République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي و البحث العلمي

جامعة عبد الرحمان ميرة بجاية
Université A. Mira de Béjaia
Faculté des Sciences Exactes Département d'Informatique

Mémoire de fin de cycle

En vue d'obtention du diplôme de Master en Informatique

Option : Systèmes d'Information Avancées (SIA)

Détection de visages en ligne basée sur le modèle Deep Learning VGG16

Réalisé par
Mme RAID LYNDA
Mme TAZIBET MANEL

Soutenu le 30 Juin 2024, Devant le jury composé de :

M. Hachem SLIMANI	Université de Bejaia	Président
Mme Sara MECHIOURI	Université de Bejaia	Examinateur
M. Achour ACHROUFENE	Université de Bejaia	Encadrant
M. Abdenour ZAIDRI	Cevital	Invité

Promotion : 2023/2024

Dédicace

“

Tout d'abord, je remercie Dieu pour m'avoir accordé la santé et l'intelligence, et d'avoir exaucé un vœu que je formulais dans chacune de mes prières depuis mon entrée à l'école. À ma chère maman, qui m'a inculqué l'amour de la lecture et de l'étude, ainsi que la valeur de la persévérance et de la sagesse, je te suis infiniment reconnaissante. À mon père, dont le soutien indéfectible et les encouragements constants ont été essentiels à ma réussite, merci de toujours croire en moi. À ma petite sœur, amie et pilier de ma vie, merci pour ton soutien inestimable et ta capacité à me faire sourire même dans les moments les plus difficiles. À ma grande sœur, source d'inspiration et guide précieuse depuis l'école primaire jusqu'à l'université, merci pour ta confiance et ton aide inconditionnelle. À mon grand frère, mentor visionnaire, merci pour tes conseils avisés et ton soutien constant qui ont éclairé mon chemin vers l'avenir. À mes cousines, que je considère comme des sœurs, merci pour votre confiance et votre soutien qui m'ont permis de croire en moi-même. Enfin, À mes amis et proches, dont l'amitié et l'amour ont été des piliers essentiels de ma réussite, merci pour votre présence et votre encouragement. À vous tous, ce mémoire est devenu une réalité, et je suis fière de la personne que je suis aujourd'hui. Merci.

”

- Lynda

“

Louange à Dieu, le Tout-Puissant, pour m'avoir accordé la force, la patience et la sagesse nécessaires pour mener à bien ce travail. À ma chère mère, je te dédie ce travail avec tout mon amour et ma gratitude. Ta patience, ton soutien inébranlable et tes sacrifices ont été les pierres angulaires de ma réussite.

À mon père, Votre présence indéfectible, vos efforts inlassables et vos sacrifices incommensurables ont fait de moi la personne que je suis aujourd'hui. Votre foi indéfectible en mes capacités et votre soutien constant m'ont toujours guidé vers de nouveaux sommets.

À mon cher grand frère, tu as été bien plus qu'un frère pour moi : un soutien constant, une présence rassurante et un modèle inspirant. À chaque étape de ma vie, tu as été à mes côtés, me guidant avec sagesse et bienveillance. Ta générosité d'esprit et ta force intérieure ont été pour moi des sources inépuisables d'encouragement et d'inspiration.
À ma famille, pour votre soutien indéfectible, vos encouragements constants, et votre foi inébranlable en moi.
Votre aide précieuse et vos conseils avisés ont été des piliers dans cette aventure académique.
Pour l'ame de Khalti leila (ASSOUL Leila née BELHARETH), partie trop tôt, vous avez laissé derrière vous un vide immense et une peine incommensurable. Vous étiez bien plus qu'une tante pour moi ; vous étiez une confidente, un soutien indéfectible, une présence réconfortante dans les moments de doute et de tristesse. Votre amour et votre bienveillance m'ont toujours accompagné, et votre souvenir restera gravé dans mon cœur à jamais.
Je dédie ce mémoire à vous tous, ce mémoire est devenu une réalité, et je peux me tenir fièrement devant vous avec la personne que je suis devenue.
Merci.

”

- Manel

Remerciements

Nous exprimons tout d'abord notre gratitude envers Allah, le Tout Miséricordieux, qui nous a accordé la patience et le courage nécessaires pour mener à bien ce mémoire.

Nous tenons à remercier particulièrement notre encadrant, Monsieur Achour Achroufene, pour son dévouement sans faille. Ses conseils avisés et ses critiques constructives nous ont guidés et incités à nous surpasser à chaque étape de ce projet.

Nos remerciements s'adressent ensuite aux membres de notre jury, dont la lecture attentive et l'évaluation sincère et constructive ont enrichi ce modeste travail.

Nous témoignons également notre profonde reconnaissance à nos professeurs d'université, auprès de qui nous avons acquis des connaissances inestimables et développé des compétences essentielles.

Enfin, nous remercions toutes les personnes ayant joué un rôle, aussi grand ou petit soit-il, dans la réalisation de cet écrit.

Résumé

La vérification et l'identification des individus dans des systèmes physiques ou numériques sont essentielles dans de nombreuses applications, telles que l'accès aux comptes bancaires, la reconnaissance des personnes via des caméras de surveillance, ou encore la sécurisation des locaux privés ou professionnels et diminuer le risque d'usurpation d'identité d'autrui. La biométrie offre un moyen d'identifier les individus en leur attribuant une forme d'identification unique et universelle. C'est dans cette perspective que s'inscrit ce travail, nous avons développé un système d'authentification biométrique de reconnaissance faciale se basant sur le modèle de classification d'images VGG16 dont on a ajusté la dernière couche de classification, nous avons entraîné notre modèle sur un jeux de données assez volumineux que nous avons construit nous même . Ce modèle a été employé pour détecter les visages du système. Les résultats des tests expérimentaux sur le jeux de données collecté ont prouvé que l'utilisation du visage comme modalité à l'aide du modèle VGG16 améliore les performances d'authentification comparativement à la version originale du modèle.

Mot clés : Authentification ; Biométrie ; Visage ; Reconnaissance faciale ; Modèle VGG16; Augmentation de données ; Détection de visage ; CNN

Abstract

Verification and identification of individuals in physical or digital systems are essential for many applications, such as accessing bank accounts, recognizing individuals via surveillance cameras, or securing private and professional premises to reduce the risk of identity theft. Biometrics provides a means to identify individuals by assigning them a unique and universal form of identification. This work fits into this perspective, we developed a facial recognition biometric authentication system based on the VGG16 image classification model, which has been fine-tuned by adjusting the last classification layer and trained on our own large datasetwhich we contructed. This model was employed to detect faces within the system. Experimental test results on collected datasets demonstrated that using facial recognition with the VGG16 model improves authentication performance compared to the original version of the model.

Keywords : Authentication ; Biometrics ; Face ; Facial Recognition ; VGG16 Model ; Data Augmentation ; Face Detection ; CNN

Table des matières

Table des figures

Liste des tableaux

Introduction générale

L'authentification et l'identification des individus jouent un rôle crucial dans la sécurisation des accès physiques et numériques, ainsi que dans la reconnaissance personnalisée [1]. Elles sont essentielles pour restreindre l'accès à des comptes bancaires, assurer la surveillance via caméras de sécurité, et protéger les installations privées des entreprises [2]. Traditionnellement, ces besoins sont comblés par des méthodes comme les mots de passe, les clés, les cartes magnétiques ou les codes PIN, mais ces solutions sont vulnérables au vol, à la perte, à l'oubli et à d'autres risques de sécurité.

La biométrie repose sur le principe fondamental de la singularité des individus à travers une gamme de caractéristiques physiques telles que l'iris, les empreintes digitales, et les traits du visage [3]. Cette approche constitue une méthode robuste pour l'authentification et l'identification, en attribuant à chaque personne une identification distincte et universellement reconnue.

Dans la pratique, les systèmes biométriques peuvent être restrictifs en termes de performances et difficiles à développer à cause des dispositifs à hautes performances. Le visage est largement reconnu pour sa commodité et sa sécurité accrue en tant que trait biométrique. Ainsi que la disponibilité des données et leur collecte avec des dispositifs disponibles et ressources matérielles et computationelles accessibles à tous.

Cependant, les systèmes de reconnaissance faciale ont des limitations, telles que les variations entre les échantillons d'un même individu et les changements que peuvent subir ces caractéristiques au fil du temps ainsi que la difficulté à détecter les visages dans des conditions environnementales complexes. C'est dans cette direction que s'insère notre travail qui consiste à relever ces défis pour améliorer davantage la fiabilité et l'efficacité des systèmes de reconnaissance faciale. Dans ce cadre nous proposons un système biométrique se basant sur la détection de visages entraîné sur un jeux de données volumineux que nous avons nous même construit avec l'accord des participants .

Ce document est structuré en quatre chapitres :

- Le premier chapitre introduit les généralités et définit les concepts clés de la biométrie et des systèmes d'authentification biométriques.
- Le deuxième chapitre présente un état de l'art des recherches récentes sur les systèmes de reconnaissance faciale ainsi que les tendances et défis du domaine de l'authentification biométriquie.
- Le troisième chapitre introduit la problématique, ensuite justifie les choix de la solution proposée, avant de détailler les différentes phases du système de reconnaissance

faciale proposé.

- Enfin, le quatrième chapitre présente le processus suivi pour l'implémentation de l'approche proposée et expose les résultats obtenus à partir des expérimentations et des tests réalisés, offrant une analyse approfondie des performances du système développé.

Enfin, le mémoire se termine par une conclusion générale qui résume le travail effectué, les résultats obtenus, et quelques perspectives.

Chapitre 1

Généralités sur les Systèmes d'authentification

1.1 Introduction

La biométrie est une science qui utilise les caractéristiques biologiques et comportementales des individus pour les identifier de manière unique et précise. Dans un monde de plus en plus digitalisé, la sécurité et l'authentification sont devenues des préoccupations majeures. La biométrie offre une solution robuste pour répondre à ces défis grâce à ses diverses applications dans les domaines de la sécurité informatique, de l'accès aux bâtiments, des transactions financières, et bien plus encore.

Ce chapitre vise à fournir une compréhension complète des concepts fondamentaux de la biométrie et de son fonctionnement. Nous commencerons par définir ce qu'est la biométrie et pourquoi elle est importante. Ensuite, nous explorerons la structure et le fonctionnement des systèmes biométriques, en détaillant les différentes modalités biométriques utilisées. Nous examinerons ensuite les spécificités de chaque modalité, avant de plonger dans l'architecture générale des systèmes biométriques. Par la suite, nous catégoriserons ces systèmes et discuterons des niveaux de fusion dans les systèmes multimodaux. Enfin, nous analyserons les mesures de performance essentielles pour évaluer ces systèmes et présenterons les modèles les plus utilisés dans le domaine biométrique. Ce chapitre se terminera par une conclusion qui résumera les points clés abordés et introduira les perspectives du chapitre suivant.

1.1.1 Définition de la biométrie

Elle concerne l'identification et l'authentification uniques d'un individu grâce à une analyse de ses caractéristiques physiques ou biologiques. Ces caractéristiques peuvent inclure les empreintes digitales, la reconnaissance faciale, la reconnaissance de l'iris, la voix, et même les comportements tels que la démarche.

La biométrie est un domaine innovant qui a pris de l'ampleur rapidement. Elle est largement utilisée dans les domaines de la sécurité, de l'accès aux systèmes informatiques, et de l'authentification des utilisateurs pour divers services et applications [3].

1.1.2 Définition d'un système biométrique

Un système biométrique est une technologie utilisée pour identifier les personnes en utilisant des caractéristiques physiques uniques, comme les empreintes digitales, le visage, ou l'iris des yeux. Ce système collecte ces données biométriques et les compare avec celles déjà enregistrées dans une base de données. Si les données correspondent, la personne est identifiée ou authentifiée ; sinon, l'accès est refusé. Un système biométrique peut fonctionner en mode inscription(enrôlement), vérification ou identification selon l'environnement de l'application [4]

Maintenant que nous avons une idée claire de ce qu'est un système biométrique, examinons les différentes modalités biométriques qui peuvent être utilisées au sein de ces systèmes.

1.2 Modalités biométriques

Les modalités biométriques permettant de vérifier l'identité d'un individu peuvent être classées en trois catégories [5] : physiques, biologiques ou comportementales. Ces dernières sont des caractéristiques uniques pour chaque personne et donc très fiables et sécurisées et difficilement réplicables.

1.2.1 Modalités physiques

Les modalités physiques sont des traits biométriques qui sont uniques à chaque individu permettant ainsi de le différencier des autres. Parmi ces traits nous pouvons citer l'empreinte digitale, le visage et l'iris.

Empreinte digitale

L'empreinte digitale est unique à chaque individu, les chances qu'une autre personne ait les mêmes empreintes digitales qu'autre sont de 1 sur 64 trillions [6]. La reconnaissance par empreinte digitale s'est rapidement étendue, passant de son utilisation dans la criminologie à une utilisation dans différents domaines tels que les banques, les entreprises et la sécurisation des processus de déverrouillage [7], comme le montre la figure 1.1 [8] . Ce processus doit sa précision et son efficacité aux détails de l'empreinte digitale, expliqués dans la figure 1.2 [9]. Une minutie est un point situé sur le changement de continuité des lignes papillaires.

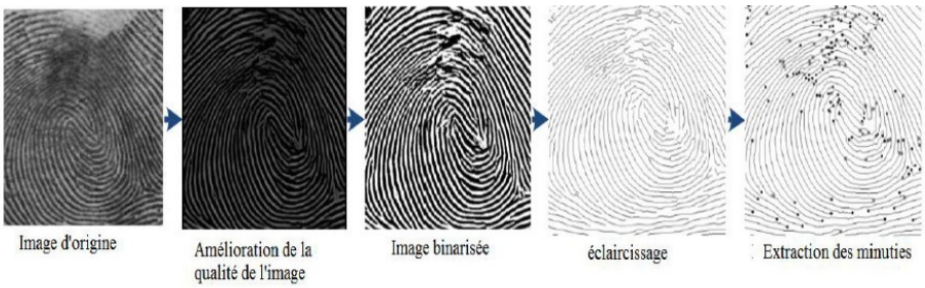


FIG. 1.1 : Processus de la reconnaisance d'empreinte

FIG. 1.2 : Les différents minutiers de m'empreinte

Parmi les défis liés à cette modalité, il est essentiel d'assurer une lecture précise et cohérente des empreintes digitales pour éviter les erreurs d'authentification.

Les variations dans la qualité de l'image, les conditions environnementales et les caractéristiques individuelles de l'empreinte digitale peuvent affecter la précision des systèmes biométriques. Un autre défi est que certains individus peuvent avoir des empreintes digitales difficiles à lire en raison de facteurs tels que l'âge, les blessures, ou les conditions médicales.

Visage

Dans ce système, la reconnaissance faciale est réalisée en construisant un modèle visuel en deux ou trois dimensions à partir de multiples images capturées de la même personne à l'aide d'une caméra ou d'une vidéo. Ce modèle facial est ensuite analysé en tenant compte de caractéristiques distinctives telles que l'orientation du nez, la forme des lèvres, la taille des yeux, etc. Par exemple, le modèle illustré dans la figure 1.3 [8] divise le visage en plusieurs points, avec des mesures spécifiques correspondant à la position de ces caractéristiques. En étudiant ces mesures, le système peut créer un modèle facial unique pour chaque individu et le vérifier avec le fichier référence.

Fig. 1.3 : Reconnaissance faciale

Iris

La biométrie par l'iris est une des technologies (avec la rétine) qui assure un haut niveau de sécurité. L'iris procure une unicité très élevée (1 sur 10 puissance 72) et sa stabilité est étendue jusqu'à la mort des individus, d'où une fiabilité extraordinaire [10].

L'iris se réfère à la structure colorée et annulaire de l'œil située entre la cornée et le cristallin. L'iris est unique pour chaque individu en termes de motifs, de textures et de couleurs. Ces caractéristiques sont modélisées dans la figure 1.4 [11] .

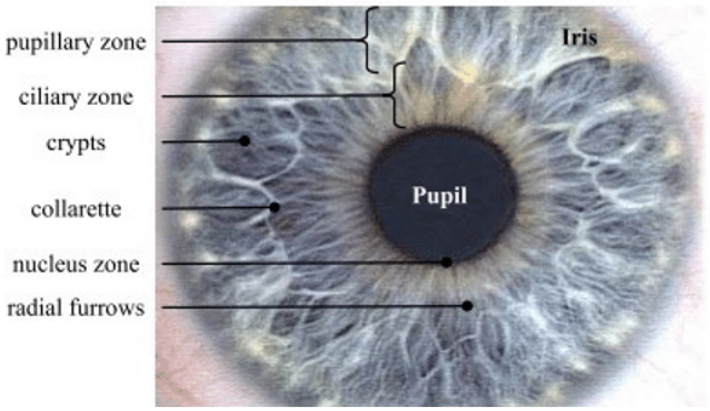


Fig. 1.4 : Dispositif capturant l'image de la rétine

Les dispositifs destinés à capturer les images de la rétine sont souvent onéreux, nécessitant des équipements spécialisés pour projeter un faisceau lumineux à faible intensité sur l'œil, généralement dans les fréquences visibles ou infrarouges [12]. De plus, l'utilisateur doit rapprocher son œil très près du dispositif de lecture et maintenir son regard fixé sur

un point spécifique pendant plusieurs secondes, ce qui exige une coopération considérable de sa part. Toutefois, en raison de la délicatesse de l'œil et de la réticence naturelle des individus à approcher cet organe près d'un appareil, cette méthode est souvent mal perçue par le grand public.

1.2.2 Modalitées comportemanetales

Les modalités comportementales sont des caractéristiques biométriques basesé sur des comportements uniques à chaque individu. Ces comportements sont l'exécution de certaines activités telles que la façon de marcher, la vitesse et la distance entre les pas. Dans cette sous-section nous allons détailler les modalités comportementales les plus utilisées.

Dynamique de la signature

Les systèmes biométriques utilisant cette modalité se basent généralement sur le fait que l'utilisateur signe avec un stylo électronique sur une palette graphique et au même temps elle examine l'ensemble de la dynamique (voire figure 1.5 [13]) comme la vitesse, la direction, et la pression de l'écriture. On distingue deux façons de capturer une signature, soit avec des capteurs qui s'assimilent à de simples scanners, soit par l'usage d'une tablette graphique et d'un stylet sensible. La figure 1.6 [13] explique ce processus de capture de la signature biométrique .

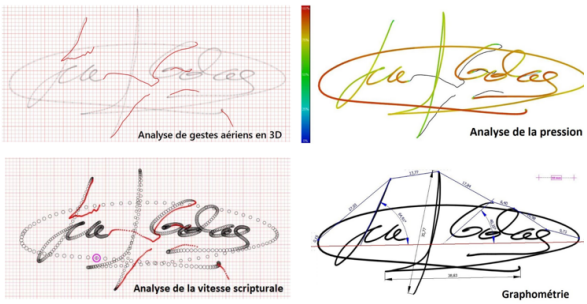


FIG. 1.5 : Exemples des caractéristiques soulevées dans une signature

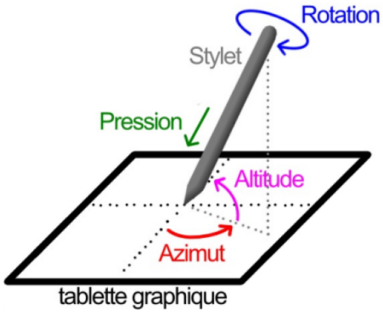


FIG. 1.6 : Tablette graphiques capturant la signature biométrique

Les difficultés liées à la capture d'une signature viennent du fait qu'une personne ne signe jamais deux fois de la même façon, même à quelques secondes d'intervalle. En effet suivant les émotions ou la fatigue, une signature peut fortement évoluer. D'où la mise au point d'algorithmes très complexes capables de prendre en compte ces évolutions possibles.

Démarche

Plusieurs solutions biométriques s'intéressent à l'utilisation de la démarche comme modalité pour l'identifications des individus, notamment quand il est sujet de reconnaître les personnes par caméra de surveillance. En analysant la vitesse moyenne, l'inclination du dos et les déformations des jambes et bras au niveau des articulations. La démarche serait en effet étroitement associée à la musculature naturelle, donc, elle est très personnelle. La figure 1.7 [14] illustre un exemple de l'analyse de la démarche.

FIG. 1.7 : Exemple d'analyse de la démarche

Elle peut, aussi, détecter les comportements suspects (par vidéo-surveillance), donc utilisée pour le contrôle d'accès aux bâtiments ou aux zones réglementées mais elle est facilement modifiable par l'individu.

Dynamique de frappe au clavier

Cette modalité vise l'identification et authentification des individus en analysant leur façon de taper. Etant ainsi un processus de mesure et d'évaluation du rythme de frappe sur des appareils numériques, notamment sur : claviers d'ordinateur, téléphones mobiles et écrans tactiles.

Le mode de fonctionnement d'un système biométrique repose sur l'utilisation d'un logiciel qui recueille les données collectées automatiquement par le système d'exploitation. Ce logiciel prend en compte les temps de pression, de vol et de relâchement des touches qui sont propres à chaque individu lorsqu'il tape sur son clavier . L'étape de vérification dans ce système consistera alors à capturer le comportement courant de l'utilisateur et de le comparer au modèle de son comportement usuel. Grâce à un certain seuil, le système peut classifier le comportement comme étant normale, ou bien anormale, selon les données collectées. La reconnaissance par la frappe au clavier est illustré par la figure 1.8 [8].

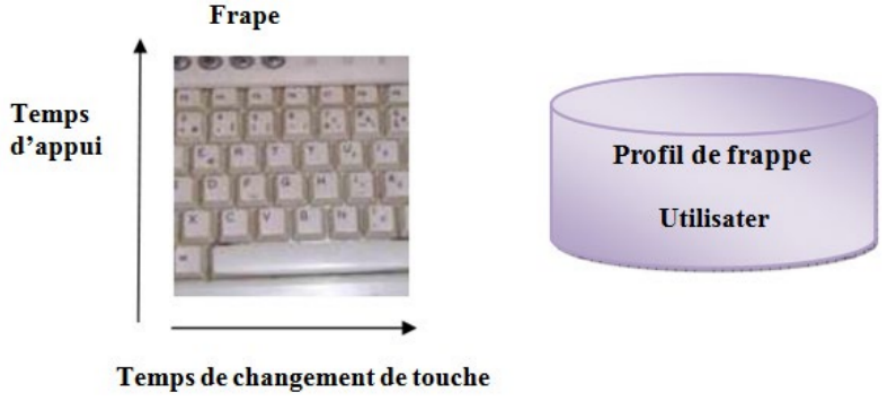


FIG. 1.8 : Reconnaissance par la frappe au clavier

Chaque modalité biométrique possède des caractéristiques spécifiques qui influencent leur choix et leur application. Explorons en détail ces spécificités.

1.2.3 Spécificités des modalités biométriques

Pour que les caractéristiques, propre à chaque individu, puissent être qualifiées de modalités biométriques, elles doivent correspondre aux spécificités suivantes [15] :

- Universalité : Le trait biométrique existe chez tous les individus.
- Unicité : Le trait biométrique doit être suffisamment différent par rapport aux autres individus dans une population donnée.
- Stabilité : Le trait biométrique d'un individu doit être suffisamment stable et invariant au cours du temps. Un trait qui change n'est pas préférable dans l'authentification biométrique.
- Mesurabilité : Le trait biométrique doit être facilement mesurable et quantifiable à l'aide d'un capteur approprié qui ne cause aucun désagrément à l'individu.
- Performance : Signifie que l'authentification doit être précise et rapide
- Acceptabilité : Indique que le trait biométrique utilisé doit être bien accepté par les utilisateurs du système.
- Non-reproductibilité : Concerne la facilité ou non à falsifier un trait biométrique.

Le tableau 1.1 [8] présente une comparaison des différentes caractéristiques des traits biométriques couramment utilisés pour l'authentification en ligne. Cette comparaison vise à identifier les options les plus adaptées selon divers critères résume les qualités des traits biometriques une à une [16] :

Modalité	Universalité	Unicité	Performance	Stabilité	Acceptabilité	Mesurabilité	Circonvention
Visage	Élevée	Élevée	Moyenne	Élevée	Élevée	Élevée	Faible
Empreinte digitale	Moyenne	Élevée	Moyenne	Moyenne	Moyenne	Élevée	Élevée
Iris	Moyenne	Élevée	Moyenne	Faible	Faible	Élevée	Faible
Empreinte palmaire	Moyenne	Moyenne	Moyenne	Moyenne	Moyenne	Élevée	Moyenne
Signature	Faible	Élevée	Moyenne	Moyenne	Moyenne	Élevée	Élevée
Voix	Élevée	Moyenne	Moyenne	Élevée	Élevée	Élevée	Élevée
Démarche	Élevée	Moyenne	Moyenne	Élevée	Élevée	Moyenne	Faible
Frappe de clavier	Moyenne	Moyenne	Moyenne	Moyenne	Élevée	Moyenne	Moyenne

Tab. 1.1 : Comparaison entre les traits biométriques

D'après la comparaison des traits biométriques nous pouvons constater que chaque méthode a ses avantages et inconvénients. Par exemple, les empreintes digitales et l'iris sont très précises, mais la collecte des données de l'iris peut être intrusive. La reconnaissance faciale, bien que moins précise, est plus acceptable et facile à intégrer en ligne grâce aux caméras disponibles. Elle est également plus abordable. Ainsi, la reconnaissance faciale est souvent le meilleur compromis pour l'authentification en ligne.

