

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR,  
DE LA RECHERCHE SCIENTIFIQUE ET DE LA TECHNOLOGIE

## UNIVERSITÉ SAAD DAHLEB BLIDA 1

Faculté des Sciences  
Département d'Informatique



### MEMOIRE DE FIN D'ETUDES EN VUE DE L'OBTENTION DU DIPLOME DE MASTER EN INFORMATIQUE

Option : Système Informatique et Réseaux (SIR)

## Thème

**Authentification Faciale Par l'Apprentissage Profond**

Organisme d'accueil

CDTA | Centre de Développement des Technologies Avancées



**Réalisé par :**

- *Mlle. AMEUR Kaouther*
- *Mme. AZOUZ Sarah*

**Promotrice :**

- *Mme. ZAHRA F.*

**Encadré par :**

- *Mme. AIT SAADI K.*
- *Mr. BENGHERABI M.*

**Soutenu le :** 23/10/2019

**Devant le jury composé de :**

- Mme. ARKAM présidente
- Mme. HADJ HENNI

Promotion : 2018 / 2019

# *Dédicaces*

*Ce travail est dédié*

*A mes chers parents "LATIFA " et "ZOUBIR" qui m'ont soutenu et encouragé et leurs prières leur soutien durant ces années d'études que dieu vous bénisses et vous protèges.*

*A mes chères sœurs SOUMIA et SELME , et mon chère frère YACOUB , mon neveu ABDERRAHIME.*

*A toute ma famille "AMEUR" et "MEBARKI".*

*A mon futur Mari Mr. MEZRAR ABDEKGHANI pour ses encouragements permanents, et son soutien moral.*

*A mes amies surtout de l'université.*

*A ma binôme SARAH.*

*A tous ceux qui m'ont aidé dans mon travail.*

*Et à tous ceux que j'ai connu durant mon cycle d'étude.*

*Merci d'être toujours là pour moi, Puisse Dieu vous donne santé, bonheur, courage et réussite.*

*ما توفيقى إلا بالله*

*Kaouther*

# *Dédicaces*

*Ce travail est dédié*

*A ma mère qui est décédée trop tôt, qui sera contente d'apprendre que sa fille a enfin terminé ses études J'espère que, du monde qui est sien maintenant.*

*A mon père NASER-EDINE qui m'a toujours poussé et motivé dans mes études, leur prière durant ces années d'études que dieu vous bénisse et vous protège.*

*A mon mari ELHOUARI MOHAMED qui n'a jamais cessé de me soutenir pour que je puisse finir mes études et avoir une bonne formation.*

*A mes chères sœurs FAIZA et HADJER, mon chère frère MAHDI, et mes neveux pour leur amour, leur tendresse, leur soutien.*

*A ma belle famille.*

*A toute ma famille "AZOUZ" et "AMER-ELKHEDOUD".*

*A Mme K.AIT SAADI qui ma aider dans mon travail.*

*A ma binôme KAOUTHER.*

*Et à tous ceux que j'ai connu durant mon cycle d'étude.*

*ما توفيقى إلا بالله*

**Sarah**

# *Remerciements*

*Nous tenons à remercier nos encadreurs Mme AIT SAADI et Mr BENGHERABI pour le temps qu'ils ont consacré et pour les précieuses informations qu'ils nous ont donné tout au long de notre travail.*

*Nous tenons à remercier notre enseignante et promotrice  
Mme ZAHRA  
pour ses recommandations, sa confiance et son aide.*

*Nous tenons aussi à remercier l'ensemble des enseignants de l'université  
SAAD DAHLEB.*

*Nous remercions les membres du jury pour avoir bien voulu évaluer et examiner notre travail.*

*Enfin, nous adressons nos remerciements les plus sincères à nos familles, tous nos amis ainsi que toute personne ayant contribué de près ou de loin à la réalisation de ce projet.*

*En espérant que ce modeste travail soit à la hauteur et reflète ce que nous avons pu acquérir pendant ce travail.*

# Résumé

L'authentification faciale joue un rôle essentiel dans les systèmes biométriques et de sécurité. Elle consiste à authentifier une ou plusieurs personnes d'une manière automatique, en fonction des bases de données stockées au préalable.

Ce projet vise à utiliser les images biométriques afin d'améliorer les performances d'authentification des modalités choisies en général et les modalités faciale en particulier.

En règle générale, Le principe des approches d'authentification des visages présentées dans la littérature extrait les caractéristiques (descripteurs) du visage d'un individu et les compare à celles extraites et stockées au préalable dans la base de données. Le but est de trouver la meilleure ressemblance (matching) <sup>1</sup>.

En second lieu, ce projet aussi vise à développer un système d'authentification biométrique multimodale. Ce dernier sera possible en utilisant les outils de reconnaissance appropriés qui respectent un protocole bien précis.

Le langage Python est le langage de programmation utilisé via la plate-forme PySide ainsi que les différentes bibliothèques offertes par ce langage de programmation. En outre, la technique "Deep-Learning", s'est mise à l'exploiter, ce qui surenchère des meilleures performances par rapport au techniques existante dans l'état de l'art.

---

1. établissement de correspondances

# Abstract

Face recognition plays a vital role in both biometric and security applications, which consists to authenticate one or many subject in an automatic manner depending on the facial databases stored.

This project aims to enhance face images in order to improve the authentications rate and the performance of the face modality. An in-depth study of the different solutions existed in the field of biometric guided us on a new approach of face recognition.

Typically, face authentication approaches extract the facial landmarks of an individual in the seek to compare it to a database and finally to find the best Matching.

Another goal of this project is to develop a biometric authentication system. The latter will be possible by using the appropriate recognition programs that respect a given protocol. Finally, platform as well as Python libraries have been used. Furthermore, Deep-Learning approach will be operated that others the best performance compared to the state of the art.

## ملخص

تلعب مصادقة الوجه دورًا حيويًا في أنظمة الأمن الحيوي. تتكون من مصادقة شخص أو أكثر بطريقة تلقائية ، وفقًا لقواعد البيانات المخزنة مسبقًا.

يهدف هذا المشروع إلى استخدام الصور البيومترية لتحسين أداء المصادقة من الطرائق المختارة بشكل عام وطرائق الوجه بشكل خاص.

عادة، يقدم مبدأ أساليب مصادقة الوجه وجه الفرد ويقارنها بتلك المستخرجة والمخزنة في قاعدة البيانات السابقة. الهدف هو العثور على أفضل.

ثانيًا ، يهدف هذا المشروع أيضًا إلى تطوير نظام مصادقة بيومترية متعدد الوسائط. سيكون الأخير ممكنًا باستخدام أدوات التعرف المناسبة التي تحترم بروتوكولًا محددًا

لغة Python هي لغة البرمجة المستخدمة عبر منصة PySide بالإضافة إلى المكتبات المختلفة التي توفرها لغة البرمجة هذه. بالإضافة إلى ذلك ، بدأت تقنية "التعلم العميق" في استغلالها ، والتي تفوقت على أفضل أداء مقارنة بالتقنيات الموجودة في أحدث التقنيات.

# Table des matières

<b>Remerciements</b> .....	iii
Introduction générale.....	1
<b>Chapitre 1: Systèmes Biométriques de reconnaissance et d'authentification</b> ....	3
1.1 Introduction .....	3
1.2 Historique .....	3
1.3 Procédé de vérification de l'identité d'une personne .....	4
1.4 Systèmes de sécurité physiques existants .....	5
1.4.1 Quelques système de sécurités physiques existants .....	5
a. L'accès par Badges .....	5
b. Clés .....	5
c. Carte crédit (bancaire) .....	6
d. Accès au clavier .....	6
e. Mot de passe .....	6
1.4.2 Les inconvénients des systèmes de sécurité physiques .....	6
1.5 Systèmes de sécurité biométriques .....	7
1.5.1 Les modalités biométriques .....	7
a. Les mesures biologiques .....	7
b. Les mesures morphologiques .....	7
1. Empreinte digitale.....	8
2. Visage .....	8
3. Iris.....	8
4. Signature .....	9
5. Géométrie de la main .....	9
6. Réseaux veineux .....	9



7.La voix .....	9
c. Les mesures comportementales .....	10
1.5.2 Applications des systèmes biométriques .....	10
a. L'identification .....	10
b. L'authentification .....	10
c. Reconnaissance .....	10
1.6 Le système de reconnaissance de visage .....	11
1.6.1 Détection de visage .....	12
a. La pose .....	12
b. La présence ou absence des composantes structurales .....	13
c. Les occlusions .....	13
d. Les conditions d'illumination .....	13
1.6.2 Extraction de caractéristiques du visage .....	14
1.6.3 Authentification .....	15
1.7 Reconnaissance faciale hétérogène .....	15
1.8 Mesure des performances d'un système biométrique .....	16
1.8 Conclusion .....	17
<b>Chapitre 2: Apprentissage basé sur les réseaux de neurones convolutifs.....</b>	<b>18</b>
2.1 Introduction .....	18
2.2 Apprentissage automatique .....	18
2.2.1 L'apprentissage supervisé .....	19
2.2.2 L'apprentissage Non supervisé .....	20
2.2.3 L'apprentissage semi-supervisé .....	20
2.2.4 L'apprentissage par renforcement .....	21

2.3. Les réseaux de neurones artificiels .....	21
2.3.1. Définition d'un neurone .....	22
2.3.2. Définition des réseaux de neurones .....	23
2.3.3 Fonction d'activation .....	23
a. Fonction d'activation linéaire .....	24
b. Fonction d'activation non linéaire .....	24
2.3.4 Les limitations des réseaux de neurones classiques .....	25
2.4 Apprentissage profond .....	26
2.4.1. Différence entre intelligence artificiel, l'apprentissage profond et l'apprentissage automatique .....	27
a. Intelligence Artificiel .....	27
b. Apprentissage automatique ou Machine Learning .....	28
c. Apprentissage profond ou Deep Learning .....	28
2.4.2. Historique .....	29
2.4.3. Domaines d'application .....	30
2.4.4. Réseaux de neurones profonds .....	31
2.4.4.1. Réseaux de neurones convolutionnels .....	32
a. Étape de convolution .....	32
b. Non linéarité (Relu) .....	32
c. Couche de Pooling .....	33
d. Couche entièrement connecté ou Fully Connected .....	34
2.4.4.2. Réseaux de neurones récurrents (RNN) .....	34
2.5 Concepts de l'apprentissage profond .....	34
• Le gradient .....	34
• Backpropagation (Rétropropagation) .....	35
• Vanishing Gradient (Disparition des gradients) .....	35
• Gradient Descent .....	36
• Sur-apprentissage(Overfitting) .....	36
2.6 Conclusion .....	37

<b>Chapitre 3: Etat de l'art sur les systèmes de détection et de la reconnaissance faciale .....</b>	<b>38</b>
3.1 Introduction .....	38
3.2 Les méthodes Standard de la détection .....	38
3.2.1 Méthode de Viola and Jones .....	38
3.2.2 Méthode basé sur LBP (Local Binary Patterns) .....	43
3.3 Les méthodes basées sur CNN les plus connus .....	46
3.3.1. FaceNet .....	46
3.3.2. LeNet-5 .....	48
3.3.3. <i>L'architecture AlexNet</i> .....	49
3.3.4. <i>GoogleNet</i> .....	50
3.3.5. <i>L'architecture VGG16</i> .....	52
3.4 Comparaison .....	53
3.5 Conclusion .....	53
<b>Chapitre 4: Mise en œuvre du système de reconnaissance et l'authentification faciale .....</b>	<b>54</b>
4.1 Introduction .....	54
4.2 Description de l'application .....	55
4.2.1 Phase de détection basée sur MTCNN .....	57
a. Le fonctionnement du bloc P-Net .....	57
b. Fonctionnement du bloc R-Net .....	60
c. Fonctionnement du bloc de traitement O-Net .....	61
4.2.2 La phase d'apprentissage .....	62
4.2.3 La phase de comparaison de reconnaissance et d'authentification.....	65
4.3 Conclusion .....	65

<b>Chapitre 5: Expérimentation et résultats .....</b>	<b>66</b>
5.1 Introduction.....	66
5.2 Environnement du travail .....	66
5.3 Technologies utilisées .....	66
Langage Python .....	66
Logiciel PyCharm .....	67
SQLite .....	67
Logiciel DB Browser (SQLite) .....	68
5.4 Outils utilisés .....	68
5.4.1 TKinter (Tool kit interface) .....	68
5.4.2 OpenVC (Open Source Computer Vision) .....	68
5.4.3 haarcascade_frontalface_default.xml .....	69
5.4.4 OS .....	69
5.4.5 NumPy .....	69
5.4.6 TensorFlow .....	69
5.4.7 Keras .....	70
5.5 Présentation des interfaces de l'application .....	70
5.6 Détection standard de visage .....	71
5.6.1 Détection et reconnaissance de visage par les modèles MTCNN et VGG-16 .....	72
5.7 Conclusion .....	77
 Conclusion général .....	 78

# Liste des Figures

Figure 1.1 : Mécanisme et procédé de vérification de l'identité d'une personne[6].....	4
Figure 1.2 : Accès par badge [8].....	5
Figure 1.3 : Représentation de la carte crédit [8] .....	6
Figure 1.4 : Les modalités biométriques morphologique [11] .....	7
Figure 1.5 : Dispositif représente l'empreinte digitale [13] .....	8
Figure 1.6 : Différentes formes de l'iris de l'œil humain [15].....	11
Figure 1.7 : dispositif de vérification de signature [8].....	11
Figure 1.8 : Exemple d'un visage d'une même personne subissant des variations de pose hors plan [19].....	12
Figure 1.9 : Exemples d'occlusion du visage [19] .....	13
Figure 1.10 : Exemple d'un visage d'une même personne subissant un changement de luminosité [19].....	13
Figure 1.11 : L'image de gauche de chaque paire est la photo d'identité et sur sa droite, une de ses photos correspondant [20] .....	15
Figure 1.12 : La relation entre les métriques FAR et FRR et EER évaluant un système biométrique.....	17
Figure 2.1 : Représentation graphique d'un neurone [23] .....	23
Figure 2.2 : Fonction d'activation linéaire [49] .....	24
Figure 2.3 : Fonction d'activation non linéaire [49] .....	24
Figure 2.4 : Données linéaire vs données non linéaire [38] .....	26
Figure 2.5 : Différence entre "Neural network" et "Deep neural network" [42] .....	27
Figure 2. 6 : Différence entre intelligence artificiel deep learning et machine .....	31
Figure 2. 7 : Différence entre le Deep Learning et le Machine Learning [44] .....	31
Figure 2. 8 : Historique de deep learning [47].....	30
Figure 2. 9 : Relu Non-Linéarité[28] .....	33
Figure 2. 1 : Étape Max/average Pooling [27] .....	33
Figure 2. 11 : Architecture d'un réseau de neurones récurrents [41] .....	34
Figure 2. 12 : Visualisation du gradient. Les flèches rouges représentent le gradient de la fonction tracée en bleu [81] .....	35

Figure 2. 13 : Visualisation du ‘gradient descent’ en fonctionnement [82] .....	36
Figure 2. 14 : Exemple sur l'apprentissage [86] .....	37
Figure 3.1 : Exemple de quatre caractéristiques de Haar [58] .....	39
Figure 3.2 : Exemple de l’algorithme Adaboost [59] .....	41
Figure 3.3 : Cascade de classifieurs forts. A chaque étage, uniquement les candidats classifiés positifs sont transmis a l’étage suivant [59] .....	42
Figure 3.4 : Fonctionnement de l’opérateur LBP basique [65] .....	44
Figure 3.5 : Extension de l’opérateur lbp basique. Les entourages (8,1), (16,2), (8,2) sont représentés [66] .....	44
Figure 3.7 : Illustrations de FaceNet [70] , [71] .....	46
Figure 3.8 : La fonction de perte de triplets [73] .....	47
Figure 3. 9 : Illustre des exemples d’images ancre, vérité et négative [74] .....	48
Figure 3.10 : Architecture du réseau LeNet [77] .....	48
Figure 3.11 : Architecture du réseau AlexNet [79] .....	50
Figure 3.12 : Architecture du réseau Google net [80] .....	51
Figure 3.13 : Architecture d’un module ‘Inception’ [36] .....	51
Figure 4.1 : Diagramme blocs du système développé .....	56
Figure 4.2 : L’architecture du modèle MTCNN [88] .....	57
Figure 4.3 : Image pyramide .....	58
Figure 4.4 : Le noyau et la fenêtre glissante .....	59
Figure 4.5 : Le réseau analyse la sortie P-Net .....	60
Figure 4.6 : Le réseau rejette un grand nombre de faux candidats .....	61
Figure 4.7 : le cadre de sélection, les 5 points de repère faciaux .....	62
Figure 4.8 : Architecture du réseau VGG16 [36].....	63
Figure 5.1: Logo de Python .....	66
Figure 5.2 : Logo de PyCharm .....	67
Figure 5.3 : Logo de DB Browser .....	68
Figure 5.4 : Logo de TKinter .....	68
Figure 5.5 : Logo de OenCV .....	68

Figure 5.6 : Logo de NumPy .....	69
Figure 5.7 : Logo de TensorFlow .....	70
Figure 5.8 : Logo de Keras .....	70
Figure 5.9 : Interface principale de l'application.....	71
Figure 5.10 : Détection des visages par Haar cascade .....	72
Figure 5.11 : Personne non reconnu par le système .....	73
Figure 5.12 : L'image alignée et stockée dans la base de données images .....	73
Figure 5.13 : La fenêtre montrant l'ajout d' un nouveau personne .....	74
Figure 5.14 : La personne ajoutée dans la base .....	75
Figure 5.15 : Affichage de la liste de personnes stockées .....	75
Figure 5.16 : Authentification de la personne .....	76
Figure 5.17 : Image sur un téléphone portable .....	77

# Liste des Tableaux

Tableau 2.1 : Fonctions d'activation [50] .....	25
Tableau 3.1 : Tableau résumant les différentes couches de Inception 'GoogleNet [84].....	52
Tableau 3.2 : Tableau de comparaison entre VGG16 et autres architectures .....	53
Tableau 4.1 : Architecture VGG16 [93].....	64



## Introduction Générale

Le monde dans lequel nous vivons aujourd'hui s'est marqué par son insécurité, et la recrudescence des attentats [1]. La sécurité est devenue donc une préoccupation des institutions tant en étatique que en privé.

