

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique

UNIVERSITE SAAD DAHLEB DE BLIDA

Faculté des sciences

Département d'informatique



MEMOIRE DE MASTER

En Informatique

Option : Ingénierie du Logiciel

THÈME :

La Reconnaissance Faciale En 3D

Réalisé par

Ghellem Douaa Zohra

Zaaf Ikram Rania

Supervisé par : Dr.Lahiani Nesrine

Encadré par : Hamza Mekide

Devant les membres du jury

President : Dr.Midoune

Examinatrice : Dr.Bouras

2023-2024

Remerciement

Nous tenons avant tout à exprimer notre profonde gratitude envers Dieu, le Tout-Puissant, pour nous avoir accordé la volonté et la force de poursuivre nos études universitaires et de mener à bien ce projet jusqu'à son achèvement, malgré les obstacles rencontrés. Nous souhaitons également remercier Madame Lahiani Nesrine, notre encadrante interne, pour son accompagnement attentif, sa confiance et son soutien inestimable, qui nous ont permis de concrétiser et de finaliser ce travail.

Nos remerciements vont aussi aux membres du jury, qui nous font l'honneur d'évaluer notre modeste contribution. Nous vous prions d'accepter l'expression de notre sincère reconnaissance.

Enfin, nous désirons remercier notre famille, nos amis et toutes les personnes qui ont apporté leur aide, de près ou de loin, à la réalisation de ce projet. Leur soutien a été essentiel à notre réussite.

Dédicaces

Je dédie ce mémoire

À mes parents Tahar et Naziha,

Votre présence dans ma vie a été une source inépuisable de soutien, d'encouragement et d'amour. Votre confiance indéfectible et votre soutien constant ont été le socle de mon parcours et le moteur de ma réussite académique. Vous avez été les premiers à m'orienter sur le chemin de l'excellence, avec un dévouement et des sacrifices qui restent ma plus grande source d'inspiration. Vous avez nourri mes rêves avec tendresse et m'avez transmis des valeurs qui sont devenues la boussole de mon quotidien. Pour tout cela, je vous suis infiniment reconnaissant.

À ma soeur Asmaa,

Tu as toujours cru en moi, encouragé mes rêves et été là pour moi dans les moments difficiles. Ta force, ta gentillesse et ton amour inconditionnel m'ont soutenu et inspiré tout au long de ce parcours..

À mon oncle Mustapha et ma famille,

Je tiens à vous exprimer ma profonde gratitude pour l'encouragement et le soutien constants que vous m'avez apportés. Votre foi en moi a été la source de ma force et de ma détermination.

À ma binome Douaa,

Je souhaite exprimer ma profonde gratitude envers ma binome, car elle a été une source constante d'encouragement grâce à son travail, ses idées et sa patience inébranlable.

Dédicaces

Je dédie ce mémoire

À mes parents Sid Ali et Aziza,

Votre présence aimante a été un pilier inébranlable dans ma vie. Votre soutien indéfectible, vos encouragements constants et votre amour inconditionnel ont été le terreau fertile de ma réussite académique. Depuis le début, vous m'avez guidé sur la voie de l'excellence avec un dévouement et des sacrifices incommensurables, qui demeurent pour moi une source intarissable d'inspiration. Avec tendresse, vous avez nourri mes rêves et m'avez inculqué des valeurs qui sont devenues ma boussole au quotidien. Pour tout cela, je vous suis infiniment reconnaissant. Mon cœur déborde d'un amour profond pour vous.

À mon oncle Sofiane et ma tante Djamila,

Je tiens à vous exprimer ma profonde gratitude pour l'encouragement et le soutien constants que vous m'avez apportés.

À ma meilleure amie Marwa

Ton inébranlable confiance en moi, tes encouragements perpétuels à poursuivre mes rêves et ta présence réconfortante dans les moments d'adversité ont été des phares lumineux guidant mon cheminement. Ta force intérieure, ta bienveillance naturelle et l'amour inconditionnel que tu m'as prodigué ont été un ancrage solide, une source d'inspiration intarissable tout au long de cette aventure. Ton soutien indéfectible a été le vent dans mes voiles, me permettant de voguer vers des rivages insoupçonnés.

À ma binôme Ikram,

Je tiens à remercier sincèrement ma binôme pour son soutien indéfectible tout au long de notre collaboration. Ses encouragements constants ont été une véritable source d'inspiration, nous permettant d'atteindre des objectifs ambitieux en cultivant une synergie fructueuse. Son dévouement et son appui ont été essentiels à notre réussite collective.

Résumé

Pour affronter et résoudre efficacement tous les problèmes d'insécurité que l'on rencontre quotidiennement dans tous les domaines, à savoir le travail, l'accès aux administrations, aux résidences, la reconnaissance faciale a été conçue. La reconnaissance faciale, un domaine phare de la vision par ordinateur. Cette technique est basée sur la capture des données faciales d'un individu pour permettre son identification à partir d'images de leurs visages. Cependant, la reconnaissance faciale 2D traditionnelle présente des limites liées aux variations d'éclairage et d'expressions faciales, etc. qui affectent négativement les performances de ces systèmes.

Pour surmonter ces défis et résoudre ce problème, ce projet propose une nouvelle méthode de vérification de visage en 3D exploitant la modélisation tensorielle des données RGB-D. Cette approche combine les informations de texture (RGB) et de profondeur (depth 3D). La méthodologie proposée repose sur l'extraction de caractéristiques par CNN et le regroupement utilisant triplet loss function. La similarité entre les représentations d'images est mesurée par la distance Euclidienne. L'exploitation des données 3D vise à surpasser les limites de la reconnaissance 2D classique, offrant ainsi une solution sécurisée d'identification dans divers contextes professionnels et personnels.

En somme, cette recherche vise à exploiter de manière optimale les avantages des données de profondeur et des méthodes d'apprentissage profond afin de développer notre modèle avec CNN et 3D CNN. Nous utiliserons ce modèle afin de mettre en évidence le fonctionnement d'un système de reconnaissance faciale 3D (RGB-D) en respectant trois étapes fondamentales : le stockage des données, la comparaison et la prise de décision. Nous testons notre modèle, on obtient un taux de reconnaissance (accuracy) atteint 77.53% sur le jeu de donnée Texas 3D Faces [1] qui contient des variations difficiles dans la pose, l'occlusion, l'expression et l'illumination .

Mots clés : Reconnaissance faciale en 3D, RGB-D, CNN, Image en profondeur, apprentissage profond, Depth Sensors, Extraction des caractéristiques, Texas 3D faces dataset, Tensorflow

Abstract

To effectively confront and solve all the security problems encountered daily in various domains, such as work, access to administrations, residences, facial recognition has been developed. Facial recognition is a key field in computer vision. This technique is based on capturing an individual's facial data to enable their identification from images of their faces. However, traditional 2D facial recognition has limitations related to variations in lighting and facial expressions, etc., which negatively affect the performance of these systems.

To overcome these challenges and solve this problem, this project proposes a new 3D face verification method using tensor modeling of RGB-D data. This approach combines texture (RGB) and depth (3D depth) information. The proposed methodology is based on feature extraction using CNN and clustering using triplet loss function. The similarity between image representations is measured by Euclidean distance. The use of 3D data aims to surpass the limitations of classic 2D recognition, thus offering a secure identification solution in various professional and personal contexts.

In summary, this research aims to optimally exploit the advantages of depth data and deep learning methods to develop our model with CNN and 3D CNN. We will use this model to highlight the functioning of a 3D facial recognition system (RGB-D) by respecting three fundamental steps : data storage, comparison, and decision-making. We test our model and obtain a recognition rate (accuracy) of 77.53% on the Texas 3D Faces dataset [1], which contains challenging variations in pose, occlusion, expression, and illumination.

Keywords : 3D face recognition, RGB-D, Convolutional Neural Network (CNN), Depth Image, Depth Sensors, Feature Extraction, Texas 3D faces dataset, Tenserflow.

من أجل مواجهة وحل جميع مشاكل انعدام الأمن التي تواجه باستمرار في جميع المجالات، أي العمل، والوصول إلى الإدارات، والمساكن، تم تصميم التعرف على الوجه. التعرف على الوجه، مجال رؤية الكمبيوتر. تعتمد هذه التقنية على التقاط بيانات وجه unindividu للسماح بتحديد ما من صور وجوههم. ومع ذلك، فإن التعرف على الوجه ثنائي الأبعاد التقليدي له قيود تتعلق باختلافات الإضاءة وتعبيرات الوجه، وما إلى ذلك، والتي تؤثر سلباً على أداء هذه الأنظمة.

للتغلب على هذه التحديات وحل هذه المشكلة، يقترح هذا المشروع طريقة جديدة للتحقق من الوجه ثلاثي الأبعاد باستخدام نمذجة الموتر لبيانات RGB-D. يجمع هذا النهج بين معلومات الملص (RGB) والعمق (العمق 3D). تستند المنهجية المقترحة إلى استخراج الخصائص بواسطة CNN والتجميع باستخدام وظيفة فقدان ثلاثة توائم. يتم قياس التشابه بين تمثيلات الصور بواسطة ladistance الإقليدي. يهدف استغلال البيانات ثلاثية الأبعاد إلى التغلب على حدود الاعتراف التقليدي ثنائي الأبعاد، مما يوفر حلاً آمناً لتحديد الهوية في مختلف السياقات المهنية والشخصية.

باختصار، يهدف هذا البحث إلى الاستغلال الأمثل لفوائد بيانات العمق وطرق التعلم العميق من أجل تطوير نموذجنا مع CNN و 3D CNN. سنستخدم هذا النموذج لتسليط الضوء على عمل نظام التعرف على الوجه ثلاثي الأبعاد (RGB-D) الذي يحترم ثلاث خطوات أساسية: ترجيح البيانات والمقارنة واتخاذ القرار. نختبر نموذجنا، حيث وصل معدل التعرف (الدقة) إلى 77.53 على مجموعة البيانات [1] Texas 3D Faces والتي تحتوي على اختلافات صعبة في الوضع والانسداد والتعبير والإضاءة

الكلمات الرئيسية: التعرف على الوجه ثلاثي الأبعاد، CNN، RGB-D، الصورة العميقة، التعلم العميق، مستشعرات العمق، استخراج الميزات، يواجه تكساس 3D مجموعة بيانات، Tensflow

Liste des acronymes et abréviations

2D deux Dimensions

3D Trois Dimensions

RGB Red Green Blue

D Depth

RGB-D Red Green Blue-Depth

SVM Machines à Vecteurs de Support

PCA Analyse en Composantes Principales

IBM International Business Machines Corporation

ICP IBM Cloud Pak

LBP Modèles Binaires Locaux

CNN Convolutional Neural Network

DCNN/3D CNN 3DDeep Convolutional Neural Network

CPU Central Processing Unit

GPU Graphics Processing Units

TPU Tensor Processing Units

AWS Amazon Web Services

GCP Google Cloud Platform

API Application Programming Interface

Table des matières

Table des figures	X
Liste des tableaux	XII
Introduction Générale	1
Contexte de travail	1
Problématique	1
Objectifs du travail	2
Organisation du mémoire	2
1 État de l’art	4
1.1 Introduction	4
1.2 Historique	4
1.3 Les défis de la reconnaissance faciale en 2D	5
1.4 La reconnaissance faciale en 3D	8
1.5 Domaines d’application	9
1.6 Techniques de reconnaissance faciale 3D	10
1.7 Architecture générale d’un système de reconnaissance faciale 3D	14
1.7.1 L’acquisition de visages 3D	14
1.7.2 Pré-traitement.....	16
1.7.3 Extraction des caractéristiques.....	16
1.7.4 Classification et décision	16
1.8 Notions de base sur les réseaux de neurones	17
1.8.1 Les réseaux de neurones convolutionnels (CNN).....	17
1.8.2 Les réseaux de neurones convolutionnels en 3D (3D CNN/ Deep CNN)	23
1.9 Travaux existants	25
1.10 Conclusion.....	27
2 Conception du système	28

2.1	Introduction	28
2.2	Motivation	28
2.3	Schéma global de notre architecture de modèle proposée.....	30
2.4	Conception du modèle	31
2.4.1	La phase d'apprentissage	31
2.4.1.1	Les entrées (Input)	32
2.4.1.2	Pré-traitement des données	34
2.4.1.3	Extraction des caractéristiques.....	35
2.4.1.4	Output de modèle : une signature biométrique	38
2.4.1.5	Le rôle de la fonction de perte par triplets : triplet loss function	39
2.4.2	La phase de test :	40
2.5	Fonctionnement de notre système	40
2.5.1	Conservation des données.....	41
2.5.2	Comparaison et Décision	42
2.6	Conclusion.....	42
3	Implémentation et résultats	43
3.1	Introduction	43
3.2	Environnement et outils de travail.....	43
3.2.1	Matériels	43
3.2.2	Langages de programmation et logiciels	43
3.3	Texas 3D face Dataset	47
3.4	Présentation des interfaces	48
3.5	Résultats et Tests	51
3.5.1	Configuration des hyperparamètres	51
3.5.2	Métriques d'évaluation	51
3.5.3	Étude comparative et discussion.....	52
3.5.3.1	Par rapport aux méthodes conventionnelles.....	53
3.5.3.2	Par rapport aux méthodes basées sur le deep learning	54
3.6	Conclusion.....	55
	Conclusion et perspectives	56
	Bibliographie	59

Table des figures

1.1	Variation de poses	6
1.2	La variation de l'éclairage	6
1.3	La variation d'expression	7
1.4	3D face recognition	9
1.5	Taxonomie des méthodes de reconnaissance faciale 3D.....	10
1.6	Architecture générale d'un système de reconnaissance faciale 3D.....	14
1.7	illustration d'un neurone	17
1.8	Réseau neurone (a) simple (b) en profondeur	18
1.9	Les couches de CNN	19
1.10	La couche de convolution.....	19
1.11	La Couche de sous-échantillonnage	21
1.12	La Couche de flattening	22
1.13	Architecture générale des couches de 3D CNN	23
1.14	La couche conv3D.....	24
1.15	La couche de pooling 3D.....	24
2.1	Les défis de la reconnaissance faciale en 2D	29
2.2	l'application whowiaty	30
2.3	Architecture proposée du modèle de reconnaissance faciale RGB-D.....	31
2.4	Depth Sensors.....	34
2.5	La normalisation des images	35
2.6	application du DCNN sur image rgb.....	36
2.7	application du DCNN sur une image de profondeur	37
2.8	Illustration des couches entièrement connectées fusionnant les caractéristiques RGB et de profondeur.....	38
2.9	Schéma explicatif sur triplet loss function	40
2.10	Architecture proposée d'un système de reconnaissance faciale RGB-D.....	41
3.1	Logos de langage de programmation Python	44

3.2	Bibliothèques utilisées.....	45
3.3	Éditeurs de texte et logiciels.....	46
3.4	Un exemplaire du Texas 3D face dataset	47
3.5	Les images du texas 3D face dataset	48
3.6	Interface du test	49
3.7	Interface d'importation des images	49
3.8	Affichage du cas accepté	50
3.9	Affichage du cas refusé	50

Liste des tableaux

1.1	Les différentes représentations des images en 3D.....	15
1.2	Les travaux existants	26
3.1	Évaluation des performances du modèle proposé avec différentes configurations des hyperparamètres	51
3.2	Tableau comparatif des méthodes basées sur le deep learning	54

Introduction Générale

Contexte de travail

Dans un monde de plus en plus numérique, la sécurité et l'authentification sont devenues des préoccupations majeures. Ainsi, plusieurs techniques biométriques permettent d'identifier ou de vérifier l'identité d'une personne sur la base de ses spécificités physiologiques (par exemple, empreintes digitales, iris, visage, rétine, empreinte de la paume, veines de la main, voix, oreille, etc.) et de ses comportements biologiques (par exemple, signature et démarche), offrant une grande unicité et permanence pour la reconnaissance, tout en réduisant le risque de perte ou de falsification.

La reconnaissance faciale est une de ces techniques biométriques qui permet d'identifier ou de vérifier l'identité d'une personne à partir de son visage. C'est l'une des modalités biométriques les plus utilisées, car elle est facile à accepter par les utilisateurs. Cependant, les systèmes de reconnaissance faciale 2D traditionnels, bien que largement utilisés, présentent des vulnérabilités. Ces vulnérabilités peuvent compromettre la sécurité dans des applications critiques.

Pour pallier ces problèmes, les systèmes de reconnaissance faciale en trois dimensions exploitent la structure tridimensionnelle et les caractéristiques du visage en profondeur. Ces systèmes utilisent des capteurs de profondeur pour capturer et analyser le visage en 3D, ce qui leur permet de s'adapter aux conditions environnementales et d'éclairage dynamiques, de résister aux tentatives d'usurpation d'identité, et de contribuer au domaine de la reconnaissance faciale 3D, en offrant une solution performante et applicable à diverses industries.

Problématique

Malgré les avancées significatives dans le domaine de la reconnaissance faciale 2D, les systèmes traditionnels présentent des vulnérabilités qui compromettent leur fiabilité et leur sécurité. Les attaques de contrefaçon utilisant des images numériques ou des photos posent un défi majeur à ces systèmes, menaçant leur utilisation dans des applications critiques où une authentification précise, fiable et nécessaire. En outre, ces techniques en deux dimensions ont des limites, notamment leur difficulté à comprendre les détails complexes des traits faciaux.

Cette contrainte les rend moins efficaces dans des environnements où les conditions d'éclairage varient fréquemment. De plus, en se basant sur des images statiques, elles ne parviennent pas à capturer l'intégralité des caractéristiques tridimensionnelles et dynamiques d'un visage, éléments essentiels pour une reconnaissance précise et vivante.

La nécessité d'assurer une sécurité renforcée et une précision accrue dans les systèmes de reconnaissance faciale souligne l'importance de résoudre ces défis actuels. L'intégration de capteurs de profondeur, combinée à l'utilisation de technologies innovantes, est essentielle pour développer des solutions plus robustes et fiables. Dans ce contexte, ce projet de thèse vise à élaborer et à optimiser une méthode de deep learning capable de surmonter les problèmes existants en exploitant les données RGB-D fournies par les capteurs de profondeur.

Objectifs du travail

Notre projet vise à concevoir une méthode de reconnaissance faciale 3D qui combine les avantages des capteurs de profondeur et du deep learning. Cette méthode permet de capturer et d'analyser la structure tridimensionnelle du visage, offrant plus de précision et de sécurité que les méthodes 2D traditionnelles, qui se basent uniquement sur des images 2D. L'objectif est de développer une méthode capable de s'adapter à diverses conditions environnementales et scénarios d'éclairage, grâce aux données fournies par les capteurs de profondeur. Ce projet s'inscrit dans le domaine de la vision par ordinateur et présente de nombreuses applications potentielles dans les domaines de l'authentification, de la surveillance et des applications immersives.

- Comprendre les principes et les techniques de la reconnaissance faciale 3D et de l'utilisation des capteurs de profondeur.
- Gérer des jeux de données diversifiés, comprenant des images 2D et des informations dérivées des capteurs de profondeur, afin d'améliorer les systèmes de reconnaissance faciale en 2D.
- Développer, optimiser et valider une architecture de deep learning intégrant les données issues des capteurs de profondeur pour réaliser une reconnaissance faciale 3D.
- Évaluer la performance, la fiabilité et l'adaptabilité du modèle de deep learning en le comparant aux systèmes de reconnaissance faciale 3D existants.

Organisation du mémoire

Pour mener à bien notre mémoire, nous avons organisé notre travail en trois chapitres

- **Chapitre 1 État de l'art** : Dans le premier chapitre, nous allons montrer un aperçu sur l'historique de la reconnaissance faciale ainsi que sur les améliorations qu'elle a subies jusqu'à la reconnaissance faciale en 3D et montrer les domaines d'application et les travaux existants déjà sur la reconnaissance faciale en 3D.

- **Chapitre 2 Conception de la solution** : Dans le deuxième chapitre, nous allons discuter sur notre conception, solution et les méthodes que nous avons choisies parmi toutes les méthodes de reconnaissance 3D existantes.

- **Chapitre 3 Implémentation et résultats** : Dans le dernier chapitre, nous présenterons et discuterons des résultats obtenus et la réalisation de notre projet.

Enfin, nous allons clôturer notre travail par une conclusion générale résumant tout ce qui a été mentionné

Chapitre 1

État de l'art

1.1 Introduction

Avec l'avancement incessant de la technologie, la question de la sécurité devient un enjeu majeur dans tous les domaines. Les méthodes traditionnelles de sécurité, telles que l'utilisation de mots de passe, de pièces d'identité ou de clés, peuvent se révéler insuffisantes et inefficaces.

Pour cette raison, les chercheurs ont orienté leurs travaux vers des méthodes de sécurité plus efficaces, comme la biométrie, qui est devenue populaire dans les domaines nécessitant une sécurité élevée. Parmi toutes les technologies biométriques, la reconnaissance faciale est l'une des plus utilisées et adaptées. Plusieurs méthodes de reconnaissance faciale ont été proposées au cours des vingt dernières années.

