

RÉPUBLIQUE DE CÔTE D'IVOIRE



INTERNATIONAL DATA SCIENCE INSTITUTE



Rapport de Projet de Deep Learning

Doppleganger Finder

RÉALISÉ PAR :

CAMARA Farouck
DIABY Aboubacar
NIADA Ulrich Boris
ZONGO Alphé Colombe

ENSEIGNANT :

Mohamed Abbas KONATE

MASTER 2 DATA SCIENCE & SÉCURITÉ

Année académique 2025-2026

Résumé

Ce projet présente la conception et la mise en oeuvre d'un système de reconnaissance faciale basé sur le Deep Learning, nommé *Doppleganger Finder*. L'objectif est d'identifier, au sein d'une large base de données d'images, les visages présentant la plus forte ressemblance avec une image requête fournie par l'utilisateur.

Notre approche s'appuie sur la méthode du **Transfer Learning** via l'architecture **Inception-ResNet V1** pré-entraînée sur le dataset **VGGFace2**, transformant les visages en vecteurs de haute dimension appelés *embeddings*. La similarité entre ces vecteurs est ensuite calculée à l'aide de la distance cosinus.

Le système final intègre une interface utilisateur interactive réalisée avec **Streamlit**, permettant l'acquisition d'images (upload ou webcam), la détection automatique des visages via **MTCNN**, et la visualisation instantanée des "sosies" trouvés avec des statistiques détaillées incluant l'analyse démographique des visages similaires identifiés.

Ce projet démontre l'efficacité des modèles pré-entraînés pour résoudre des tâches complexes de vision par ordinateur sans nécessiter de phase d'entraînement coûteuse.

Table des matières

1	Introduction	3
2	Objectif du Projet	3
3	Problématique	3
4	Fondements théoriques	3
4.1	Le transfert learning	3
4.2	Les embeddings faciaux et FaceNet	4
4.3	Architecture Inception-ResNet	4
5	Données utilisées	4
5.1	Dataset FairFace	4
5.2	Images personnelles	4
6	Méthodologie et Justification	5
6.1	Justification des choix méthodologiques	5
6.2	Prétraitement des images (Détection et Alignement)	5
6.3	Extraction des embeddings	6
6.4	Fusion des données	6
6.5	Recherche de similarité	6
7	Interface utilisateur	7
7.1	Choix de l'outil : Streamlit	7
7.2	Fonctionnalités avancées	7
8	Résultats et analyse	7
8.1	Analyse des scores de similarité	7
8.2	Influence du prétraitement des images	7
8.3	Analyse critique des résultats	8
8.4	Analyse qualitative	8
8.5	Limites et biais	8
9	Conclusion et perspectives	8
	Références	10

1 Introduction

La reconnaissance faciale est devenue une technologie omniprésente, trouvant des applications allant de la sécurité biométrique à l'organisation d'albums photos personnels, en passant par le divertissement et les systèmes de contrôle d'accès.

Cependant, développer un système performant *à from scratch* (à partir de zéro) représente un défi majeur. Cela nécessite des puissances de calcul considérables et des jeux de données massifs contenant des millions d'images annotées, ressources souvent inaccessibles en contexte académique ou pour des équipes disposant de ressources limitées.

C'est dans ce contexte que le Deep Learning et particulièrement le Transfer Learning offrent une solution élégante et efficace. En réutilisant la *à connaissance* acquise par des modèles entraînés sur des tâches gigantesques, il devient possible de créer des applications performantes et fiables avec les ressources disponibles.

2 Objectif du Projet

L'objectif central de ce projet est de concevoir un système de recherche de visages similaires, capable de fonctionner efficacement sans phase d'apprentissage coûteuse, grâce à l'exploitation de modèles pré-entraînés. Concrètement, l'application développée est capable de trouver les sosies d'un utilisateur au sein d'une base de données publique enrichie de photographies personnelles, en mettant l'accent sur la précision des résultats et la qualité de l'expérience utilisateur.

3 Problématique

Dans quelle mesure un modèle de reconnaissance faciale pré-entraîné peut-il être utilisé efficacement, sans phase d'apprentissage supplémentaire, pour réaliser une recherche de similarité faciale fiable et précise ?

Cette question soulève plusieurs enjeux : la transférabilité des connaissances acquises sur des tâches de reconnaissance généraliste à la tâche spécifique de recherche de similarité, la qualité et la robustesse des résultats face à des variations de conditions de prise de vue, et enfin la possibilité de déployer efficacement un tel système avec des ressources computationnelles limitées.

4 Fondements théoriques

4.1 Le transfert learning

Le transfert learning est une technique d'apprentissage automatique où un modèle développé pour une tâche spécifique est réutilisé comme point de départ pour un modèle sur une seconde tâche. Plutôt que de commencer l'apprentissage avec des poids aléatoires, on initialise le réseau avec des poids pré-entraînés.