La reconnaissance automatique des visages humains continue d'attirer des chercheurs du domaine de la vision artificielle, et reste l'une des techniques les plus étudiées au cours des dernières décennies [2].

Dans son environnement quotidien, un individu a besoin de s'identifier. Donc Il est évident que nous cherchons une solution efficace et ergonomique. Cette solution est basée sur l'analyse de l'image faciale en vue de développer une application capable de sécuriser la vie privée des personnes.

La reconnaissance automatique des visages sert à identifier ou authentifier une ou plusieurs personnes à partir d'images fixes ou dynamique (vidéo), en utilisant une base de données de visages stockée au préalable. Dans le cas où la scène est encombrée, la détection des visages s'avère nécessaire pour segmenter et extraire des caractéristiques des régions des visages. Dans les problèmes d'identification, le système d'identification renvoie une identité déterminée à partir d'une base de données d'individus déjà connus, alors que dans les problèmes de vérification, le système doit confirmer ou rejeter l'identité [3].

C'est dans ce contexte que s'insère notre mémoire de Master 2, il s'agit de développer une application d'authentification et de reconnaissance faciale qui répond à des besoins de gestion et de contrôle d'accès des employés du Centre de Développement des Technologies Avancées (CDTA).

Dans ce mémoire, nous allons discuter les différents travaux développés dans le domaine de l'authentification biométrique d'une manière générale et de la modalité faciale d'une manière particulière. Nous présentons d'abord le contexte dans lequel notre projet se déroule. Par la suite, on survole les motivations ainsi que l'intérêt de

l'identification des personnes. Nous présentons ensuite, les problèmes majeurs à résoudre et nous terminerons par nos contributions apportées.

Ce mémoire est organisé comme suit :

Le chapitre I «**Systèmes Biométriques de reconnaissance et d'authentification**» : ce chapitre expose les notions essentielles pour la compréhension de certains concepts basiques des systèmes biométriques en général tels que les différentes modalités biométriques, et les métriques utilisées pour évaluer un système biométrique. Il inclut aussi les différents traitements que peut inclure un système biométrique en citant la détection, la reconnaissance et l'identification.

Le chapitre II «**Apprentissage basé sur les réseaux de neurones convolutifs**» . Ce chapitre introduit l'apprentissage et exposé la différence entre les réseaux de neurones classiques et l'apprentissage profond.

Le chapitre III «**Etat de l'art sur les systèmes de détection et de la reconnaissance faciale**» : Ce chapitre vise à passer en revue les différentes étapes et méthodes de détection automatique de visage en passant en commençant par les méthodes standard de la reconnaissance faciale qui procèdent par l'extraction des caractéristiques locales des visages puis une classification pour une éventuelle reconnaissance et authentification des personnes, suivis par les méthodes récentes basées sur la technologies réseau de neurones convolutifs (CNN)

Le chapitre IV «**Mise en œuvre du système de reconnaissance et d'authentification faciale** » : Ce chapitre présente en détails l'architecture VGG 16 adoptée dans l'application réalisée de reconnaissance et d'authentification, ainsi que les modules qui le composent tels que l'acquisition et l'alignement des visages pour la détection.

Le chapitre IV «**Expérimentation et résultats**» : ce chapitre fournit une vision détaillée de l'implémentation de notre système, puis l'analyse des résultats obtenus par les différents tests effectués.

# **Chapitre1**

## **Systèmes Biométriques de reconnaissance et d'authentification**

### **1.1 Introduction**

Actuellement l'insécurité, l'authentification des individus est devenue un enjeu majeur pour la protection des lieux sensible. En effet, les progrès enregistrés dans le domaine de la reprographie numérique ont ouverts des portes à des malfaiteurs pour exceller dans la falsification des pièces d'identité de type (carte d'identité et passeport). Ceci a poussé les scientifiques et gouvernements à réfléchir à des nouveaux moyens qui permettraient d'authentifier efficacement et avec un grand pourcentage de certitude une personne.

C'est à partir de là qu'une nouvelle tendance visant à remplacer l'identité d'une personne (définie comme un ensemble d'informations reportées sur un document officiel) par un ensemble d'informations extraites de la personne elle-même a vu le jour. Cette nouvelle science appelée : l'authentification biométrique qui permet à l'avenir d'authentifier les personnes à partir de leurs caractéristiques physiologiques ou comportementales.

Dans ce qui suit, nous allons d'abord définir et présenter quelques systèmes d'authentification biométrique monomodaux en citant les avantages et inconvénients de chaque modalité. Enfin, nous justifierons notre choix de la reconnaissance faciale comme modalité de travail [4].

### **1.2 Historique**




La reconnaissance faciale automatique est un concept relativement nouveau. Le premier système semi-automatisé de la reconnaissance faciale a été développé dans les années 1960, il nécessite à l'administrateur de localiser les yeux, les oreilles, le nez et la bouche sur la photo et de saisir les distances calculées et les ratios à un point de référence commun, qui ont ensuite été comparés aux données de référence.

## Chapitre 1 : Systèmes Biométriques de reconnaissance et d'authentification

Dans les années 1970, Goldstein, Harmon et Lesk ont utilisé 21 marqueurs spécifiques tels que la couleur des cheveux et l'épaisseur de la lèvre pour automatiser la reconnaissance. Le problème avec ces deux premières solutions, c'est que les mesures et les emplacements ont été calculés manuellement. En 1988, Kirby et Sirovich ont appliqué l'analyse en composantes principales (ACP). Une technique standard de l'algèbre linéaire. Cela a été considéré comme une étape importante, car elle a montré qu'au moins une centaine de valeurs ont été nécessaires pour coder convenablement et avec précision une image alignée et normalisée. En 1991, Turk et Pentland ont découvert que lorsque vous utilisez la technique Eigenfaces (ACP), l'erreur résiduelle peut être utilisée pour détecter un visage dans une image, une découverte qui a permis la reconnaissance faciale automatique en temps réel. Bien que l'approche fût peu limitée par des facteurs environnementaux, elle a néanmoins créé un intérêt significatif pour promouvoir le développement des technologies de la reconnaissance faciale automatique. Cette technologie a été mise en essai en janvier 2001 lors de la finale du championnat de football américain SUPER BOWL en capturant des images de surveillance puis comparées à une base de données numérique. Aujourd'hui la reconnaissance faciale est utilisée dans plusieurs domaines [5].

### 1.3 Procédé de vérification de l'identité d'une personne

La vérification automatique de l'identité d'une personne peut faire appel à l'une des trois types d'informations suivantes (Figure 1.1) :

	 Posseder	 Mémoriser	 Biométrie
Copie	■■■■	■■■■	■■■■■
Volée	■■■■	■■■■	■■■■■
Oublié	■■■■	■■■■	■■■■■
Perdue	■■■■	■■■■■	■■■■■

**Figure 1.1 : Mécanisme et procédé de vérification de l'identité d'une personne [6].**

## Chapitre 1 : Systèmes Biométriques de reconnaissance et d'authentification

1. Ce que détient une personne (Une carte avec puce électronique, l'accès par Badges, etc...).
2. Ce que mémorise une personne (mot de passe, code PIN , code etc...).
3. Ce qui caractérise une personne (empreinte digitale, visage, iris, signature etc...).

### 1.4 Systèmes de sécurité physiques existants

Le contrôle d'accès est une technique visant à minimiser les risques d'accès des intrus dans des entreprises ou des organisations. L'objectif du contrôle d'accès c'est déterminé qui peut accéder dans un environnement ou un milieu bien déterminé [7]. La vérification automatique de l'identité d'une personne peut faire appel à l'une des types d'informations suivantes :

#### 1.4.1 Quelques système de sécurités physiques existants

- a. **L'accès par Badges** : est une solution informatique pour un système de surveillance permet, par un procédé électronique, de gérer les flux entrants et sortants à certains locaux ou installations. Le badge de sécurité se présente généralement sous la forme d'une carte ou d'une clé que le salarié introduit dans un appareil lui permettant l'accès s'il fait partie des personnes autorisées. Il s'agit donc d'un système d'identification des personnes entrant dans un lieu donné ou utilisant un équipement particulier (Figure1.2).



**Figure 1.2 : Accès par badge [8].**

- b. **Clés** : Clé cryptographique, composante de la bicyclette, qui est connue de son unique propriétaire et utilisée par lui seul pour déchiffrer un message dont il est le destinataire, ou pour signer un message dont il est l'expéditeur.

## Chapitre 1 : Systèmes Biométriques de reconnaissance et d'authentification

- c. **Carte crédit (bancaire)** : Une carte de crédit peut être délivrée par une banque, ou par exemple un grand magasin, ou un supermarché. Une carte de crédit est associée une réserve d'argent. Les cartes de crédit les plus courantes sont les cartes Visa et Mastercard (Figure 1.3). On peut utiliser une carte de crédit pour retirer de l'argent à un distributeur, ou bien pour régler des achats dans magasins. Ces montants ne sont pas débités directement du compte bancaire, mais sont débités plus tard.



*Figure 1.3 – Représentation de la carte crédit [8].*

- d. **Accès au clavier** : Les systèmes de contrôle d'accès au clavier sont une forme électronique de la serrure à combinaison.
- e. **Mot de passe** : C'est un mot secret ou expression connue seulement d'un groupe restreint de personnes [9]. Il peut être aussi une séquence de caractères ou mot qu'un sujet donne à un système pour authentification, validation ou vérification. Un mot de passe fort permet d'avantage de sécurité.

### **1.4.2 Les inconvénients des systèmes de sécurité physiques**

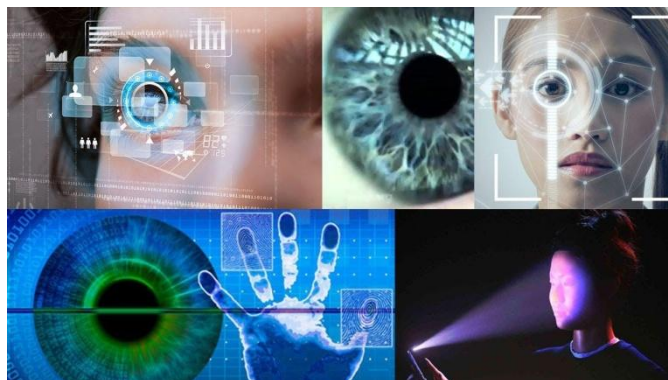
Bien qu'ils ornent de nombreux avantages, les systèmes d'accès physiques ne sont pas exempts de risques ni complètement sécurisés. La plupart des vulnérabilités, des systèmes physiques, sont dues à des risques de vol et de perte, Les cartes électroniques posent actuellement un réel problème surtout pour les cartes bancaires. En outre, la perte, l'oubli et la découverte des mots de passe associés rendent les deux premiers types d'information peu et moins sécurisés. Le fait d'inclure les propriétés physiologiques et comportementales des individus dans tout

## Chapitre 1 : Systèmes Biométriques de reconnaissance et d'authentification

processus d'authentification des personnes augmenterait considérablement leur niveau de sécurité.

### 1.5 Systèmes de sécurité biométriques

Les systèmes de sécurité biométriques repose la biométrie des personnes. Leur principe se base sur la reconnaissance des caractéristiques physiologiques et comportementale d'une personne (Figure 1.4). De surcroît, ces systèmes peuvent offrir plus de sécurité et de commodité que méthodes de sécurité traditionnelles. Par ailleurs, les systèmes biométriques reposent sur des caractéristiques acquissent par un capteur dédié [10].



*Figure 1.4 : Les modalités biométriques morphologique [11].*

#### 1.5.1 Les modalités biométriques

On distingue trois types de mesures biométriques [12] :

##### **a. Les mesures biologiques**

Ces mesures sont basées sur les caractéristiques biologiques avec trace de type : analyse l'ADN, la salive, etc... Ces procédés restent contraignants à mettre en œuvre et sont favorisés aux domaines médicaux et judiciaire.

##### **b. Les mesures morphologiques**

Ces métriques sont extraites d'une partie du corps humain. La Figure 1.5 présente les propriétés des principaux descripteurs biométriques susceptibles d'être utilisés dans les systèmes biométriques.

## **Chapitre 1 : Systèmes Biométriques de reconnaissance et d'authentification**

1. Empreinte digitale: Les empreintes digitales sont les marques laissées par les sillons des pulpes digitales (Figure 1.5). Elles permettent d'identifier de façon unique un individu. Le dessin qu'elles forment est propre à chacun d'entre nous, ce qui explique pourquoi les empreintes digitales servent à l'identification des personnes.



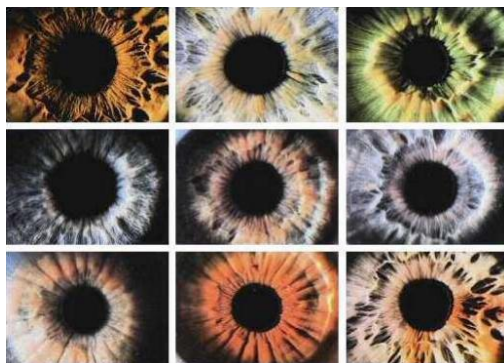
**Figure 1.5 : Dispositif représente l'empreinte digitale [13].**

En effet, la probabilité que deux individus partagent la même empreinte a été estimée à une sur 64 milliards. Les empreintes digitales possèdent une propriété intéressante qui ne dépend du patrimoine génétique. Elles diffèrent donc chez les vrais jumeaux. Elles ne s'altèrent pas avec le temps. L'analyse consiste à reconnaître une douzaine de points caractéristiques. Le problème principal des empreintes digitales est qu'elles peuvent être volées, car on risque de les laisser derrière soi à chaque fois que l'on touche un objet (empreinte latente) [14].

2. Visage: le visage d'une personne, en référence à la forme ou aux proportions des traits du visage censés révéler des qualités de tempérament ou de caractère.

3. Iris :L'iris est la partie colorée de l'œil qui entoure la pupille (Figure 1.6). Contrairement à ce que l'on pourrait croire, la reconnaissance d'iris ne dépend nullement de la couleur de celle-ci mais se base plutôt sur des points caractéristiques. On peut considérer que l'iris en possède un nombre infini. De plus, ils varient très peu au cours de la vie. Pour empêcher que l'appareil ne soit trompé par une photographie posée devant le capteur, ce dernier peut utiliser les infrarouges et vérifier que la pupille se dilate correctement en présence de lumière.





**Figure 1.6 : Différentes formes de l'iris de l'œil humain [15].**

4. Signature: La signature est une représentation manuscrite d'une personne écrite sur des documents comme preuve d'identité.



**Figure 1.7 : Dispositif de vérification de signature [8].**

5. Géométrie de la main : Il s'agit de reconnaître la forme de la main. C'est un mode de reconnaissance peu utilisé.
6. Réseaux veineux : Il est possible de repérer les réseaux veineux pour en extraire quelques points caractéristiques. Cette méthode est souvent utilisée en combinaison avec l'analyse des formes de la main. Il est également possible de réaliser une mesure du réseau veineux de la rétine (la paroi au fond de l'œil). Cette mesure, à ne pas confondre avec la reconnaissance de l'iris.
7. La voix : Cette biométrie est en train de devenir une nouvelle forme d'empreinte digitale. C'est une technologie de plus en plus sophistiquée, capable de détecter des nuances sonores inaudibles pour les humains. Elle permet de relever des indices sur le possible endroit où se trouve une personne, son état de santé, voire des caractéristiques physiques. Le

## **Chapitre 1 : Systèmes Biométriques de reconnaissance et d'authentification**

problème de la voix est qu'elle est assez changeante. où la difficulté tient plutôt à la nécessité de reconnaître n'importe quelle voix dans des conditions acoustiques.

### **c. Les mesures comportementales :**

C'est mesures sont basées sur des données dérivées d'une action exécutée par l'utilisateur tels que la manière de signer, la démarche, le geste).

### **1.5.2 Applications des systèmes biométriques**

Les systèmes biométriques sont exploités pour divers applications la reconnaissance et l'identification et l'authentification.

- (a) L'identification : Est le premier mécanisme de défense contre une entité non-autorisée d'accès aux ressources d'un système informatique [16]. Il s'agit de saisir une donnée biométrique d'une personne, en prenant par exemple une photo de son visage, en enregistrant sa voix, ou en captant l'image de son empreinte digitale. Le système automatique doit déterminer qui est la personne par la comparaison de ces données biométriques avec toutes les autres données des personnes qui figurent dans une base prédéfini d'utilisateurs connus. Dans cette approche, on pose une question simple : « qui vous êtes ? ».
- (b) L'authentification : La personne doit fournir une identité et que le système doit faire correspondre la caractéristique de la personne à reconnaître à l'ensemble des caractéristiques stockées dans la base de données, déterminant ainsi si l'une d'elles est suffisamment similaire à celle considérée pour accepter ou refuser [17].
- (c) Reconnaissance: la reconnaissance faciale consiste à trouver le modèle qui correspond le mieux au visage pris en entrée à partir de ceux stockés

## **Chapitre 1 : Systèmes Biométriques de reconnaissance et d'authentification**

dans la base de données, il est caractérisé par son taux de reconnaissance.

Dans notre travail, nous nous intéressons à l'application de reconnaissance et d'authentification faciale, dans le prochain paragraphe, nous aborderons en détails les différentes étapes à suivre pour réaliser système de reconnaissance et d'authentification de visage standard (traditionnel) et présenterons les principales difficultés rencontrées.

### **1.6 Le système de reconnaissance de visage**

La reconnaissance automatique de visage s'effectue en quatre étapes principales à savoir:

1. Détection de visages.
2. Extraction et normalisation des caractéristiques du visage.
3. Authentification.
4. La reconnaissance.

Certaines techniques de traitements d'images peuvent être communes à plusieurs étapes. Par exemple, l'extraction des caractéristiques faciales (yeux, nez, bouche) est utilisée aussi bien pour la détection que pour l'identification de visages. Par ailleurs, les étapes de détection de visage et d'extraction de caractéristiques peuvent être exécutées simultanément. Cela dépend notamment de la nature de l'application, de la taille de la base d'apprentissage, et des conditions de prise de vue (bruit, occultation, etc.). Enfin, les techniques de traitement utilisées dans chaque étape sont très critiques pour les applications biométriques, et doivent, par conséquent, être optimisées pour améliorer les performances du système global [18].