Au cours des dix dernières années, la reconnaissance faciale a fait un saut qualitatif et une révolution technique en utilisant des images tridimensionnelles (3D) du visage d'un individu, ce qui permet une précision accrue par rapport aux méthodes 2D traditionnelles. Elle est capable de capturer des informations plus détaillées et structurelles du visage, qui sont difficiles à obtenir à partir d'une seule image 2D

1.2 Historique

La recherche en reconnaissance faciale peut être datée des années 1960 [2]. De 1964 à 1966, Woodrow W. Bledsoe, Helen Chan et Charles Bisson ont développé un programme informatique pour la reconnaissance faciale[2]. Le programme nécessite que l'administrateur localise les caractéristiques du visage sur la photo, qui sont ensuite utilisées pour la comparaison avec les distances et les mesures.

Malgré des inconvénients initiaux, la recherche sur la reconnaissance faciale a progressé. Peter Hart de Stanford a obtenu des résultats prometteurs en utilisant des images plutôt que des points caractéristiques. Dans les années 1970, Goldstein, Harmon et Lesk ont utilisé 21

marqueurs spécifiques pour identifier les visages avec une bonne précision [2]. Cependant, cette méthode nécessitant une mesure manuelle des caractéristiques, elle n'est pas pratique pour une utilisation à grande échelle.

En 1991, Turk et Pentland ont proposé une méthode utilisant l'analyse en composantes principales (ACP) pour traiter les données faciales[3]. C'est ce qu'on appelle l'algorithme des eigenfaces qui est déjà devenu un standard doré pour la reconnaissance faciale. Plus tard, inspirés par les eigenfaces, de nombreux algorithmes de ce type ont été proposés [4]. En 1997, Christoph von der Malsburg a conçu un système capable d'identifier les personnes sur des photos lorsque les photos ne sont pas claires [5]. À la suite de ce travail, la recherche en reconnaissance faciale a divergé en deux voies. La reconnaissance faciale par vue en 3D est proposée et implémentée dans des systèmes tels que Polar et FaceIt. Bien que la reconnaissance faciale en 2D ait connu un succès considérable, sa précision est encore significativement affectée par les changements de pose et les conditions d'éclairage [6]. De nombreux chercheurs se sont tournés vers la reconnaissance faciale en 3D en raison de ses capacités potentielles à surmonter les limitations inhérentes et les inconvénients de la reconnaissance faciale en 2D. De plus, les informations géométriques fournies par les données faciales en 3D peuvent entraîner une précision de reconnaissance plus élevée que dans le cas en 2D lorsque les conditions de pose et d'éclairage sont les mêmes [7].

À la fin des années 1980, [8] a utilisé des méthodes basées sur la courbure pour tester une petite base de données faciales en 3D et a atteint une précision de reconnaissance de 10/10. En 1996, les expériences de reconnaissance faciale de Gordon ont montré que la combinaison des vues de face et de profil peut améliorer la précision de reconnaissance [9]. Après cela, de plus en plus de recherches en reconnaissance faciale en 3D ont été proposées, en raison de la disponibilité croissante d'équipements de numérisation 3D (principalement basés sur la technologie laser et la lumière structurée).

En 2012, le deep learning a révolutionné la reconnaissance faciale 2D grâce à de grands ensembles de données. Néanmoins, la reconnaissance 3D est limitée par le manque de données. Des solutions ont été proposées, comme l'utilisation de modèles 2D pré-entraînés avec des données 3D limitées, ou la création de larges corpus synthétiques. Un modèle CNN dédié à la 3D a été introduit, formé sur 3,1 millions de scans faciaux 3D. Cela a permis à l'algorithme FR3DNet d'atteindre des performances élevées en reconnaissance faciale 3D.

1.3 Les défis de la reconnaissance faciale en 2D

Les défis de la reconnaissance faciale en 2D qui ont motivé le passage à la reconnaissance faciale en 3D sont principalement liés à la précision et à la robustesse des systèmes.

1. Variation de la pose :

La variation de pose est considérée comme un problème majeur pour les systèmes de reconnaissance faciale. Lorsque le visage est de profil dans le plan de l'image (orientation $< 30^\circ$), il peut être normalisé en détectant au moins deux traits faciaux (passant par les yeux). Cependant, lorsque la rotation est supérieure à 30° , la normalisation géométrique n'est plus possible. Comme illustré dans la Figure 1.1 [10]



FIGURE 1.1 – Variation de poses [10]

2. La variation de l'éclairage :

Les variations d'éclairage rendent la tâche de reconnaissance faciale très difficile. La reconnaissance faciale en 2D peut être particulièrement sensible à ces variations. En effet, le changement d'apparence d'un visage dû à l'illumination peut parfois être plus critique que les différences physiques entre les individus, entraînant une mauvaise classification des images d'entrée et des erreurs de reconnaissance.

En 3D, les modèles peuvent mieux prendre en compte les propriétés de réflexion et de diffusion de la lumière sur les surfaces faciales, réduisant ainsi l'impact des variations d'éclairage [11]. Comme illustré dans la Figure 1.2



FIGURE 1.2 – La variation de l'éclairage [11]

3. Limitations de la résolution des images :

Les images 2D peuvent être limitées en termes de résolution, ce qui peut affecter la précision de la reconnaissance faciale, en particulier pour les détails fins du visage. En 3D, les modèles peuvent exploiter des données tridimensionnelles plus riches pour capturer des caractéristiques faciales plus précises.

4. Limites dans la détection et la reconnaissance des expressions faciales :

Un autre facteur qui affecte l'apparence du visage est l'expression faciale (voir figure 1.3). La déformation du visage qui est due aux expressions faciales est localisée principalement sur la partie inférieure du visage. L'information faciale se situant dans la partie supérieure du visage reste quasi invariable. Elle est généralement suffisante pour effectuer une identification. Toutefois, étant donné que l'expression faciale modifie l'aspect du visage, elle entraîne forcément une diminution du taux de reconnaissance [10]. L'identification de visage avec expression faciale est un problème difficile qui est toujours d'actualité et qui reste non résolu. L'information temporelle fournit une connaissance additionnelle significative qui peut être utilisée pour résoudre ce problème [10]. Comme illustré dans la Figure 1.3 [11]



FIGURE 1.3 – La variation d'expression [11]

5. Vieillesse et le changement d'aspect :

Les visages changent d'apparence au fil du temps. Les modifications concernent la tension des muscles, l'apparence de la peau (apparition de rides), le port de lunettes, éventuellement le maquillage ou la présence d'une frange occultant une partie du front [10].

Ces défis ont incité les chercheurs et les développeurs à explorer la reconnaissance faciale en 3D comme une solution pour améliorer la précision, la robustesse et la fiabilité des systèmes de reconnaissance faciale.

1.4 La reconnaissance faciale en 3D

Au cours des dernières décennies, de nombreuses techniques de reconnaissance faciale 2D ont obtenu de bons résultats dans des environnements contrôlés, en particulier grâce à l'émergence de l'apprentissage profond. Cependant, ces techniques sont toujours confrontées aux limitations intrinsèques des images 2D, dues à des variations d'éclairage, de pose, d'expression, d'occlusion, de déguisements, de retards temporels et de qualité d'image.

La reconnaissance faciale 3D peut surpasser la reconnaissance faciale 2D avec une plus grande précision et robustesse, car elle est moins sensible à la pose, à l'éclairage et à l'expression. En d'autres termes, la reconnaissance faciale 3D est capable de gérer plus efficacement les variations d'éclairage, de pose et d'expression, ce qui la rend plus précise et robuste que la reconnaissance faciale 2D. [12]

La reconnaissance faciale 3D est devenue un sujet de recherche actif ces dernières années. Elle implique principalement l'extraction de caractéristiques à partir du visage entrant, la correspondance de ces caractéristiques avec des bases de données existantes et la prédiction de l'identité de la personne à partir du visage entrant.

L'utilisation de caractéristiques faciales riches est essentielle pour obtenir un bon résultat de reconnaissance. Dans la reconnaissance faciale 3D, les données de visage 3D sont utilisées à des fins d'entraînement et de test. Par rapport aux images 2D, les visages 3D contiennent des informations géométriques plus riches, qui peuvent fournir des caractéristiques plus discriminantes et aider les systèmes de reconnaissance faciale à surmonter les défauts et les inconvénients inhérents à la reconnaissance faciale 2D. [12]

De plus, les données 3D restent relativement inchangées après mise à l'échelle, rotation et changement d'éclairage. La plupart des scanners 3D peuvent acquérir à la fois des maillages/nuages de points 3D et des textures correspondantes. Cela nous permet d'intégrer des algorithmes avancés de reconnaissance faciale 2D dans les systèmes de reconnaissance faciale 3D pour de meilleurs résultats.[12]



FIGURE 1.4 – 3D face recognition [13]

1.5 Domaines d'application

La reconnaissance faciale en 3D a de nombreuses applications potentielles dans divers domaines.

1. Sécurité et Surveillance [14] :

- Contrôle d'accès : Les systèmes de sécurité peuvent utiliser la reconnaissance faciale en 3D pour autoriser ou refuser l'accès à des zones restreintes.
- Surveillance vidéo : La technologie peut être utilisée pour détecter et suivre les personnes dans des espaces publics ou privés, aidant à prévenir la criminalité ou les intrusions.
- Contrôle des frontières : Utilisation dans les points de contrôle d'aéroports pour créer des modèles biométriques précis à partir de scans faciaux.

2. Identification et Authentification [14] :

- La reconnaissance faciale en 3D permet d'authentifier les utilisateurs et de déverrouiller l'accès aux smartphones, ordinateurs et autres dispositifs.
- Vérification de l'identité des clients lors de transactions sécurisées : Cette technologie est utilisée par les banques et d'autres institutions financières pour s'assurer de l'identité des clients lors de transactions en ligne ou en personne.
- Identification criminelle : Acquisition de scans de visage auprès de criminels enregistrés.

- Identification à distance : Défi lors de la détection de sujets dans des situations arbitraires. Application pour l'identification des consommateurs dans un contexte commercial

Donc Ces domaines d'application démontrent le large éventail de possibilités offertes par la reconnaissance faciale en 3D, allant de la sécurité et de la surveillance à la personnalisation de l'expérience utilisateur dans divers secteurs.

1.6 Techniques de reconnaissance faciale 3D

Les techniques de reconnaissance faciale 3D peuvent être divisées en deux catégories : les méthodes conventionnelles et les méthodes basées sur le deep learning (voir la Figure 1.5).

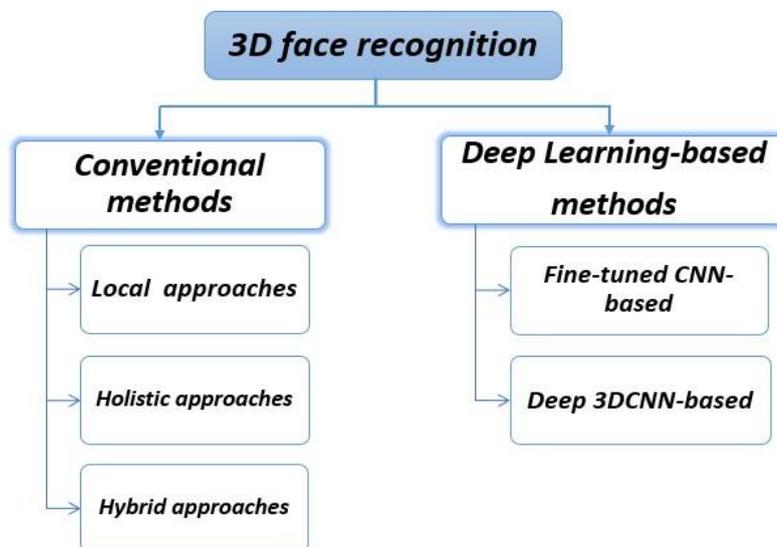


FIGURE 1.5 – Taxonomie des méthodes de reconnaissance faciale 3D [15]

Pour extraire les caractéristiques des visages, les méthodes conventionnelles utilisent toujours des algorithmes traditionnels, linéaires ou non linéaires, par exemple, l'analyse des composants principaux (ACP). Ils peuvent être divisés en trois types : local, global et hybride. En ce qui concerne les méthodes basées sur le deep learning, presque toutes utilisent des réseaux pré-entraînés (pre-trained networks), puis affinent ces réseaux avec des données converties (par exemple, des cartes 2D à partir de visages 3D). Les réseaux populaires de reconnaissance faciale basés sur l'apprentissage profond comprennent, ResNet18/ResNet50 [16], ANN [17] et les récents CNN légers tels que MobileNetV2 [18] ou bien Pointnet,.. etc.

1. Les méthodes conventionnelles : Selon l'examen de la littérature de la dernière décennie, les algorithmes conventionnels de reconnaissance faciale en 3D peuvent être classés en trois types en fonction de leurs approches d'extraction de caractéristiques, locales, globales et hybrides, Les approches locales se concentrent principalement sur les traits faciaux locaux comme le nez et les yeux [19]. Les approches globales utilisent la face entière pour générer des vecteurs de caractéristiques pour la classification des caractéristiques. Les méthodes hybrides utilisent des traits faciaux locaux et globaux. voici les détails des trois approches classiques de reconnaissance faciale en 3D :

- **les approches locales :** Au cours de la dernière décennie, des approches locales ont été développées pour décrire les informations faciales en 3D. Ces méthodes sont classées en trois types : basées sur les points clés, basées sur les courbes et basées sur les régions.[20]
 - Les méthodes basées sur les points clés (Keypoint-based methods) : détectent des points clés 3D basés sur la géométrie du visage et construisent des descripteurs en calculant les relations entre ces points, quelques modèles qui appliquent cette méthode : SIFT, SURF, BRIE, etc.
 - Les méthodes basées sur les courbes (Curve-based methods) : utilisent des courbes sur chaque surface du visage comme vecteurs de caractéristiques[20].
 - Les méthodes basées sur les régions (Local surface-based/Region-based methods) : extraient des caractéristiques de zones spécifiques du visage, quelques modèles qui appliquent cette méthode : LBP, HOG, LPQ, etc.
- **les approches globales :** Les approches globales sont conçues pour traiter l'ensemble du visage, sans nécessiter l'extraction de régions spécifiques (yeux, bouche, nez, etc.). Elles représentent l'image du visage par une matrice de pixels, convertie ensuite en vecteurs de caractéristiques. Cependant, ces méthodes sont sensibles aux variations (expressions faciales, éclairage, poses). Elles peuvent être classées en techniques linéaires et non linéaires, selon la méthode de représentation du sous-espace.
 - a) Techniques linéaires : Les méthodes linéaires sont largement utilisées pour la reconnaissance faciale et visent à améliorer la performance du système. Les techniques linéaires les plus utilisées pour les systèmes de reconnaissance faciale en 3D sont les suivantes :
 - Analyse en composantes principales (PCA)[21] : Réduit la dimensionnalité des données en projetant les images du visage sur un sous-espace de caractéristiques. Capture les variations globales dans les données.
 - Analyse discriminante linéaire (LDA) [21] : Cherche à maximiser la séparation entre les classes tout en minimisant la variance intra-classe. Tient

compte des relations entre les classes pour trouver des axes discriminants.

— Analyse en composantes indépendantes (ICA) [21] : Identifier des composantes indépendantes dans les données. Peut-être utilisée pour extraire des caractéristiques discriminantes

b) Techniques non linéaires : Les techniques non linéaires sont souvent utilisées dans la reconnaissance faciale 3D pour améliorer la précision et la robustesse. Voici quelques-unes des techniques non linéaires les plus utilisées : Kernel Principal Component Analysis (KPCA) est une extension de PCA qui détecte les relations non linéaires. KPCA est utilisé pour extraire des caractéristiques non linéaires des images de visages. Les SVM sont des modèles d'apprentissage supervisé pour la classification et la régression. Ils sont aussi utilisés en reconnaissance faciale, en s'entraînant sur des vecteurs de caractéristiques extraits des images de visage.

- **les approches hybrides** : Les systèmes hybrides de reconnaissance faciale combinent les caractéristiques locales et globales pour améliorer la précision et la robustesse. Par exemple, [22] ils utilisent l'ICP pour enregistrer une image 3D et le PCA pour restaurer la zone occluse. Ces techniques sophistiquées offrent une meilleure précision de reconnaissance, mais sont plus coûteuses en termes de calcul.

2. Les méthodes basées sur le deep learning : L'apprentissage profond a transformé la reconnaissance faciale 3D, exploitant la capacité des réseaux de neurones à apprendre et à identifier automatiquement des traits faciaux significatifs. Ces techniques surpassent les méthodes traditionnelles qui dépendent d'algorithmes dédiés pour détecter des repères faciaux spécifiques. Il permet une analyse détaillée et une adaptation précise face aux variations de pose, d'expression et d'éclairage. Cette percée offre une exactitude et une fiabilité inégalées, cruciales pour les applications avancées en matière de sécurité et de surveillance.

Récemment, le deep learning a gagné en popularité dans le domaine de la reconnaissance faciale en 3D. Cette approche extrait des caractéristiques de bas niveau pour élaborer des attributs faciaux de haut niveau via des réseaux de neurones comportant plusieurs couches cachées et s'appuyant sur d'importantes quantités de données d'entraînement. Nous résumons ici les principales stratégies d'apprentissage profond pour la reconnaissance faciale 3D (3DFR), qui sont classées en deux catégories :

- **Basé sur CNN affiné**

Les modèles de reconnaissance faciale 3D basés sur des CNN affinis représentent une avancée significative dans le domaine de l'identification biométrique. Ces modèles

exploitent des réseaux de neurones convolutionnels pré-entraînés, tels que VGG-16 ou MobileNetV2, qui ont été initialement formés sur de vastes ensembles de données d'images 2D. Cette formation initiale leur permet d'acquérir une compréhension approfondie des caractéristiques visuelles, qui sert de fondement à l'extraction de caractéristiques plus complexes. Lorsqu'il s'agit de reconnaissance faciale 3D, ces modèles subissent un processus d'affinement où ils sont ré-entraînés sur des scans faciaux 3D. Ce processus d'affinement est essentiel pour adapter le modèle aux défis uniques de la reconnaissance 3D, tels que les variations de pose, d'expression et les occlusions. Pendant l'affinement, le modèle ajuste ses poids pour se concentrer sur les caractéristiques spatiales invariantes des visages, en apprenant des informations de profondeur et des caractéristiques géométriques absentes des images 2D [23].

En fin de compte, le CNN affiné produit un ensemble de caractéristiques qui décrivent efficacement un visage en trois dimensions. Ces caractéristiques sont ensuite utilisées pour comparer et reconnaître les visages avec une grande précision, même dans des bases de données contenant un grand nombre de sujets, mais un nombre limité de scans. Ainsi, les modèles basés sur des CNN affinis pour la reconnaissance faciale 3D tirent parti de la puissance de l'apprentissage profond pour offrir des capacités de reconnaissance robustes, surpassant les limitations inhérentes aux technologies 2D.

- **Basé sur Deep 3DCNN**

Les modèles basés sur Deep 3DCNN pour la reconnaissance faciale 3D représentent une fusion de l'apprentissage profond et de l'analyse spatiale, exploitant des couches conventionnelles 3D pour traiter les données tridimensionnelles des visages. Ces réseaux apprennent à partir de scans faciaux 3D, capturant les caractéristiques superficielles ainsi que la profondeur et la forme du visage, ce qui est essentiel pour reconnaître les contours et les caractéristiques uniques d'un individu [12]. Par exemple, en utilisant des architectures comme PointNet, qui traitent directement les nuages de points, les Deep 3DCNN peuvent identifier des éléments distinctifs tels que la courbure des lèvres ou la hauteur des arcades sourcilières sans convertir les données 3D en images 2D, préservant ainsi l'intégrité des informations spatiales[24].

À mesure que les données traversent les couches du réseau, des caractéristiques de plus en plus complexes sont extraites, formant une représentation numérique détaillée du visage. Cette représentation peut être utilisée pour comparer et identifier les visages avec une grande précision, offrant une robustesse accrue même face aux variations de pose ou d'expression. Entraînés sur un ensemble diversifié d'images faciales 3D, les modèles Deep 3DCNN deviennent capables de gérer une large gamme de variations faciales, rendant ces modèles particulièrement efficaces pour les tâches de reconnaissance faciale 3D. Leur capacité à apprendre et à interpréter les données faciales complexes en trois dimensions fait des modèles basés sur Deep

3DCNN une solution de pointe dans le domaine de la biométrie faciale [25].

1.7 Architecture générale d'un système de reconnaissance faciale 3D

Un schéma général du processus de reconnaissance faciale en 3D est présenté à la Figure 1.6. Il contient quatre étapes clés : L'acquisition de visages 3D, le prétraitement des données, l'extraction de caractéristiques et la décision.

