyifaa Athirah, A. S. Jaafar, A. H. Norihan, M. S. Saleha
- [3] 'An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment
Author links open overlay panel'
- D. Yang a, Abeer Alsadoon a, P.W.C. Prasad a, A.K. Singh b, A. Elchouemi c
- [4] 'WIDER FACE: A Face Detection Benchmark' Shuo Yang¹ Ping Luo^{2,1} Chen Change Loy^{1,2} Xiaoou Tang^{1,2}
- [5] 'HEAD POSE ESTIMATION IN FACE RECOGNITION ACROSS POSE SCENARIOS' M. Saquib Sarfraz and Olaf Hellwich,
- [6] TP7 facial recognition Laurent NAJMAN
- [7] W. Zhao, R. Chellappa, P. J. Phillips, Face recognition: A literature survey, "ACM Computing Surveys (CSUR)", December 2003
- [8] Understanding Bias in Machine Learning Jindong Gu ^{1, 2}, Daniela Oelke ²