RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abderrahmane Mira – Bejaia

Faculté des Sciences Exactes

Département d'Informatique

Mémoire de fin de cycle

En vue de l'obtention d'un diplôme en Master en Informatique recherche

Option : Intelligence Artificielle

Détection des attaques dans l'authentification biométriques par le visage avec Deep Learning.

Présenté par :

Mlle CHALGOU Iliza.

Devant le jury :

Président	A. Akilal	Maitre-Assistant. A	U. A/Mira Béjaia.
Examinateur	M. Moktefi	Maitre-Assistant. A	U. A/Mira Béjaia.
Encadrant	M. Khammari	Maitre de conf. A	U. A/Mira Béjaia.

2020-2021

REMERCIEMENTS

Dieu merci pour la santé, la volonté, le courage et la détermination qui m'ont accompagnée tout au long de la préparation de ce mémoire de Master et qui m'ont permis d'achever ce mémoire. Je tiens à remercier mon encadrant M. Mohammed Khammari pour ses précieuses orientations. Je remercie également les membres du jury d'avoir consacré de leur temps pour l'évaluation de mon modeste travail. En ce moment précis, toutes mes pensées vont vers mes honorables parents en reconnaissance à leur esprit de sacrifice et de dévouement ainsi qu'à leur soutien constant –moral et matériel- et ce, pour m'avoir permis de construire un avenir certain et en même temps réaliser mes rêves. A la fin, Je remercie tous ceux qui ont contribué à la réalisation de ce projet.

Merci

DÉDICACE

Je dédie ce modeste travail :

A **mes chers parents** qui m'ont aidé à réaliser ce parcours. Votre soutien et amour m'ont donné la force constamment pour réussir et prospérer dans la vie.

« Vos prières et vos bénédictions m'ont été d'un grand secours pour mener à bien mes études »

A **ma sœur SARA** pour ses gestes gracieux, ses encouragements et ses conseils.

A **mes amis** pour leur soutien infini et leur aide incessante, à qui je souhaite un meilleur avenir.

Enfin, je témoigne toute mon affection et gratitude à **Yanis** qui m'a inlassablement encouragé. Son aide et soutien moral étaient d'une si grande importance pour que ce mémoire soit achevé.

A LA MÉMOIRE DE MES CHERS DÉFUNTS

MES GRANDS-PARENTS

MA TANTE

AIMED QUI A TANT DONNÉ POUR MOI

Table des matières

Table des figures

Liste des tableaux

Liste des abréviations

Introduction générale : 1
1) Chapitre 1 : Présentation de la biométrie 4
1.1) Introduction : 4
1.2) Notions de base de la sécurité informatique : 4
1.3) Menaces et solutions : 5
1.4) Mécanismes de sécurité : 6
1.4.1) Premier niveau : 6
1.4.2) Second niveau : 6
1.5) Présentation de la biométrie : 7
1.6) Caractéristiques biométriques : 7
1.7) Modalités biométriques : 7
1.7.1) Modalités biologiques : 8
1.7.2) Modalités comportementales : 8
1.7.3) Modalités morphologiques : 10
1.8) Systèmes de sécurité avec la reconnaissance de visage : 13
1.9) Fonctionnement d'un système biométrique : 14
1.9.1) Phase d'enrôlement : 14
1.9.2) Phase d'authentification : 14
1.9.3) Phase d'identification : 15
1.10) Attaques contre un système de reconnaissance de visage : 15
1.10.1) Attaques à l'intérieur du système : 15
1.10.2) Attaques au niveau du capteur : 16
1.11) Conclusion : 17
2) CHAPITRE 2 : La reconnaissance faciale 18
2.1) Introduction : 18
2.2) Détection de visage : 18
2.2.1) Méthodes basées sur les connaissances : 18
2.2.2) Méthodes basées sur les caractéristiques invariantes : 18
2.2.3) Méthodes basées sur la mise en correspondance modèle : 19
2.2.4) Méthodes basées sur l'apparence globale : 19
2.3) Méthode de Viola et Jones : 19
2.4) Caractéristiques pseudo-Haar : 19

2.5) Image Intégrale : 20
2.6) Algorithme d'apprentissage Adaboost : 22
2.7) Reconnaissance de visage : 24
2.8) Fonctionnement de la reconnaissance faciale : 25
2.8.1) Étape 1 : détection du visage 25
2.8.2) Étape 2 : analyse du visage 25
2.8.3) Étape 3 : conversion de l'image en données 26
2.8.4) Étape 4 : trouver une correspondance 26
2.9) Cas d'usage de la reconnaissance faciale : 26
2.10) Avantages de la reconnaissance faciale : 28
2.11) Inconvénients de la reconnaissance faciale : 30
2.12) Conclusion : 31
3) CHAPITRE 3 : les attaques biométriques par le visage et solutions existantes 32
3.1) INTRODUCTION : 32
3.2) Attaques au niveau du capteur : 33
3.2.1) Attaques par photos : 33
3.2.2) Attaques par vidéo : 38
3.2.3) Attaque par masque 3D : 41
3.3) Détection de la vivacité : 43
3.4) Etat de l'art de la détection de vivacité 44
3.5) Apprentissage profond : 46
3.6) Apprentissage automatique traditionnel VS apprentissage profond : 47
3.7) Réseaux de neurones convolutifs : 48
3.8) Couche de convolution : 48
3.9) Couche d'échantillonnage (Pooling) : 50
3.10) Couche complètement connectée : 51
3.11) Quelques réseaux convolutifs célèbres : 53
3.12) Conclusion : 54
4) Chapitre 4 : Approche proposée 55
4.1) Introduction : 55
4.2) Objectif de notre approche : 55
4.3) Organigramme : 56
4.4) Méthodologie : 57
4.5) Implémentation de notre détecteur de vivacité basé sur l'apprentissage profond : 61
4.6) Conclusion 62
5) CHAPITRE 5 : Test et résultat 63
5.1) Introduction : 63
5.2) Environnement de développement : 63

5.3) Mesures d'évaluation : 63
5.4) Présentation des outils : 65
5.4.1) Logiciels (software) : 65
5.4.2) Matériels (hardware) : 67
5.5) Base de données utilisée : 67
5.6) Structure de l'application : 68
5.7) TEST : 70
5.8) Résultats et discussion 71
5.9) Conclusion 72
Conclusion générale et perspectives: 79
Bibliographie :
Résumé :
Abstract :

Table des figures

Figure 1-1 ADN [25]. 08
Figure 1-2 La voix [7]. 09
Figure 1-3 Signature [7]. 09
Figure 1-4 La dynamique de frappe sur un clavier [7]. 10
Figure 1-5 Les empreintes digitales [8]. 10
Figure 1-6 La main [7]. 11
Figure 1-7 L'iris [7]. 12
Figure 1-8 La rétine [7]. 12
Figure 1-9 Le visage [11]. 13
Figure 1-10 L'enrôlement [18]. 14
Figure 1-11 L'authentification [18]. 15
Figure 1-12 L'identification [18]. 15
Figure 2-1 caractérisques peudo-Haar [60]. 20
Figure 2-2 Image intégrale [61]. 21
Figure 2-3 somme des pixels dans la région est A-B-C+D [61]. 21
Figure 2-4 Calcul des caractéristiques à partir de l'image intégrale [62]. 22
Figure 2-5 Entrainement d'un classifieur faible [63]. 23
Figure 2-6 Construction du classifieur fort [63]. 23
Figure 2-7 Schéma général d'un système de reconnaissance de visage. 24
Figure 2-8 schéma général du fonctionnement d'un système de reconnaissance de visage. ... 25
Figure 3-1 Système sécurisé par la reconnaissance du visage [1]. 33
Figure 3-2 Piratage d'un système sécurisé par la reconnaissance du visage avec photo [1]. .. 34
Figure 3-3 Neutralisation du piratage avec photo, par la détection des clignements des yeux, dans un système sécurisé par la reconnaissance du visage [1]. 34
Figure 3-4 Exemple de régions œil binarisées de faux visages (a) et de visage réels (b)[2]. .. 35
Figure 3-5 Modèle graphique d'un CRF à chaîne linéaire [3]. 36
Figure 3-6 calcul LBP pour un pixel [20]. 37
Figure 3-7 Deux images : un vrai visage(a), une photo(c) et leurs images LBP correspondantes (b, d) [20]. 37
Figure 3-8 Caractéristiques d'illumination d'un faux visage et un vrai visage [21]. 38
Figure 3-9 Piratage d'un système sécurisé par la reconnaissance du visage avec vidéo [1]. ... 38
Figure 3-10 Défaillance du système de vérification par l'approche de clignement des yeux contre les attaques par vidéo [1]. 39
Figure 3-11 Comparaison de la façon dont les repères se comportent entre un visage authentique et la photo d'un visage [26]. 39
Figure 3-12 Exemple de la façon dont le défi est présenté à l'utilisateur [28]. 40
Figure 3-13 Piratage avec Masque 3D d'un système sécurisé par la reconnaissance du visage [27]. 41
Figure 3-14 Réflectance albédo pour une peau réelle et une peau en silicone [27]. 42
Figure 3-15 Exemple à partir de la base de données de masques créée par MORPHO [15]. .. 43
Figure 3-16 La relation entre l'intelligence artificielle, le ML et le deep learning [56]. 46

Figure 3-17 Le procédé du ML classique comparé à celui du deep learning [56]. 47
Figure 3-18 Exemple d'une convolution 2D [42]. 49
Figure 3-19 Exemple d'une opération de pooling. 51
Figure 3-20 Un réseau de neurones convolutif [42]. 52
Figure 3-21 taux d'erreur dans ImageNet Visual recognition Challenge [45]. 54
Figure 4-1 organigramme de notre système anti-spoofing. 57
Figure 4-2 Détection de visage. 58
Figure 4-3 Schéma fonctionnel pour la préparation du jeu de données pour système anti-spoofing. 58
Figure 4-4 Jeu de données pour la détection de la vivacité des visages. 59
Figure 4-5 Résultat de la formation(entrainement). 59
Figure 4-6 Architecture basée sur les CNN pour détecter les vrais et faux visages. 60
Figure 4-7 Notre modèle d'apprentissage profond basé sur CNN pour détecter les vrais et faux visages. 61
Figure 5-1 La croissance de la popularité de TensorFlow[71]. 66
Figure 5-2 Exemple de la base CASIA-FASD. 68
Figure 5-3 Le résultat de l'entraînement de notre détecteur de vivacité basé sur les CNN. 69
Figure 5-4 Courbes de sortie de l'historique de formation et de perte. 70
Figure 5-5 Chargement de notre de visage et vivacité de visage. 71
Figure 5-6 Notre détecteur de vivacité basé sur les CNN distingue avec succès les visages réels des visages faux/falsifiés. 71

Liste des tableaux

Tableau 1 Avantages et inconvénients de Theano. 65
Tableau 2 Avantages et inconvénients de TensorFlow. 65
Tableau 3 Avantages et inconvénients de Keras. 66
Tableau 4 Caractéristiques de la machine. 67

Liste des abréviations :

ADN Acide Désoxyribo Nucléique

CNN Convolutional Neural Network

CRF Conditional Random Fields

DNN Deep Neural Networks

FAR False Acceptance Rate

FBI Federal Bureau of Investigation

FC Fully Connected

FRR False Rejection Rate

HTER Half Total Error Rate

IDA Image Difference Analysis

ILSVRC ImageNet Large Scale Visual Recognition Challenge

LBP Local Binary Patterns

LDP Label Distrubution Protocol

ML Machine Learning

NBC National Broadcasting Company

NLP Naturel Language Processing

PMC Perceptron Multi Couches

ReLU Rectified Linear Unit

RNN Recurrent Neural networks

RVB Rouge, vert et bleu

SGD Stochastic Gradient Descent

STR Short Tandem Repeats

SVM Support Vector Machines

Introduction générale :

La biométrie cherche à identifier une personne à partir de la mesure d'éléments biologiques, comportementaux ou physiologiques uniques et propres à chaque individu, elle est utilisée dans plusieurs domaines, tels que l'interaction Homme Machine, la surveillance, l'indexation, la sécurisation des transactions, le Contrôle d'accès, etc.

L'une des techniques les plus utilisées dans les systèmes biométriques est la reconnaissance de visage qui vient en deuxième position après les empreintes digitales car la reconnaissance faciale ne nécessite pas une grande coopération des utilisateurs. Cette modalité biométrique est sans contact, naturelle, bien acceptée et ne nécessite en plus qu'un capteur très bon marchés (Webcam) qui est pratiquement disponible sur tous les appareils électroniques récents.
On voit de plus en plus l'utilisation du visage comme une modalité biométrique dans les systèmes de reconnaissance comme par exemple pour :
- Protéger des appareils électroniques personnels, comme les ordinateurs portables et les téléphones portables.
- Vérifier l'accès à un lieu privé réservé pour une population particulière, comme les laboratoires, les bureaux et les entreprises.
- Sécuriser les transactions bancaires, les paiements électroniques et l'activation des applications personnelles.
Un système automatique de reconnaissance de visages se décompose en trois sous-systèmes, qui sont la détection du visage, l'extraction automatique des caractéristiques et la vérification pour discriminer les vrais visages des faux visages(attaques).

On peut trouver plusieurs travaux et à tous les niveaux du système de reconnaissance.
Pour l'étape de détection du visage plusieurs méthodes de détection de visage ont été proposées ; on trouve les approches qui se basent sur des connaissances acquises comme celle introduite par Yang et Huang dans [1], les approches basées sur des caractéristiques invariables, qui utilisent la propriété de la peau humaine pour capter les régions contenant des visages [2], les approches basées sur la mise en correspondance « Template-matching » comme celle de Yuille et al. Dans [3] qui utilise un modèle déformable pour représenter les caractéristiques faciales. L'une des méthodes les plus performantes de la détection de visage est celle proposée par Viola et Jones [23], plusieurs systèmes de reconnaissance de visage sont basés sur cette méthode notamment les systèmes en temps réel.

Malheureusement les progrès accomplis jusqu'à présent pour réaliser un système de reconnaissance fiable et performant sont menacés par l'apparition du problème du piratage.
Le piratage d'un système de reconnaissance de personnes est l'accès non autorisé à ce système par usurpation d'identité.
L'usurpation d'identité ou « spoofing » est l'un des problèmes majeurs actuels auquel fait face la biométrie.
On parle de spoofing lorsqu'une personne essaie de se faire passer une autre personne pour accéder à un système de reconnaissance. Étant donné que les données faciales peuvent être facilement acquises sans contact, le spoofing est une menace réelle pour les systèmes de reconnaissance faciale.
Il existe trois types d'attaques par spoofing dans les systèmes de reconnaissance, des attaques par photos, les attaques par vidéo et les attaques par masques.
Pour tromper le système biométrique facial dans les attaques par photos, les pirates peuvent simplement présenter à la caméra une photo du visage de l'utilisateur ; l'efficacité de ce type d'attaque a été améliorée par l'apparition des imprimantes à haute résolution. Une autre possibilité c'est d'utiliser des photos très détaillées sur des écrans à haute définition (smart phones).
Les attaques par vidéo consistent à présenter une vidéo devant l'objectif de la caméra afin de tromper le système de reconnaissance faciale ; à présent il est possible de présenter la vidéo d'une personne au système de reconnaissance avec des supports électroniques comme par exemple un Ipad ou un Smartphone.
Pour tromper le système de reconnaissance faciale dans les attaques par masque, les pirates se présentent devant la caméra avec un masque 3D d'une personne appartenant à la base de données du système ; ce type d'attaque par masque est devenu une tâche facile grâce à l'apparition des imprimantes 3D et des caméras avec une haute résolution.

Nous avons choisi d'organiser notre mémoire en cinq chapitres principaux :
Chapitre 1 : On va étudier les notions de base de la sécurité, puis certaines menaces existantes et des mécanismes qui ont été mis en avant comme solutions.
On donnera ensuite un aperçu sur des technologies biométriques utilisées actuellement ou en cours de développement, telles que les empreintes, le visage, la rétine, etc.., puis nous présenterons un système de sécurité avec la reconnaissance de visage ainsi le fonctionnement d'un système biométrique et quelques différentes attaques contre un système de reconnaissance de visage.

Chapitre 2 : dans ce chapitre on va parler sur la détection de visage et ses étapes on détaillera la méthode proposée par Viola et Jones, ensuite on abordera la reconnaissance faciale et son fonctionnement et les différents domaines d'utilisation, puis on citera des avantages et inconvénients de la reconnaissance faciale.

Chapitre 3 : dans ce chapitre, nous allons parler sur des attaques (spoofing en anglais) et les techniques principales de spoofing dans les systèmes de reconnaissance du visage. Dans un second temps, on présentera les différentes méthodes proposées au cours de ces dernières années pour contrer ces attaques ensuite, nous allons parler sur les réseaux de neurones convolutifs (CNN).

Chapitre 4 : dans ce chapitre nous allons expliquer l'approche qu'on a proposé comme étant une contre-mesure lors d'une attaque sur un système de reconnaissance de visage.

Chapitre 5 : dans ce dernier chapitre nous allons décrire l'environnement de travail et la base de données qu'on a utilisé, ensuite on va donner un aperçu sur notre application, enfin nous discuterons sur les résultats des tests effectués.

1) Chapitre 1 : Présentation de la biométrie

1.1) Introduction :

Les technologies biométriques de reconnaissance apportent la simplicité et le confort aux utilisateurs et un très bon niveau de sécurité. Elles procurent une ergonomie non négligeable dans leur utilisation et sont une brique dans tout système de sécurité actuel et futur. Cette technologie est applicable à un large champ d'applications (contrôle d'accès, gestion horaire, paiement sécurisé sur Internet, login sur ordinateur, etc.).

Pour cela, nous consacrons ce chapitre pour l'étude des notions de base de la sécurité informatique dans un premier temps, quelques menaces existantes et les mécanismes qui ont été proposés comme solutions, on donnera ensuite un aperçu sur des technologies biométriques utilisées actuellement ou en cours de développement, telles que les empreintes, le visage, la rétine etc, puis nous présenterons un systèmes de sécurité avec la reconnaissance de visage ainsi le fonctionnement d'un système biométrique et quelques différentes attaques contre un système de reconnaissance de visage et nous terminerons par une conclusion.

1.2) Notions de base de la sécurité informatique :

Les notions les plus utilisées dans la sécurité informatique sont [4] :

Identification : l'obtention de l'identité d'une personne ou d'une entité.

Authentification : vérification qu'une personne ou une entité correspond bien à son identité déclarée.

Autorisation : vérification qu'une personne ou une entité possède les droits nécessaires pour accéder ou modifier une ressource.

Disponibilité : capacité d'une ressource à être accessible.

Intégrité : propriété d'une information qui n'a pas été modifiée où altérée lors d'un échange (par exemple sur un réseau) ou pendant sa conservation.

Maintenant que nous connaissons toutes ces notions, nous pouvons définir la sécurité informatique comme étant la discipline qui nous permet d'identifier et d'authentifier les entités qui possèdent les droits d'accès nécessaires aux ressources qui doivent être disponibles et intègres, et peuvent même contenir des données confidentielles.

1.3) Menaces et solutions :

Les différentes menaces peuvent être énumérées comme suit [6] :

Usurpation d'identité (Spoofing identity) un utilisateur non accrédité usurpe l'identité d'un utilisateur valide de l'application.

Altération des données (Tampering with data) un utilisateur détruit ou modifie les informations sans autorisation.

Répudiation (Repudiability) la possibilité qu'un utilisateur puisse nier avoir effectué telle ou telle opération.

Divulgation d'information (Information disclosure) des données confidentielles sont rendues visibles à des utilisateurs non accrédités.

Déni de service (Denial of service) l'application ou la ressource est rendue indisponible pour l'entité qui veut l'utiliser.

Elévation de privilège (Elevation of privilege) un utilisateur dispose un niveau d'accès à l'application supérieur à celui qui devrait lui être accessible.

Pour se prémunir de ces menaces, différentes techniques liées à la mise en œuvre de la sécurité sont envisageables :

Authentification identification des applications clientes. Les mécanismes d'authentification permettent de se prémunir contre l'usurpation d'identité (signature).

Sécurité des communications il faut faire en sorte que les messages circulant sur le réseau restent privés (chiffrement) et non altérés (signature du message).

Audit enregistrement des actions de l'application cliente, qu'elles soient autorisées ou non, pour garantir la non répudiation.

Autorisation définition des droits d'un client authentifié vis-à-vis d'une application ou d'une ressource.

1.4) Mécanismes de sécurité :

Les différents mécanismes de sécurité sont repartis en deux niveaux [4] :

1.4.1) Premier niveau :

A ce niveau de la sécurité, ou l'accès à une ressource est permis ou bien il ne l'est pas.

1.4.1.1) L'authentification :

Authentifier quelqu'un signifie s'assurer de son identité, cette authentification n'est possible que si les deux protagonistes partagent un secret commun, ce secret peut prendre diverses formes, et comme conséquence on distingue deux modes d'authentification :

Authentification directe : Lorsque l'entité A qui veut s'authentifier auprès de l'entité B, s'adresse directement à elle et lui divulgue le secret qu'ils partagent.

Authentification indirecte : Lorsque les deux entités ne partagent aucune relation de confiance, dans ce cas il est nécessaire d'utiliser un agent intermédiaire qui va vérifier l'identité de l'entité qui veut s'authentifier.

1.4.2) Second niveau :

(Contrôle d'accès) Concerne les autorisations, ce qui veut dire : qui a le droit de faire Quoi ? et sur quelle ressource ?

En général les systèmes de contrôle d'accès peuvent être représentés en termes de sujets, d'objets, d'opérations et de règles dont les définitions sont les suivantes :

- Un sujet peut être une application, un humain . . . etc. il interagit avec des données informatiques.
- Un objet est la donnée avec laquelle le sujet veut interagir (la ressource).
- Une opération représente ce que veut faire le sujet sur l'objet (lecture, exécution. . .).
- Une règle définit une relation entre un sujet, un objet et une opération. (Un ensemble de règles représente une politique de contrôle d'accès).

Le défaut commun à tous les systèmes d'authentification est que l'on identifie un objet (ordinateur, carte, code...) et non la personne elle-même. Il est pourtant plus acceptable d'authentifier une personne, plutôt qu'une machine (biométrie).

1.5) Présentation de la biométrie :

Il s'agit de toute caractéristique physique ou trait personnel automatiquement mesurable, robuste ou distinctif qui peut être employé pour identifier un individu (identification) ou pour vérifier l'identité d'un individu (authentification) [19].

La biométrie consiste à vérifier ou déterminer l'identité d'un individu à partir de ses caractéristiques biologiques tel que l'ADN, comportementales comme la voix ou morphologiques comme l'empreinte digitale [7].

1.6) Caractéristiques biométriques :

Le choix des caractéristiques physiques est important. Il faut qu'elles soient toutes à la fois [16] :

-**Universelles** : existent chez tous les individus.

-**Uniques** : possibilité de différencier un individu par rapport à un autre.

-**Permanentes** : stables et invariantes au cours du temps.

-**Enregistrables** : possibilité d'enregistrer les caractéristiques d'un individu à l'aide d'un capteur approprié qui ne cause aucun dérangement pour l'individu.

-**Performance** : Signifie que l'authentification doit être précise et rapide.

1.7) Modalités biométriques :

Les modalités biométriques sont les caractéristiques par lesquelles il est possible de vérifier l'identité d'un individu. Ces modalités sont basées sur l'analyse des données liées à l'individu et sont généralement classées en trois catégories : biologiques, comportementales et morphologiques. [5]

- La biométrie biologique se base sur l'analyse de données biologiques liées à l'individu (salive, ADN…).
- La biométrie comportementale se base sur l'analyse du comportement de l'individu (démarche, dynamique de frappe sur un clavier…).
- La biométrie morphologique se base sur les traits physiques particuliers qui sont chez toute personne permanents et uniques (empreinte digitale, visage …).

Comme cité, il existe trois types de modalités biométriques : biologiques, comportementales et morphologiques :

1.7.1) Modalités biologiques :

1.7.1.1) L'ADN :

Ce qu'on appelle également l'empreinte génétique, l'ADN est propre à chaque individu, ce qui le place parmi les modalités biométriques les plus robustes.

L'analyse la plus commune de l'ADN est basée sur les séquences répétées en tendem courtes STR (Short tandem repeats) [25].