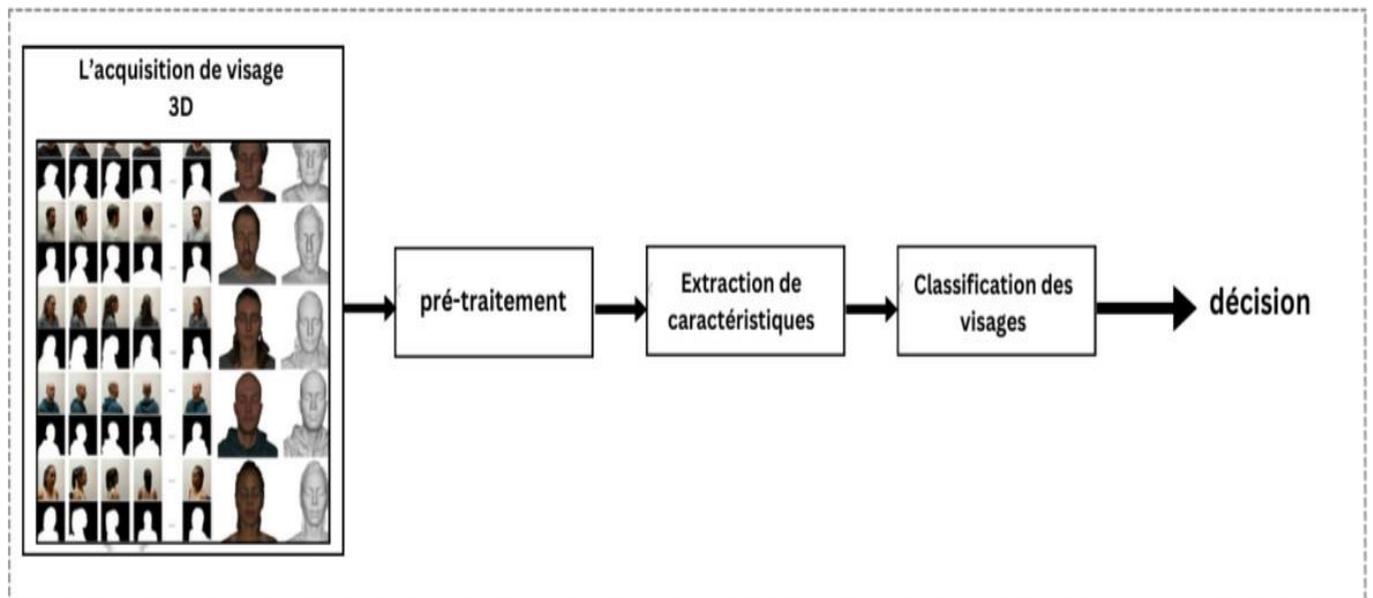


FIGURE 1.6 – Architecture générale d'un système de reconnaissance faciale 3D [15]

1.7.1 L'acquisition de visages 3D

L'étape de capture d'images est une étape essentielle dans les systèmes de reconnaissance. Avoir des images de référence de haute qualité est primordial pour améliorer les performances de reconnaissance, car cela permet de saisir l'information pertinente tout en réduisant le bruit [26]. Cette phase consiste à extraire l'image de l'utilisateur du monde réel, ce qui peut être effectué de manière statique à l'aide de capteurs 3D et de systèmes stéréo, ou de manière dynamique avec une caméra[24]. L'objectif est de capturer les données nécessaires avec la plus grande précision possible, en éliminant tout élément superflu qui pourrait interférer avec le processus de reconnaissance.

L'image obtenue peut être représentée sous différents formats, selon le type et le niveau de détail souhaités, le tableau ci-dessous illustre ces formats :

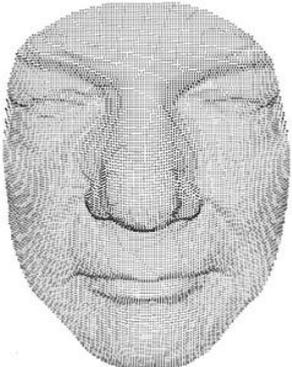
Type de Re-présentation	Images de profondeur	Nuages de points	Maillages
Définition	inclut des informations sur la distance des objets à partir d'une perspective ou d'un point de référence spécifique (comme un objectif de caméra). Une valeur est attribuée à chaque pixel pour représenter la distance de ce pixel par rapport au point de référence, ce qui crée une représentation 3D de la scène pour son image RVB ou sa scène [27]	Un nuage de points est un ensemble de points dans l'espace 3D sans connexion explicite entre eux. Chaque point contient son ensemble de coordonnées cartésiennes (x, y et z) qui définissent la position du point. Le nuage de points représente la forme ou l'objet 3D enveloppant[23].	inclut une collection de sommets (ou vertex), d'arêtes et de facettes qui définissent la géométrie d'une forme ou d'un objet en 3D. Les maillages peuvent être constitués de triangles (maillage triangulaire), de quadrilatères (quads). Chaque sommet est défini par des coordonnées cartésiennes x, y et z. Les facettes représentent les surfaces entre les sommets et sont définies par les connectivités entre eux. Pour un maillage triangulaire, chaque facette est composée de trois sommets précisément. Une arête est formée par la connexion de deux sommets[23].
Utilisation/ Application	Vision par ordinateur : la détection et la reconnaissance d'objets	Modélisation 3D, la reconstruction 3D	Graphiques 3D, jeux vidéo, simulation.
Avantages	Simple à comprendre et à utiliser.	Précision des détails, Flexibilité, Facilité d'acquisition	Visualisation de qualité, Manipulation et édition ,Interactivité
Inconvénients	Moins de détails spatiaux comparé aux autres types	Temps de traitement, Absence de formats normalisés, Grandes tailles de fichiers	Plus complexe et coûteux à générer et à stocker.
Exemple			

TABLE 1.1 – Les différentes représentations des images en 3D

1.7.2 Pré-traitement

Les données faciales 3D acquises ne peuvent pas être directement utilisées comme entrées des algorithmes d'extraction de caractéristiques, car les données contiennent les visages humains, mais aussi de nombreuses caractéristiques distrayantes comme les cheveux, les oreilles, le cou, les lunettes et les bijoux. Il est vrai que lorsque nous, les êtres humains, nous identifions les uns les autres, ces caractéristiques peuvent être utiles. Cependant, les ordinateurs ne sont pas aussi intelligents que nous pour l'instant. Les caractéristiques comme les cheveux, les lunettes et les bijoux peuvent changer de temps en temps. Les caractéristiques de l'oreille et du cou ne sont pas identifiables de manière fiable pour différentes poses de tête. Ces caractéristiques peuvent induire en erreur les algorithmes de reconnaissance faciale 3D les plus avancés et doivent donc être supprimées avant l'extraction des caractéristiques. La première étape du prétraitement consiste à détecter la position et l'orientation du visage humain. Des transformations géométriques sont utilisées pour « tourner » le visage humain directement contre l'axe de la caméra. Ensuite, le prétraitement utilise l'aide de parties faciales clairement identifiables comme le nez pour isoler la zone du visage humain des zones des caractéristiques distrayantes. Cette opération s'appelle la segmentation [28].

1.7.3 Extraction des caractéristiques

Cette étape représente le coeur du système de reconnaissance, on extrait de l'image les informations qui seront sauvegardées en mémoire pour être utilisées plus tard dans la phase de décision. Le choix de ces informations utiles revient à établir un modèle pour le visage, elles doivent être discriminantes et non redondantes. L'efficacité de cette étape a une influence directe sur la performance du système de reconnaissance de visage [29].

1.7.4 Classification et décision

La classification des visages est l'étape finale de la reconnaissance faciale. Elle consiste à évaluer le degré de similarité entre le visage cible et les visages de la galerie, qui sont les visages préalablement enregistrés dans le système. Pour cela, il faut comparer les vecteurs de caractéristiques, qui sont des représentations numériques des visages basées sur leurs attributs distinctifs. Il existe différentes mesures de distance pour effectuer cette comparaison, comme la distance euclidienne, la distance d'Hausdorff ou la distance de Mahalanobis angulaire. Ces mesures permettent de calculer un score de correspondance. Le visage qui a le score le plus élevé (ou la distance la plus faible) est considéré comme le plus proche du visage cible. Si ce score dépasse un seuil prédéfini, le visage cible est reconnu comme appartenant à la galerie. Sinon, le visage cible est rejeté comme inconnu [28].

1.8 Notions de base sur les réseaux de neurones

Dans cette partie, nous décrivons la méthode des réseaux de neurones convolutionnels (CNN) qui est utilisée de manière générale pour de nombreuses tâches de vision par ordinateur, notamment la reconnaissance faciale.

1.8.1 Les réseaux de neurones convolutionnels (CNN)

Les réseaux de neurones convolutionnels (CNN) sont les modèles les plus performants par rapport aux modèles classiques.[30]

Le réseau de neurones lui-même est une méthode populaire d'apprentissage automatique (machine learning) et d'apprentissage profond (deep learning). L'unité de base est le neurone artificiel, une fonction mathématique qui analyse les données d'entrée et produit un signal en fonction de sa sensibilité aux informations traitées . Un neurone artificiel ou formel (perceptron est défini par : $f(w+w_1x_1+w_2x_2) = y$, voir la figure 1.7 [30] .

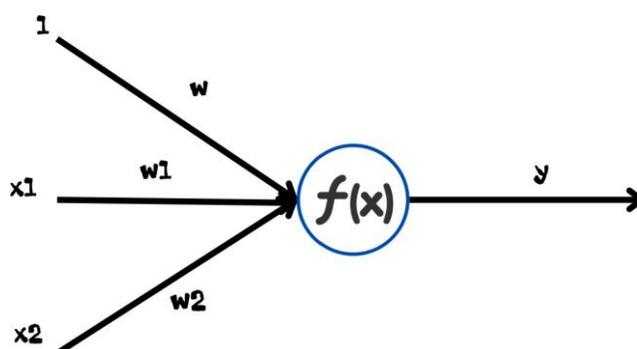
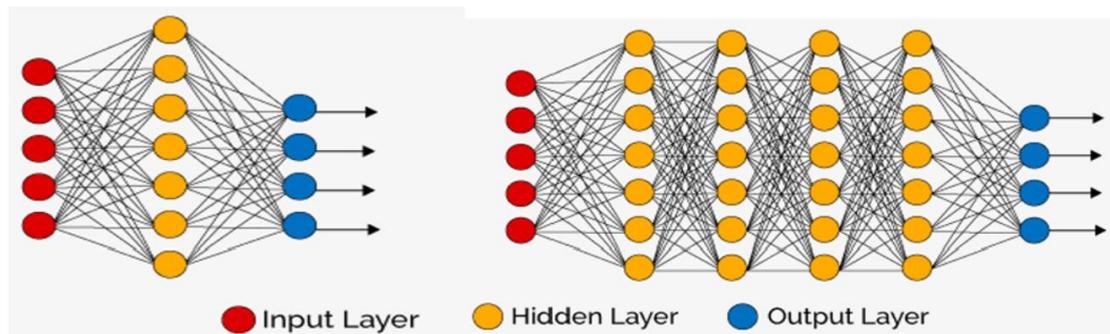


FIGURE 1.7 – illustration d'un neurone

Un réseau de neurones convolutifs (CNN) est généralement composé de couches convolutionnelles, de couches de pooling et de couches entièrement connectées. Les couches convolutionnelles permettent d'extraire des caractéristiques des données d'entrée. [30]

Chaque couche convolutionnelle effectue une opération de convolution avec un filtre et applique une fonction d'activation non linéaire. Les couches de pooling réduisent les dimensions des cartes de caractéristiques en regroupant les sorties de neurones voisins. Les caractéristiques discriminantes apprises par le CNN améliorent significativement les performances de la reconnaissance faciale. Un processus typique de reconnaissance faciale utilise un réseau de neurones convolutifs profond (DCNN) pour gérer les données en 3D. Comme illustré dans la Figure 1.8 [10]



(a) Simple neural Network

(b) Deep Learning neural Network

FIGURE 1.8 – Réseau neurone (a) simple (b) en profondeur [10]

En 2012, une révolution se produit : lors de la compétition annuelle de vision par ordinateur ILSVRC, un nouvel algorithme de Deep Learning fait exploser les records! Il s'agit d'un réseau de neurones convolutif [31].

Tensor Flow est un framework créé par Google pour la création de modèles Deep Learning. Deep Learning est une catégorie de modèles d'apprentissage automatique utilisant des réseaux de neurones multicouches [31].

Les réseaux de neurones convolutifs ont une méthodologie similaire à celle des méthodes traditionnelles d'apprentissage supervisé : ils reçoivent des images en entrée, détectent les features de chacune d'entre elles, puis entraînent un classifieur dessus [31]. Les réseaux de neurones convolutionnels ont été proposés initialement par LeCun. Ce choix a été motivé principalement par le fait qu'ils intègrent implicitement une phase d'extraction de caractéristiques et ils ont été utilisés avec succès dans de nombreuses applications [32].

Ils sont réputés pour leur robustesse aux faibles variations d'entrée et le faible taux de prétraitement nécessaire à leur fonctionnement [32].

L'architecture de CNN repose sur plusieurs réseaux de neurones profonds consistant en une succession de couches de convolution et d'agrégation (pooling) est dédié à l'extraction automatique de caractéristiques, tandis que la seconde partie, composée de couches de neurones complètement connectés, est dédiée à la classification [32]. Chaque cellule des couches de convolution est connectée à un ensemble de cellules regroupées dans un voisinage rectangulaire sur la couche précédente. Les champs récepteurs locaux permettent d'extraire des caractéristiques basiques [32]. On détaille les couches de CNN dans la figure 1.10[33] :

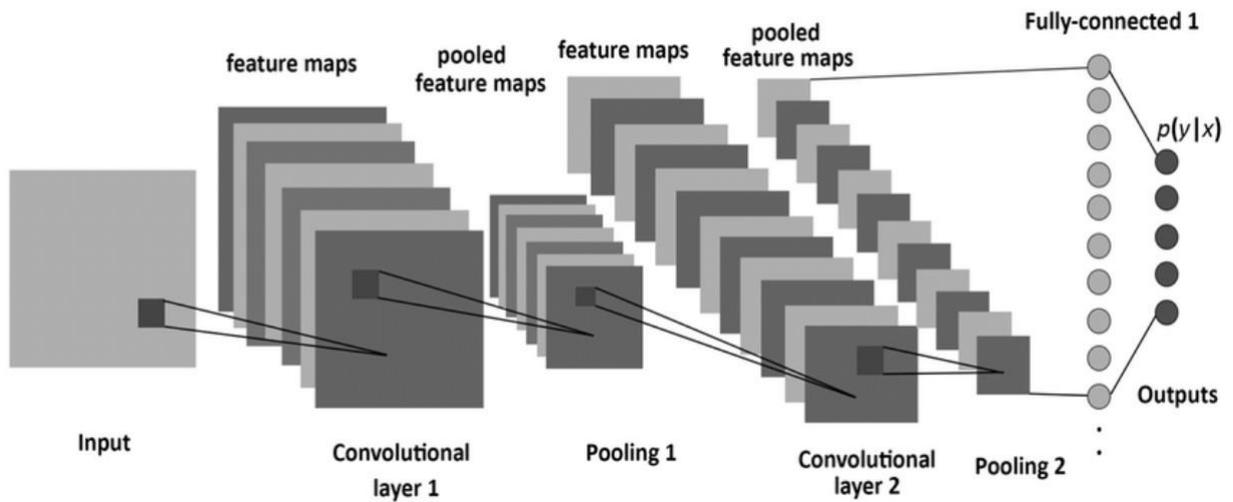


FIGURE 1.9 – Les couches de CNN [33]

1. Couche de convolution :

La couche de convolution est la composante clé des réseaux de neurones, convolutifs, elle constitue toujours au moins leur première couche.

Les couches de convolution sont formées de ce qu'on appelle des filtres. Les filtres sont des tableaux de valeurs appelées feature maps. Chaque couche de convolution intègre une image et génère une carte des caractéristiques. On obtient chaque carte des caractéristiques en appliquant le filtre à l'image.

Par exemple, si la taille de l'image est de 5x5 et que le filtre est de 3x3, la carte des caractéristiques sera de 3x3. Ainsi, la couche de convolution reçoit plusieurs images en entrée et prépare la convolution de chacune d'entre elles en utilisant chaque filtre. Les filtres correspondent exactement aux features que l'on souhaite retrouver dans les images[34]. Voir la Figure 1.10 [35]

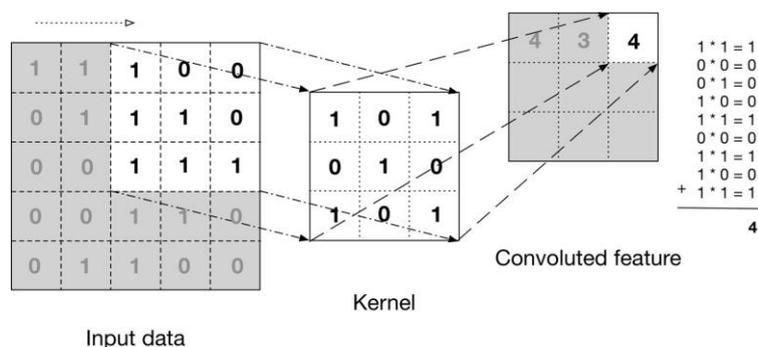


FIGURE 1.10 – La couche de convolution [35]

Fonctions d'activation

Les fonctions d'activation jouent un rôle crucial dans les réseaux de neurones convolutifs (CNN). Elles introduisent de la non-linéarité dans le modèle, ce qui lui permet de modéliser des relations complexes entre les données d'entrée et de sortie. L'une des fonctions d'activation les plus populaires pour les couches de convolution dans les CNN est la fonction ReLU (Rectified Linear Unit). La fonction ReLU est définie comme suit [30] :

$$f(x) = \max(0, x) \quad (1.1)$$

En d'autres termes, la fonction ReLU renvoie la valeur de x si x est positif, et 0 si x est négatif[30] . C'est une fonction très simple, mais elle présente plusieurs avantages par rapport à d'autres fonctions d'activation comme la sigmoïde .

- a) Non-linéarité : la ReLU introduit une non-linéarité qui permet aux CNN de modéliser des relations complexes.
- b) Calcul efficace : La ReLU est très efficace en termes de calcul, car elle ne nécessite aucune opération complexe comme l'exponentielle.
- c) Propriété de parcimonie : Lorsque les neurones reçoivent des entrées négatives, leur sortie est mise à zéro. Cela entraîne une représentation parcimonieuse des données, ce qui peut aider à éviter le sur-apprentissage.

Cependant, la fonction ReLU présente également quelques inconvénients, comme lorsque les neurones ne s'activent jamais et la non-linéarité non lisse autour de zéro. Pour résoudre ces problèmes, des variantes de la ReLU ont été proposées, comme la Leaky ReLU, la ReLU randomisée, ou encore la fonction d'activation ELU (Exponential Linear Unit)[30] .