### 1.6.1 Détection de visage

L'efficacité des systèmes biométriques basés sur l'authentification de visage dépend essentiellement de la méthode utilisée pour localiser le visage dans l'image. Dans la littérature scientifique, le problème de localisation de visages est aussi désigné par la terminologie "détection de visages". Plusieurs travaux de recherches ont été effectués dans ce domaine. Ils ont donné lieu au développement d'une multitude de techniques allant de la simple détection du visage, à la localisation précise des régions caractéristiques du visage, tels que les yeux, le nez, les narines, les sourcils, la bouche, les lèvres, les oreilles. Les solutions proposées jusqu'à maintenant sont loin d'être satisfaisantes car elles fonctionnent uniquement dans des environnements contrôlés, et par conséquent elles ne gèrent pas la variabilité des conditions d'acquisition de la vie quotidienne, notamment :

- a. **La pose** : La pose d'un visage définit la rotation qu'a pu subir un visage lors de la capture. Les variations de pose peuvent être de deux types selon le type de rotation (Figure 1.8) : rotation dans le plan où l'axe de rotation est l'axe de la caméra et rotation hors-plan. Les variations de pose affectent grandement les systèmes de reconnaissance automatique de visages.



**Figure 1.8 : Exemple d'un visage d'une même personne subissant des variations de pose hors plan [19].**

b. **La présence ou absence des composantes structurales** : Les caractéristiques faciales tels que la barbe, la moustache, et les lunettes causent une grande variabilité des composantes structurales du visage, notamment au niveau de la forme, de la couleur, et de la taille.

c. **Les occlusions** : Les partielles apparaissent fréquemment dans des applications réelles, comme illustré sur la figure 1.9. Elles peuvent être causées par une main cachant une partie du visage, par des cheveux longs, des lunettes de vue, de soleil, par tout autre objet (foulard, etc...), ou encore par une autre personne. Il arrive également qu'une partie du visage en cache une autre, comme dans le cas d'une rotation hors plan par exemple.



**Figure 1.9 : Exemples d'occlusion du visage[19].**

d. **Les conditions d'illumination**: Les variations d'illumination viennent entraîner des variations considérables dans l'apparence d'un visage. La figure 1.10 présente un exemple de visage dont la source lumineuse l'éclairant se déplace. De nombreuses approches ont été proposées pour gérer ces problèmes de luminosité. Une modélisation implicite de la luminosité dans la création d'un modèle de visage peut être réalisée. L'extraction de caractéristiques invariantes aux changements de luminosité est également une approche largement décrite dans la littérature. Notons enfin qu'un certain nombre d'approches traitent le problème de la luminosité en amont de la reconnaissance par une étape de prétraitement dont l'objectif principal est bien souvent de corriger les artefacts dû aux variations de luminosité.



***Figure 1.10 : Exemple d'un visage d'une même personne subissant un changement de luminosité [19]***

### **1.6.2 Extraction de caractéristiques du visage**

L'extraction des caractéristiques telles que les yeux, le nez, la bouche est une étape prétraitement nécessaire à la reconnaissance faciale. On peut distinguer deux pratiques différentes :

1. la première repose sur l'extraction de régions entières du visage, elle est souvent implémentée avec une approche globale de reconnaissance de visage.
2. La deuxième pratique extrait des points particuliers des différentes régions caractéristiques du visage, tels que les coins des yeux, de la bouche et du nez. Elle est utilisée avec une méthode locale de reconnaissance et aussi pour l'estimation de la pose du visage.

Par ailleurs, plusieurs études ont été menées afin de déterminer les caractéristiques qui semblent pertinentes pour la perception, la mémorisation et la reconnaissance d'un visage humain. Les caractéristiques pertinentes rapportées sont : les cheveux, le contour du visage, les yeux et la bouche.

Cette étude a également démontré le rôle important que joue le nez dans la reconnaissance faciale à partir des images de profil. La forme du nez est plus intéressante que les yeux ou la bouche. Dans les auteurs ont particulièrement établi que la partie supérieure du visage est plus utile pour la reconnaissance faciale que la partie inférieure.

### 1.6.3 Authentification

Le module de reconnaissance exploite les caractéristiques du visage ainsi extraites pour créer une signature numérique qu'il stocke dans une base de données. Ainsi, à chaque visage de la base est associée une signature unique qui caractérise la personne correspondante. La reconnaissance d'un visage requête est obtenue par l'extraction de la signature requête correspondante et sa mise en correspondance avec la signature la plus proche dans la base de données. La reconnaissance dépend du mode de comparaison utilisé [18].

### 1.7 Reconnaissance faciale hétérogène

La reconnaissance faciale Hétérogène est un sujet d'actualité. Il se réfère généralement à la reconnaissance faciale entre deux modalités différentes, y compris des images du spectre visible, des images infrarouges proche, les images infrarouge thermique, des croquis composites [20] (Figure 1.11).



**Figure 1.11 : L'image de gauche de chaque paire est la photo d'identité et sur sa droite, une de ses photos correspondant [20].**

## 1.8 Mesure des performances d'un système biométrique

Les performances du système biométrique sont principalement définies par le niveau de sécurité. Généralement deux mesures d'évaluation des performances sont couramment utilisées à savoir :

- **Le taux de fausse acceptation FAR (False Acceptance Ratio)** : cette métrique est une unité utilisée pour mesurer le nombre moyen de fausses acceptations dans un système de sécurité biométrique. Il mesure et évalue l'efficacité et la précision d'un système biométrique en déterminant le taux de vérification des utilisateurs non autorisés ou illégitimes sur un système donné. FAR est également appelé "False Acceptance Rate" ou "type 2 error rate/ratio". FAR aide principalement à déterminer le niveau de précision d'un système de sécurité biométrique. Il est donné par :

$$FAR = \frac{FA}{TA} \quad (1.1)[21]$$

où :

**FA** est le nombre de fausses acceptations,

**TA** est correspond au nombre total de tentatives.

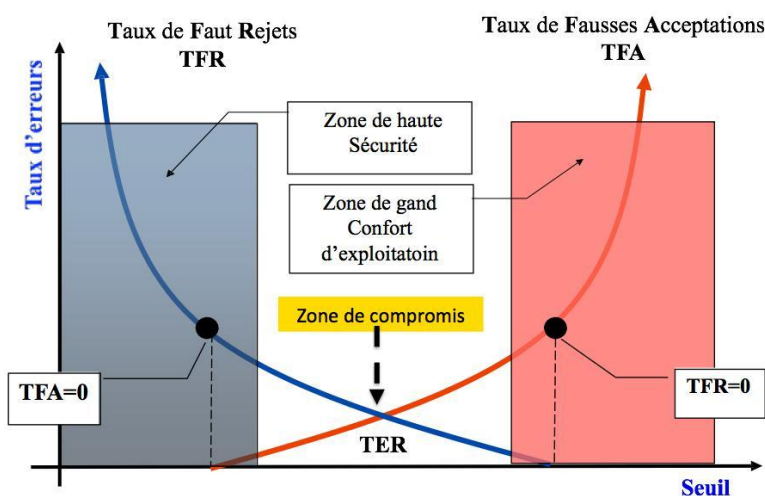
Par exemple, si le nombre de fausses acceptations est de 100 sur un total de 1 000 tentatives d'identification, le FAR est de 0,10. Cela signifie que le système biométrique autorisera 10%, ou 1 tentative sur 10, d'accéder au système [22].

- **FRR (False Rejection Rate)** : cette mesure correspond à la fréquence des rejets par rapport aux personnes qui doivent être correctement vérifiées. Notons qu'un faux rejet ne signifie pas nécessairement une erreur du système ; Par exemple, dans le cas des systèmes basés sur les empreintes digitales, le faux rejet peut être dû à un doigt mal placé sur le capteur ou par des modifications de la peau dues à des problèmes de peau ou à des plaies.

En règle générale, FAR et FRR dépendent du seuil d'acceptation  $t$ , utilisé pour définir le niveau de sécurité souhaité. Ces deux métriques sont strictement liés l'une de



l'autre par le taux d'erreur EER (Equal Error Rate). Ce dernier indique l'erreur du système lorsque FRR est égale à FAR [4] (Figure 1.12).



**Figure 1.12 : La relation entre les métriques FAR et FRR et EER évaluant un système biométrique.**

### 1.8 Conclusion :

Dans ce chapitre, nous avons présenté les concepts de base utilisés dans les systèmes biométriques pour la reconnaissance et l'authentification. Nous avons aussi donné un aperçu sur les techniques de mesure de leurs performances. Cette étude nous a permis de constater que la reconnaissance de visage ainsi la vision par ordinateur suscite de plus en plus l'intérêt de la communauté scientifique, vu qu'elle présente plusieurs défis. Enfin, nous avons mis en évidence les différentes difficultés inhérentes aux systèmes de détection, de reconnaissance et d'authentification visages, ce qui nous a permis de bien définir les problématiques traitées dans ce mémoire. Dans le chapitre suivant, nous aborderons les la technologie de l'apprentissage automatique et particulièrement les réseaux de neurones profonds.

## Chapitre2

# Apprentissage basé sur les réseaux de neurones convolutifs

### 2.1 Introduction

L'intelligence artificielle souvent abrégée avec le sigle IA, elle est une branche de l'informatique fondamentale, qui cherche à résoudre des problèmes logiques ou algorithmiques complexes qui ne peuvent pas être résolus avec des algorithmes classiques ou qui sont très difficiles à programmer.

Plus précisément IA est une technique informatique qui permet à la machine de reproduire des fonctions que seul l'être humain peut effectuer du fait de leur complexité. En d'autres termes elle constitue des méthodes imitant ou remplaçant l'être humain. Mais c'est à l'homme d'entraîner la machine ce qu'elle doit apprendre pour l'aider à reproduire sa propre représentation du monde et c'est ce qu'on appelle l'apprentissage automatique, en anglais « Le machine Learning ».

Dans ce chapitre nous présentons une étude sur les réseaux de neurones, ainsi qu'une introduction au apprentissage profond, suivie des architectures actuellement utilisées.

### 2.2. Apprentissage automatique

L'apprentissage automatique fait habituellement référence aux changements dans les systèmes qui exécutent des tâches associées à l'intelligence artificielle (IA). De telles tâches comprennent la reconnaissance, le diagnostic, la planification, la commande de robot, la prédiction, etc. Les " changements " pourraient être soit des améliorations à des systèmes déjà performants, soit une synthèse de nouveaux systèmes [19].

Nous pourrions dire, qu'une machine apprend chaque fois qu'elle modifie sa structure, son programme ou ses données (sur la base de ses entrées ou en réponse

## Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs

à des informations externes) d'une manière telle que ses performances futures s'améliorent. Certains de ces changements, comme l'ajout d'un document à une base de données, se situent confortablement dans la province d'autres disciplines et ne sont pas nécessairement mieux compris pour ce qu'on appelle apprentissage. Mais, par exemple, lorsque la performance d'une machine de reconnaissance vocale s'améliore après avoir entendu plusieurs échantillons de la parole d'une personne, nous nous sentons tout à fait justifiés dans ce cas de dire que la machine a appris [19].

Les méthodes et techniques d'apprentissage automatique sont divisées en plusieurs catégories, adoptant chacune d'elle un type d'apprentissage. Ces types d'apprentissage sont explicités dans ce qui suit :

### 2.2.1.L'apprentissage supervisé

L'apprentissage supervisé ajuste les paramètres du réseau par une comparaison directe entre la sortie réseau réelle et la sortie souhaitée. L'apprentissage supervisé est un système de rétroaction en boucle fermée, où l'erreur est le signal de rétroaction. La mesure d'erreur montre la différence entre la sortie réseau et la sortie des échantillons de l'entraînement. La mesure de l'erreur est habituellement définie par l'erreur quadratique moyenne (EQM).

$$E = \frac{1}{N} \sum_{p=1}^N \|y_p - \hat{y}_p\|^2, \quad (3.1)$$

où  $N$  est le nombre de paires de motifs dans l'ensemble d'échantillons,  $y_p$  est la partie de sortie de la  $p^{\text{ième}}$  paire de motifs, et  $\hat{y}_p$  est la sortie du réseau correspondant à la paire de motifs  $p$ .

L'erreur  $E$  est calculée à nouveau après chaque époque. Le processus d'apprentissage se termine lorsque  $E$  est suffisamment petit ou lorsqu'un critère d'échec est satisfait [20].

### **2.2.2.L'apprentissage Non supervisé**

L'apprentissage non supervisé n'implique aucune valeur cible. Il essaie d'associer automatiquement l'information des entrées avec une réduction intrinsèque de la dimensionnalité des données ou de la quantité totale des données d'entrée. L'apprentissage non supervisé est uniquement basé sur les corrélations entre les données d'entrée, et est utilisé pour trouver les motifs ou les caractéristiques significatives dans les données d'entrée sans l'aide d'un enseignant. L'apprentissage non supervisé est particulièrement adapté à l'apprentissage biologique en ce sens qu'il ne dépend pas d'un enseignant et qu'il utilise des primitives intuitives comme la compétition et la coopération neuronale [20].

### **2.2.3. L'apprentissage semi-supervisé**

Dans de nombreuses applications d'apprentissage automatique, telles que la bio-informatique, le web et le texte mining, la catégorisation de texte, le marketing de base de données, la détection du spam, la reconnaissance des visages et l'indexation vidéo, des quantités abondantes de données non étiquetées peuvent être collectées automatiquement. Cependant, l'étiquetage manuel est souvent lent, coûteux et sujet aux erreurs. Lorsque seul un petit nombre d'échantillons étiquetés est disponible, des échantillons non étiquetés pourraient être utilisés pour prévenir la dégradation de la performance due à un sur apprentissage.

L'objectif de l'apprentissage semi-supervisé est d'employer une grande collection de données non étiquetées. Avec quelques exemples étiquetés pour améliorer la performance de généralisation. Certaines méthodes d'apprentissage semi-supervisé sont basées sur des hypothèses selon lesquelles relier la probabilité  $P(x)$  à la distribution conditionnelle  $P(Y = 1|X = x)$ . L'apprentissage semi-supervisé est lié au problème de l'apprentissage transductif [20].

### **2.2.4. L'apprentissage par renforcement**

L'apprentissage par renforcement est un cas particulier d'apprentissage supervisé, dans lequel la valeur exacte de la sortie désirée n'est pas connue. L'enseignante ou l'enseignant ne fournit qu'une évaluation de la réussite ou l'échec d'une réponse. Cela est plus plausible sur le plan cognitif que l'apprentissage supervisé. Étant donné qu'une réponse correcte et entièrement spécifiée n'est pas toujours à la disposition de l'apprenant ou même l'enseignant [20].

Il est fondé uniquement sur l'information permettant de savoir si la sortie réelle est proche ou non de l'estimation. L'apprentissage par renforcement est une procédure d'apprentissage qui récompense le réseau neuronal pour son bon résultat de sortie et le punit pour le mauvais résultat de sortie [20].

Le calcul explicite des dérivés n'est pas nécessaire. Ceci, cependant, présente un processus d'apprentissage plus lent. Pour un système de contrôle, si le contrôleur fonctionne toujours correctement après une entrée, la sortie est jugée bonne ; sinon, elle est considérée comme mauvaise. L'évaluation de la sortie binaire, appelée renforcement externe, est utilisée comme signal d'erreur [20].

### **2.3. Les réseaux de neurones artificiels**

Les réseaux de neurones artificiels sont communément appelés systèmes connexionnistes. Ces systèmes sont capables d'apprendre tout en améliorant progressivement les performances afin de réaliser différentes tâches en considérant quelques exemples mais sans programmation spécifique à la tâche [22], par exemple, dans la reconnaissance d'images, les réseaux de neurones artificiels pourraient facilement apprendre à identifier les images, qui contiennent le visage simplement en analysant différentes images d'exemple qui ont été étiquetées manuellement comme des images de visage. Les réseaux de neurones (RNAs) utilise en outre ces résultats d'analyse afin d'identifier les visages dans d'autres images obtenues.

Les RNAs sont couramment utilisés dans différentes applications en particulier lorsque c'est difficile d'utiliser la programmation basée sur des règles avec un

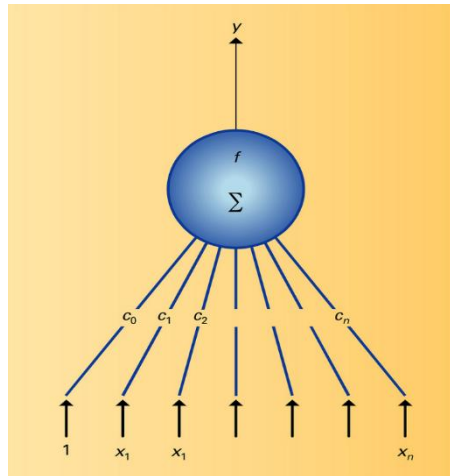
## **Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs**

algorithme traditionnel [22]. Généralement, les neurones sont organisés en couches. Il est noté que les différentes couches ont la capacité d'effectuer différents types de transformations quand il s'agit de leurs entrées. Les signaux partent de l'entrée initiale à la dernière couche appelée sortie, en générale après avoir traversé ces couches plusieurs fois.

Le but ultime des réseaux de neurones artificiels est de résoudre différents problèmes exactement de la même manière qu'un cerveau humain [21]. Le modèle le plus simple qui illustre le fonctionnement d'un réseau de neurones est le perceptron, qui est un réseau composé d'une seule couche, Notons l'existence de réseaux de neurones à multiple-couches dits "acycliques" feedforward neural network. Dans ce réseau particulier, les différentes données ne peuvent se déplacer que dans une seule direction qui est l'avant. L'information est transférée de l'entrée à travers des nœuds cachés et finalement aux nœuds de sortie. Il n'y a aucun cycle formé dans cette structure.

### **2.3.1. Définition d'un neurone**

Un neurone, est une fonction algébrique non linéaire et bornée, dont la valeur dépend de paramètres appelés coefficients ou poids. Les variables de cette fonction sont habituellement appelées « entrées » du neurone, et la valeur de la fonction est appelée « sortie ». Un neurone est donc avant tout un opérateur mathématique, dont on peut calculer la valeur numérique par quelques lignes de programme informatique. Il est très rarement réalisé physiquement sous la forme d'un objet (circuit électronique par exemple). Il est cependant pratique de le représenter graphiquement [23].