Une représentation de l'ADN est montrée dans la figure 1.1

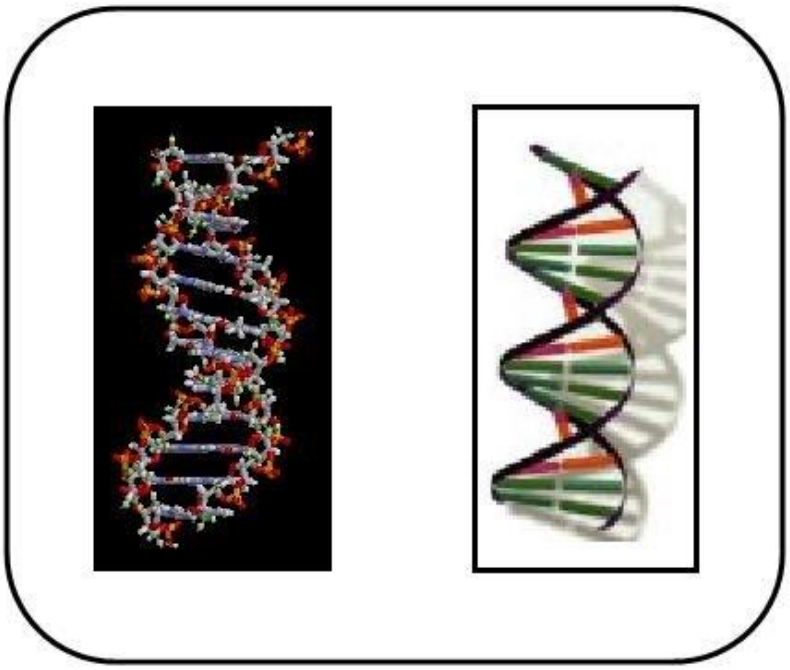


Figure 1-1 ADN [25].

1.7.2) Modalités comportementales :

1.7.2.1) La voix :

La technologie de l'analyse de la voix s'applique avec succès là où les autres technologies sont difficiles à employer. Elle est utilisée dans des secteurs tels que les centres d'appels, les opérations bancaires, l'accès à des comptes, sur un PC domestique, l'accès à un réseau ou encore pour des applications judiciaires.

La plupart des systèmes d'authentification de la voix utilisent l'affichage d'un texte, des mots spécifiques doivent être lu puis parlés afin de vérifier que la personne à authentifier est bien présente et qu'il ne s'agit pas d'un enregistrement [7].

Figure 1-2 La voix [7].

1.7.2.2) La signature :

Chaque personne possède un style d'écriture unique, on peut donc définir à partir de la signature d'une personne, un modèle qui pourra être employé pour effectuer une identification. De plus, la signature est utilisée dans beaucoup de pays comme élément juridique ou administratif [7]. La figure 1.3 nous montre un exemple de signature.

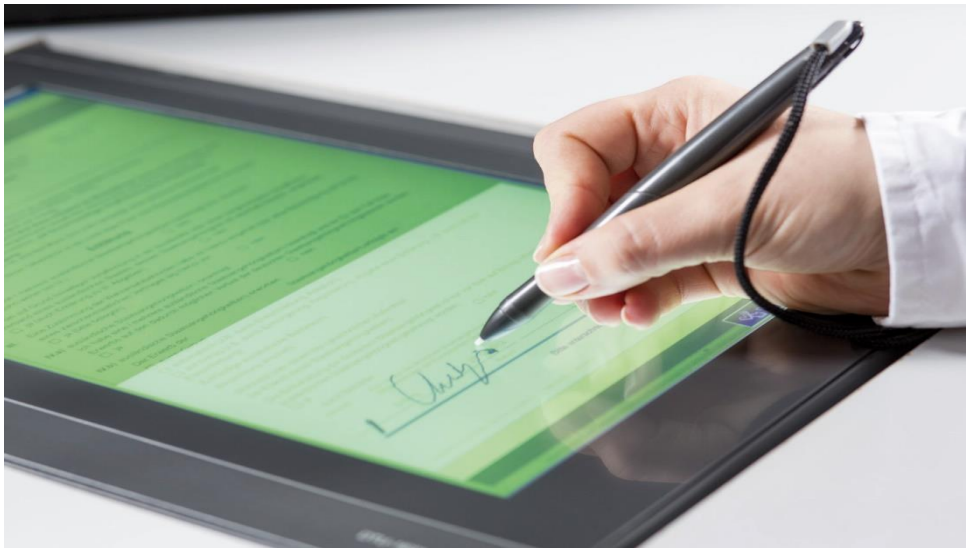


Figure 1-3 Signature [7].

1.7.2.3) La dynamique de frappe sur un clavier :

Il s'agit d'une technique de reconnaissance des personnes basée sur le rythme de frappe qui leurs est propre, elle est appliquée au mot de passe qui devient ainsi beaucoup plus difficile à imiter.

Lors de la mise en place de cette technique, il est demandé à l'utilisateur de saisir son mot de passe une dizaine de fois de suite, puis à l'aide d'un algorithme qui exploite le temps d'appui

sur chaque touche, la dizaine de saisie est moyennée pour bâtir un profil de frappe de l'utilisateur qui servira de référence aux accès suivants [7].

Un utilisateur entrain de saisir sur un clavier est montré dans la figure 1.4

Figure 1-4 La dynamique de frappe sur un clavier [7].

1.7.3) Modalités morphologiques :

1.7.3.1) Les empreintes digitales :

Une empreinte digitale est le dessin formé par les lignes de la peau des doigts, ce dessin se forme durant la période fœtale, les empreintes digitales sont uniques et immuables, elles ne se modifient donc pas au fil du temps (sauf en cas d'accident comme une brulure par exemple).

Les empreintes digitales sont regroupées en trois catégories principales : l'arche, le tourbillon et la boucle, à l'intérieur de chacune de ces catégories, il y'a un très grand nombre d'éléments qui différencient entre chaque personne, en plus des cicatrices, il y'a les fourches, les ilots et les espaces qui donnent un caractère unique aux empreintes [8].

Un exemple des trois catégories d'empreintes est présenté dans la figure 1.5

Figure 1-5 Les empreintes digitales [8].

1.7.3.2) La main :

La biométrie par la forme de la main est une technologie populaire qui est largement employée pour le contrôle d'accès physique ou le pointage horaire. Elle est très simple et bon marché, l'exactitude d'un système biométrique basé sur la forme de la main est tout à fait raisonnable.
Les éléments pris en compte ne reposent que sur la géométrie de la main et non sur l'empreinte palmaire.
Le système prend en photo la main et examine 90 caractéristiques, y compris la forme tridimensionnelle de la main, la longueur et largeur des doigts et la forme des articulations [7].
La figure 1.6 nous montre un exemple de l'examination d'une main.

Figure 1-6 La main [7].

1.7.3.3) L'iris :

L'identification par l'iris utilise plus de paramètres que les autres méthodes d'identification et la fiabilité résultante est suffisante non seulement pour faire de l'identification mais aussi l'authentification.
Pour distinguer l'iris, on utilise les sillons de contraction, les cryptes, les anneaux, etc [7].
La figure 1.7 représente l'iris.

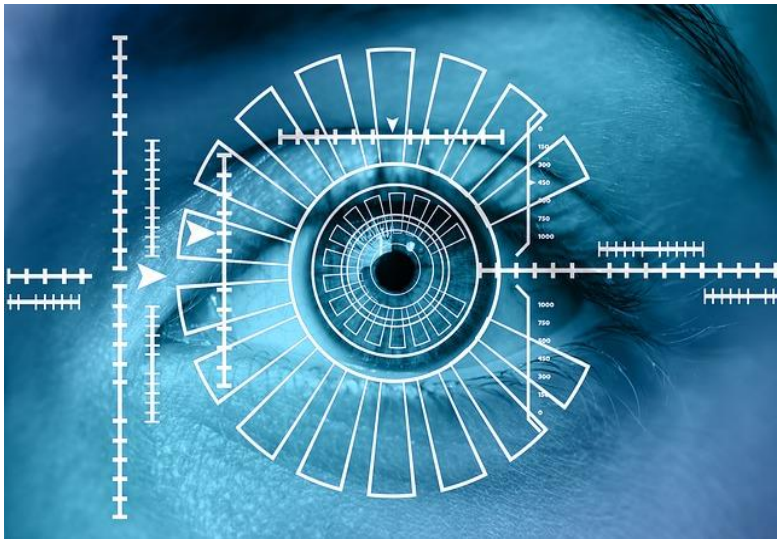


Figure 1-7 L'iris [7].

1.7.3.4) La rétine :

La rétine est la pellicule photographique de l'œil, elle est constituées de quatre couches de cellules et située au fond de l'œil.

Les éléments qui permettent de distinguer deux rétines sont les veines qui les tapissent, la disposition de ces veines est stable et unique pour chaque individu.

Après la capture d'une image de la rétine, le logiciel du dispositif de lecture découpe un anneau autour de la fovéa (la zone centrale), dans cet anneau il repère l'emplacement des veines et leur orientation, puis il les codifie dans un gabarit. L'opération en elle-même est assez simple à décrire mais les algorithmes restent relativement complexes [7].

Une image de la rétine est représentée dans la figure 1.8

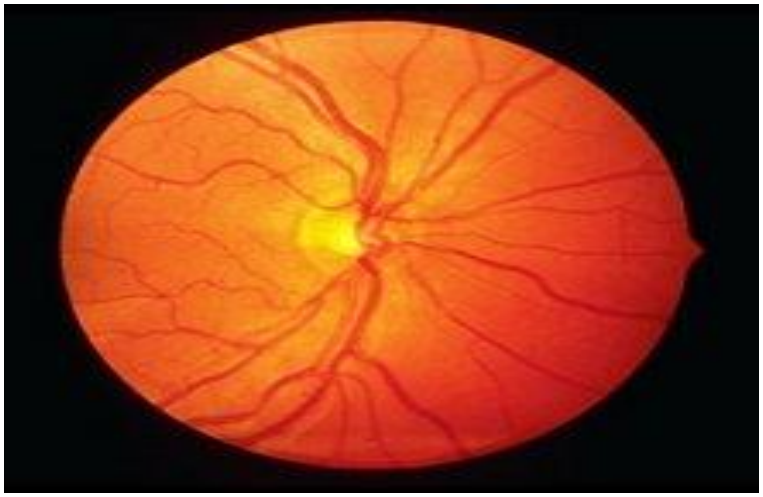


Figure 1-8 La rétine [7].

1.7.3.5) Le visage :

Les images faciales sont probablement la caractérise biométrique la plus communément employée par l'homme pour effectuer une identification personnelle.

L'utilisation d'une caméra permet de capter la forme du visage d'un individu et d'en dégager certaines particularités, selon le système utilisé, l'individu doit être positionné devant l'objectif de l'appareil ou peut être en mouvement à une certaine distance, les données biométriques qui sont obtenues par la suite, sont comparées au fichier de référence.

Il existe plusieurs techniques de reconnaissance par l'analyse du visage, mais pour la plupart il est d'intérêt que ces techniques se basent sur les éléments du visage qui sont le moins susceptibles aux changements (les grands traits supérieurs des orbites, les secteurs entourant les pommettes, les côtés de la bouche et d'autres caractéristiques similaires) de façon à ignorer les changements en identification sur les bases de milliers voir de centaines de milliers de personnes [11].

La figure 1.9 représente un exemple d'un visage humain.

Figure 1-9 Le visage [11].

1.8) Systèmes de sécurité avec la reconnaissance de visage :

Un système de sécurité avec la reconnaissance de visage est utilisé pour sécuriser l'accès à des ressources ou seulement les personnes autorisées ont le droit d'y accéder et toute autre personne est rejetée, son fonctionnement est similaire au fonctionnement des autres systèmes biométriques avec les modalités précédemment énumérées, que ça soit dans l'enrôlement, l'identification et l'authentification [18].

1.9) Fonctionnement d'un système biométrique :

Les systèmes biométriques peuvent fournir trois modes de fonctionnement, à savoir : l'enrôlement, l'authentification (vérification) et l'identification.

1.9.1) Phase d'enrôlement :

C'est la première phase de tout système biométrique, il s'agit de l'étape pendant laquelle un utilisateur est enregistré dans le système pour la première fois et ou une ou plusieurs modalités biométriques sont capturées et enregistrées dans une base de données, l'ensemble de données de cette base qui sont utilisées pour présenter un utilisateur est appelé un modèle biométrique [18].

Le processus de l'enrôlement est représenté dans la figure 1.10.

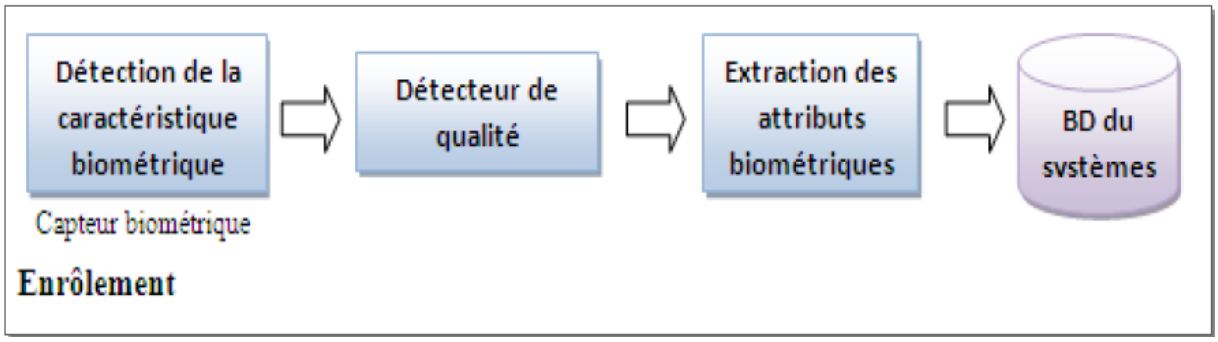


Figure 1-10 L'enrôlement [18].

1.9.2) Phase d'authentification :

Lorsqu'un système biométrique opère en mode authentification, l'utilisateur affirme son identité et le système vérifie si cette affirmation est valide ou non, on parle alors de correspondance 1 :1 ,si l'entrée biométrique de l'utilisateur et le modèle enregistré dans la base de données correspondant à l'identité affirmée possèdent un degré de similitude élevé, l'affirmation est validée et l'utilisateur est considéré comme étant authentique, dans le cas contraire l'affirmation est rejetée et l'utilisateur est considéré comme étant un imposteur [18].

Le processus de l'authentification est représenté dans la figure 1.11.

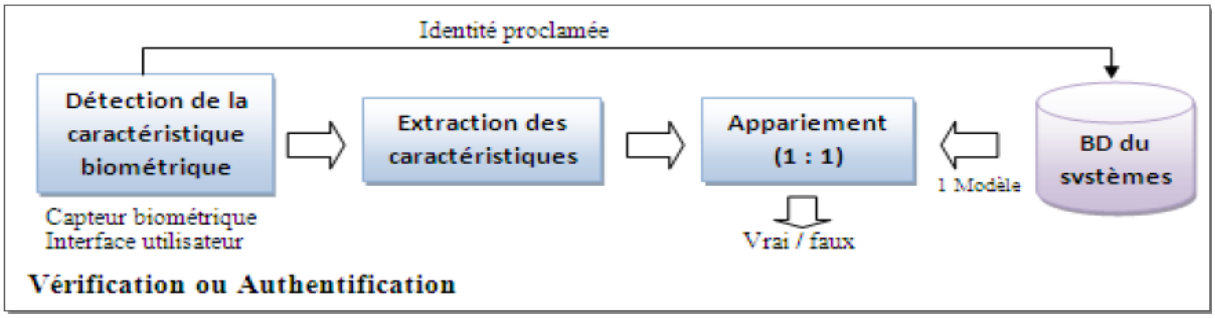


Figure 1-11 L'authentification [18].

1.9.3) Phase d'identification :

Lorsqu'on parle d'identification, l'utilisateur ne dévoile pas son identité et le système vérifie si l'entrée biométrique de l'utilisateur correspond à un modèle déjà enrôlé dans la base de données on parle alors de correspondance 1 : n, La sortie du système biométrique est constituée par l'identité de la personne dont le modèle possède le degré de similitude le plus élevé avec l'échantillon biométrique présenté en entrée. Typiquement, si la plus grande similarité entre l'échantillon et tous les modèles est inférieure à un seuil de sécurité minimum fixé, la personne est rejetée, ce qui implique que l'utilisateur n'était pas une des personnes enrôlées par le système, dans le cas contraire, la personne est acceptée [18]. Le processus de l'identification est représenté dans la figure 1.12.

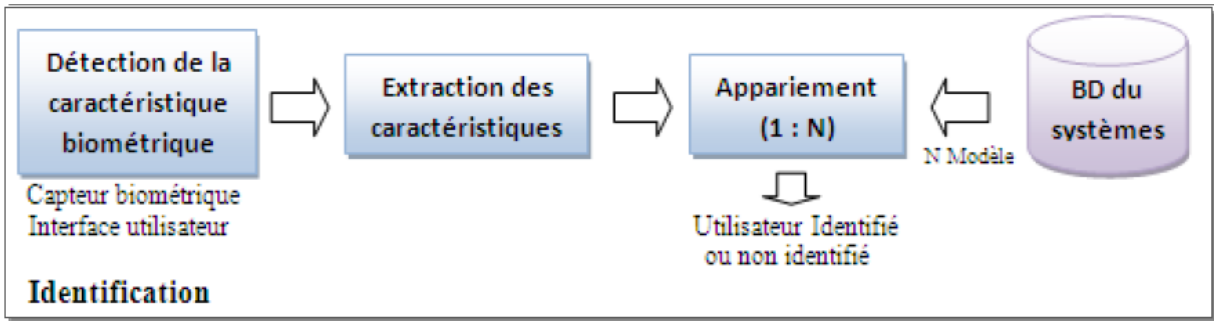


Figure 1-12 L'identification [18].

1.10) Attaques contre un système de reconnaissance de visage :

Un système de sécurité doit faire face aux attaques qui se divisent en deux catégories [10] :

1.10.1) Attaques à l'intérieur du système :

Cette catégorie se fait à l'intérieur du système en modifiant les données biométriques pour permettre aux personnes non autorisées d'accéder au système ou pendant la communication en remplaçant les décisions négatives d'authentification [10].

1.10.2) Attaques au niveau du capteur :

Cette catégorie d'attaque se fait au niveau du capteur pendant l'acquisition, ces attaques consistent principalement à présenter devant la caméra de faux visages, Ces visages peuvent être réalisés à l'aide d'une image d'une autre personne (imprimée sur papier ou affichée sur un écran), une vidéo ou un masque 3D (spofing).

On parle de spoofing lorsqu'une personne essaie de se faire passer à une autre personne pour accéder à un système de reconnaissance. Étant donné que les données faciales peuvent être facilement acquises sans contact, le spoofing est une menace réelle pour les systèmes de reconnaissance faciale [1].

C'est à ce niveau que nous souhaitons développer un système de détection contre ce type d'attaques.

Ce type d'attaques peuvent être réalisées par trois méthodes :

1.10.2.1) Attaques par photos :

Dans ce type, le pirate se présente devant le capteur avec une photo de la personne qui est autorisée à accéder à la ressource voulue, la photo peut être imprimée sur papier ou affichée sur un équipement tel qu'une tablette [1].

Pour distinguer entre le visage réel et une usurpation d'identité, les chercheurs se sont focalisés sur la détection de la vivacité du visage, par exemple le clignement des yeux.

1.10.2.2) Attaques par vidéos :

Dans ce type, le pirate utilise des supports vidéos comme : les Smartphones, tablette…sur lesquels une vidéo du visage de la personne autorisée à accéder à la ressource est jouée.

Face à ce type d'attaque, les méthodes de détection de vivacité ont échoué, et pour résoudre ce problème les chercheurs se sont focalisés sur l'analyse du mouvement du visage pour distinguer les mouvements 3D d'un visage réel des mouvements 2D d'un faux visage [1].

1.10.2.3) Attaques par masques 3D :

C'est le troisième type d'attaques, dans lequel le pirate tente de tromper le système de reconnaissance de visage en utilisant un masque 3D du visage de la personne autorisée à l'accès de la ressource sécurisée, le pirate porte le masque puis se présente devant le capteur [1].

Les méthodes utilisées dans les deux types d'attaques précédentes ne font pas face à ce dernier type, les chercheurs sont toujours d'actualité pour trouver des contres mesures pour ce type d'attaque.

1.11) Conclusion :

On peut constater que la biométrie est une véritable alternative aux mots de passe et autres identifiants. Elle permet de vérifier que l'usager est bien la personne qu'il prétend être. Cette technologie est en pleine croissance, particulièrement la reconnaissance de visage qui suscite de plus en plus d'intérêt de la communauté scientifique, et cela grâce à sa popularité et sa simplicité ainsi que le faible coût pour réaliser ce système, car il ne suffit que d'une caméra pour capter les images de visages et d'un ordinateur pour faire les calculs.

Dans ce chapitre nous avons présenté quelques notions et définitions de base sur les différentes technologies biométriques utilisées pour l'identification des personnes.

Dans le chapitre suivant, nous allons étudier la reconnaissance faciale et la détection de visage.

2) CHAPITRE 2 : La reconnaissance faciale

2.1) Introduction :

Des progrès techniques importants ont été réalisés ces dernières années dans le domaine du traitement d'images, en particulier en reconnaissance faciale. Les déploiements et expérimentations de ce type de systèmes sont de plus en plus nombreux.

Le développement dans ce domaine se dirige vers des applications de vision plus généralisées, telles que la reconnaissance de visage. Dans ce chapitre nous allons parler sur la détection de visage et ses étapes, nous allons aussi aborder la reconnaissance faciale et son fonctionnement Ensuite citer les différents domaines d'utilisation, et enfin parler des avantages et inconvénients de la reconnaissance faciale.

2.2) Détection de visage :

La détection de visages est la première étape dans le processus de reconnaissance faciale. Son efficacité a une influence directe sur les performances du système de reconnaissance de visages. Il existe plusieurs méthodes pour la détection de visages, certaines utilisent la couleur de la peau [65] [66], la forme de la tête et la couleur de la peau [67], l'apparence faciale [68], alors que d'autres combinent plusieurs de ces caractéristiques [69]. Les méthodes de détection de visage peuvent être classifiées en quatre catégories [64] :

2.2.1) Méthodes basées sur les connaissances :

Ces méthodes sont basées sur des réglés qui tentent de modéliser la connaissance de ce qui caractérise un visage. Par exemple un visage apparait souvent dans une image avec deux yeux symétriques entre un nez, et une bouche. Classiquement, ces règles représentent des relations en caractéristiques faciales.

2.2.2) Méthodes basées sur les caractéristiques invariantes :

Ces méthodes se basent sur des caractéristiques structurelles (Traits faciaux, Texture, Couleur de la peau) qui existent même quand la pose, le point de vue, ou les conditions d'illumination varient.

2.2.3) Méthodes basées sur la mise en correspondance modèle :

Plusieurs modèles standards de visages sont prédéfinis manuellement ou paramètres par des fonctions. L'un de ces modèles est comparé à une image en entrée. La corrélation entre une image présentée et la base des modèles est évaluée pour détecter la présence de visage.

2.2.4) Méthodes basées sur l'apparence globale :

Les modèles sont ici appris à partir d'un ensemble d'images d'apprentissage qui doivent permettre de caractériser la variabilité de l'apparence d'un visage. Ces méthodes se basent sur des techniques telles que l'analyse statistique et l'apprentissage automatique pour trouver les caractéristiques appropriées des images de visage et de non-visage. L'une des méthodes les plus performantes est celle proposée par Viola et Jones [23], cette méthode est même utilisée dans la détection en temps réel, et cela est dû grâce à son taux élevé de détection et au temps qu'elle prend pour l'exécution.

2.3) Méthode de Viola et Jones :

La méthode de Viola et Jones consiste à balayer une image avec une fenêtre de détection de taille initiale 24px par 24px (dans l'algorithme original) pour déterminer la présence d'un visage, Le balayage consiste à décaler la fenêtre dans l'image avec un pixel et si on veut accélérer le processus on augmente le nombre des pixels d´emplacement mais cela va affecter la précision, ensuite lorsque l'image est entièrement parcourue avec cette fenêtre sa taille va augmenter avec un facteur multiplicatif de 1.25. Cette méthode est une approche basée sur l'apparence, qui consiste à parcourir l'ensemble de l'image en calculant un certain nombre de caractéristiques dans des zones rectangulaires qui se chevauchent. Elle a la particularité d'utiliser des caractéristiques très simples mais très nombreuses. Et avant cela l'apprentissage du classifieur est une étape préliminaire très importante, Il s'agit d'entrainer le classifieur afin de le sensibiliser à ce que l'on veut détecter, ici des visages, à l'aide d'un ensemble d'images (positives et négatives) [9].

2.4) Caractéristiques pseudo-Haar :

Une caractéristique est une représentation synthétique et informative, calculée comme la différence des sommes de pixels de deux ou plusieurs zones rectangulaires adjacentes.