2. Couche de sous-échantillonnage (Pooling)

Ce type de couche est souvent placé entre deux couches de convolution : elle reçoit en entrée plusieurs feature maps, et applique à chacune d'entre elles l'opération de pooling. Une couche de pooling, agit comme une couche de réduction. Elle divise l'image en blocs et ne garde que le maximum de chaque bloc. Cela permet de réduire la dimension de l'image tout en conservant les caractéristiques les plus importantes. On obtient en sortie le même nombre de feature maps qu'en entrée, mais celles-ci sont bien plus petites[34]. Voir la Figure 1.11 [34]

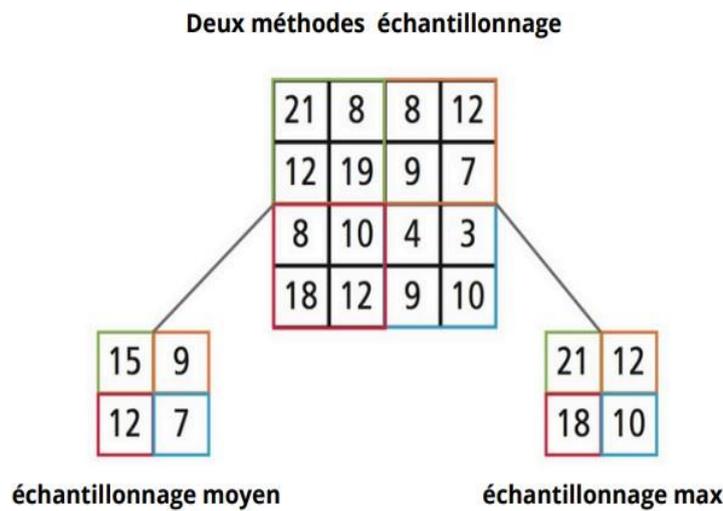


FIGURE 1.11 – La Couche de sous-échantillonnage [34]

3. Couche de flattening

Les réseaux de neurones convolutifs utilisent une couche de flattening (ou d'aplatissement). Il a pour fonction de « découper » les données en une dimension afin de les transmettre aux couches entièrement connectées (fully connected layers) qui suivent habituellement les couches convolutives. Cela permet de passer d'une représentation en plusieurs dimensions de l'image provenant des convolutions à un vecteur en une seule dimension qui pourra être traité par les couches finales fully connected. Il n'y a pas de paramètres entraînaables dans cette couche, elle simplement modifie les dimensions des données. Il s'agit d'une étape cruciale dans l'organisation des réseaux neuronaux convolutifs afin de pouvoir fusionner les couches convolutives qui extraient les caractéristiques spatiales. Voir la Figure 1.12 [34]

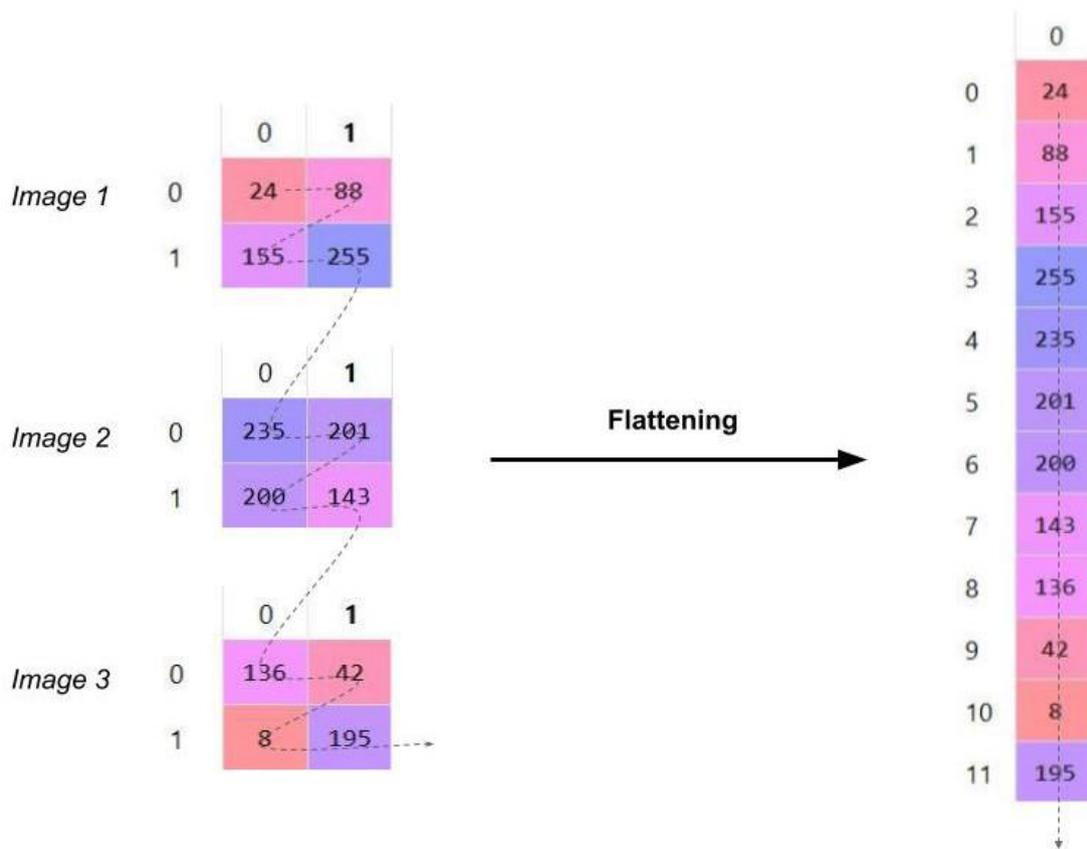


FIGURE 1.12 – La Couche de flattening [34]

4. Couches complètement connectées

Les réseaux de neurones convolutifs (CNN) sont généralement composés de plusieurs couches de convolution et de pooling, suivies d'une couche entièrement connectée qui combine les caractéristiques extraites par les couches précédentes pour classifier l'image, mais généralement n'est pas utilisée pour l'extraction de caractéristiques. Cette couche renvoie un vecteur de taille N, où N représente le nombre de classes dans notre problème de classification d'images. Chaque élément du vecteur indique la probabilité que l'image en entrée appartienne à une classe spécifique [34].

1.8.2 Les réseaux de neurones convolutionnels en 3D (3D CNN/ Deep CNN)

Un réseau de neurones convolutif 3D (3D CNN) est une extension des réseaux de neurones convolutifs 2D utilisés pour analyser et traiter des données tridimensionnelles, telles que des vidéos, les données volumineuses en 3D.

Architecture generale de 3D CNN

Les CNN 3D [36] utilisent des couches de convolution en 3D, qui appliquent des filtres 3D aux données d'entrée afin de saisir les caractéristiques spatiales. Ils intègrent également des couches de fusion 3D afin de diminuer la taille de l'espace et à la différence des CNN 2D qui traitent des images fixes, les CNN 3D prennent en entrée des données tridimensionnelles, comme illustré dans la figure 1.13 ci-dessous :

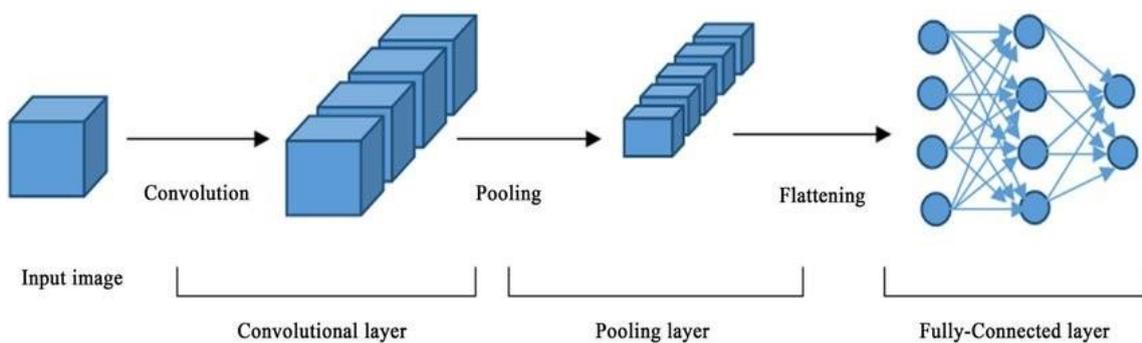


FIGURE 1.13 – Architecture générale des couches de 3D CNN [36]

Les couches d'un réseau de neurones convolutif 3D (3D CNN) et leur mécanisme fonctionnent pratiquement de la même manière que les CNN en 2D :

1. Conv3D : Couche de convolution en 3D.

La couche Conv3D applique des filtres 3D (noyaux) aux données d'entrée pour extraire les caractéristiques spatiales et temporelles[37]. Tout filtre Con3D parcourt les données d'entrée en 3D (hauteur, largeur, profondeur) et effectue un produit scalaire entre les poids du filtre et la région correspondante des données d'entrée. Les filtres 3D apprennent à détecter des motifs spécifiques dans les données, par exemple les objets, les textures, les mouvements, etc. La sortie d'une telle couche est un volume 3D appelé feature map[37]. Voir la figure 1.14 ci-dessous :

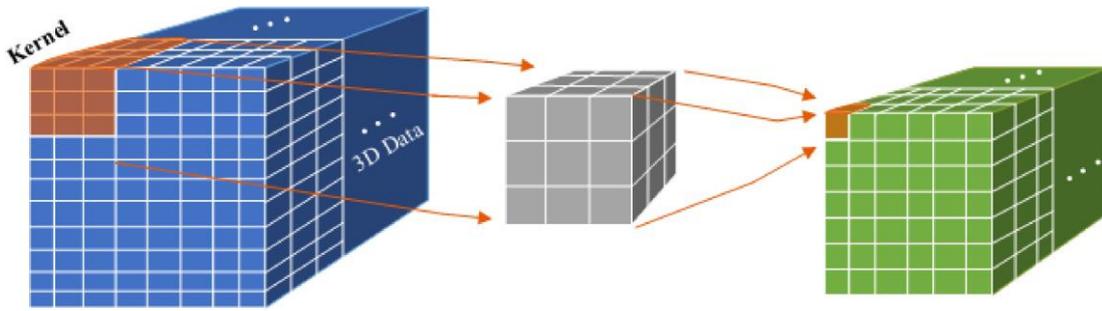


FIGURE 1.14 – La couche conv3D [37]

2. Couche de pooling 3D (MaxPool3D, AvgPool3D) [38] :

- La couche de regroupement 3D (Pooling 3D) réduit les dimensions spatiales et temporelles des cartes d'entités en appliquant une opération de regroupement (max ou moyenne) aux régions 3D.
- MaxPool3D sélectionne la valeur maximale dans chaque région, tandis qu'AvgPool3D calcule la valeur de la moyenne.
- Le regroupement réduit la taille des cartes d'entités, le nombre de paramètres dans le modèle et la complexité du réseau, tout en conservant les caractéristiques importantes.
- La mise en commun introduit également une invariance spatiale et temporelle au modèle, ce qui le rend moins sensible aux petites traductions et à la variabilité des données.

Voir la figure 1.15 [38] ci-dessous :

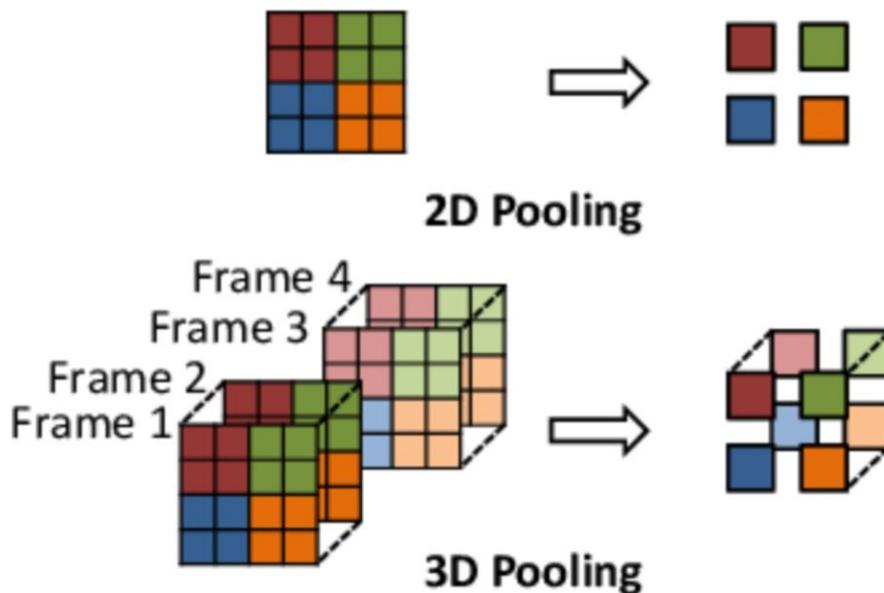


FIGURE 1.15 – La couche de pooling 3D [38]

3. La couche entièrement connectée (Dense) [36] :

- La couche entièrement connectée consiste à aplatir les feature maps 3D en un vecteur 1D après plusieurs couches de convolution et de pooling.
- Ces dernières utilisent ce vecteur comme entrée et étudient les liens entre les caractéristiques de haut niveau extraites par les couches précédentes.
- Les prédictions finales du modèle sont obtenues par la dernière couche entièrement connectée, comme les probabilités de classe pour une tâche de classification.

Ces couches sont empilées les unes sur les autres pour former l'architecture complète d'un 3D CNN. Les données d'entrée passent à travers ces couches, subissant des transformations et des extractions de caractéristiques à chaque étape. Le modèle apprend en ajustant les poids des filtres de convolution et des couches entièrement connectées pour minimiser une fonction de perte.

1.9 Travaux existants

Le tableau ci-dessous présente une synthèse des travaux récents en matière de reconnaissance faciale 3D.

Titre de travail	objectif	Points négatives	Méthodes et Technologie	taux de reconnaissance
Face recognition with the RGB-D sensor [39]	Améliorer la reconnaissance faciale en fusionnant les informations provenant de deux modalités d'image : la couleur (RGB) et la profondeur (depth).	la méthode de la fusion des caractéristiques RGB et depth ne donne pas de bon résultat [beaucoup d'information manquante]	programmation orientée objet (POO) en utilisant java	89%
PointFace : Point Set Based Feature Learning for 3D Face Recognition[40]	Apprentissage de similarités/dissimilarités entre visages 3D par réseau de triplets pour une reconnaissance faciale 3D plus précise et robuste.	Complexité des nuages de points, Dépendance à la qualité de la capture de profondeur , Limitations matérielles	python et model PointFace	85%
Face-GCN : A Graph Convolutional Network for 3D Dynamic Face Identification/Recognition [41]	Utiliser données 3D BU4DFE pour reconnaissance faciale 3D, tirant parti des informations de géométrie faciale 3D. Pour une reconnaissance robuste aux variations d'identité, d'expression et de pose	Capturer les déformations géométriques 3D du visage. Surmonter le manque de données 3D annotées en utilisant des techniques d'apprentissage par transfert, semi-supervisé ou auto-supervisé.	python et model Face-GCN	88%
Improving 2D face recognition via fine-level facial depth generation and RGB-D complementary feature learning[42]	Utilisation de la profondeur générée pour améliorer la reconnaissance faciale 2D dans des scènes complexes avec déformations de pose, éclairage inadéquat et occlusions partielles.	La profondeur générée par certaines méthodes peut souffrir de flou d'image, introduisant du bruit dans les tâches de reconnaissance faciale RGB-D ultérieures.	Python Model PointMLP Framework PyTorch	89%
Learning similarity and dissimilarity in 3D faces with triplet network [43]	Apprendre des représentations similaires et dissimilaires pour les visages 3D. Le réseau de triplets vise à rapprocher les visages similaires et à séparer les visages différents dans l'espace d'incorporation.	la sensibilité de réseaux de triplets aux dés équilibrage de données. Les réseaux triplets appropriés peut être coûteuse en termes de calcul. type de données point cloud	PointNet++ et PointNet sont des architectures de CNN pour extraire des caractéristiques à partir de nuages de points 3D. le framework PyTorch.	90%

TABLE 1.2 – Les travaux existants

1.10 Conclusion

Ce chapitre a traité de la reconnaissance faciale en 3D, un procédé biométrique offrant plus de précision et de robustesse que la reconnaissance faciale en 2D. Nous avons examiné son historique, ses avantages face aux variations de pose, d'échelle, d'éclairage et parfois d'expressions faciales. Les principales techniques et approches de reconnaissance de visages 3D ont été passées en revue, ainsi que les principaux systèmes existants et leur architecture générale. Nous avons également abordé les applications de la reconnaissance faciale en 3D. Ainsi, ce chapitre a présenté l'état de l'art, les défis et les opportunités de ce procédé.

Chapitre 2

Conception du système

2.1 Introduction

Dans le but d'améliorer la reconnaissance faciale 3D, ce sujet explore diverses méthodes pour la reconnaissance faciale, la détection du visage et l'extraction des caractéristiques, etc. Nous avons choisi de nous concentrer sur l'utilisation des images RGB-D, qui combinent des images couleur traditionnelles avec des informations de profondeur. Ces images offrent une richesse de données inégalée par les images 2D, permettant de capturer la structure tridimensionnelle du visage pour une meilleure discrimination et une meilleure résistance aux variations de pose, d'éclairage et d'expression faciale. Nous croyons que l'exploitation des images RGB-D dans la reconnaissance faciale 3D ouvre de nouvelles possibilités pour le développement de systèmes de reconnaissance plus fiables et précis, capables de fonctionner efficacement dans de réelles conditions variées.

2.2 Motivation

La technologie de la reconnaissance faciale 3D constitue une avancée majeure dans le domaine de la biométrie. Contrairement à l'identification faciale 2D, affectée par divers facteurs tels que l'âge, les changements de couleur de la peau, les changements d'expression, et à des facteurs environnementaux tels que l'éclairage, l'occlusion, comme présenté dans la figure 2.1, la reconnaissance faciale 3D surmonte ces obstacles et est de plus en plus utilisée pour l'identification humaine. Grâce à sa capacité à capturer la forme et la structure spatiale du visage, la reconnaissance faciale 3D offre une puissance discriminante supérieure, essentielle pour des applications critiques telles que la sécurité et la surveillance.

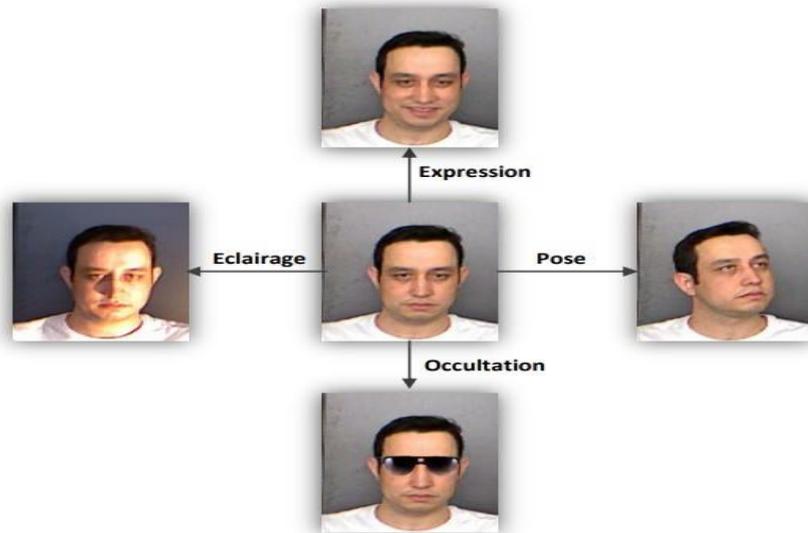


FIGURE 2.1 – Les défis de la reconnaissance faciale en 2D [11]

Whowiaty [44] représente un progrès notable dans la simplification et la sécurisation du processus d'identification en ligne. Cette plateforme propose une solution intégrale qui tire parti de l'intelligence artificielle (IA) pour l'identification des personnes à distance et en ligne, permettant ainsi une authentification instantanée et fiable des documents administratifs essentiels tels que la carte d'identité nationale et le registre du commerce. Cette vérification rapide garantit la conformité légale et une précision totale des données.

Chez Whowiaty, l'objectif est de fournir la création d'identité électronique la plus précise et la plus fiable en tirant parti des technologies avancées d'intelligence artificielle et de vision par ordinateur. Actuellement, notre système d'appariement de visage 2D fournit de bons résultats, mais il présente des limitations inhérentes affectant la précision et la fiabilité de notre processus d'identification ce qui rend cette application incomplète et contient des vulnérabilités. Pour surmonter ces défis, nous travaillons à intégrer la reconnaissance faciale 3D dans notre solution. Contrairement aux méthodes 2D, la reconnaissance faciale 3D capture la profondeur et les contours du visage, offrant une représentation plus complète et détaillée. Cela permet une correspondance plus précise, même dans des conditions environnementales et des orientations faciales variables. De plus, la technologie 3D améliore la capacité de détecter et de prévenir la fraude, car il est beaucoup plus difficile de falsifier ou de manipuler un modèle de visage 3D par rapport à une image 2D.

En adoptant la reconnaissance faciale 3D, Whowiaty améliorera non seulement la précision de la vérification d'identité, mais garantira également un niveau supérieur de sécurité et de confiance dans notre système. Cette avancée permettra d'offrir une expérience utilisateur transparente et sécurisée, répondant aux demandes croissantes de solutions d'identité numérique fiables dans un monde de plus en plus numérique. Notre engagement envers l'innovation et la sécurité fait de

Whowiaty un leader dans le domaine de la création d'identité électronique, et l'intégration de la reconnaissance faciale 3D témoigne de notre engagement envers l'excellence et l'amélioration continue.



FIGURE 2.2 – l'application whowiaty [44]

La reconnaissance faciale RGB-D exploite la forme et la structure spatiale du visage pour obtenir une identification plus précise et fiable. Cette technologie capture à la fois l'image visuelle du visage et sa profondeur, permettant ainsi de surmonter les limitations des systèmes 2D. Les algorithmes de reconnaissance faciale RGB-D utilisent les informations discriminantes des images RGB et de profondeur à travers des algorithmes de fusion d'informations sophistiqués. Cette technologie prometteuse pave la voie vers des systèmes de reconnaissance plus fiables et précis, capables de fonctionner efficacement dans des conditions réelles et variées. C'est cette technologie qu'Icosnet envisage d'intégrer dans son application Whowiaty, afin de proposer une solution d'identification plus sûre et plus pratique.

2.3 Schéma global de notre processus de modèle proposée

L'un des objectifs principaux de notre recherche est de développer un modèle innovant basé sur une architecture d'apprentissage profond. Cette architecture est conçue pour intégrer de manière fluide les caractéristiques des images 2D avec les données issues des capteurs de profondeur. L'architecture globale de notre modèle, illustrée dans la figure ci-dessous, se compose de deux phases principales : la phase d'apprentissage et la phase de test. Chacune de ces phases

comprend trois étapes clés : l'entrée (input), le processus de traitement des images et la sortie (output).