**Figure 2.1 : Représentation graphique d'un neurone [23].**

### 2.3.2. Définition des réseaux de neurones

Un neurone réalise une fonction non linéaire de ses entrées. Ces dernières peuvent être les sorties d'autres neurones : on réalise ainsi un réseau de neurones, qui n'est donc rien d'autre que la composition de fonctions non linéaires élémentaires constituées par les neurones individuels. La manière dont on « relie » entre elles les sorties des neurones aux entrées d'autres neurones définit l'architecture du réseau [23].

### 2.3.3 Fonction d'activation

Les fonctions d'activation sont une composante extrêmement importante des réseaux de neurones artificiels. Ils décident si un neurone doit être activé ou non. Si l'information que le neurone reçoit est pertinente pour l'information donnée ou si elle doit être ignorée [48].

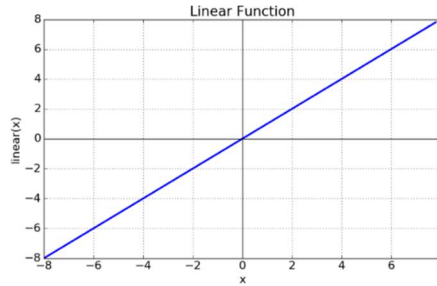
$$Y = \text{Activation} (\sum (w \times x) + b)$$

Où **W** est le poids, **X** : entrée et **B** représente le biais  
Les fonctions d'activation peuvent être divisées en 2 types à savoir :

### a. Fonction d'activation linéaire

La fonction est une ligne (Figure 2.2) ou linéaire ( $f(x) = x$ ) par conséquent, la sortie des fonctions ne sera pas confinée entre n'importe quelle range ( $-\infty$ ,  $+\infty$ ).

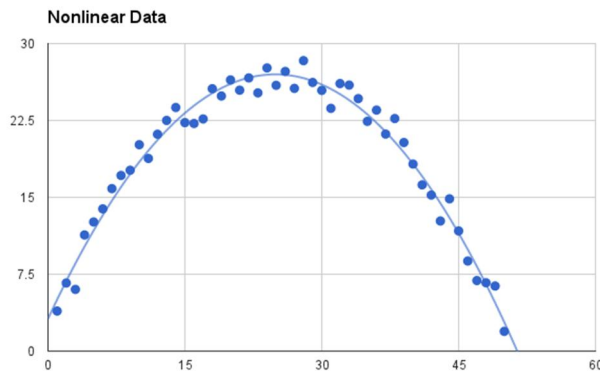
Cela n'aide pas avec la complexité ou les divers paramètres des données habituelles qui sont introduites dans les réseaux de neurones [49].



**Figure 2.2 : Fonction d'activation linéaire [49].**

### b. Fonction d'activation non linéaire

Les fonctions d'activation non linéaire sont les fonctions d'activation les plus utilisées (Figure 2.3). Il permet au modèle de généraliser ou de s'adapter à une variété de données et de différencier les résultats [10].

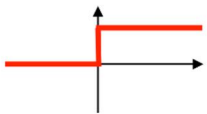
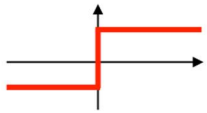
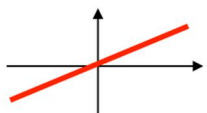
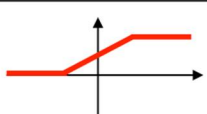
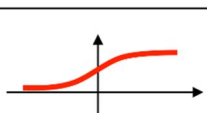

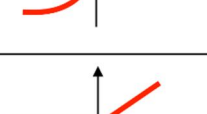



**Figure 2.3 : Fonction d'activation non linéaire [49].**



## Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs

Le tableau 3.1 regroupe les fonctions d'activation les plus connus.

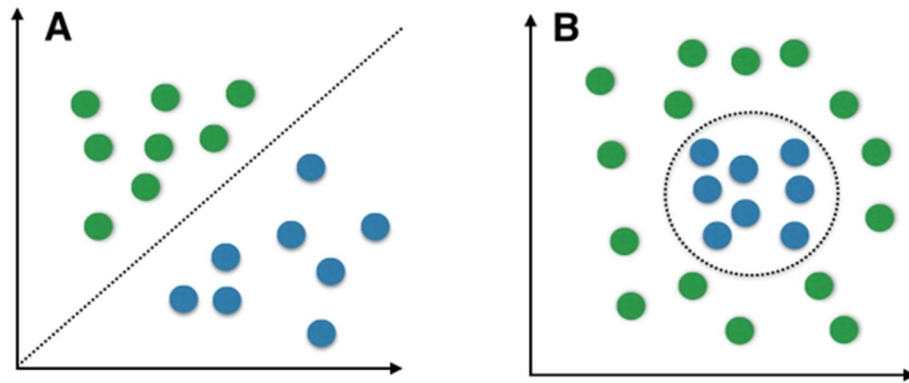
Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	Multi-layer Neural Networks	
Rectifier, ReLU (Rectified Linear Unit)	$\phi(z) = \max(0, z)$	Multi-layer Neural Networks	
Rectifier, softplus	$\phi(z) = \ln(1 + e^z)$	Multi-layer Neural Networks	

Copyright © Sebastian Raschka 2016  
(<http://sebastianraschka.com>)

**Tableau 2.1 : Fonctions d'activation [50].**

### 2.3.4 Les limitations des réseaux de neurones classiques

La convergence est l'un des plus grands problèmes du perceptron. Frank Rosenblatt a prouvé mathématiquement que la règle d'apprentissage perceptron converge si les deux classes peuvent être séparées par un hyperplan linéaire (Figure 2.4), mais des problèmes surgissent si les classes ne peuvent pas être séparées parfaitement par un classificateur linéaire [24].



**Figure 2.4 : Données linéaire vs données non linéaire [38].**

### 2.4 Apprentissage profond

Le Deep Learning ou L'apprentissage profond est un domaine relativement nouveau de méthodes plus larges d'apprentissage automatique et de recherche, qui a été introduit principalement dans le but de rapprocher la recherche en apprentissage automatique de l'intelligence artificielle. Les méthodes d'apprentissage profond consistent essentiellement à apprendre de grandes représentations de données ou, en d'autres termes, à apprendre des algorithmes de réseau neuronal [22].

Il convient de noter que l'apprentissage profond pourrait être effectué dans trois variétés : supervisé, non supervisé ainsi que l'apprentissage profond semi-supervisé [4]. De nombreuses interprétations d'apprentissage profond sont basées sur divers traitements de l'information et différents modèles de communication.

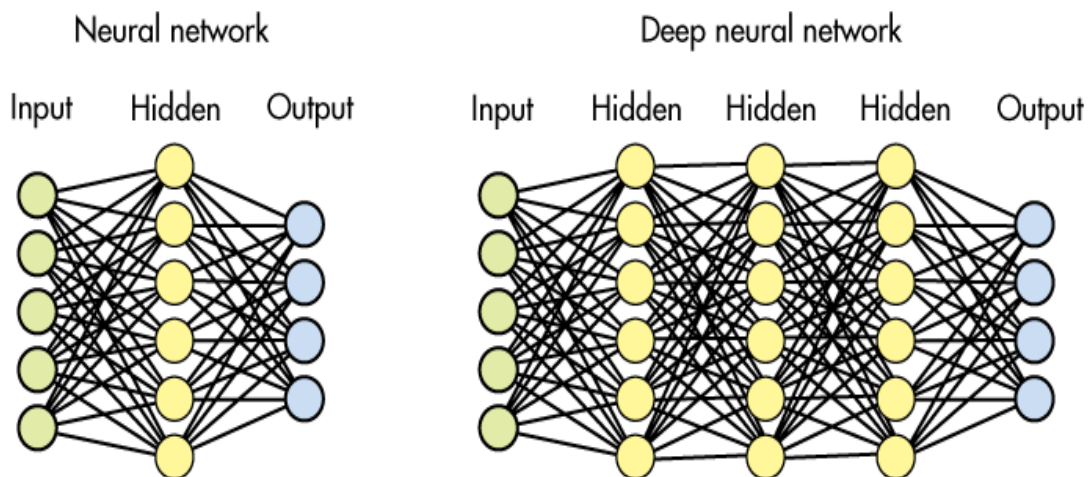
Les chercheurs en apprentissage profond se consacrent également à la création de systèmes capables d'apprendre diverses représentations d'apprentissage profond à partir de grandes collections de données ou de données volumineuses. Diverses structures d'apprentissage profond, y compris les réseaux de neurones profonds comme le « convolutional neural networks » (CNN) sont largement utilisées dans différents domaines, y compris la reconnaissance de la parole, reconnaissance audio, traduction automatique de traitement de langage neuronal, filtrage de réseaux sociaux et autres.

L'apprentissage profond fait partie d'une plus grande famille d'apprentissage automatique, et les techniques d'apprentissage profond comprennent une cascade

## Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs

contenant de nombreuses couches d'unités cachées et en arrière-plan qui sont représentées de manière non linéaire.

Pour résumer l'apprentissage profond nous pouvons simplement dire que les méthodes d'apprentissage profond utilisent différentes unités de couche et que l'apprentissage supervisé et non supervisé de diverses représentations d'entités de données est représenté dans chacune de ces couches. Leur composition dépend principalement du problème qui va être résolu [18] [22] [25] (Figure 2.5).



**Figure 2.5 : Différence entre “Neural network” et “Deep neural network” [42].**

### **2.4.1. Différence entre intelligence artificiel, l'apprentissage profond et l'apprentissage automatique**

#### **a. Intelligence Artificiel**

Créée en 1956 par John McCarthy, l'intelligence artificielle fait appel à des machines capables d'accomplir des tâches caractéristiques de l'intelligence humaine. Bien que cela soit plutôt général, cela inclut des choses comme la planification, la compréhension du langage, la reconnaissance des objets et des sons, l'apprentissage et la résolution de problèmes [43].

L'IA est classée en deux catégories, générale et étroite. L'IA générale aurait toutes les caractéristiques de l'intelligence humaine, y compris les capacités mentionnées ci-dessus. L'intelligence artificielle étroite présente certaines facettes de l'intelligence

## **Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs**

humaine, et peut très bien faire cette facette, mais elle fait défaut dans d'autres domaines. Une machine qui reconnaît très bien les images, mais rien d'autre, serait un exemple d'IA étroite [43].

### **b. Apprentissage automatique ou Machine Learning**

Au fond, l'apprentissage automatique n'est qu'un moyen d'atteindre l'IA. Arthur Samuel a inventé la phrase peu de temps après AI, en 1959, la définissant comme " la capacité d'apprendre sans être explicitement programmé ". Vous pouvez obtenir l'IA sans utiliser l'apprentissage automatique, mais cela nécessiterait la construction de millions de lignes de codes avec des règles et des arbres de décision complexes [43].

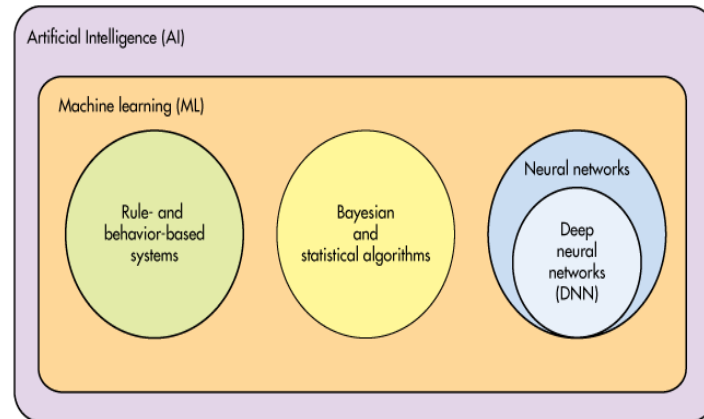
Au lieu de coder des logiciels avec des instructions spécifiques pour accomplir une tâche particulière, l'apprentissage automatique est une façon de " entraîner " un algorithme afin qu'il puisse apprendre comment. " L'entraînement " consiste à fournir d'énormes quantités de données à l'algorithme et à permettre à l'algorithme de s'ajuster et de s'améliorer [43].

### **c. Apprentissage profond ou Deep Learning**

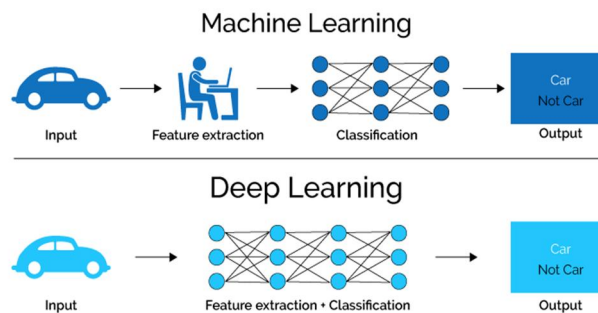
L'apprentissage profond est l'une des nombreuses approches de l'apprentissage automatique. D'autres approches comprennent l'apprentissage par arbre de décision, la programmation logique inductive, le clustering, l'apprentissage par renforcement et les réseaux bayésiens, entre autres [43]. L'apprentissage profond a été inspiré par la structure et la fonction du cerveau, à savoir l'interconnexion de nombreux neurones. Les réseaux de neurones artificiels (RNA) sont des algorithmes qui imitent la structure biologique du cerveau [43]. Dans les RNA, il y a des "neurones" qui ont des couches discrètes et des connexions à d'autres "neurones". Chaque couche choisit une caractéristique spécifique à apprendre, comme les courbes/bords dans la reconnaissance d'image. C'est cette superposition qui donne à l'apprentissage profond son nom, la profondeur est créée en utilisant plusieurs couches plutôt qu'une seule couche [43].

## Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs

La figure 2.6 illustre la différence entre l'IA l'apprentissage profond et l'apprentissage automatique alors que la figure 2.7 montre la différence entre l'apprentissage automatique et l'apprentissage profond.



**Figure 2.6 : Différence entre intelligence artificiel deep learning et machine learning [42].**



**Figure 2.7 : Différence entre le Deep Learning et le Machine Learning [44].**

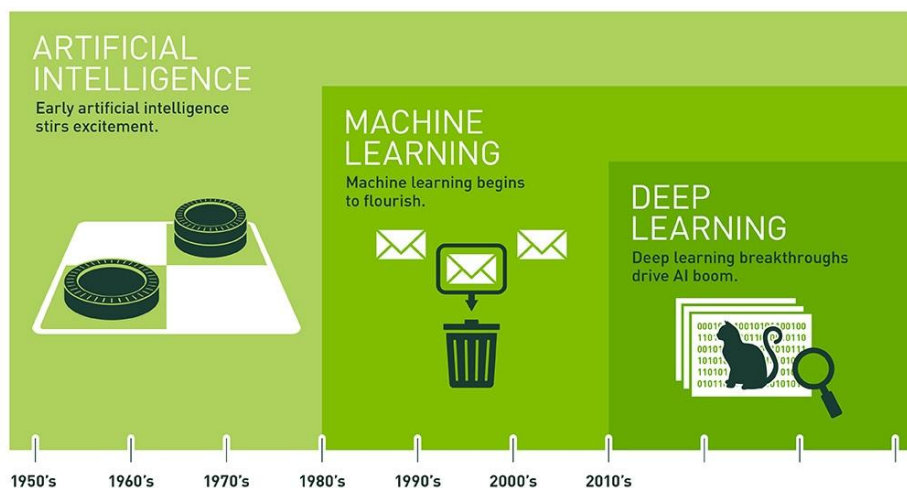
### 2.4.2. Historique

Les réseaux de neurones artificiels existent depuis longtemps. Leur application a été historiquement appelée cybernétique (années 1940-1960), connexionnisme (années 1980-1990), puis en tant qu'apprentissage profond vers 2006, lorsque les réseaux de neurones ont commencé à s'approfondir. Mais ce n'est que récemment que nous avons vraiment commencé à gratter la surface de leur plein potentiel [46]. Comme le décrit Andrej Karpathy, il y a généralement "quatre facteurs distincts qui freinent le deep learning [45] :

## Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs

1. Puissance de calcul (la plus évidente : loi de Moore, GPU).
2. Données (sous une bonne forme,c'est-à-dire pas seulement quelque part sur Internet - par exemple ImageNet).
3. Algorithmes (recherche et idées, par exemple. Backpropagation, Gradient descent).
4. Infrastructure (software,Linux, Git, AWS, TensorFlow, café, theano etc.).

Au cours de la dernière décennie environ (Figure 2.8), le plein potentiel de l'apprentissage profond est finalement débloqué par les progrès réalisés dans les facteurs (1) et (2), ce qui a conduit à d'autres percées pour les facteurs (3) et (4) [45].



**Figure 2.8 : Historique de deep learning [47].**

### 2.4.3. Domaines d'application

L'apprentissage profond est en train de transformer le monde dans pratiquement tous les domaines. Voici quelques exemples des choses incroyables que l'apprentissage en profondeur peut faire

- **Voitures autonome :** Les voitures qui s'auto-conduisent dépendent de l'apprentissage profond pour des tâches visuelles comme la compréhension des panneaux routiers, la détection des voies et la reconnaissance des obstacles [46].
- **Traduction automatique :** La traduction automatique existe depuis longtemps, mais l'apprentissage profond permet d'obtenir les meilleurs résultats dans deux

## Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs

domaines spécifiques [52] [53] [54] à savoir : (1) Traduction automatique de texte et (2) Traduction automatique des images.

- **Classification et détection d'objets** : Des résultats de haute qualité ont été obtenus sur des exemples de benchmarking en utilisant de très grands réseaux de neurones convolutionnaires. Une percée dans ce problème par Alex Krizhevsky [55] des résultats sur le problème de classification d'ImageNet appelé AlexNet [56] [57].

### **2.4.4. Réseaux de neurones profonds**

Les réseaux de neurones profonds sont principalement des paradigmes remarquables d'inspiration biologique qui permettent à n'importe quel ordinateur d'apprendre une large gamme d'informations utiles et pertinentes simplement en observant différentes données. L'apprentissage profond et les réseaux de neurones fournissent actuellement les meilleurs exemples de la qualité des données. Ils fournissent également les meilleures solutions possibles à divers problèmes dans différents domaines, y compris le traitement du langage naturel, la reconnaissance de la parole et de l'image, ect [22].