La figure 2.1 représente les différentes caractéristiques pseudo-Haar.

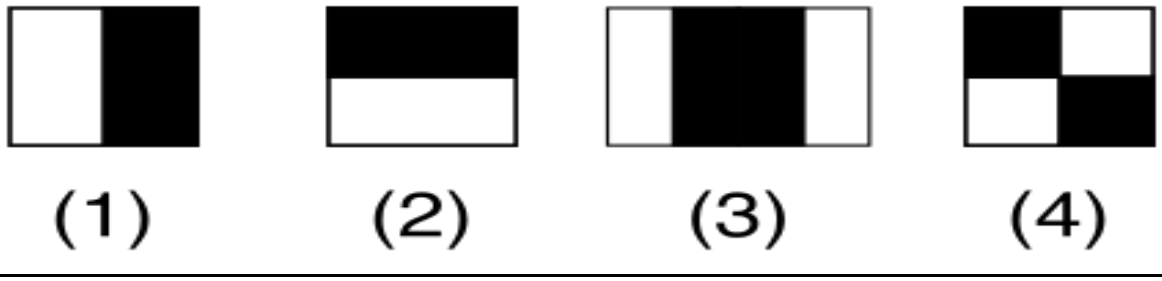


Figure 2-1 caractérisques peudo-Haar [60].

Ces masques sont appliqués à différentes positions d'une fenêtre avec une taille variable, Les caractéristiques seraient calculées en soustrayant la somme des pixels noirs à la somme des pixels blancs, dans une fenêtre de taille 24*24 pixels un très grand nombre de caractéristiques peut être généré d'après Viola et Jones [23], si on considère tous les paramètres possibles des caractéristiques de haar comme la position, l'échelle et le type, on aura plus de 180 000 caractéristiques potentielles. Le calcul de ces caractéristiques d'une manière classique coûte beaucoup de temps c'est pour cela qu'ils ont utilisé l'image intégrale.

2.5) Image Intégrale :

Les images intégrales permettent de gagner du temps dans le calcul des caractéristiques, cette image est de la même taille que l'image d'origine, et la valeur de chaque pixel de cette image est égale à la somme des pixels situés au-dessus et à gauche de ce pixel dans l'image d'origine, le pixel rouge dans la figure 2.2 est égal à la somme de tous les pixels bleus [57]. Mathématiquement, la valeur du pixel (x,y) de l'image intégrale I associée à l'image f est :

$$I(x,y) \doteq \sum_{m \doteq 1}^{x} \sum_{n \doteq 1}^{x} f(m,n) \qquad (1)$$

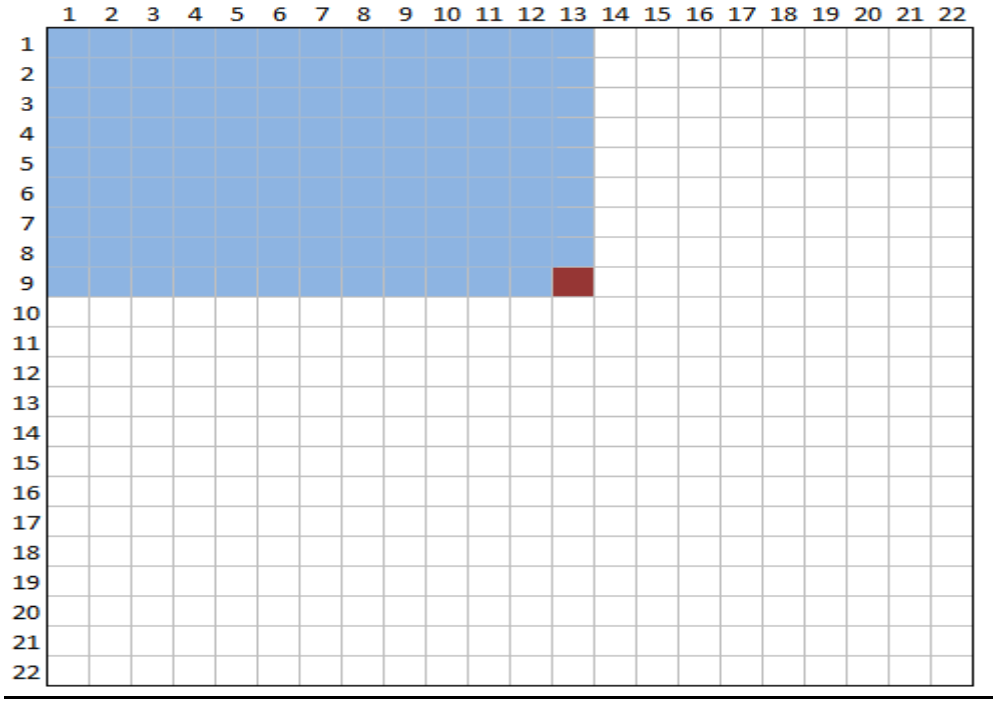


Figure 2-2 Image intégrale [61].

En utilisant l'image intégrale le calcul de la somme des pixels d'une région de l'image se fait en un nombre fixe d'accès seulement quatre, ce qui nous fera gagner du temps [57]. La figure 2.3 représente un exemple sur le calcul de la somme des pixels d'une région.

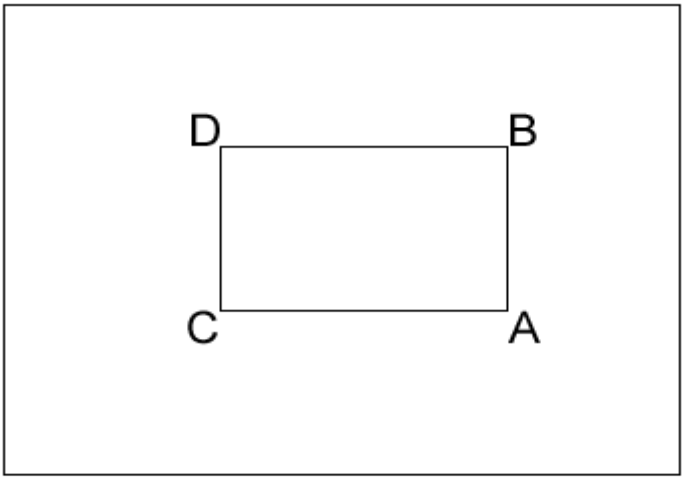


Figure 2-3 somme des pixels dans la région est A-B-C+D [61].

La valeur de la somme des pixels en chacun des points est connue grâce à l'image intégrale, donc pour calculer la somme des pixels du rectangle ABDC il suffit de faire : A − B − C + D. Ainsi une caractéristique pseudo-Haar à deux rectangles peut alors être déterminé en seulement

6 accès (2 points sont partagés) à l'image, et une caractéristique à 3 rectangles en seulement 8 accès [57], un exemple est donné dans La figure 2.4

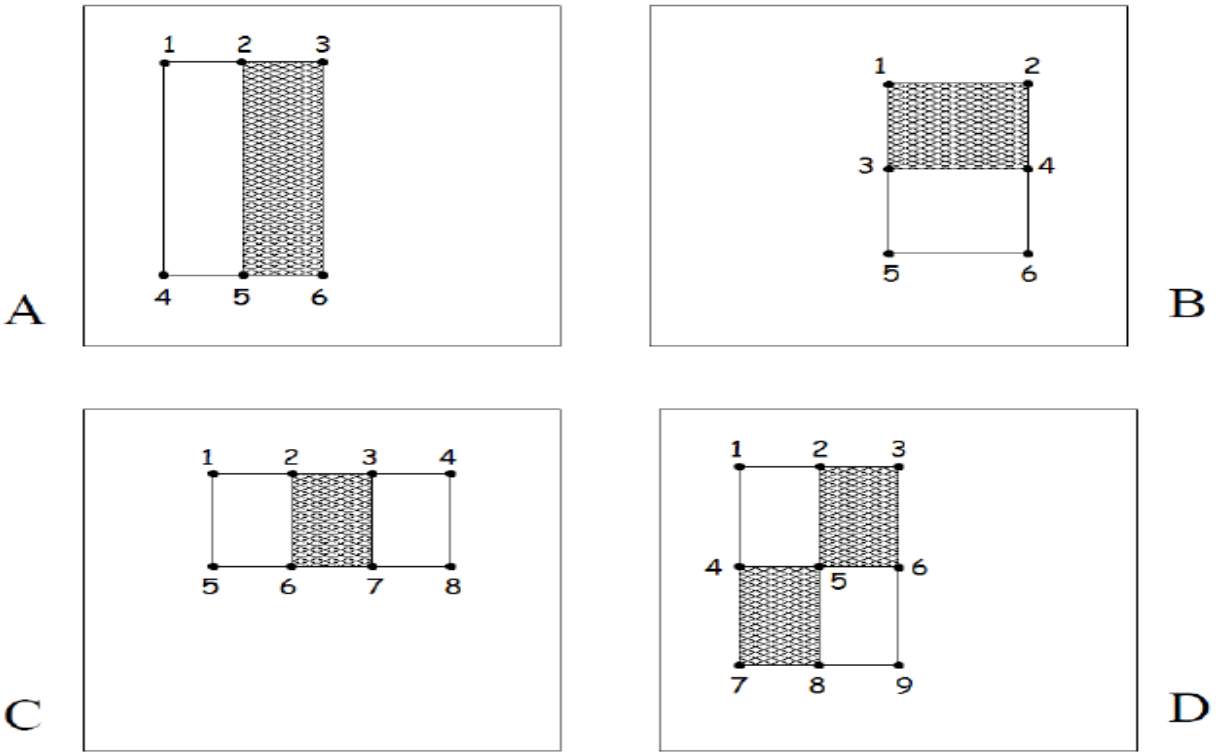


Figure 2-4 Calcul des caractéristiques à partir de l'image intégrale [62].

Calcul de la caractéristique A : 2 + 6 - 3 - 5 - (1 + 5 - 2 - 4)

Calcul de la caractéristique B : 1 + 4 - 2 - 3 - (3 + 6 - 4 - 5)

Calcul de la caractéristique C : 2 + 7 - 3 - 6 - (1 + 6 - 2 - 5) - (3 + 8 - 4 - 7)

Calcul de la caractéristique D : 2 + 6 - 3 - 5 + 4 + 8 - 5 - 7 - (1 + 5 - 2 - 4) - (5 + 9 - 6 - 8)

2.6) Algorithme d'apprentissage Adaboost :

Adaboost est un principe d'apprentissage automatique qui consiste à construire un classifieur fort à partir d'une combinaison pondérée de classifieurs faibles et d'en déterminer les meilleures caractéristiques [63].

Dans leur méthode de Viola et Jones un classifeur est assimilé à une seule caractéristique de Haar fi, l'apprentissage du classifieur faible hj(x) consiste à trouver un seuil Θj de la caractéristique qui permet de séparer les exemples positifs et négatifs.

Chaque classifieur faible se réduit au tuple (fj ,Θj , pj), où pj est la parité permettant de définir le sens d'application du seuil.

$$h_j(x) = 1 \text{ si } p_j f_j(x) < p_j\Theta_j \quad (2)$$

$$h_j(x) = 0 \text{ sinon} \quad (3)$$

La figure 2.5 montre un exemple d'entrainement d'un classifieur faible avec une image positive et une autre négative.

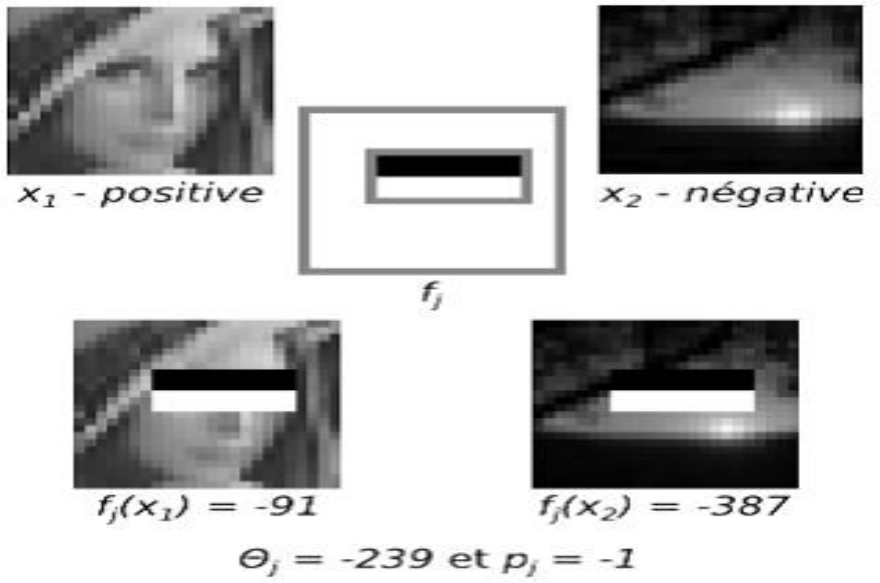


Figure 2-5 Entrainement d'un classifieur faible [63].

Après avoir construit les classifieurs faibles le classifieur fort h est construit (Figure 2.6) avec la somme pondérée des classifieurs faibles, et un seuil est d´défini pour ce classifieur fort.

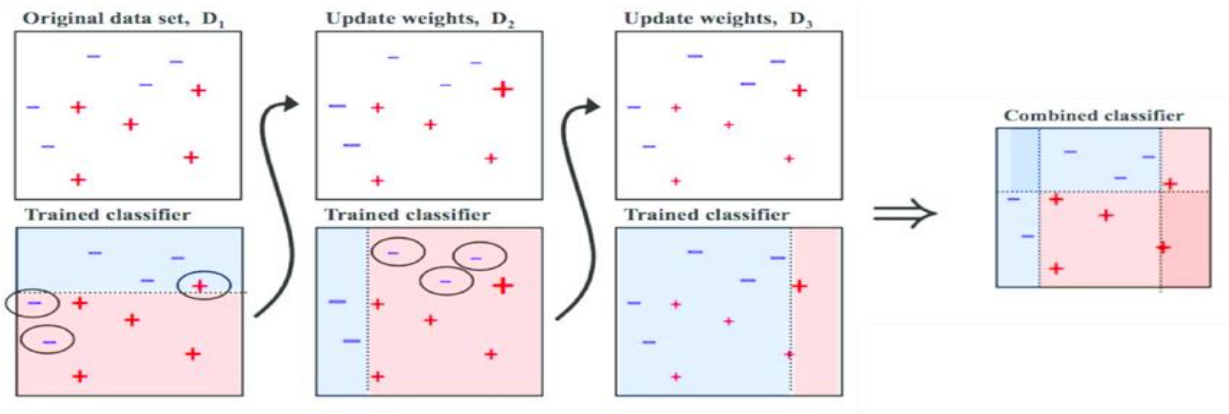


Figure 2-6 Construction du classifieur fort [63].

Après avoir effectué la détection de visage, il est désormais possible de faire la reconnaissance de visage.

2.7) Reconnaissance de visage :

La reconnaissance faciale est un moyen d'identifier ou de confirmer l'identité d'un individu grâce à son visage. Les systèmes de reconnaissance faciale peuvent servir à l'identification de personnes sur des photos, dans des vidéos ou en temps réel.

Les visages constituent une catégorie de stimulus unique par la richesse des informations qu'ils véhiculent, le but de la reconnaissance de visage est de concevoir des systèmes informatiques capables de copier les facultés de reconnaissance du cerveau humain, Idéalement, un système de reconnaissance faciale doit pouvoir identifier des visages présents dans une image ou une vidéo de manière automatique [19] :

Un schéma général qui décrit un système de reconnaissance de visage est présenté dans la figure 2.7

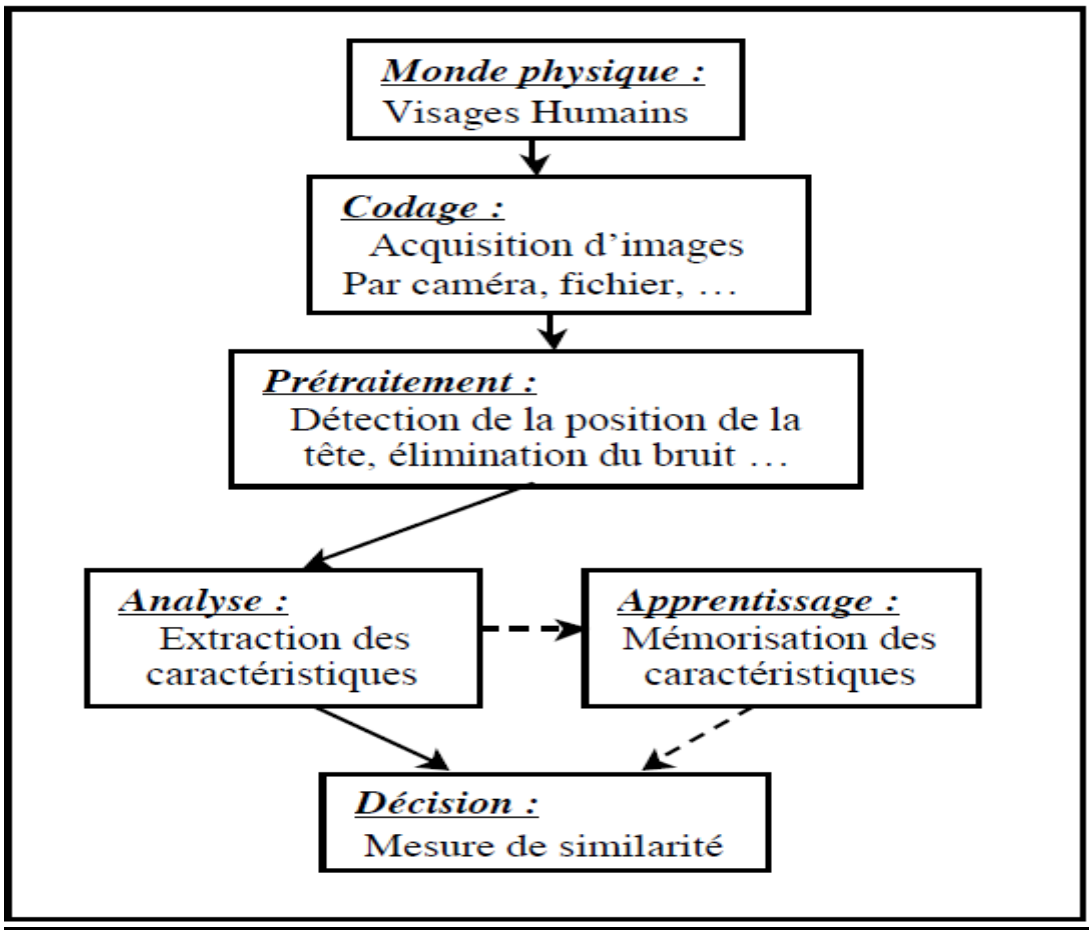


Figure 2-7 Schéma général d'un système de reconnaissance de visage [19].

2.8) Fonctionnement de la reconnaissance faciale :

Beaucoup de gens connaissent la reconnaissance faciale grâce à l'utilisation de FaceID qui permet de déverrouiller les iPhone (ce n'est qu'une utilisation parmi d'autres). De manière générale, la reconnaissance faciale ne repose pas sur une énorme base de données de photos pour déterminer l'identité d'une personne : elle ne fait qu'identifier et reconnaître une personne comme propriétaire unique de l'appareil, pour empêcher autrui d'y avoir accès [12].

Au-delà du déverrouillage des téléphones, la reconnaissance faciale fonctionne en comparant le visage des personnes qui passent devant des caméras spéciales à des photos de personnes surveillées. Ces listes de personnes surveillées peuvent contenir des photos de n'importe qui, même de personnes qui n'ont jamais rien fait de mal, et les photos peuvent provenir de n'importe où, même des comptes de réseaux sociaux. Les systèmes de reconnaissance faciale peuvent varier, mais ils fonctionnent généralement de la manière suivante :

Un schéma général qui décrit le fonctionnement d'un système de reconnaissance de visage est présenté dans la figure 2.8

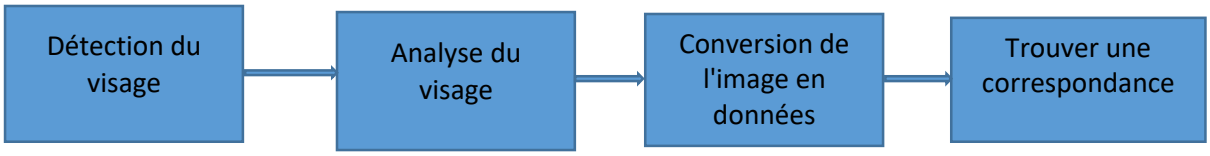


Figure 2-8 schéma général du fonctionnement d'un système de reconnaissance de visage.

2.8.1) Étape 1 : détection du visage

La caméra détecte et localise l'image d'un visage, seul ou dans une foule. L'image peut montrer la personne de face ou de profil.

2.8.2) Étape 2 : analyse du visage

Ensuite, une image du visage est capturée et analysée. La plupart des technologies de reconnaissance faciale utilisent la 2D plutôt que la 3D, car il est plus pratique de comparer une image en 2D à des photos ou aux images d'une base de données. Le logiciel analyse la géométrie du visage. Les facteurs clés incluent la distance entre les yeux, la profondeur des orbites, la distance entre le front et le menton, la forme des pommettes, ainsi que le contour des lèvres, des oreilles et du menton. Le but est d'identifier les spécificités de votre visage [12].

2.8.3) Étape 3 : conversion de l'image en données

Le processus de capture du visage transforme les informations analogues (un visage) en un ensemble d'informations numériques (les données) selon les caractéristiques du visage de la personne. En fait, l'analyse de votre visage est transformée en formule mathématique. Le code numérique est appelé une empreinte faciale. De la même manière que les empreintes digitales sont uniques, tout le monde a sa propre empreinte faciale.

2.8.4) Étape 4 : trouver une correspondance

Votre empreinte faciale est ensuite comparée à une base de données avec d'autres visages connus. Par exemple, le FBI a accès à près de 650 millions de photos sur de nombreuses bases de données d'État[12]. Sur Facebook, les photos avec des personnes identifiées rejoignent leur base de données qui peut aussi être utilisée pour la reconnaissance faciale. Si votre empreinte faciale correspond à une image de la base de données utilisée par la reconnaissance faciale, une correspondance est effectuée.

La reconnaissance faciale est considérée comme la mesure biométrique la plus naturelle. Ce qui est logique au niveau intuitif, car nous nous reconnaissons nous-mêmes et les autres en observant le visage, plutôt que les empreintes digitales ou l'iris. On a déterminé que plus de la moitié de la population du monde est concernée régulièrement par la technologie de reconnaissance faciale [12].

2.9) Cas d’usage de la reconnaissance faciale :

Cette technologie sert à bien des choses. Cela inclut :

Le déverrouillage des téléphones :

De nombreux téléphones, y compris les iPhones les plus récents, utilisent la reconnaissance faciale pour déverrouiller l'appareil. Cette technologie fournit une manière efficace de protéger les données personnelles et de veiller à ce que les données sensibles restent inaccessibles si le téléphone est volé [12]. Apple déclare que les chances de déverrouiller votre téléphone avec un visage au hasard sont d'environ une sur un million.

Application de la loi :

La reconnaissance faciale est souvent utilisée par les autorités. Selon ce rapport de la FBI, cette technologie est de plus en plus utilisée par les autorités aux États-Unis et dans d'autres pays

[13]. La police collecte des photos anthropométriques des personnes arrêtées et les compare aux bases de données locales, étatiques et fédérales. Une fois prise, la photo du suspect est ajoutée aux bases de données pour être scannée dès que la police effectue une autre recherche de criminel.

La reconnaissance faciale mobile permet également aux policiers d'utiliser les smartphones, les tablettes ou d'autres appareils portables pour prendre en photo des conducteurs ou des piétons sur le terrain et de comparer immédiatement la photo à une ou plusieurs bases de données pour essayer de les identifier [13].

Contrôles dans les aéroports et aux frontières

La reconnaissance faciale est de plus en plus courante dans les aéroports du monde entier. De nombreux voyageurs utilisent des passeports biométriques qui leur permettent de ne pas faire la queue et de passer par un contrôle des passeports automatisé pour rejoindre plus rapidement leur terminal. La reconnaissance faciale réduit les files d'attente et permet aux aéroports d'améliorer la sécurité. Le Département de la Sécurité intérieure des États-Unis prédit que la reconnaissance faciale sera utilisée sur 97 % des voyageurs d'ici 2023 [13]. En plus des aéroports et des frontières, cette technologie sert à améliorer la sécurité lors d'événements à grande échelle comme les Jeux olympiques.