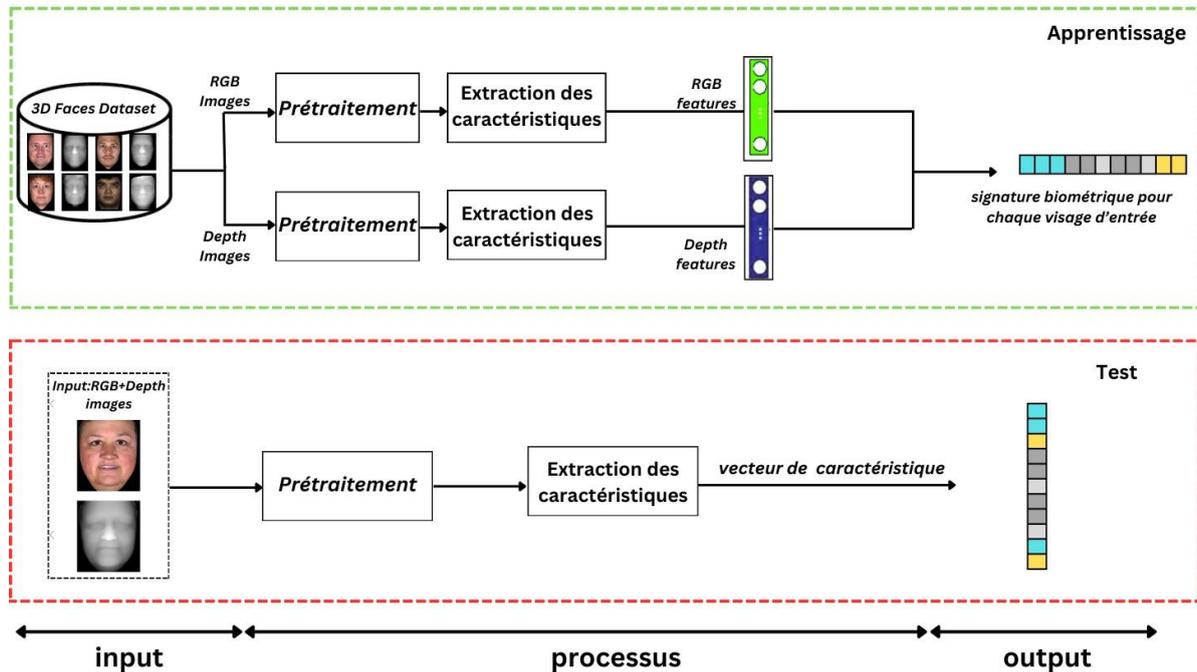


FIGURE 2.3 – Architecture proposée du modèle de reconnaissance faciale RGB-D

Dans ce qui suit, nous allons décrire les étapes détaillées du processus de conception du système, en mettant en lumière les phases essentielles de conception et de mise en œuvre nécessaires pour créer une solution optimale et intégrée.

2.4 Conception du modèle

Cette section explore la conception du modèle d'extraction des caractéristiques. Elle est divisée en trois sous-sections, chacune abordant des aspects spécifiques liés à la reconnaissance faciale RGB-D. Ces sous-sections offrent des aperçus précieux et des analyses approfondies.

2.4.1 La phase d'apprentissage

Dans la phase d'apprentissage, notre modèle est entraîné pour reconnaître les visages en utilisant à la fois des images RGB (couleur) et des données de profondeur (D). L'objectif est de minimiser la fonction de perte pour améliorer la précision de reconnaissance des visages, en exploitant les informations complémentaires fournies par les données de profondeur pour une meilleure robustesse face aux variations de pose et d'éclairage [45]. Le modèle cherche à atteindre une haute précision permettant une reconnaissance fiable même dans des conditions

difficiles [46]. Pendant cette phase, le modèle est rigoureusement entraîné pour qu'il puisse apprendre de manière efficace, améliorer progressivement sa capacité et être applicable à divers ensembles de données.

Cette phase implique les quatre étapes clés suivantes :

2.4.1.1 Les entrées (Input)

L'entrée de notre modèle conçu pour la reconnaissance faciale RGB-D se compose d'une fusion de données visuelles en couleurs (RGB) et d'informations spatiales tridimensionnelles (Depth), permettant une analyse détaillée et une perception accrue de la profondeur des traits faciaux. Les images RGB dans le modèle de reconnaissance faciale RGB-D permettent d'obtenir des informations détaillées sur la couleur et la texture du visage, tandis que les images de profondeur fournissent des informations sur la forme et la structure tridimensionnelle du visage. Grâce à cette association, la reconnaissance devient plus précise et solide, car elle permet de distinguer les caractéristiques faciales dans diverses conditions d'éclairage et d'angles de vue, et peut également améliorer la performance de reconnaissance en cas de tentatives de tromperie, telles que l'utilisation de photos ou de masques.

- **Image rvb (rgb image) :** Les images RGB des visages sont créées en utilisant le modèle de couleur Rouge-Vert-Bleu (RVB) pour représenter et stocker les images. Chaque pixel dans une image RGB est composé de trois composants de couleur : rouge, vert et bleu, et la combinaison de ces trois couleurs peut créer une large gamme de couleurs visibles. Ces images sont capturées à l'aide de caméras numériques standard qui enregistrent la lumière dans ces trois canaux de couleur [47]. Dans le modèle de la reconnaissance faciale RGB-D, les images RGB sont indispensables car elles fournissent des informations riches en couleur et en texture sur le visage d'une personne. Ces détails sont cruciaux pour identifier, détecter et reconnaître les caractéristiques uniques d'un individu [48].
- **Image de profondeur (depth image) :** Les images de profondeur sont des images en niveaux de gris 2.5D qui ont la même taille que les images RGB auxquelles elles sont associées. Le niveau de gris de chaque pixel de la carte de profondeur indique la distance à la caméra de son pixel RGB correspondant. Dans la carte de profondeur, chaque pixel détermine la position sur l'axe Z du pixel RGB correspondant. La valeur 0 des pixels de profondeur correspond aux points les plus éloignés de la scène 3D par rapport à la caméra de capture. La valeur 255 indique que les pixels se trouvent dans les plans les plus proches [49].

Les images de profondeur, dans le contexte de la reconnaissance faciale RGB-D, offrent plusieurs avantages significatifs tel que :

- 1. Amélioration de la précision :** Elles fournissent une mesure précise de la distance entre les points du visage et la caméra, ce qui permet de créer une représentation tridimensionnelle du visage. Cela aide à améliorer la précision de la reconnaissance faciale en fournissant une compréhension plus profonde de la forme et de la structure du visage.
- 2. Robustesse aux variations d'éclairage :** Contrairement aux images RGB qui peuvent être affectées par les changements de lumière, les images de profondeur sont moins sensibles aux variations d'éclairage, ce qui rend la reconnaissance plus fiable dans différentes conditions d'éclairage.
- 3. Détection des tentatives de fraude :** Les images de profondeur peuvent aider à détecter les tentatives de spoofing, comme l'utilisation de photographies ou de masques, car elles permettent de distinguer entre un visage réel et une réplique en deux dimensions.
- 4. Complémentarité avec les images RGB :** Lorsqu'elles sont combinées avec les images RGB, les images de profondeur enrichissent les données disponibles pour la reconnaissance, permettant ainsi une analyse plus complète des caractéristiques faciales.
- 5. Expériences immersives :** Elles permettent des expériences 3D immersives et offrent des affichages stéréoscopiques réalistes, ce qui peut être utilisé pour des applications de réalité augmentée ou virtuelle en plus de la reconnaissance faciale [49].

La capture des images RGB-D est réalisée par des capteurs RGB-D, qui sont des dispositifs conçus pour capturer non seulement les informations de couleur traditionnelles (RGB), mais aussi les données de profondeur (D). Il existe principalement deux types de capteurs actifs RGB-D : les capteurs à temps de vol (Time-of-Flight, ToF) et les capteurs à lumière structurée. Les capteurs à temps de vol, tels que l'Azure Kinect, mesurent la profondeur en calculant le temps que la lumière met pour voyager du capteur à l'objet et revenir. Les capteurs à lumière structurée, comme le Microsoft Kinect, projettent un motif connu sur la scène et analysent sa déformation pour calculer les informations de profondeur[49].

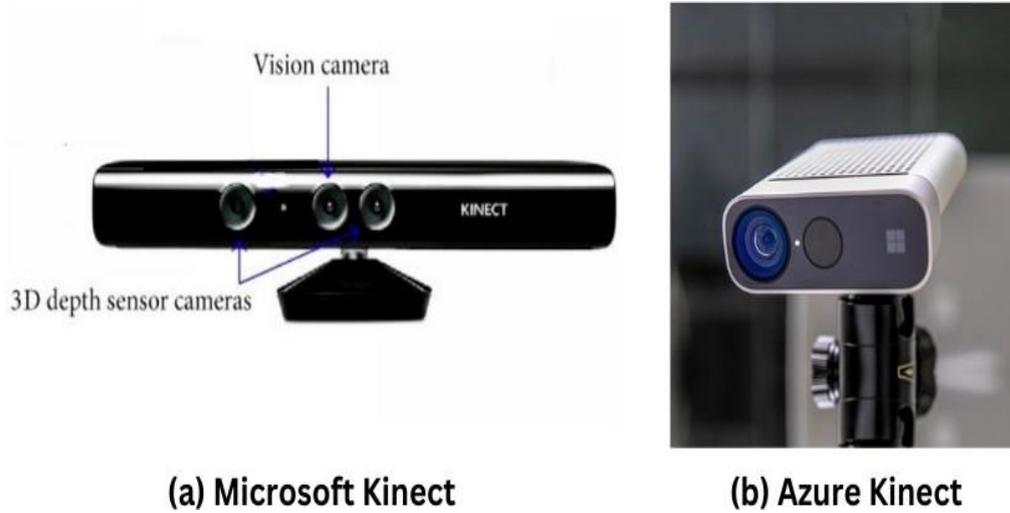


FIGURE 2.4 – Depth Sensors [49]

2.4.1.2 Pré-traitement des données

Dans notre cas, nous avons utilisé la normalisation des données comme un prétraitement selon nos besoins. Étant donné qu'elle restitue un format préétabli à toutes les images extraites de l'image brute, cette étape implique deux étapes distinctes : une normalisation géométrique en raison de la variation de la taille du visage à l'intérieur de l'image acquise en fonction de la distance entre le module d'acquisition et la personne, et une normalisation photométrique pour gérer l'illumination dans l'image [50]. Les images d'entrée sont d'abord évaluées afin de déterminer leur pertinence pour un traitement ultérieur et retirer les parties contenant des informations inutiles. Ensuite, les données collectées sont ajustées pour améliorer la qualité de la modalité, car il est important que toutes les images soient de taille identique, à la même échelle et au même format concernant les couleurs, puis un ensemble de caractéristiques saillantes et discriminatives sera extrait pour représenter/modéliser le trait biométrique [51]. Une illustration de l'étape de prétraitement d'une image RGB et d'une image de profondeur est présentée dans la figure suivante.

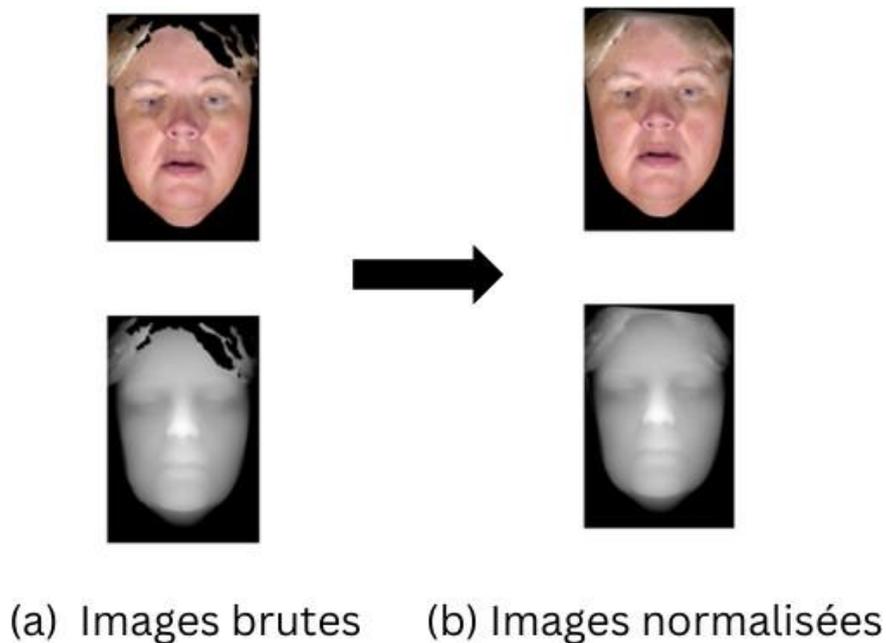


FIGURE 2.5 – La normalisation des images

2.4.1.3 Extraction des caractéristiques

Dans cette partie, nous décrivons notre solution de reconnaissance faciale 3D qui s'appuie sur approche Deep 3DCNN.

A) Extraction des caractéristique depuis image RGB :

Pour extraire les caractéristiques d'une image de visage RGB à des fins de reconnaissance faciale avec un DCNN, le processus est décrit :

1. L'image du visage RGB est utilisée en entrée, où chaque canal correspond à une composante de couleur (rouge, vert, bleu).
2. Cette image passe d'abord par les couches convolutionnelles. Les premiers filtres détectent des caractéristiques de bas niveau comme les contours, les bords, les textures, etc.
3. Au fur et à mesure que l'on progresse dans les couches convolutionnelles, les filtres apprennent à détecter des caractéristiques plus complexes et spécifiques aux visages, comme les yeux, le nez, la bouche, etc.
4. Les couches de pooling réduisent la dimension spatiale tout en conservant les informations les plus pertinentes.
5. Suite à différentes étapes de convolution et de pooling, un vecteur de caractéristiques est obtenu qui encode des informations essentielles sur le visage, telles que la forme, la texture, la profondeur, etc.

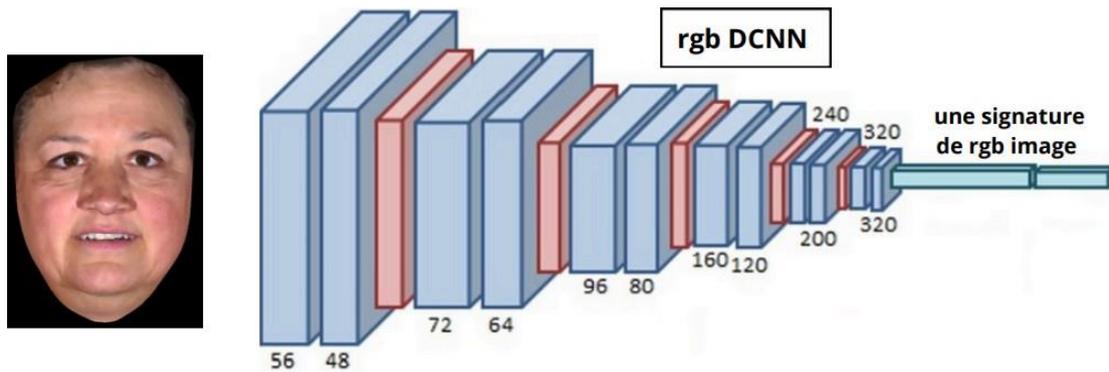


FIGURE 2.6 – application du DCNN sur image rgb

Prenons l'exemple d'une image de visage RGB de 224x224 pixels, où le vecteur de caractéristiques final pourrait être de $1 \times 1 \times 1024$, où 1024 est le nombre de caractéristiques extraites par le DCNN après toutes les couches convolutionnelles et de pooling sous forme binaire (0 et 1). le vecteur de caractéristiques appris par le DCNN pour la reconnaissance faciale 3D inclure :

- La forme globale du visage (contours, proportions, etc.)
- Les traits locaux tels que les yeux, le nez, la bouche et les sourcils sont présents sous la forme de trinômes X, Y et Z=0.
- La texture de la peau (rides, taches, etc.)
- Les expressions faciales
- information d'éclairage ... etc .

B) Extraction des caractéristiques depuis image depth :

Pour extraire les caractéristiques d'une image de visage en profondeur (depth image) à des fins de reconnaissance faciale 3D avec un 3D CNN (3D Convolutional Neural Network), le processus est similaire à celui utilisé pour les images RVB classiques, mais avec quelques différences clés liées à l'exploitation des informations de profondeur le processus décrit est :

1. L'information de profondeur (Z) est cruciale pour capturer la structure tridimensionnelle du visage ; Chaque pixel de l'image de profondeur contient une valeur scalaire représentant la distance entre ce point et le plan de la caméra. Cette information encode la géométrie et la structure 3D du visage; Le DCNN peut exploiter cette information Z de différentes manières. Par exemple, avec des filtres 3D qui prennent en compte les 3 dimensions spatiales (X, Y, Z) au lieu de 2 seulement.
2. Cette image passe d'abord par les premières couches convolutionnelles. Les filtres de ces couches sont entraînés pour détecter des caractéristiques de bas niveau liées à la profondeur, comme les bords, les discontinuités de profondeur, les surfaces planes ou courbes, etc.
3. Au fur et à mesure que l'on progresse dans les couches convolutionnelles, les filtres apprennent à détecter des caractéristiques plus complexes et spécifiques aux visages en

3D, comme la forme globale du visage, les creux (yeux, nez, bouche), les protubérances (pommettes, menton), etc.

4. L'utilisation de couches de pooling permet de diminuer la taille spatiale tout en préservant les informations de profondeur les plus importantes.
5. Après avoir effectué diverses étapes de convolution et de pooling, on obtient un vecteur de caractéristiques qui contient des informations cruciales sur le visage en 3D, y compris les données de profondeur.

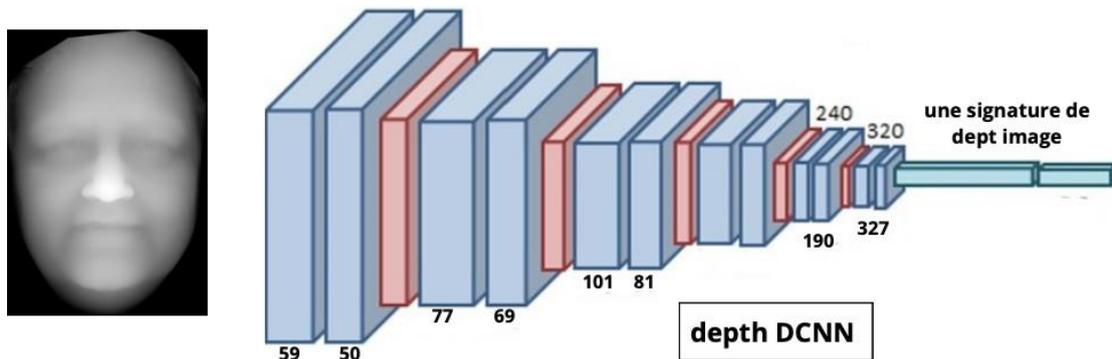


FIGURE 2.7 – application du DCNN sur une image de profondeur

Par exemple, pour une image de visage en profondeur de 224x224 pixels, le vecteur de caractéristiques final pourrait avoir une taille de 1x1x1024, où 1024 représente le nombre de caractéristiques extraites par le 3D CNN après toutes les couches convolutionnelles en 3D et de 3D pooling. Les caractéristiques apprises par le DCNN pour la reconnaissance faciale 3D à partir d'images de profondeur peuvent inclure :

- La forme globale du visage en 3D (contours, proportions, relief).
- Les caractéristiques locales en 3D comme les creux des yeux, du nez, de la bouche
- Les protubérances comme les pommettes, le menton
- Les textures de surface liées à la profondeur (rides, bosses, etc.)
- Les informations de profondeur absolue (coordonnée Z) pour chaque point du visage
- Les relations spatiales et géométriques entre les différentes parties du visage en 3D ...etc.

Le 3D CNN apprend à encoder toutes ces informations de profondeur dans un vecteur de caractéristiques compact et discriminant, permettant d'identifier de manière précise les différents visages en 3D. Un entraînement sur un grand ensemble de données d'images de profondeur de visages étiquetés est essentiel pour que le DCNN puisse apprendre les caractéristiques pertinentes liées à la profondeur.

2.4.1.4 Output de modèle : une signature biométrique

En utilisant les couches entièrement connectées (*fully connected layers*) qui jouent un rôle crucial dans un 3D CNN pour la reconnaissance faciale 3D en combinant les informations de caractéristiques extraites à partir des images RGB et de profondeur, le processus décrit est :

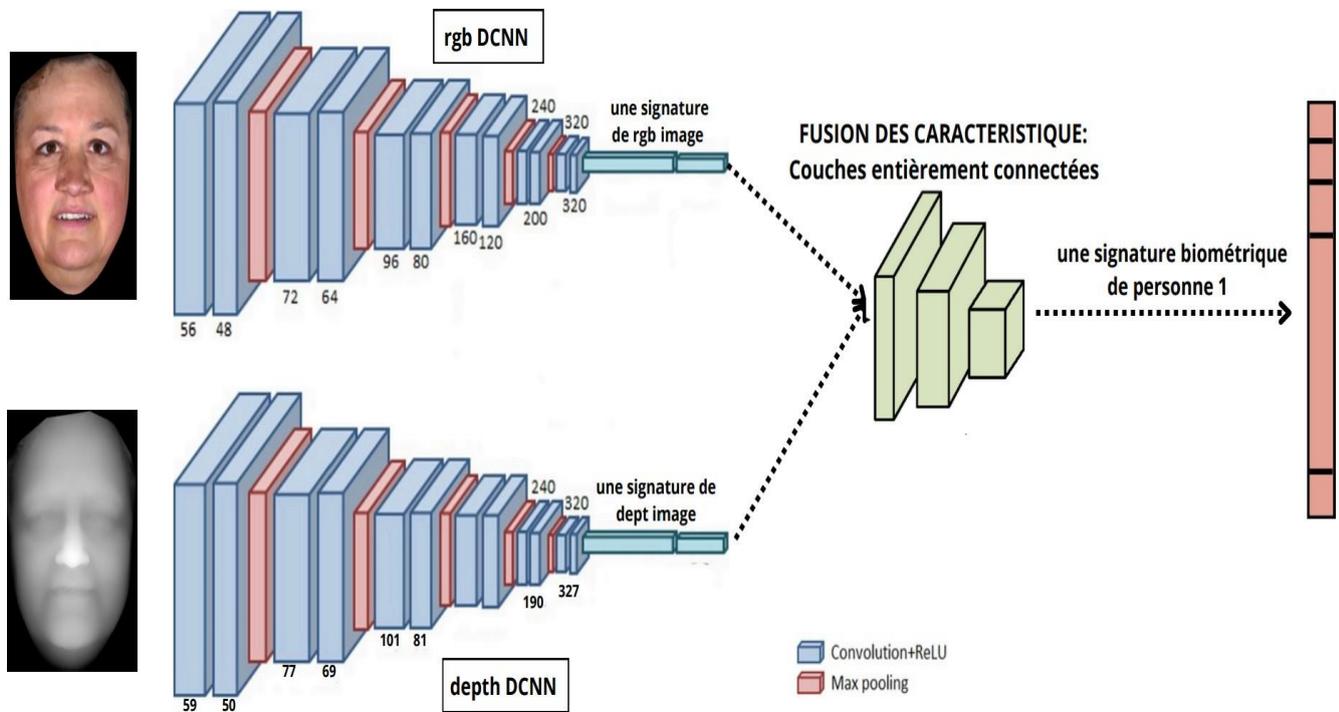


FIGURE 2.8 – Illustration des couches entièrement connectées fusionnant les caractéristiques RGB et de profondeur.