Les réseaux de neurones sont conçus pour nous aider à classer et à regrouper une grande quantité de données. En d'autres termes, pensez aux réseaux de neurones comme couche de classification et de clustering qui sont placés au-dessus des données qui seront gérées et stockées. Les réseaux de neurones sont également conçus pour aider à regrouper diverses données non étiquetées en fonction de certaines similitudes existant dans différentes collections de données [26].

On peut également dire que les réseaux de neurones extraient diverses caractéristiques de données qui sont transmises à différents algorithmes pour une classification et un regroupement plus poussés. Il faut penser au réseau de neurones en tant que divers composants de la plus grande famille de modèles et d'applications d'apprentissage automatique. Les réseaux de neurones impliquent différents algorithmes conçus pour la régression des données, la classification des données et le renforcement de l'apprentissage profond [26].

## **Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs**

Les techniques d'apprentissage en profondeur sont également capables de nous fournir une approximation de la fonction entre n'importe quelle entrée possible et n'importe quelle sortie potentielle. Nous devons juste supposer que les entrées et les sorties sont liées via une certaine corrélation [26].

Il y'a plusieurs type des réseaux de neurones profonds.

### **2.4.4.1. Réseaux de neurones convolutionnels**

Les réseaux de neurones à convolution sont utilisés dans l'apprentissage en profondeur non supervisé. Ces techniques d'apprentissage étaient, en fait, le premier modèle artificiel utilisé pour la reconnaissance afin d'atteindre une grande performance dans diverses tâches [27]. Les réseaux de neurone à convolution sont créés afin de traiter diverses données qui sont sous la forme de tableaux numériques, par exemple, une image colorée qui est composée de trois tableaux bidimensionnels qui contiennent également différentes intensités de pixel dans ces trois couches de couleur [28]. Derrière le réseau de neurone a convolution, il y a quatre idées fondamentales :

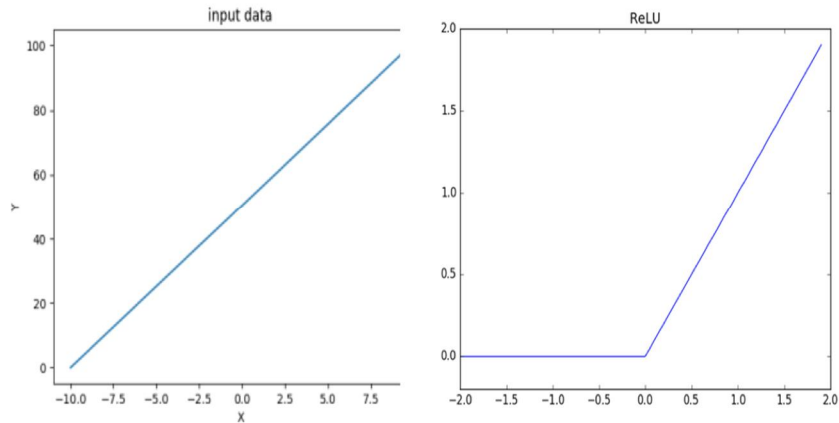
**A. Étape de convolution :** Le but principal de la convolution est d'extraire des caractéristiques de la donnée d'entrée. Par exemple une image en couleur qui contient 3 couches devra passer pas un traitement à convolution qui consiste à faire glisser un filtre ou bien le détecteur de caractéristiques (Feature detector) sur l'image d'origine pixel par pixel. On calcule la multiplication par éléments (entre les deux matrices) et on ajoute les sorties de multiplication pour obtenir l'entier final qui forme un seul élément de la matrice de sortie qui est appeler 'Convolved Feature' ou 'Feature Map' [28].

**B. Non linéarité (Relu) :** Une opération supplémentaire appelée ReLU est utilisée après chaque opération de convolution, ReLU signifie unité linéaire rectifiée et est une opération non linéaire.sa formule de sortie est :  $\text{Sortie} = \text{Max}(\text{Zéro}, \text{Entré})$ .

ReLU est une opération appliquer par pixel et remplace toutes les valeurs de pixels négatifs dans la Feature map par zéro (Figure 2.9). Le but de ReLu est d'introduire la notion de non-linéarité[28].

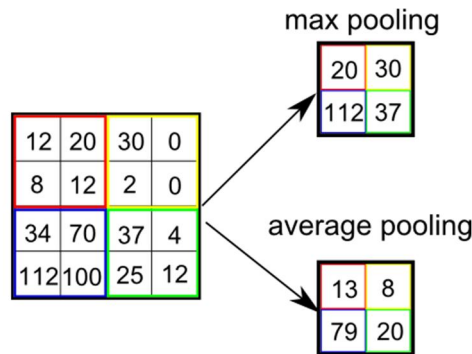


## Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs



**Figure 2.9 : Relu Non-Linéarité[28].**

**C. Couche de Pooling** : L'étape de Pooling réduit la dimensionnalité de chaque « Feature Map » et conserve les informations les plus importantes. Les deux formes les plus utilisées sont le 'Average Pooling' qui consiste à faire la moyenne des valeurs de la zone, ou bien 'Max Pooling' qui prend l'élément le plus grand de la « Feature Map » (Figure 2.10). Le Pooling rend les représentations d'entrée plus petites et plus faciles à gérer. Il réduit aussi le nombre de paramètres et de calculs dans le réseau, contrôlant ainsi le sur-apprentissage [28].



**Figure 2.4 : Étape Max/average Pooling [27].**

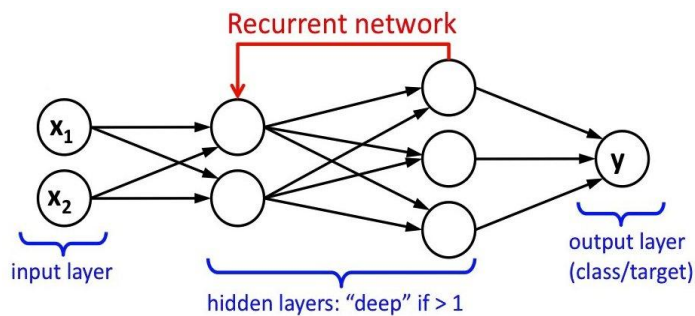
**D. Couche entièrement connecté ou Fully Connected** : La couche Fully Connected est configurée exactement comme son nom l'indique : elle est entièrement connectée à la sortie de la couche précédente. Une couche entièrement

## Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs

connectée prend tous les neurones de la couche précédente (qu'elle soit entièrement connectée, en pool ou convolutionnelle) et la connecte à chaque neurone qu'elle possède [28].

### 2.4.4.2. Réseaux de neurones récurrents (RNN)

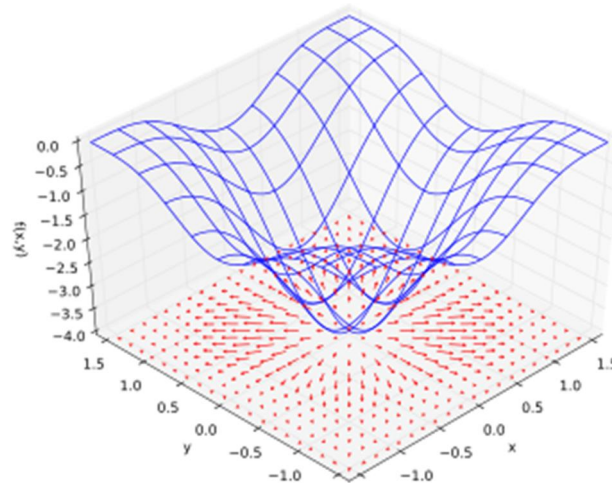
L'idée général derrière les réseaux de neurones récurrent est d'utiliser des informations séquentielles (Figure 2.11). Dans un réseau neuronal traditionnel, nous supposons que toutes les entrées (et sorties) sont indépendantes les unes des autres. Mais si on veut prédire le mot suivant dans une phrase, on doit savoir quels mots sont antérieurs. Les RNN sont appelés récurrents car ils effectuent la même tâche pour chaque élément d'une séquence, avec la sortie étant dépendante des calculs précédents [29]. Voici à quoi ressemble un RNN typique :



**Figure 2.11 : Architecture d'un réseau de neurones récurrents [41].**

## 2.5. Concepts de l'apprentissage profond

- **Le gradient** : Le gradient est la dérivée partielle d'une fonction qui prend plusieurs vecteurs et fournit une valeur unique (les fonctions de coût dans les réseaux de neurones) (Figure 2.12). Si nous augmentons les variables d'entrée (l'entrée et les poids de réseau) le gradient nous indiquera dans quelle direction aller sur le graphe pour augmenter notre sortie. Puisque nous souhaitons diminuer notre fonction de coût nous utilisons le gradient et allons dans la direction opposée [81].



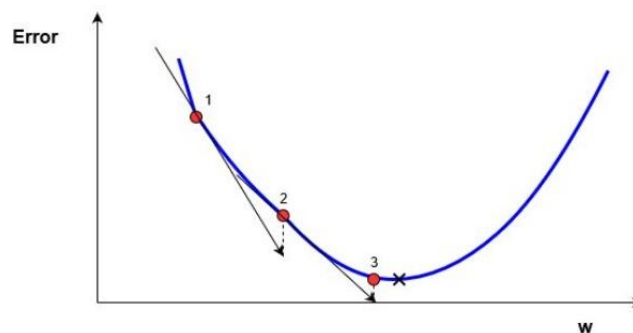
**Figure 2.12 : Visualisation du gradient. Les flèches rouges représentent le gradient de la fonction tracée en bleu [81].**

- **Backpropagation (Rétropropagation) :** La rétropropagation est une méthode efficace pour calculer les gradients dans les graphes de computations, comme les réseaux de neurones. Ce n'est pas une méthode d'apprentissage, mais plutôt une forme de calcul qui est souvent utilisée dans les méthodes d'apprentissage. Il s'agit en fait d'une simple implémentation de la règle de la chaîne des dérivés, celle-ci donne simplement la possibilité de calculer toutes les dérivées partielles requises dans un temps linéaire en termes de taille du Graphe, alors que les approches naïves le calculer de gradient se développeraient exponentiellement par rapport à la profondeur du réseau[83].
- **Vanishing Gradient (Disparition des gradients) :** Lorsque nous faisons la rétropropagation, c'est-à-dire en reculant dans le réseau et en calculant les gradients de la fonction de coût (erreur) par rapport aux poids, les gradients tendent à devenir de plus en plus petits à mesure que nous continuons à reculer dans le réseau. Cela signifie que les neurones des couches antérieures apprennent très lentement par rapport aux neurones des couches ultérieures de la Hiérarchie. Les couches antérieures du réseau sont les plus lentes à s'entraîner. C'est donc ce que le problème de la disparition de

## Chapitre2 Apprentissage basé sur les réseaux de neurones convolutifs

gradient fait à notre modèle. Le processus d'entraînement prend trop de temps et la précision de prédiction du modèle diminuera [84].

- **Gradient Descent** : C'est un algorithme d'optimisation qui vise à minimiser certaines fonctions de coût en fonction du gradient de cette fonction. Des itérations successives sont employées pour s'approcher progressivement d'un minimum local ou global de la fonction de coût. La figure 2.13 ci-dessous montre un exemple du « **Gradient descent** » fonctionnant dans une seule dimension [82].

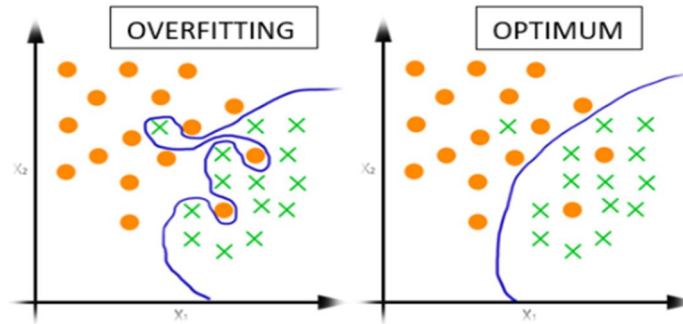


**Figure 2.13 : Visualisation du 'gradient descent' en fonctionnement [82].**

- **Sur-apprentissage(Overfitting)**

Le mot overfitting fait référence à un modèle qui apprend par cœur les données d'entraînement. Au lieu d'apprendre la distribution générale des données, le modèle apprend la sortie attendue pour chaque point de données et c'est pour cette raison que le modèle ne pourra pas prédire par rapport à des données généralisées

La difficulté réside dans le fait qu'à première vue, il peut sembler que le modèle fonctionne correctement parce qu'il comporte une très petite erreur sur les données d'entraînement. Cependant, dès qu'on lui demande de prédire de nouveaux points de données, il va échouer [85]. La figure 2.13 illustre un exemple de sur-apprentissage.



**Figure 2.14 : Exemple sur-apprentissage[86].**

## **2.6. Conclusion**

Dans ce chapitre nous avons introduit l'apprentissage et exposé la différence entre les réseaux de neurones classiques et l'apprentissage profond. Dans le prochain chapitre, nous passerons en revue les différentes étapes et méthodes de détection automatique de visage en passant en commençant par les méthodes standard de la reconnaissance faciale suivis par les méthodes récentes basées sur la technologies réseau de neurones convolutifs (CNN).

## **Chapitre 3**

### **Etat de l'art sur les systèmes de détection et de la reconnaissance faciale.**

#### **3.1 Introduction**

Ce chapitre vise à passer en revue les différentes étapes et méthodes de détection automatique de visage en commençant par les méthodes standard de la reconnaissance faciale qui procèdent par l'extraction des caractéristiques locales des visages puis une classification pour une éventuelle reconnaissance et authentification des personnes, suivis par les méthodes récentes basées sur la technologies réseau de neurones convolutifs (CNN)

#### **3.2 Les méthodes Standard de la détection**

La détection des visages pose le problème de la localisation des visages présents dans une image d'entrée. Idéalement, la détection fournit aussi leurs dimensions pour un éventuel traitement ultérieur.

##### **3.2.1 Méthode de Viola and Jones**

Une avancée majeure dans le domaine a été réalisée par Viola et Jones en 2001 [56]. Ces derniers ont proposé une méthode basée sur l'apparence ("Appearance-based methods") robuste à 15 fps pour des images de 384 x 288 pixels sur un PC Intel Pentium III 700Mhz. Ce fut la première méthode présentée en temps réel. Cette approche est faite sur trois concepts [57]:

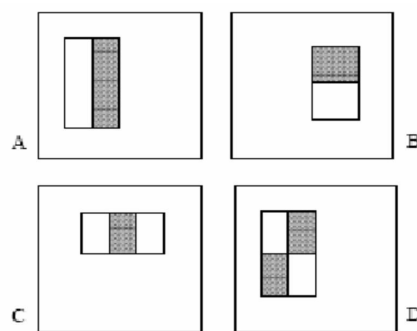
- **L'image intégrale**

L'algorithme se base sur les caractéristiques de Haar (Haar features) pour localiser les visages présents sur une image d'entrée. Dans le but d'extraire rapidement ces

caractéristiques, l'image est représentée sous forme intégrale. En effet, sous cette forme, l'extraction d'une caractéristique à n'importe quel endroit et à n'importe quelle échelle est exécutée en un temps constant tandis que le temps de conversion vers la représentation intégrale ne remet pas en cause ce gain de temps offert par l'utilisation de la représentation en image intégrale.

La définition des caractéristiques de Haar et la manière dont la représentation intégrale accélère considérablement leur extraction sont présentées ci-après :

Pour une image en niveaux de gris. Dans toute image, une zone rectangulaire peut être délimitée et la somme des valeurs de ses pixels calculée. Une caractéristique de Haar est une simple combinaison linéaire de sommes ainsi obtenues. Plusieurs caractéristiques de Haar peuvent être définies selon le nombre, les échelles, les positions et les dimensions des zones rectangulaires considérées [57]. Un exemple des caractéristiques de Haar est illustré sur la figure 3.1.



**Figure 3.1 : Exemple de quatre caractéristiques de Haar [58].**

Selon l'exemple de la figure 3.1, la somme des valeurs des pixels appartenant aux zones encadrées claires est soustraite à la somme des valeurs des pixels appartenant aux zones encadrées sombres pour obtenir la caractéristique de Haar. Chacune des quatre caractéristiques de Haar est représentée avec son cadre de détection respectif [57].

Le calcul d'une caractéristique de Haar demande à chaque fois l'accès aux valeurs de tous les pixels contenus dans les zones rectangulaires considérées. Cela devient vite contraignant temporellement dès que les caractéristiques de Haar sont définies par des zones rectangulaires de grandes dimensions. L'image intégrale permet de surmonter ce problème en rendant constant le temps de calcul d'une caractéristique

de Haar à n'importe quelle échelle [57].. L'image intégrale est représentée mathématiquement par:

$$u(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (3.1)[57]$$

$$\forall 0 < x \leq width, 0 < y \leq height \quad (3.2)[57]$$

où  $i(x; y)$  est l'image d'origine et  $i(x'; y')$  représente l'image sous sa nouvelle représentation Ainsi chaque pixel est la somme des valeurs des pixels compris dans le rectangle défini par le coin supérieur gauche de l'image et lui-même[57].

Le calcul de la somme des valeurs des pixels appartenant à une zone rectangulaire s'effectue en accédant seulement à quatre pixel de l'image intégrale : Soit un rectangle ABCD dont les sommets sont nommés dans le sens des aiguilles d'une montre en commençant par le sommet supérieur gauche et soit  $x$  la valeur sous la représentation intégrale d'un sommet  $X$  du rectangle ( $X \in \{A; B; C; D\}$ ). La somme des valeurs des pixels appartenant à ABCD est, quelle que soit sa taille, donnée par  $(c - b - d + a)$ . Une caractéristique de Haar étant une combinaison linéaire de tels rectangles ABCD, son calcul se fait alors en un temps indépendant de sa taille[57].

- **Algorithme d'apprentissage base sur Adaboost**

Pour localiser les visages sur l'image d'entrée, cette dernière est scannée par une fenêtre de dimension déterminée. La fenêtre parcourt l'image, par la suite son contenu est analysé pour savoir s'il s'agit d'un visage ou non. Comme il a été mentionné auparavant, les caractéristiques de Haar sont extraites pour effectuer la classification et de ce fait la représentation intégrale de l'image accélère l'analyse. Mais, pour une fenêtre de 24x24 pixels il y a 45 396 caractéristiques de Haar. Traiter toutes ces caractéristiques prendrait beaucoup de temps pour une application en temps réel. Pour surmonter ce problème, une variante de la méthode de boosting Adaboost est utilisée.