1.3 Architecture d'un système biométrique

L'architecture d'un système biométrique comprend plusieurs composants clés qui fonctionnent ensemble pour assurer une reconnaissance précise et fiable des individus. L'architecture d'un système biométrique est illustrée sur la figure 1.9 [8].

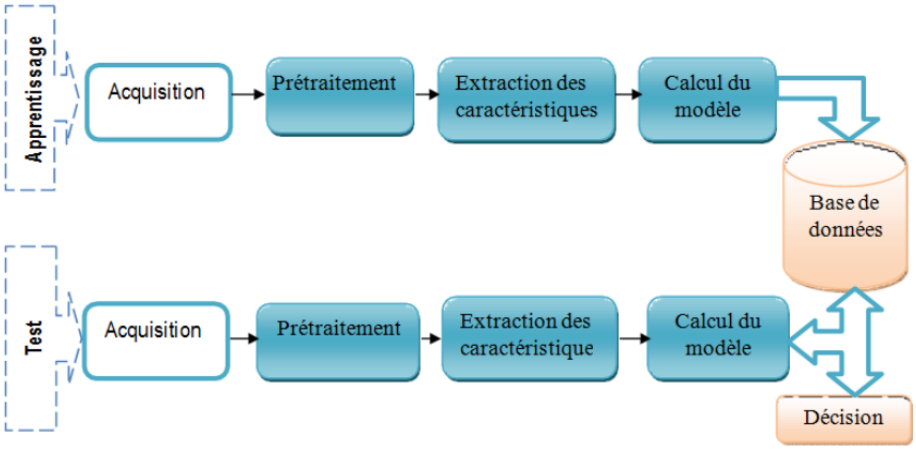


FIG. 1.9 : Architecture d'une système biométrique

1.3.1 Principaux modules d'un système biométrique

Un système biométrique peux être essentiellement composé de six modules [17] :

Module de capture : Assure l'acquisition des traits caractéristiques de l'intrus et en extrait un représentation de données biométriques brutes. Cette représentation est ensuite utilisée pour l'enrôlement, la vérification ou l'identification. Il s'agit d'un capteur biométrique qui peut être de type sans ou avec contact.

Module de prétraitement : Les données brutes acquises sont prétraitées pour améliorer leurs qualité et facilité l'extraction des données discriminatoires pertinentes (caractéristiques).

Module d'extraction de caractéristiques : Il détermine à partir de la donnée prétraitée la nouvelle représentation des données dite modèle ou Template. Cette nouvelle représentation doit être pertinente et idéalement unique pour chaque personne.

Module du stockage : contient les modèles biométriques des utilisateurs enrôlés dans le système. Le système de stockage peut être un simple fichier dans une carte à puce, ou bien une grande base de données gérée par un système de gestion de base de données

Module de de similarité : Compare les données biométriques extraites par le module d'extraction de caractéristiques à un ou plusieurs modèles préalablement enregistrés. Ce module éxtermine ainsi le degré de similarité ou de divergence entre deux vecteurs biométriques.

Module de décision : Une décision est prise en fonction de l'indice de similarité retourné, si celui-ci est suffisant pour déterminer l'identité d'un individu. Généralement, l'indice peut varier de 0 à 1, l'indice 0 représentant une différence totale et un indice de 1 indiquant une correspondance parfaite.

La diversité des modalités biométriques conduit à une nécessité de comprendre l'architecture d'un système biométrique. La prochaine section se concentre sur les éléments constitutifs de ce système et leur rôle dans le processus d'authentification.

1.3.2 Modes de fonctionnement d'un système biométrique

Les systèmes biométriques peuvent offrir trois modes de fonctionnement : le mode d'enrôlement, le mode de vérification et le mode d'identification.

Mode d'enrôlement

La phase initiale est crucial de tout système biométrique. Il s'agit de l'enregistrement de l'utilisateur dans la base de données du système. Ce processus est décrit en détail dans la figure 1.10 [17]. A l'aide d'un capteur les caractéristiques biométriques sont mesurées pour en extraire une représentation numérique. Cette dernière est par la suite réduite par un algorithme d'extraction dans le but de faciliter le stockage, la vérification et l'identification.

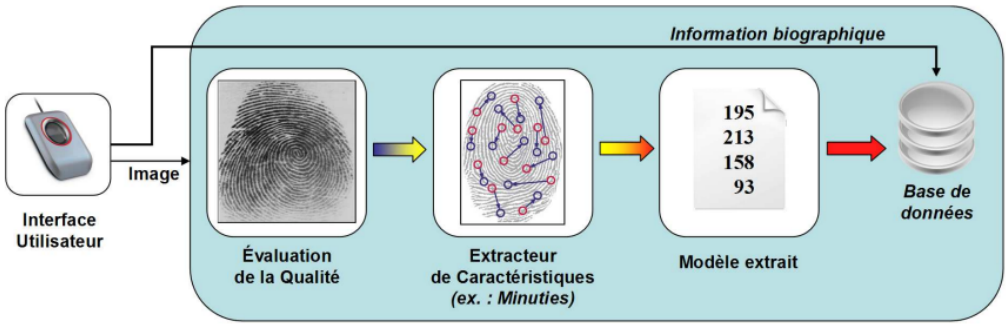


FIG. 1.10 : Enrôlement d'un individu dans un système biométrique

Mode d'identification

C'est une comparaison "1 : N", dans lequel le système reconnaît un individu en l'appariant avec un des modèles de la base de données. La personne peut ne pas être dans la base de données. Ce mode consiste à associer une identité à une personne. Voici une représentation visuelle du mode d'identification dans la figure 1.11 [17].

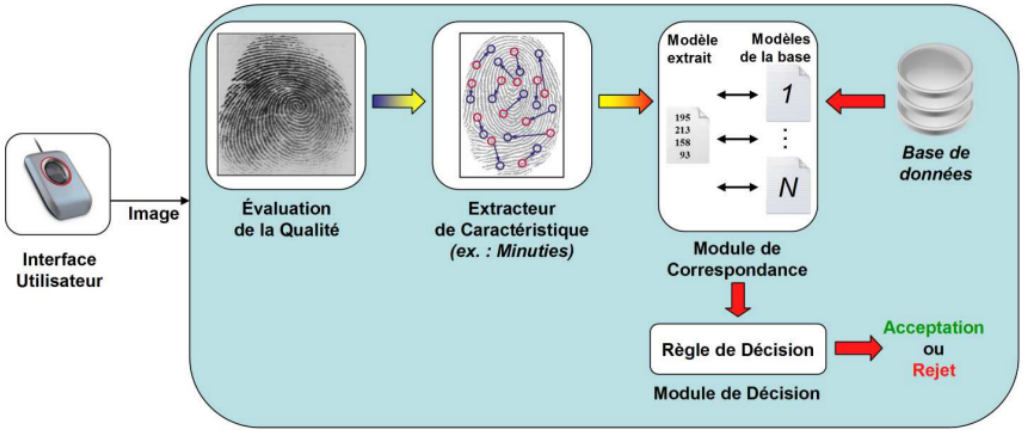


FIG. 1.11 : Identification d'un individu dans un système biométrique

Mode d'authentification

Contrôler que l'utilisateur est bien la personne qu'il prétend être. Le système illustré dans la figure 1.12 [17] va réaliser une comparaison entre les caractéristiques biométriques fraîchement acquises avec le modèle biométrique préalablement stocké dans la base de

données. Ce mode de test s'appel un test 1 : 1 ou le système renvoie une réponse binaire (oui ou non).

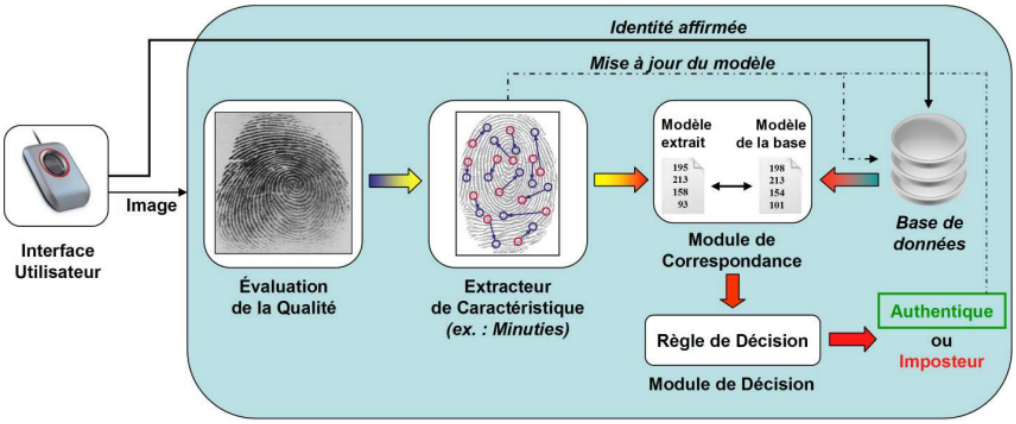


FIG. 1.12 : Authentification d'un individu dans un système biométrique

Dans la pratique, les systèmes biométriques peuvent utiliser une ou plusieurs modalités biométriques. Ils sont classer en conséquence dans différentes catégories en fonction de leur complexité et de leur utilisation.

1.4 Catégories des systèmes biométriques

Un système biométrique peut-être mono-modal ou multimodale. La fusion des modalité se fait à différent niveau du système.

1.4.1 Systèmes biométriques unimodeaux

Les systèmes biométriques ne se basant que sur une seule source d'information pour le processus d'authentification ou d'identification sont appelés "systèmes biométriques unimodaux", et ils se spécialisent dans le traitement d'un seul trait biométrique, comportant l'acquisition des caractéristiques, leur pré-traitement, l'approche utilisée et en se basant sur ces étapes une décision finale sur l'identité de l'individu examinée est générée. soit physiologique ou comportemental.

1.4.2 Systèmes biométriques multimodaux

Les systèmes biométriques multimodaux sont des solutions biométriques à modalité multiples provenant de plusieurs sources, échantillons, capteur ou algorithme [4]. Les systèmes biométriques multimodaux combinent plusieurs modalités biométriques pour améliorer la performance. Le choix d'une stratégie de fusion efficace est primordial pour assurer la combinaison des différents traits biométriques acquis de différents sources [4].

Fusion au niveau des capteurs

Les données brutes sont fusionnées au niveau des capteurs directement avant l'extraction des caractéristiques [18].

Fusion au niveau des décisions

La prise de décision dans cette approche de fusion est faite en prenant en entrée plusieurs autres décisions émises individuellement par chaque sous-système. Une décision est

généralement sous forme de OUI, ou NON, qu'on peut représenter par 1 ou 0. Pour arriver à la décision finale, il existe plusieurs manières telles que les règles « OU et ET », le vote par majorité, etc.

Fusion au niveau des caractéristiques

Ce niveau de fusion consiste à combiner les vecteurs des caractéristiques des traits biométriques en un seul vecteur commun. Contrairement à la fusion au niveau capteur, cette fusion ne nécessite pas une homogénéité entre les données. Le vecteur des caractéristiques résultant de cette fusion est calculé soit comme une somme pondérée si les vecteurs de caractéristiques en entrée sont homogènes, ou comme une concaténation si les vecteurs de caractéristiques à fusionner sont hétérogènes.

Fusion au niveau des rangs

Ce niveau de fusion est utilisé dans le cas des systèmes biométriques qui à leur sortie des matchers associent un rang a chaque identité inscrite. La correspondance est plus forte si le rang est élevé. Une fois ces rangs attribués, ils sont combinés en un nouveau rang associé à une identité spécifique, ce qui conduit à la prise de décision finale.

Fusion au niveau des scores

Ce niveau de fusion est le plus répondu et utilisé dans le marché de la biométrique étant donné la richesse des informations et leur implémentation facile. Cette approche fait références à des techniques qui combinent les scores de correspondance générés par différentes comparaisons.

Pour évaluer l'efficacité des systèmes biométriques, il est crucial de comprendre et d'utiliser les mesures de performance appropriées.

1.5 Mesures de performance d'un système biométrique

Il existe deux types d'utilisateurs qui peuvent interagir avec un système biométrique : les clients, qui sont autorisés à l'utiliser, et les fraudeurs, qui tentent de s'y connecter sans autorisation. Pour cette raison, il est essentiel d'évaluer les performances d'un système biométrique, car de nombreux critères sont en jeu [19]. Voici quelques-unes des définitions de ces critères [20].

- Taux de faux rejet (FRR) : ce taux représente le pourcentage de personnes faussement rejetées par le système :

$$\text{FRR} = \frac{\text{nombre de faux rejets}}{\text{nombre total de requêtes clients}} \tag{1.1}$$

- Taux de fausses acceptations (FAR) : ce taux représente le pourcentage de personnes acceptées alors qu'elles ne sont pas sensées être reconnues :

$$\text{FRR} = \frac{\text{nombre de fausses acceptations}}{\text{nombre total de requêtes imposteurs}} \tag{1.2}$$

- Équivalent du taux d'erreurs (FAR) : Ce taux est calculé à partir des deux premiers critères utilisés et constitue un critère de performance courant. Il correspond

au point FRR = FAR, c'est-à-dire qu'il représente le meilleur compromis entre la précision d'identification et le risque de faux positifs (voir figure 1.13 [21]).

$$\text{ERR} = \frac{\text{nombre de fausses acceptations + nombre de f aux rejets}}{\text{nombre total d accès}} \tag{1.3}$$

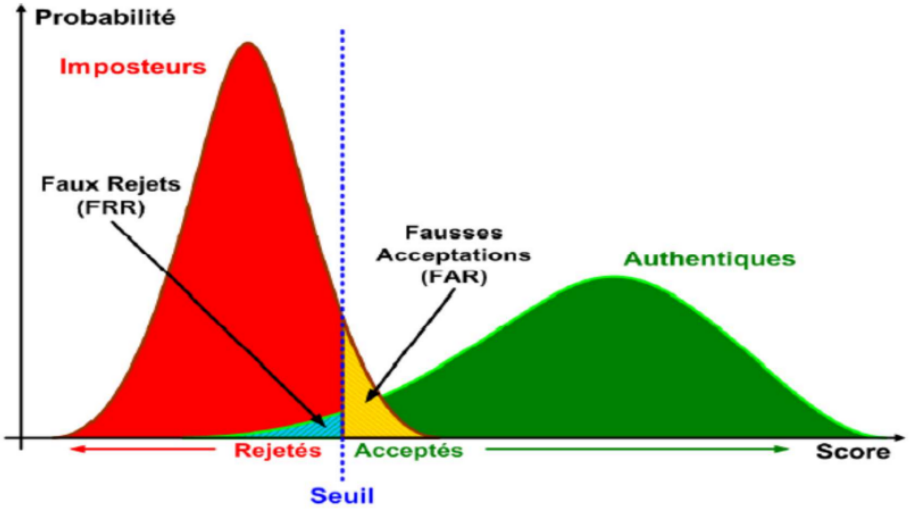


Fig. 1.13 : Représentation du FRR, FAR et ERR

Pour les applications nécessitant une grande sécurité, le choix du seuil tend vers un taux de fausses acceptations (positifs) faible, c'est-à-dire limiter au maximum les possibilités d'intrusions. Dans le cas des applications qui ne nécessitent pas de sécurité élevée, mais plutôt de commodité, c'est-à-dire le besoin de confort et utilisation aisée du système, dans ce cas, on cherche à minimiser le rejet par erreurs des personnes, ce qui veut dire que le choix du seuil tend vers un taux de faux rejet (négatifs) faible [22]. La relation entre le FRR et le FAR, ainsi que le choix du seuil sont illustrés par la figure 1.14 [4].

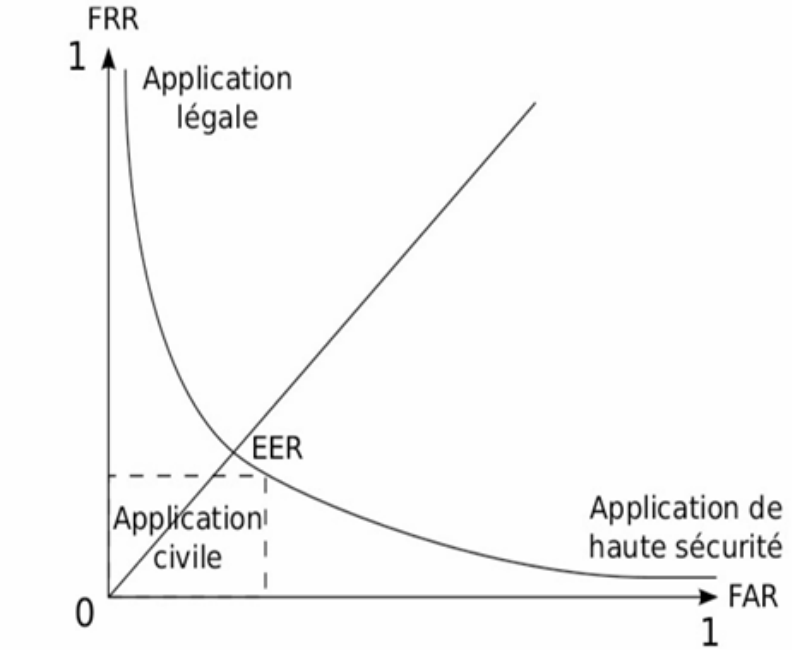


Fig. 1.14 : La courbe ROC

La courbe "ROC"(Receiver Operating Characteristic) est celle qui permet de montrer la

relation entre les critères cités précédemment. Le taux d'égale erreur peut être facilement iden- tifiable puisqu'il s'agit de l'intersection de cette courbe avec la droite d'équation y = x [68]. Il est très couramment utilisé dans le mode de fonctionnement "authentification". Pour le mode "identification", c'est la courbe "CMC" (Cumulative Match Characteristic) qui est souvent utilisée. Cette courbe donne le pourcentage de personnes reconnues en fonction du rang, plus le choix d'image de reconnaissance est proche plus la valeur du rang est petite donc le système est plus fiable [17].

Un autre critère de performance souvent utilisé est la précision [23]. Il représente la fraction des bonnes prédictions. Deux formules de calcul existent pour ce critère :

- La précision

$$\text{Précision} = \frac{\text{nombre de prédictions correctes}}{\text{nombre total de prédictions}} \tag{1.4}$$

- Pour la classification binaire, la précision peut également être calculée en termes de positifs et négatifs :

$$\text{Précision} = \frac{TP + TN}{TP + TN + FN + FP} \tag{1.5}$$

où :
— TP : nombre de vraies acceptations.
— TN : nombre de vrais rejets.
— FP : nombre de fausses acceptations.
— FN : nombre de faux rejets.

Enfin, pour obtenir de bonnes performances, les systèmes biométriques utilisent divers modèles et algorithmes. Dans la section suivante, nous explorons quelques exemples de réseaux de neurones profonds couramment utilisés dans la reconnaissance biométrique, ce qui aidera à comprendre le reste de notre travail.

1.6 Modèles deep learning utilisés dans les systèmes biométriques

Nous discuterons quelques modèles d'apprentissage profond, leur architecture, leur fonctionnement et leur application à des tâches spécifiques telles que la vérification d'identité et la détection d'anomalies. Ces exemples illustrent l'impact significatif de l'apprentissage profond dans le domaine de la biométrie et ouvrent la voie à de nouvelles avancées dans ce domaine en constante évolution.

Les réseaux de neurones profonds, ou Deep Neural Networks (DNN), représentent une avancée significative dans le domaine de l'apprentissage automatique. Ces réseaux sont constitués de plusieurs couches de neurones artificiels, ce qui leur permet de modéliser des relations complexes et d'apprendre des représentations de données à différents niveaux d'abstraction. Ils fonctionnent en utilisant des processus similaires à ceux des neurones biologiques, où chaque neurone artificiel reçoit des entrées, applique des poids et des biais, puis passe le résultat à travers une fonction d'activation pour produire une sortie.[24].

1.6.1 Réseau de neurones convolutif (CNN)

Les réseaux de neurones convolutifs (CNN) sont une classe de réseaux neuronaux profonds particulièrement adaptés à l'analyse d'images. Ils sont composés de couches de neurones organisées en couches convolutives, de pooling et de couches entièrement connectées. Les couches convolutives filtrent l'image d'entrée pour extraire des caractéristiques pertinentes, tandis que les couches de pooling réduisent la dimensionnalité de la sortie. Les CNN sont largement utilisés dans la reconnaissance d'images et la vision par ordinateur en raison de leur capacité à apprendre des caractéristiques hiérarchiques à différentes échelles. Ces caractéristiques sont ensuite utilisées pour la classification, la segmentation ou la détection d'objets dans les images. L'architecture d'un réseau de neurones convolutif est illustré par la figure 1.15 [25].

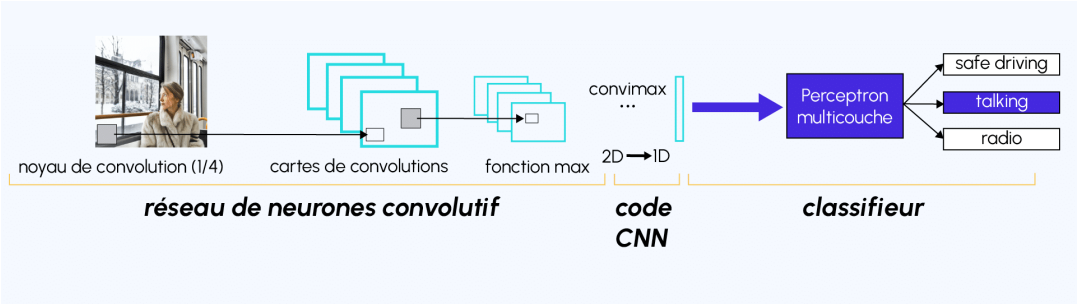


FIG. 1.15 : Architecture d'un réseau de neurones convolutif.

Voici quelques autres exemples de réseaux CNN qui ont été largement utilisés dans le domaine de la reconnaissance d'images et la vision par ordinateur :

LeNet-5

Développé par Yann LeCun et ses collègues en 1998, LeNet-5 est l'un des premiers CNNs utilisés pour la reconnaissance de chiffres manuscrits dans les codes postaux américains. Il a été largement utilisé dans le domaine de la reconnaissance optique de caractères (OCR). L'architecture LeNet-5 est illustrée par la figure 1.16 [26].

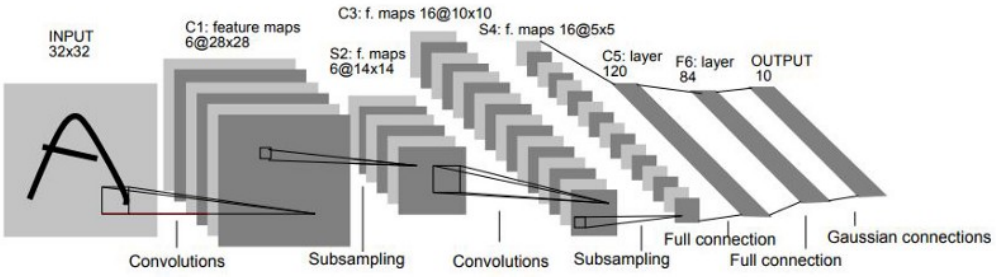


FIG. 1.16 : Architecture LeNet-5.

L'architecture LeNet-5 est organisée en deux ensembles distincts de couches de convolution, chacun suivi d'une couche de pooling moyenné. Ensuite, une couche de convolution spéciale, qui aplatit les données, est utilisée pour préparer l'entrée des deux couches entièrement connectées qui suivent. Enfin, un classifieur softmax est utilisé pour produire les prédictions finales.

AlexNet

Développé par Alex Krizhevsky, Ilya Sutskever et Geoffrey Hinton en 2012, AlexNet a été le premier CNN à remporter le concours ImageNet Large Scale Visual Recognition

Challenge (ILSVRC) en 2012. Il a contribué à populariser l'utilisation des CNN pour la classification d'images.L'architecture AlexNet est illustrée par la figure 1.17 [27].