1. Nous avons deux vecteurs de caractéristiques précédemment :
 - un vecteur de caractéristiques d’images RGB \vec{f}_{RGB} ,
 - un vecteur de caractéristiques d’image en profondeur \vec{f}_{depth} .
2. Ces deux vecteurs sont concaténés pour former un seul vecteur combiné de caractéristiques $\vec{f}_{comb} = [\vec{f}_{RGB}, \vec{f}_{depth}]$. Ce vecteur combiné est alors fourni en entrée à la première couche entièrement connectée. Chaque neurone de cette couche reçoit une connexion avec toutes les caractéristiques du vecteur d’entrée.
3. Les neurones de cette couche effectuent une combinaison linéaire pondérée de toutes les caractéristiques d’entrée, suivie d’une fonction d’activation non linéaire (ReLU) :

$$\vec{h}^{(1)} = \phi \left(W^{(1)} \vec{f}_{comb} + \vec{b}^{(1)} \right) \quad (2.1)$$

où $W^{(1)}$ et $\vec{b}^{(1)}$ sont les poids et biais appris de cette couche.

4. Cette première couche entièrement connectée permet d'apprendre des représentations de plus haut niveau en fusionnant les informations RGB et de profondeur.
5. Le vecteur de sortie $\vec{h}^{(1)}$ de cette première couche est ensuite transmis à une ou plusieurs autres couches entièrement connectées supplémentaires.
6. Ces couches supplémentaires permettent d'extraire des représentations encore plus abstraites et discriminantes en combinant de manière non linéaire les caractéristiques apprises précédemment.
7. La dernière couche entièrement connectée produit un vecteur de sortie \vec{y} de taille égale au nombre de classes (identités de visages) C à reconnaître.
8. Ce vecteur final \vec{y} est passé à une fonction de perte (*loss function*) comme le Triplet loss function pour obtenir des scores de probabilité pour chaque classe.

Donc, les couches entièrement connectées jouent un rôle clé en fusionnant les informations RGB et de profondeur dans une représentation unifiée et hautement discriminatoire pour la reconnaissance faciale 3D.

2.4.1.5 Le rôle de la fonction de perte par triplets : triplet loss function

La perte par triplets a diverses applications dans les tâches de reconnaissance, de vérification et de clustering. [43]

L'entrée d'un réseau de triplets est un triplet de trois numérisations faciales 3D, dénoté par un triplet $\langle \text{positif}, \text{anc}, \text{négatif} \rangle$ où les échantillons faciaux positif et anc appartenent à la même classe, tandis que l'échantillon facial négatif appartient à une classe différente de celle de l'anc [43]. L'idée derrière la perte par triplets est de distinguer les visages 3D selon les distances entre un échantillon 3D anc, un échantillon 3D positif et un échantillon 3D négatif dans un espace de plongement. La tâche de la perte par triplets est de minimiser la distance au carré entre les échantillons de visage 3D anc et positif et de maximiser la distance entre les échantillons de visage 3D anc et négatif.[43]

Soit "anc" une numérisation faciale 3D anc, "pos" une numérisation faciale 3D positive et "neg" une numérisation faciale 3D négative.[43] Soit $d(\text{anc}, \text{pos})$ la distance au carré entre les numérisations anc et positive, tandis que $d(\text{anc}, \text{neg})$ représente la distance au carré entre les numérisations anc et négative. Alors ces distances doivent suivre l'inégalité donnée ci-dessous[43] :

$$d(\text{anc}, \text{pos}) < d(\text{anc}, \text{neg}) \quad (2.2)$$

Dans certains cas, le réseau cesse d'apprendre et calcule zéro pour tout. Pour éviter ce genre de situations, une marge (m) est introduite, qui est une frontière de décision entre $d(\text{anc}, \text{pos})$ et $d(\text{anc}, \text{neg})$. Avec cela, l'inégalité ci-dessus peut être écrite comme suit[43] :

$$d(\text{anc}, \text{pos}) - d(\text{anc}, \text{neg}) + m < 0 \quad (2.3)$$

Donc, l'objectif est de minimiser la perte par triplets. Cela signifie indirectement que la dissimilarité entre l'ancre et les numérisations positives doit être minimisée, tandis qu'elle doit être maximisée entre l'ancre et les numérisations négatives.

Pour y arriver, on calcule les gradients de la fonction de perte par triplets, c'est-à-dire comment cette perte varie en fonction des poids du réseau. Ensuite, on met à jour les poids du réseau dans la direction qui diminue la perte, à l'aide de ces gradients. Ainsi, au fur et à mesure des mises à jour, le réseau apprend à encoder les visages similaires par des représentations proches, et les visages différents par des représentations éloignées dans l'espace de représentation appris. Cet espace permet alors de bien discriminer et classifier les identités, en regroupant les images proches pour une même identité et en séparant les identités différentes. Voir la Figure 2.9 [52].

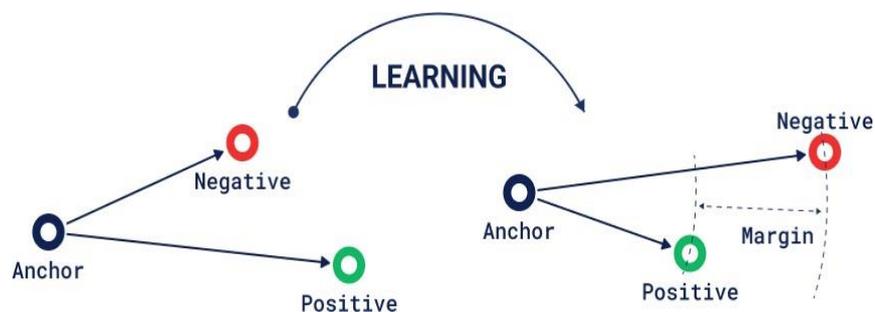


FIGURE 2.9 – Schéma explicatif sur triplet loss function [52]

2.4.2 La phase de test :

Dans cette phase, le visage cible est soumis à des étapes similaires à celles de la phase d'entraînement. Cette étape est essentielle pour évaluer si le modèle peut appliquer les connaissances acquises pendant l'entraînement à des données inédites. Son objectif est de vérifier si le modèle peut réellement généraliser et reconnaître de nouveaux visages avec précision, en exploitant les caractéristiques extraites lors de l'apprentissage.

2.5 Fonctionnement de notre système

Dans cette section, nous allons expliquer comment fonctionne le système de reconnaissance faciale RGB-D en utilisant le modèle développé comme expliqué dans la figure 2.10 .

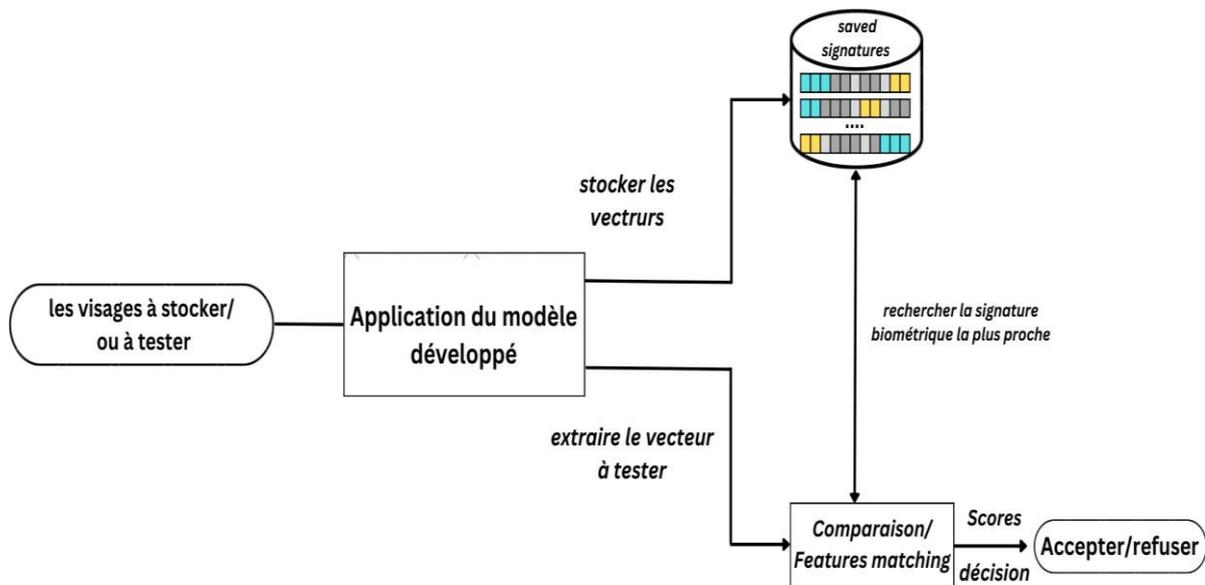


FIGURE 2.10 – Architecture proposée d’un système de reconnaissance faciale RGB-D

2.5.1 Conservation des données

Au cours de cette étape, nous mettons en œuvre attentivement notre modèle de reconnaissance faciale afin d’extraire des caractéristiques distinctives à partir d’une base de données complète, incluant obligatoirement des images RGB et leurs cartes de profondeur correspondantes. Dans notre scénario spécifique, ce processus sera mis en place pour les employés d’Icosnet, afin de générer des vecteurs RGB-D précis pour chaque individu. Ces vecteurs, uniques à chaque personne, constitueront l’identifiant exclusif de chacun, assurant ainsi une identification à la fois fiable et sécurisée (voir la figure).

Une fois ces caractéristiques identifiées, elles sont sécurisées et stockées dans une base de données dédiée. Cette base de données ne sert pas seulement de lieu de stockage ; elle est la pierre angulaire du système, servant de référence centrale pour toutes les opérations de reconnaissance faciale ultérieures. Les vecteurs de caractéristiques sont enregistrés avec soin et indexés de manière à être associés à l’identité correspondante de chaque individu. Cela crée une bibliothèque de référence robuste et fiable, permettant au système de comparer rapidement et précisément les caractéristiques extraites avec celles stockées lors des tentatives de reconnaissance.

La sécurité de cette base de données est primordiale, car elle contient des informations sensibles. Nous prenons des mesures rigoureuses pour garantir la protection et l’intégrité de ces données, assurant ainsi la confidentialité et la sécurité des informations personnelles des employés.

2.5.2 Comparaison et Décision

Quand on veut identifier un nouveau visage cible, on l'envoie aussi dans le réseau pour extraire son vecteur de caractéristiques.

Pour déterminer à quelle identité de la galerie ce visage cible correspond, on calcule une mesure de similarité/dissimilarité entre son vecteur et les vecteurs de tous les visages connus.

La mesure de distance la plus couramment utilisée est la distance Euclidienne dans l'espace des caractéristiques. Plus la distance euclidienne entre deux vecteurs est petite, plus les visages correspondants sont considérés proches/similaires selon les caractéristiques apprises. On a noté que :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.4)$$

x_i et y_i sont les composantes des vecteurs x et y respectivement.

Donc, on prédit que le visage cible correspond à l'identité du visage connu ou bien non, dont le vecteur de caractéristiques a la distance euclidienne minimale avec celui du visage cible.

2.6 Conclusion

Dans ce chapitre, nous avons présenté la conception globale et détaillée de notre modèle. Nous avons commencé par une description générale, puis nous avons mis en avant la méthode suggérée pour extraire les caractéristiques. Ensuite, nous nous sommes concentrés sur le réseau de neurones convolutifs, qui constitue le fondement de ce domaine. Enfin, nous avons discuté du fonctionnement du système de reconnaissance faciale RGB-D en appliquant notre modèle. Le chapitre suivant portera sur l'implémentation de cette solution et l'évaluation de son efficacité.

Chapitre 3

Implémentation et résultats

3.1 Introduction

Dans ce chapitre, nous allons présenter la partie de réalisation de notre projet. Notre objectif dans le développement était de parvenir à gérer des jeux de données diversifiés, comprenant des images 2D et des informations dérivées des capteurs, et d'optimiser et de valider une architecture de deep learning qui intègre les données issues des capteurs de profondeur pour réaliser une reconnaissance faciale 3D (RGB-D). Tout d'abord, nous présenterons l'environnement et les outils de travail utilisés. Ensuite, nous présenterons la métrique d'évaluation, ainsi que les résultats et les tests.

3.2 Environnement et outils de travail

3.2.1 Matériels

Le matériel utilisé consiste en 2 ordinateurs personnels :

- Le premier poste de travail dispose d'un système d'exploitation Windows 10 Professionnel 64 bits. Il est équipé d'un processeur Intel(R) Core (TM) i7-3635 OM CPU @ 2.40 GHz, et d'une mémoire RAM de 8 Go et ROM : 128 Go (SSD).
- Le deuxième dispose d'un système d'exploitation Windows 10 de type 64 bits. Les caractéristiques de cet ordinateur sont un processeur Intel(R) Core i5-5200U CPU @ 2.3 GHz, une RAM de 4 Go et une ROM : 800 Go (HDD).

3.2.2 Langages de programmation et logiciels

Au cours du développement de notre système, nous avons employé divers langages de programmation et logiciels. Voici une courte description de ces outils utilisés :

- Python : un langage de programmation de haut niveau qui prend en charge la programmation impérative structurée, fonctionnelle et orienté objet. Il se distingue par son typage dynamique solide et sa gestion automatique de la mémoire. De plus, il est livré avec plusieurs bibliothèques qui simplifient le processus de développement[53].

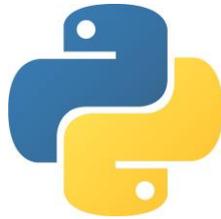


FIGURE 3.1 – Logos de langage de programmation Python

Outils et bibliothèques

- TensorFlow : À l’origine, TensorFlow s’appelait DistBelief et était un système propriétaire dédié aux réseaux de neurones de Deep Learning. En 2015, Google a rendu le code source de DistBelief open source et l’a renommé TensorFlow.
TensorFlow est une boîte à outils qui permet de résoudre des problèmes mathématiques complexes liés au Machine Learning. Elle permet aux chercheurs de développer des architectures d’apprentissage expérimentales et de les transformer en logiciels. TensorFlow propose un grand nombre de modèles et d’algorithmes pour la classification, la reconnaissance d’image, les réseaux de neurones récurrents, la traduction automatique [54]
- Keras : est une API de réseau de neurones écrite en langage Python.[55] Il s’agit d’une bibliothèque Open Source, exécutée par-dessus des frameworks tels que Theano et TensorFlow. Conçue pour être modulaire, rapide et simple d’utilisation, Keras a été créée par l’ingénieur François Chollet de Google. Elle offre une façon simple et intuitive de créer des modèles de Deep Learning. Elle permet de créer très facilement des "layers" pour les Neural Networks ou de mettre en place des architectures complexes. Un modèle Keras est constitué d’une séquence ou d’un graphique indépendant, et il existe plusieurs modules entièrement configurables et pouvant être combinés pour créer de nouveaux modèles. Keras est donc très flexible, et adapté à la recherche et à l’innovation[56].
- OpenCV : qui signifie “Open Computer Vision”, est une bibliothèque libre initialement développée par Intel. Elle est utilisée pour travailler sur des images et des vidéos. OpenCV est très importante pour la vision par ordinateur, l’apprentissage automatique et la manipulation d’images[57].
OpenCV propose également des outils pour la reconnaissance faciale. Elle permet d’encoder des images de visages, de comparer ces encodages pour trouver des correspondances,

et même de réaliser une reconnaissance faciale en temps réel avec une webcam[57].

- NumPy : qui signifie "Numerical Python", est une bibliothèque Python extrêmement populaire conçue principalement pour les calculs mathématiques et scientifiques. Elle fournit une interface permettant de stocker et d'effectuer des opérations sur les données. Les tableaux NumPy sont similaires aux listes Python, mais ils offrent des performances nettement améliorées, en particulier pour les tableaux de grande taille qui jouent un rôle central dans l'écosystème de la Data Science[58].
- Matplotlib : Matplotlib est une bibliothèque de visualisation de données en Python. Elle a été créée par John D. Hunter en 2003 et c'est une bibliothèque open source. Matplotlib est capable de créer une grande variété de graphiques, y compris des histogrammes, des graphiques à barres, des graphiques de dispersion, des graphiques en courbes, des graphiques 3D, des graphiques géographiques et bien plus encore. Elle est généralement utilisée avec d'autres bibliothèques Python comme NumPy et Pandas pour manipuler des données et créer des visualisations.[59]
- Tkinter : Tkinter est une bibliothèque Python permettant de créer facilement des interfaces graphiques (GUI) composées de widgets. Elle offre un moyen simple d'intégrer des APIs, notamment dans le domaine du machine learning ou Deep learning. Tkinter peut servir de frontend visuel où l'utilisateur saisit des entrées, qui sont ensuite transmises à l'API. Les résultats renvoyés par l'API sont alors affichés de manière conviviale dans l'interface Tkinter. Cette approche combinant Tkinter et des APIs de ML comme TensorFlow permet de déployer et d'exposer interactivement des modèles d'apprentissage automatique via une application graphique. [60]

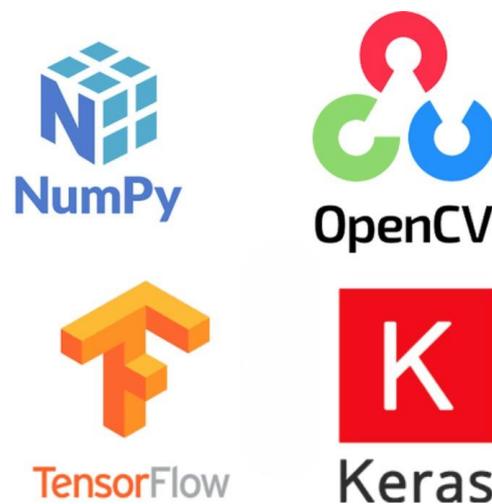


FIGURE 3.2 – Bibliothèques utilisées

Logiciels et éditeurs de textes

- Pycharm : est un environnement de développement intégré (IDE) utilisé pour la programmation en Python. Il est développé par la société tchèque JetBrains et est disponible pour windows, macOS et Linux. PyCharm propose une analyse de code, un débogueur graphique, un testeur unitaire intégré, une intégration avec les systèmes de contrôle de version, et prend en charge le développement web avec Django. Il existe deux versions de PyCharm : la Community Edition, gratuite et open source, et la Professional Edition, qui est un service d'abonnement payant. Que vous soyez débutant ou développeur expérimenté, PyCharm offre des fonctionnalités telles que l'achèvement intelligent du code, l'inspection du code, le débogage, les tests et la gestion de version, ce qui en fait un choix populaire pour les projets Python[61].
- Google colab : également connu sous le nom de Colab, est un service hébergé par Google qui permet d'exécuter des Jupyter Notebooks sans aucune configuration préalable. Il offre un accès gratuit à des ressources de calcul, notamment des GPU et des TPU (Tensor Processing Units) Colab est particulièrement adapté au Machine Learning, à l'analyse de données et à l'éducation. Les notebooks Colab sont stockés dans Google Drive et peuvent être facilement partagés avec d'autres utilisateurs pour une collaboration en temps réel [62].



FIGURE 3.3 – Éditeurs de texte et logiciels

Pourquoi les GPU?

Pour la plupart des approches de l'apprentissage en machine ou l'apprentissage en profondeur, les GPU sont essentiels en raison de la quantité de données sur laquelle nous devons opérer. Le calcul mathématique de Deep Learning sur un CPU peut prendre des mois! Mais ces calculs peuvent être assignés à des GPU pour un calcul plus rapide. Il n'est pas recommandé de former un réseau de neurones sur un processeur, un GPU est nécessaire pour un calcul élevé.[63]

Comme nous le savons tous, les GPU sont chers, de même que les plates-formes cloud telles qu'AWS et GCP. Alors, où pouvons-nous obtenir un GPU? Colab vient à la rescousse! Colab nous fournit un GPU Nvidia Tesla K80 gratuit dont le prix est d'environ 1,7 lakh Roupie et qui est une véritable bête dans la performance. nous avons utiliser ce GPU pour former notre réseau de neurones sur le cloud et profiter des performances rapides et sans tracas.[63]

3.3 Texas 3D face Dataset

L'ensemble de données de reconnaissance faciale Texas 3D Face est à la disposition des chercheurs en trois dimensions (3D) pour la reconnaissance faciale et d'autres domaines connexes. Cette base de données contient 1149 paires d'images haute résolution, de pose en couleurs et en profondeur de 118 sujets humains adultes acquises à l'aide d'une caméra stéréo. Les images sont accompagnées d'informations sur le sexe, l'ethnicité et l'expression faciale du sujet (voir la figure 3.4. [1][64][65].