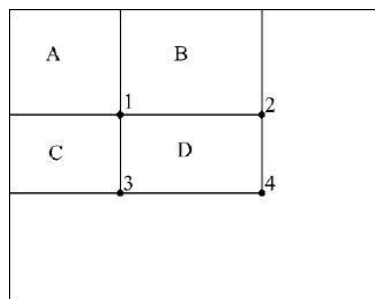


Ci-dessous Adaboost est brièvement présent suivi de sa variante qui constitue le deuxième apport du travail de Viola & Jones [60].

Adaboost est une méthode d'apprentissage permettant de "booster" les performances d'un classifieur quelconque nommé "classifieur faible". L'idée est de faire passer les candidats à classifier à travers plusieurs classifieurs faibles, chacun étant entraîné en portant plus d'attention sur les candidats mal classifiés par le classifieur précédent. Pour arriver à ce résultat des poids  $W_i$  sont associés aux échantillons de l'ensemble *d'entraînement*  $((x_i; y_i) \ i = 1; \dots; m)$ , tout d'abord de manière équilibrée :

$$W_i^0 = \frac{1}{m} \quad (3.3)[57]$$

Adaboost sert à booster un classifieur déjà existant et a priori chaque classifieur faible possède le même espace d'entrée. Dans la variante d'Adaboost de Viola et Jones, Les classifieurs faibles ont pour entrée une caractéristique de Haar différente. Adaboost s'apparente alors à une sélection de caractéristiques (feature selection). Cette variante d'Adaboost est utilisée lors de l'apprentissage pour sélectionner les caractéristiques de Haar et pour détecter un visage et permet de surmonter le problème du nombre élevé de caractéristiques de Haar existant pour une fenêtre de recherche. La figure 3.2 illustre un exemple de l'algorithme d'Adaboost.

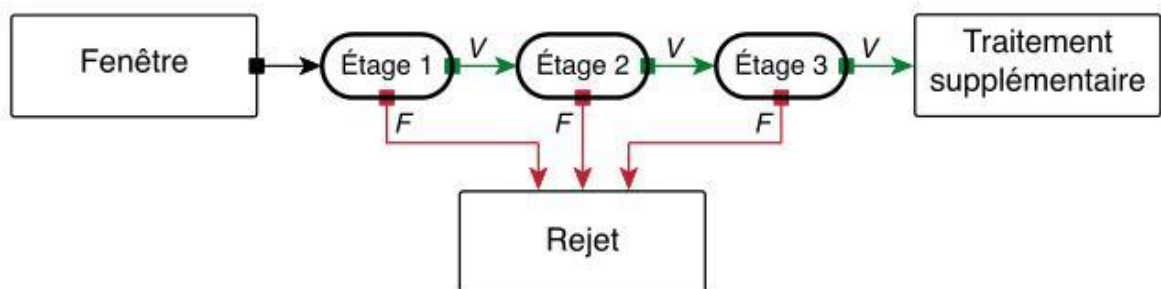


**Figure 3.2 : Exemple de l'algorithme Adaboost [59].**

L'emplacement 1 dans l'exemple de la figure 3.2 est la somme des pixels dans le rectangle A, 2 est la somme de A et B ( $A + B$ ), 3 est la somme de A et C ( $A + C$ ) et rectangle D correspond à :  $D = (4 + 1) - (2 + 3)$ [59].

- **Cascade**

L'idée de base derrière le concept de cascade est que parmi l'ensemble des candidats, c'est à dire l'ensemble des états de la fenêtre de recherche, une partie peut être éliminée sur base de l'évaluation de seulement quelques caractéristiques de Haar. Une fois cette élimination effectuée, les candidats restants sont analysés par des classifieurs forts plus complexes (utilisant plus de caractéristiques de Haar) demandant un plus grand temps de traitement. En utilisant plusieurs 'étages' de ce type, le processeur évite d'effectuer des analyses lourdes en temps de calcul sur des échantillons pour lesquels il est rapidement possible de se rendre compte qu'ils sont négatifs. Le processus de classification apparait alors comme une cascade de classifieurs forts de plus en plus complexes ou à chaque étage les échantillons classifiés négatifs sont sortis tandis que les échantillons classifiés positifs sont envoyés aux classifieurs suivants. Ceci est représenté sur la figure 3.3.[57]



**Figure 3.3 : Cascade de classifieurs forts. A chaque étage, uniquement les candidats classifiés positifs sont transmis à l'étage suivant. [59].**

Si le premier étage rejette un faux négatif, c'est un gros problème car il ne sera jamais recouper par la cascade. Autrement dit c'est un visage qui ne sera pas détecté. Par contre, si le premier étage transmet un faux positif, il pourra toujours être éliminé aux étages suivants de la cascade. Ce petit raisonnement permet de mettre en évidence que les premiers nœuds constitutifs de la cascade peuvent se permettre d'avoir un taux de faux positifs élevés (de l'ordre de 40-50%) mais doivent absolument assurer un taux de détection maximum. La para-métrisation de l'algorithme en ce sens sera détaillée dans la prochaine section [57].

Ce concept permet donc à l'algorithme de consacrer son temps à de longues analyses complexes uniquement lorsque cela en vaut la peine. Il s'agit à nouveau d'un mécanisme qui accélère la vitesse d'exécution de la méthode proposée par Viola & Jones [57].

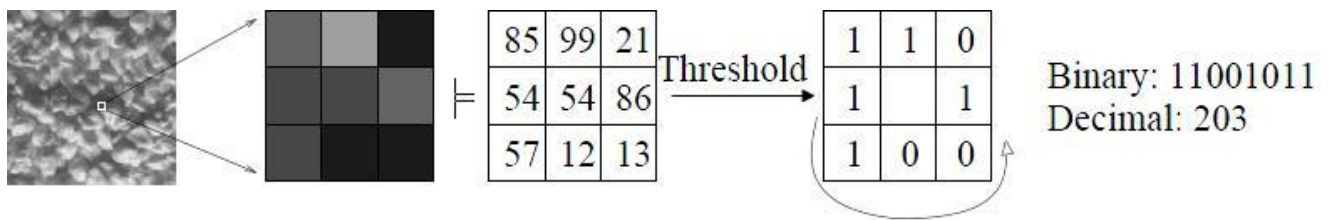
- **Dernières avancées de l'algorithme Viola & Jones.**

Lors de l'entraînement de la cascade, un nouvel ensemble de non-visages est généré par l'algorithme de Adaboost à la fin de la construction de chaque étage [46]. Les auteurs Xiaohua et al ont remarqué qu'au fur et à mesure que la cascade grandissait, les non-visages collectés sont de plus en plus similaires aux visages et que le taux d'erreur des nœuds les plus profonds sur un ensemble de validation composé uniquement d'échantillons négatifs, Tendait vers 0.5. Cela signifie qu'il devient difficile de trouver des caractéristiques encore capables de distinguer les non-visages des visages. Ils observent aussi que le nombre de caractéristiques nécessaires pour atteindre les objectifs locaux en termes de taux de détection et de faux positifs augmente selon une pente de plus en plus raide pour les nœuds profonds. Ceci confirme la difficulté de trouver des bonnes caractéristiques. Notons que les échantillons positifs utilisés sont des carrés couvrant la zone yeux-nez-bouche. Xiaohua et al proposent qu'au delà d'une certaine profondeur soient considérées les parties entourant la zone yeux-nez-bouche en se basant sur le fonctionnement du système visuel humain [63]. Ils avancent que les informations de ces zones peuvent être utilisées lorsque la zone centrale est épuisée. Ainsi une hiérarchie est mise en place où 3 dimensions standards d'échantillons positifs sont utilisées successivement pour entraîner la cascade. Les expériences menées par l'équipe ont montré des améliorations en termes de vitesse et de taux de détection [57].

### **3.2.2 Méthode basé sur LBP (Local Binary Patterns)**

La méthode LBP a été créée par Ahonen et al. L'opérateur LBP basique prend comme entrée un carré de 9 pixels et a pour sortie un nombre binaire 8 bits. Ce dernier est illustré sur la figure 3.4 . La motivation qui a poussé à utiliser cet opérateur est qu'un visage peut être vu comme un assemblage de micro-modèle

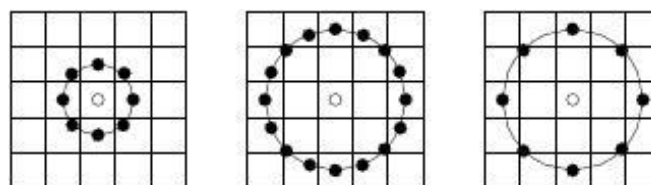
dont la description par LBP est à la fois bonne, robuste face aux variations de gris et rapide à générer.



**Figure 3.4 : Fonctionnement de l'opérateur LBP basique [65].**

En se basant sur l'exemple illustré sur la figure 3.4, l'entrée est une matrice carré de 9 pixels. Un seuil est appliqué tel que tous les pixels périphériques dont la valeur est supérieure à la valeur du pixel central se voient attribuer la valeur 1 tandis que les autres reçoivent la valeur 0. La valeur LBP obtenue est le nombre binaire se lisant dans le sens des aiguilles d'une montre autour du pixel central. La conversion de toute une image se fait en déplaçant une fenêtre de conversion de 3x3 pixels sur l'image entière. Pour obtenir une image convertie de même dimension que l'image d'entrée, les pixels du bord de l'image sont répétés [65]

L'opérateur simple (LBP) a été étendu pour rester fiable à différentes échelles. Ainsi, P points décrivent le nombre binaire et ceux-ci sont distribués le long d'un cercle de rayon R. Ceci est noté (P; R). Comme les P points ne tombent pas nécessairement au centre d'un pixel de l'image, leurs valeurs sont obtenues par interpolation bilinéaire. Ceci est illustré à la figure 3.5.



**Figure 3.5 : Extension de l'opérateur lbp basique. Les entourages (8,1), (16,2), (8,2) sont représentés [66].**

Pour inclure une information quant à la disposition spatiale des textures et éviter de se limiter à une description holistique des textures qui souffrirait des limitations connues des méthodes de ce type, l'image convertie est divisée en plusieurs sous-régions pour lesquelles autant d'histogrammes LBP seront faits. Ainsi pour 4 sous-

régions, 4 histogrammes seront générés. Ces derniers seront concaténés pour former une matrice a 2 dimensions appelée 'histogramme spatialement amélioré. Notons que les sous-régions peuvent se recouvrir et ne doivent pas nécessairement être rectangulaires[57].

La comparaison de deux histogrammes spatialement améliorés suppose d'une part d'établir une méthode de mesure de distance entre deux histogrammes simples et d'autre part l'utilisation de poids pour rassembler les distances obtenues pour chaque sous-région.

Notons maintenant l'ensemble des éléments para-métrisables tout d'abord le choix de l'entourage (P; R) à utiliser pour la conversion LBP, ensuite la manière dont les sous-régions sont délimitées, puis le poids accordé à chacune d'elles et en enfin la méthode de calcul de distance qui sera utilisée pour comparer deux histogrammes. Ces paramétrages bien effectués amélioreront les performances du classifieur final.

- **Dernières avancées de la méthode LBP**

Le travail de **Tan et Triggs [7]** a fait suite à celui de Ahonen et al. Et. Il propose 3 nouveaux concepts qui permettent d'améliorer significativement les performances (plus de 40% sur la database FRGC-104 selon leur publication). Ces trois concepts sont les processus «Local Ternary Patterns » (LTP), une méthode de prétraitement de l'image et enfin une méthode de mesure de distance pour la comparaison d'échantillons au format LBP ou LTP. Ces concepts sont détaillés ci-dessous. Il s'agit de la généralisation des LBP au système ternaire. Elle a été proposée par Tan et Triggs comme solution au problème de sensibilité qu'éprouve, Le LBP face au bruit aléatoire et celui de quantification. Le principe est le suivant :

La méthode LBP appliquait un seuil égal à la valeur du pixel central, quand à la conversion LTP attribue à la valeur 0 aux pixels dont la valeur se trouve dans un voisinage de la valeur du pixel central, 1 à ceux dont la valeur est au delà de ce voisinage et -1 à ceux dont la valeur est en dessous. La formulation mathématique de la conversion LTP est donnée par la comme suit :

$$s(u, i_c, t) = \begin{cases} 1 & \text{si } u \geq i_c + t \\ 0 & \text{si } |u - i_c| < t \\ -1 & \text{si } u \leq i_c - t \end{cases} \quad (3.4)[68]$$

Où

$u$  un pixel de périphérie,

$i_c$  la valeur du pixel central ,et

$t$  le voisinage

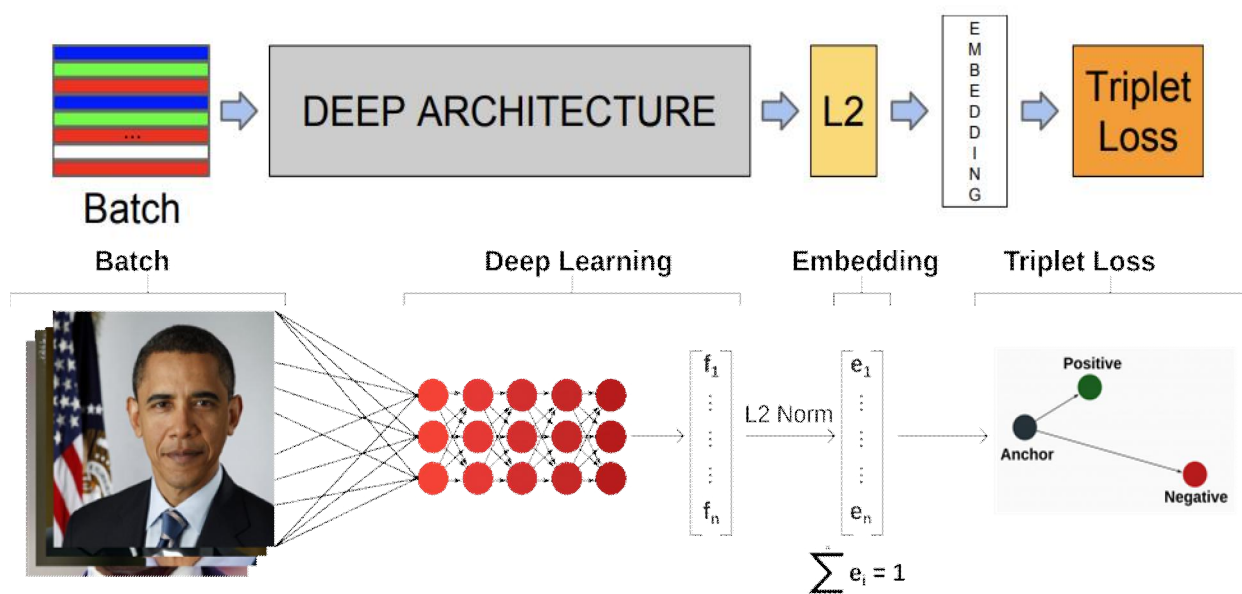
### 3.3 Les méthodes basées sur CNN les plus connus

Plusieurs architectures de reconnaissance faciale basée sur les réseaux convolutifs existent, les plus utilisées sont :

#### 3.3.1. FaceNet

FaceNet [69] est parmi les premières architectures de la reconnaissance faciale mises en œuvre. C'est est un réseau neuronal de reconnaissance, de vérification et de clustering novateur. Il possède un réseau neuronal profond de 22 couches qui entraine directement sa sortie dans un en capsulage à 128 dimensions (embedding). La fonction de perte utilisée à la dernière couche est appelée perte de triplets.[78]

L'architecture FaceNet regroupe trois schémas blocs (Figure 3.6):



**Figure 3.7 : Illustrations de FaceNet [70],[71]**

- **La fonction de perte Triplets**

La perte de triplets est une fonction où une entrée de base (ancree) est comparée à une entrée positive (vérité) et à une entrée négative (faux). Cette distance est minimisée dans le cas de la comparaison entre image ancre et image vérité et elle est maximisée dans la comparaison entre image ancre et image fausse. La fonction de perte peut être décrite à l'aide d'une fonction de distance euclidienne:

$$L(A, P, N) = \max(\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha, 0) \quad (3.5)[72]$$

où

**A** est une entrée d'ancrage,

**P** est une entrée positive de la même classe que A,

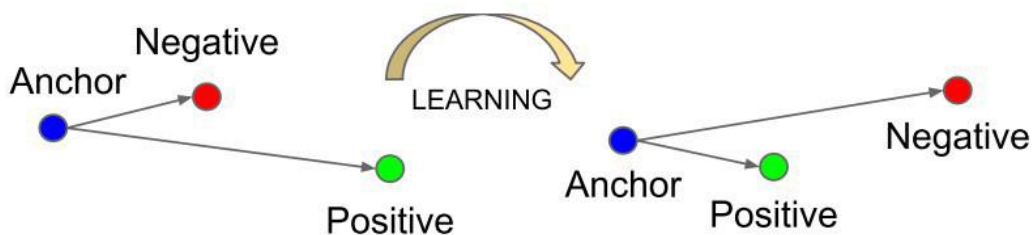
**N** est une entrée négative d'une classe différente de A,

$\alpha$  est une marge entre les paires positives et négatives, et

**f** est une intégration.

La figure (3.8 ) montre le schéma bloc de la fonction de perte de triplets et la figure (3.9) illustre des exemples d'images ancre, vérité et négative respectivement.

- Ancree : Image de Elon Mask, avec laquelle nous voulons comparer,
- Positif : Une autre image d'Elon Mask, exemple positif,
- Négatif : Image de John Travolta, exemple négatif.