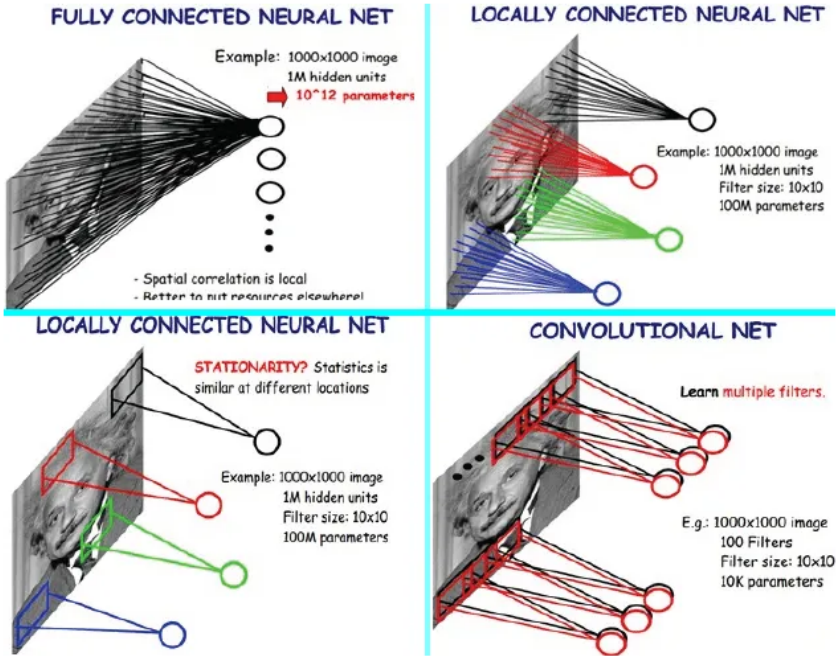


Fig. 1.17 : CNN AlexNet vu d'un point de vue de réseau neuronal.

C'était la première architecture à utiliser un GPU pour accélérer les performances d'entraînement. AlexNet se compose de 5 couches de convolution, 3 couches de max-pooling, 2 couches normalisées, 2 couches entièrement connectées et 1 couche SoftMax. Chaque couche de convolution se compose d'un filtre de convolution et d'une fonction d'activation non linéaire appelée "ReLU". Les couches de pooling sont utilisées pour effectuer la fonction de max-pooling et la taille de l'entrée est fixée en raison de la présence de couches entièrement connectées. La taille de l'entrée est mentionnée à la plupart des endroits comme étant 224x224x3, mais en raison de certains remplissages qui se produisent, elle s'avère être de 227x227x3. En outre, AlexNet compte plus de 60 millions de paramètres.

VGGNet

Développé par le Visual Geometry Group (VGG) à l'Université d'Oxford en 2014, VGGNet se distingue par sa profondeur. Il utilise principalement des couches de convolution 3x3 empilées les unes sur les autres, suivies de couches de pooling et de quelques couches entièrement connectées pour la classification. L'architecture VGG16 est illustré par la figure 1.19 1.19 et 1.18 1.18 [28].

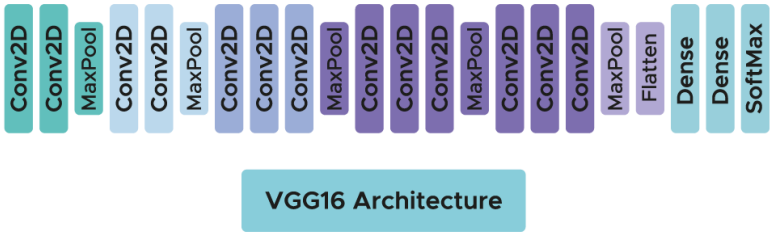


FIG. 1.18 : Description des couches du VGG16.

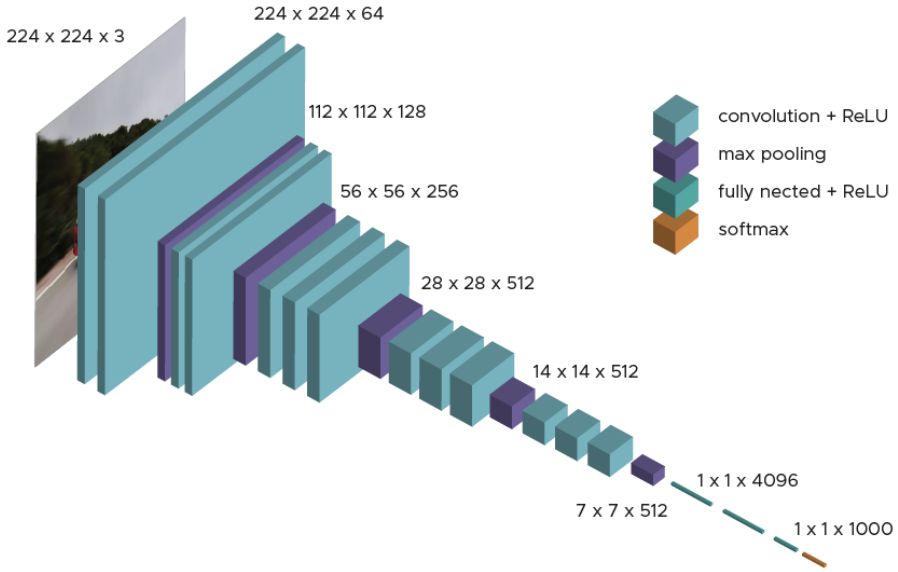


FIG. 1.19 : Architecture VGG16.

Le modèle nécessite un prétraitement consistant à soustraire la valeur moyenne RGB de chaque pixel. Pendant l'apprentissage, l'entrée de la première couche de convolution est une image RGB de taille 224 x 224. Les noyaux de convolution sont de taille 3×3, permettant de capturer différentes notions dans l'image. Le modèle utilise des couches de Max-Pooling de taille 2×2 pour réduire la taille des filtres au cours de l'apprentissage. À la sortie des couches de convolution et de pooling, il y a trois couches de neurones entièrement connectées. Les deux premières ont 4096 neurones, et la dernière a 1000 neurones avec une fonction d'activation softmax pour déterminer la classe de l'image.

GoogLeNet (Inception)

Développé par Google Research en 2014, GoogLeNet a introduit l'architecture Inception, caractérisée par l'utilisation de modules d'inception qui combinent différentes tailles de noyaux de convolution et de pooling pour améliorer l'efficacité computationnelle et la performance du réseau. L'architecture GoogleNet est illustré par la figure 1.20 [29].

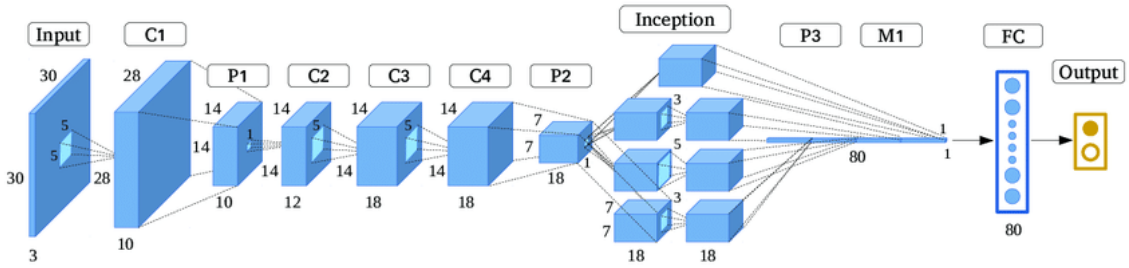


FIG. 1.20 : Architecture GoogleNet.

ResNet

Développé par Microsoft Research en 2015, ResNet (Residual Network) a introduit des connexions résiduelles qui permettent de former des réseaux encore plus profonds tout en évitant le problème de la disparition du gradient. Cela a permis de construire des réseaux de neurones avec des centaines de couches. L'architecture ResNet est illustré par la figure 1.21 [29]

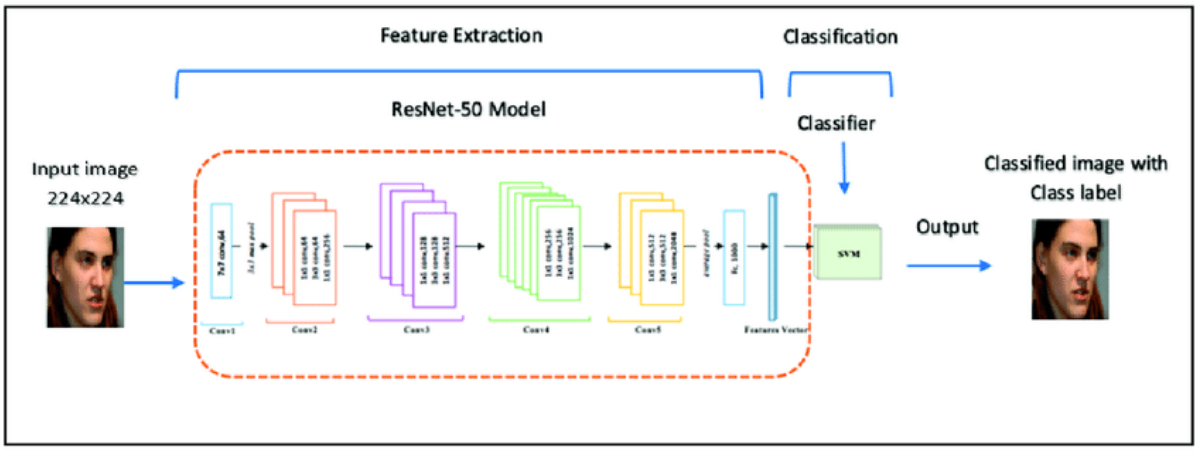


FIG. 1.21 : Architecture ResNet.

1.6.2 Réseau de neurones siamois

Un réseau de neurones siamois est une architecture spéciale de réseau neuronal utilisée pour la comparaison de paires d'entrées. Il est souvent utilisé dans des tâches de vérification ou de comparaison, telles que la vérification d'identité ou la comparaison de similitude entre deux éléments. Les réseaux de neurones siamois utilisent une architecture avec deux branches identiques, partageant les mêmes poids, et les sorties des deux branches sont comparées pour évaluer la similarité entre les entrées.

1.6.3 Réseau de neurones récurrent (RNN)

Un réseau de neurones récurrent (RNN) est un type de réseau neuronal qui traite des séquences de données en tenant compte des dépendances temporelles entre les éléments de la séquence. Contrairement aux réseaux de neurones classiques, les RNN ont des connexions récurrentes qui leur permettent de conserver une mémoire à court terme. Ils sont souvent utilisés dans des tâches telles que la génération de texte, la traduction automatique et la reconnaissance vocale.

1.6.4 Réseau de neurones adversarial (GAN)

Un réseau de neurones antagonistes génératifs (GAN) est une architecture de réseau neuronal composée de deux réseaux : un générateur et un discriminateur. Le générateur crée de nouvelles données réalistes, tandis que le discriminateur essaie de distinguer entre les données réelles et générées. Les deux réseaux s'entraînent de manière antagoniste, où le générateur cherche à tromper le discriminateur en produisant des données indiscernables des données réelles. Les GAN sont souvent utilisés dans la génération d'images réalistes et la synthèse de données.

1.7 Conclusion

Ce chapitre nous a plongés au cœur de la biométrie et de ses multiples facettes. Il nous a permis de comprendre les fondements de la biométrie et des systèmes biométriques. Nous avons défini la biométrie et exploré les diverses modalités biométriques, telles que les empreintes digitales, la reconnaissance faciale et la reconnaissance de l'iris, en discutant des avantages et des inconvénients de chacune. Nous avons également examiné les spécificités de ces modalités et la façon dont elles influencent le choix des systèmes biométriques.

Nous avons ensuite analysé l'architecture des systèmes biométriques, détaillant les différentes couches et configurations possibles. Les systèmes biométriques ont été classifiés en différentes catégories, et nous avons discuté des niveaux de fusion dans les systèmes multimodaux pour améliorer leur performance. Enfin, nous avons abordé les mesures de performance critiques pour évaluer ces systèmes, avant de présenter quelques modèles de réseaux de neurones profonds couramment utilisés dans les systèmes biométriques.

Dans le prochain chapitre, nous explorerons l'état de l'art des modèles de détection et de reconnaissance faciale, en mettant en lumière les avancées récentes dans ce domaine. Nous analyserons en détail les modèles VGG16, FaceNet et VGG50, ainsi que les articles de recherche qui ont contribué à l'évolution de ces technologies. En examinant les performances et les applications de ces modèles, nous continuerons à approfondir notre compréhension des systèmes biométriques et de leur potentiel dans divers domaines d'application.

Chapitre 2

Etat de l'art de la recherche sur l'authentification biométrique

2.1 Introduction

Dans ce deuxième chapitre, nous plongerons dans le domaine de la détection et de la reconnaissance faciale, en explorant les modèles de l'état de l'art qui ont façonné ce domaine en pleine expansion. La détection et la reconnaissance faciale ont acquis une importance croissante dans de nombreux secteurs, de la sécurité à la gestion des identités, en passant par les applications mobiles et les systèmes de surveillance.

Notre objectif ici est d'effectuer une revue critique des articles de recherche les plus pertinents, mettant en lumière les avancées récentes, les tendances émergeâtes et les défis persistants dans ce domaine en constante évolution. Nous nous concentrerons sur les modèles clés tels que VGG16, FaceNet et ResNet, en explorant leurs architectures, performances et applications dans divers contextes, tout en incluant également les travaux supplémentaires sur l'augmentation de données et la multimodalité ainsi que les défis et tendances actuels du domaine de la reconnaissance facial.

2.2 Études connexes

Afin de se situer dans la littérature autour des systèmes biométriques spécialisé en reconnaissance faciale, cette section a pour but de mettre en évidence les points les plus importants du domaine en question, tout en faisant référence de façon générale, à quelques travaux de recherche lus à ce sujet.

Selon l'architecture adoptée, les systèmes biométriques de reconnaissance faciale font usage de divers algorithmes et modèles, dont le modèle VGG16 [30] [31] [32], FaceNet [33] [34], MTCNN [35] [36]. Actuellement d'autres techniques sont utilisées afin d'améliorer les performances d'un système de reconnaissance facial telle que l'augmentation des données [37] [34].

De plus, des jeux de données contenant un grand nombre d'images de visages sont utilisés : Labeled Faces in the Wild (LFW) [38], CASIA-FASD [31], et ce, afin d'évaluer les systèmes en question et les comparer aux autres. Les chercheurs s'appuient aussi pour cela, sur plusieurs mesures de performances, dont la précision, le taux d'erreur, des courbes de performances telles que DET et ROC, le taux de fausses acceptations et rejet et recall, etc. Ces mesures de performances sont souvent citées de façon détaillée dans de nombreux documents [17].

Le but d'évaluation de chaque système proposé, est bien entendu d'en étudier l'efficacité vis-à-vis des autres systèmes, le but étant de réussir à mettre en œuvre un système surpassant les autres dans le cadre des mesures de performances déjà citées. Divers documents réalisés, se spécialisent dans l'évaluation des solutions proposées dans ce domaine, dont les comparatifs qui implémentent et testent de multiples approches réalisées [39].

2.3 Synthèse des documents

Comme nous avons pue le voir à travers la section précedente les systèmes biométriques de détection faciale , et les travaux réalisés là-dessus dans la littérature, sont très variés et nombreux. De ce fait„ notre synthèse des documents offre un état de l'art détaillé sur les articles et travaux récents en matière de détection faciale, mettant en lumière les avancées technologiques et les défis persistants dans ce domaine en constante évolution.

2.3.1 Travaux VGG16

Le travail de Vimal et al. [30]

Cet article explore la création d'un système de détection de visages et de masques en temps réel utilisant l'architecture VGG-16, un réseau de neurones convolutifs (CNN) performant dans les tâches de vision par ordinateur. La motivation principale réside dans la nécessité accrue de systèmes de détection faciale fiables, surtout avec la pandémie de COVID-19 qui a rendu le port du masque obligatoire dans de nombreux lieux publics.

Le système proposé repose sur un modèle pré-entraîné VGG-16, connu pour sa capacité à extraire des caractéristiques complexes à partir d'images, ce qui est crucial pour la détection précise des visages avec ou sans masque. Le modèle VGG-16 est modifié en remplaçant la couche entièrement connectée supérieure par des couchespar des couches d'aplatissement (*flatten*), dense, et dense softmax pour une meilleure classification. L'apprentissage par transfert est utilisé pour tirer parti des poids pré-entraînés sur Imagenet, facilitant ainsi une adaptation rapide aux nouvelles données.

Les images de visages avec et sans masque ont été utilisées, totalisant 10 000 images, avec un partage de 80 % pour l'entraînement et 20% pour la validation. Le processus d'entraînement comprenait des techniques d'augmentation des images pour améliorer la robustesse du modèle face aux variations de position, d'éclairage et de taille des visages.

Les résultats montrent que le modèle peut détecter avec précision 94% sur les ensembles de données utilisés. Les visages masqués et non masqués, même dans des positions non frontales. Les tests ont démontré des performances prometteuses, suggérant que le modèle peut être utilisé efficacement dans des applications de surveillance, de sécurité et de personnalisation technologique. Pour améliorer encore la précision et l'efficacité, des travaux futurs envisagent d'entraîner le modèle sur un ensemble de données plus large et d'utiliser des caméras de meilleure qualité pour une détection en conditions de faible luminosité. Cette méthode est illustrée par la figure 2.1 [30].

FIG. 2.1 : Diagramme de l'architecture Proposée

Le travail de Muhtasim et al. [31]

Cet article propose une méthode innovante pour la détection de vivacité faciale, visant à renforcer la sécurité des systèmes d'authentification biométrique contre les attaques par usurpation d'identité. L'approche utilise un réseau CNN basé sur l'architecture VGG-16, modifiée pour traiter des images faciales segmentées en patchs. Cette segmentation permet de préserver la résolution d'origine des images et d'éviter le sur apprentissage en augmentant le nombre d'exemples d'apprentissage. Après le traitement initial par le CNN, les sorties sont envoyées à un réseau de mémoire à long terme (LSTM) pour capturer les structures temporelles des séquences d'images et déterminer si elles sont authentiques ou fausses. Cette architecture combinée CNN-LSTM permet d'extraire des informations spatiales et temporelles riches, améliorant ainsi la capacité de détection de vivacité. La méthode a été testée sur deux ensembles de données, REPLAY-ATTACK et CASIA-FASD, montrant des performances supérieures avec des taux d'erreur EER et HTER (Half-Total Error Rate) très bas, respectivement 0,67% et 0,71% pour CASIA-FASD, et 0,30% et 1,52% pour REPLAY-ATTACK. Les résultats indiquent que cette approche pourrait être utilisée efficacement dans des systèmes d'authentification biométrique pour améliorer la détection de vivacité et, par conséquent, la sécurité globale du système. Cette méthode est expliqué en détail dans la figure 2.2 [31].

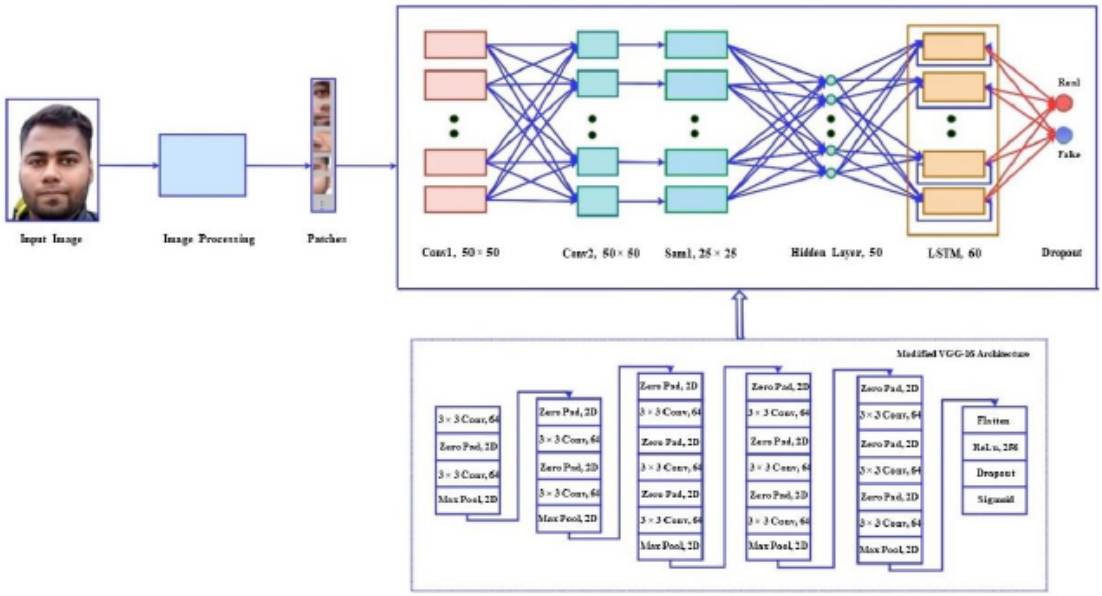


Fig. 2.2 : Le flux de travail de l'architecture CNN–LSTM basée sur des patchs avec le VGG-16 modifié [31].

Le travail de Li. [40]

La recherche présentée explore une architecture CNN à double canal visant à améliorer la détection des visages partiellement cachés. Cette architecture sera intégré dans le réseau VGG16 pour former un réseau neuronal perceptron d'occlusion

L'auteur commence d'abord par l'utilisation de Face++ pour détecter les points clés du visage et d'OpenCV pour sélectionner les régions d'intérêt. Les images sont dimensionnées à 128x128 pixels et découpées en sous-régions via une fenêtre de taille fixe avec un pas de glissement. Ensuite, deux canaux CNN sont utilisés. Un perceptron d'occultation ayant 12 couches convolutionnelles, c'est un réseau conçu pour extraire les caractéristiques des

images faciales occultées afin d'identifier et de traiter les caractéristiques importantes malgré les obstructions. Un réseau résiduel qui s'occupe de l'extraction des caractéristiques de l'image du visage dans son ensemble. Il utilise une architecture de réseau résiduel de 102 couches modifiée pour capturer les détails de l'ensemble du visage, même lorsque certaines parties sont occultées. Les deux ensembles de caractéristiques extraits des deux canaux CNN sont ensuite fusionnés pour améliorer la performance globale du modèle en termes de précision de détection des visages. Le mode de fonctionnement de cette méthode est expliqué dans la figure 2.3 [32].

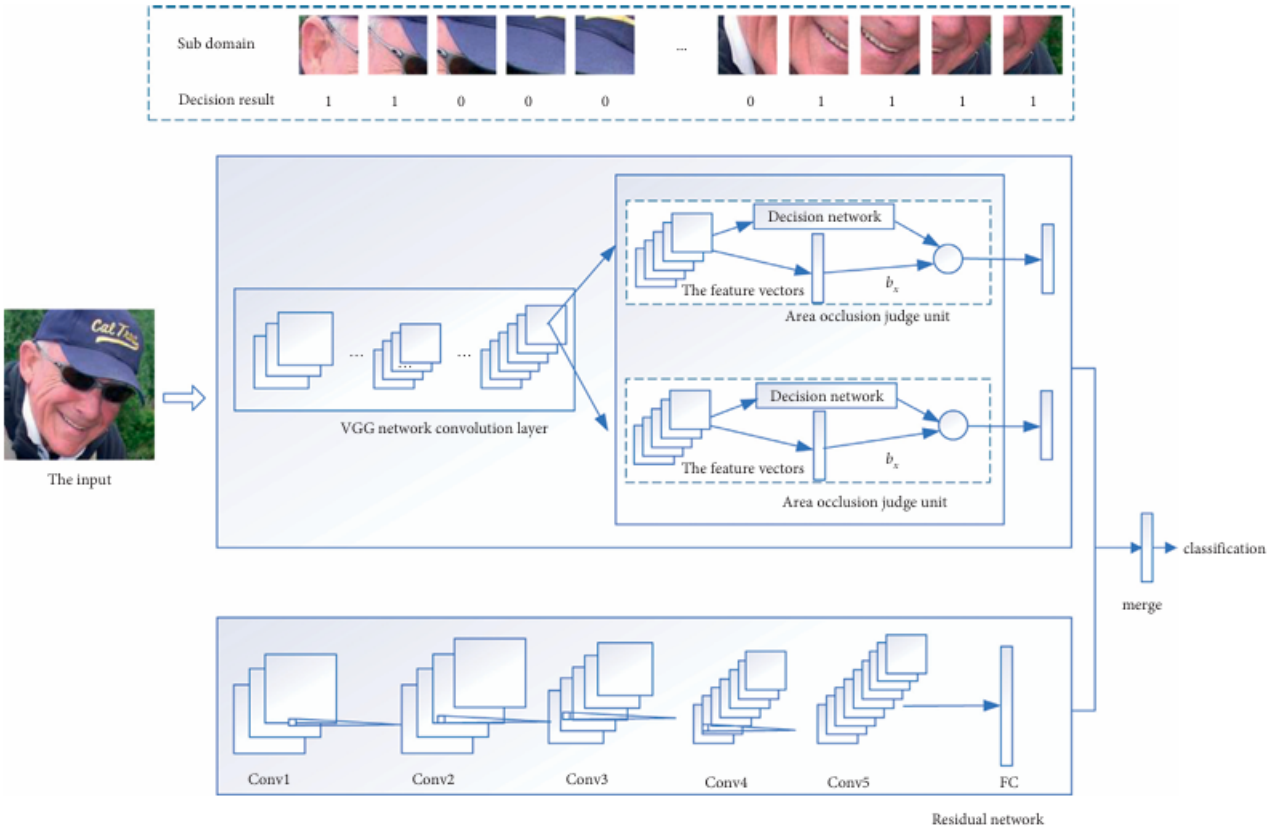


FIG. 2.3 : Approche proposée par Li [32].

Selon les tests effectués sur les datasets AR et MAFA, le modèle a montré une amélioration en terme de vitesse et de précision par rapport aux méthodes existantes, les résultats expérimentaux ont enregistré une précision de détection supérieure à 90% sur les ensembles de données utilisés.