FIGURE 3.4 – Un exemple du Texas 3D face dataset

L'acquisition de données :

Les images 3D de la base de données de reconnaissance faciale 3D du Texas ont été acquises à l'aide d'un système d'imagerie stéréo MU-2 fabriqué par 3Q Technologies Ltd. Les images 3D acquises ont été transformés avec succès en une orientation frontale avec le front incliné de 10° par rapport à l'axe vertical. Cela a été réalisé en alignant itérativement les modèles

faciaux dans des poses arbitraires sur un visage modèle dans une pose frontale canonique à l'aide de l'algorithme ICP. Incliner le front du visage 3D de 10° par rapport à l'axe vertical a assuré que chaque emplacement (x, y) était associé à une valeur z unique, et donc la surface faciale pouvait être représentée comme $z = f(x, y)$.

Dans la base de données, chaque visage est représenté par une paire d'images RGB +Depth. Comme présenté dans la figure 3.5. Ces images sont obtenues en projetant orthographiquement les modèles 3D sur une grille rectangulaire. Les images de portée sont construites en ajustant un maillage triangulé au nuage de points 3D et en estimant les valeurs z. Les images de couleur sont construites de manière similaire. Chaque image a une taille de 751 * 501 pixels avec une résolution de 0,32 mm. Les valeurs z sont représentées dans un format de 8 bits, avec la valeur la plus élevée de 255 attribuées au bout du nez. Les images en couleur sont représentées dans un format RVB non compressé de 8 bits.[1][64][65].

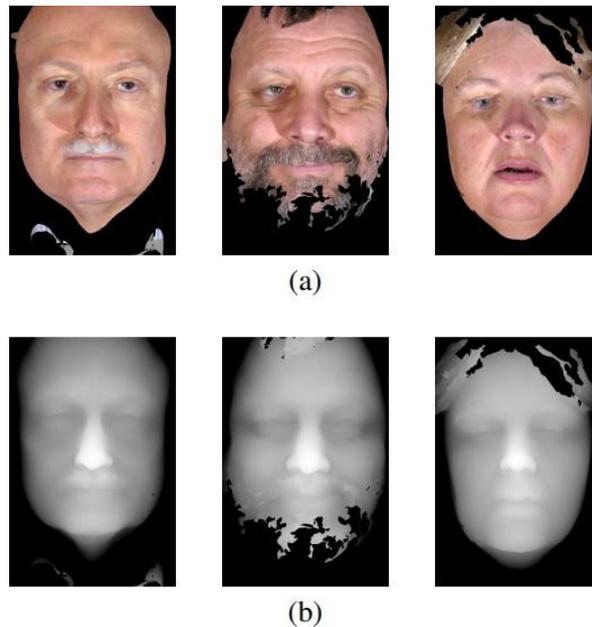


FIGURE 3.5 – Les images du texas 3D face dataset

3.4 Présentation des interfaces

La première interface présentée dans la figure 3.6 est conçue de manière intuitive pour faciliter le test de la technologie de reconnaissance faciale en RGB-D, avec une présentation visuelle claire et des options de navigation permettant de vérifier un visage.

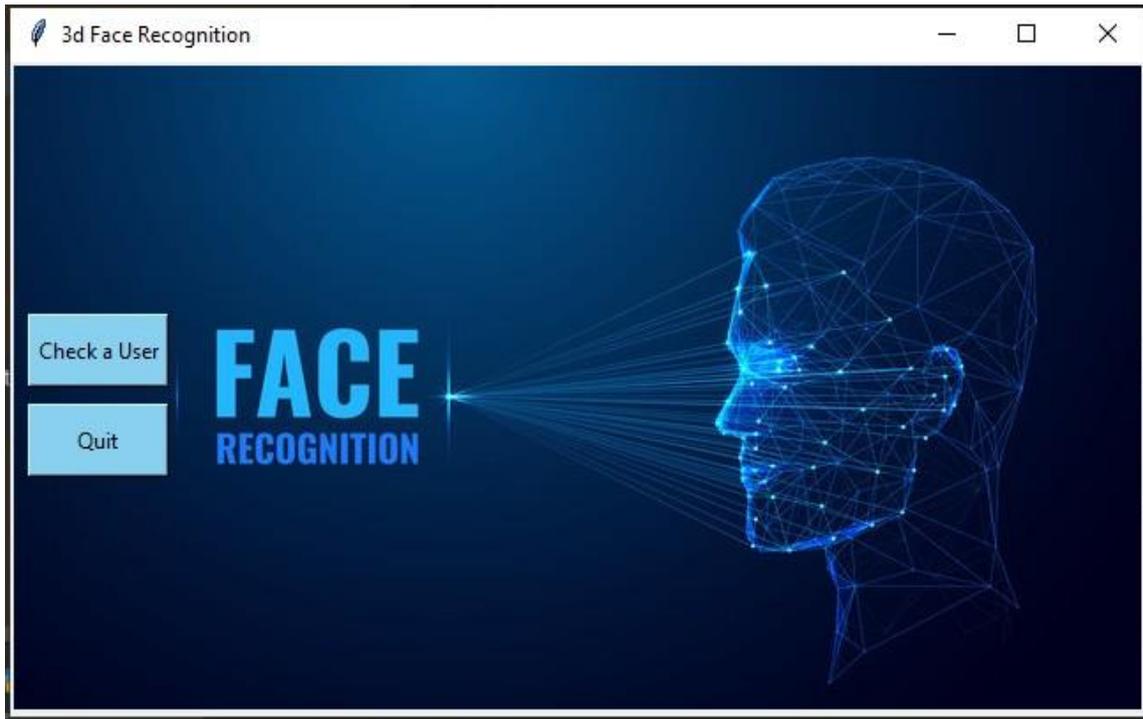


FIGURE 3.6 – Interface du test

La deuxième interface permet de tester la reconnaissance faciale en utilisant des images RGB et des images de profondeur. L'objectif principal est de permettre de sélectionner et d'importer ces deux types d'images, puis de vérifier l'image faciale pour une analyse précise et efficace (voir les figures 3.7).

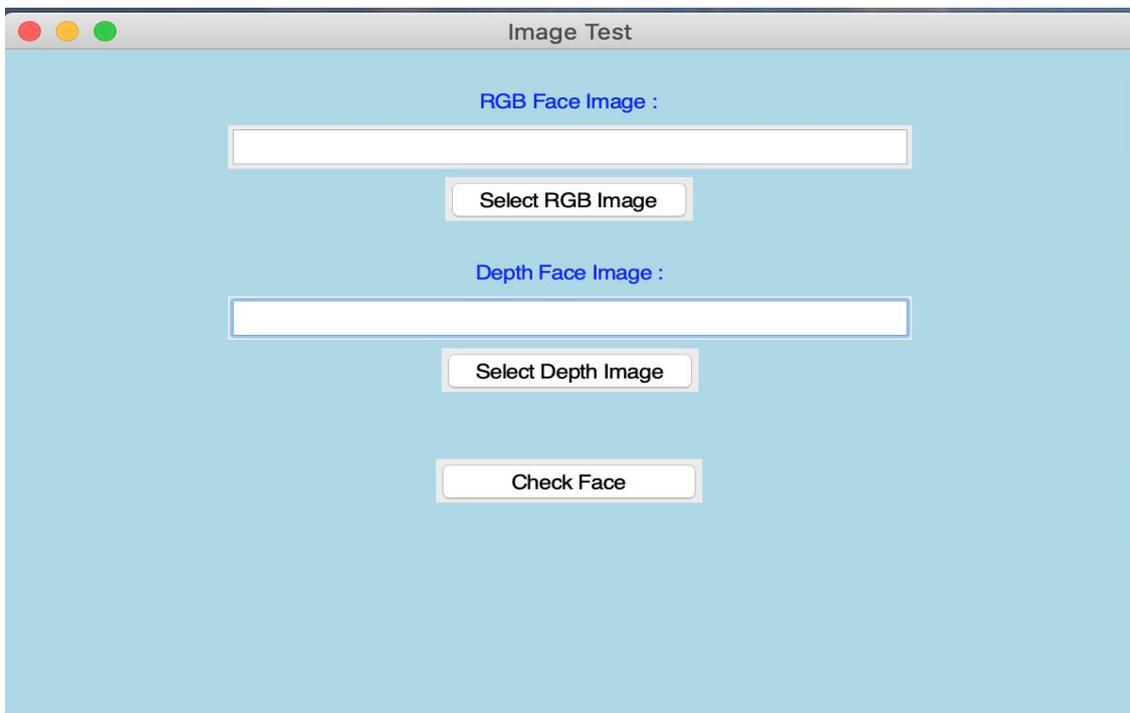


FIGURE 3.7 – Interface d'importation des images

Ces deux interfaces illustrées dans la figure 3.8 et la figure 3.9 permettent de vérifier l'identité de la personne représentée en comparant ces images avec une base de données. Si aucune similarité n'est trouvée, un message d'erreur sera affiché, comme illustré par la fenêtre pop-up indiquant que « Face not accepted » (Visage non accepté), sinon le visage sera accepté.

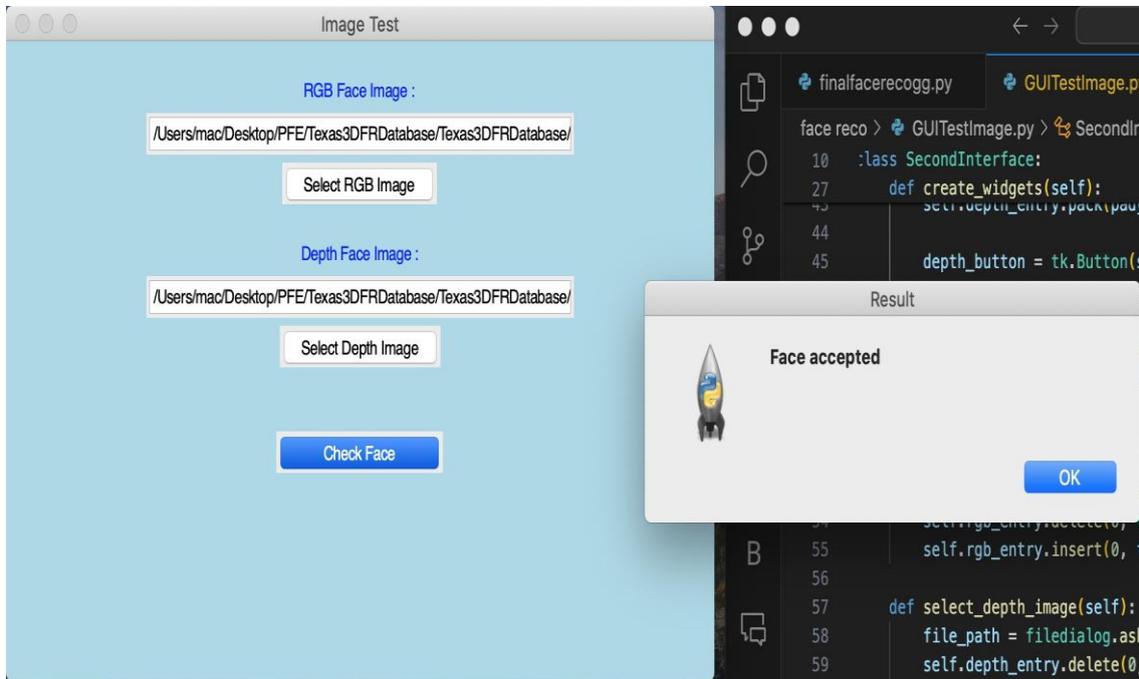


FIGURE 3.8 – Affichage du cas accepté

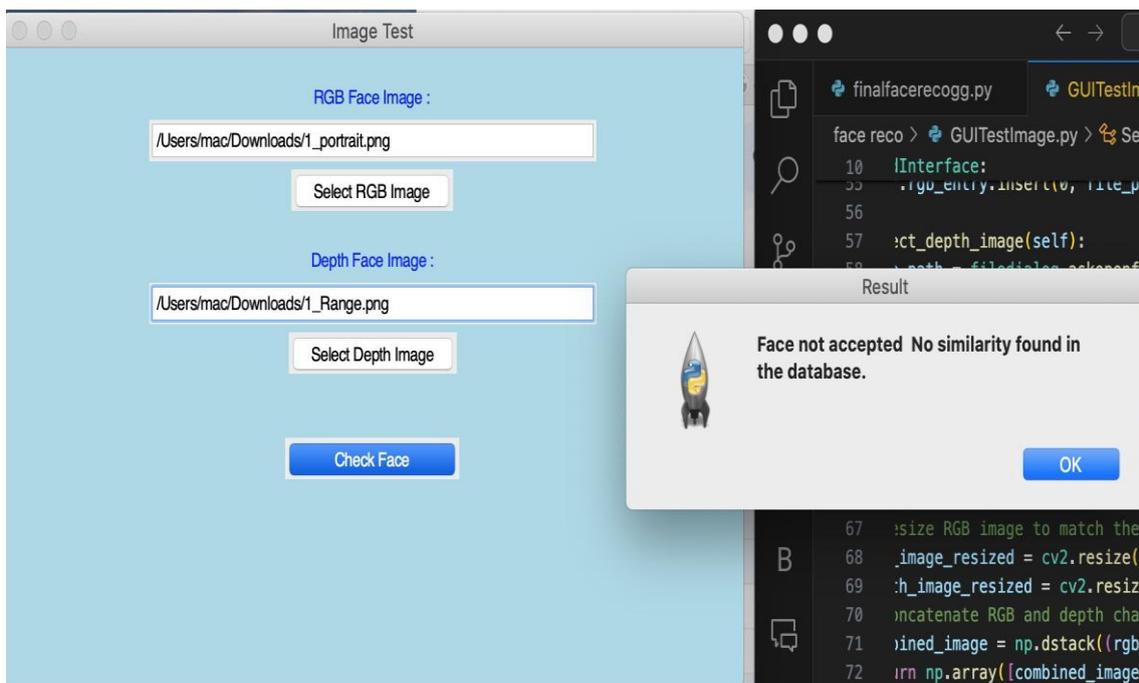


FIGURE 3.9 – Affichage du cas refusé

3.5 Résultats et Tests

Afin de tester la validité de notre modèle et de mettre en œuvre notre approche, il est nécessaire que le modèle passe par la phase d'apprentissage et de test. La phase d'apprentissage et de test de ce modèle commence par la préparation des données. Les images RGB et de profondeur sont traitées pour en extraire les caractéristiques pertinentes. Ces caractéristiques servent ensuite à constituer un ensemble de données, permettant de comparer les individus et de prendre des décisions éclairées.

Les résultats obtenus démontrent que le taux de reconnaissance (accuracy) de notre modèle atteint 77.53% et La précision obtenue est 58.79% sur le jeu de données Texas 3DF.

3.5.1 Configuration des hyperparamètres

Le tableau 3.1 récapitule les hyperparamètres testés de notre modèle proposé pour la reconnaissance faciale en 3D. La configuration de ces hyperparamètres s'est avéré être une tâche délicate que nous avons abordée de manière expérimentale par essais et ajustements successifs des hyperparamètres suivants : nombre d'époques (Epoque), taille du lot (batch size), taux d'apprentissage (learning rate), nombre de qubits (Qubits), et nombre de couches.

Epoque	Taille du lot	taux d'apprentissage	Validation
30	16	0.0004	54.44%
30	16	0.001	60.78%
50	24	0.001	43.64%
100	24	0.001	77.53%

TABLE 3.1 – Évaluation des performances du modèle proposé avec différentes configurations des hyperparamètres

Selon le tableau 3.1, nous constatons qu'en augmentant la taille du lot et le taux d'apprentissage, notre modèle atteint une exactitude supérieure sur l'ensemble de validation.

3.5.2 Métriques d'évaluation

Pour une analyse plus approfondie de la qualité des classes générées par notre modèle, nous proposons les indications suivantes sur les mesures de performance pour les tâches de reconnaissance faciale 3D. La notation utilisée pour l'évaluation est la suivante :

- TP — le nombre de cas où le modèle prédit correctement la classe positive
- FN — le nombre de cas où le modèle prédit à tort la classe négative.
- FP — le nombre de cas où le modèle prédit à tort la classe positive.
- TN — le nombre de cas où le modèle prédit correctement la classe négative.

En analysant ces éléments, nous pouvons évaluer de manière plus précise la performance du modèle de catégorisation. Cela nous permet d'identifier les cas où le modèle prédit correctement ou incorrectement chaque classe. Les matrices de confusion offrent donc une vision détaillée de la qualité de la classification effectuée par le modèle.

Accuracy

Cette mesure évalue l'efficacité globale de l'algorithme en comparaison avec les données de test.

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn}$$

Le taux d'erreur

$$Erreur = \frac{Fn + Fp}{Tp + Tn + Fp + Fn} = 1 - accuracy$$

Précision

Elle évalue la capacité prédictive du modèle en mesurant sa capacité à effectuer des prédictions correctes des classes.

$$Précision = \frac{Tp}{Tp + Fp}$$

Rappel

Cette mesure évalue l'efficacité globale de l'algorithme en comparaison avec les données de test.

$$Rappel = \frac{Tp}{Tp + Fn}$$

3.5.3 Étude comparative et discussion

L'objectif de cette section est de fournir une analyse critique et exhaustive des différentes méthodes et des résultats obtenus. L'étude commence par une comparaison détaillée des algorithmes de reconnaissance faciale en 3D, en examinant leurs avantages, inconvénients et performances respectives en termes de précision, de rapidité et de robustesse. Cette analyse critique permet de proposer des améliorations possibles et des pistes pour des travaux futurs et propose des recommandations pour la recherche dans le domaine de la reconnaissance faciale. En somme, cette section sert à fournir une analyse critique et approfondie des méthodes et des résultats. Elle permet non seulement de comprendre les performances des différentes techniques, mais aussi de situer ces performances dans un cadre pratique et théorique plus large, tout en identifiant des pistes pour l'amélioration future.

3.5.3.1 Par rapport aux méthodes conventionnelles

Dans l'étude mentionnée [66], les auteurs ont mis en œuvre une approche conventionnelle fondée sur les algorithmes Pattern Locale Binaire (LBP) et Transformée en Cosinus Discrète (DCT). Les résultats obtenus sont les suivants :

- lorsque la méthode est appliquée aux images RGB, un taux de reconnaissance de 87% est atteint.
- En revanche, si elle est uniquement appliquée aux images de profondeur, les performances chutent à 60%.
- Néanmoins, la combinaison des modalités RGB et profondeur permet de retrouver un taux de reconnaissance de 87%, équivalent à celui obtenu sur les images RGB seules.

Dans l'étude citée [67], les auteurs ont également appliqué une méthode classique de réduction de dimensionnalité, l'Analyse en Composantes Principales (ACP ou PCA pour Principal Component Analysis), sur le jeu de données Texas 3DF en utilisant les images RGB et de profondeur en entrée.

- L'application de la PCA seule a permis d'obtenir un taux de reconnaissance de 58%.
- Les auteurs ont ensuite combiné différentes approches : l'utilisation de la PCA symbolique a conduit à un taux de 59%.
- Tandis que la combinaison de la méthode de Régression des Transferts (RT pour Regression Transfer) avec la PCA a atteint 77% de reconnaissance.
- Enfin, l'association de la RT, de la PCA symbolique a permis d'obtenir le meilleur taux de reconnaissance de 78%.

Les travaux de l'étude citée [68] proposent une méthode de reconnaissance faciale 2D-3D basée sur un algorithme CCA-PCA amélioré.

- Les résultats expérimentaux montrent que les algorithmes de fusion CCADouble et CCA-PCA sont meilleurs que les méthodes traditionnelles, avec un taux de reconnaissance de plus de 80% sur la base de données TEXAS.
- Les bonnes performances de ces algorithmes indiquent qu'ils sont utiles dans les applications du monde réel qui nécessitent une reconnaissance faciale précise et fiable.
- Comparé aux CCA et KCCA communs, l'algorithme proposé montre une amélioration significative du taux de reconnaissance, ouvrant ainsi une nouvelle perspective pour des applications plus avancées de reconnaissance faciale.
- Ces résultats encouragent une exploration plus approfondie de la possibilité de combiner différentes méthodes pour améliorer la robustesse et la précision des systèmes de reconnaissance faciale, ouvrant ainsi la voie à des applications plus efficaces dans divers domaines.

L'avantage majeur de notre travail par rapport aux travaux antérieurs réside dans l'utilisation de l'information faciale dans son intégralité, et non seulement de parties spécifiques du visage (points de repère). Cette approche basée sur l'apprentissage profond permet d'assurer une reconnaissance robuste face aux variations de pose, d'éclairage et autres conditions environnementales.