Retrouver des personnes disparues :

La reconnaissance faciale peut servir à trouver des personnes disparues et les victimes de la traite d'êtres humains. Imaginons que les personnes disparues soient ajoutées à une base de données. Les autorités peuvent alors être alertées dès que ces personnes sont repérées par la reconnaissance faciale, que ce soit dans un aéroport, un magasin ou un espace public [13].

Réduire la criminalité dans les magasins :

La reconnaissance faciale sert à identifier les voleurs, les criminels organisés ou les personnes connues pour frauder lorsqu'ils pénètrent dans un magasin [13]. Les photos des individus sont comparées dans de vastes bases de données de criminels pour que les professionnels de prévention des pertes et de sécurité soient avertis immédiatement lorsque des clients potentiellement dangereux entrent dans le magasin.

Améliorer les expériences d'achat :

Cette technologie a le potentiel d'améliorer l'expérience d'achat pour les clients. Par exemple, les kiosques dans les magasins peuvent reconnaître les clients, leur recommander des produits selon leurs achats précédents et leur indiquer où se rendre. Le paiement avec la reconnaissance faciale permet aux clients de ne plus faire la queue aux caisses [12].

Services bancaires :

Les services biométriques de banque en ligne sont possibles grâce à la reconnaissance faciale. Au lieu d'utiliser des mots de passe à usage unique, les clients peuvent autoriser des transactions simplement en regardant leur smartphone ou leur ordinateur [12]. Avec la reconnaissance faciale, les cybercriminels ne peuvent plus compromettre de mots de passe. Si des pirates dérobent votre photo, la détection du « vivant » – une technique qui sert à déterminer si la source de l'objet biométrique est un être humain vivant ou une fausse représentation – devrait (en théorie) les empêcher de l'utiliser pour usurper votre identité. La reconnaissance faciale pourrait rendre obsolètes les cartes de débit et les signatures.

Surveiller la présence des employés ou des élèves :

Certaines institutions scolaires en Chine utilisent la reconnaissance faciale pour vérifier que les enfants ne font pas l'école buissonnière. Elles utilisent des tablettes pour scanner le visage des élèves et les comparer aux photos d'une base de données pour valider leur identité. Plus largement, cette technologie peut être utilisée à l'entrée et à la sortie des lieux de travail afin que les employeurs puissent vérifier l'assiduité des employés [13].

Reconnaissance des conducteurs :

Selon ce rapport de consommateurs, les concessionnaires sont en train d'expérimenter avec la reconnaissance faciale afin de remplacer les clefs de voiture. Cette technologie remplacerait la clef pour entrer dans la voiture et la faire démarrer, et mémoriserait les préférences du conducteur concernant le siège et les rétroviseurs, ainsi que les stations de radio [64].

2.10) Avantages de la reconnaissance faciale :

En dehors du déverrouillage de votre smartphone, la reconnaissance faciale comporte d'autres avantages :

Une meilleure sécurité :

Au niveau gouvernemental, la reconnaissance faciale permet de mieux identifier les terroristes et autres criminels. Au niveau personnel, la reconnaissance faciale peut servir d'outil de sécurité pour verrouiller les appareils personnels et pour les caméras de surveillance des particuliers[63].

Diminution de la criminalité :

La reconnaissance faciale facilite le traçage des cambrioleurs, des voleurs et des intrus. La simple connaissance de l'existence de systèmes de reconnaissance faciale a un effet dissuasif, surtout dans le cadre d'infractions mineures. En plus de la sécurité physique, elle apporte aussi des avantages pour la cyber sécurité. Les entreprises peuvent utiliser la reconnaissance faciale pour remplacer les mots de passe qui sécurisent les ordinateurs [13]. En théorie, cette technologie ne peut pas être piratée puisqu'il n'y a rien à voler ou à modifier, comme c'est le cas pour les mots de passe.

Rationaliser la fouille des suspects :

Les inquiétudes du public concernant les fouilles injustifiées sont un sujet controversé pour la police : la reconnaissance faciale pourrait améliorer ce processus. En isolant des suspects dans les foules grâce à un processus qui serait automatisé au lieu d'être humain, la reconnaissance faciale pourrait réduire les arrestations pour délit de faciès et réduire les fouilles effectuées sur les citoyens respectueux de la loi [13].

Un traitement plus rapide :

Le processus de reconnaissance du visage ne dure qu'une seconde, ce qui représente un avantage pour les entreprises qui utilisent la reconnaissance faciale. Dans notre ère de cyberattaques et d'outils de piratage avancés, les entreprises ont besoin de technologies à la fois sécurisantes et rapides. La reconnaissance faciale permet de vérifier rapidement et efficacement l'identité des personnes [64].

Intégration avec d’autres technologies :

La plupart des solutions de reconnaissance faciale sont compatibles avec les grands logiciels de sécurité. En fait, elles sont très faciles à intégrer. Cela permet de réduire les dépenses liées à leur intégration [64].

2.11) Inconvénients de la reconnaissance faciale :

Si cela ne dérange pas certains d'être filmé en public et que la reconnaissance faciale soit utilisée pour de bonnes raisons, cette technologie peut également créer des réactions vives chez d'autres. Voici certains inconvénients et les problèmes que peut poser la reconnaissance faciale :

La surveillance

Certains s'inquiètent de l'utilisation de la reconnaissance faciale et de l'omniprésence des caméras de surveillance, de l'intelligence artificielle et des analyses de données qui créent une surveillance de masse potentielle et restreignent les libertés individuelles. Si la reconnaissance faciale permet aux gouvernements de retrouver des criminels, elle peut aussi leur permettre de pister des gens ordinaires et innocents à tout moment [13].

Violation de la vie privée

Le problème de l'éthique et de la vie privée se pose encore et toujours. On sait que certains gouvernements possèdent des photos de plusieurs citoyens sans leur consentement. En 2020, la Commission européenne a laissé entendre qu'elle envisageait d'interdire la reconnaissance faciale dans les espaces publics pour les cinq années à venir, pour avoir le temps d'établir un cadre réglementaire afin d'empêcher les abus au niveau de l'éthique et de la confidentialité[63].

Stockage massif des données :

La reconnaissance faciale repose sur le Machine Learning, une technologie qui nécessite d'énormes ensembles de données pour « apprendre » à donner des résultats précis. De tels ensembles de données ont besoin d'un stockage tout aussi massif. Les petites et les moyennes entreprises n'ont souvent pas les ressources suffisantes pour stocker les données nécessaires [64].

Attaques biométriques par le visage :

Ce type d'attaque se produit au niveau du capteur lors de l'acquisition, et elle consiste principalement à présenter des faux visages à la caméra [1]. Ces faux visages peuvent être créés à l'aide d'une image d'une autre personne (imprimée sur papier ou affichée sur un écran), une vidéo ou un masque 3D (spofing).

Ce type d'attaques peuvent être réalisées par trois méthodes : Attaques par photos, attaques par vidéos, attaques par masques 3D.

Les recherches pour trouver des contres mesures sont toujours d'actualité.

2.12) Conclusion :

Dans ce chapitre nous avons parlé sur la détection de visage et ses étapes, nous avons aussi abordé la reconnaissance faciale, son fonctionnement et les différents domaines d'utilisation de la reconnaissance faciale, puis cité des avantages et inconvénients de la reconnaissance faciale.

Actuellement, il existe de nouvelles attaques dans les systèmes de reconnaissance des personnes par leurs visages, ou même dans d'autres modalités biométriques qui nécessite pas l'accès à l'intérieur du système interne. Cependant, ils peuvent être exécutés de l'extérieur ; elles interviennent au niveau du capteur lors de la phase d'acquisition. Dans le cas d'un système de reconnaissance faciale, ces attaques consistent principalement à présenter des faux visages à la caméra.

Dans le chapitre suivant, on parlera des attaques (spoofing en anglais) et les techniques principales de spoofing dans les systèmes de reconnaissance du visage seront présentées dans un premier temps. On présentera les différentes méthodes proposées au cours de ces dernières années pour contrer ces attaques dans un second temps. Enfin, parler sur notre approche qui est la détection de la vivacité de visage en utilisant les réseaux de neurones convolutifs (CNN) comme étant une mesure de sécurité contre les attaques et le piratage.

3) CHAPITRE 3 : les attaques biométriques par le visage et solutions existantes

3.1) INTRODUCTION :

Auparavant le piratage de tout type de systèmes de reconnaissance de personnes se faisait à l'intérieur du système en modifiant les données biométriques pour permettre aux personnes non autorisées d'accéder au système ou pendant la communication en remplaçant les décisions négatives d'authentification. Afin de se protéger contre ce type d'attaques les recherches étaient focalisées sur la sécurisation des données sauvegardées dans la base de données et protection des communications par des algorithmes de cryptage.

Les nouvelles attaques dans les systèmes de reconnaissance de personnes par visage par exemple ou une autre modalité biométrique ne nécessitent pas l'accès au fonctionnement interne du système, mais peuvent être effectuées de l'extérieur ; elles se font au niveau du capteur pendant la phase d'acquisition.

Dans le cas d'un système de reconnaissance faciale, ces attaques consistent principalement à présenter à la caméra de faux visages.

Ces visages peuvent être réalisés à l'aide d'une image d'une autre personne (imprimée sur papier ou affichée sur un écran), une vidéo ou un masque 3D.

Nous avons choisi d'organiser notre chapitre en deux parties principales :

Dans la première partie, nous allons diviser les axes de recherche dans le domaine du spoofing (piratage) dans les systèmes de reconnaissance du visage en trois axes selon les types d'attaques, puis parler des attaques (spoofing en anglais) et les techniques principales de spoofing dans les systèmes de reconnaissance du visage ensuite, on présentera les différentes méthodes proposées au cours de ces dernières années pour contrer ces attaques.

Dans la seconde partie, nous aborderons la détection de la vivacité de visage et ses différentes approches, nous passerons également en revue notre approche de détecteur de vivacité où nous

allons expliquer l'architecture des réseaux de neurones profonds capable de distinguer les vrais des faux visages.

3.2) Attaques au niveau du capteur :

Cette catégorie d'attaque se produit au niveau du capteur lors de l'acquisition, ces attaques consistent principalement à présenter devant la caméra de faux visages, ce type d'attaque peut être réalisée par trois méthodes :

3.2.1) Attaques par photos :

La première méthode pour pirater un système sécurisé utilisant la reconnaissance faciale consistait à utiliser une image du visage de la victime. Comme le montre la figure 3.1 une personne peut utiliser son visage comme un mot de passe pour activer son ordinateur, mais d'autres utilisateurs seront tout simplement rejetée car ils n'ont pas le même visage [1].

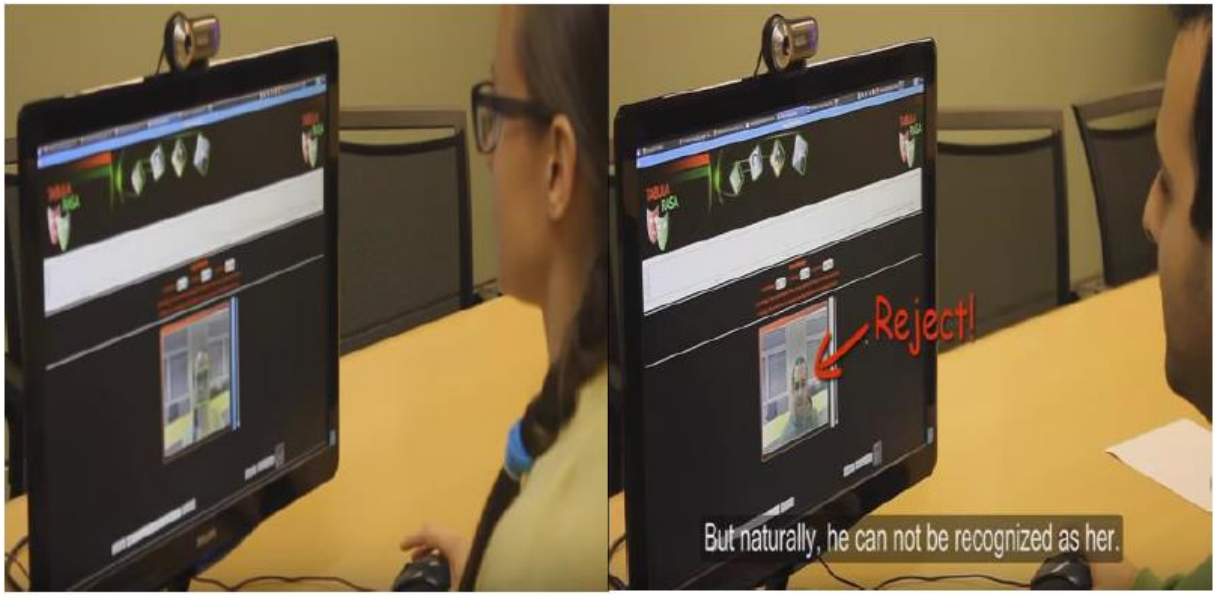


Figure 3-1 Système sécurisé par la reconnaissance du visage [1].

Afin de tromper le système d'authentification un pirate peut simplement transporter le visage valable pour activer le système sur un support en papier, comme le montre l'image 3.2.

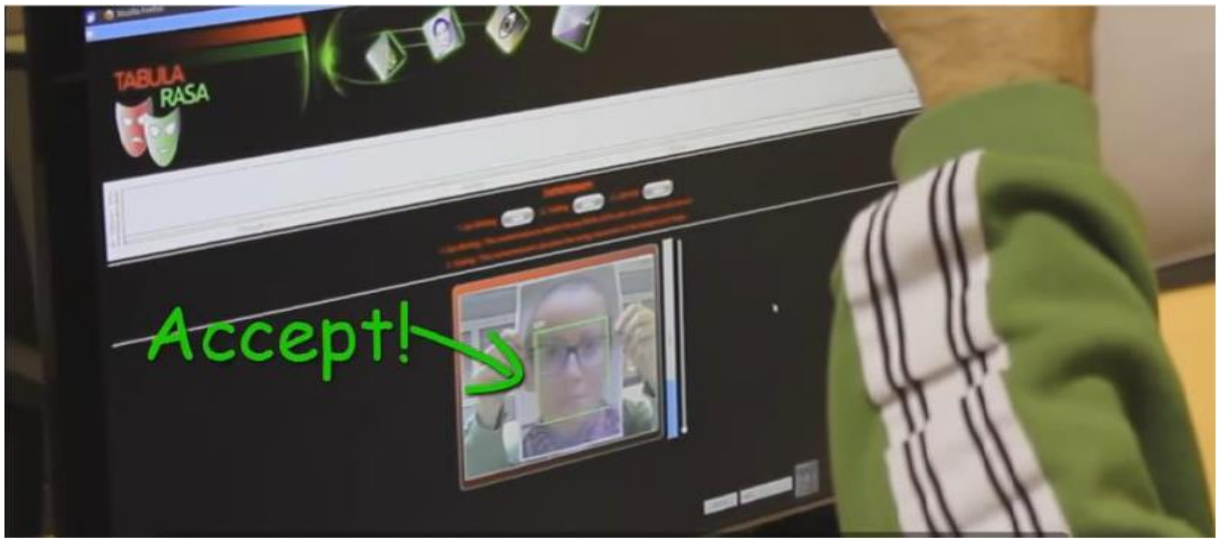


Figure 3-2 Piratage d'un système sécurisé par la reconnaissance du visage avec une photo [1].

Pour résoudre le problème de piratage par image, la plupart des recherches se sont focalisées sur la détection de la vivacité du visage, en se basant sur les clignements des yeux [3] comme le montre la figure 3.3, ou par l'analyse de texture [20]. Pour distinguer entre un visage réel et une usurpation d'identité par photos les méthodes de détection de spoofing peuvent être classées en deux classes :

- Détection de mouvement ou détection de vivacité.

- Analyse de texture ou détection de la qualité d'image.

Pour contourner l'authentification système, un pirate peut simplement transmettre le valide visage pour activer le système sur un papier support, comme vu dans la figure 3.3.

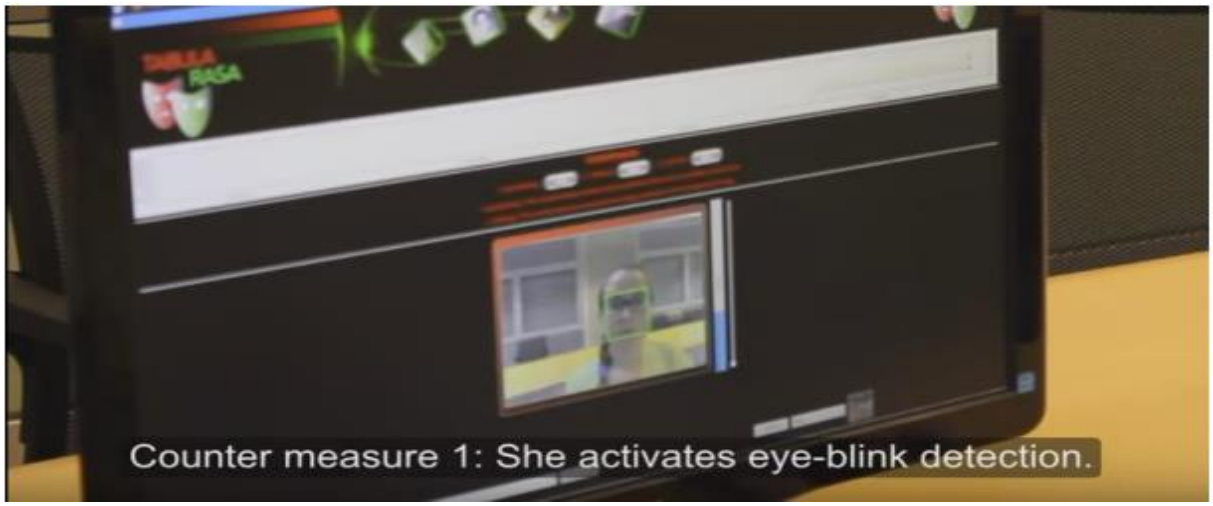


Figure 3-3 Neutralisation du piratage avec photo, par la détection des clignements des yeux, dans un système sécurisé par la reconnaissance du visage [1].

Dans leur travail, Jee et al. [2] ont détecté les yeux dans les images d'entrée séquentielles et ont calculé la variation de chaque région oculaire pour déterminer si le visage d'entrée est un vrai visage ou non. Après normalisation de la région du visage, les régions des yeux sont extraites avec une taille 10x20 en fonction du centre des yeux. Ensuite, les régions oculaires sont binarisées en utilisant un seuil qui est calculé à partir de la valeur moyenne des pixels de chaque région oculaire.

La Figue 3.4 montre des exemples des régions d'œil binarisées extraites de 5 images séquentielles pour des visages réels et des visages faux. Comme le montre cette figure, les régions oculaires des faux visages changent très peu, mais les régions oculaires des visages réels ont une variation beaucoup plus grande, en raison du clignotement ou du mouvement de la pupille.

Figure 3-4 Exemple de régions œil binarisées de faux visages (a) et de visage réels (b)[2].

Toujours pour la détection de la vivacité Pan et al. Dans [3] ont proposé une méthode qui analyse le comportement du clignotement des yeux, qui est représenté comme une séquence d'images temporelles après avoir été capturée numériquement par la caméra. Ils ont proposé un champ aléatoire conditionnel (CRF) pour modéliser une suite d'observations, en supposant qu'il existe une séquence sous-jacente d'états tirés d'un ensemble d'états finis. Une activité de clignotement peut être représentée par une séquence d'image S composée de T images, où S = {Ii , i = 1, ..., T}. Typiquement, les yeux dans les images s'ouvrent et se ferment, en outre il y a un état ambigu lorsque l'œil clignote en passant d'un état ouvert à un état fermé ou d'un état fermé à un état ouvert. Ils ont défini un ensemble de trois états pour les yeux. Q = {o: open, c: close, b: ambiguous}.

Ainsi, une activité de clignotement typique comme montre la figure 3.5 peut être décrite comme un motif de changement d'état de o → b → c → b → o, les cercles dans la figure3.5 représente les états et les rectangles représente les transitions entre états.

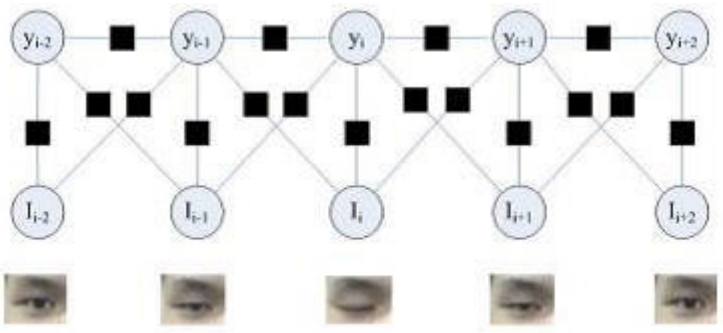


Figure 3-5 Modèle graphique d'un CRF à chaîne linéaire [3].

A partir de ce modèle ils peuvent vérifier s'il y a un clignement des yeux dans les séquences d'images capturées ou non. Les performances de la méthode proposée sont très bonnes, avec 1% d'échecs de détection d'attaques par photos. Les autres approches utilisent l'analyse de texture et la détection de la qualité de l'image. Les méthodes dans cette catégorie tentent de mesurer les différences dans les détails d'images issues des visages réels et celles issues des écrans d'ordinateur ou du papier. Cela repose sur de multiples hypothèses, telles que : les images récapitulatives entraînent une diminution de la qualité, la réflectance de la lumière n'est pas la même entre les différentes surfaces, ou que l'impression sur papier crée des artefacts détectables.

Maatta et al. Dans leurs travaux [20] adoptent les modèles binaires locaux (LBP) introduis par Ojala et al. [17] pour extraire les vecteurs caractéristiques. Le LBP est un opérateur de texture puissant, pour décrire non seulement les micros textures mais aussi leur information spatiale. Les vecteurs dans l'espace caractéristique alimentent alors un classificateur SVM qui détermine si les modèles de micro-texture caractérisent une personne réelle ou une fausse image.

L'opérateur LBP original forme des étiquettes pour les pixels d'image en comparant chaque pixel aux pixels dans voisinage 3 × 3; les résultats de la comparaison, qui sont soit 0 soit 1, sont regroupés pour former un code binaire pour chaque pixel. La figure 3.6 montre un exemple de calcul LBP.

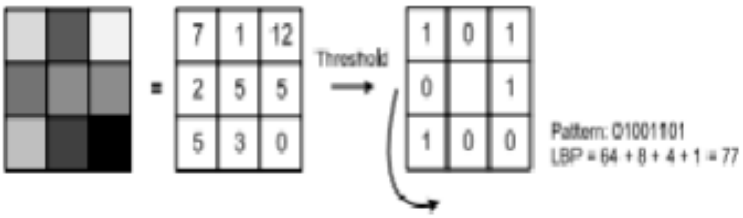


Figure 3-6 calcul LBP pour un pixel [20].

La Figue 3.7 montre des exemples de deux images (un visage réel et une photo) dans l'espace d'origine et les images LBP correspondantes en utilisant le LBP de base.

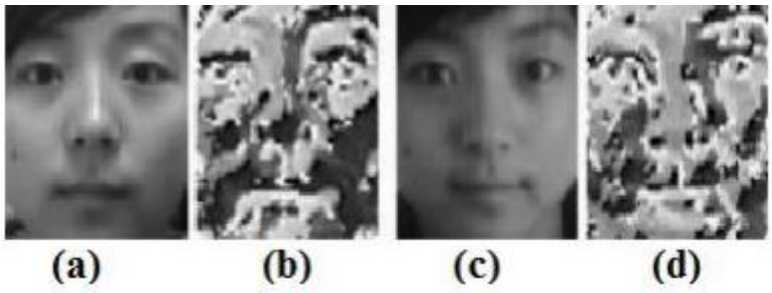


Figure 3-7 Deux images : un vrai visage(a), une photo(c) et leurs images LBP correspondantes (b, d) [20].

Maatta et al. Ont remarqué que la photo imprimée semble assez similaire à l'image du visage originale alors que les images LBP présentent certaines différences. Le taux d'erreur de discrimination entre des images de visages réels et de faux visages obtenu est égal à 2.9 %.

Dans le même contexte Gautam et Jayash dans [21] ont proposé une technique qui exploite la différence dans les caractéristiques d'illumination entre le visage réel et le faux visage. Ils ont montré que la lumière qui tombe sur un visage réel se reflète au hasard dans différentes directions en raison de sa surface 3D (lèvre, nez, etc.), par contre la lumière se reflète uniformément sur une surface planaire 2D, telle que celle d'une photo, comme le montre la Figure 3.8.