Le travaille de Perdana et Prahara [41]

Dans ce travail, les auteurs ont proposé une méthode de reconnaissance faciale en appliquant un réseau de neurones convolutifs léger (Light-CNN) basé sur le modèle VGG16 auquel ils ont apporté des modifications. Leur approche expliquée dans la figure 2.4 [41] consiste à simplifier l'architecture du VGG16 dans un objectif de rendre le modèle plus compact et efficace. Pour atteindre cet objectif une couche des couches convolutionnelles à 64 filtres est supprimée et les couches convolutionnelles à 256 filtres et 512 filtres sont complètement retirées. De plus, la taille des couches fully connected est modifiée. Cette modification vise à maintenir des performances élevées tout en utilisant des ensembles de

données et des ressources computationnelles limités.

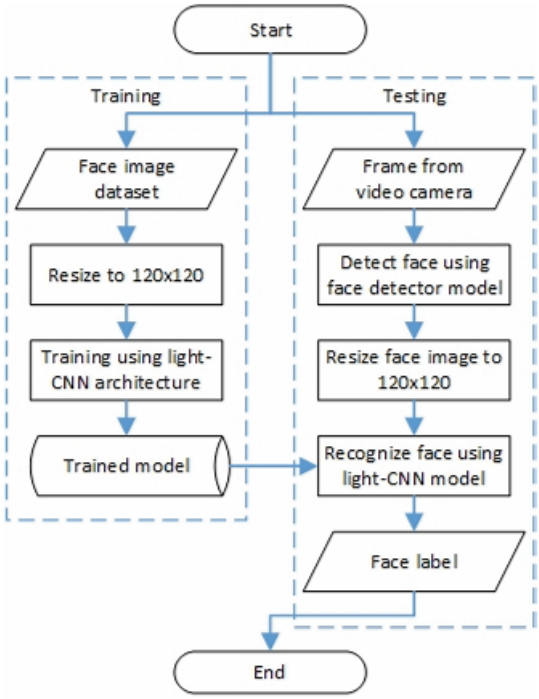


FIG. 2.4 : Approche proposée par Perdana et Prahara [41].

Pour évaluer leur travail, les auteurs ont utilisé un ensemble de données restreint. La performance du modèle à été mesurée en termes d'exactitude, de rappel, de précision et de score F1, en s'appuyant sur la matrice de confusion. Les résultats montrent que le modèle proposé atteint une précision de 94,4%.

Le travail de Zhiqi. [42]

Cette étude met en avant le développement du modèle MicroFace, qui s'agit d'une version améliorée du modèle VGG16 pour fixer ses limitations du grand nombre de couches et de la quantité du stockage requis .

Pour réduire le nombre de paramètres du réseau VGG-16, les auteurs commencent d'abord par supprimer deux couches fully connected ce qui fais que le nombre de noyaux de convolution dans le module 5 est augmenté de 512 à 560.

Une autre modification importante concerne la couche de pooling. La couche de pooling maximale est remplacée par la couche de sous-échantillonnage maximale a pass-bas (max-pooling lowpass), inspirée biologiquement et offrant de meilleurs résultats expérimentaux. Cette couche de pooling Lp est définie par une formule spécifique qui permet de maintenir l'extraction maximale des caractéristiques tout en réduisant le nombre de paramètres. Un fois toutes les modifcations appliquées au modèle VGG16 la structure détaillée du modèle amélioré, nommé MicroFace, est présentée dans la figure 2.5 [42].

Pour entraîner et tester l'algorithme amélioré, l'auteur a utilisé le dataset CASIA Web-Face. Les résultats ont enregistré une précision de 96.26%. Sur le plan technique, les paramètres du réseau VGG16 sont considérablement réduits, passant de 138 millions à 17 millions, ce qui permet d'économiser de l'espace mémoire et du temps d'entraînement tout en maintenant une capacité d'extraction de caractéristiques d'image efficace.

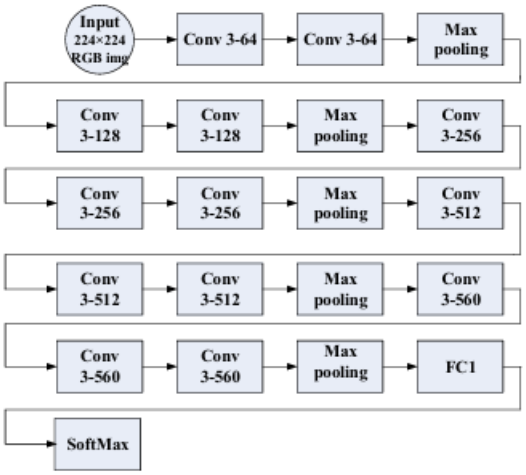


FIG. 2.5 : Structure du modèle Microface [42].

Le travail de Zhang et al. [43]

Bien que les détecteurs de visages existants sont prometteurs, ils ont du mal à détecter les visages capturés dans des scénarios complexes. Alors les auteurs proposent une méthode améliorée pour la détection des visages entraînée en particulier sur ceux qui sont petits et difficiles à détecter en utilisant le CNN et le modèle VGG16.

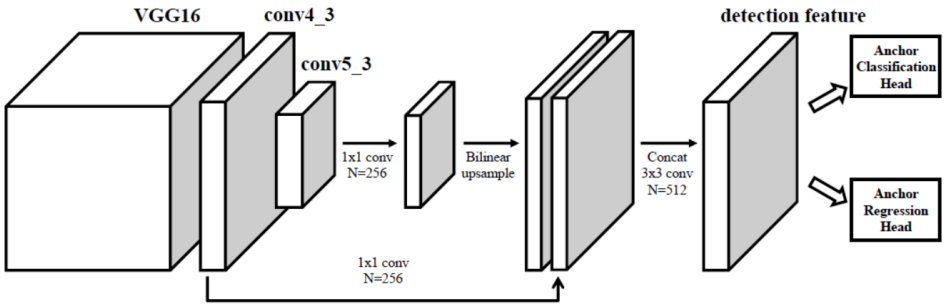


FIG. 2.6 : Approche proposée par Zhang et al [43].

Pour atteindre leurs objectifs, le détecteur de visage proposé a été construit en utilisant le modèle VGG16 comme base de CNN. L'architecture schématisée dans la figure 2.6 [43] consiste à modifier le VGG16 en combinant les caractéristiques des deux couches conv4_3 et conv5_3 afin de créer une carte de caractéristiques de détection qui intègre des informations sémantiques de bas et de haut niveau. Cette carte est ensuite utilisée pour la classification et la régression des boîtes englobantes des visages.

Ils ont mené de nombreuses expériences sur les datasets WIDER FACE, FDDB, Pascal Faces et AFW avec des précisions moyennes de 95,7%, 94,9% et 89,7% sur les sous-ensembles cités précédemment.

Le travail de Najibi et al. [44]

Le travail présenté dans cet article est le détecteur de visage "Single Stage Headed"(SSH), ce dernier parvient à détecter des visages en appliquant leur version modifié du VGG16.

L'architecture du SSH (voir la figure 2.7 [44]) est entièrement convolutionelle sans les couches fully connected "head" du VGG16 qui permet la détection des visages de différents format avec un seul passage sans avoir à utiliser une pyramide d'images. Des modules de

détection on été ajoutés aux couches convolutionneles qui sont : M1 ajoutée sur la couche conv4-3 (stride de 8) pour détecter les petits visages, M2 ajoutée sur la couche conv5-3 (stride de 16) pour détecter les visages moyens, M3 ajouté sur la couche conv5-3 avec une couche de max-pooling (stride de 32) pour détecter les grands visages. Pour assurer la robustesse de la détection, SSH incorpore des modules de contexte utilisant des couches convolutionnelles de plus grande taille (5x5 et 7x7).

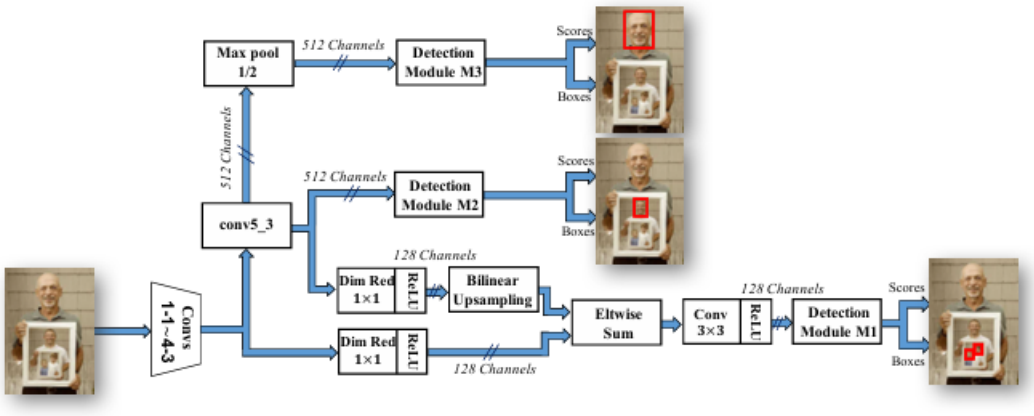


Fig. 2.7 : Approche proposée par Najibi et al [44].

Pour la phase d'évaluation, ils ont choisis de tester le SSH sur les datsets WIDER, FDDB et Pascal-Faces. Concernant le dataset WIDER surpasse le ResNet-101 en termes de précision tout en étant 5 fois plus rapide. Pour les datasets FDDB et Pascal-Faces, les expériences ont montré qu'avec une taille d'entrée relativement petite : SSH atteint une vitesse de 50 images par seconde sur un GPU.

2.3.2 Travaux Fecenet

Le travail de Manna et al. [33]

Les auteurs présente leur travail consistant en un système biométrique de reconnaissance faciale à partir des vidéo à distance, dont l'objectif est d'identifier les criminels. Le modèle pré-entraîné utilisé dans cette étude est FaceNet (FN).

D'après l'architecture (voire la figure 2.8 [33]) de leur système, celui-ci commence d'abord par la collecte de données en télé-chargeant les images des acteurs indiens de google. Ensuite, l'identification et préparation des visages dans les images de la base de données. En dernier, la vérification et l'identification des visages à partir des images ou des vidéos pour lesquelles le système a été formé en utilisant le modèle pré-entrainé Face Net dont l'architecture comprend une couche d'entrée par lot, un CNN profond suivi d'une normalisation L2, ce qui conduit à l'embedding (représentions vectorielles) des visages. La distance entre deux points dans l'espace euclidien est alors calculée pour déterminer la similarité des visages.

Pour la phase d'évaluation des performances, le modèle a été testé et a montré une précision d'environ 90%. FaceNet a montré une précision supérieure par rapport à d'autres modèles comme DeepFace et DeepID1.

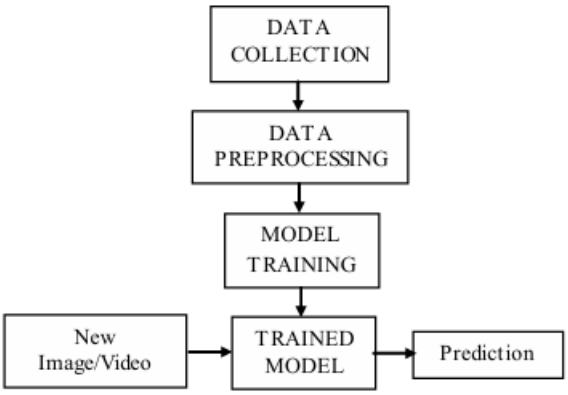


FIG. 2.8 : Approche proposée par Manna et al [33].

Travail de Kuand et Baul [34]

Dans cet article les auteurs proposent un système de présence en temps réel utilisant la reconnaissance faciale basée sur l'apprentissage en profondeur pour améliorer les méthodes de présence traditionnelles. Le système capture les images des étudiants via une webcam, stocke les images de référence dans une base de données qui fait que la détection des visages est réalisée en utilisant des classificateurs en cascade de caractéristiques de Haar, ce qui permet de localiser précisément la région du visage dans les images capturées. Ensuite, le réseau neuronal profond pré-entraîné FaceNet est utilisé pour extraire des encodages faciaux de dimension 128 à partir des visages détectés. Une comparaison entre ces encodages et ceux de la base de données pour déterminer les correspondances. Enfin, les résultats de la reconnaissance faciale sont utilisés pour mettre à jour les enregistrements de présence dans un fichier. L'architecture de leur système est illustrée par la figure 2.9 [34].

Le système a atteint une précision d'environ 95 % dans des conditions constantes, démontrant ainsi son efficacité dans l'automatisation des contrôles de présence et la réduction des fraudes.

Travail de Schroff et al. [38]

FaceNet propose une solution efficace pour la vérification et la reconnaissance faciale à grande échelle en apprenant un espace de représentation compact des visages. Le système utilise des réseaux de convolution profonds pour optimiser directement l'espace d'embedding, nécessitant seulement 128 octets par visage, et atteint des performances de pointe avec une précision de 99,63% sur le dataset LFW et de 95,12% sur YouTube Faces DB. FaceNet introduit également les concepts d'embedded harmonique et de perte de triplet harmonique pour permettre la comparaison directe entre différentes versions des embeddings de visage. La structure du modèle proposé est illustrée par la figure 2.10 [38].

2.4 Modèles de Détection Combinés et Avancés

Pour enrichir notre étude et offrir une perspective plus complète sur les techniques de détection de visage, nous allons maintenant explorer d'autres modèles de détection, y compris ceux qui combinent plusieurs technologies avancées pour améliorer la précision et l'efficacité du processus de reconnaissance faciale.

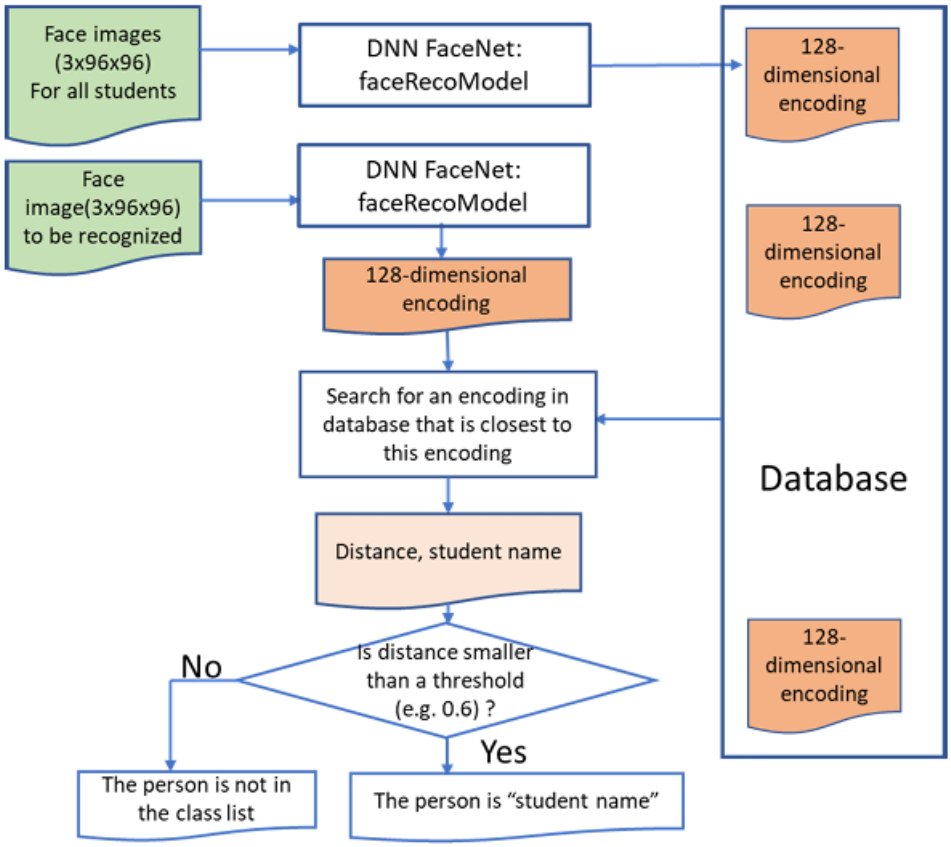


FIG. 2.9 : Approche proposée par Kuand et Baul [34]

FIG. 2.10 : Approche proposée par Schroff et al [38].

Travail de Thai-Viet Dang et al. [35]

L'article présente une nouvelle approche de la reconnaissance faciale combinant l'estimation de la pose de la tête avec les réseaux de neurones MTCNN et FaceNet pour améliorer l'efficacité de la reconnaissance et éviter les faux positifs. En utilisant un ordinateur embarqué Jetson Nano et une caméra IMX-219, le système capture des images sous cinq poses principales (face droite, tête tournée à gauche, à droite, en haut et en bas) pour vérifier l'identité de l'utilisateur. Le MTCNN détecte la zone du visage sur les images alignées, tandis que FaceNet convertit ces images en vecteurs de 128 caractéristiques pour la reconnaissance finale. L'entraînement de FaceNet implique un processus de sélection et de comparaison d'images ancrées, positives et négatives pour maximiser la séparation entre les différentes identités. Le système atteint une précision de 90-95% avec une vitesse de traitement de 21-25 fps, démontrant une performance robuste sur des images complexes, tout en restant assez simple pour être exécuté sur du matériel embarqué à faible puissance.

Travail de Huang et Luo [36]

Dans le domaine éducatif, les auteurs propose un système de reconnaissance faciale en temps réel pour la gestion des présences, en combinant des réseaux de neurones convolutionnels multitâches (MTCNN), FaceNet et l'algorithme Gradient Boosting Decision Tree

(GBDT).

Le processus d'identification illustré par la figure 2.11 [36] passe par trois étapes : la première étape, consiste à capturer les images des visages dans une base de données et utiliser le réseau MTCNN pour détecter et aligner les visages dans les images capturées effectuant plusieurs tâches simultanément : il détecte la présence d'un visage, identifie les points clés (comme les yeux, le nez, la bouche) et aligne le visage pour une meilleure reconnaissance. Dans la deuxième étape, la présence en temps réel prend lieux en utilisant les trois technologies cité précédemment : MTCNN est utlisé pour détecter les visages en temps réel à partir des flux vidéo ou des images prises à des intervalles réguliers, les visages détectés sont ensuite passés par le modèle FaceNet, qui pour chaque visage génère des embeddings (représentations vectorielles). Ces embeddings sont comparés avec ceux stockés dans la base de données pour identifier les individus présents, l'algorithme GBDT est utilisé pour améliorer la précision de la classification des visages en exploitant les embeddings générés par FaceNet.

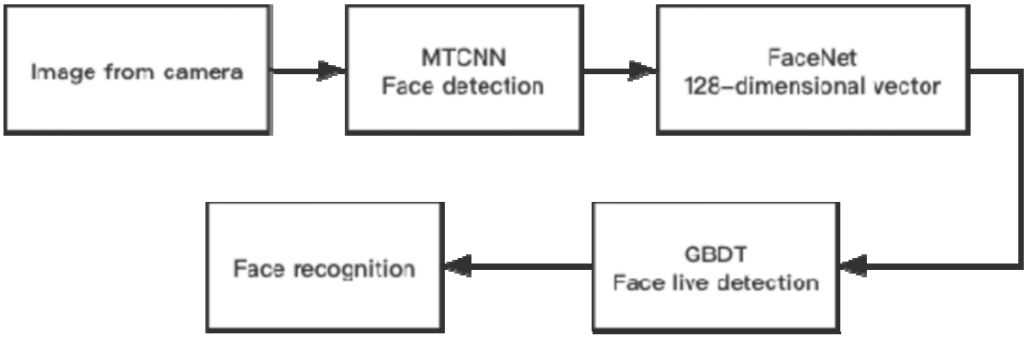


Fig. 2.11 : Processus d'identification propposé par Huang et Luo [36].

Les résultats montrent que le taux de fausses acceptations (FAR) est de 1.85% et le taux de faux rejets (FER) est de 1.92%. Le taux de reconnaissance peut être stable autour de 20 images par seconde (FPS).

Le travail de Nida et al. [45]

Avec les avancées technologiques la reconnaissance de la falsification des images devient difficile alors les auteurs proposent dans leur travail la reconnaissance faciale des réelles ou fausses images utilisant le CNN et l'apprentissage profond avec l'analyse du niveau d'erreur (ELA).

D'après la structure de leur approche expliquée dans la Figure 2.12 [45], le mode de fonctionnement passe par quartes étapes. La première étape consiste à normaliser les images acquises 128x128 pixels pour reconnaître les fausses images des vraies. Dans la deuxième étape, l'analyse du niveau d'erreur prend place permettant : la détection des images réelles ou manipulée en enregistrant un niveau de qualité. La troisième étape se résume à diviser le jeux de données en ensembles d'entraînement et de test, puis passer à des modèles d'apprentissage profond pré-entraînés (VGG-16, ResNet-50, InceptionV3 et VGG-19). Dans la dernière étape, le CNN est appliqué, les images entraînées sont passées à une couche aplatie et une couche dense séquentielle. La couche fully connected (dense) est utilisée pour la reconnaissance des motifs, avec une fonction d'activation SoftMax pour convertir les vecteurs de caractéristiques en probabilités, comparant l'ensemble d'entraî-

nement avec l'ensemble de test pour fournir une distribution de probabilité sur les images réelles et falsifiées.

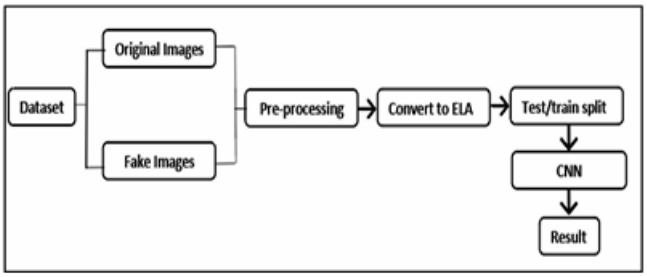


FIG. 2.12 : Approche proposée par Nida et al [45].

Pour évaluer les performance de leur méthode, les auteurs ont utilisé le dataset Real and Fake Face detection du Computational Intelligence Photography Lab de l'Université Yonsei [46]. Ils ont testé plusieurs modèles pré-entraînés parmi eux, les modèles VGG donnent les meilleures précisions d'entraînement avec 91,97% pour VGG-16 et 92,09% pour VGG-19.

Le travail de Kumar et al. [47]

Les chercheur présente leur travail qui a consisté à développé une méthode améliorée pour la détection des masques faciaux dans les images et les vidéos en temps réel en utilisant le modèle Caffe-MobileNetV2 (CMNV2), le modèle modifié CMNV2 intègre cinq nouvelles couches pour améliorer la classification. Le processus utilise OpenCV pour détecter les visages et applique des techniques d'apprentissage profond pour identifier les régions d'intérêt (ROI). Les expériences montrent que le modèle atteint une précision de 99,64% sur les images de photos et des performances solides en temps réel, surpassant les méthodes existantes en termes de précision.

L'architecture du modèle CMNV2 modifié est illustreé par la Figure 2.13 [47].

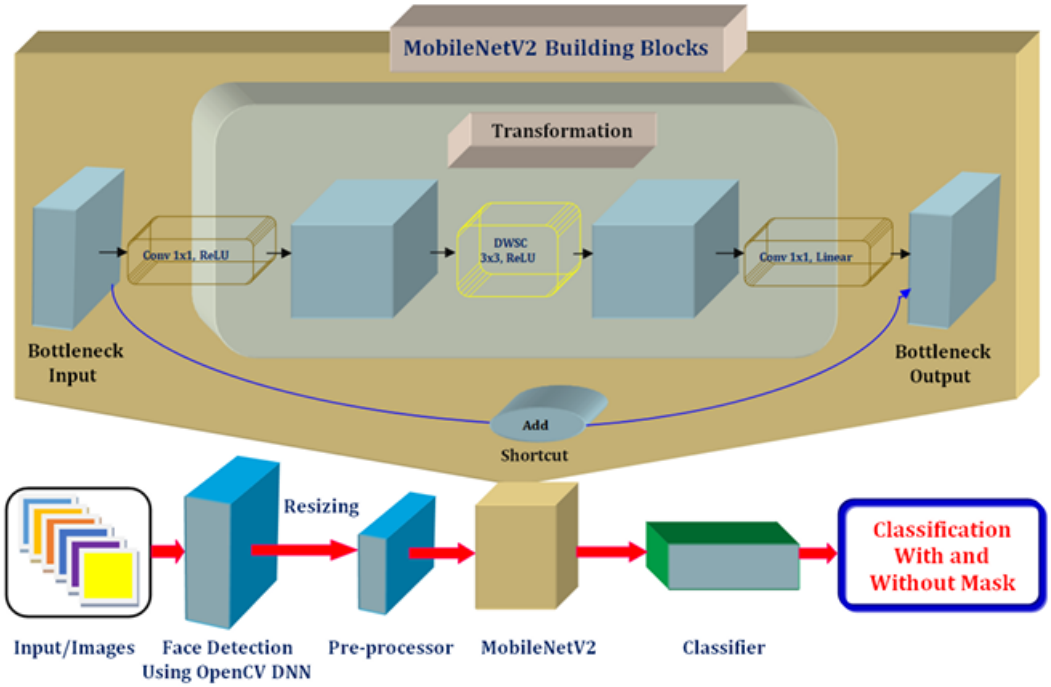


FIG. 2.13 : Architecture du modèle modifié CMNV2 [47].