**Figure 3.8 : La fonction de perte de triplets [73].**

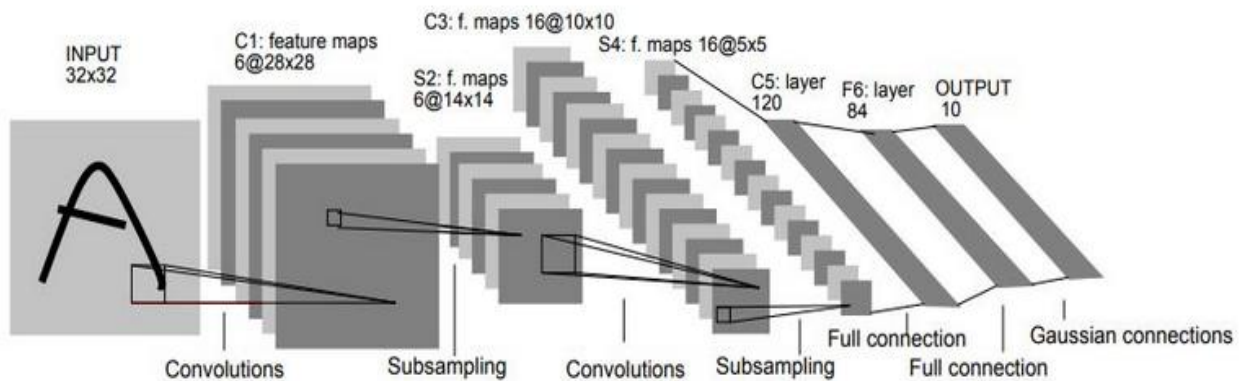


**Figure 3. 9 : Illustre des exemples d'images ancre, vérité et négative [74].**

En théorie, l'image de l'ancre devrait être plus proche de l'image vérité et éloignée de l'image négative dans l'espace euclidien.

### 3.3.2. LeNet-5

L'architecture LeNet [75] est l'un des premiers réseaux de neurones convolutifs, développé par LeCun et al. en 1998 [76]. Il a été utilisé pour la classification et la reconnaissance des caractères manuscrits dans les banques et les postes. La figure 3.10 illustre l'architecture LeNet, elle comprend des couches de convolution, des sous-échantillonnages et des couches entièrement connectées à la sortie.



**Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.3.10 : Architecture du réseau LeNet [77].**



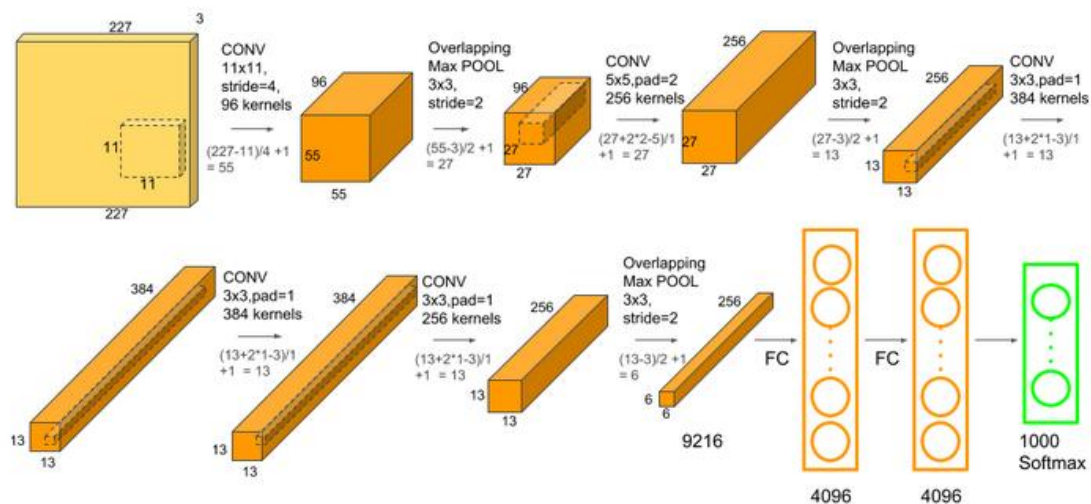
### **Description des Couche du réseau *LeNet***

- **La première couche** : Cette première couche est appliquée sur une image en niveau de gris de taille 32x32, avec six filtres de convolution de taille 5x5 et un pas (stride) égal à 1. La dimension de l'image passe de 32x32x1 à 28x28x6.
- **La deuxième couche** : Le LeNet-5 applique une couche de sous-échantillonnage avec une taille de filtre de 2x2 et un pas de deux. Les dimensions de l'image résultante seront réduites à 14x14x6.
- **La troisième couche** : C'est une deuxième couche de convolution avec 16 cartes de caractéristiques (filtres) de taille 5x5 et un pas de 1, les dimensions de l'image passent à 10x10x16.
- **La quatrième couche** : La quatrième couche est à nouveau une couche de sous échantillonnage avec une taille de filtre de 2 x 2 et un pas de 2. Cette couche est identique à la deuxième couche, la différence est qu'elle comporte 16 cartes de caractéristiques de sorte que la sortie soit réduite à 5x5x16.
- **La cinquième couche** : La cinquième couche (C5) est une couche convolutionnelle entièrement connectée avec 120 cartes de caractéristiques de taille 1x1. Chacune des 120 unités est connectée à tous les 400 nœuds (5x5x16) de la quatrième couche.
- **La sixième couche** : La sixième couche est une couche entièrement connectée avec 84 unités.
- **La couche de sortie**: Enfin, il existe une couche de sortie softmax entièrement connectée avec 10 valeurs possibles correspondant aux chiffres de 0 à 9 [78].

### **3.3.3. L'architecture AlexNet**

L'entrée est une image en couleur RGB de taille 256x256 pixels. Différentes couches de convolutions sont utilisées pour extraire les caractéristiques de l'image : la première couche de convolution contient 96 filtres de taille 11x11x3 et un pas de deux et la seconde contient 256 filtres de taille 5x5 et un pas égal à deux. Chaque couche des deux premières couches de convolution est suivie par une couche de

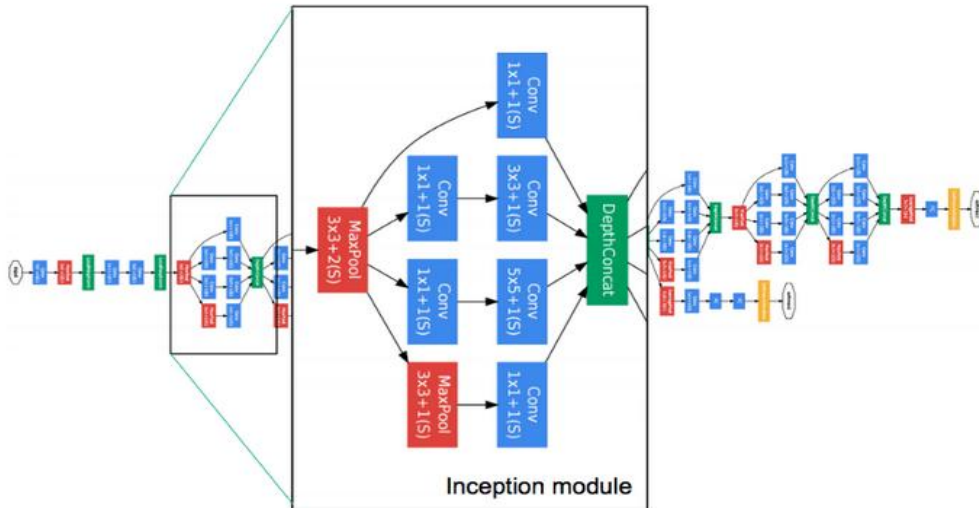
sous échantillonnage similaire qui utilise des fenêtres de taille 3x3 avec un pas de deux. Cependant, les trois couches de convolution suivantes (3<sup>ème</sup>, 4<sup>ème</sup> et 5<sup>ème</sup> couches) sont connectées directement suivies par une couche de sous-échantillonnage. La sortie contient deux couches entièrement connectées et une couche softmax avec 1000 éléments pour la classification. La figure 3.11 montre l'architecture du réseau AlexNet [81].



**Figure** Erreur ! Il n'y a pas de texte répondant à ce style dans ce document. **3.11 :**  
**Architecture du réseau AlexNet.[79]**

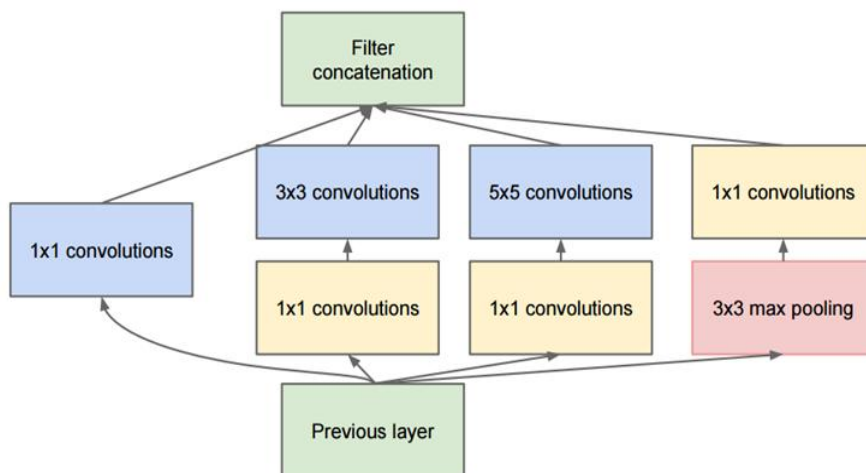
### 3.3.4. GoogleNet

GoogleNet [82] appelé aussi Inception Module, il contient 22 couches, ces couches ne sont pas toutes séquentielles comme dans les architectures précédentes. Il comporte des couches qui se traitent en parallèle. Son architecture est montrée sur la figure 3.12.



**Figure 3.12 : Architecture du réseau Google net[80].**

La partie encadré s'appelle « inception module », la figure 3.13 l'énonce avec plus précision. Le tableau 3.1 suivant représente l'architecture Inception 'Google Net'.



**Figure 3.13 : Architecture d'un module 'Inception'. [36]**

<i>Type</i>	<i>taille/ pas</i>	<i>taille de l'output</i>
Convolution	7x7/2	112x112x64
Max pool	3x3/2	56x56x64
Convolution	3x3/1	56x56x192
Max pool	3x3/2	28x28x192
Inception(3a)		28x28x256
Inception(3b)		28x28x480
Max pool		14x14x480
Inception(4a)		14x14x512
Inception(4b)		14x14x512
Inception(4c)		14x14x512
Inception(4d)		14x14x528
Inception(4e)		14x14x832
Max pool	3x3/2	7x7x832
Inception(5a)		7x7x832
Inception(5b)		7x7x1024
Avg pool	7x7/1	1x1x1024
Dropout		1x1x1024
Linear		1x1x1000
Softmax		1x1x1000

**Tableau 3.1 : Tableau résumant les différentes couches de Inception 'GoogleNet [84].**

### **3.3.5. L'architecture VGG16**

VGG16 est un modèle de réseau de neurones convolutif proposé par Simonyan et al [85]. de l'Université d'Oxford dans l'article intitulé «Réseaux convolutionnels très profonds pour la reconnaissance d'images à grande échelle». Le modèle atteint une précision de 92,7% dans le top 5 des tests dans ImageNet, qui est un jeu de données de plus de 14 millions d'images appartenant à 1000 classes. C'était l'un des

fameux modèles soumis à la conférence ILSVRC-2014. Ses résultats sont beaucoup plus élevés par rapport à ceux obtenus par l'architecture AlexNet.

### **3.4 Comparaison**

Le tableau 3.2 c'est une étude de l'architecture VGG16 avec autres architectures qui montre que malgré le nombre énorme de paramètres il est capable de détecter moins d'erreur comparons les autres architectures.

year	CNN	Devlopped by	Top 5 error rate	N of paramètres
1998	LeNet(8)	Yan LeCun et al		60 milles
2012	AlexNet(7)	Alex Krizhevzky, Geoffrey Hinton, Ilya Stutkever	15.03%	60 millions
2013	ZFNet()	Matthew Zeili, Rob Fergus	14.08%	
2014	GoogleNet(19)	Google	8.01%	4 millions
2014	VGGNet(16)	Simonyan, Zisserman	7.3%	132 millions

***Tableau 3.2 : Tableau de comparaison entre VGG16 et autres architectures***

Dans notre application, nous avons exploité le modèle VGG-16 pour la reconnaissance de visage. La structure et les différentes couches de ce modèle seront détaillées dans le chapitre 4.

### **3.5 Conclusion**

Dans ce chapitre, nous avons présenté quelques méthodes standard de détection et de reconnaissance faciale ainsi que les modèles les plus connus des CNN afin d'avoir un aperçu et comprendre comment fonctionne les CNN. Dans le chapitre suivant, nous aborderons en détails l'architecture VGG-16 que nous avons exploité pour développer notre application de détection et de classification de visage en temps réel.

## **Chapitre 4**

### **Mise en œuvre du système de reconnaissance et l'authentification faciale**

#### **4.1 Introduction**

Ce chapitre vise à expliquer plus en détail le modèle MTCNN utilisé dans notre application pour la détection et l'alignement des visages ainsi que l'architecture VGG-16 prise comme modèle pour la reconnaissance faciale. L'application de reconnaissance et d'authentification élaborée comprend quatre modules principaux :

- Création de la base de données :
- Phase de détection, alignement et d'extraction de caractéristique. Pour cette étape de travail, nous avons utilisé le modèle basé sur les CNN qui est MTCNN (Multi-task Cascade Convolutional Neural Network)
- Phase d'apprentissage : Cette phase de travail n'a pas été effectuée, nous avons exploité un modèle VGG-16 prédéfini de Keras.
- Phase de comparaison et de décision qui est la phase de reconnaissance et d'authentification. Elle consiste à reconnaître le visage de la personne et le comparer à la base de données établit en moyennant la métrique de similarité de cosinus.

## **4.2 Description de l'application**

Comme tout système d'authentification et de reconnaissance de visage, l'application à réaliser comprendra quatre modules essentiels à savoir :

1. Création de la base de données des employés CDTA en effectuant l'acquisition de leur visage à l'aide d'une webcam.
2. Amélioration de la qualité et l'alignement des visages
3. Applications du modèle VGG 16 pour l'extraction des descripteurs, classification et reconnaissance à partir d'images faciales 2D,
4. Recherche de similarité dans la base de données du **personnelle du CDTA** et prise de décision du visage en question.
5. Affichage les diverses informations concernant l'individu en question.

Le schéma bloc de l'application est illustré sur la figure 4.1.

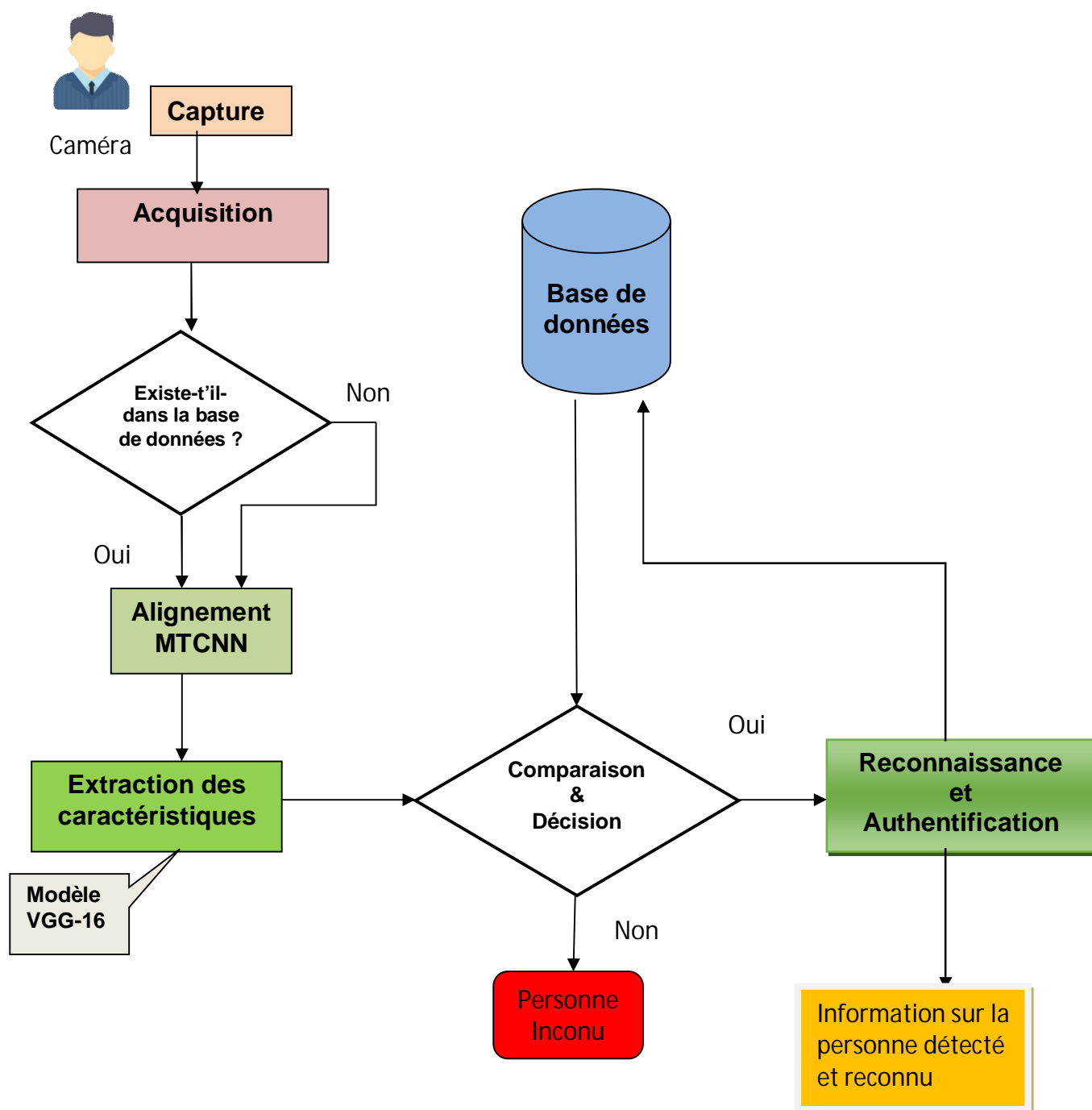
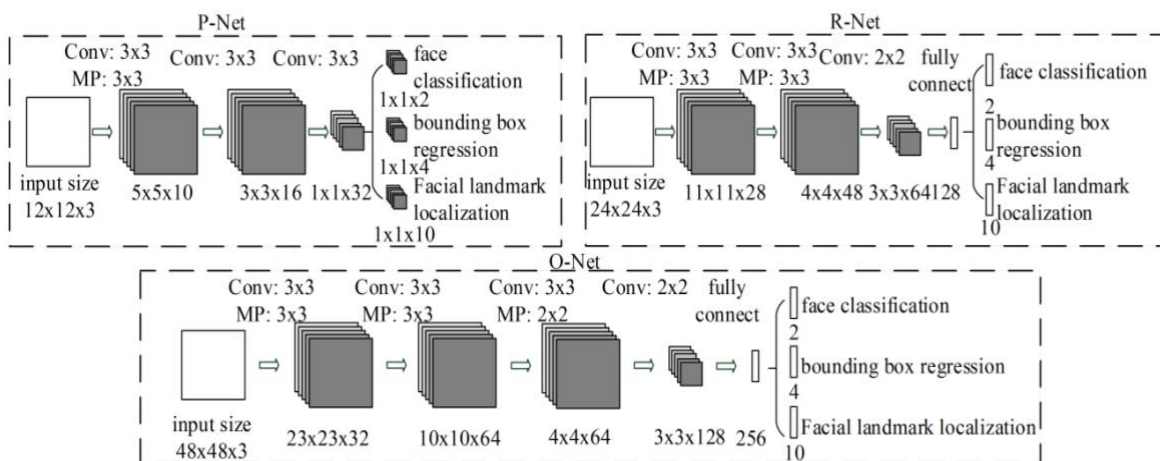


Figure 4.1 : Diagramme blocs du système développé.



