IA & Deep Learning : Face Recognition

AMARA Yanis
IA et Deep Learning
ESIEE - 2025
France
yanis.amara@edu.esiee.fr

APPUDURAI Achveiya
IA et Deep Learning
ESIEE - 2025
France
achveiya.appudurai@edu.esiee.fr

Présentation du projet

Notre travail sur la reconnaissance faciale s'est déroulé en plusieurs étapes : extraction des visages d'un dataset et stockage dans un fichier pickle, division des données (70 % entraînement, 30 % test) avec normalisation, alignement des visages pour corriger les poses et encodage en vecteurs pour entraîner un réseau de neurones. Enfin, nous avons testé divers classificateurs (régression logistique, SVM, KNN) pour choisir le meilleur. Malgré des défis sur Kaggle, nous avons réussi la détection vidéo et son téléchargement.

Keywords:

- Réseaux de neurones
- Détection faciale
- Entraînement du modèle
- Normalisation des données
- Apprentissage automatiq***ue***

I. Introduction

Ce projet consiste à créer un modèle d'intelligence artificielle dédié à la reconnaissance faciale. Cette technologie est aujourd'hui largement utilisée dans des domaines très variés, allant de la défense et de la sécurité biométrique à des applications plus courantes comme le tri de photos selon les visages familiers.

Un tel modèle présente plusieurs défis. Une intelligence artificielle de reconnaissance faciale doit être en mesure de reconnaître une personne même dans des conditions peu idéales. Elle doit pouvoir faire abstraction de détails comme le port d'accessoires, une nouvelle coupe de cheveux ou une expression faciale différente. Nous verrons par la suite comment pallier ces challenges.

L'essor de l'IA laisse présager une évolution continue de la technologie de reconnaissance faciale. Comprendre les enjeux et le fonctionnement de ces modèles est essentiel pour y contribuer efficacement.

III. Travaux connexes

Il existe de nombreuses approches différentes à ce problème, telles que : le DeepFace, FaceNet, Dlib, LBP (*Local Binary Patterns*) ou HOG (*Histogram of Oriented Gradients)*. De manière générale, nous pouvons scinder la plupart des méthodes connues en deux catégories : celles basées sur le *Deep Learning*, et celles plus classiques reposant sur des caractéristiques manuelles (*hand-crafted features*).

Un exemple assez représentatif d'une approche axée sur le Deep Learning est le système de reconnaissance faciale FaceNet. Cette méthode est introduite pour la première fois à la communauté scientifique en 2015 dans l'article *FaceNet* : *A Unified Embedding for Face Recognition and Clustering* rédigé par Florian Scroff, Dmitry Kalenichenko et James Philbin de Google Inc.

FaceNet est un extracteur de caractéristiques, il permet de représenter des visages dans un espace euclidien dans lequel, les vecteurs associés aux visages présentant des similarités sont plus ou moins proches les uns des autres. Ce système repose sur un réseau de neurones convolutif CNN prenant en entrée une image sous forme de matrice 2D ou 3D, qui sera normalisée. La sortie de ce réseau est un vecteur de 128 dimensions décrivant le visage dans un espace

géométrique multidimensionnel. Plus deux visages se ressembleront, plus la distance euclidienne entre les deux vecteurs résultants sera petite. FaceNet repose sur de l'apprentissage supervisé avec la fonction de coût *Triplet Loss*. Un système comme celui-ci peut très bien être employé en combinaison avec un classificateur *(k-NN, SVM, softmax)* pour faire de la prédiction et/ou de la vérification d'identité.