Il existe principalement deux stratégies :

- ❑ **Fine-tuning** : On ré-entraîne tout ou partie du modèle sur le nouveau jeu de données.
- ❑ **Feature extraction (Extraction de caractéristiques)** : Le modèle pré-entraîné est utilisé tel quel pour transformer les données brutes (pixels) en une représentation abstraite pertinente (vecteurs).

Dans ce projet, nous avons opté pour le **feature extraction**. Nous utilisons le réseau de neurones comme un extracteur de signature faciale, sans modifier ses poids internes. Cette approche est justifiée par la similitude entre notre tâche (comparer des visages) et la tâche d'origine du modèle (reconnaître des visages).

4.2 Les embeddings faciaux et FaceNet

Un *embedding* facial est une représentation vectorielle compacte d'un visage. L'idée centrale introduite par des systèmes comme FaceNet est de mapper les images de visages vers un espace euclidien où les distances correspondent directement à la similarité des visages.

- ❑ Deux images de la même personne auront des embeddings proches (distance faible).
- ❑ Deux images de personnes différentes auront des embeddings éloignés.

Dans notre projet, chaque visage est converti en un vecteur de 512 nombres réels.

4.3 Architecture Inception-ResNet

Nous utilisons l'architecture **Inception-ResNet V1** pré-entraînée sur le dataset **VGG-Face2**. Cette architecture combine deux innovations majeures :

- ❑ **Modules Inception** : Permettent de capturer des détails à différentes échelles grâce à des filtres de convolution de tailles variées.
- ❑ **Connexions Résiduelles (ResNet)** : Facilitent l'entraînement de réseaux très profonds en permettant au gradient de mieux circuler.

Ce choix garantit une excellente capacité à extraire des traits faciaux discriminants, même avec des variations de pose ou d'éclairage.

5 Données utilisées

5.1 Dataset FairFace

Le dataset FairFace a été retenu en raison de son objectif explicite de réduction des biais en reconnaissance faciale. Contrairement à d'autres jeux de données largement utilisés, FairFace propose une répartition plus équilibrée des images selon l'origine ethnique, le genre et l'âge, ce qui en fait un choix pertinent pour une analyse qualitative équitable.

- ❑ **Description** : Il contient plus de 100 000 images de visages annotées avec l'âge, le genre et l'origine ethnique. Pour des raisons de performance locale, nous avons utilisé un sous-ensemble représentatif (environ 10 000 images).
- ❑ **Justification** : Contrairement à de nombreux datasets historiques biaisés vers des visages à peau claire, FairFace a été conçu pour être équilibré racialement (7 groupes : White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, Latino). Cela assure que notre application fonctionne équitablement pour tous les utilisateurs.

5.2 Images personnelles

Pour rendre la démonstration plus concrète et ludique, nous avons intégré un petit jeu de données personnelles (photos des membres du projet, amis, célébrités). **Précautions** : Ces

images ont été traitées avec rigueur. Elles sont intégrées uniquement au niveau des embeddings (représentation mathématique) dans le système de recherche, tout en restant séparées au niveau du stockage brut. Cela permet de tester le système sur des visages connus pour valider son bon fonctionnement.

6 Méthodologie et Justification

6.1 Justification des choix méthodologiques

Plusieurs choix méthodologiques ont été effectués dans le cadre de ce projet afin de répondre aux contraintes académiques tout en garantissant la cohérence du système proposé.

Concernant le modèle de reconnaissance faciale, le choix s'est porté sur **InceptionResnetV1**, utilisé comme extracteur d'embeddings faciaux.

D'autres architectures convolutionnelles, telles que VGG16 ou ResNet classiques, auraient pu être envisagées. Toutefois, ces modèles ne sont pas spécifiquement optimisés pour la reconnaissance faciale et nécessitent généralement une phase d'adaptation ou de ré-entraînement. À l'inverse, InceptionResnetV1, inspiré des travaux FaceNet, est conçu pour produire directement des représentations faciales discriminantes, ce qui le rend particulièrement adapté à une approche basée sur le transfert learning sans apprentissage supplémentaire.

Concernant le choix du dataset, **FairFace** a été privilégié par rapport à d'autres jeux de données populaires tels que CelebA. Bien que CelebA soit largement utilisé en vision par ordinateur, il est principalement orienté vers la reconnaissance d'attributs faciaux et présente des déséquilibres en termes de représentation ethnique. FairFace, en revanche, a été conçu explicitement pour offrir une distribution plus équilibrée des profils faciaux, ce qui constitue un avantage majeur dans un contexte d'analyse qualitative et éthique des résultats.