2.5 Comparatif des travaux

Dans le domaine en évolution rapide de la détection faciale, de nombreuses études ont été menées pour évaluer et comparer diverses techniques. Parmi celles-ci, plusieurs articles récents se sont concentrés sur la comparaison de différentes approches, mettant en évidence leurs forces, leurs faiblesses et leurs performances dans des scénarios spécifiques. Par exemple, l'étude faite par Hasin et ses collègues [48] a comparé les performances de plusieurs réseaux neuronaux convolutifs (VGG19, VGG16, VGGFace, DenseNet, DenseNetDenseNet, ResNet50) pour la détection et la vérification de visages. De même, le travail de Gwyn et al. [49] a analysé l'efficacité des méthodes basées sur des architectures de réseaux profonds (AlexNet, Xception, Inception, Inception, ResNet50, ResNet101, VGG16, VGG19) pour la reconnaissance faciale dans divers contextes.

Dans cette section, nous résumons et comparons les principales conclusions de ces études, ainsi que d'autres travaux pertinents, afin d'éclairer davantage les choix et les défis dans le domaine de la détection faciale. Ceci permettra d'avoir une vision plus claire et de pouvoir comparer les approches, ainsi que leurs points distinctifs, selon des mesures de performance fixes. Les tableaux suivants (voir Tableaux 2.1, 2.2, 2.3) présentent une comparaison détaillée des différents modèles de détection faciale en termes d'auteur, année, ensemble de données utilisés, techniques employées, limites identifiées et performances obtenues.

Auteur	Année	Dataset	techniques	Limites	Performance
Perdana et Prahara[41]	2019	ROSE-Youtu Face Liveness Detection	VGG16	Nécessite plus de ressources computationelles pour l'entraînement et l'inférence, Jeu de Données Limité	Acc= 94,4%
Zhiqi. [42]	2021	CASIA Web-Face	VGG16	L'efficacité du modèle peut diminuer si les images de test présentent des arrière-plans plus complexes ou des conditions de lumière variables.	Acc =96%
Zhang et al [43]	2018	WIDER FACE FDDB Pascal Faces	VGG16	Concentrer l'entraînement sur les images difficiles fais que le modèle devient très bon pour détecter des visages dans des conditions extrêmes mais moins performant dans des conditions normales ou faciles.	Wider Face Average precision(AP)=95,7% FDDB AP =94,9% Pascal Faces AP=89,7%
Najibi et al. [44]	2017	WIDER FACE FDDB Pascal Face	VGG16	Le SSH bien qu'il détecte des visages de différentes tailles, la détection de visages très petits peut encore poser des défis.	WIDER 5 fois plus rapide que ResNet-101.Pour FDDB et Pascal-Faces, avec une taille d'entrée relativement petite : SSH atteint une vitesse de 50 images par seconde sur un GPU

Nida et al. [45]	2021	REPLAY-ATTACK, CASIA-FASD	Patch-based CNN, VGG-16, LSTM	Vulnérabilité aux attaques par usurpation, dépendance aux datasets spécifiques	HTER de 0.71% (CASIA-FASD), EER de 0.67% (CASIA-FASD), HTER de 1.52% (REPLAY-ATTACK), EER de 0.30% (REPLAY-ATTACK)
Kumar et al. [47]	2023	Photo images et vidéos en temps réel	Caffe-MobileNetV2 Transfer Learning	Précision variable selon le modèle, coût computationnel, complexité de formation	Exactitude sur images photo : 99.64%, Précision : 100%, Rappel : 99.28%, F1-score : 99.64%, Taux d'erreur : 0.36%
Manna et al [33]	2020	"Real and Fake Face detection" par Yonsei University	ELA (Error Level Analysis), CNN (VGG-16, VGG-19, ResNet-50, InceptionV3)	Précision variable selon le modèle, problèmes de surapprentissage, sensibilité aux paramètres de compression	Précision d'entraînement : 91.97% (VGG-16), 92.09% (VGG-19) ; Précision de test : 64.49% (VGG-16), 60.63% (VGG-19)
Kuand et Baul [34]	2020	Images des étudiants capturées par webcam	Haar Cascade, FaceNet p	Sensibilité aux changements de lumière et de distance, nécessite des conditions similaires pour une reconnaissance précise	Exactitude de reconnaissance faciale de 95%

Dawen.A.M et al. [31]	2022	REPLAY-ATTACK, CASIA-FASD	Patch-based CNN, VGG-16, LSTM	Vulnérabilité aux attaques par usurpation, dépendance aux datasets spécifiques	HTER de 0.71% (CASIA-FASD), EER de 0.67% (CASIA-FASD), HTER de 1.52% (REPLAY-ATTACK), EER de 0.30% (REPLAY-ATTACK)
C Vimal et al. [30]	2023	10,000 images (5,000 avec masque, 5,000 sans masque)	VGG-16, Apprentissage par transfert, Augmentation d'images	Précision variable pour les petits visages, performances réduites en faible luminosité, besoin de ressources informatiques importantes	Précision=94% Re-call=95% F1-Score=94%
Thai-Viet Dang et al. [35]	2023	IJB-A, IJB-B, IJB-C, CS5	MTCNN, FaceNet, estimation de pose de tête	Complexité de l'intégration du système sur des dispositifs à faible puissance	Précision : 90-95%, FPS : 21-25
Scroff et al. [38]	2015	Labeled Faces in the Wild (LFW), YouTube Faces DB	FaceNet	- Besoin de grande puissance de calcul pour l'entraînement - Performance dépendante de la qualité des triplets sélectionnés	LFW ACC =99,63% youtube face DB ACC 95,12‘%

Huang et Luo [36]	2020	Dataset interne	Des conditions d'éclairage insuffisantes ou des angles de vue défavorables peuvent affecter la précision de la reconnaissance.	MTCNN FaceNet Algorithme GBDT	FAR = 1.85% FER=1.92% Taux de reconnaissance stable à 20 FPS
Yueying Li [40]	2022	Dataset AR Dataset MAFA	Réseau de neurones à double canal avec perceptron d'occlusion intégré dans VGG16	La méthode peut être affectée par des conditions d'occlusion complexes	Précision de détection plus élevée et vitesse plus rapide que les autres méthodes

2.6 Analyse des tendances et défis actuels

2.6.1 Incorporation des techniques de prétraitement et d'augmentation

Travail de Siru Chen [37]

Cette etude explore l'utilisation d'un CNN pour la reconnaissance faciale, pour améliorer les performances avec l'ensemble de données limité Olivetti, l'augmentation des données a été utilisée. Le modèle comprend deux couches de convolution, des couches de max-pooling, une couche entièrement connectée et un classificateur softmax, utilisant l'optimiseur Adam et un taux de dropout de 0,25. Des techniques d'augmentation, telles que la rotation, le flip horizontal et le décalage, sont appliquées pour diversifier les données d'entraînement. Pour tester les performances de son modèle, les expérimentations on été faites en la présence de l'augmentation des données et son absence. Sans augmentation, le modèle atteint une précision de 91,67% après 100 époques, mais présente des signes de sous-apprentissage. Avec l'augmentation, la précision s'améliore à 94,17% après 150 époques, malgré une convergence plus lente.

Travail de Kuang et Baul[34]

Cet article, dans le but d'améliorer la reconnaissance faciale, propose une méthode d'augmentation de données et d'apprentissage par transfert. L'approche consiste à appliquer des transformations 2D linéaires aux images, telles que des ajustements de luminosité, de contraste et de saturation, générant ainsi 11 ensembles de données augmentés. Ces ensembles sont traités par le CNN VGG-Face pré-entraîné pour extraire des caractéristiques, qui sont ensuite utilisées pour entraîner un classificateur KNN. Les tests effectués sur les ensembles de données LFW et un ensemble propriétaire ont montré une amélioration significative de la précision, atteignant jusqu'à 98,43% sur LFW.

Travail de Mungra et al. [32]

L'article propose une méthode de reconnaissance des émotions faciales en utilisant le CNN. Pour enrichir le jeu de données et améliorer la précision de la classification, les auteurs appliquent diverses techniques d'augmentation des données, comme l'ajustement de la luminosité, le contraste, l'inversion des couleurs, le bruit aléatoire, le zoom, la rotation et le retournement des images. L'approche inclut le prétraitement des images et la détection et le recadrage des visages, l'utilisation d'un CNN à sept couches, l'optimisation des hyperparamètres et l'évaluation de la performance. Cette méthode permet d'atteindre une précision de 79,19% avec l'égalisation d'histogramme, démontrant l'importance de l'augmentation des données et du pré-traitement pour améliorer la robustesse des modèles de reconnaissance des émotions faciales.

2.6.2 Multimodalité

Le travail de Khemar et al. [50]

Khemar et al. proposent une plate-forme d'authentification biométrique innovante destinée a être intégrée dans les applications de contrôle d'accès basées sur la reconnaissance faciale. L'objectif des chercheurs est de développer un fauteuil roulant intelligent pour les personnes handicapées. La plateforme s'agit d'un système de vision multicapteurs qui fusionne un capteur catadioptrique et une caméra pan tilt zoom (PTZ), le processus de reconnaissance est basé sur l'algorithme d'Eigenfaces.

Pour atteindre leur objectif, les auteurs on utilisé leur propre base de données hétérogène comprenant des images de démarche, de visage et d'iris. Les images faciales ont été collectées selon un protocole spécifique (position de la tête par rapport aux yeux, expression faciale et conditions d'éclairage) hétérogène, combinée avec la base de données ORL (Olivetti Research Laboratories). Pour évaluer leur plateforme plusieurs test on été réalisés et en analysant les résultats de l'histogramme la combinaison des algorithmes CS (Compressing Sensing) et SS (Sub-Sampling) offre les meilleures performances. Sur dix itérations, neuf ont atteint un taux de reconnaissance supérieur à 0,9, et dans chaque itération, le taux de reconnaissance de cette combinaison est supérieur à celui des deux autres combinaisons d'algorithmes (CS + LBP et LBP + NN).

Le travail de Hussain et al. [51]

L'article présente un système biométrique multimodal utilisant les modalités du visage, de l'iris et de l'empreinte digitale. Les auteurs ont employé plusieurs modèles de réseaux CNN, notamment VggNet16, ResNet50, MobileNet, DenseNet et GoogleNet, pour extraire

des caractéristiques à partir de ces modalités. La fusion des caractéristiques est effectuée à deux niveaux : au niveau des caractéristiques et au niveau des scores. La fusion au niveau des caractéristiques combine les informations riches des différentes modalités avant le module de classification softmax. D'autre part, la fusion au niveau des scores agrège les résultats des modèles CNN, en utilisant la règle de somme arithmétique. Ces deux méthodes de fusion visent à améliorer la précision de l'identification biométrique multimodale. L'architecture du modèle proposé est illustré dans la figure 2.14 [51].

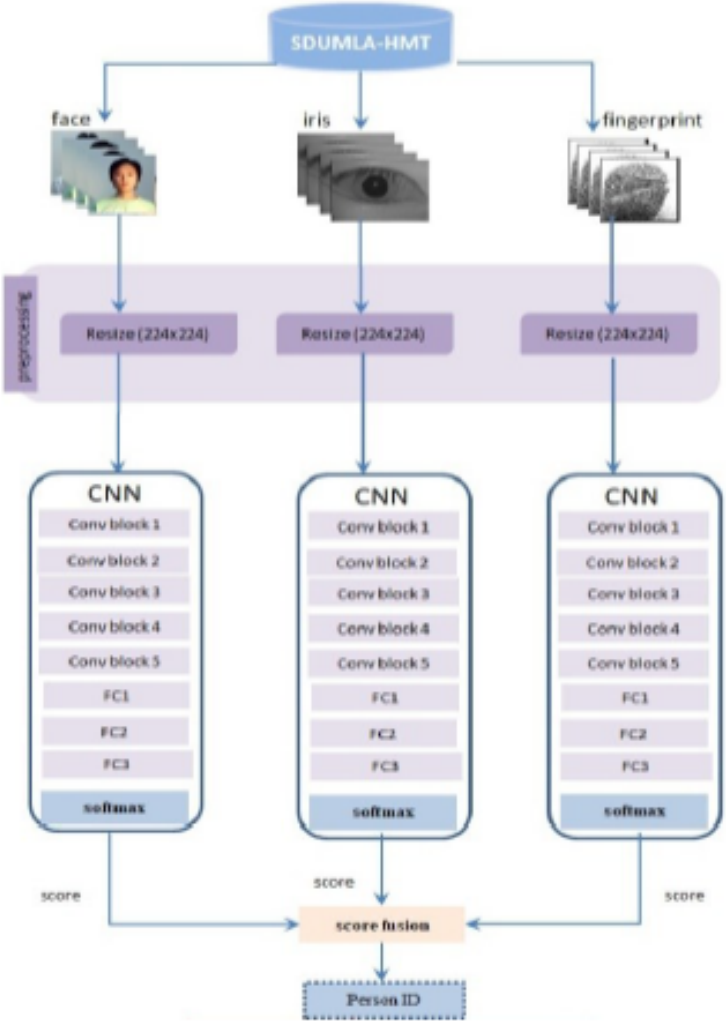


Fig. 2.14 : La structure du modèle biométrique multimodale utilisant une approche de fusion au niveau des scores [51].

Le travail de Alin Majed Alkadi et al. [52]

L'article en question explore la mise en œuvre d'un système novateur d'authentification biométrique destiné aux distributeurs automatiques de billets (DAB), se concentrant spécifiquement sur la reconnaissance faciale. L'approche préconisée repose sur l'utilisation d'images capturées simultanément dans les spectres visible, thermique et infrarouge, ainsi qu'une fusion d'infrarouge et de spectre visible. Le modèle de machine learning ResNet-50, minutieusement formé sur le jeu de données Sejong [53], est déployé pour garantir une reconnaissance faciale précise. Le système a atteint une reconnaissance de plus de 99%, même avec l'ajout d'accessoires courants tels que des lunettes de soleil, des masques, des bonnets et des écharpes. La reconnaissance des sujets féminins portant le hijab a présenté

des lacunes, surtout avec l'ajout d'accessoires. Les résultats ont été mitigés, en particulier pour les cas avec des ajouts tels que les casquettes. L'expansion du jeu de données avec des sujets féminins portant le hijab a amélioré la reconnaissance, mais certaines combinaisons d'ajouts ont posé des problèmes, soulignant la complexité de la reconnaissance avec des accessoires.

Les résultats ont montré une capacité limitée à reconnaître des sujets féminins qui ont ajouté le hijab après avoir été initialement photographiées sans celui-ci.

FIG. 2.15 : Images complémentaires du spectre visible pour le sujet féminin non-hijabi [52].

Le travail de Sarangi et al.[54]

L'étude des auteurs Partha Saranghi et al. [54] présente une solution biométrique multimodale se basant sur deux traits biométriques physiques : l'oreille et le profil faciale en utilisant une fusion au niveau des caractéristiques. Selon l'architecture proposé, le système commence d'abord par efficacement extraire les caractéristiques des deux modalités en les représentant séparément combinant deux descripteurs de caractéristiques : l'un qui est LPQ (Local Phase Quantization) et l'autre LDP (Local Directional Patern). Une fois ces derniers combinés en des vecteurs de caractéristiques, une réduction de dimensions avec la technique PCA et une normalisation ont été appliquées en utulisant la technique Z-score. Ces vecteur résultants sont alors fusionnés en un seul ensemble de caractéristiques. Comme dernière étape, dans un but de réduction de l'ensemble des caractéristiques obtenues de la fusion, l'approche KDCV (Kernel Discriminative Common Vector) a été appliquée. Le mode de fonctionnement de cette approche est détaillé dans le schéma explicatif de la figure 2.16 [54].

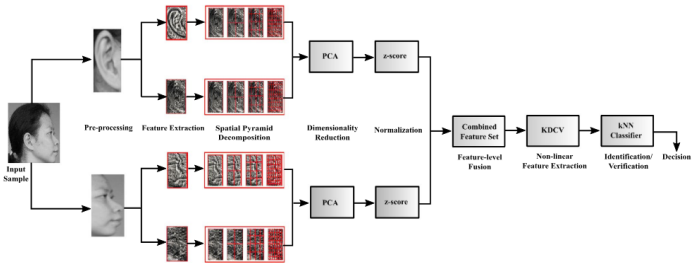


FIG. 2.16 : Architecture du modèle proposé par Sarangi et al [54].

Pour la phase d'évaluation de leur système, les test de performances on été organisés sur chaque modalité de façon individuelle. Pour réaliser une analyse comparative avec

les méthode de l'état de l'art, ils ont utilisé deux bases de données UND-E, et UND-J2. Les résultats démontrent un taux d'identification maximum de 99.42 % pour la base de données UND-E et 98.53% pour la base de données UND-J2.

Le travail de Mohamed et al. [55]

Les auteurs Mohamed et al. proposent une solution biométrique aux modalités : Empreinte digitale et visage. Dans un objectif d'améliorer la précision de l'identification Diseigen a été utilisé, un algorithme de fusion au niveau des caractéristiques.

Tous d'abord le système proposé commence par l'acquisition des caractéristiques et leur pré-traitement, ensuite la technique "AUMI (Aspect United Moment Invarient) " est appliquée pour extraire les caractéristiques. Une fois ces étapes réalisées les caractéristiques sont fusionnées grâce à la technique de l'algorithme DisEigen. L'algorithme proposé convertit les multi-représentations des caractéristiques individuelles en une représentation unique. Les caractéristiques généralisées d'un individu ont été présentées de manière significative. Une fois la fusion terminé, il vient une recherche de correspondance entre les bases de données et une décision en résulte sur l'identité de l'individu. La figure 2.17 [55] illustre le mode de fonctionnement de l'approche proposée.

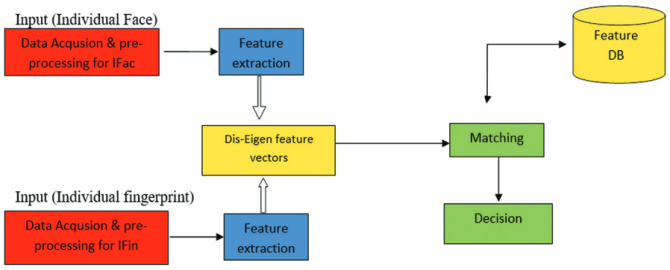


FIG. 2.17 : Architecture de l'approche proposée par Mohamed et al. [55]

Le travail de Gunasekarana [56]

Dans cet article [56], l'auteur Gunasekarana propose un système multimodale biométriques aux modalités : empreinte digital, iris et visage. Le mode de fonctionnement de ce système consiste en plusieurs étapes. Pour commencer, les données recueillies subissent un prétraitement pour en réduire la complexité. Ensuite, les caractéristiques pertinentes sont extraites à l'aide du modèle Local Derivative Ternary Pattern, impliquant l'utilisation d'histogrammes pour chaque trait biométrique. Les images sont ensuite fusionnées en utilisant une méthode de fusion pondérée au niveau des rangs, où les pondérations reflètent l'importance relative de chaque type de données. Enfin, la mise en correspondance est réalisée à l'aide de techniques d'apprentissage profond .

Pour ce qui est de la phase de l'évaluation et des tests, le chercheur a employé le dataset CCASIA Biometric Idea. Les résultats obtenus révèlent un taux de reconnaissance maximal de 96%, avec un temps de traitement moyen de 49,2 ms pour les échantillons biométriques provenant de 500 individus.

Le travail de Tharewal et al. [57]

Le travail de Tharewal et al. [57] consiste au développement d'un système biométrique aux modalités image et oreille format 3D utilisant une fusion au niveau des scores. Leurs objectifs est de résoudre les limitations qu'offrent un système unimodal aux modalités sous format 2D et sa non fiabilité.

Pour atteindre cet objectif, les chercheurs ont commencé par l'acquisition des images en utilisant les base de données FRGC pour le trait visage en 3D et la base de données UND pour les images en 3D de l'oreille. Ensuite, les fichiers ".abs" ont été lus et prétraités pour détecter la pointe du nez et recadrer les images après des opérations de lissage et de remplissage des trous. Les caractéristiques on été extraites à l'aide d'une analyse en composantes principales pour le visage et (Iteretive Closest Point) pour la modalité oreille. Douze caractéristiques ont été utilisées pour les visages 3D et neuf pour les oreilles 3D. En utilisant la distance euclidien les caractéristiques extraites sont alignées et comparées pour après avoir fusioné les scores générés. En dernier lieu, en se basant sur le score final, une décision est prise. L'architecture de l'approche proposée est visualisée dans la figure 2.18 [57].

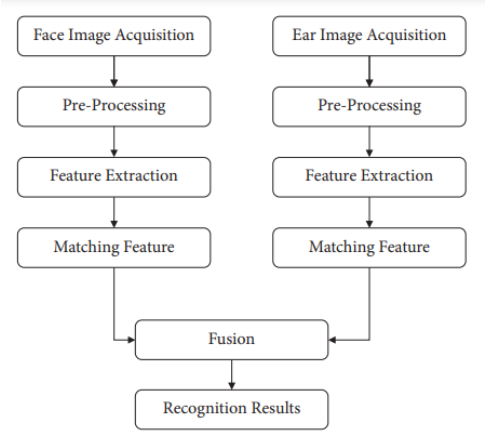


FIG. 2.18 : Modèle proposé du système biometrique multimodal [57].

Les résultats expérimentaux ont révélé une précision de 63,44% visage en 3D avec un seuil d'erreur de 36,56%. Pour l'oreille en 3D, une précision de 86,36% a été obtenue avec un taux d'erreur de 13,64%. Tandis que le modèle de fusion au niveau du score proposé a atteint une précision de 99,25% avec un seuil d'erreur de 0,75%. Des analyses de performances approfondies ont révélé que le modèle proposé a réalisé une amélioration moyenne de 1,2847% par rapport aux modèles concurrents.

Le travail de Hattan et al. [58]

Les deux auteurs [58] proposent un système biométrique multimodale basé sur la reconnaissance visage et deux iris. Trois architectures on été proposé dans leur travail, chaque architecture est basée sur son propre niveau de fusion : la fusion au niveau des images, fusion au niveaux des caractéristiques et la fusion au niveau des scores.

Le système proposé est basé sur un enchaînement de trois étapes. Les régions d'intérêt sont d'abord détectées en utilisant le modèle YOLOV-4 prés-entrainé avec le dataset MSCOCO. Ensuite, un réseau d'extraction de caractéristiques inspiré du modèle pré-entrainé Xception a été conçue. Ce réseau simplifié ne comprend que huit blocs résiduels, pour un total de vingt couches convolution, suivi d'une couche GAP. Le modèle proposé, utilise environ 7 millions de paramètres et se sert de la couche GAP pour réduire la dimensionnalité des caractéristiques de 361*728 à 728. La dernière étape classification est réalisée avec le classificateur LinearSVC qui sépare une classe des autres (One-Versus-All) en créant un hyperplan linéaire. Il cherche à maximiser les marges entre les classes en se basant sur une transformation des dimensions et sur les vecteurs supports.

Ces quatre base de données on été utilisées ORL, FERET, Geo tech. SDUMLA-HMT est utilisée pour tester la robustesse du système de reconnaissance visage obtenant ainsi les résultats 95.63%, 99.90%, 93.08% et 63.3 % sur les bases de données ORL, FERET, and GEORGIA TECH et SDUMLA-HMT. L'évaluation du système de reconnaissance iris a atteint des niveaux de précision remarquables pendant le protocole de validation croisée, dépassant 99% sur les bases de données IITD et CASIA Interval. De plus, la base de données iris SDUMLA-HTM a montré un taux de précision minimum de 98,68% et un maximum de 100 %.

2.7 Conclusion

Ce chapitre a offert une exploration approfondie des recherches récentes sur les modèles de détection et de reconnaissance faciale, mettant en lumière les progrès significatifs ainsi que les défis persistants dans ce domaine crucial. À travers l'analyse critique des articles de recherche pertinents, nous avons pu saisir l'évolution des techniques et des approches utilisées pour la détection des visages.

La revue des modèles clés tels que VGG16, FaceNet et ResNet a permis de comprendre leurs forces et leurs limitations dans différents contextes d'application. Nous avons également identifié les tendances émergentes, telles que l'importance croissante des réseaux de neurones profonds, ainsi que les défis actuels tels que la robustesse aux variations d'éclairage et les contraintes de données.

En comprenant les lacunes des modèles existants, analyse approfondie nous permettra d'acquérir une perspective éclairée sur l'état de l'art de la détection et de la reconnaissance faciale, jetant ainsi les bases pour le développement de solutions innovantes dans les chapitres à venir.

Chapitre 3

Amélioration de la Détection des Visages avec une Modification de la Couche de Classification du Modèle VGG16

3.1 Introduction

Comme exploré précédemment dans les chapitres un et deux, la reconnaisssance faciale a été adoptée comme solution dans les systèmes d'authentification barométriques afin pallier les limitations des méthodes traditionnelles. Les avancées dans les algorithmes de traitement d'images et les technologies de reconnaissance faciale ont considérablement élargi ce domaine. Plusieurs propositions différentes de systèmes, notamment celles présentées précédemment dans la synthèse de documents, ont vu le jour selon divers objectifs d'études.