Une méthode plus ancienne basée sur des caractéristiques manuelles est l'approche Eigenface.
Introduite au début des années 90 par Matthew Turk et Alex Pentland du *Massachusetts Institute of Technology* dans leur article *Eigenfaces for Recognition.* Il s'agit d'une méthode statistique basée sur l'approche PCA *(Principal Component Analysis)*
Nous partons ici aussi sur une base d'images de visages vectorisés, alignés et normalisés. Le principe de la méthode PCA repose sur le fait de déterminer les principaux paramètres dont la variance est la plus élevée dans le jeu de données. Ce sont ces paramètres qui sont les *eigenfaces*. Une fois ces *eigenfaces* identifiés, un visage peut être approximativement représenté par une combinaison linéaire de vecteurs propres (les *eigenfaces*) coefficientés par des poids déterminés par projection matricielle de l'image normalisée sur le paramètre que l'on veut pondérer. À partir de cela, comparer deux visages revient finalement à calculer leur différence de poids *(distance euclidienne entre les vecteurs de poids)*.

IV. Proposition

Nous nous concentrerons dans la suite sur une approche plus moderne suivant de l'apprentissage supervisé en utilisant des réseaux de neurones convolutifs et/ou fully-connected. Cette approche est l'une des plus modernes et nous offrira des résultats plus acceptables.

Nous commencerons par modéliser et entraîner un réseau de convolution sur un dataset de visages simples. Pour perfectionner notre système, nous adapterons notre dataset pour entraîner, toujours un réseau de convolution, cette fois, sur un dataset de visages normalisés *(face alignment)*. Pour finir, nous considérerons une approche basée sur un réseau de neurones fully-connected dont l'entrée ne sera plus des images de visages mais des mesures servant de repères faciaux.
Pour chacun des cas, une étude de validité du modèle sera effectuée. L'objectif général est d'améliorer la précision et la validité de notre système de reconnaissance faciale en ajustant notre dataset et notre modèle.
Pour finir, nous évaluerons le meilleur classificateur parmi de nombreux exemples de modèles existants afin de l'appliquer à un dataset que nous aurons créé. Notre objectif est de considérer les différentes options et méthodes évaluées positivement au cours du projet afin d'obtenir un système optimal.

A. Face detection

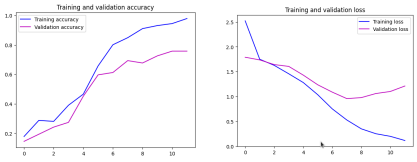
La détection faciale est un système répandu présent dans de nombreuses applications. Nous le retrouvons par exemple dans la plupart des appareils photos, il permet de détecter le visage d'une personne afin d'effectuer un bon focus dessus. De manière générale, cela consiste à trouver des visages sur une image donnée *(cf. Human face detection techniques: A comprehensive review and future research directions)*.

Le dataset fournit est un set de photographies de 6 personnes totalisant 160 images en couleur de taille *128x128*. Nous allons utiliser la détection faciale sur ce dataset afin de former un jeu de données constitué exclusivement de visages. L'objectif est de permettre à notre modèle de se concentrer sur ce qu'on veut qu'il évalue, à savoir le visage de personnes.
Laisser des paysages, le reste du corps ou d'autres éléments superficiels pourrait fausser les résultats, donnant la possibilité à l'IA de les considérer comme déterminant dans la classification.

Nous commençons, grâce à la fonction *face_extraction()* par créer un nouveau dataset constitué des images modifiées pour ne montrer que le visage. Nous utilisons pour cela la fonction *extract_faces()*.

Une fois notre dataset modifié, nous pouvons entraîner un réseau de convolution avec un Dropout à un ratio de 0.5.

Nous avons obtenus avec ce modèle en moyenne, une précision, en entraînement, de 98% et en validation, de 75%



B. Pose estimation

L'estimation de pose humaine (HPE) consiste à localiser les parties du corps humain, comme les bras, jambes ou tête, à partir d'images ou vidéos, souvent sous forme de squelette. Elle désigne plus précisément la localisation des articulations ou points de repère prédéfinis (Hong et al., 2018). Essentielle en intelligence artificielle, cette technique est utilisée en analyse du mouvement, réalité augmentée et interaction homme-machine. Les méthodes d'apprentissage profond ont considérablement amélioré sa précision en 2D et 3D (Zheng et al., 2023).