3.5.3.2 Par rapport aux méthodes basées sur le deep learning

Nous comparerons Dans cette section les performances des modèles GoogleNet-SNet [69] et CNN-SNet [70] et ResNet-SNet [71] sur le jeu de données Texas 3DF, où les critères d'évaluation sont la précision, les points positifs et négatifs comme présenté dans le tableau 3.2. Nous examinerons également les technologies utilisées pour chaque modèle. Ces modèles, conçus spécifiquement pour la reconnaissance faciale 3D, ont été développés par des chercheurs et doctorants réputés, mais ne sont pas ouverts à une utilisation publique pour le moment.

Facteurs de comparaison	GoogleNet-SNet [69]	CNN-SNet [70]	ResNet-SNet [71]
Accuracy	96.84%	97.98%	97.36%
precision	95.40%	93.82%	91.77%
Les points forts	<ul style="list-style-type: none"> - Architecture plus profonde et plus complexe, capable de capturer des caractéristiques de plus haut niveau. - Utilisation de modules "Inception" permettant une extraction de caractéristiques multi-échelle. - Bonne capacité à gérer les variations d'éclairage et les occultations partielles. 	<ul style="list-style-type: none"> - Par rapport à des architectures plus complexes comme GoogleNet et ResNet, CNN-SNet peut être plus facile à interpréter et à ajuster pour des applications spécifiques. - Utilisation d'un réseau Siamois permettant une comparaison efficace des représentations d'images. 	<ul style="list-style-type: none"> - Proposition d'une amélioration de la loss softmax pour une meilleure reconnaissance des visages et des expressions. - Architecture résiduelle permettant une meilleure propagation du gradient et une convergence plus rapide.
Les points faibles	<ul style="list-style-type: none"> - Complexité accrue entraînant une augmentation des besoins en calcul et en mémoire. -Potentielle surparamétrisation pour des jeux de données de taille modérée comme Texas 3DF. 	<ul style="list-style-type: none"> - Potentiellement moins performant que les architectures plus récentes (GoogleNet, ResNet) pour les tâches de reconnaissance faciale 3D. -Difficulté à gérer les variations d'éclairage et les occultations partielles du visage. 	<ul style="list-style-type: none"> - Complexité accrue par rapport à CNN-SNet, entraînant des besoins en calcul et en mémoire plus importants. -Besoin de données d'entraînement volumineuses .

TABLE 3.2 – Tableau comparatif des méthodes basées sur le deep learning

3.6 Conclusion

Dans ce chapitre, nous avons exploré l'environnement et les outils de travail essentiels pour notre étude. Nous avons examiné en détail les différents matériels utilisés, ainsi que les langages de programmation et les logiciels nécessaires. En outre, nous avons présenté le Texas 3D Face Dataset, en mettant particulièrement l'accent sur le processus d'acquisition des données. Enfin, nous avons discuté des métriques d'évaluation utilisées et des résultats des tests effectués ainsi que de leurs différentes interfaces.

Conclusion et perspectives

Ce projet s'inscrit dans le domaine de la biométrie, plus particulièrement de la reconnaissance faciale 3D. Cette technologie vise à vérifier l'identité d'un individu en exploitant les caractéristiques uniques de son visage capturées par des images ou séquences vidéo RGB-D, combinant les modalités couleur (RGB) et de profondeur (D). Largement déployés pour des raisons de sécurité et de confidentialité, les systèmes de reconnaissance faciale sont de plus en plus présents pour contrôler l'accès à des zones privées. L'objectif de ce travail est de développer, optimiser et valider une architecture de deep learning intégrant les données issues des capteurs de profondeur pour réaliser une reconnaissance faciale 3D et éviter les limites de la reconnaissance faciale 2D dans l'application "Whowiyati".

Ces dernières années, le domaine de la reconnaissance faciale a connu des progrès remarquables, principalement grâce aux avancées de l'apprentissage profond, et plus particulièrement des réseaux de neurones convolutionnels (CNN). Ces avancées encourageantes ne reposent pas sur des améliorations matérielles, des jeux de données plus volumineux ou des modèles plus importants, mais résultent plutôt de nouvelles idées algorithmiques, d'architectures réseau innovantes et d'optimisations méthodologiques. Il est à noter que ces progrès significatifs ont été obtenus sans recourir à de nouvelles sources de données étendues, démontrant ainsi la puissance des innovations algorithmiques et conceptuelles dans le domaine de l'apprentissage profond. Les chercheurs ont su exploiter avec succès le potentiel des CNN existants en les optimisant et les adaptant de manière innovante pour résoudre des problèmes spécifiques de reconnaissance d'images.

Au cours de notre étude, nous avons présenté notre système de reconnaissance faciale 3D, exposé les étapes essentielles du processus ainsi que les principales difficultés rencontrées dans la reconnaissance faciale 2D, comme la diversité des expressions faciales, les variations d'éclairage et les occlusions, qui ont été évitées dans la résolution des systèmes de reconnaissance faciale 3D. Après avoir dressé un état de l'art, constatant les performances supérieures des systèmes basés sur l'apprentissage profond, nous avons présenté les concepts fondamentaux de cette approche, puis introduit en détail les réseaux de neurones convolutionnels. Sur le plan pratique, nous avons développé et implémenté un modèle d'extraction de caractéristiques innovant, basé

sur des images RGB et de profondeur et utilisant des réseaux de neurones convolutifs (CNN), qui démontre une précision remarquable après la phase d'apprentissage et de test. Ce modèle tire parti des recherches et études antérieures pour optimiser son efficacité. En appliquant ce modèle à un ensemble de données, nous avons pu valider le fonctionnement robuste d'un système de reconnaissance faciale RGB-D, ouvrant ainsi la voie à des applications pratiques pour intégrer ce type de techniques. Dans la dernière étape du système, la distance euclidienne est utilisée pour la comparaison des caractéristiques extraites et la prise de décision.

Dans un futur proche, notre système de reconnaissance faciale RGB-D de pointe sera intégré à l'application « Whoiyati ». Cette intégration améliorera considérablement le processus d'identification en ligne en fournissant une couche de sécurité supplémentaire grâce à l'authentification biométrique 3D. Ainsi, les utilisateurs pourront s'authentifier de manière sécurisée et pratique, notamment pour accéder à leurs comptes et effectuer des transactions sensibles.

Nous cherchons constamment à perfectionner notre solution de système RGB-D afin d'offrir une expérience utilisateur optimale. Pour ce faire, nous envisageons d'exploiter les données fournies par Icosnet SPA et d'acquérir un capteur 3D « 3D sensor » supplémentaire. Ces ressources nous permettront de réaliser des tests approfondis et d'affiner notre technologie de reconnaissance faciale, garantissant ainsi une précision et une fiabilité accrues.

L'intégration de cette technologie de reconnaissance faciale RGB-D révolutionnaire positionnera « Whoiyati » comme un leader du marché, offrant une plateforme sécurisée et conviviale. Notre dévouement à l'innovation et à l'amélioration constante de nos solutions biométriques nous permettra de fournir à nos utilisateurs une expérience optimale et de renforcer notre réputation en tant qu'acteur majeur de l'industrie.

Au-delà, les perspectives d'utilisation sont prometteuses dans divers domaines clés : contrôle d'accès et sécurité grâce à une authentification biométrique précise, identification rapide des victimes en cas d'urgence, soins de santé personnalisés grâce à une identification précise des patients, identification des clients, etc. Les applications commerciales permettront de sécuriser les paiements dans les magasins en utilisant son visage pour s'identifier.

Nous résumons notre perspective future dans les points suivants :

- Pour améliorer la précision et l'efficacité du système de reconnaissance faciale 3D, il est nécessaire de mener davantage de recherches et de développer des techniques de deep learning de pointe.
- Extension du jeu de données d'évaluation pour inclure un ensemble d'images plus diversifié et représentatif, destiné à l'entraînement et à l'évaluation. Mise à l'échelle du système développé pour des applications pratiques afin de tester ses performances dans des scénarios réels.
- Étude de l'intégration de systèmes biométriques multimodaux pour améliorer les capacités de sécurité et d'authentification.

- Application de la technologie de reconnaissance faciale 3D dans des domaines émergents tels que la réalité virtuelle et augmentée pour des applications innovantes et des expériences utilisateur améliorées.
- Pour répondre aux préoccupations concernant la vie privée et la sécurité, il est nécessaire de mener davantage d'études sur les questions éthiques et l'application appropriée des technologies biométriques.
- Tester notre solution proposée sur des jeux de données plus volumineux pour meilleur résultat.

Bibliographie

- [1] S. Gupta, M. K. Markey, and A. C. Bovik. Anthropometric 3d face recognition. *International Journal of Computer Vision*, 90(3) :331–349, 2010.
- [2] Woodrow Wilson Bledsoe. The model method in facial recognition. *Panoramic Research Inc., Palo Alto, CA, Rep. PR1*, 15(47) :2, 1966.
- [3] Baback Moghaddam, Tony Jebara, and Alex Pentland. Bayesian face recognition. *Pattern recognition*, 33(11) :1771–1782, 2000.
- [4] Matthew A Turk and Alex P Pentland. Face recognition using eigenfaces. pages 586–587, 1991.
- [5] Laurenz Wiskott, Jean-Marc Fellous, Nobert Krüger, and Christoph Von Der Malsburg. Face recognition by elastic bunch graph matching. pages 355–396, 2022.
- [6] Andrea F Abate, Michele Nappi, Daniel Riccio, and Gabriele Sabatino. 2d and 3d face recognition : A survey. *Pattern recognition letters*, 28(14) :1885–1906, 2007.
- [7] Chenghua Xu, Yunhong Wang, Tieniu Tan, and Long Quan. Depth vs. intensity : Which is more important for face recognition? In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 1, pages 342–345. IEEE, 2004.
- [8] Jean-Yves Cartoux, Jean-Thierry LaPresté, and Marc Richetin. Face authentication or recognition by profile extraction from range images. In *Proceedings. Workshop on Interpretation of 3D Scenes*, pages 194–195. IEEE Computer Society, 1989.
- [9] G Gordon. Face recognition from frontal and profile views. In *Proceedings of the International Workshop on Automatic Face-and Gesture-Recognition, IWAAGR*, volume 95, pages 47–52. Citeseer, 1995.
- [10] Hayit Greenspan, Jacob Goldberger, and Itay Eshet. Mixture model for face-color modeling and segmentation. *Pattern Recognition Letters*, 22(14) :1525–1536, 2001.

- [11] P. J. Phillips, P. J. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone. Face recognition vendor test 2002 : Evaluation report. Technical Report NISTIR 6965, National Institute of Standards and Technology, 2003.
- [12] Elena Carlotta Olivetti, Jacopo Ferretti, Giansalvo Cirrincione, Francesca Nonis, Stefano Tornincasa, and Federica Marcolin. Deep cnn for 3d face recognition. In Caterina Rizzi, Angelo Oreste Andrisano, Francesco Leali, Francesco Gherardini, Fabio Pini, and Alberto Vergnano, editors, *Design Tools and Methods in Industrial Engineering*, pages 665–674, Cham, 2020. Springer International Publishing.
- [13] TechLog360. Artificial intelligence judges your face : This is how facial recognition works, 2023. Accessed : 2024-06-19.
- [14] Berk Gokberk, Albert Salah, Nese Alyuz, and Lale Akarun. *3D Face Recognition : Technology and Applications*, pages 217–246. 01 2009.
- [15] Yaping Jing, Xuequan Lu, and Shang Gao. 3d face recognition : A comprehensive survey in 2022. *Computational Visual Media*, 9(4) :657–685, 2023.
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [17] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. Face recognition : A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1) :98–113, 1997.
- [18] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2 : Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.
- [19] Wenyi Zhao, Rama Chellappa, P Jonathon Phillips, and Azriel Rosenfeld. Face recognition : A literature survey. *ACM computing surveys (CSUR)*, 35(4) :399–458, 2003.
- [20] Sima Soltanpour, Boubakeur Boufama, and QM Jonathan Wu. A survey of local feature methods for 3d face recognition. *Pattern Recognition*, 72 :391–406, 2017.
- [21] Marcella Peter, Jacey-Lynn Minoi, and Irwandi Hipni Mohamad Hipiny. 3d face recognition using kernel-based pca approach. In *Computational Science and Technology : 5th ICCST 2018, Kota Kinabalu, Malaysia, 29-30 August 2018*, pages 77–86. Springer, 2019.
- [22] Parama Bagchi, Debotosh Bhattacharjee, and Mita Nasipuri. Robust 3d face recognition in presence of pose and partial occlusions or missing parts. *arXiv preprint arXiv :1408.3709*, 2014.

- [23] Marwa Jabberi, Ali Wali, Bilel Neji, Taha Beyrouthy, and Adel Alimi. Face shapenets for 3d face recognition. *IEEE Access*, PP :1–1, 01 2023.
- [24] Yaping Jing, Xuequan Lu, and Shang Gao. 3d face recognition : A comprehensive survey in 2022. *Computational Visual Media*, 9(9) :657–685, 2023.
- [25] Donghyun Kim, Matthias Hernandez, Jongmoo Choi, and Gerard Medioni. Deep 3d face identification, 2017.
- [26] Walid Hizem. *Capteur Intelligent pour la Reconnaissance de Visages*. Thèse de doctorat, Institut National des Télécommunications et Université Pierre et Marie Curie, Paris, 2009.
- [27] Looking Glass Factory. Depth map. <https://lookingglassfactory.com/blog/depth-map>, 2024.
- [28] Song Zhou and Sheng Xiao. 3d face recognition : a survey. *Human-centric Computing and Information Sciences*, 8(1) :35, 2018.
- [29] M. Belahcene. *Authentification et Identification en Biométrie*. Thèse de doctorat, Université Mohamed Khider, Biskra, 1 2013. Thèse pour obtenir le titre de Docteur en Sciences en Automatique.
- [30] Zineddine Sellami. Optimisation du cnn par l’algorithme génétique pour la reconnaissance de visage. Master’s thesis, Université de Biskra, 2019.
- [31] OpenClassrooms. Qu’est-ce qu’un réseau de neurones convolutif (ou cnn) ?, 2019.
- [32] Samiha Beggari and Khaoula Khamra. Système de reconnaissance de visage par un réseau de neurone convolutionnel (cnn). Master’s thesis, Université Kasdi Merbah, Ouargla, 2017.
- [33] OpenGenus. Object detection using region-based convolutional neural network (rcnn). *OpenGenus IQ*, 2024.
- [34] Équipe Blent. Réseaux convolutifs (cnn) : comment ça marche ? *Blent.ai*, juin 2022.
- [35] Josh Patterson and Adam Gibson. *Deep Learning : A Practitioner’s Approach*. O’Reilly Media, Inc., 2017.
- [36] Tasnim Ahmed, Mst. Shahnaj Parvin, Mohammad Reduanul Haque, and Mohammad Shorif Uddin. Lung cancer detection using ct image based on 3d convolutional neural network. *Journal of Computer and Communications*, 8(3), 2020.
- [37] Róberta Hlavatá, Robert Hudec, Patrik Kamencay, and Peter Sykora. Human activity classification using the 3dcnn architecture. *Applied Sciences*, 12 :931, 01 2022.

- [38] Hongxiang Fan, Ho-Cheung Ng, Shuanglong Liu, Zhiqiang Que, Xinyu Niu, and Wayne W. C. Luk. Reconfigurable acceleration of 3d-cnns for human action recognition with block floating-point representation. *2018 28th International Conference on Field Programmable Logic and Applications (FPL)*, pages 287–2877, 2018.
- [39] Jiyun Cui, Hu Han, Shiguang Shan, and Xilin Chen. Rgb-d face recognition : A comparative study of representative fusion schemes. In *Biometric Recognition (CCBR 2018)*, pages 358–366. Springer, 2018.
- [40] Changyuan Jiang, Shisong Lin, Wei Chen, Feng Liu, and Linlin Shen. Pointface : Point set based feature learning for 3d face recognition. In *2021 IEEE International Joint Conference on Biometrics, IJCB 2021*. Institute of Electrical and Electronics Engineers Inc., 2021.
- [41] Konstantinos Papadopoulos, Anis Kacem, Abdelrahman Shabayek, and Djamila Aouada. Face-gcn : A graph convolutional network for 3d dynamic face identification/recognition. *arXiv preprint arXiv :2104.09145*, 2021.
- [42] Wenhao Hu. Improving 2d face recognition via fine-level facial depth generation and rgb-d complementary feature learning. *arXiv preprint arXiv :2305.04426*, 2023.
- [43] Anagha R. Bhopale and Surya Prakash. Learning similarity and dissimilarity in 3d faces with triplet network. *Multimedia Tools and Applications*, 80(23) :35973–35991, 2021.
- [44] Whowiyati app, 2024.
- [45] Arindam Chaudhuri. *Deep Learning Models for Face Recognition : A Comparative Analysis*, pages 99–140. Springer International Publishing, Cham, 2020.
- [46] Meng-Tzu Chiu, Hsun-Ying Cheng, Chien-Yi Wang, and Shang-Hong Lai. High-accuracy rgb-d face recognition via segmentation-aware face depth estimation and mask-guided attention network. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–8. IEEE, 2021.
- [47] Visionplatform. Comprendre la reconnaissance d’images : algorithmes, apprentissage automatique et utilisations. <https://visionplatform.ai/fr/comprendre-la-reconnaissance-dimages-algorithmes-apprentissage-automatique> 2024.
- [48] Yuta Inoue, Shun Nishide, and Fuji Ren. Facial expression recognition adaptive to face pose using rgb-d camera. In Hamido Fujita, Moonis Ali, Ali Selamat, Jun Sasaki, and Masaki Kurematsu, editors, *Trends in Applied Knowledge-Based Systems and Data Science*, pages 422–427, Cham, 2016. Springer International Publishing.
- [49] Dorsaf Sebai. Depth in focus : D component of rgb-d images and videos. *Electrical Electronic Engineering*, Oct 2023. Published by Springer Nature.

- [50] R. Heitmeyer. Biometric identification promises fast and secure processing of airline passengers. *ICAO Journal*, 55(9) :10–11, 2000.
- [51] B. Amir. *Identification Biométrique par Descripteurs de Texture Locaux : Application au Visage & Oreille*. PhD thesis, Université 08 Mai 1945 - Guelma, 2014/2015. Doctorat 3ème Cycle, Electronique et Télécommunications.
- [52] Yusuf Sarıgöz. Triplet loss - advanced intro. *Qdrant*, 2022.
- [53] Wikilivres. Programmation python/introduction, 2022. Consulté le 28 décembre 2022.
- [54] TensorFlow Libraries and Extensions.
url :<https://www.tensorflow.org/resources/libraries-extensions>.
- [55] Keras : tout savoir sur l’API de Deep Learning.
url :<https://datascientest.com/keras>.
- [56] Keras : The high-level API for TensorFlow | TensorFlow Core.
url :<https://www.tensorflow.org/guide/keras?hl=fr>.
- [57] OpenCV. Face recognition with opencv, 2024.
- [58] Raphael Kassel. Numpy : la bibliothèque python la plus utilisée en data science, 2021.
- [59] 3d plotting — matplotlib 3.8.4 documentation, 2024. Accessed : 2024-05-15.
- [60] PythonTutorial.net. Python tutorial - tkinter. <https://www.pythontutorial.net/tkinter/>. Accessed : 2024-06-19.
- [61] Pycharm. : <https://www.jetbrains.com/pycharm/>.
- [62] google colab. : <https://colab.google>.
- [63] Deep Learning Turkey. Google colab free gpu tutorial. <https://medium.com/deep-learning-turkey/google-colab-free-gpu-tutorial-e113627b9f5d>. Accessed : 2024-06-19.
- [64] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik. Texas 3d face recognition database. In *IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 97–100, Austin, TX, 2010.
- [65] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik. Texas 3d face recognition database. <http://live.ece.utexas.edu/research/texas3dfr/index.htm>. Accessed : 2024-06-03.

- [66] Sunil Kumar BL and Sharmila Kumari. Rgb-d face recognition using lbp-dct algorithm. *Applied Computer Science*, 17(3) :73–81, 2021.
- [67] PS Hiremath and Manjunath Hiremath. 3d face recognition using radon transform and symbolic lda. *International Journal of Computer Applications*, 975 :8887, 2013.
- [68] Patrik Kamencay, Robert Hudec, Miroslav Benco, and Martina Zachariasova. 2d–3d face recognition method based on a modified cca-pca algorithm. *International Journal of Advanced Robotic Systems*, 11(3) :1–9, 2014. Received 3 October 2013 ; Accepted 17 January 2014.
- [69] K. Qiu, Y. Ai, B. Tian, B. Wang, and D. Cao. Siamese-resnet : implementing loop closure detection based on siamese network. In *2018 IEEE intelligent vehicles symposium (IV)*, pages 716–721. IEEE, 2018.
- [70] W. Cui, W. Zhan, Y. Jingjing, C. Sun, and Y. Zhang. Face recognition via convolutional neural networks and siamese neural networks. In *2019 international conference on intelligent computing, automation and systems (ICICAS)*, pages 746–750. IEEE, 2019.
- [71] J. Zhou, X. Jia, L. Shen, Z. Wen, and Z. Ming. Improved softmax loss for deep learning-based face and expression recognition. *Cognitive Comput Syst*, 1(4) :97–102, 2019.