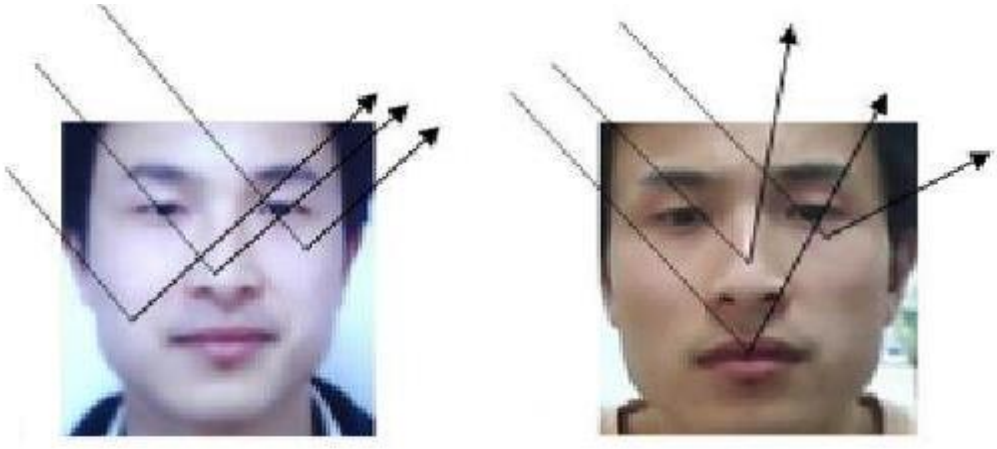


Figure 3-8 Caractéristiques d'illumination d’un faux visage et un vrai visage [21].

Le modèle local distribué (LDP) est suffisamment capable d'obtenir les petites différences entre les cartes de vitesse de diffusion des vrais et des faux visages [21].

Avec cette méthode le taux de reconnaissance des visages réels est de 88.67%, et 92.53% pour les faux visages.

3.2.2) Attaques par vidéo :

Afin de tromper le système d’authentification dans ce type d’attaques, les pirates utilisent des supports vidéo comme (smartphone, ipad) sur lequel il y a un enregistrement vidéo d’une séquence d’un visage valable pour activer le système comme montre la figure 3.9.

Figure 3-9 Piratage d’un système sécurisé par la reconnaissance du visage avec vidéo [1].

Les méthodes pour détecter le spoofing par photo qui utilisent la détection de la vivacité du visage ou le clignement des yeux ont échoué face à ce type de piratage comme le montre la Figure 3.10.

Figure 3-10 Défaillance du système de vérification par l'approche de clignement des yeux contre les attaques par vidéo [1].

Pour résoudre le problème de piratage par vidéo, la plupart des recherches se sont focalisés sur l'analyse du mouvement du visage pour distinguer les mouvements 3D d'un visage réel des mouvements 2D d'un faux visage. La méthode proposée par Wang et al [26] repose sur l'hypothèse que lorsqu'un vrai visage tourne, les points de repère sur le visage se déplacent d'une manière différente par rapport à une face imprimée ou un support vidéo (voir la figure 3.11).

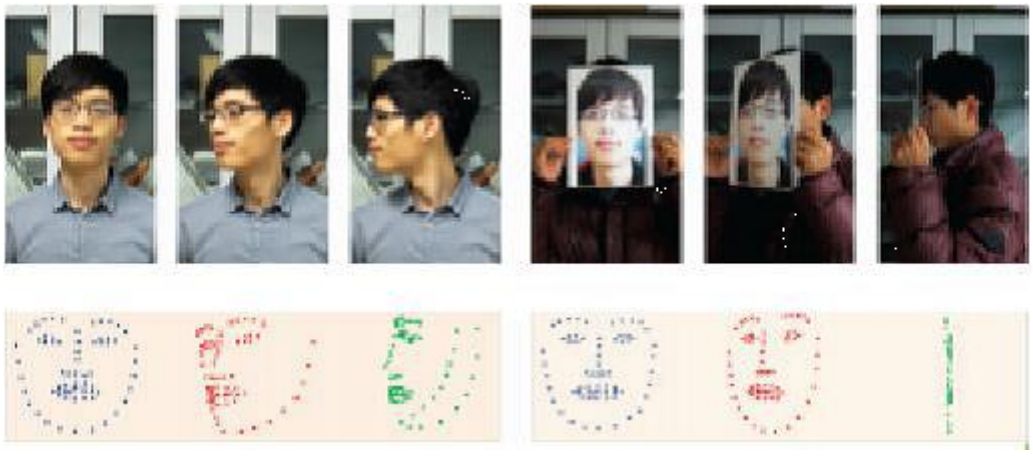


Figure 3-11 Comparaison de la façon dont les repères se comportent entre un visage authentique et la photo d'un visage [26].

Ils repèrent tout d'abord l'emplacement du visage à l'aide d'un détecteur global ensuite ils déterminent l'emplacement global des repères après ils font effectuer environ quatre itérations des opérations suivantes :

– Obtenir la partie de l'image entourant un point de repère.

– Comparer la position aux positions d'un modèle de forme global, connu pour ce point de repère.

– Améliorer le positionnement du repère.

– Confirmer la forme nouvellement proposée de tous les points de repère.

Dans le travail réalisé par Fokkema[28], l'utilisateur dois gagner un challenge, la détection du spoofing dépend de l'utilisateur qui doit pivoter la tête en suivant un motif donné avec un schéma spécifique qui sont générés aléatoirement ce qui rend le défi plus robuste, réussir à faire ce défi dans le temps donné prouve que l'utilisateur est authentique, par contre s'écarter du schéma ou manquer de temps va faire perdre le défi donc l'utilisateur est considérer comme étant non authentique, un exemple d'un défi est représenté dans la Figure 3.12.

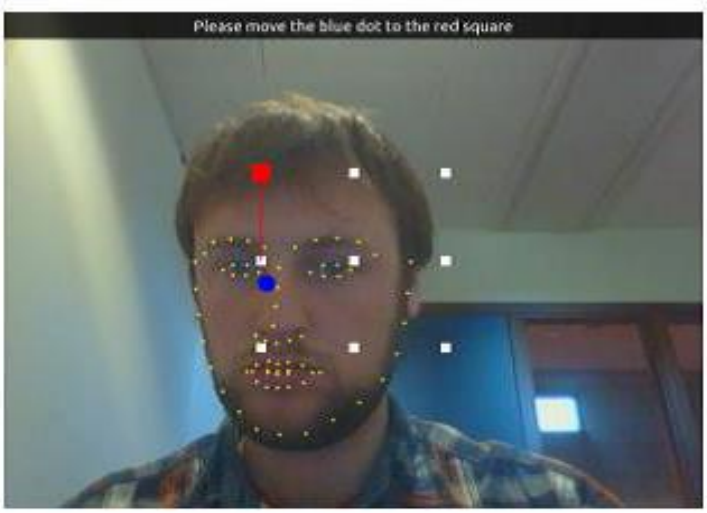


Figure 3-12 Exemple de la façon dont le défi est présenté à l'utilisateur [28].

Dans leur travail W. Di et al. [22] proposent d'effectuer une détection du spoofing par support vidéo à l'aide de l'analyse de distorsion d'image (IDA). Quatre types de caractéristiques IDA (réflexion spéculaire, flou, moments de couleur et diversité des couleurs) ont été conçus pour capturer la distorsion de l'image. Les quatre caractéristiques sont concaténées ensemble, pour former un vecteur de fonctionnalité IDA d'une dimension égale à 121, ensuite un classificateur

d'ensemble composé de deux classificateurs SVM conçus pour différentes attaques est utilisé pour la discrimination entre les visages authentiques et les visages sur un support mobile.

Ils ont également construit une base de données, appelée MSU MFSD [24], en utilisant deux appareils mobiles (Android Nesus 5 et MacBook Air 13). Il s'agit de la première base de données de la parodie mobile. Un sous-ensemble de cette base de données, composé de 35 sujets est mis à la disposition du public [22].

W. Di et al. [22] ont obtenu un taux d'erreur total moyen (HTER en anglais) égal à 7.42%. Le HTER (Half Total Error Rate) est la moyenne entre le Taux de Faux Rejets (FRR, pour False Rejection Rate en anglais) et le Taux de Fausse Acceptation (FAR, pour False Acceptance Rate, en anglais) [14].

3.2.3) Attaque par masque 3D :

Afin de tromper le système de reconnaissance faciale ce troisième type d'attaques consiste à utiliser devant la caméra un masque 3D du visage d'une personne appartenant à la base de données du système comme le montre la Figure 3.13.

Figure 3-13 Piratage avec Masque 3D d'un système sécurisé par la reconnaissance du visage [27].

Les contres mesures se basant sur le clignement des yeux peuvent être vaincu seulement en laissant les yeux apparentes juste en faisant des trous, Aussi les contremesures basées sur les mouvements dépendent des différences entre les mouvements des surfaces 2D et 3D et ne s'appliquent pas lorsque des masques sont utilisés à la place de photos ou de vidéos.

C'est pour cela que les chercheurs doivent trouver d'autres méthodes afin de contrer les attaques avec un masque 3D.

L’un des premiers travaux est celui de Kim et al [27],

où ils ont introduit un nouvel espace de fonctionnalité où les vrais et les faux visages peuvent être distingué, leur méthode repose sur la disparité de réflectance basée sur l’albédo entre la peau réelle et les matériaux spécifiques avec lesquels on fait les masques, le vecteur de caractéristiques utilisé pour la classification est composé des mesures de rayonnement de la région du front sous des illuminations de 850 à 950 nm, les peaux et les matériaux sont linéairement séparable dans l’espace de caractéristiques proposé comme le montre la Figure 3.14.

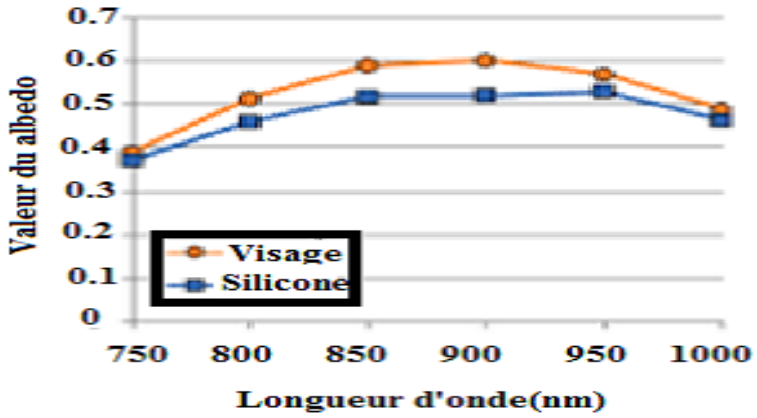


Figure 3-14 Réflectance albédo pour une peau réelle et une peau en silicone [27].

Dans leurs expériences ils ont obtenu une précision de 97.78% dans la détection dans la face fausse, mais ces expériences sont effectuées sur les matériaux non sur des vrais masques, mais cette méthode est peu pratique car des radiations doivent être utilisées pour se situer à 30cm de l’émetteur des rayons, ajouter à ça la possibilité d’occlusion frontale ainsi que la limitation de portée.

N. Kose et J-L. Dugelay[15] ont calculé les performance de spoofing par masque en utilisant des données 3D issues du balayage 3D par un scanner, dans leur étude ils ont utilisé une base de données crée par MORPHO, cette base de données est composée de masques de sujets réels de haute qualité, les scan des sujets ont été acquis avec un scanner 3D, ensuite les masques sont imprimés avec une imprimante 3D, la base contient des images de texture et des balayage 3D pour les visage réels et les masques. La Figure 3.15 montre un exemple à partir de la base de données créée par MORPHO.

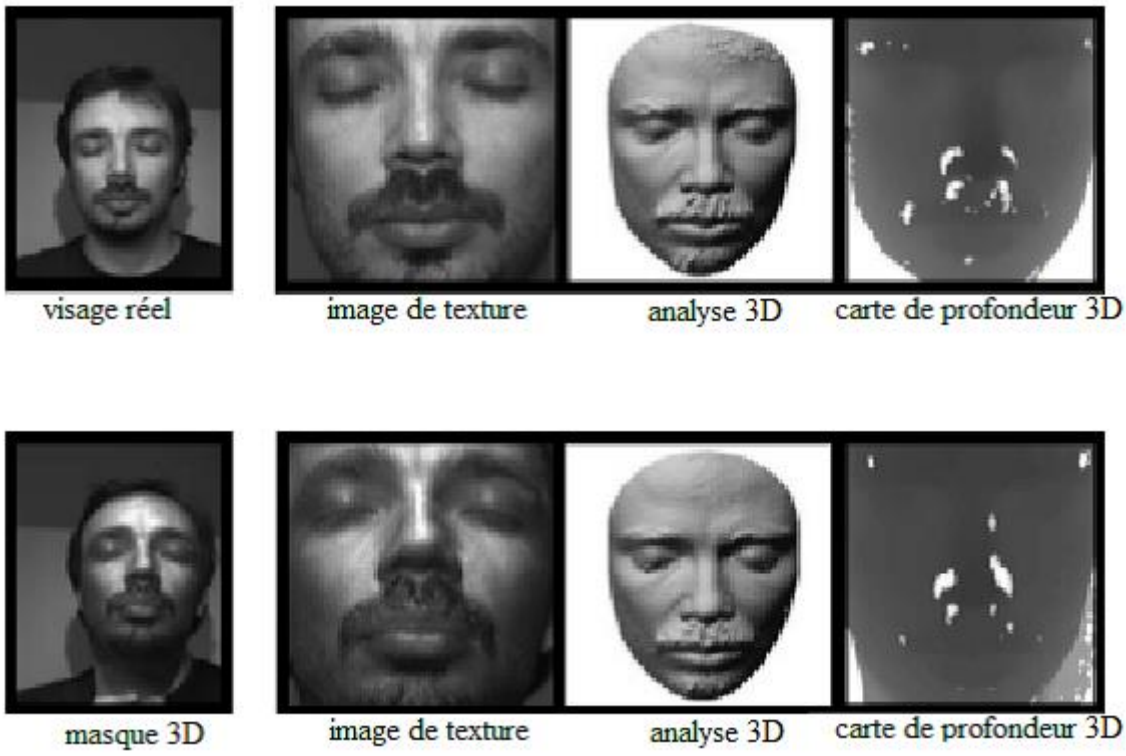


Figure 3-15 Exemple à partir de la base de données de masques créée par MORPHO [15].

Cette base leur a permis d´évaluer l'impact de spoofing par masque en 2D et en 3D afin de développer des contres mesures, les résultats montrent que les masques utilisés ont une texture très similaire à celle des visages réel, les masques sont efficace pour tromper les systèmes de reconnaissance de visage, dans leur travail les auteurs ont appliqué une analyse de micro-texture sur des images en niveau de fris et des images en profondeur par la méthode LBP, la précision de la classification avec les images en niveau de gris est de 88.12% et avec les images en profondeur est de 86% .

Pour obtenir un meilleur résultat les auteurs ont mené une autre étude où ils ont fusionné les informations extraites des images de texture et de profondeur et ils ont obtenu un résultat de 93.5%.

3.3) Détection de la vivacité :

La capacité de détecter les usurpant de situation et les comportements frauduleux est essentielle à toute technique biométrique faciale. La forme la plus courante d'usurpé est de présenter une image statique précédemment obtenue d'un individu pour la comparaison avec l'image de source de confiance.

Afin de contrer cela et d'assurer la présence de la personne, une certaine forme de détection de vie peut être utilisée.

De nombreuses méthodes différentes de détection de la vivacité sont disponibles sur le marché aujourd'hui. La forme la plus courante de détection de la vivacité demande à l'utilisateur d'effectuer une série de mouvements de la tête pour prouver la vivacité. Les techniques plus avancées, telles que la reconnaissance 3D et l'imagerie thermique, nécessitent du matériel spécialisé et ne conviennent pas aux applications commerciales quotidiennes.

3.4) Etat de l'art de la détection de vivacité

La reconnaissance des visages est aujourd'hui considérée comme l'une des meilleures approches pour identifier une personne, car elle ne nécessite aucune intervention humaine dans le processus de reconnaissance des visages.

L'architecture des réseaux neuronaux peut être dérivée sous la forme d'une simple portion du neurone d'un cerveau humain. Il s'agit nécessairement d'unions pondérées et d'intégration d'entrées qui procèdent par de nombreuses fonctions non séquentielles ou non linéaires. Ces réseaux neuronaux sont basés sur une méthode d'apprentissage itérative, communément appelée rétropropagation et optimiseur (tel que la descente de gradient stochastique (SGD)). Les réseaux neuronaux profonds (DNN) sont basés sur des architectures de réseaux neuronaux simples et se composent de plusieurs couches cachées. Ces réseaux sont très couramment utilisés pour la catégorisation. Les réseaux neuronaux convolutifs (CNN) présentent un type de proposition structurelle tout à fait différent de l'étude des réseaux neuronaux. L'objectif principal du CNN est d'utiliser des réseaux feed-forward couplés à des couches convolutionnelles qui consistent en des couches de mise en commun locale et globale. A. Krizhevsky a utilisé le CNN, mais il a utilisé des couches convolutionnelles bidimensionnelles couplées à un espace caractéristique bidimensionnel de l'image [50]. Les couches convolutionnelles unidimensionnelles sont utilisées pour les textes et les séquences avec l'intégration des mots comme espace caractéristique d'entrée. Les réseaux neuronaux récurrents (RNN) sont le dernier type d'architecture d'apprentissage profond où les sorties des neurones sont renvoyées dans le réseau comme entrées pour l'étape suivante. Cette architecture a été utilisée efficacement pour le traitement du langage naturel (NLP) [53].

Viola et Jones ont été les premiers à en proposer un qui était basé sur l'application de boîtes rectangulaires pour un visage. Cependant, il présentait de nombreuses limites, car la taille de ses caractéristiques était importante. Le nombre total d'éléments de type Haar_ était de 0,16 million dans une image 24 × 24 et, de plus, il ne traitait pas les visages sauvages et les visages frontaux. Après avoir analysé et identifié le problème ci-dessus, des personnes de diverses spécialisations ont déployé beaucoup d'efforts pour commencer à utiliser des caractéristiques complexes supplémentaires. Dlib est une autre méthode bien connue de détection des visages qui utilise la machine à vecteurs de support comme classificateur. En outre, la combinaison de plusieurs détections qui doivent être entraînées séparément dans différentes vues est l'une des méthodes les plus simples [51].

Des modèles déformables multiples ont été appliqués par Zhu et al pour capturer des visages dans des vues contrastées. De même, Shenet et al ont présenté un modèle basé sur la récupération et intégré à différents apprentissages. Ces modèles ont nécessité une formation et des tests qui ont pris plus de temps et ont été moins efficaces. Garcia et al ont développé un réseau neuronal en 2002 pour localiser les visages humains semi-frontaux dans des images complexes. L'approche décrite dans cet article prend en considération toutes ces idées et théories et l'approche proposée fournit des résultats bien meilleurs et plus précis [51].

La reconnaissance d'objets est considérée comme l'une des technologies informatiques les plus populaires, qui intègre le traitement d'images et la vision par ordinateur. Elle est liée à la détection d'exemples d'une entité comme les visages humains, les bâtiments, les arbres, les voitures, etc. L'objectif principal des algorithmes de reconnaissance des visages est de décider et d'analyser si un visage existe ou non dans une image.

Depuis des décennies, beaucoup d'études et de travaux ont été réalisés dans le domaine de la reconnaissance et de la détection des visages pour les rendre plus avancés et plus précis, mais une révolution a eu lieu dans ce domaine lorsque Viola-Jones a mis au point un détecteur de visages en temps réel, capable de détecter les visages en temps réel avec une grande précision. L'étape initiale et primordiale de la reconnaissance des visages est la détection des visages, qui est utilisée pour détecter les visages dans les images. Elle fait partie intégrante de la reconnaissance d'objets et peut être utilisée dans plusieurs domaines importants tels que la sécurité, la biométrie, le droit, l'analyse statistique, le divertissement, la sûreté, etc. Elle est utilisée pour reconnaître les visages en temps réel afin de surveiller et de suivre l'emplacement d'une personne ou d'autres entités. Elle est le plus souvent utilisée dans les appareils photo

mobiles et les reflex numériques pour identifier des apparitions multiples dans le cadre. Le meilleur exemple qui peut être tiré d'une application quotidienne populaire est Facebook, qui utilise également des algorithmes de reconnaissance des visages pour détecter les visages dans les images et les reconnaître [52].

L'objectif principal de l'approche proposée est de reconnaître les données du visage de différentes personnes et de les identifier avec des résultats précis et de manière efficace ou nous allons entraîner un réseau de neurones profond capable de distinguer les vrais des faux visages.

3.5) Apprentissage profond :

L'apprentissage profond (Deep Learning) est un nouveau domaine de recherche du machine learning (ML), qui a été introduit dans le but de rapprocher le ML de son objectif principal : l'intelligence artificielle, Il concerne les algorithmes inspirés par la structure et le fonctionnement du cerveau. Ils peuvent apprendre plusieurs niveaux de représentation dans le but de modéliser des relations complexes entre les données [55].

La figure 3.16 montre La relation entre l'intelligence artificielle, le ML et le deep learning.

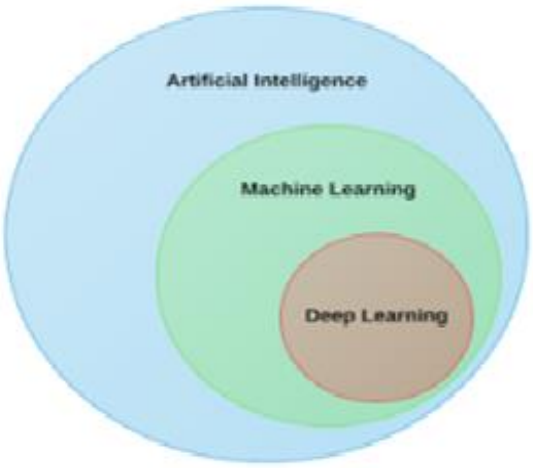


Figure 3-16 La relation entre l'intelligence artificielle, le ML et le deep learning [56].

Le Deep Learning est basé sur l'idée des réseaux de neurones artificielles et il est taillé pour gérer de larges quantités de données en ajoutant des couches au réseau. Un modèle de deep learning a la capacité d'extraire des caractéristiques à partir des données brutes grâce aux multiples couches de traitement composé de multiples transformations linéaires et non linéaires et apprendre sur ces caractéristiques petites à petit à travers chaque couche avec une intervention humaine minime [56].

Sur les cinq dernières années, le deep learning est passé d'un marché de niche ou seulement une poignée de chercheurs s'y intéressait au domaine le plus prisé par les chercheurs. Les recherches en relation avec le deep learning apparaissent maintenant dans les top journaux comme Science [29], Nature [30] et Nature Methods [31] pour ne citer que quelques-uns. Le deep learning a coquerie le jeu de GO [32], appris à conduire une voiture [33], diagnostiquer le cancer [34] et l'autisme [35] et même devenu un artiste [36].

Le terme "Deep Learning" a été introduit pour la première fois au ML par Dechter (1986) [37], et aux réseaux neuronaux artificiels par Aizenberg et al (2000) [38].

3.6) Apprentissage automatique traditionnel VS apprentissage profond :

La principale différence entre l'apprentissage automatique traditionnel et les algorithmes d'apprentissage en profondeur réside dans l'ingénierie des fonctionnalités. Dans les algorithmes d'apprentissage automatique traditionnels, nous devons créer les fonctionnalités à la main. En revanche, dans les algorithmes d'apprentissage en profondeur, l'ingénierie des fonctionnalités est effectuée automatiquement par l'algorithme. L'ingénierie des fonctionnalités est difficile, prend du temps et nécessite une expertise du domaine. La promesse de l'apprentissage en profondeur réside dans des algorithmes d'apprentissage automatique plus précis que l'apprentissage automatique traditionnel avec peu ou pas d'ingénierie de fonctionnalités.

Figure 3-17 Le procédé du ML classique comparé à celui du deep learning.

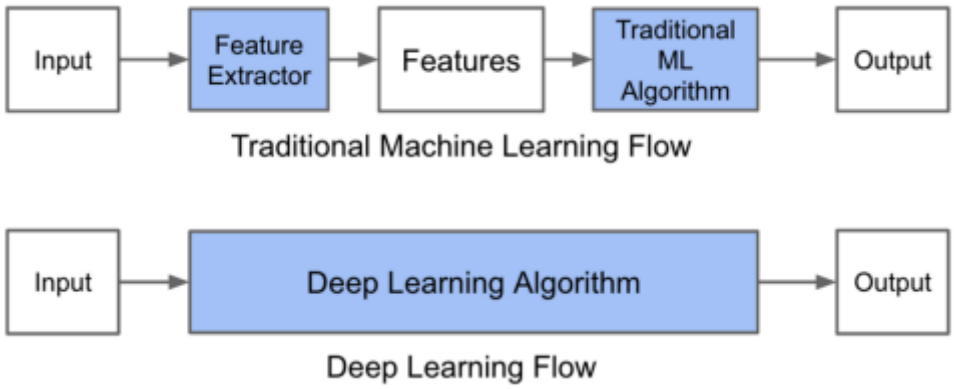


Figure 3-17 Le procédé du ML classique comparé à celui du deep learning [56].