Pour l'estimation de pose, nous utilisons 160 images colorées (128x128 pixels) de 6 personnes, offrant un traitement rapide tout en conservant les détails nécessaires pour localiser les traits du visage..

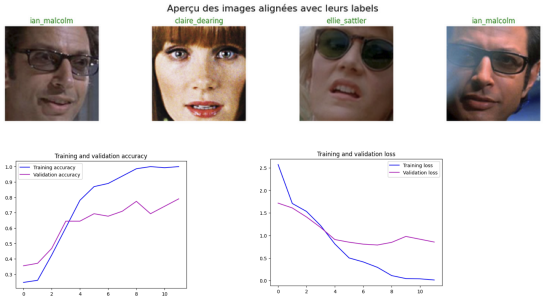
Dans notre projet, l'estimation de pose corrige l'orientation des visages pour améliorer la reconnaissance faciale. La méthode des landmarks identifie 68 points clés pour aligner les visages, standardiser leur représentation et augmenter la précision du modèle.

Pour évaluer la pose dans notre projet, nous avons chargé des images d'un fichier.pkl, puis utilisé la fonction face_landmarks pour identifier les zones majeures du visage, comme les yeux, le nez et la bouche. Ensuite, la fonction *align_faces* a été utilisée pour aligner chaque image selon les points identifiés, standardisant ainsi la position des visages et optimisant leur représentation pour une meilleure reconnaissance faciale.

Le modèle ConvNet a atteint une précision de 97% et une perte de 0,12 en entraînement, mais à 64% de précision et une perte de 1 en validation, indiquant un surapprentissage. Le nouveau modèle CNN, avec des couches supplémentaires, Dropout et l'optimiseur Adam, a amélioré la généralisation, atteignant 100 % de précision en entraînement, 79 % en validation, et des pertes réduites à 0,01 et 0,8343 respectivement.

Les améliorations sont dues à l'ajout de couches comme Dense(128, activation='relu'), qui a enrichi les représentations apprises, et à l'intégration de Dropout et de l'optimiseur Adam, permettant une meilleure régularisation et généralisation. Ces changements ont augmenté la précision et ont diminué la perte par rapport au modèle précédent.

les visages alignés:



C. Face encoding

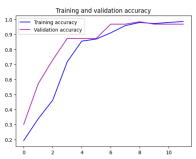
L'encodage facial est la procédure par laquelle un visage devient une représentation unique dans le cerveau. Des recherches ont démontré que certaines régions du cortex temporal inférieur stimulent un petit groupe de neurones en fonction de caractéristiques faciales particulières (comme la distance interoculaire), permettant ainsi une reconnaissance efficace. Chaque visage est lié à une « combinaison » neuronale spécifique, comme un code.(cf CNRS – Le code cérébral décrypté).

L'encodage facial au niveau du cerveau fournit une reconnaissance rapide et stable de visages, y compris dans des cas de variations. Grâce à une activation neuronale spécifique, il crée un "code" spécifique pour tout visage, optimisant la mémoire visuelle sociale et différenciant des milliers de visages.

Dans notre projet, nous utilisons 160 images colorées (128x128 pixels) de 6 personnes pour l'encodage des visages.En complément de dlib, ces images ont été utilisées pour entraîner le modèle, traitées et encodées pour extraire des caractéristiques faciales destinées à la reconnaissance et à la comparaison avec d'autres visages du dataset.Après l'encodage, les images sont transformées en vecteurs numériques représentant les traits du visage, associés à des labels via one-hot encoding. Chaque individu est identifié par un

vecteur binaire où un seul élément vaut 1 (ex. [1. 0. 0. 0. 0. 0.] pour *alan_grant*). Ces vecteurs sont ensuite utilisés pour entraîner les classificateurs. Pour l'entraînement du modèle, nous avons utilisé ces encodages comme données d'entrée. Le modèle a appris à reconnaître et différencier les visages à partir de ces représentations.