Dans ce troisième chapitre, nous présentons notre approche méthodologique. Il abordera tout d'abord la problématique étudiée, puis mettra en avant la solution proposée en justifiant les choix techniques. Enfin, il présentera de manière détaillée les différentes phases du système de reconnaissance faciale proposé.

3.2 Problématique

Avec l'essor fulgurant du commerce en ligne et des services numériques, le nombre d'internautes a augmenté. En janvier 2024, il y avait 5,35 milliards d'internautes dans le monde, soit 66,2% de la population mondiale. Sur ce total, 5,04 milliards, soit 62,3% de la population mondiale [59], de ce fait la transactions en ligne a considérablement augmenté. Cette croissance entraîne une exposition accrue aux risques de sécurité et aux cyberattaques.

Face à ces enjeux, l'authentification biométrique a émergé comme une solution alternative plus sécurisée. Parmi les différentes techniques biométriques, la reconnaissance faciale largement étudiée depuis plus de quatre décennies dans le domaine de la vision par ordinateur [60]. La reconnaissance faciale trouve de nombreuses applications pratiques, notamment dans la surveillance de la sécurité [61], reconnaissance des occlusions de visages [32], ainsi que dans l'identification des victimes et des personnes disparues [62].

La reconnaissance faciale, bien qu'offrant des avantages indéniables dans divers domaines, fait face à plusieurs défis majeurs. Parmi ceux-ci, la précision demeure un enjeu crucial, nécessitant une capacité à distinguer avec fiabilité les visages autorisés des intrus malgré les variations d'éclairage et d'expression. De plus, la robustesse du système est essentielle pour garantir son efficacité face aux changements d'apparence et aux tentatives de contournement, comme l'utilisation de masques [63]. La performance du système est également un défi, en particulier dans des environnements à grande échelle où la réactivité rapide est essentielle. Enfin, des préoccupations éthiques et de protection de la vie privée entourent l'utilisation de données biométriques, nécessitant des protocoles stricts pour éviter les abus potentiels tout en respectant les normes éthiques et juridiques [64].

Comment surmonter les défis techniques et éthiques de la reconnaissance faciale pour développer une authentification biométrique en ligne fiable et respectueuse de la vie privée, garantissant la sécurité des transactions en ligne ?

3.3 Choix du modèle de base

Pour déterminer le modèle approprié pour notre approche de détection faciale, nous avons choisi d'utiliser le modèle VGG en raison de plusieurs facteurs remarquables. Le modèle VGG16 est un réseau de neurones convolutifs proposé par Simonyan et Zisserman [65] de l'Université d'Oxford, remportant la première place du défi de localisation de l'ILSVRC-2014 avec une erreur de 25,3% et la deuxième place du défi de classification d'images .

L'architecture du modèle VGG16, illustrée dans la figure 3.1 [66], comporte un total de 16 couches, dont 13 sont des couches de convolution avec des filtres de convolution de 3×3 et un pas de 1. Elles utilisent toujours un rembourrage identique et une couche de max-pooling avec un filtre de 2×2 pour simplifier l'architecture tout en maintenant sa profondeur. La fonction d'activation rectifiée linéaire (ReLU) est utilisée pour sélectionner les valeurs les plus élevées dans la région de l'image [32]. En ce qui concerne les couches entièrement connectées, les deux premières couches contiennent chacune 4096 neurones, offrant une grande capacité d'apprentissage des caractéristiques abstraites. La dernière couche entièrement connectée comporte 1000 neurones, correspondant aux 1000 classes de l'ensemble de données ImageNet [67]. Elle utilise une fonction d'activation softmax pour produire une distribution de probabilité sur les classes, permettant une classification précise [68].

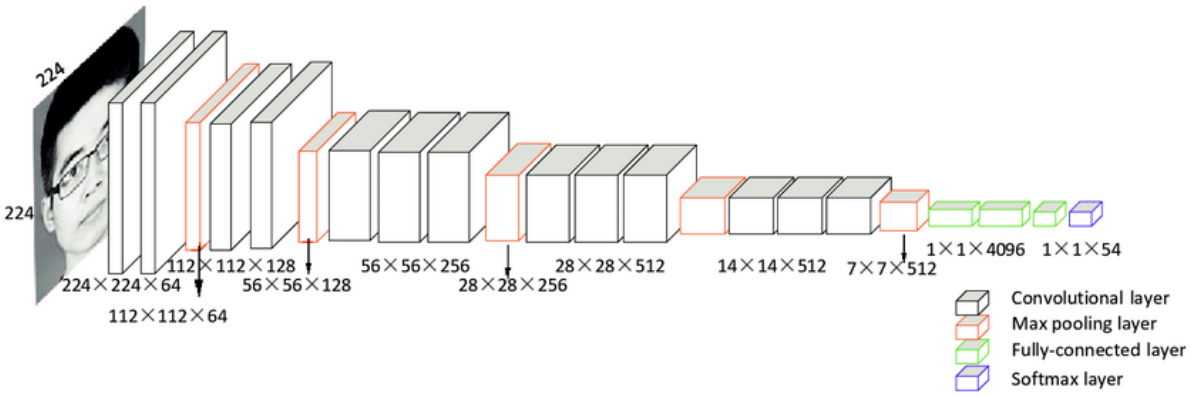


Fig. 3.1 : Architecture du modèle VGG16.

Le choix de VGG16 nous permet de tirer partie d'un modèle qui a démontré une grande précision par rapport à d'autres modèles existants tels que ResNet-50, Inception V3, ELA et CNN, comme indiqué dans l'article [69], où la précision d'entraînement a atteint 91,97% avec un nombre restreint d'époques. De plus, VGG16 est compatible avec la plupart des frameworks d'apprentissage en profondeur tels que TensorFlow, Keras et PyTorch, et grâce à son architecture modulaire, il permet des modifications et des améliorations, facilitant le développement de versions personnalisées pour de multiples applications variantes comme la détection de visages falsifiés [63], la détection des visages occultés [40] et la reconnaissance des petits visages [43].

3.4 Méthodologie

3.4.1 Modification du modèle VGG

Avant d'explorer les modifications apportées à l'architecture VGG16, il est essentiel de comprendre sa conception originale.

La modification de l'architecture VGG16 implique le remplacement de la couche de classification originale par une couche personnalisée conçue pour produire à la fois un score de classification (indiquant la présence d'un visage) et des valeurs de régression (coordonnées de la boîte englobante).Cette approche vise à améliorer les performances du modèle en augmentant considérablement la taille de l'ensemble de données d'entraînement. En fournissant une représentation plus complète et diversifiée des données, notre approche permet de détecter efficacement les visages dans des scénarios complexes.

3.4.2 Transformation à deux sorties

La modification de l'architecture VGG16 consiste au remplacement des dernières couches entièrement connectées par une couche personnalisée conçue pour produire à la fois des scores de classification et des valeurs de régression. Cette modification permet au modèle de prédire simultanément la présence d'un visage dans l'image d'entrée et d'estimer les coordonnées de la boîte englobante si un visage est détecté.

La couche personnalisée se compose de deux branches :

- **Branche de Classification :** Cette branche produit un score de classification binaire indiquant la présence ou l'absence d'un visage dans l'image d'entrée.
- **Branche de Régression :** Cette branche produit quatre valeurs représentant les coordonnées de la boîte englobante entourant le visage détecté (coordonnée x, coordonnée y, largeur et hauteur).

En incorporant ces modifications, le modèle devient capable d'effectuer à la fois la détection de visage et la régression de boîte englobante en une seule passe avant.

3.4.3 Fonctions de coût et Optimisation

Les équations mathématiques utilisées pour les pertes de classification et de régression sont les suivantes :

Classification Loss (L_C) :

$$L_C = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)], \tag{3.1}$$

où y_i représente l'étiquette réelle (0 ou 1) indiquant la présence d'un visage, et p_i représente la probabilité prédite de la présence d'un visage.

Regression Loss (L_R) :

$$L_R = \frac{1}{N}\sum_{i=1}^{N}\left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2\right]. \tag{3.2}$$

où (x_i, y_i, w_i, h_i) sont les coordonnées de la boîte englobante réelle, et $(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i)$ sont les coordonnées prédites.

Total Loss (L_{Total}) :

$$L_{Total} = \alpha L_C + \beta L_R. \tag{3.3}$$

La fonction de perte totale L_{Total} est une combinaison des pertes de classification et de régression, pondérée par des facteurs appropriés (α et β) pour équilibrer leurs contributions.

3.4.4 Optimisation utilisant Adam

L'optimiseur Adam est largement reconnu pour son efficacité dans l'entraînement des modèles d'apprentissage en profondeur [70]. Le processus d'optimisation est régi par les équations suivantes :

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1-\beta_1) g_t, \\ v_t &= \beta_2 v_{t-1} + (1-\beta_2) g_t^2, \\ \hat{m}_t &= \frac{m_t}{1-\beta_1^t}, \\ \hat{v}_t &= \frac{v_t}{1-\beta_2^t}. \\ \theta_{t+1} &= \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \end{aligned} \tag{3.4}$$

où :

- m_t et v_t sont les estimations du premier et du deuxième moment.
- β_1 and β_2 sont des hyperparamètres pour les estimations des moments.
- η est le taux d'apprentissage.
- ϵ est une petite constante pour éviter la division par zéro.
- g_t est le gradient du time step t.
- θ_t représente les paramètres du modèle.

3.4.5 Architecture du modèle proposé

L'architecture du modèle proposé est illustrée dans la figure **??**. Ce diagramme montre le flux de données à travers le réseau, mettant en évidence la couche personnalisée ajoutée qui est responsable de la production des sorties de classification et de régression. Ce changement architectural permet au modèle d'effectuer simultanément la détection des visages et la régression des boîtes englobantes.

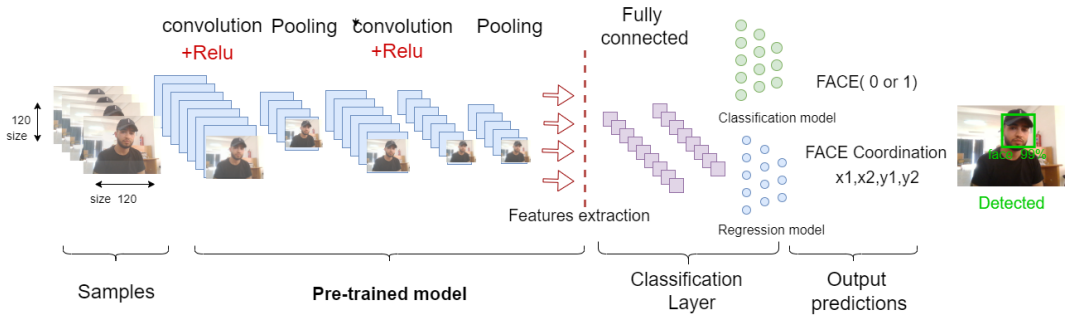


FIG. 3.2 : Modified VGG Architecture

3.4.6 Considérations éthiques et légales

Dans le cadre de notre étude, nous avons pris des mesures rigoureuses pour garantir le respect des principes éthiques et légaux concernant l'utilisation des images collectées. Tout d'abord, un processus de consentement éclairé a été établi pour les étudiants participants, où ils ont été informés de manière transparente de l'objectif de la collecte d'images et de l'utilisation de leurs données. Ce processus garantissait que les étudiants étaient pleinement conscients de leur participation et donnaient leur consentement éclairé. De plus, nous avons scrupuleusement respecté toutes les lois et réglementations pertinentes concernant l'utilisation d'images dans la recherche, garantissant la conformité aux normes de confidentialité et de protection des données. Enfin, des mesures techniques ont été mises en œuvre pour garantir la confidentialité des données des participants, telles que la sécurisation des bases de données et la restriction de l'accès aux seuls chercheurs autorisés. Ces pratiques reflètent notre engagement envers l'éthique et la légalité dans la conduite de nos recherches.

3.5 Collecte et préparation des données

En raison de la difficulté à trouver un ensemble de données existant répondant à toutes les conditions de distances faciales variables, d'éclairage, d'expressions et de positions de la tête, nous avons décidé de créer notre propre ensemble de données.

3.5.1 Processus de collecte des données

Pour collecter les images nécessaires, nous avons utilisé une webcam avec une résolution de 720p (1280x720) d'ordinateur, en exploitant la bibliothèque OpenCV en Python [71] pour la capture d'images. La collecte d'images a eu lieu à la bibliothèque de l'Université Abderrahmane Mira de Béjaïa , offrant un environnement contrôlé avec un éclairage constant et peu de distractions de fond. Cette configuration garantissait que le visage de chaque étudiant était capturé dans des conditions similaires, contribuant à l'uni²formité de l'ensemble de données. Un total de 16 étudiants se sont portés volontaires pour l'étude, donnant leur consentement explicite pour que leurs images soient utilisées à des fins de recherche. Chaque étudiant a été photographié 30 fois, ce qui a donné un ensemble complet de 480 images. La variété des images visait à couvrir un large éventail d'expressions faciales et de positions de tête pour créer un ensemble de données robuste. Les images capturées comprenaient différentes expressions faciales et positions de tête, garantissant ainsi la diversité. Les poses spécifiques comprenaient des visages entièrement visibles à la caméra, des visages partiellement obscurcis par les mains, des profils avec le visage

tourné à gauche, à droite, vers le haut et vers le bas, et des expressions allant du neutre à des émotions extrêmes telles que le sourire et la grimace. L'objectif était de simuler des scénarios réalistes où les traits du visage pourraient être partiellement cachés ou affichés sous différents angles.

Pour fournir une compréhension visuelle de la gamme diversifiée d'expressions faciales et de positions de tête capturées lors du processus de collecte de données, la section suivante présente un échantillon d'images de notre ensemble de données. Ces exemples illustrent la variété des conditions dans lesquelles les images ont été prises, garantissant la robustesse et la diversité de l'ensemble de données.

Les images de la figure 3.3 présentent différentes poses et expressions de notre ensemble de données, notamment des visages entièrement visibles, des visages partiellement obscurcis et diverses orientations de la tête. Cette représentation visuelle met en évidence notre collecte de données, visant à capturer des positions faciales réalistes et variées.

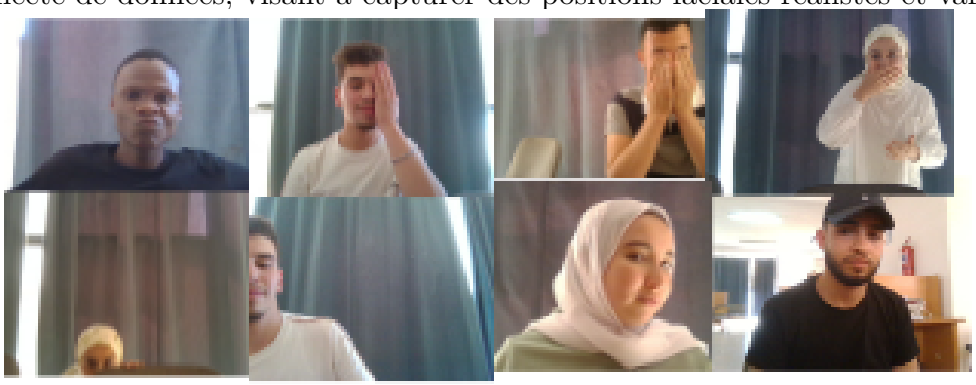


Fig. 3.3 : Exemples de l'ensemble de données construit.

3.5.2 Construction et augmentation de l'ensemble de données

Dans le cadre de notre étude, nous avons accordé une attention méticuleuse à la construction et à l'augmentation de notre ensemble de données pour garantir des résultats fiables et représentatifs. Initialement, notre ensemble de données comprenait un total de 480 images, avec 30 images collectées pour chaque étudiant participant. Cependant, nous avons rapidement réalisé que cette quantité d'images était insuffisante pour entraîner efficacement notre modèle. Pour remédier à cela, nous avons utilisé des techniques avancées d'augmentation de données en exploitant la puissante bibliothèque de traitement d'images de Python. En appliquant des filtres d'augmentation tels que la rotation, le retournement et l'ajustement des couleurs, nous avons pu multiplier notre ensemble de données initial pour obtenir une taille impressionnante de 28 800 images. Cette augmentation significative de la taille de l'ensemble de données a permis d'améliorer la robustesse et la généralisation de notre modèle, en fournissant une représentation plus complète et diversifiée des données d'entraînement. Les modèles de deep learning nécessitent en effet des ensembles de données volumineux pour être efficacement entraînés, ce qui justifie l'utilité de cette collecte étendue.

La figure 3.4 présente des exemples d'images augmentées générées à partir de l'ensemble de données initial. Chaque image représente une technique d'augmentation différente ap-

pliquée aux images d'origine, notamment la rotation, le retournement et l'ajustement des couleurs. Les images augmentées démontrent des variations d'orientation, de perspective et de couleur, étendant efficacement la diversité et la richesse de l'ensemble de données.

FIG. 3.4 : Exemples d'ensemble de données augmentées.

3.5.3 Préparation et division des données

Dans cette section, nous fournissons un compte rendu détaillé du processus méticuleux de préparation et de division des données, essentiel pour garantir la robustesse et la fiabilité de notre modèle. Initialement, chaque image a été soigneusement étiquetée et annotée à l'aide de la bibliothèque Python LabelMe, facilitant la récupération précise et l'encadrement du visage dans chaque image. Ces annotations étaient essentielles pour générer des cibles pour notre modèle, comprenant des vecteurs détaillant la présence ou l'absence d'un visage, ainsi que les coordonnées exactes des points constituant le cadre facial. La figure 3.5 présente le processus d'annotation faciale à l'aide de l'outil LabelMe, où des boîtes englobantes sont méticuleusement dessinées autour des visages sur les images. Chaque boîte englobante est associée à des annotations spécifiques détaillant les caractéristiques et les traits du visage. Cette image sert de représentation visuelle du processus d'étiquetage méticuleux utilisé pour annoter précisément l'ensemble de données.

Par la suite, des étapes de pré-traitement ont été appliquées à l'ensemble du jeu de données, comprenant le redimensionnement et la normalisation des images, afin d'assurer une cohérence et une qualité optimales. Enfin, l'ensemble de données a été divisé de manière stratégique en trois ensembles distincts : 70% pour l'entraînement, 15% pour la validation et 15% pour les tests . L'ensemble d'entraînement est utilisé pour former le modèle et constitue la majorité des données. L'ensemble de validation est utilisé pendant le processus d'entraînement pour ajuster les hyperparamètres du modèle et prévenir le surajustement. Enfin, l'ensemble de test est utilisé pour évaluer les performances du modèle sur des données non vues, fournissant une évaluation impartiale du modèle final. Le Tableau 3.1 résume le nombre d'images attribuées à chaque sous-ensemble.

Entraînement	Validation	Test
20 160	4320	4320

TAB. 3.1 : Répartition des images pour les ensembles d'entraînement, de validation et de test.

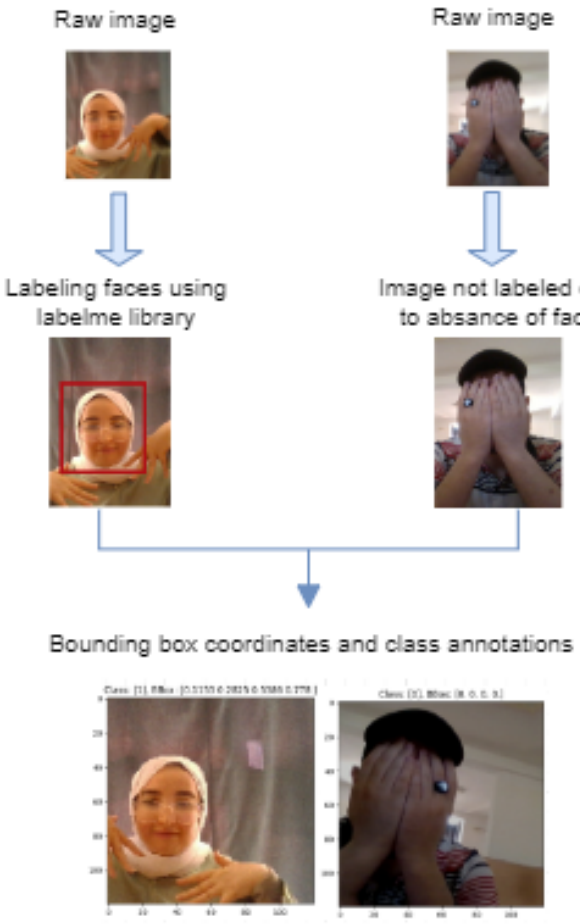


FIG. 3.5 : Processus d'annotation faciale utilisant LabelMe.

Avec les données collectées et préparées selon les procédures éthiques et techniques décrites ci-dessus, nous sommes passés à l'étape cruciale de la mise en œuvre de notre modèle.

3.6 Conclusion

Dans ce chapitre, nous avons exploré en détail notre approche méthodologique pour la détection de visages tout en justifiant minutieusement nos choix de modèles et techniques. Nous avons exposé les détails théoriques de notre système d'authentification proposé, en décrivant chaque étape en profondeur et en clarifiant les concepts essentiels qui sous-tendent notre approche.

Chapitre 4

Implémentation et Validation

4.1 Introduction

L'objectif de ce chapitre est de détailler le processus de mise en œuvre des modèles de détection de visage basés sur une version modifiée de l'architecture VGG16 et d'analyser les résultats obtenus. Nous allons commencer par une description de l'environnement de travail utilisé pour le développement et l'entraînement des modèles. Ensuite, nous présentons les résultats expérimentaux en discutant des performances des modèles modifiés par rapport aux modèles standards. Enfin, nous allons aborder les limitations de notre approche et proposerons des perspectives d'amélioration pour des travaux futurs.

4.2 Environnement de Travail

Dans cette section, nous décrivons les outils et technologies utilisés pour le développement et l'évaluation de nos modèles de détection de visage. Nous présentons également les spécifications matérielles et logicielles.

4.2.1 Matériel

Pour garantir des performances optimales lors de l'entraînement et de l'évaluation de nos modèles, nous avons utilisé Google Colab, qui offre un accès gratuit à des GPU puissants, notamment la carte graphique A100.

Google Colab : Utilisation du GPU A100 Google Colab fournit un environnement de développement basé sur le cloud avec des capacités de calcul avancées, idéal pour l'entraînement de modèles de deep learning. Le GPU A100 utilisé sur Google Colab est particulièrement performant pour accélérer les tâches de machine learning.

Fig. 4.1 : Google Colab

- **Spécifications de la GPU A100** :
  - CUDA Cores : 6,912
  - Mémoire : 40 GB HBM2
  - Performance : Jusqu'à 19.5 TFLOPS (FP32)

4.2.2 Logiciels

Nous avons utilisé plusieurs logiciels et bibliothèques pour développer, entraîner, et évaluer nos modèles de détection de visage.

Python : Version 3.11 Python est le langage de programmation principal utilisé pour ce projet, offrant une large gamme de bibliothèques pour le développement de modèles de

machine learning.

FIG. 4.2 : Python

TensorFlow : Version 2.12 TensorFlow est une bibliothèque open-source de deep learning développée par Google, utilisée pour la création et l'entraînement de modèles de machine learning.

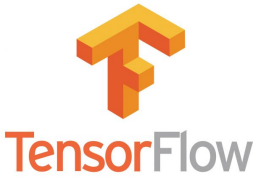


FIG. 4.3 : TensorFlow

Keras : Version 2.3 Keras est une API de haut niveau pour TensorFlow, permettant de construire et d'entraîner des modèles de manière simple et intuitive.

FIG. 4.4 : Keras

Bibliothèques et frameworks

En plus des principaux logiciels, nous avons utilisé des bibliothèques supplémentaires pour l'évaluation des modèles et la visualisation des résultats.

Scikit-learn Utilisé pour l'évaluation des métriques, Scikit-learn fournit des outils simples et efficaces pour l'analyse des données et le machine learning.

FIG. 4.5 : Scikit-learn

Matplotlib Utilisé pour la visualisation des données, Matplotlib est une bibliothèque de visualisation de données en 2D pour Python, permettant de créer des graphiques et des plots interactifs.

Fig. 4.6 : Matplotlib

NumPy Utilisé pour les opérations numériques, NumPy est une bibliothèque fondamentale pour le calcul scientifique avec Python, offrant un support pour les tableaux et les matrices multidimensionnels.

Fig. 4.7 : NumPy

OpenCV Utilisé pour le traitement d'images, OpenCV (Open Source Computer Vision Library) est une bibliothèque de vision par ordinateur qui permet de traiter et d'analyser des images de manière efficace.

Fig. 4.8 : OpenCV

4.3 Description du Modèle et des Couches

Dans cette section, nous allons détailler l'architecture du modèle de détection de visage basé sur VGG16 modifié, en expliquant les différentes couches et leur fonction. Nous incluons également une visualisation de la structure du modèle pour aider à mieux comprendre la disposition et les connexions entre les couches.