3.7) Réseaux de neurones convolutifs :

Réseaux de neurones convolutifs (Convolutional Neural Network) CNN sont un type de réseau de neurones spécialisés pour le traitement de données ayant une topologie semblable à une grille. Les exemples comprennent des données de type série temporelle, qui peuvent être considérées comme une grille 1D en prenant des échantillons à des intervalles de temps réguliers et des données de type image, qui peuvent être considérées comme une grille 2D de pixels. Les réseaux convolutifs ont connu un succès considérable dans les applications pratiques. Le nom « réseau de neurones convolutif » indique que le réseau emploie une opération mathématique appelée convolution. La convolution est une opération linéaire spéciale. Les réseaux convolutifs sont simplement des réseaux de neurones qui utilisent la convolution à la place de la multiplication matricielle dans au moins une de leurs couches.

Ils ont de larges applications dans la reconnaissance de l'image et de la vidéo, les systèmes de recommandations [39] et le traitement du langage naturel [40].

3.8) Couche de convolution :

La convolution s'appuie sur trois idées importantes qui peuvent aider à améliorer un système de Machine Learning (ML) : Interactions éparses (sparse interactions), partage de paramètres (parameter sharing) et représentations équivariantes (equivariant representations)[59].

• **Interactions convolutives** (Sparse interactions) : Les réseaux de neurones traditionnelles utilisent la multiplication matricielle par une matrice de paramètres avec des paramètres séparés décrivant l'interaction entre chaque unité d'entrée et chaque unité de sortie. Cela signifie que chaque unité de sortie interagit avec chaque unité d'entrée ce qui n'est pas le cas des réseaux de neurones convolutifs. Ceci est accompli en rendant le noyau plus petit que l'entrée. Par exemple, lors du traitement d'une image, l'image d'entrée peut avoir des milliers ou des millions de pixels, mais nous pouvons détecter de petites caractéristiques significatives telles que les bords avec des noyaux qui n'occupent que des dizaines ou des centaines de pixels. Cela signifie que nous pouvons stocker moins de paramètres, ce qui réduit les besoins en matière de mémoire du modèle et améliore son efficacité. Cela signifie également que le calcul des sorties nécessite moins d'opérations. Ces améliorations en matière d'efficacité sont généralement assez importantes. Figure 3.18.

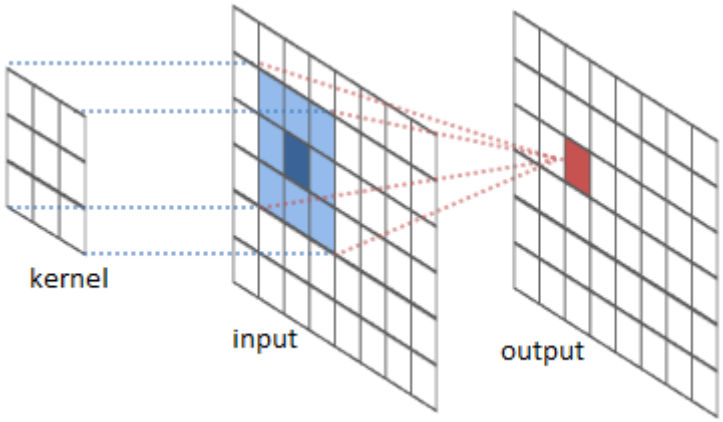


Figure 3-18 Exemple d'une convolution 2D [42].

Pour calculer la valeur du pixel (x,y) de l'image output (en rouge) on superpose le centre du kernel (ici une matrice 3×3) sur le pixel source en position (x,y) (en bleu foncé) et on multiplie deux à deux les valeurs du kernel et de la zone couverte en bleue entre elles puis on les additionnent pour trouver la valeur de la sortie.

La valeur du pixel sera donnée, dans le cas d'un noyau 3×3, par la formule :

$$\begin{aligned} &O(x,y) = \\ &K(0,0) \times I(x-1,y-1) + K(0,1) \times I(x,y-1) + K(0,2) \times I(x+1,y-1) + \\ &K(1,0) \times I(x-1,y) + K(1,1) \times I(x,y) + K(1,2) \times I(x+1,y) + \\ &K(2,0) \times I(x-1,y+1) + K(2,1) \times I(x,y+1) + K(2,2) \times I(x+1,y+1) \end{aligned} \quad (4)$$

Que l'on peut aussi écrire :

$$O(x,y) = \sum_{i=0}^{2} \sum_{j=0}^{2} K(i,j) \times I(x-1+j, y-1+i) \quad (5)$$

Il reste ensuite à diviser le résultat par le nombre d'éléments du noyau, cette dernière opération n'appartient pas au produit de convolution proprement dit, mais elle est nécessaire pour maintenir la dynamique de l'image ainsi que sa linéarité.

Cette formule est à appliquer sur les trois canaux RVB.

• **Le partage des paramètres** (parameter sharing**) :** Se réfère à l'utilisation du même paramètre pour plus d'une fonction dans un modèle. Dans un réseau de neurones convolutif, chaque élément du noyau est utilisé à chaque position de l'entrée (sauf peut-être quelques-uns des pixels des bords, selon le choix de conception concernant la frontière). Le parameter sharing utilisé par l'opération de convolution signifie que, plutôt que d'apprendre un ensemble de

paramètres distincts pour chaque emplacement, nous n'apprenons qu'un ensemble ce qui réduit encore davantage les exigences de stockage du modèle.

• **Représentations équivariantes (**equivariant representations**)** : La couche de convolution équivariante à la translation. Une fonction est équivariante signifie que si l'entrée change, la sortie change de la même manière. Autrement, une fonction f (x) est équivalente à une fonction g (x) si f (g (x)) = g (f (x)). Dans le cas d'une couche de convolution, une translation de l'image d'entrée entraîne la translation des cartes d'activation.

Les couches de convolution sont généralement suivies d'une couche de ReLU (introduite précédemment) afin de modifier les cartes de caractéristiques de sortie appelées ainsi cartes de caractéristiques rectifiées. L'application de la fonction ReLU à la sortie des couches de convolution possède plusieurs avantages :

— la convolution réalise des opérations d'additions et de multiplication, ce qui conserve la linéarité de la sortie par rapport à l'entrée. Le fait d'appliquer la fonction ReLU, introduit une certaine non-linéarité (en supprimant les pixels de valeur négative).

— suite à la suppression d'une partie des données, la fonction ReLU permet l'accélération des calculs.

— la mise en valeur des caractéristiques extraites par les couches de convolution en accentuant l'écart entre eux (valeurs négatives).

3.9) Couche d'échantillonnage (Pooling) :

Semblable à la couche de convolution, la couche d'échantillonnage est chargée de réduire la taille spatiale des cartes de caractéristiques, mais elle conserve les informations les plus importantes. Il existe différents types d'échantillonnage dont l'échantillonnage maximum ou Max Pooling-, l'échantillonnage moyen ou Average Pooling, etc.

L'échantillonnage consiste à appliquer un noyau de taille n ×n sur la carte d'activation en le faisant glisser avec un pas préalablement défini (le pas est généralement égal à la taille du noyau n pour éviter le phénomène de chevauchement). Le Max Pooling renvoie la valeur maximale de la partie de l'image couverte par le noyau. Au lieu de prendre le maximum, nous pourrions prendre la moyenne de tous les éléments couverts par le noyau, cela est assuré par le Average Pooling. Le Max Pooling permet de supprimer le bruit. D'autre part, le Average Pooling effectue simplement une réduction de la dimensionnalité en tant que mécanisme de suppression

de bruit. Par conséquent, en pratique, le Max Pooling est beaucoup plus utilisé que le Average Pooling puisqu'il fonctionne mieux.

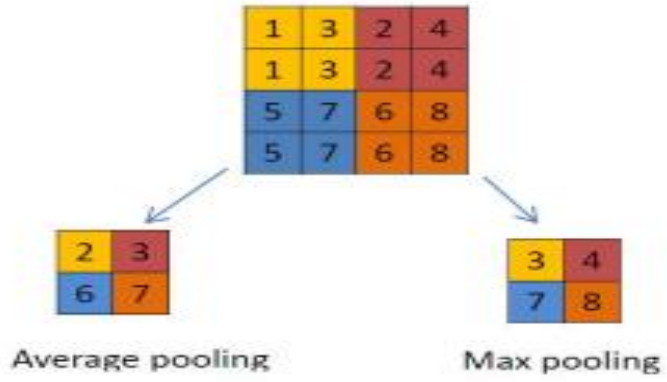


Figure 3-19 Exemple d'une opération de pooling.

Figure 3.19 – (à gauche) Average pooling : chaque case correspond à la moyenne du carré d'entrée de la même couleur, ex de la case jaune : (1+ 3+ 1+ 3)/4 = 2. (à droite) Max pooling : chaque case correspond à la valeur maximum du carré d'entrée de la même couleur, ex de la case bleu : max(5, 7, 5, 7) = 7.

L'échantillonnage a pour fonction de réduire progressivement la taille spatiale de la donnée d'entrée. En effet, il permet de :

— rendre les représentations d'entrée (carte des caractéristiques) plus petites et plus faciles à gérer.

— réduire la puissance de calcul requise pour traiter les données par réduction des dimensions et donc il contrôle le sur-apprentissage.

— extraire les caractéristiques dominantes qui sont invariantes en rotation et en position, maintenant ainsi l'efficacité du processus d'apprentissage du modèle.

— aider à obtenir une représentation équivariante de l'image. Ceci est très puissant puis qu'il permet de détecter des objets dans une image, peu importe où ils se trouvent.

3.10) Couche complètement connectée :

La couche complètement connectée est un Perceptron multi-couches traditionnel qui utilise une fonction d'activation (par exemple softmax) sur le vecteur de sortie afin d'ajouter la non-linéarité. Le terme « complètement connecté » implique que chaque neurone de la couche

précédente est connecté à chaque neurone de la couche suivante. Leurs activations peuvent donc être calculées avec une multiplication matricielle suivie d'un offset de biais.

La sortie des couches de convolution et d'échantillonnage représente des caractéristiques de haut niveau de l'image d'entrée. L'objectif de la couche complètement connectée est d'utiliser ces caractéristiques pour classer l'image d'entrée du réseau en différentes classes en fonction de la base de données d'apprentissage.

Outre la classification, l'ajout d'une couche complètement connectée est généralement un moyen peu coûteux pour faire apprendre des combinaisons non-linéaires des caractéristiques. La plupart des cartes de caractéristiques des couches de convolution et d'échantillonnage peuvent être utiles pour la tâche de classification, mais une combinaison de ces fonctionnalités peut être encore meilleure.

Les couches entièrement connectées sont généralement suivies d'un Dropout [58]. Ce dernier agit sur les poids de ces couches afin de désactiver un certain nombre de neurones pour réduire le nombre de paramètres. Cela permet de contrôler le sur-apprentissage qui peut être causé par un nombre important de paramètres.

La somme des probabilités en sortie de la couche entièrement connectée est P i yˆ = 1. Cela est garanti en utilisant Softmax comme fonction d'activation dans la couche en sortie de la couche entièrement connectée. La fonction Softmax prend un vecteur de scores arbitraires à valeurs réelles et le réduit à un vecteur de valeurs comprises entre zéro et un qui égal à un. Elle s'écrit sous la forme :

$$\hat{y}_i = \mathbf{softmax}(x)_i = \frac{e^{x_i}}{\sum_j e^{x_j}} \qquad (6)$$

La figure 3.20 représente un exemple d'un réseau de neurones convolutif.

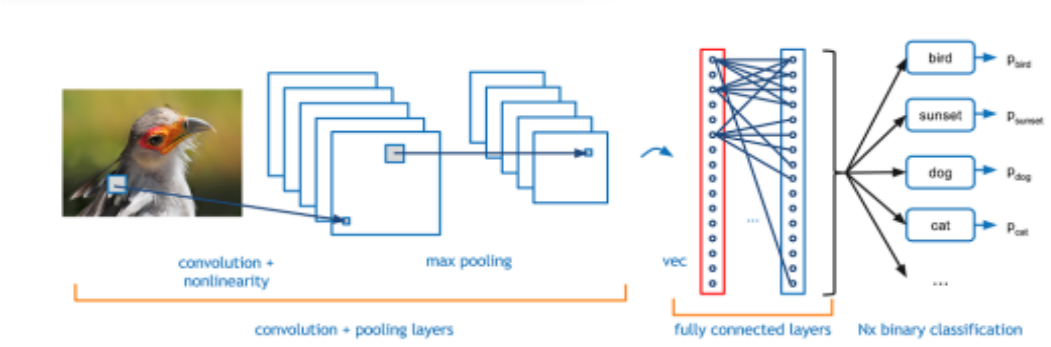


Figure 3-20 Un réseau de neurones convolutif [42].

Figure 3.20 – Un réseau de neurones convolutif qui reçois une image 2D comme entrée et qui est composé d'une couche convolutive, une fonction d'activation non lineaire, une couche MAX pooling et enfin un perceptron multi couche.

3.11) Quelques réseaux convolutifs célèbres :

• **LeNet** [43] : Les premières applications réussies des réseaux convolutifs ont été développées par Yann LeCun dans les années 1990. Parmi ceux-ci, le plus connu est l'architecture LeNet utilisée pour lire les codes postaux, les chiffres, etc.

• **AlexNet**[44] : Le premier travail qui a popularisé les réseaux convolutifs dans la vision par ordinateur était AlexNet, développé par Alex Krizhevsky, Ilya Sutskever et Geoff Hinton. AlexNet a été soumis au défi ImageNet ILSVRC [45] en 2012 et a nettement surpassé ses concurrents. Le réseau avait une architecture très similaire à LeNet, mais était plus profond, plus grand et comportait des couches convolutives empilées les unes sur les autres (auparavant, il était commun de ne disposer que d'une seule couche convolutifs toujours immédiatement suivie d'une couche de pooling).

• **ZFnet**[46] : Le vainqueur de ILSVRC challenge 2013 était un réseau convolutif de Matthew Zeiler et Rob Fergus. Il est devenu ZFNet (abréviation de Zeiler et Fergus Net). C'était une amélioration de AlexNet en ajustant les hyper-paramètres de l'architecture, en particulier en élargissant la taille des couches convolutifs et en réduisant la taille du noyau sur la première couche.

• **GoogLeNet**[47] : Le vainqueur de ILSVRC challenge 2014 était un réseau convolutif de Szegedy et al. De Google. Sa principale contribution a été le développement d'un module inception qui a considérablement réduit le nombre de paramètres dans le réseau (4M, par rapport à AlexNet avec 60M). En outre, ce module utilise le **global AVG pooling** au lieu du PMC à la fin du réseaux, ce qui élimine une grande quantité de paramètres. Il existe également plusieurs versions de GoogLeNet, parmi elles, Inception-v4 [48].

• **ResNet**[49] : Residual network développé par Kaiming He et al. A été le vainqueur de ILSVRC 2015. Il présente des sauts de connexion et une forte utilisation de la batch normalisation. Il utilise aussi le global AVG pooling au lieu du PMC à la fin.

La figure 3.21- représente le taux d'erreur dans ImageNet Visual recognition Challenge :

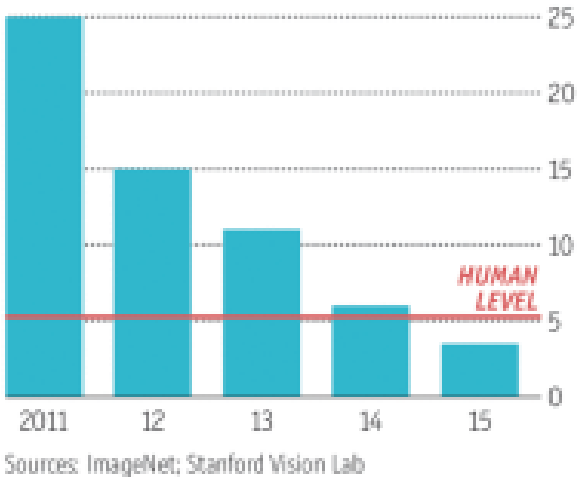


Figure 3-21 taux d'erreur dans ImageNet Visual recognition Challenge [45].

Le Deep Learning dépasse la performance humaine.

3.12) Conclusion :

Avec l'utilisation de la reconnaissance faciale qui se répand, la marge de manœuvre des cybercriminels qui volent les données sur votre visage pour frauder ne fait qu'augmenter.

C'est pourquoi une solution complète de cybersécurité est essentielle pour protéger votre vie privée et votre sécurité en ligne.

La technologie biométrique offre des solutions de sécurité très performantes malgré les risques, les systèmes sont pratiques et difficiles à reproduire. Ces systèmes vont continuer de se développer à l'avenir. Le défi sera d'optimiser leurs avantages tout en minimisant les risques.

Dans ce chapitre nous avons parlé des attaques sur un système de reconnaissance de visage au niveau du capteur et les méthodes existantes pour contrer ces attaques, puis nous avons abordé la détection de la vivacité de visage et ses quelques approches déjà existantes, nous avons parlé ensuite sur l'apprentissage profond, les réseaux de neurones convolutifs, couche convolutif en détail et citer quelques réseaux convolutifs célèbres.

Dans le chapitre suivant nous allons donner une description détaillée sur les étapes de l'approche qu'on a proposé comme étant une contre mesure en cas d'attaque.

4) Chapitre 4 : Approche proposée

4.1) Introduction :

Si les données biométriques sont généralement considérées comme la méthode d'authentification la plus fiable, elles comportent néanmoins des risques. Car si les informations de carte de crédit d'une personne sont volées, celle-ci peut alors bloquer son compte et prendre les mesures nécessaires pour changer les informations personnelles qui ont été piratées. Que faire si vous perdez votre « visage numérique » ?

Partout dans le monde des informations biométriques sont collectées, stockées et analysées en grandes quantités, souvent par des sociétés et des gouvernements, pas toujours dans les meilleures conditions de cybersécurité. La question qui revient souvent est la suivante : l'infrastructure qui possède et traite toutes ces données est-elle sûre ?

La reconnaissance faciale est une forme populaire et efficace d'authentification biométrique utilisée dans de nombreuses applications logicielles. Un inconvénient de cette technique est qu'elle est sujette à des attaques d'usurpation d'identité, où un imposteur peut accéder au système en présentant une photo ou vidéo ou un masque 3D d'un utilisateur valide au capteur ainsi, la détection de la vivacité du visage est une étape nécessaire avant d'accorder l'authentification à l'utilisateur.

Dans ce chapitre, nous allons parler de l’approche que nous avons proposée comme étant une contre mesure lors d’une attaque.

4.2) Objectif de notre approche :

La reconnaissance d'objets est considérée comme l'une des technologies informatiques les plus populaires qui intègre le traitement d'images et la vision par ordinateur, et elle est liée à la détection d'exemples d'une entité comme les visages humains, les bâtiments, les arbres, les voitures, etc.

L'objectif principal des algorithmes de reconnaissance des visages est de décider et d'analyser si un visage existe ou non dans une image.

Au cours des dernières décennies, beaucoup d'études et de travaux ont été réalisés dans le domaine de la reconnaissance et de la détection des visages pour les rendre plus avancés et plus précis, mais une révolution a eu lieu dans ce domaine lorsque Viola-Jones a mis au point un détecteur de visage en temps réel, qui était capable de détecter les visages en temps réel avec une grande précision.

L'étape initiale et primordiale de la reconnaissance des visages est la détection des visages, qui est utilisée pour détecter les visages dans les images. Elle fait partie intégrante de la reconnaissance d'objets et peut être utilisée dans plusieurs domaines importants tels que la sécurité, la biométrie, le droit, l'analyse statistique, le divertissement, etc.

Elle est utilisée pour reconnaître les visages en temps réel pour surveiller et suivre L'emplacement d'une personne ou d'autres entités.

Elle est le plus souvent utilisée dans les appareils photo mobiles et les reflex numériques pour identifier des apparitions multiples dans le cadre.

Le meilleur exemple que l'on puisse tirer d'une application quotidienne populaire est Facebook qui utilise également des algorithmes de reconnaissance faciale pour détecter les visages dans les images et les reconnaître [52].

L'objectif principal de l'approche proposée est de reconnaître les données du visage de différentes personnes et de les identifier avec des résultats précis et de manière efficace.

4.3) Organigramme :

La figure 4.1 montre l'organigramme de notre système anti spoofing en utilisant les deux ensembles de données, à savoir les vraies et les fausses images :

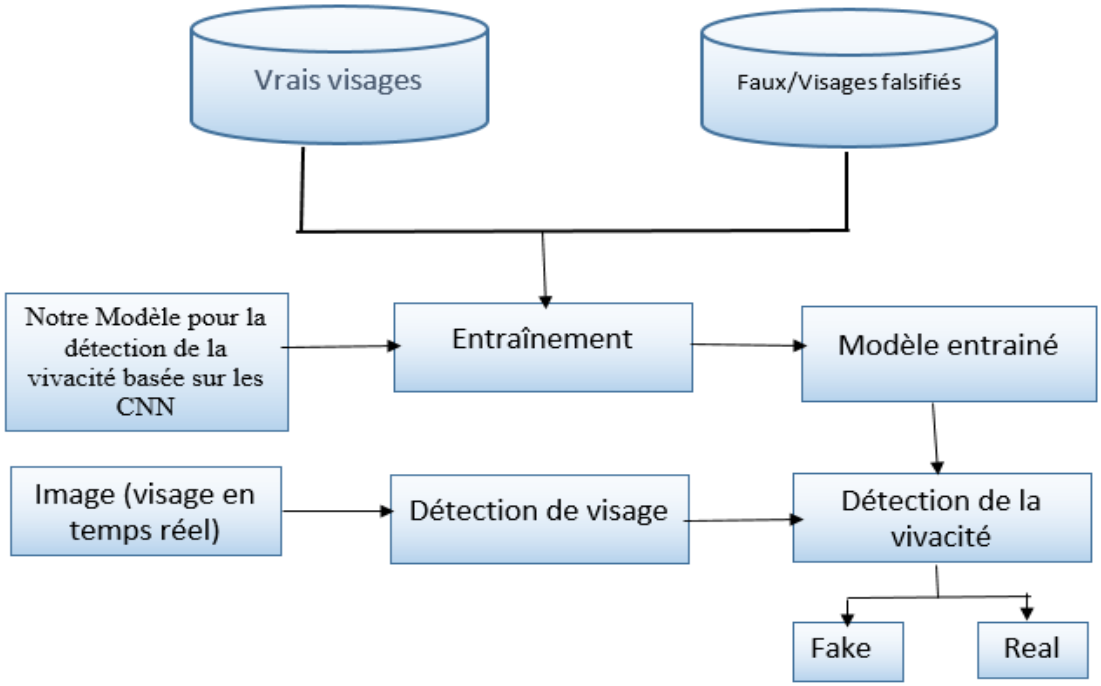


Figure 4-1 organigramme de notre système anti-spoofing.

4.4) Méthodologie :

● **Ensemble de données** : La première étape de la construction de notre système de reconnaissance des visages consiste à collecter des exemples de chaque visage à identifier. Plusieurs méthodes peuvent être utilisées pour créer les ensembles de données. La méthode la plus pratique consiste à recueillir des images vidéo de personnes en leur demandant de se rendre dans une zone ou un lieu défini à cet effet. Cette méthode peut être répétée pendant un certain temps dans différentes conditions, humeurs, situations, etc. pour ajouter de la variabilité et de l'unicité à l'ensemble de données. L'autre méthode consiste à collecter des vidéos et des images de personnes sur des plateformes en ligne dans le cas de personnalités publiques ou célèbres.

Dans notre cas, nous avons utilisé la base d'images CASIA-FASD.

● **Détection de visage** : La toute première étape est celle de détection de visage, cette étape consiste à extraire le visage à partir de l'image donnée en entrée, La détection de visage peut se faire avec plusieurs méthodes, nous nous avons opté pour la méthode de Viola et Jones qui est une méthode très performante en vue du taux de détection élevé dans un temps minimal, La figure 4.2 montre un exemple de détection de visage (image issue de la base CASIA).

Figure 4-2 Détection de visage.

Après avoir repérer le visage une imagette contenant uniquement le visage va être extraite, vu qu'on a besoin que du visage, et les traitements futurs vont être fait sur cette imagette.