Les résultats se sont améliorés grâce à l'ajout de landmarks pour aligner les visages et normaliser les données, réduisant les erreurs de pose. L'utilisation d'un modèle MLP après l'encodage a également augmenté la précision, car il est mieux adapté pour traiter ces représentations numériques et effectuer des classifications complexes.



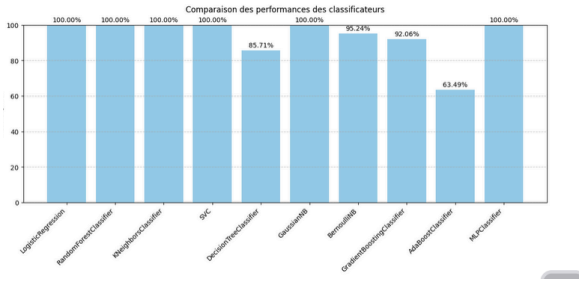

```
[1. 0. 0. 0. 0. 0.] = 1 --> ellie_sattler
[0. 1. 0. 0. 0. 0.] = 2 --> claire_dearing
[0. 0. 1. 0. 0. 0.] = 3 --> alan_grant
[0. 0. 0. 1. 0. 0.] = 4 --> john_hammond
[0. 0. 0. 0. 1. 0.] = 5 --> owen_grady
[0. 0. 0. 0. 0. 1.] = 6 --> ian_malcolm
```

D. Face recognition

La reconnaissance faciale est une technologie qui permet d'identifier une personne en utilisant les caractéristiques uniques de son visage. Selon un article sur le site Think Global Health, elle peut être utilisée pour diverses applications, notamment dans le domaine de la santé pour surveiller les signes vitaux et améliorer la sécurité des patients. La différence entre la reconnaissance et la détection faciale réside dans le fait que la détection ne va que identifier un visage dans une image ou vidéo, tandis que, la reconnaissance fait un pas supplémentaire dans le procédé en identifiant la personne.

Plusieurs classificateurs ont été évalués dans cette tâche de reconnaissance faciale, tous atteignant une précision de 96,83 %. Logistic Regression s'est démarqué par sa simplicité et fiabilité, Random Forest par sa réduction du surapprentissage, SVC par son efficacité dans les espaces complexes, Gaussian Naive Bayes par son indépendance des caractéristiques, et le MLPClassifier, par sa capacité à traiter les relations non linéaires.

Les autres classificateurs ont rapporté des résultats variés: KNN (95,24 %), sensible à la métrique et au nombre de voisins ; Decision Tree (77,78 %), fragile face au surapprentissage ; Bernoulli Naive Bayes (88,89 %), prévu pour la donnée binaire ; Gradient Boosting (85,71 %), basé sur une approche séquentielle. Logistic Regression a marqué par sa vitesse, facilité et précision, ce qui en fait le modèle de préférence.



E. Personal dataset

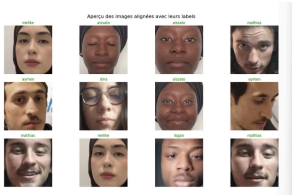
Pour tester notre modèle dans un contexte plus réaliste, nous avons créé notre propre dataset à partir de personnes de notre entourage. Nous leur avons demandé de se filmer pendant environ 30 à 40 secondes. À partir de ces vidéos, nous avons ensuite extrait les visages à l'aide d'un script Python. Cela nous a permis de constituer un jeu de données contenant environ une cinquantaine d'images.

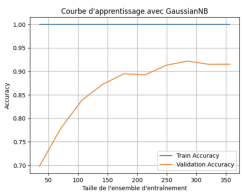
En utilisant la même approche de la partie 4, nous avons sélectionné le modèle qui avait eu les meilleures performances et nous l'avons ensuite entraîné sur notre jeu de données. L'objectif était de vérifier si le modèle était capable de reconnaître correctement les visages et de prédire les bons labels.