Notre modèle de détection de visage est basé sur une architecture VGG16 modifiée, composée de plusieurs couches de traitement. La Table 4.1 ci-dessous présente un résumé de la structure du modèle.

Couche d'entrée (input_2) :

- **Forme de sortie :** (None, 120, 120, 3)
- **Description :** La couche d'entrée accepte des images de taille 120x120 avec 3 canaux de couleur (RGB).

Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	(None, 120, 120, 3)	0	[]
vgg16 (Functional)	(None, None, None, 512)	14714688	['input_2[0][0]']
global_max_pooling2d	(None, 512)	0	['vgg16[0][0]']
global_max_pooling2d_1	(None, 512)	0	['vgg16[0][0]']
dense	(None, 2048)	1050624	['global_max_pooling2d[0][0]']
dense_2	(None, 2048)	1050624	['global_max_pooling2d_1[0][0]']
dense_1	(None, 1)	2049	['dense[0][0]']
dense_3	(None, 4)	8196	['dense_2[0][0]']
Total params : 16826181 (64.19 MB)			
Trainable params : 16826181 (64.19 MB)			
Non-trainable params : 0 (0.00 Byte)			

Tab. 4.1 : Architecture du modèle VGG modifié

VGG16 (vgg16) :

- **Forme de sortie :** (None, None, None, 512)
- **Description :** Cette couche représente le modèle pré-entraîné VGG16, coupé après la dernière couche de convolution. Elle est responsable de l'extraction des caractéristiques importantes des images.

Global Max Pooling (global_max_pooling2d et global_max_pooling2d_1) :

- **Forme de sortie :** (None, 512)
- **Description :** Ces couches effectuent une réduction des dimensions spatiales en prenant le maximum sur chaque canal, réduisant ainsi la complexité tout en conservant les caractéristiques les plus importantes.

Couches denses (dense et dense_2) :

- **Forme de sortie :** (None, 2048)
- **Description :** Ces couches entièrement connectées transforment les caractéristiques extraites par le modèle VGG16 en un format approprié pour la détection de visage. Elles sont suivies par des fonctions d'activation pour ajouter de la non-linéarité au modèle.

Couches de sortie (dense_1 et dense_3) :

- **Forme de sortie :**
  - **dense_1 :** (None, 1)
  - **dense_3 :** (None, 4)
- **Description :** Ces couches génèrent les sorties finales du modèle :
  - **dense_1** produit une seule valeur indiquant la probabilité de la présence d'un visage.
  - **dense_3** produit quatre valeurs représentant les coordonnées de la boîte englobante pour localiser le visage dans l'image.

4.4 Entraînement du Modèle

Dans cette section, nous décrivons en détail le processus d'entraînement du modèle de détection de visage basé sur VGG16 modifié. Nous abordons les paramètres d'entraînement utilisés, les méthodes de régularisation appliquées, et l'algorithme d'optimisation choisi. Le modèle a été entraîné sur un ensemble de données construit à partir d'images contenant des visages. Les images ont été prétraitées pour normaliser les valeurs des pixels et ont été redimensionnées à une taille de 120x120 pixels pour être compatibles avec l'entrée du modèle.

4.4.1 Paramètres d'entraînement

Le processus d'entraînement de nos modèles de détection de visage a été effectué sur Google Colab, en utilisant des GPU A100 pour accélérer les calculs. Nous avons suivi une approche systématique pour assurer un entraînement efficace et éviter le surapprentissage. Les paramètres d'entraînement suivants ont été utilisés pour entraîner nos modèles :

- **Taux d'apprentissage** : 0.001, un taux d'apprentissage approprié est crucial pour garantir la convergence du modèle sans oscillations excessives.
- **Nombre d'époques** : 6, nous avons choisi d'entraîner nos modèles pendant 6 époques, ce qui a montré un bon équilibre entre la performance et le temps de calcul.
- **Taille de batch (batch size)** : 32, cette taille de batch permet de tirer parti des capacités de calcul du GPU tout en assurant une bonne généralisation du modèle.

Ces paramètres ont été choisis après plusieurs expérimentations pour trouver la configuration optimale qui permet au modèle de converger tout en évitant le surapprentissage.

4.4.2 Méthodes de régularisation

Pour améliorer la généralisation du modèle et réduire le surapprentissage, nous avons utilisé les méthodes de régularisation suivantes :

- **Dropout** : Une probabilité de dropout de 0.5 a été appliquée aux couches denses pour prévenir le surapprentissage en désactivant aléatoirement certains neurones durant l'entraînement.
- **Data Augmentation** : Des techniques d'augmentation de données, telles que les rotations, les translations, et les inversions horizontales, ont été appliquées aux images d'entraînement pour augmenter la diversité des données et améliorer la robustesse du modèle.

4.4.3 Algorithme d'optimisation

Nous avons utilisé l'algorithme d'optimisation **Adam** (Adaptive Moment Estimation) pour entraîner notre modèle. Adam combine les avantages des méthodes AdaGrad et RMSProp, et il est bien adapté pour les problèmes de classification et de régression avec de grands ensembles de données.

4.5 Résultats et Évaluation

Dans cette section, nous présentons les résultats obtenus après l'entraînement du modèle de détection de visage. Nous discutons les performances du modèle en utilisant divers métriques d'évaluation, y compris la précision, le rappel, et la perte. Nous comparons également les performances de notre modèle modifié VGG16 avec celles du modèle standard VGG16. Les résultats sont analysés en détail pour évaluer l'efficacité de notre approche et identifier les domaines potentiels d'amélioration.

4.5.1 Description des Métriques d'Évaluation

Pour évaluer les performances de notre modèle, nous avons utilisé plusieurs métriques couramment employées dans l'évaluation des modèles machine learning et en particulier dans la classification d'objets.

Exactitude (Accuracy)

L'exactitude, ou accuracy, est une métrique de classification qui mesure le pourcentage de prédictions correctes par rapport au nombre total de prédictions. Elle est définie par la formule (4.1) :

$$\text{Accuracy} = \frac{\text{Nombre de Prédictions Correctes}}{\text{Nombre Total de Prédictions}} \tag{4.1}$$

Cette métrique est particulièrement utile pour évaluer la capacité du modèle à distinguer entre les images contenant des visages et celles n'en contenant pas.

Précision (Precision)

La précision est la proportion des prédictions positives correctes parmi l'ensemble des prédictions positives faites par le modèle. Elle est donnée par la formule (4.2) :

$$\text{Précision} = \frac{\text{VP}}{\text{VP} + \text{FP}} \tag{4.2}$$

où VP est le nombre de vrais positifs et FP est le nombre de faux positifs. Une précision élevée indique que le modèle fait peu de fausses alertes.

Rappel (Recall)

Le rappel, ou sensibilité, mesure la capacité du modèle à identifier correctement toutes les instances positives. Il est défini par (4.3) :

$$\text{Rappel} = \frac{\text{VP}}{\text{VP} + \text{FN}} \tag{4.3}$$

où FN est le nombre de faux négatifs. Un rappel élevé indique que le modèle détecte la plupart des instances positives.

Courbe ROC (Receiver Operating Characteristic)

La courbe ROC est une représentation graphique de la performance d'un modèle de classification à différents seuils. Elle trace le taux de vrais positifs (rappel) en fonction du taux de faux positifs (1 - spécificité) pour différents seuils de décision.

La métrique associée à la courbe ROC est l'AUC-ROC (Area Under the Curve - Receiver Operating Characteristic), qui mesure la capacité du modèle à discriminer entre les classes positives et négatives. Une AUC-ROC élevée (proche de 1) indique une meilleure performance du modèle.

4.5.2 Résultats d'Entraînement et de Validation

Dans cette section, nous présentons les résultats obtenus après l'entraînement du modèle de détection de visage. Nous discutons les performances du modèle en utilisant les métriques d'évaluation précédentes et la perte. Nous comparons également les performances de notre modèle modifié VGG16 avec celles du modèle standard VGG16. Les résultats sont analysés en détail pour évaluer l'efficacité de notre approche et identifier les domaines potentiels d'amélioration.

L'évolution de la fonction de perte au cours des époques pendant l'entraînement et la validation est illustrée à la Fig. 4.9. Les courbes fournissent des informations sur la convergence et les performances du modèle pour les pertes de classification et de régression.

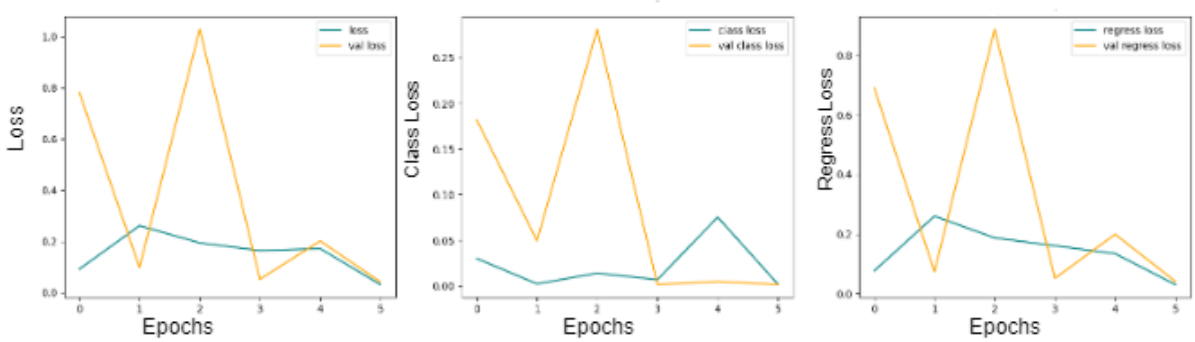


Fig. 4.9 : Évolution de la perte pendant l'entraînement et la validation

Les performances obtenues par le modèle VGG modifié en termes de précision, rappel et exactitude sont présentées dans le Tableau 4.2. Ces valeurs représentent la moyenne sur plusieurs itérations pour les ensembles de validation et d'entraînement. L'exactitude atteinte de 99,95 % démontre l'efficacité de l'approche proposée pour détecter précisément les visages dans diverses conditions.

Métriques	Entraînement	Validation
Précision	0.9997	1.0
Rappel	0.9997	1.0
Exactitude	0.9995	1.0

Tab. 4.2 : Résultats de validation et d'entraînement retournés par les métriques.

De plus, la Fig. 4.10 montre l'évolution de l'exactitude du modèle sur les ensembles d'entraînement et de validation au cours des époques.

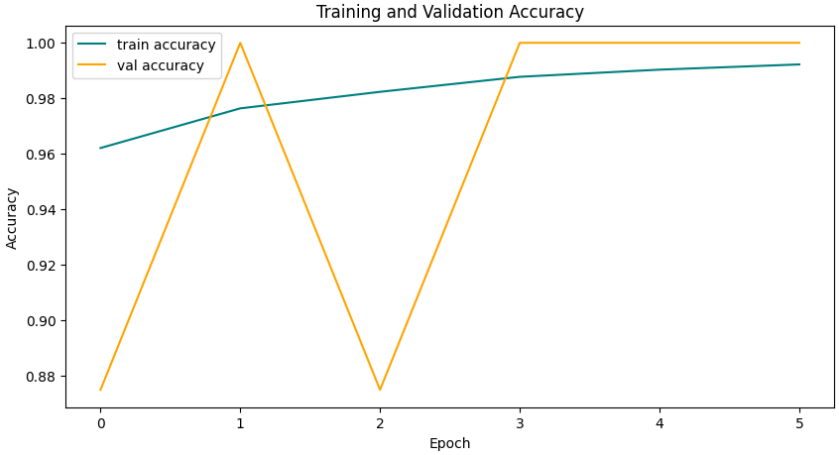


FIG. 4.10 : Évolution de l'exactitude pendant l'entraînement et la validation

4.5.3 Résultats et Visualisation sur le Jeu de Test

Les résultats présentés dans le Tableau 4.3 montrent la performance de notre modèle à travers différentes métriques sur le jeu de test, indiquant l'efficacité du modèle.

Métrique	Précision	Rappel	Exactitude
Jeu de Test	100%	100%	100%

TAB. 4.3 : Les performance obtenus pour sur le jeu de test.

De plus, la Fig. 4.11 illustre les performances du modèle sur des images d'échantillon du jeu de test, où les boîtes englobantes sont correctement placées sur les visages. Le modèle a montré des capacités de détection robustes, identifiant et plaçant avec précision les boîtes englobantes autour des visages dans les images de test.

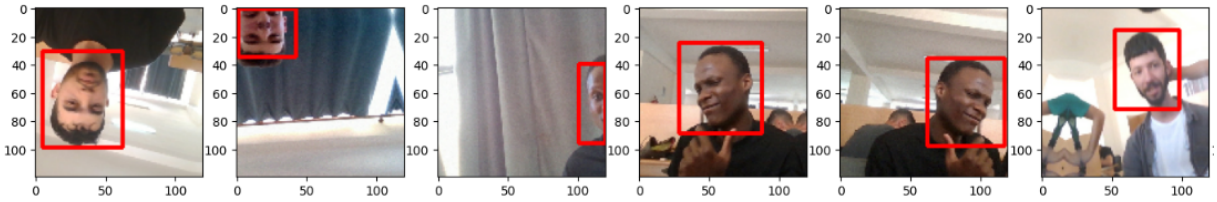


FIG. 4.11 : Détections d'échantillons du jeu de test avec des boîtes englobantes placées avec précision sur les visages

Pour analyser plus en détail les performances de notre modèle, nous avons visualisé (voir Fig. 4.12) les courbes de perte pour les tâches de classification et de régression afin d'identifier d'éventuels problèmes tels que le surapprentissage. En effet, les courbes de perte montrent des processus d'entraînement et de validation cohérents. Nous pouvons confirmer la robustesse et la fiabilité du modèle dans les tâches de détection de visage. Les boîtes englobantes sont placées avec précision, et les métriques de performance valident encore l'efficacité du modèle.

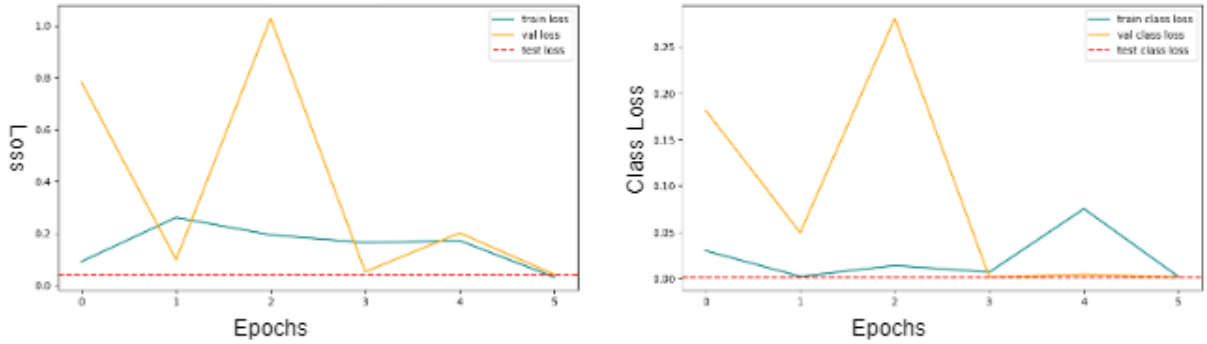


FIG. 4.12 : Courbes de perte de classification et de régression sur le jeu de test

4.5.4 Comparaison avec le Modèle VGG16 Standard

Pour valider l'efficacité de notre modèle de détection de visage, nous l'avons comparé au modèle VGG16 standard.

Le Tableau 4.4 présente les résultats obtenus par le modèle VGG16 modifié et le modèle VGG16 standard après un entraînement de 6 époques sur notre ensemble de données.

Modèle	Perte Totale	Rappel	Exactitude	Précision
VGG16 Standard	0.1787	0.9922	0.9899	0.9959
VGG16 Modifié	0.0905	0.9938	0.9926	0.9976

TAB. 4.4 : Résultats des deux modèles sur notre ensemble de données

Nous avons également comparé les matrices de confusion du modèle VGG16 standard et du modèle VGG16 modifié. Les résultats montrent que le modèle modifié réduit significativement le nombre de faux positifs et de faux négatifs par rapport au modèle VGG16 standard, comme le montre la Fig. 4.13.

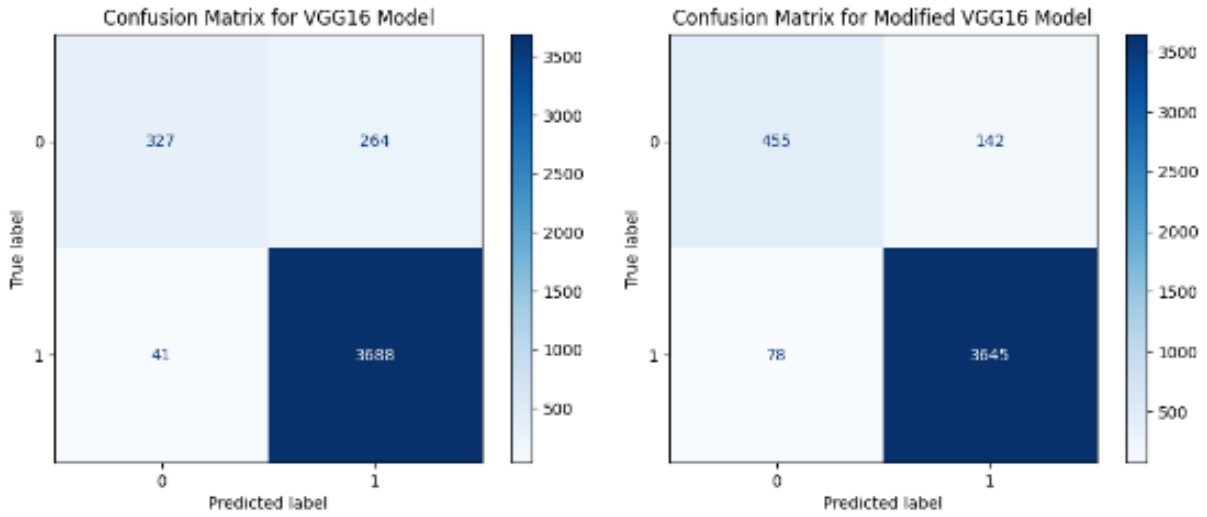


FIG. 4.13 : Comparaison des matrices de confusion

Métrique ROC

La courbe ROC et l'AUC pour les modèles standard VGG16 et modifié VGG16 sont présentées ci-dessous. Une valeur AUC plus élevée indique une meilleure performance du modèle en termes de distinction entre les classes positives et négatives., comme le montre

la Fig. 4.14.

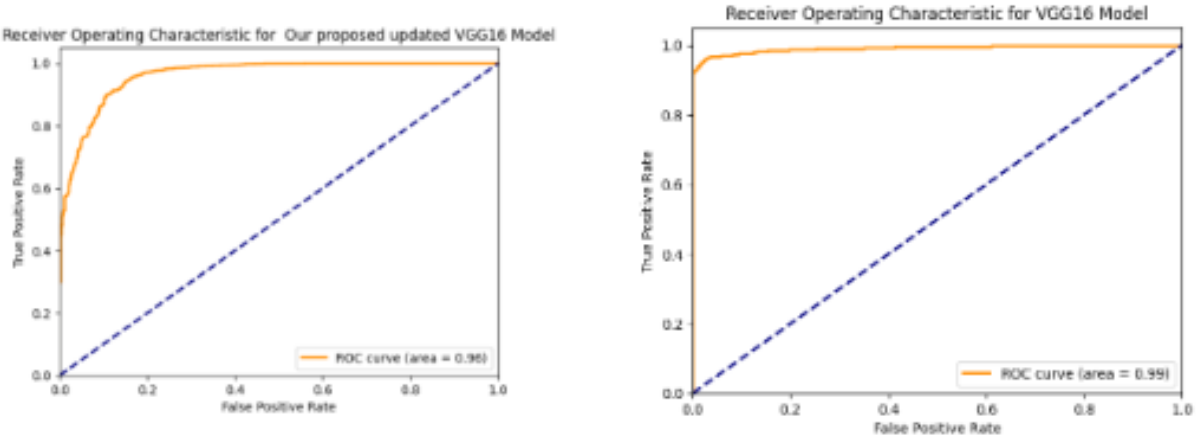


FIG. 4.14 : Courbes ROC pour les Modèles VGG16 Standard et Modifié

4.6 Évaluation des modèles avec un Jeu de Données Externe

Pour assurer la robustesse de notre modèle et éviter le surapprentissage, nous avons testé à la fois les modèles VGG16 standard et VGG16 modifié en utilisant un ensemble de données externe contenant 670 images [72].

4.6.1 Préparation de l'Ensemble de Données de Test

Cette section décrit en détail le processus de préparation de cet ensemble de test.

Redimensionnement des Images

Toutes les images de l'ensemble de données supplémentaires ont été redimensionnées à une taille de 120x120 pixels. Cette étape est cruciale pour assurer que les images soient compatibles avec l'architecture d'entrée de notre modèle. Le redimensionnement a été effectué en utilisant des bibliothèques telles que OpenCV et PIL, qui permettent un ajustement précis des dimensions tout en maintenant la qualité des images.

Ajout de Labels aux Images

Chaque image a été annotée avec des labels indiquant la présence ou l'absence de visages, ainsi que les coordonnées des boîtes englobantes pour les visages détectés. Ces annotations ont été ajoutées en utilisant des bibliothèques Labelme, qui facilite le processus de labellisation manuelle en fournissant une interface utilisateur intuitive.

Construction du Dataset Final

Après le redimensionnement et l'annotation des images, nous avons construit le dataset final en combinant les nouvelles images avec celles déjà présentes dans notre ensemble de données de test. Ce processus a impliqué la concaténation des fichiers d'annotations et des images, ainsi que la vérification de l'intégrité des données pour s'assurer que chaque image avait bien ses annotations correspondantes.

4.6.2 Résultats de l'Évaluation sur le Jeu de Données Externe

Après la préparation de l'ensemble de données, nous avons appliqué les deux modèles (VGG16 standard et VGG16 modifié) pour évaluer leur performance.Les performances des modèles ont été évaluées en utilisant plusieurs métriques, y compris la précision, le rappel, et l'exactitude. Ces métriques ont permis de comparer l'efficacité des deux modèles pour la détection de visages dans un contexte externe. Le Tableau 4.5 résume les résultats obtenus pour cet ensemble de données.

Modèle	**Perte Totale**	**Précision**	**Rappel**	**Exactitude**
VGG16 Standard	0.6623	0.9398	0.9861	0.9289
VGG16	0.0161	0.99	0.9901	0.9888

Tab. 4.5 : Résultats des deux modèles sur l'ensemble de données externe

Un autre facteur important pour les applications en temps réel est le temps de prédiction par image. Notre modèle VGG16 modifié a démontré des performances supérieures, avec un temps de prédiction moyen par image de 0.02 s comparé à 0.04 s pour le modèle VGG16 standard comme on peut le voir dans le Tableau 4.6). Les résultats obtenus ont

Modèle	Temps de Test (s)	Temps de Prédiction par Image (s)
VGG16 Standard	22	0.04
VGG16 Modifié	21	0.02

Tab. 4.6 : Performances temporelles des modèles de détection de visage

été analysés pour identifier les points forts et les limitations de chaque modèle. Cette analyse a également permis de déterminer si les modèles sont adaptés à des applications réelles de détection de visages.

4.7 Conclusion

Ce chapitre a présenté le processus suivi pour évaluer et valider l'approche de détection de visage proposée. Après présentation de l'environnement du développement et la description des différentes couches du modèle VGG16 modifié et implémenté, nous avons procédé un ensemble de tests sur le dataset que nous avons construit et un dataset disponible publiquement pour valider notre approche.

En résumé, les résultats obtenus démontrent l'efficacité de notre modèle modifié VGG16 pour la détection de visage, surpassant le modèle standard avec plusieurs métriques à savoir la précision, le rappel et l'exactitude, ainsi que le temps de prédiction. Ces performances mettent en évidence les améliorations apportées par notre approche et souligne son potentiel pour des applications temps réel de détection de visage.

Conclusion générale

Cette étude vise à mettre en œuvre un système d'authentification biométrique en ligne basé sur le deep learning, se basant sur le modèle de classification d'images VGG16 et en appliquant des techniques d'augmentation de données.

Tout avons d'abord, nous avons entamé notre travail en explorant les généralités sur la biométrie, traits biométriques, les systèmes biométriques et ses mesures de performances. L'architecture globale d'un système biométrique est composée de plusieurs modules : module d'acquisition, de prétraitement, d'extraction des caractéristiques, de mise en correspondance, et module de décision. Ensuite, nous avons fait une étude des modèle du deep learning utilisés dans les systèmes biométriques .Nous avons également distingué les types de systèmes biométriques en fonction de leur modalité, comme les systèmes unimodaux et multimodaux.

Puisque certains systèmes sont plus performants que d'autres et sont variés selon leur type de modalité, nous avons établi un état de l'art critique sur le domaine de l'authentification biométrique. Nous avons constaté que la modalité du visage est particulièrement efficace, car elle est accessible via des dispositifs courants et offre une grande précision. En outre, le modèle VGG16 a été utilisé pour sa capacité à extraire des caractéristiques pertinentes des images faciales tout en améliorant ainsi la performance et la robustesse du système d'authentification.