En ce qui concerne la mesure de similarité, la similarité cosinus a été retenue plutôt que la distance euclidienne. Cette métrique est particulièrement adaptée aux embeddings faciaux normalisés, car elle mesure la proximité angulaire entre deux vecteurs indépendamment de leur norme. Ce choix est cohérent avec les recommandations issues des travaux sur FaceNet et permet une comparaison stable et interprétable des visages dans l'espace des embeddings.

Enfin, le recours à un prétraitement systématique basé sur la détection et le recadrage des visages à l'aide de MTCNN s'est imposé comme une étape essentielle. Des expériences qualitatives ont montré que l'absence de prétraitement conduit à des embeddings moins discriminants, notamment lorsque l'image contient un fond complexe ou des variations importantes de pose. L'application d'un pipeline homogène à toutes les images garantit ainsi la validité des comparaisons réalisées.

La pipeline de traitement de notre application se décompose en quatre étapes clés.

6.2 Prétraitement des images (Détection et Alignement)

Avant toute analyse, il est crucial d'isoler le visage du reste de l'image. Nous utilisons pour cela **MTCNN** (Multi-task Cascaded Convolutional Networks).

1. **Détection** : MTCNN localise le visage et fournit une "bounding box".

2. **Alignement et Recadrage** : Le visage est découpé et redimensionné à une taille standard de 160×160 pixels.
3. **Normalisation** : Les pixels (0-255) sont normalisés pour correspondre aux attentes du modèle Inception (valeurs entre -1 et 1).

Cette étape standardise toutes les entrées, garantissant que le modèle de reconnaissance se concentre uniquement sur les traits du visage et non sur le fond ou la taille de l'image.

6.3 Extraction des embeddings

Une fois le visage préparé (tenseur 160×160), il traverse le réseau **InceptionResnetV1**.

- ❑ Le modèle fonctionne en mode *évaluation* (pas d'apprentissage).
- ❑ La sortie est un vecteur brut de dimension 512.
- ❑ **Normalisation L2** : Étape critique où le vecteur est divisé par sa norme. Cela place tous les visages sur une base de rayon 1, rendant le calcul de similarité cosinus équivalent à un simple produit scalaire.

6.4 Fusion des données

Nous avons créé un processus automatisé qui :

1. Parcourt les dossiers d'images (FairFace et Personnelles).
2. Calcule les embeddings pour chaque image valide.
3. Stocke les vecteurs dans un fichier `.npy` (NumPy) optimisé pour la lecture rapide.
4. Génère un fichier de métadonnées `meta.csv` contenant les chemins, les sources et les identités associées.

Récemment, nous avons enrichi ce fichier `meta.csv` avec les données démographiques de FairFace (âge, genre) pour permettre des analyses statistiques avancées dans l'application.

6.5 Recherche de similarité

Lorsqu'un utilisateur envoie une photo, son embedding u est comparé à tous les embeddings de la base v_i via la **similarité cosinus** :

$$\text{Similarit}(u, v_i) = \frac{u \cdot v_i}{\|u\| \|v_i\|}$$

Comme nos vecteurs sont déjà normalisés ($\|u\| = \|v_i\| = 1$), la formule se simplifie en :

$$\text{Similarit}(u, v_i) = u \cdot v_i$$

Nous trions ensuite les résultats par score décroissant pour afficher les k visages les plus proches (Top-K).

7 Interface utilisateur

7.1 Choix de l'outil : Streamlit

Nous avons choisi **Streamlit** pour développer l'interface graphique. C'est un framework Python open-source spécifiquement conçu pour la Data Science. Il nous a permis de transformer nos scripts d'analyse en une application web interactive très rapidement, sans avoir à gérer HTML, CSS ou JavaScript complexe.

7.2 Fonctionnalités avancées

L'interface a été conçue pour être moderne et intuitive (thème sombre, animations). Les fonctionnalités principales incluent :

- ❑ **Modes d'entrée flexibles** : L'utilisateur peut uploader une image ou prendre un selfie directement via sa webcam.
- ❑ **Visualisation en temps réel** : Affichage de l'image originale et du visage détecté/recadré par l'IA.
- ❑ **Résultats enrichis** : Affichage des sosies avec des badges de confiance colorés (Vert/Jaune/Rouge) selon le degré de ressemblance.
- ❑ **Statistiques analytiques** :
 - Graphique de distribution des scores pour voir où se situe le "match" par rapport à la moyenne.
 - **Analyse démographique** : Graphiques montrant la répartition d'âge et de genre des sosies trouvés, donnant une indication sur les traits perçus par l'IA.

8 Résultats et analyse

8.1 Analyse des scores de similarité

Le système de reconnaissance faciale repose sur la comparaison de embeddings faciaux à l'aide de la similarité cosinus. Pour chaque image requête, les visages de la base sont classés selon leur score de similarité, compris entre 0 et 1.