● **Entraînement de la reconnaissance du visage** Tout d'abord, un script python est créé pour la formation des données et placées dans le même dossier que les ensembles de données, comme le montre la figure 4.3.

En outre, la méthode de viola et Jones est utilisée afin de capturer les visages à partir des jeux de données. Comme le montre la figure 4.4.

Après avoir terminé le script python et l'exécuter, cela entraînera l'ensemble des données pour la reconnaissance faciale comme le montre la figure 4.5.

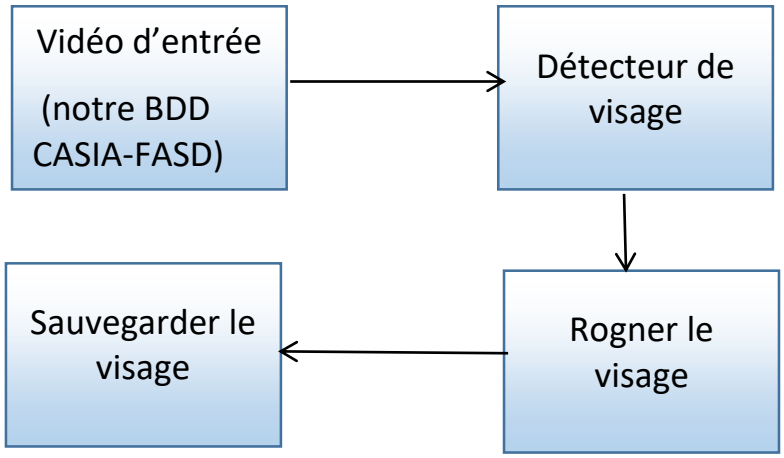


Figure 4-3 Schéma fonctionnel pour la préparation du jeu de données pour système anti-spoofing.

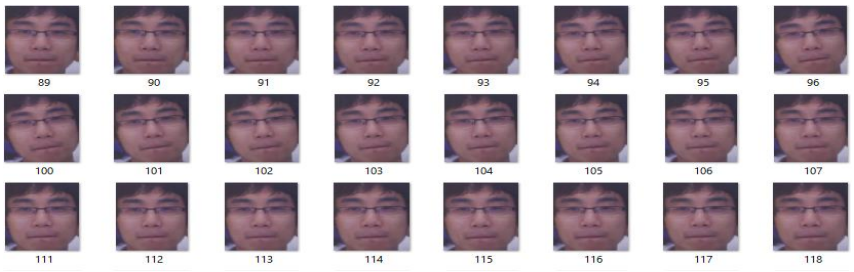


Figure 4-4 Jeu de données pour la détection de la vivacité des visages.

● **Intégration des caméras** La caméra est l'un des composants les plus importants du processus de reconnaissance faciale. La caméra sera utilisée pour capturer les images d'une personne en temps réel afin qu'elles puissent être comparées à l'ensemble des données entraînées.

Figure 4-5 montre le résultat de la formation(entrainement).

```
y: 0.9142
Epoch 45/50
37/37 [==============================] - 1s 21ms/step - loss: 0.4426 - accuracy: 0.8058 - val_loss: 0.1713 - val_accurac
y: 0.9373
Epoch 46/50
37/37 [==============================] - 1s 21ms/step - loss: 0.4915 - accuracy: 0.7692 - val_loss: 0.1869 - val_accurac
y: 0.9307
Epoch 47/50
37/37 [==============================] - 1s 20ms/step - loss: 0.3803 - accuracy: 0.8316 - val_loss: 0.1748 - val_accurac
y: 0.9307
Epoch 48/50
37/37 [==============================] - 1s 20ms/step - loss: 0.4190 - accuracy: 0.7945 - val_loss: 0.2054 - val_accurac
y: 0.9208
Epoch 49/50
37/37 [==============================] - 1s 21ms/step - loss: 0.4449 - accuracy: 0.7951 - val_loss: 0.2130 - val_accurac
y: 0.9274
Epoch 50/50
37/37 [==============================] - 1s 21ms/step - loss: 0.3898 - accuracy: 0.8669 - val_loss: 0.1987 - val_accurac
y: 0.9373
[INFO] evaluating network...
              precision    recall  f1-score   support

       fake       1.00      0.88      0.94       163
       real       0.88      1.00      0.94       140

    accuracy                           0.94       303
   macro avg       0.94      0.94      0.94       303
weighted avg       0.94      0.94      0.94       303
```

Figure 4-5 Résultat de la formation(entrainement).

● **Mise en œuvre de la reconnaissance des visages** Après la création d'un jeu de données productif, l'entraînement de ce dernier et l'intégration du module caméra, et l'implémentation de notre reconnaisseur de visage. Dès qu'une personne arrive devant la caméra, celle-ci va capturer l'image de la personne en temps réel et comparer son visage avec le dataset entraîné, on obtiendra ensuite un résultat avec une valeur de précision appropriée.

● **Détection de la vivacité pour éviter l'usurpation d'identité** : Notre architecture CNN présente des qualités VGGNet.

La figure 4.6 présente notre architecture de détecteur de vivacité basé sur les CNN.

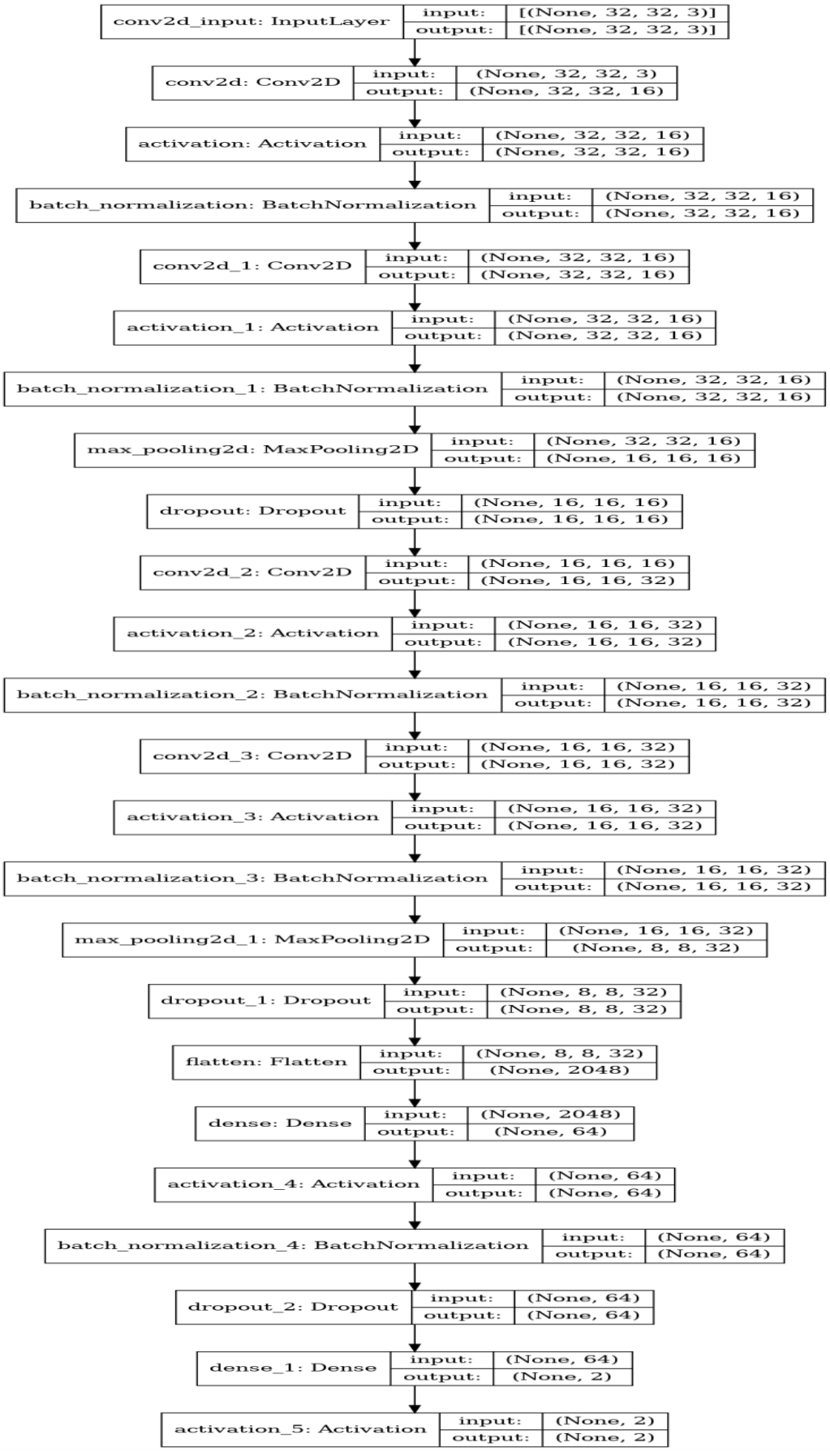


Figure 4-6 Architecture CNN pour détecter les vrais et faux visages.

4.5) Implémentation de notre détecteur de vivacité basé sur l'apprentissage profond :

Le réseau proposé permet de garantir que notre détecteur de vivacité soit rapide, capable de fonctionner en temps réel (même sur des appareils à ressources limitées, tels que le Raspberry Pi).

La figure 4-7 représente l'implémentation de notre modèle d'apprentissage profond basé sur CNN pour détecter les vrais et faux visages.

```
1.  # import the necessary packages
2.  from tensorflow.keras.models import Sequential
3.  from tensorflow.keras.layers import BatchNormalization
4.  from tensorflow.keras.layers import Conv2D
5.  from tensorflow.keras.layers import MaxPooling2D
6.  from tensorflow.keras.layers import Activation
7.  from tensorflow.keras.layers import Flatten
8.  from tensorflow.keras.layers import Dropout
9.  from tensorflow.keras.layers import Dense
10. from tensorflow.keras import backend as K
11.
12. class Notremodele :
13.     @staticmethod
14.     def build(width, height, depth, classes):
15.         # initialize the model along with the input shape to be
16.         # "channels last" and the channels dimension itself
17.         model = Sequential()
18.         inputShape = (height, width, depth)
19.         chanDim = -1
20.
21.         # if we are using "channels first", update the input shape
22.         # and channels dimension
23.         if K.image_data_format() == "channels_first":
24.             inputShape = (depth, height, width)
25.             chanDim = 1
```

Figure 4-7 Notre modèle d'apprentissage profond basé sur CNN pour détecter les vrais et faux visages.

Notre classe (notremodele)se compose d'une méthode statique qui accepte quatre paramètres :

- Largeur : largeur de l'image/du volume.
- La taille : la hauteur de l'image.
- Profondeur : Le nombre de canaux pour l'image (dans ce cas 3 puisque nous allons travailler avec des images RVB).
- Des classes : Nous avons deux classes au total 2 : « vrai » et « faux ».

4.6) Conclusion

En utilisant l'approche proposée, on peut reconnaître les visages avec une bonne précision. Elle inclut également le détecteur de vivacité pour détecter les faux et les vrais visages en utilisant l'architecture d'apprentissage profond basé sur les CNN.

La première étape de la construction de notre système antispoofing consiste à mettre à notre disposition un ensemble de données pour l'entraînement du modèle d'apprentissage profond, nous avons utilisé la base d'images CASIA-FASD.

Après avoir préparé le jeu de données, la vivacité de l'image est d'abord détectée à l'aide du modèle d'apprentissage profond basé sur les CNN. Ensuite, le visage réel détecté est reconnu en utilisant ce dernier construit sur le jeu de données d'entraînement.

La méthode proposée est capable de reconnaître les visages et détecter des images réelles et fausses.

Dans le chapitre suivant nous allons décrire l'environnement de travail ainsi que la base de données utilisée, ensuite définir la structure de notre application et nous finirons par des tests et résultats de notre système antispoofing en utilisant notre approche basée sur les CNN.

5) CHAPITRE 5 : Test et résultat

5.1) Introduction :

Dans ce dernier chapitre on va exposer notre implémentation de la contre-mesure lors d'une attaque contre un système de reconnaissance de visage en utilisant notre modèle basé sur les CNN comme expliqué précédemment dans notre approche proposée (chapitre 4) ensuite, présenter l'environnement de développement et décrire la base d'image qu'on a utilisé, on terminera par les tests et discussion des résultats.

5.2) Environnement de développement :

Pour la réalisation de notre application nous avons eu recours à PYTHON.

Python : est un langage de programmation interprété, multi-paradigme et multiplateformes. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet. Il est doté d'un typage dynamique fort, d'une gestion automatique de la mémoire par ramasse-miettes et d'un système de gestion d'exceptions.

Le langage Python est placé sous une licence libre proche de la licence BSD et fonctionne sur la plupart des plates-formes informatiques, des smartphones aux ordinateurs centraux, de windows à Unix avec notamment GNU/Linux en passant par macOS, ou encore Android, iOS, et peut aussi être traduit en Java ou NET. Il est conçu pour optimiser la productivité des programmeurs en offrant des outils de haut niveau et une syntaxe simple à utiliser.

Il est également apprécié par certains pédagogues qui y trouvent un langage où la syntaxe, clairement séparée des mécanismes de bas niveau, permet une initiation aisée aux concepts de base de la programmation.

5.3) Mesures d'évaluation :

Diverses mesures d'évaluation sont utilisées pour déterminer la performance du classificateur. Les mesures d'évaluation utilisées dans l'étude sont expliquées ci-dessous [70].

La matrice de confusion est l'une des techniques les plus utilisées dans l'apprentissage automatique comprend des informations sur les classes réelles et prédites obtenues par un

système de classification. La matrice de confusion a deux dimensions : les classes réelles et les classes prédites. Alors que chaque ligne représente un exemple de classe réelle, chaque colonne représente l'état d'une classe prédite. Dans la matrice de confusion, TP est le nombre de vrais positifs, TN est le nombre de vrais négatifs, FP est le nombre de faux positifs et FN est le nombre de faux négatifs, FP est le nombre de faux positifs et FN est le nombre de faux négatifs.

ACCURACY :

La précision du classificateur est quantifiée avec cette métrique. Le nombre de données correctement classées est divisé par le nombre total de données pour calculer la précision.

$$\text{Accuracy} = TP + TN / TP + TN + FP + FN. \quad (7)$$

Précision :

La précision indique la proportion des données prédites comme positives qui sont prédites correctement. En d'autres termes, une précision élevée signifie moins de faux positifs.

$$\text{Précision} = TP / TP + FP. \quad (8)$$

Recall :

Le rappel (recall en anglais) est la métrique permettant de déterminer la complétude du classificateur. Un rappel plus élevé indique des faux négatifs plus faibles, tandis qu'un rappel plus faible indique des faux négatifs plus élevés. La précision diminue souvent avec une amélioration du rappel.

$$\text{Recall} = TP / TP + FN. \quad (9)$$

F1- score :

$$\text{F1-Score} = 2 \times \text{Recall} \times \text{Precision} / \text{Recall} + \text{Precision}. \quad (10)$$

Pour obtenir le score F1, le produit du rappel et de la précision est divisé par la somme du rappel et de la précision.

Cette métrique est le membre le plus utilisé de la famille paramétrique des mesures F, nommée d'après la valeur du paramètre $\beta = 1$. Le score F1 est défini comme la moyenne harmonique de la précision et du rappel, (pire valeur : 0 ; meilleure valeur : 1) F1 s'étend sur [0, 1], où le minimum est atteint pour TP=0, c'est-à-dire lorsque tous les échantillons positifs sont mal classifiés, et le maximum pour FN=FP=0, c'est-à-dire pour une classification parfaite.

5.4) Présentation des outils :

5.4.1) Logiciels (software) :

Plusieurs framework open sources sont disponibles dans la littérature, la grande majorité supporte le langage Python. Voici une liste non exhaustive de quelques framework :

Theano :

Créé par Frédéric Bastien et l'équipe de recherche derrière le laboratoire de l'Université de Montréal, MILA.

• **Avantages**	• **Inconvénients**
— Python — Performant si utilisé correctement	— Le supporte du Multi GPU nécessite une solution de contournement — Les grands modèles peuvent nécessiter un long temps de compilation — API bas niveau

Tableau 1 Avantages et inconvénients de Theano.

TensorFlow :

TensorFlow fut créé par l'équipe Google Brain pour mener des recherches sur le ML et le Deep Learning. Il est considéré comme une version moderne de Theano.

• **Avantages**	• **Inconvénients**
— Python — supporté par Google — Une très grande communauté — Le support du multi-GPU	— Plus lent que les autres framework dans de nombreux benchmarks, bien que Tensorflow se rattrape. — Le soutien des RNN est encore surclassé par Theano

Tableau 2 Avantages et inconvénients de TensorFlow.

Keras :

Le framework le plus haut niveau, le plus convivial de la liste. Il permet aux utilisateurs de choisir si les modèles qu'ils construisent sont exécutés sur Theano ou TensorFlow. Il est écrit et entretenu par Francis Chollet, un autre membre de l'équipe Google Brain [71].

• **Avantages**	• **Inconvénients**
— Python — Le backend par excellence pour Theano ou TensorFlow — Interface haut niveau, intuitive	— Moins flexible que les autres API

Tableau 3 Avantages et inconvénients de Keras.

Autres frameworks :

• **CNTK :** par Microsoft

• **Neon :** par Nervana Systems. Il a récemment été classé comme le frameworks le plus rapide dans plusieurs catégories.

• **Deeplearning4j** : Il supporte le langage java

• **Caffe :** par Berkeley Vision and Learning Center

Après ce petit tour d'horizon des différents frameworks disponibles, notre choix s'est porté sur le framework TensorFlow de Google. La raison principale de ce choix c'est la très grande et aussi très active communauté qui est derrière cette librairie.

La figure 5.1 nous montre la croissance de la popularité de TensorFlow.

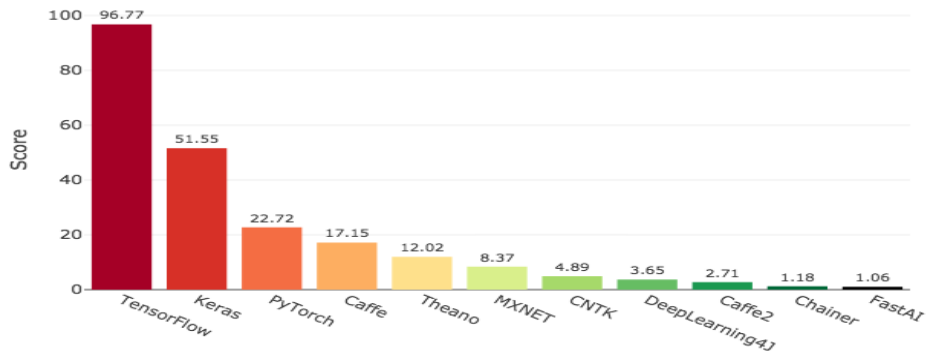


Figure 5-1 La croissance de la popularité de TensorFlow [71].

5.4.2) Matériels (hardware) :

L’apprentissage profond (deep learning) est un domaine avec des exigences en calculs intenses et la disponibilité des ressources dédiés à cette tache vont fondamentalement influencer sur l’expérience de l’utilisateur car sans ses ressources, il faudra trop de temps pour apprendre de ses erreurs ce qui peut être décourageant. Les expérimentations ont tous été effectuées sur une machine qui offre des performances acceptables dont voici les caractéristiques :

CPU	Intel(R) Core(TM) i5-8250U (1.60GHz 1.80 GHz)
RAM	16 GB

Tableau 4 Caractéristiques de la machine.

Maintenant que nous avons présenté l’environnement de développement nous allons passer à la description de la base de données utilisée.

5.5) Base de données utilisée :

Comme pour toute problématique de reconnaissance automatique, les techniques doivent être validées à l’aide d’un ensemble consistant de données. La base de données que nous avons utilisée pour valider la reconnaissance de visage et détecter les attaques biométriques par le visage est présentée ci-après.

Dans notre système nous avons utilisé la base d’images CASIA-FASD.

La base de données CASIA-FASD est une base de données d’attaques qui se compose de trois types d’attaques : les photographies imprimées gondolées, les photographies imprimées avec des yeux coupés et les attaques vidéo. Les échantillons sont prélevés avec trois types de caméras : basse qualité, qualité normale et haute qualité.

Nous avons utilisé « scikit-learn » pour partitionner nos données en deux parties : 50% pour l'ensemble d'apprentissage et 50% pour l'ensemble de validation. L'ensemble d'apprentissage est utilisé pour entraîner le modèle, tandis que l'ensemble de validation est utilisé pour évaluer les performances du modèle.

La Figure 5.2 est un exemple de la base CASIA-FASD.

Figure 5-2 Exemple de la base CASIA-FASD.

5.6) Structure de l'application :

Le processus de formation de détecteur de vivacité basé sur CNN en utilisant à la fois des images « réelles » et « falsifiées / fausses » comme ensemble de données, nous avons formé un système avec un modèle de détection de vivacité basé sur l'apprentissage profond comme nous le montre la figure 4.1.

Dans cette partie on va présenter les différents aspects de notre application,

- **Base de données/ :** Notre répertoire de jeux de données CASIA-FASD se compose de deux classes d'images : Fausses images et de vraies images.
  - Ensemble de données/faux/ : Contient les faux visages extraient de nos vidéos (base de données CASIA FASD) dédiés pour la phase d'apprentissage.
  - Jeu de données/réel/ : Contient les visages réels extraient de nos vidéos (base de données CASIA FASD) dédiés pour la phase d'apprentissage.
- **Détecteur_visage/** : Se compose de notre détecteur de visage pour localiser des visages.

Cette étape consiste à extraire le visage à partir de l'image donnée en entrée, La détection de visage peut se faire avec plusieurs méthodes, nous avons opté pour la méthode de Viola et Jones qui est une méthode très performante en vue du taux de détection élevé dans un temps minimal,

- **Pyimagesearch/** : Ce module contient notre classe (Notremodele) de détecteur de vivacité de visage.
- **Vidéos/** : notre base de données CASIA-FASD (vidéos réelles et fausses).

Nous avons ensuite utilisé **trois s**cripts Python qui sont :

1)extractimg.py Ce script récupère les visages à partir du fichier vidéo (base de données CASIA-FASD) et nous aide à créer un ensemble de données d'apprentissage profond sur la vivacité du visage.

2)train.py : Comme le nom du fichier l'indique, ce script entraînera notre classificateur. Nous utiliserons Keras et TensorFlow pour entraîner le modèle. Le processus de formation se traduit par quelques fichiers :

- le .cornichon : Notre encodeur d'étiquette de classe.
- vivacité.modèle : Notre modèle basé sur les CNN qui détecte la vivacité du visage.
- plot.png : Le graphique de l'historique d'entraînement montre les courbes de précision et de perte afin que nous puissions évaluer notre modèle.

3)demo.py : Notre script de démonstration qui déclenchera notre webcam pour saisir des images afin de détecter la vivacité des visages en temps réel.

Formation de notre détecteur de vivacité :

Les figures 5.3 et 5.4 nous montre résultat après l'entraînement de notre détecteur de vivacité :

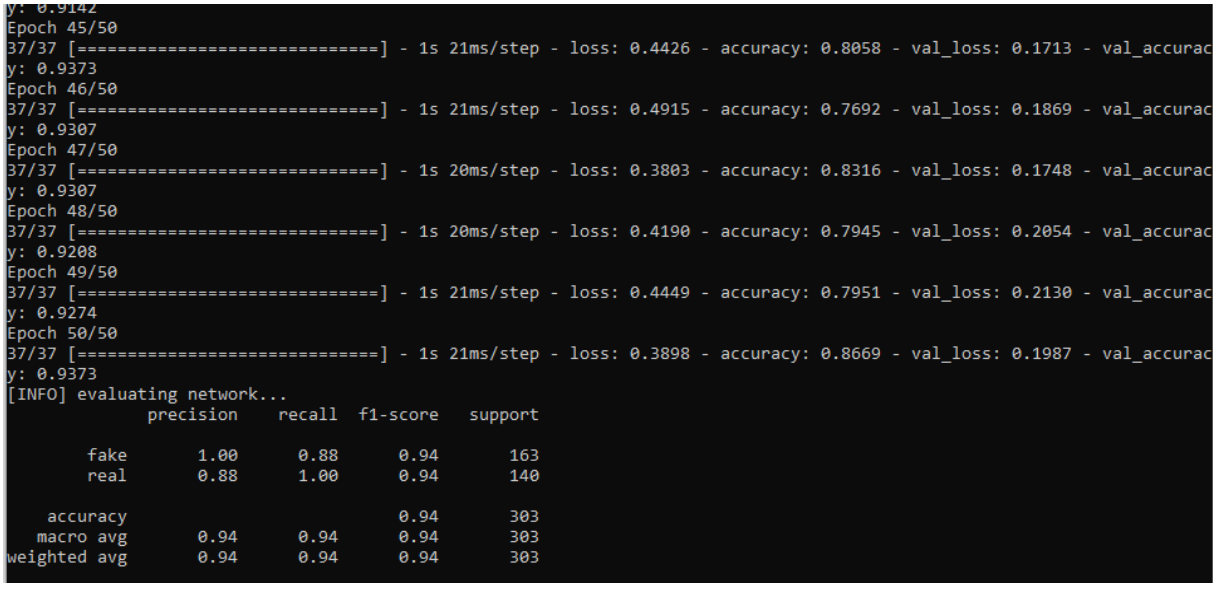


Figure 5-3 Le résultat de l'entraînement de notre détecteur de vivacité basé sur les CNN.