Durant les tests, nous avons analysé l'évolution de la performance du modèle en fonction de la taille du dataset. Nous pouvons voir que sur un petit dataset, le modèle atteint très rapidement une accuracy proche de 1, donc il y a de l'overfitting. Lorsque nous augmentons la taille du dataset, l'accuracy diminue. Cela s'explique car il y a plus d'images , ce qui rend la tâche plus difficile mais permet ainsi au modèle de mieux généraliser et d'éviter l'overfitting.

Par rapport au dataset initial utilisé au début de projet, notre dataset contient plus d'images. Cela nous a permis de mieux entraîner notre modèle. Nous avons dû adapter notre modèle pour optimiser ses performances et nous avons observer une

meilleure capacité de généralisation, car le modèle apprend à reconnaître les visages dans différents angles.





F. Extra - Bias analysis

Le biais de machine learning fait référence à l'apparition de biais en raison d'un déséquilibre dans les données d'entraînement. Cela signifie que certaines classes ou groupes peuvent être surdéveloppés ou sous-représentés, à cause de cela le modèle généralise mal sur les données réelles.
Une situation dans laquelle un modèle biaisé peut être problématique est lorsqu'il ne reconnaît pas certains individus. Par exemple, dans le domaine médical, les femmes peuvent présenter des symptômes d'infarctus différents de ceux des hommes. Or, si un modèle d'aide au diagnostic a été entraîné majoritairement sur des données masculines, il pourrait ne pas détecter correctement une crise cardiaque chez une femme. Cela pourrait entraîner un mauvais diagnostic, avec des conséquences graves, et soulève un vrai problème d'équité dans le traitement médical.
Lors de la création de notre propre dataset, nous avons demandé à plusieurs personnes de notre entourage de se filmer pendant environ 30 à 40 secondes. L'objectif était justement de ne pas limiter notre modèle à une seule personne ou un seul type de visage, mais d'inclure des individus avec des caractéristiques physiques variées.
Nous pouvons utiliser des statistiques utiles pour détecter un biais comme l'accuracy par classe ou par individu qui va donc permettre de vérifier si le modèle est plus performant pour certaines personnes.
Sur notre dataset, on pourrait comparer les performances par individu, puisque chaque visage a un label, cela permettrait de voir si le modèle fonctionne mieux pour certains que pour d'autres.

CONCLUSION

Ce projet nous a donné l'occasion de bien creuser les diverses étapes requises à la mise en place d'un système de reconnaissance faciale efficace. De la détection à l'alignement des visages, jusqu'à l'encodage et la classification, nous avons eu la possibilité de tester différentes méthodes et améliorer notre modèle pour obtenir des résultats acceptables, sur le dataset de base aussi bien que sur notre jeu de données personnel. En adaptant le modèle et en analysant les performances selon la taille du dataset, nous avons observé les effets de l'overfitting et l'importance de la généralisation.
Ce projet nous a également sensibilisés à des problématiques importantes comme les biais dans les données. En diversifiant notre dataset personnel, nous avons cherché à rendre notre modèle plus précis et représentatif. Ce projet nous donne de bonnes bases pour continuer à améliorer et mieux comprendre comment utiliser l'intelligence artificielle pour reconnaître des visages tout en garantissant une performance équitable.

RÉFÉRENCES

1. https://lejournal.cnrs.fr/nos-blogs/aux-frontieres-du-cerveau/reconnaissance-des-visages-le-code-cerebral-decrypte

2. Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition*, *111*(2), 219-247.

3. Starke, G., De Clercq, E., & Elger, B. S. (2021). Towards a pragmatist dealing with algorithmic bias in medical machine learning. *Medicine, Health Care and Philosophy*, *24*, 341-349

4. https://arxiv.org/abs/1503.03832

5. https://www.face-rec.org/algorithms/PCA/jcn.pdf

6. https://www.thinkglobalhealth.org/article/ai-and-facial-recognition-dive-global-health-care