L'analyse des résultats obtenus montre que les correspondances jugées visuellement pertinentes présentent généralement des scores de similarité élevés, typiquement supérieurs à 0,7. À l'inverse, les visages dont la similarité est inférieure à 0,6 correspondent le plus souvent à des individus visuellement distincts de l'image requête.

Ces observations permettent de définir empiriquement des zones de confiance, dans lesquelles les résultats retournés par le système sont plus fiables d'un point de vue qualitatif.

8.2 Influence du prétraitement des images

Une comparaison qualitative a été réalisée entre les résultats obtenus avec et sans application du prétraitement par détection et recadrage des visages. Les expériences montrent que l'absence de prétraitement conduit à des embeddings moins discriminants, notamment lorsque l'image contient un fond complexe ou plusieurs visages.

À l'inverse, l'utilisation systématique du même pipeline de prétraitement, basé sur la détection faciale par MTCNN, améliore la cohérence des résultats et la pertinence visuelle des correspondances retournées. Cette étape apparaît ainsi comme un élément clé du système.

8.3 Analyse critique des résultats

Malgré des résultats globalement satisfaisants, certaines limites ont été observées. La performance du système dépend fortement de la qualité de l'image requête, en particulier en termes d'éclairage, de résolution et d'angle de prise de vue. Les profils très inclinés ou partiellement visibles peuvent conduire à des scores de similarité moins fiables.

Par ailleurs, bien que le dataset FairFace vise à réduire les biais en reconnaissance faciale, certains déséquilibres peuvent subsister en fonction des caractéristiques faciales et des conditions de capture. Ces limitations soulignent l'importance d'une analyse critique des résultats et ouvrent la voie à des améliorations futures.

8.4 Analyse qualitative

Les tests effectués montrent une pertinence visuelle impressionnante.

- ❑ **Fidélité des traits** : Le système capture bien la forme du visage, la présence de lunettes, de barbe, ou le type de coiffure.
- ❑ **Robustesse** : Même avec des éclairages différents, le modèle parvient souvent à retrouver la bonne personne si elle est présente dans la base personnelle.
- ❑ **FairFace** : Les correspondances dans la base FairFace, bien que n'étant pas la "même" personne, partagent souvent une structure faciale et une ethnie très proches.

8.5 Limites et biais

Malgré ses performances, le système a des limites :

- ❑ **Qualité de l'image requête** : Une photo floue, trop sombre ou avec un visage de profil extrême (non détecté par MTCNN) échoue.
- ❑ **Dépendance au prétraitement** : La qualité de l'embedding dépend entièrement de la qualité du recadrage. Si MTCNN cadre mal (ex : coupe le menton), la similarité chute.
- ❑ **Biais de la base** : Si la base de données ne contient personne ressemblant à l'utilisateur (cas rare avec FairFace mais possible), le système renverra quand même le "moins pire" résultat, avec un score de confiance faible.

9 Conclusion et perspectives

L'objectif principal de concevoir un système de reconnaissance faciale fonctionnel est atteint avec succès. Nous avons réussi à intégrer un pipeline complet de Deep Learning (MTCNN + InceptionResnetV1) dans une interface utilisateur fluide, professionnelle et interactive.

Ce projet a été extrêmement formateur sur plusieurs dimensions :

- ❑ Manipulation avancée de tenseurs et architectures de réseaux de neurones profonds.
- ❑ Importance critique du prétraitement et du nettoyage des données en vision par ordinateur.
- ❑ Utilisation pratique et efficace du Transfer Learning en environnement de production.
- ❑ Mise en place d'une interface utilisateur accessible.

Références

1. Transfert learning, Andrew Ng Transfer Learning, <https://www.youtube.com/watch?v=yofjFQddwHE>.
2. TensorFlow Transfer Learning and Fine-Tuning, https://www.tensorflow.org/tutorials/images/transfer_learning.
3. Reconnaissance faciale et embeddings. Schroff, Kalenichenko, Philbin FaceNet : A Unified Embedding for Face Recognition and Clustering, <https://arxiv.org/abs/1503.03832>.
4. Architecture du modèle utilisé. Szegedy et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, <https://arxiv.org/abs/1602.07261>.
5. Dataset d'entraînement du modèle (pré-entraînement), https://www.robots.ox.ac.uk/~vgg/data/vgg_face2/.
6. Dataset utilisé dans le projet, <https://github.com/joojs/fairface/>.
7. facenet-pytorch Implémentation officielle PyTorch, <https://github.com/timesler/facenet-pytorch>.
8. Cosine Similarity Référence pédagogique, https://en.wikipedia.org/wiki/Cosine_similarity.
9. Détection et prétraitement des visages. Zhang et al. Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Neural Networks (MTCNN), <https://arxiv.org/abs/1604.02878>.