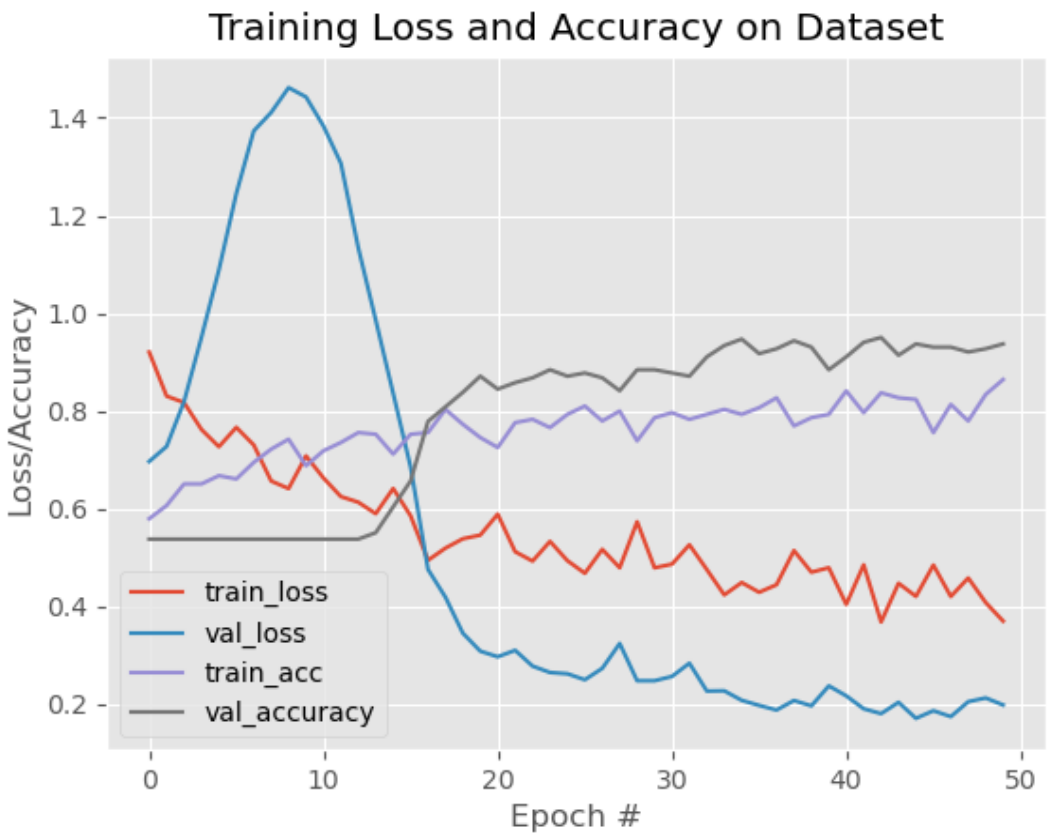


Figure 5-4 Courbes de sortie de l'historique de précision et de perte.

Train_loss et tain_acc sont des mesures de perte et de précision sur l'ensemble d'apprentissage, tandis que val_loss et val_acc sont des mesures de perte et de précision sur l'ensemble de validation.

On peut voir que notre modèle a une précision d'environ 80 % sur l'ensemble d'apprentissage et d'environ 85 % sur l'ensemble de validation. Cela signifie que notre modèle fonctionne avec une précision d'environ 85 % sur les nouvelles données.

On remarque que lorsque nos époch passent de 25, nos métriques train_acc et val_acc augmentent. Cela signifie que notre modèle s'adapte à l'ensemble d'apprentissage et capable de prédire sur de nouvelles données.

5.7) TEST :

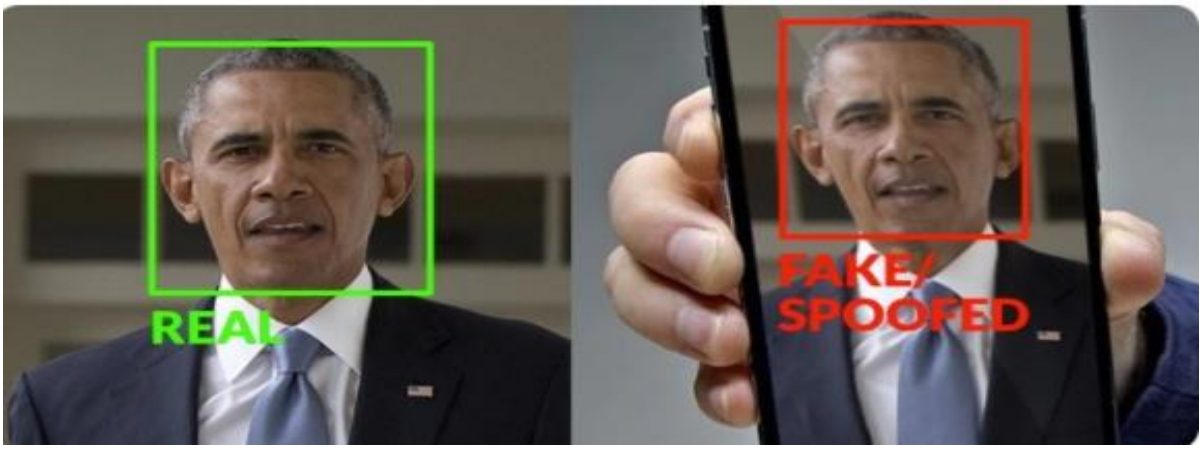
Pour le test nous avons déployé notre détecteur de vivacité sur la vidéo en temps réel :

La figure 5.5 nous montre le chargement de notre détecteur de visage et détecteur vivacité.

```
2.      --detector face_detector<font></font>
3.    Utilisation du backend TensorFlow.<font></font>
4.    [INFO] chargement du détecteur de visage...<font></font>
5.    [INFO] chargement du détecteur de vivacité...<font></font>
6.    [INFO] démarrage du flux vidéo...<font></font>
```

Figure 5-5 Chargement de notre de visage et vivacité de visage.

On peut voir que notre détecteur de vivacité distingue avec succès les visages réels des visages faux/falsifiés comme le montre la figure 5.6.



Pour Figure 5-6 Résultat de l'image traitée pour la vivacité montrant le vrai et le faux visage.

5.8) Résultats et discussion

La précision globale obtenue est d'environ 85% pour notre modèle de reconnaissance des visages sur l'ensemble de données de validation, comme le montre la figure 5.4.

En utilisant l'approche proposée, on peut reconnaître les visages avec une grande robustesse et une grande précision. Elle inclut également le détecteur de vivacité pour détecter les faux et les vrais visages en utilisant notre méthode basée sur les CNN.

Notre réseau nous permet de réduire les risques de surapprentissage sur notre ensemble de données petit (une seule base de données utilisée) et garantir que notre détecteur de vivacité soit rapide et capable de fonctionner en temps réel.

En utilisant le modèle d'apprentissage profond construit sur le jeu de données(CASIA-FASD), nous avons choisi de partitionner nos données en deux parties : 50% pour l'ensemble d'apprentissage et 50% pour l'ensemble de validation. L'ensemble d'apprentissage est utilisé pour entraîner le modèle, tandis que l'ensemble de validation est utilisé pour évaluer les

performances du modèle. La méthode proposée est capable de reconnaître les visages et détecter les images réelles et fausses comme le montre la figure 5.6.

Le principal inconvénient de la méthode que nous proposons est le nombre limité de données qui est dû à l'utilisation d'une seule base de données (CASIA-FASD), soit environ 223 vidéos dont 82 appartiennent à la vraie classe et 141 à la fausse classe ensuite nous avons ajouté une boucle qui nous permit d'extraire 6 images (visage) de chaque vidéo.

Le travail futur consistera d'entrainer notre modèle sur d'autres bases de données (la base de données ReplayAttack par exemple) pour diversifier le jeu de données dans différentes conditions et situations. Par exemple, des images contenant des visages de différentes couleurs de peau, tailles et formes. Un autre point est que le faux ensemble de données utilisé pour former notre modèle était uniquement basé sur l'exposition de photos provenant de téléphones mobiles, tablettes et sur des photos imprimées et non sur des masques 3D, Par conséquent, différents types de données devraient être inclus dans le faux ensemble de données.

5.9) Conclusion

Dans ce chapitre nous avons décrit Python comme étant notre environnement de développement avec lequel on a réalisé notre application, nous avons aussi parlé sur la base d'images CASIA-FASD qui contient des vidéos de différents sujets avec les différents types d'attaques(photo, vidéo), ensuite nous avons présenté notre application, enfin nous avons testé notre application en temps réel et on a terminé sur les résultats obtenu après l'entrainement où on a eu une précision globale d'environ 85% pour notre modèle de reconnaissance des visages sur l'ensemble de données de validation qui inclut également le détecteur de vivacité pour détecter les faux et les vrais visages en utilisant les réseaux de neurones convolutifs (CNN).

Conclusion générale et perspectives :

Il y a une augmentation continue de la quantité de population sur le globe, et ceci, à son tour, augmente le nombre d'ensembles de données complexes sur une période. Cela nécessite l'amélioration des algorithmes d'intelligence artificielle pour une catégorisation meilleure et précise des données. La caractéristique la plus déterminante du corps humain est le visage. Le visage de chaque personne est unique, bien qu'ayant la même structure comme le bruit, les yeux, les lèvres, etc. mais il peut varier de façon frappante. C'est dans cette variance que se trouvent les caractéristiques distinctives qui peuvent être utilisées pour identifier une personne par rapport à une autre. La reconnaissance du visage est un concept populaire qui est couramment utilisé dans les caméras de surveillance des lieux publics à des fins de sécurité.

Dans ce mémoire, nous avons développé une architecture profonde pour contrer les attaques biométriques par le visage avec deep learning en utilisant des réseaux de neurones convolutifs (CNN) pour classer l'image capturée comme réelle ou fausse, Les images de visage de la base de données CASIA-FASD sont utilisées pour faire l'apprentissage et les tests. Nous avons ainsi adapté notre modèle de détecteur des attaques par le visage formé et l'appliquer à la vidéo en temps réel.

Tout algorithme n'est considéré comme efficace que s'il est robuste et précis. Il fournit des résultats précis en matière d'usurpation de visage de manière rapide et efficace. Le principal avantage de notre technique est d'identifier l'unicité dans les ensembles de données en capturant les données de visage en temps réel à travers différents modes et gigue. Elle fournit également un modèle précis de reconnaissance des visages qui peut être utilisé à des fins de sûreté et de sécurité.

Nous avons obtenu une précision de 85% pour notre modèle de reconnaissance des visages sur l'ensemble de données de validation de la base de données CASIA-FASD en utilisant notre architecture approfondie basée sur les CNN.
Le principal avantage de cette approche est qu'elle est très précise et qu'elle peut être exécutée en temps réel.
Les inconvénients de la méthode que nous proposons est le nombre limité de données qui est dû à l'utilisation d'une seule base de données (CASIA-FASD), pour cela nous envisageons au futur d'entrainer notre modèle sur d'autres bases de données (la base de données ReplayAttack par exemple) pour diversifier le jeu de données dans différentes conditions et situations et approfondir notre CNN. Un autre point est que le faux ensemble de données utilisé pour former

notre modèle était uniquement basé sur l'exposition de photos provenant de téléphones mobiles, tablettes et sur des photos imprimées et non sur des masques 3D, Par conséquent, différents types de données devraient être inclus dans le faux ensemble de données.

Bibliographie :

[1] https://da.keylemon.com/(consulté le 21 mai 2021).

[2] H.-K. Jee, S.-U. Jung, J.-H. Yoo, "Liveness detection for embedded face recognition system.", International Journal of Biological and Medical Sciences, 1(4):235–238, 2006.

[3] G. Pan, L. Sun, Z. Wu, "Liveness detection for face recognition.", INTECH Open Access Publisher, 2008.

[4] AISSANI.S, Elaboration d'un cadre formel pour le renforcement de politiques de sécurité dans les programmes, thèse de doctorat, université de bejaia ;2007/2008.

[5] M. El Abed, R. Giot, B. Hemery, and C. Rosenberger. A study of users acceptance and satisfaction of biometric systems. In :IEEE International Carnahan Conference on Security Technology (ICCST), pp. 170–178, San Jose, USA, (2010).

[6] E.Maiwald: sécurité des réseaux informatiques, université de Rennes 1, Campus press 2001.

[7] Mohamad El-Abed. Evaluation de système biométrique. thèse de doctorat, Université de Caen, 2011.

[8] https://www.biometrie-online.net/technologies/ empreintes-digitales, (Consulté le 10 Juin 2019).

[9]https://www.firediy.fr/article/face-tracking-implementation-de-la-methode-de-viol (Consulté le 20 aout 2021).

[10] Bensenane Hamdan. Sécurisation des systèmes biométriques. thèse de doctorat, Université d'Oran, 2018.

[11] Benchennane Ibtissam. Etude et mise au point d'un procédé biométrique multimodale pour la reconnaissance des individus. thèse de doctorat, Université d'Oran, 2016.

[12] S.GUERFI ABABSA ; Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D ; thèse de doctorat ;Université Evry Val d'Essonne ;2008.

[13] https://www.kaspersky.fr/resource-center/definitions/what-is-facial-recognition (consulté le 13 aout 2021)

[14] A. Anjos, S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," in Proc. IJCB, pp. 1–7, 2011.

[15] N. Kose and J-L. Dugelay. Counter measure for the protection of face recognition systems against mask attacks. IEEE International Conference on Automatic Face and Gesture Recognition, (2013).

[16] P. Kumari and P. Ahlawat. A comparative study of different biometric technologies. In :International Journal of Enhanced Research in Management and Computer Applications,pp. 26–32,India, (2017).

[17] P. M. Ojala T., M. T. Multiresolution, "gray-scale and rotation invariant texture classification with local binary patterns.", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, issue 7, pp. 971–987, 2002.

[18] M. Fedias. Combinaisons de données d'espaces couleurs et de méthodes de vérification d'identité pour l'authentification de visages. PhD thesis, Université de Biskra, 2013.

[19] M**.** Nicolas. Reconnaissance Biométrique par Fusion Multimodale du Visage et de l'Iris. thèse de doctorat, Télécom ParisTech, 2009.

[20] J. Maatta, A. Hadid, M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis ", IEEE international joint conference In Biometrics (IJCB), pp. 1–7, 2011.

[21] P.Gautam, K. S. Jayash, "Face Liveness Detection using Local Diffused Patterns ", International Journal of Computer Applications, Vol. 149 – No.4, pp.0975 – 8887, 2016.

[22] W. Di, H. Hu, Anil K. Jain, "Face Spoof Detection with Image Distortion Analysis",IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.

[23] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference, vol 1, pp 511-518, (2001).

[24] http://biometrics.cse.msu.edu/pubs/databases.html (consulté le 20 juillet 2021).

[25] SACI Abdelmoumen ZITOUNI Sif Eddine. Authentification et identification biométrique des personnes par les empreintes palmaires. Thèse de master, UNIVERSITE KASDI MERBAH OUARGLA, 2016.

[26] T. Wang, J. Yang, Z. Lei, S. Liao, S. Z. Li., "Face liveness detection using 3d structure recovered from a single camera", In Biometrics (ICB), 2013.

[27] Y. Kim, J. Na, S. Yoon, and J. Yi. Masked fake face detection using radiance measurements. Journal of the Optical Society of America A, Vol. 2, issue 4, pp. 760-766, (2009).

[28] J. Fokkema. Using a challenge to improve face spoofing detection. essay.utwente.nl, (2016).

[29] H. Y. Xiong, B. Alipanahi, L. J. Lee, H. Bretschneider, D. Merico, R. K. Yuen, Y. Hua, S. Gueroussov, H. S. Najafabadi, T. R. Hughes, et al., "The human splicing code reveals new

insights into the genetic determinants of disease," Science, vol. 347, no. 6218, p. 1254806, 2015.

[30] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., "Mastering the game of go with deep neural networks and tree search," Nature, vol. 529, no. 7587, pp. 484–489, 2016.

[31] F. Buggenthin, F. Buettner, P. S. Hoppe, M. Endele, M. Kroiss, M. Strasser, M. Schwarzfischer, D. Loeffler, K. D. Kokkaliaris, O. Hilsenbeck, et al., "Prospective identification of hematopoietic lineage choice by deep learning," Nature Methods, vol. 14, no. 4, pp. 403–406, 2017.

[32] E. Gibney, "Google reveals secret test of ai bot to beat top go players.," Nature, vol. 541, no. 7636, pp. 142, 2017.

[33] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, et al., "End to end learning for self-driving cars," arXiv preprint arXiv :1604.07316, 2016.

[34] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologistlevel classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.

[35] H. C. Hazlett, "Early brain development in infants at high risk for autism spectrum disorder," in Biological Psychiatry, vol. 73, pp. 115S–115S, ELSEVIER SCIENCE INC 360 PARK AVE SOUTH, NEW YORK, NY 10010-1710 USA, 2013.

[36] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," arXiv preprint arXiv :1508.06576, 2015.

[37] R. Dechter and J. Pearl, The cycle-cutset method for improving search performance in AI applications. University of California, Computer Science Department, 1986.

[38] I. Aizenberg, N. N. Aizenberg, and J. P. Vandewalle, Multi-Valued and Universal Binary Neurons : Theory, Learning and Applications. Springer Science & Business Media, 2013.

[39] A. Van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in Advances in neural information processing systems, pp. 2643–2651, 2013.

[40] R. Collobert and J. Weston, "A unified architecture for natural language processing : Deep neural networks with multitask learning," in Proceedings of the 25th international conference on Machine learning, pp. 160–167, ACM, 2008.

[42] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org.

[43] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, pp. 1097–1105, 2012.

[45] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” International Journal of Computer Vision (IJCV), vol. 115, no. 3, pp. 211–252, 2015.

[46] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in European conference on computer vision, pp. 818–833, Springer, 2014.

[47] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9, 2015.

[48] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” arXiv preprint arXiv :1602.07261, 2016.

[49] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016.

[50] Kamran Kowsari, Donald E. Brown, Laura E. Barnes, “ RMDL : Random Multimodal Deep Learning for Classification", ICISDM '18, Lakeland, FL, US, 2018.

[51] Manik Sharma, J Anuradha, H KManne et G S Kashyap " Facial detection using deep learning ", School of Computing Science and Engineering, VIT University, Vellore - 632014, India, 2017.

[52] Divyansh Dwivedi , " Détection de visages pour les débutants ", Vers la science des données (https://towardsdatascience.com/face- detection-for-beginners- Actes de la cinquième conférence internationale sur les technologies de calcul inventif (ICICT-2020) IEEE Xplore Part Number:CFP20F70-ART ; ISBN:978-1-7281-4685-0 e58e8f21aad9), 2018.

[53] Jason Brownlee, "Comment utiliser les couches d'incorporation de mots pour l'apprentissage profond avec Keras", Machine Learning Mastery, 2017. (https://machinelearningmastery.com/use-word-embeddinglayers- deep-learning-keras.

[54] X. Tan, Y. Li, J. Liu, L. Jiang, “Face liveness detection from a single image with sparse low rank bilinear discriminative model.” in Proc. ECCV, pp. 504–517, 2010.

[55] L. Deng, D. Yu, et al., “Deep learning : methods and applications,” Foundations and Trends R in Signal Processing, vol. 7, no. 3–4, pp. 197–387, 2014. 73.

[56] Y. Bengio, A. Courville, and P. Vincent, “Representation learning : A review and new perspectives,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.

[57] P**.** Viola, M .Jones, “Robust real-time face detection”, International Journal of Computer Vision 57(2), vol 2, July 2001.

[58] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 2014.

[59] I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. MIT Press, 2016. 43.

[60] caractérisques peudo-Haar https://towardsdatascience.com/a-guide-to-face-detection-in-python-3eab0f6b9fc1 (consulté le 22 juin 2021).

[61]https://www.firediy.fr/article/face-tracking-implementation-de-la-methode-de-viola-jones-en-c (consulté le 10 septembre 2021).

[62]https://www.researchgate.net/publication/320279493_Descripteurs_de_Fourier_inspires_de_la_structure_du_cortex_visuel_primaire_humain_Application_a_la_reconnaissance_de_navires_dans_le_cadre_de_la_surveillance_maritime (consulté le 10 septembre 2021).

[63] http://www.sfu.ca/~llockhar/projects/facial-expression-recognition(consulté 10 septembre 2021).

[64] Dr Qaim Mehdi Rizvi. A review on face detection methods. Fevrier 2011.

[65] H.P. Graf, T. Chen, E. Petajan, et E. Cosatto, "Locating Faces and Facial Parts". Proc. First Intl Workshop Automatic Face and Gesture Recognition, pp. 41-46,1995.

[66] H.P. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, et E. Petajan, "Multimodal System for Locating Heads and Faces", Proc. Second Intl Conf. Automatic Face and Gesture Recognition, pp. 88-93, 1996.

[67] K. Sobottka, et I. Pitas, "Face Localization and Feature Extraction Based on Shape and Color Information", Proc. IEEE Int’l Conf. Image Processing, pp 483-486, 1996.

[68] T. K. Leung, M.C Burl, et P. Perona, "Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching", Proc. Fifth IEEE Int’l Conf. Computer Vision, pp 637-644, 1995.

[69] E. Sabert, et A. M. Tekalp, "Frontal-View Face Detection and Facial Feature Extraction using Color, Shape and Symmetry Based Cost Function", Pattern Recognition Letters, vol 17, no. 8, pp 669-680, 1998.

[70] Chicco, D. Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1-score and accuracy in binary classification evaluation. Chicco Jurman BMC Genom., 21, 1–13, 2020.

[71]https://towardsdatascience.com/deep-learning-framework-power-scores-2018-23607ddf297a (consulté le 17 septembre 2021).

Résumé :

La reconnaissance faciale est une forme populaire et efficace d'authentification biométrique utilisée dans de nombreuses applications logicielles.

L'un des inconvénients de cette technique est qu'elle est sujette aux attaques par usurpation de visage.

Les systèmes de reconnaissance de visage sont ciblés par de nombreuses attaques qui tentent de contourner ces systèmes, il existe trois types d'attaques qui sont les attaques par images, les attaques vidéo et les attaques par masque 3D du visage, pour contrer ces attaques les chercheurs se sont mis à chercher des contre-mesures.

La détection de la vivacité du visage est une étape nécessaire avant d'accorder l'authentification à l'utilisateur.

Dans ce mémoire, nous avons fait appel au deep learning pour contrer les attaques biométriques par le visage en utilisant des réseaux de neurones convolutifs (CNN) pour classer l'image capturée comme réelle ou fausse.

Nous avons donc utilisé un ensemble de données, ensuite implémenté un CNN capable d'effectuer la détection des attaques par le visage puis on a formé le réseau de détecteurs de vivacité.

Nous avons ainsi adapté notre modèle de détection des attaques par le visage formé aux vidéos en temps réel or, nous avons obtenu une précision de 85% pour notre modèle de reconnaissance des visages sur l'ensemble de données de validation de la base de données CASIA-FASD en utilisant notre architecture basée sur les réseaux de neurones convolutifs (CNN).

Abstract :

Facial recognition is a popular and effective form of biometric authentication used in many software applications.
One of the drawbacks of this technique is that it is prone to face spoofing attacks.
Face recognition systems are targeted by many attacks that try to bypass these systems, there are three types of attacks which are image attacks, video attacks and 3D face mask attacks, to counter these attacks researchers have started looking for countermeasures.

Detecting the liveliness of the face is a necessary step before granting authentication to the user. In this dissertation, we developed a deep architecture to counter face biometric attacks with deep learning using convolutional neural networks (CNN) to classify the captured image as real or fake.

So we used a dataset, then Implement a CNN capable of performing face attack detection and then the liveness detector network was trained.

We have thus adapted our model for detecting trained face attacks and apply it to real-time video.

We achieved 85% accuracy for our face recognition model on the validation dataset of the CASIA-FASD database using our deep CNN-based architecture.