Dans l'implémentation de notre système, diverses configurations ont été testées lors des différentes étapes du processus d'authentification, en particulier concernant les techniques d'augmentation des données et le traitement des images.Pour la phase d'expérimentation et comparaison avec le modèle originale nous avons utilisé plusieurs mesure de performances qui ont montré que le sytstème proposé améliore significativement la précision de la reconnaissance faciale et sa précision, réduisait le temps de traitement global et augmentait la robustesse face à diverses conditions environnementales.

En analysant de manière critique notre travail ,le jeux de donnée augmenté est encore réduit par rapport aux grandes bases de données utilisées dans le deep learning alos comme future perspective nous prévoyons d'augmenter la diversité du jeu de données en appliquant des techniques de réseaux antagonistes génératifs. De plus, nous visons à optimiser les hyperparamètres grâce à une recherche approfondie, et à valider l'efficacité et la fiabilité du modèle en le testant dans des applications réelles, telles que les systèmes de sécurité. Parvenir à faire passer notre système à détecteur de multiples visages serait une bonne perspective ;

Bibliographie

[1] LegalProd. *Authentification Numérique*. 2024. URL : https://www.legalprod.com/authentification-numerique/#:~:text=R%C3%B4le%20crucial%20dans%20la%20cybers%C3%A9curit%C3%A9,prot%C3%A9geant%20ainsi%20leurs%20ressources%20critiques..

[2] Syloe. *Glossaire : Authentification*. 2024. URL : https://www.syloe.com/glossaire/authentification/#:~:text=L'authentification%20est%20une%20proc%C3%A9dure,acc%C3%A9der%20%C3%A0%20certaines%20ressources%20s%C3%A9curis%C3%A9es..

[3] *La biométrie au service de l'identification (2021)*. fr. Juin 2023. URL : https://www.thalesgroup.com/fr/europe/france/dis/gouvernement/inspiration/biometrie (visité le 31/03/2024).

[4] Boussad Faouzi Boussa Rahim Ryad. "Développement d'un système biométrique multimodal basé sur la fusion des scores de matching". Thèse de doct.

[5] S. Saousse et K. Mayssa. *Combinaison de multiple classifieurs pour la reconnaissance biométriques des personnes*. [Online ; accessed 20-May-2024]. 2016. URL : http://dspace.univ-tebessa.dz:8080/xmlui/bitstream/handle/123456789/655/memoire.pdf?sequence=1&isAllowed=y.

[6] *La reconnaissance des empreintes digitales expliquée*. fr-FR. URL : https://www.cdvi.ch/fr/la-biometrie-expliquee/la-reconnaissance-des-empreintes-digitales-expliquee/ (visité le 13/04/2024).

[7] *Biometrie - Biometrics - Empreintes digitales*. URL : https://www.biometrie-online.net/technologies/empreintes-digitales (visité le 13/04/2024).

[8] LAADJAL Mohammed et al. "L'identification Biométrique Par Les Veines Des Doigts". fr. In : ().

[9] *Blog DUT Génie Biologique Créteil : La biométrie : un nouveau moyen de sécuriser et rendre plus rapide nos paiements ?* URL : https://depgbcreteil.blogspot.com/2016/06/la-biometrie-un-nouveau-moyen-de.html (visité le 13/04/2024).

[10] *Biometrie - Biometrics - Iris*. URL : https://www.biometrie-online.net/technologies/iris (visité le 13/04/2024).

[11] Zaheera Zainal Abidin, Mazani Manaf et Abdul Samad Shibghatullah. "ant-CBIR : A New Method for Radial Furrow Extraction in Iris Biometric". en. In : (2014).

[12] Takwa CHIHAOUI. "Système d'identification de personnes basé sur la rétine". Theses. Université Paris-Est ; Université de Tunis El Manar, déc. 2018.

[13] *Expertise biometrique de signatures*. URL : https://criminalistique.fr/services/expertise-biometrique-signatures.html (visité le 13/04/2024).

[14] *Gait recognition technology can detect you from 50m distance without seeing your face*. URL : https://news.cgtn.com/news/7767544d7a597a6333566d54/share_p.html.

[15] Abdelkrim MOUSSAOUI et al. "Présentée par : Amir BENZAOUI". fr. In : ().

[16] TIDJANI ZAKARIA, CHAA MORAD et NASRI NAJIB. "Identification des personnes par système multimodale". fr. In : ().

[17] Rahim Ryad BOUSSA et Faouzi BOUSSAD. *Boussa R.; Boussad F..pdf*.

[18] Hawraa Abed AL-KAREEM HUSSAIN et Hawraa Hassan ABBAS. "A Survey on Multi-biometric Fusion Approaches". In : *Kerbala Journal for Engineering Science* 03.02 (2023), p. 1-22.

[19] Khaled MAHDADI et El-Moundher HADJAIDJI. "Modélisation d'empreinte biométrique par un modèle flou de Sugeno optimisé". Mémoire de master. Université Kasdi Merbah-Ouargla, Mai 2017.

[20] Elhocine BOUTTLLAA. "Système biométrique de vérification de signatures manuscrites en ligne". Mémoire de magister. Ecole nationale Supérieure d'Informatique (E.S.I), mar. 2019.

[21] Ahmed HADJAR. "Identification des individus par la biométrie multimodale". Mémoire de magister. Université Des Sciences Et De La Technologie D'Oran Mohamed Boudiaf, Nouvembre 2014.

[22] Bilal ATTALLAH. "Conception d'un système de reconnaissance des empreintes digitales par apprentissage". Mémoire de magister. Université Des Sciences Et De La Technologie D'Oran Mohammed Boudiaf, 2012.

[23] GOOGLE DEVELOPERS. *Classification : Accuracy*. https://developers.google.com/machine-learning/crash-course/classification/accuracy.

[24] IBM DATA AND AI TEAM. *What are Neural Networks ?* 2024. URL : https://www.ibm.com/fr-fr/topics/neural-networks.

[25] Jérémy ROBERT. *Convolutional Neural Network : Tout ce qu'il y a à savoir*. Formation Data Science | DataScientest.com. 25 juin 2020.

[26] MUHAMMADRIZWAN. *LeNet-5 - A Classic CNN Architecture - DataScienceCentral.com*. Data Science Central. 16 oct. 2018.

[27] Chen YANHUI. *From AlexNet to NASNet : A Brief History and Introduction of Convolutional Neural Networks*. Medium. 26 fév. 2021.

[28] Raphael KASSEL. *VGG : en quoi consiste ce modèle ? Daniel vous dit tout !* Formation Data Science | DataScientest.com. 27 avr. 2021.

[29] *Figure 8. GoogleNet-like architecture.* ResearchGate.

[30] Chamandeep Vimal et Neeraj Shirivastava. “Face and face-mask detection system using vgg-16 architecture based on convolutional neural network”. In : *International Journal of Computer Applications* 183.50 (2022), p. 16-21.

[31] Dewan Ahmed Muhtasim, Monirul Islam Pavel et Siok Yee Tan. “A patch-based CNN built on the VGG-16 architecture for real-time facial liveness detection”. In : *Sustainability* 14.16 (2022), p. 10024.

[32] Dhara Mungra et al. “PRATIT : A CNN-Based Emotion Recognition System using Histogram Equalization and Data Augmentation”. In : *Multimedia Tools and Applications* (jan. 2020).

[33] Saibal Manna, Sushil Ghildiyal et Kishankumar Bhimani. “Face Recognition from Video using Deep Learning”. In : *2020 5th International Conference on Communication and Electronics Systems (ICCES)*. 2020, p. 1101-1106.

[34] Weidong Kuang et Abhijit Baul. “A Real-time Attendance System Using Deep-learning Face Recognition”. In : juin 2020.

[35] Thai-Viet Dang et Hoai-Linh Tran. “A Secured, Multilevel Face Recognition based on Head Pose Estimation, MTCNN and FaceNet”. In : *Journal of Robotics and Control (JRC)* 4.4 (2023), p. 431-437.

[36] Shizhen Huang et Haonan Luo. “Attendance System Based on Dynamic Face Recognition”. In : *2020 International Conference on Communications, Information System and Computer Engineering (CISCE)*. 2020, p. 368-371.

[37] Siru Chen. “CNN combined with data augmentation for face recognition on small dataset”. In : *Journal of Physics : Conference Series* 2634.1 (nov. 2023), p. 012040.

[38] Florian Schroff, Dmitry Kalenichenko et James Philbin. “FaceNet : A Unified Embedding for Face Recognition and Clustering”. In : *CoRR* abs/1503.03832 (2015).

[39] Hasin Shad et al. “Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network”. In : *Computational Intelligence and Neuroscience* 2021 (déc. 2021), p. 1-18.

[40] Yueying Li. “Face Detection Algorithm Based on Double-Channel CNN with Occlusion Perceptron”. In : *Computational Intelligence and Neuroscience* 2022 (jan. 2022), p. 1-10.

[41] Anugrah Bintang Perdana et Adhi Prahara. “Face Recognition Using Light-Convolutional Neural Networks Based On Modified Vgg16 Model”. In : *2019 International Conference of Computer Science and Information Technology (ICoSNIKOM)*. 2019, p. 1-4.

[42] Yang Zhiqi. “Face recognition based on improved VGGNET convolutional neural network”. In : *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. 2021, p. 2530-2533.

[43] Zhishuai Zhang et al. “Robust Face Detection via Learning Small Faces on Hard Images”. In : *CoRR* abs/1811.11662 (2018).

[44] Mahyar Najibi et al. “SSH : Single Stage Headless Face Detector”. In : *CoRR* abs/1708.03979 (2017).

[45] Qurat-ul-ain et al. “Forged Face Detection using ELA and Deep Learning Techniques”. In : *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*. 2021, p. 271-275.

[46] CIPLab. *Computational Intelligence Photography Lab de l’Université Yonsei Real and Fake Face Detection*. https://www.kaggle.com/datasets/ciplab/real-and-fake-face-detection. Retrieved 2024-06-21. Mai 2022.

[47] B. Anil Kumar et Mohan Bansal. “Face Mask Detection on Photo and Real-Time Video Images Using Caffe-MobileNetV2 Transfer Learning”. In : *Applied Sciences* 13.2 (2023).

[48] Hasin Shahed Shad et al. “Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network”. In : *Computational intelligence and neuroscience* 2021.1 (2021), p. 3111676.

[49] T Gwyn, K Roy et M Atay. *Face recognition using popular deep net architectures : a brief comparative study. Futur Internet 13 (7) : 1–15*. 2021.

[50] Redouane Khemmar et al. “Biometrie authentication platform-based multisensor fusion”. In : *2017 Seventh International Conference on Emerging Security Technologies (EST)*. 2017, p. 191-195.

[51] Hawraa A. Hussain et Hawraa H. Abbas. “Biometric Fusion Approaches Based on Deep Convolutional Neural Network”. In : *Al-Sadiq International Conference on Communication and Information Technology-2023 (AICCIT-2023)* (2023), p. 85.

[52] Alin Majed Alkadi et al. “Biometric Authentication Based on Multi-Modal Facial Recognition Using Machine Learning”. In : *2023 Advances in Science and Engineering Technology International Conferences (ASET)*. 2023 Advances in Science and Engineering Technology International Conferences (ASET). Dubai, United Arab Emirates : IEEE, 20 fév. 2023, p. 1-6.

[53] Usman Cheema et Seungbin Moon. “Sejong face database : A multi-modal disguise face database”. In : *Computer Vision and Image Understanding* 208-209 (2021), p. 103218.

[54] P.P. Sarangi et al. “A feature-level fusion based improved multimodal biometric recognition system using ear and profile face”. In : *J Ambient Intell Human Comput* 13 (2022), p. 1867-1898.

[55] Bayan Omar et al. “New feature-level algorithm for a face-fingerprint integral multibiometrics identification system”. In : *UHD Journal of Science and Technology* 6.1 (2022), p. 12-20.

[56] K. Gunasekaran, J. Raja et R. Pitchai. “Deep multimodal biometric recognition using contourlet derivative weighted rank fusion with human face, fingerprint and iris images”. en. In : *Automatika* 60.3 (juil. 2019), p. 253-265.

[57] Sumegh Tharewal et al. “Score-Level Fusion of 3D Face and 3D Ear for Multimodal Biometric Human Recognition”. en. In : *Computational Intelligence and Neuroscience* 2022 (avr. 2022). Sous la dir. de Ziya Uddin, p. 1-9.

[58] Abdessalam Hattab et Ali Behloul. “Face-Iris multimodal biometric recognition system based on deep learning”. en. In : *Multimed Tools Appl* 83.14 (oct. 2023), p. 43349-43376.

[59] Statista. *Nombre d’utilisateurs d’Internet et des réseaux sociaux dans le monde en janvier 2024*. 2024. url : https://fr.statista.com/statistiques/1350675/nombre-utilisateurs-internet-reseaux-sociaux-monde/.

[60] Paramjit Kaur et al. “Facial-recognition algorithms : A literature review”. In : *Medicine, science, and the law* (jan. 2020).

[61] Umair Butt et al. “Detecting Video Surveillance Using VGG19 Convolutional Neural Networks”. In : *International Journal of Advanced Computer Science and Applications* 11 (jan. 2020).

[62] P.D N Harsha Sai et al. “Identification of Missing Person Using Convolutional Neural Networks”. In : *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*. 2022, p. 485-489.

[63] Neeraj Shirivastava Chamandeep Vimal. “Face and Face-mask Detection System using VGG-16 Architecture based on Convolutional Neural Network”. In : *International Journal of Computer Applications* 183.50 (fév. 2022), p. 16-21.

[64] Denise R. S. Almeida, Konstantin Shmarko et Elizabeth Lomas. “The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence : a comparative analysis of US, EU, and UK regulatory frameworks”. In : *AI and Ethics* 2 (août 2022).

[65] Karen Simonyan et Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In : *International Conference on Learning Representations*. 2015.

[66] Zhao Pei et al. “Face Recognition via Deep Learning Using Data Augmentation Based on Orthogonal Experiments”. In : *Electronics* 8 (sept. 2019), p. 1088.

[67] Jia Deng et al. “ImageNet : A large-scale hierarchical image database”. In : *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2009, p. 248-255.

[68] S Bindushree et N RakshithaA. “Face Recognition Using Deep Learning”. In : 2020.

[69] Qurat-ul-ain et al. “Forged Face Detection using ELA and Deep Learning Techniques”. In : *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)* (2021), p. 271-275.

[70] Diederik P Kingma et Jimmy Ba. “Adam : A method for stochastic optimization”. In : *arXiv preprint arXiv :1412.6980* (2014).

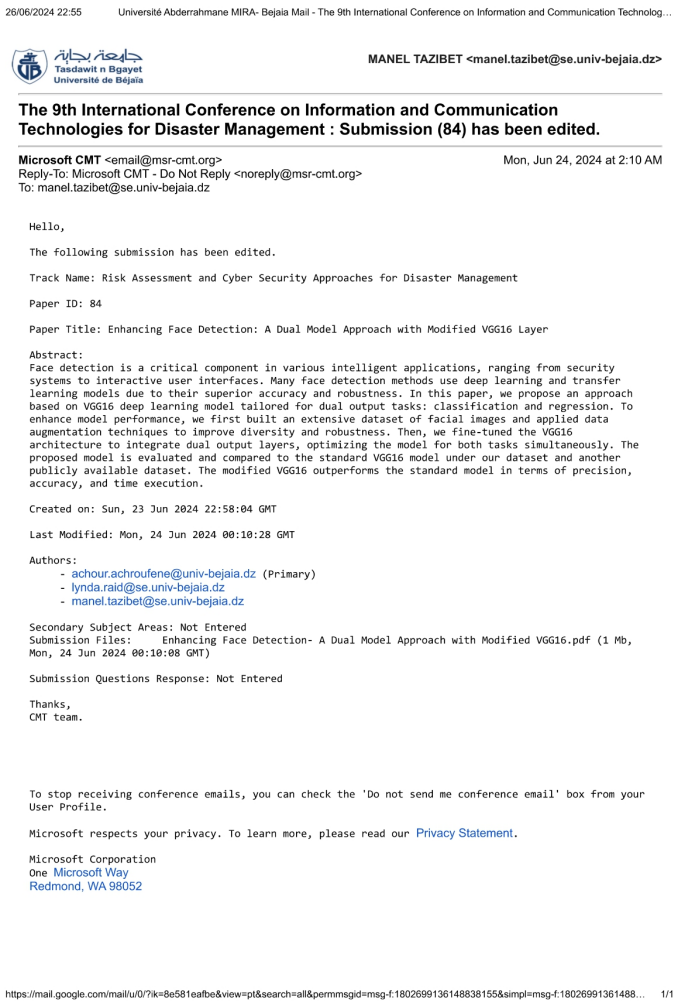
[71] Mamata S. Kalas. “REAL TIME FACE DETECTION AND TRACKING USING OPENCV”. In : 2014.

[72] Ashwin Gupta. *Human Faces*. https://www.kaggle.com/datasets/ashwingupta3012/human-faces. 2019.

Annexes

Soumission de l'article à une conférence

Nous avons soumis notre article intitulé *Enhancing Face Detection : A Dual Model Approach with Modified VGG16 Layer* à la 9ème Conférence *Technologies de l'Information et de la Communication pour la Gestion des Catastrophes* le *23 Juin 2024*. L'image ci-dessous montre l'accusé de réception de notre soumission.



26/06/2024 22:55 Université Abderrahmane MIRA- Bejaia Mail - The 9th International Conference on Information and Communication Technolog...

Tasdawit n Bgayet Université de Béjaia

MANEL TAZIBET <manel.tazibet@se.univ-bejaia.dz>

The 9th International Conference on Information and Communication Technologies for Disaster Management : Submission (84) has been edited.

Microsoft CMT <email@msr-cmt.org> Mon, Jun 24, 2024 at 2:10 AM
Reply-To: Microsoft CMT - Do Not Reply <noreply@msr-cmt.org>
To: manel.tazibet@se.univ-bejaia.dz

Hello,

The following submission has been edited.

Track Name: Risk Assessment and Cyber Security Approaches for Disaster Management

Paper ID: 84

Paper Title: Enhancing Face Detection: A Dual Model Approach with Modified VGG16 Layer

Abstract:
Face detection is a critical component in various intelligent applications, ranging from security systems to interactive user interfaces. Many face detection methods use deep learning and transfer learning models due to their superior accuracy and robustness. In this paper, we propose an approach based on VGG16 deep learning model tailored for dual output tasks: classification and regression. To enhance model performance, we first built an extensive dataset of facial images and applied data augmentation techniques to improve diversity and robustness. Then, we fine-tuned the VGG16 architecture to integrate dual output layers, optimizing the model for both tasks simultaneously. The proposed model is evaluated and compared to the standard VGG16 model under our dataset and another publicly available dataset. The modified VGG16 outperforms the standard model in terms of precision, accuracy, and time execution.

Created on: Sun, 23 Jun 2024 22:58:04 GMT

Last Modified: Mon, 24 Jun 2024 00:10:28 GMT

Authors:
- achour.achroufene@univ-bejaia.dz (Primary)
- lynda.raid@se.univ-bejaia.dz
- manel.tazibet@se.univ-bejaia.dz

Secondary Subject Areas: Not Entered
Submission Files: Enhancing Face Detection- A Dual Model Approach with Modified VGG16.pdf (1 Mb, Mon, 24 Jun 2024 00:10:08 GMT)

Submission Questions Response: Not Entered

Thanks,
CMT team.

To stop receiving conference emails, you can check the 'Do not send me conference email' box from your User Profile.

Microsoft respects your privacy. To learn more, please read our Privacy Statement.

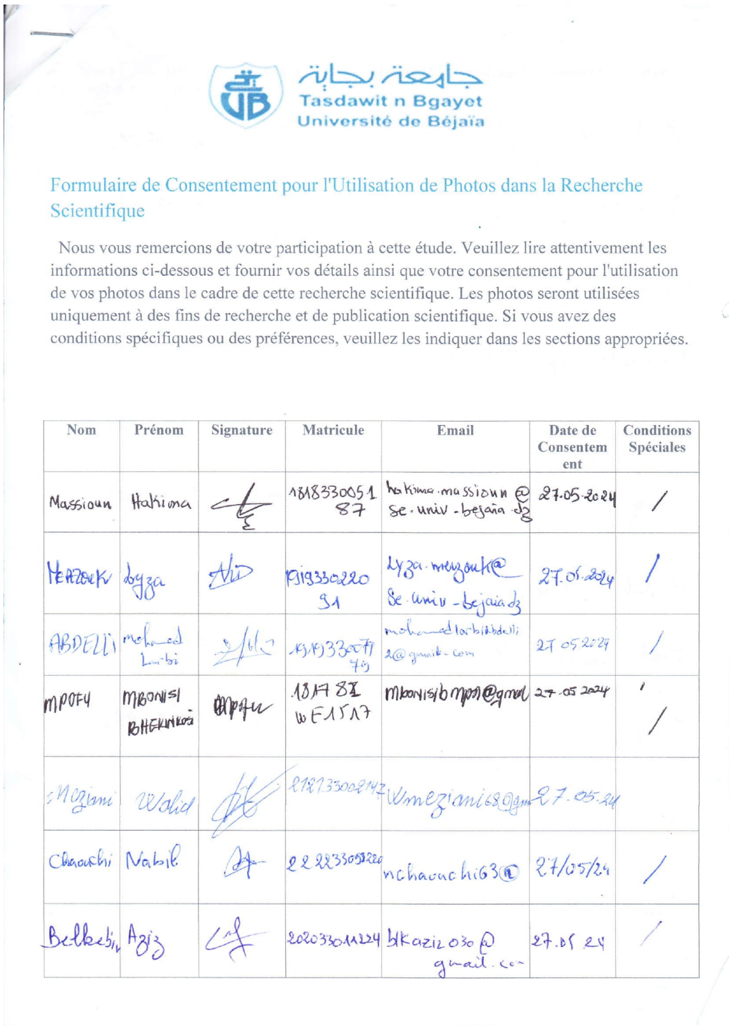
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

https://mail.google.com/mail/u/0/?ik=8e581eafbe&view=pt&search=all&permmsgid=msg-f:1802699136148838155&simpl=msg-f:18026991361488... 1/1

FIG. 4.15 : Accusé de réception de la soumission de l'article.

Signatures des étudiants participants

Nous avons collecté les signatures des étudiants qui ont accepté de participer à notre expérience et notre dataset. Ce formulaire précisait leur consentement à participer à la collecte de données, à l'utilisation des images dans le cadre de la recherche, et à la confidentialité des informations collectées.Les feuilles scannées sont disponibles ci-dessous.



جامعة بجاية
Tasdawit n Bgayet
Université de Béjaïa

Formulaire de Consentement pour l'Utilisation de Photos dans la Recherche Scientifique

Nous vous remercions de votre participation à cette étude. Veuillez lire attentivement les informations ci-dessous et fournir vos détails ainsi que votre consentement pour l'utilisation de vos photos dans le cadre de cette recherche scientifique. Les photos seront utilisées uniquement à des fins de recherche et de publication scientifique. Si vous avez des conditions spécifiques ou des préférences, veuillez les indiquer dans les sections appropriées.

Nom	Prénom	Signature	Matricule	Email	Date de Consentement	Conditions Spéciales
Massioun	Hakima		181833005187	hakima.massioun@se.univ-bejaia.dz	27-05-2024	/
Mezzouk	Lyza		191933022091	Lyza.mezzouk@se.univ-bejaia.dz	27.05.2024	/
ABDELLi	mohamed Larbi		19193300779	mohamedlarbiabdelli2@gmail.com	27.05.2024	/
MPOFU	MBONISI BHEKINKOSI		181781 WE1517	mbonisibmpofu@gmail	27.05 2024	/
Meziani	Walid		2121233002147	Wmeziani68@gm	27.05.24	
Chaouchi	Nabil		222233059220	nchaouchi63@	27/05/24	/
Belkebir	Aziz		202033011224	blkaziz030@gmail.com	27.05.24	/

FIG. 4.16 : Feuilles de signatures des étudiants participants.

Nom	Prénom	Signature	Matricule	Email	Date de Consentement	Conditions Spéciales
Zaadi	Ali			alibouzaadi02@gmail.com	25-05-2024	
Ouali	Anes			anesouali007@gmail.com.	25-05-2024	
Merighed	Aissam			merighedAissam@gmail.com	25-05-2024	
Benamma	Abdennacyl			abdennacyl.benamma1@gmail.com	25-05-2024	
Aissaoui	Sid Ahmed Adel			sidahmedaissaoui511@gmail.com	25-05-2024	
Kerma	Nacereddine			kermafn@gmail.com	25-05-2024	
Kaoudi	Nesrine			Nesrinekaoudi@gmail.com	25-05-2024	
Raad	Zyneb			Zyneb.raad@ec.univ-bejaia	25-05-2024	
TAZIB ET	Manel			Manel.TAZIBET@se.univ-bejaia.	25-05-2024	

FIG. 4.17 : Feuilles de signatures des étudiants participants.