**4.2.1 Phase de détection basée sur MTCNN**

Le modèle MTCNN est proposé par Kaipeng *Zhang et al.* [87]. Son principe consiste à détecter des boîtes englobantes de visage ("faces bounding box") dans une image et cinq points de repère qu'on appelle "Landmarks". Il comprend trois étapes de traitements (Figure 4.2) pour effectuer simultanément la détection des visages et celle des repères faciaux. Dans le premier bloc de traitement, plusieurs fenêtres candidates via un réseau CNN peu profond est proposé, puis un deuxième réseau CNN plus complexe qui consiste à affiner les fenêtres pour rejeter un grand nombre de fenêtres ne contenant pas de visage. Dans le troisième bloc de traitement, un CNN plus puissant est utilisé pour affiner le résultat et afficher les positions des repères faciaux.

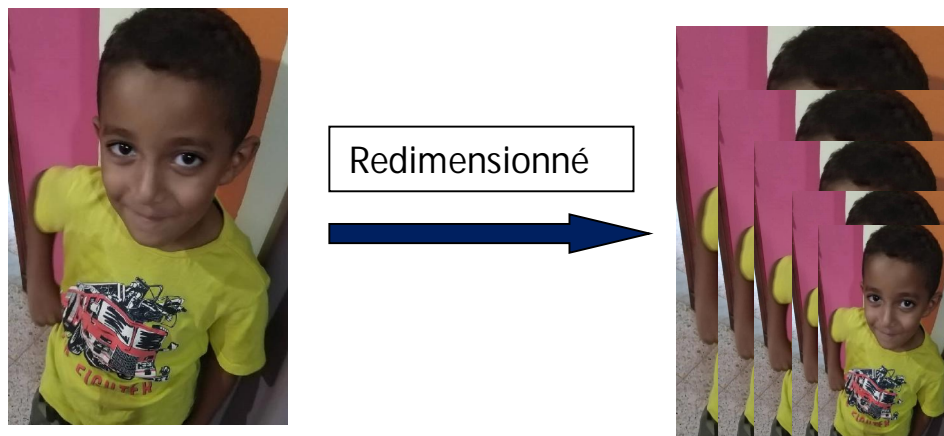


**Figure 4.2 : L'architecture du modèle MTCNN [88].**

**a. Le fonctionnement du bloc P-Net**

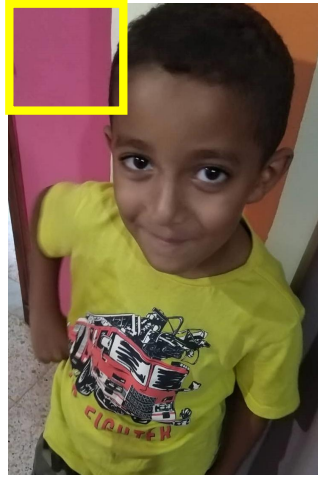
Dans la première étape, un réseau entièrement convolutionnel appelé réseau de propositions (P-Net) est utilisé pour obtenir les régions proposées et leurs vecteurs de régression par boîte de délimitation. Les vecteurs de régression obtenus sont utilisés pour calibrer les régions proposées, puis pour appliquer une suppression non maximale (NMS) afin de fusionner des régions fortement chevauchées. Le principe

de cette étape de traitement consiste à créer une pyramide d'images à partir d'une image en entrée afin de détecter les visages de différentes tailles (Figure 4.3). En d'autres termes, différentes copies de différentes tailles d'une même image sont créées. Ceci afin de chercher des visages de différentes tailles dans l'image. Pour chaque copie du visage mise à l'échelle, un filtre (noyau) de Kernel de taille 12x12 parcourt toute l'image de manière incrémentale en recherchant les visages [89].



**Figure 4.3 : Image pyramide**

Le traitement commence par une région de l'image comprise entre (0,0) et (12,12) située dans le coin supérieur gauche (Figure 4.4). Celle-ci est transmise au bloc P-Net, qui renvoie les coordonnées d'un cadre de sélection s'il détecte un visage. Ensuite, il répète ce processus avec les régions  $(0 + 2a, 0 + 2b)$  à  $(12 + 2a, 12 + 2b)$ , en décalant les pixels du noyau 12x12 avec un pas de 2 pixels vers la droite ou vers le bas. Le décalage de 2 pixels est appelé pas (stride ou pas), ou le nombre de pixels que le noyau déplace à chaque fois. Dans ce cas, le temps de traitement est réduit. Le seul inconvénient de ce processus est de recalculer tous les index liés au pas de déplacement. La méthode MTCNN permet de faire la détection et l'alignement des visages.[89]

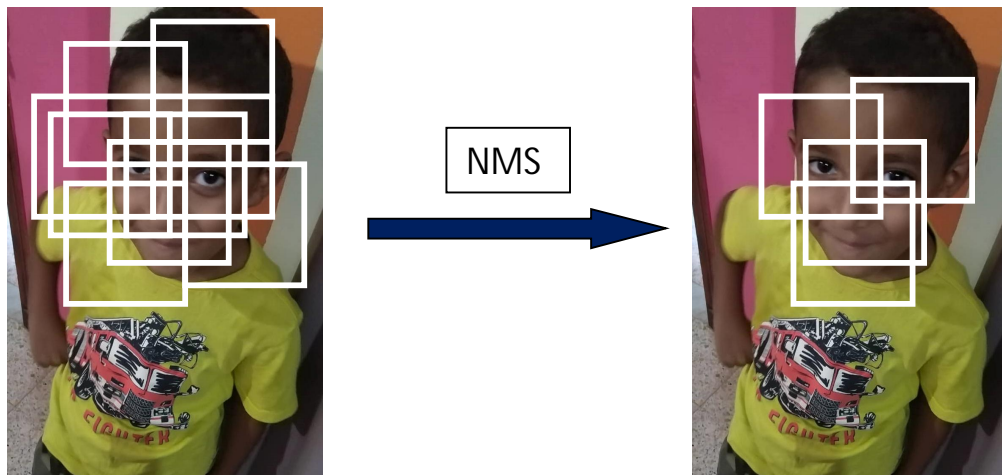


**Figure 4.4 : Le noyau et la fenêtre glissante.**

Chaque noyau serait plus petit par rapport à une image de grande taille en entrée, de sorte qu'il serait capable de trouver des visages plus petits. De la même manière, le noyau serait plus gros par rapport à une image de taille plus petite, ce qui lui permettrait de trouver des visages plus grands dans l'image de plus petite taille. Plusieurs copies de l'image de différentes tailles sont créées et transmises au premier réseau de neurones, P-Net, et rassemblées en sortie du bloc de traitement. Les poids et les biais de P-Net ont été formés de manière à produire un cadre de sélection relativement précis pour chaque noyau de 12 x 12 [89].

Cependant, le réseau est plus confiant sur certaines boîtes que sur d'autres. Par conséquent, il est nécessaire d'analyser la sortie P-Net pour obtenir une liste de niveaux de confiance pour chaque cadre de sélection et supprimer les cadres avec un niveau de confiance inférieur (c'est-à-dire que les cases dont le réseau n'est pas tout à fait sûr de contenir un visage). Après avoir sélectionné les cases avec une confiance accrue, les coordonnées des cadres de sélection sont normalisées en les convertissant en ceux de l'image réelle non mise à l'échelle. Comme la plupart des noyaux sont dans une image réduite, leurs coordonnées seront basées sur la plus petite image. Cependant, il reste encore beaucoup de cadres de sélection, qui se chevauchent souvent. Le traitement basé sur la méthode suppression non maximale (NMS) est utilisé pour réduire le nombre de cadres de sélection. Le processus de NMS est effectué en triant d'abord les boîtes englobantes (et leurs noyaux respectifs 12 x 12) en fonction de leur confiance ou de leur score. Dans d'autres modèles, le système NMS utilise la plus grande boîte englobante au lieu de celle sur laquelle le réseau a le plus confiance [89].

Par la suite, la surface de chacun des noyaux, ainsi que la zone de chevauchement entre chaque noyau et le noyau ayant le score le plus élevée sont calculées. Les noyaux qui se chevauchent beaucoup avec le noyau très performant sont supprimés. Enfin, le traitement NMS renvoie une liste des boites englobantes survivantes. Le processus NMS est effectué une fois pour chaque image mise à l'échelle, puis une fois de plus avec tous les noyaux survivants de chaque échelle. Cela supprime les cadres de sélection redondants, ce qui permet de limiter la recherche à une boîte précise par visage [89].



**Figure 4.5 : Le réseau analyse la sortie P-Net**

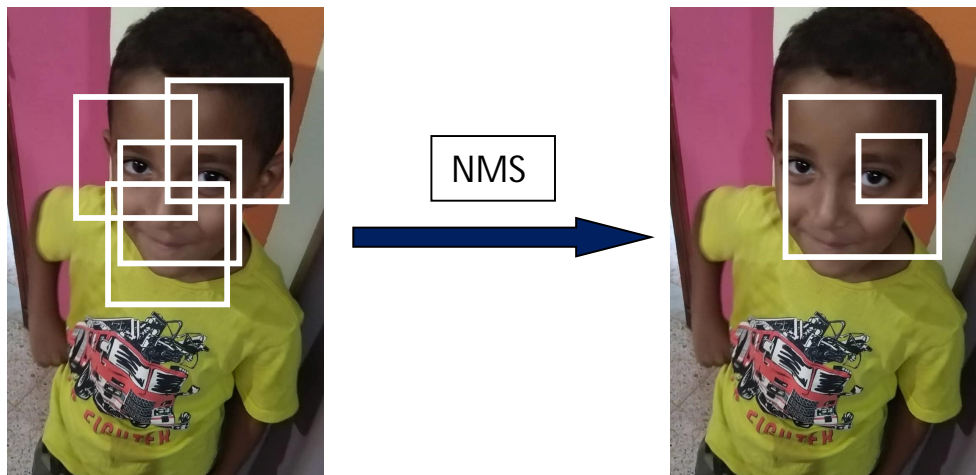
### **b. Fonctionnement du bloc R-Net**

A la sortie du bloc P-Net, tous les candidats sont acheminés vers un autre CNN, appelé réseau raffiné (R-Net), qui rejette en outre un grand nombre de faux candidats et effectue un étalonnage avec une régression du cadre de sélection et une NMS [89].

Parfois, une image peut contenir uniquement une partie du visage qui s'observe du côté du cadre. Dans ce cas, le réseau peut renvoyer une zone de sélection partielle en dehors du cadre. Pour chaque cadre de sélection, un tableau de la même taille est créé et les valeurs de pixel (l'image dans le cadre de sélection) sont copiées dans le nouveau tableau. Si le cadre de sélection est en dehors des bords, uniquement la partie de l'image du cadre de sélection est copiée dans le nouveau tableau et le

reste du tableau est complété par des zéros. Une fois que les tableaux des cadres de sélection sont remplis, ils sont redimensionnés à 24 x 24 pixels et normalisés à des valeurs comprises entre -1 et 1 [89].

Une fois que tous tableaux d'images sont de taille 24 x 24 (autant que le nombre de boites englobantes ayant survécu à l'étape 1, puisque chacune de ces boites englobantes a été redimensionnée et normalisée dans ces noyaux) ; ils sont rassemblés et transmis au réseau R-Net. La sortie de R-Net est similaire à celle de P-Net, elle inclut les coordonnées plus précises des nouveaux cadres de sélection, ainsi que le niveau de confiance de chacun de ces derniers. Une fois encore, les boites avec moins de confiance sont éliminées. Après normalisation des coordonnées, les boites englobantes sont transformées en un carré à transmettre au réseau O-Net.[89]

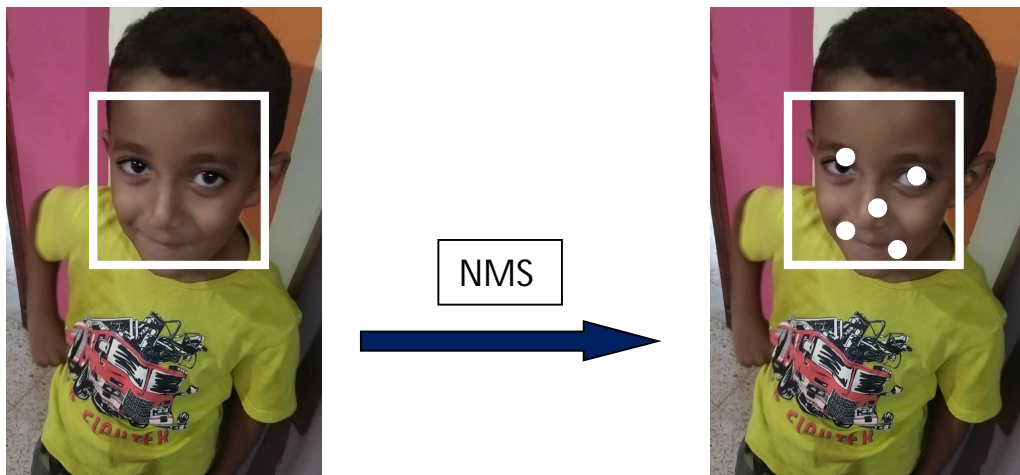


**Figure 4.6 : Le réseau rejette un grand nombre de faux candidats.**

### **c. Fonctionnement du bloc de traitement O-Net**

Cette étape est semblable à la deuxième étape, mais dans celle-ci, les régions de visage avec les cinq positions de points de repère faciaux sont identifiées (les deux yeux, le nez et les extrémités de la bouche). Avant de traiter les boites englobantes de R-Net, les boites qui sont hors limites sont remplies de zéros et sont redimensionnées à la taille 48 x 48 pixels, puis transmis au réseau O-Net. Les résultats du bloc O-Net diffèrent légèrement de ceux de P-Net et de R-Net. Le traitement effectué par le bloc O-Net fournit trois sorties : les coordonnées du cadre

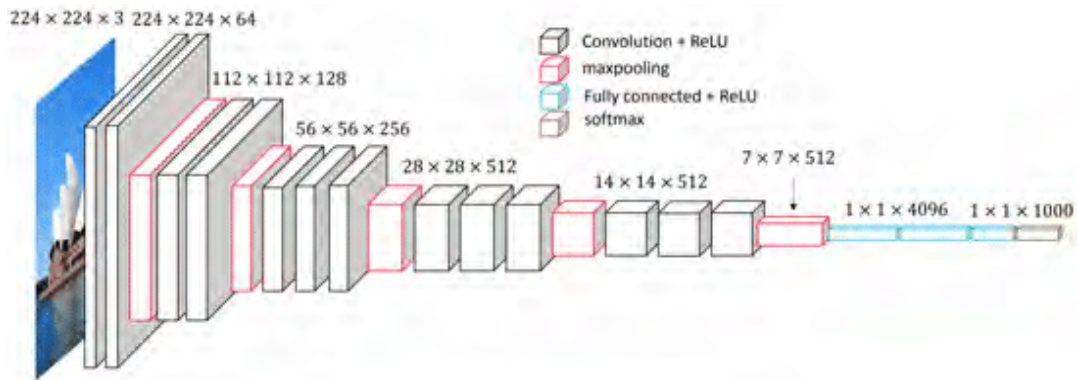
de sélection (out [0]), les coordonnées des cinq points de repère faciaux (out [1]) et le niveau de confiance de chaque cadre (out [2]). Une fois encore, les boîtes dont le niveau de confiance est faible sont éliminées et les coordonnées des boîtes englobantes et celles du repère facial sont standardisées. Enfin, le traitement NMS est appliqué. A ce stade, il ne doit exister qu'un seul cadre de sélection pour chaque visage de l'image. La toute dernière étape consiste à regrouper toutes les informations dans un dictionnaire comportant trois clés à savoir : (1) box, (2) confiance et (3) points-clés. `Box` contient les coordonnées du cadre de sélection, `confiance` contient le niveau de confiance du réseau pour chaque case, et `points-clés` inclut les coordonnées de chaque repère facial (yeux, nez et extrémités de la bouche)[89].



**Figure 4.7 : le cadre de sélection, les 5 points de repère faciaux.**

### **4.2.2 La phase d'apprentissage**

Pour la phase d'apprentissage, le modèle VGG-16 prédéfini est utilisé. Le Modèle VGG-16 permet de classer les images en 1000 catégories d'objets entre autres le visage humain. Ce réseau a une taille d'image en entrée de 224x224x3. La valeur 3 spécifie le nombre de canaux de couleur (channels en anglais) qui indique que le traitement s'effectue sur des images en RVB. L'architecture de VGG-16 est illustrée par le schéma de la figure 4.6 ci-dessous :



**Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.4.8 :**  
**Architecture du réseau VGG16 [36].**

Chaque couche de convolution utilise des filtres de taille 3\*3, déplacés avec un pas de 1 pixel. Le zero-padding vaut 1 pixel afin que les volumes en entrée aient les mêmes dimensions en sortie. Le nombre de filtres varie selon le "bloc" dans lequel la couche se trouve. De plus, un paramètre de biais est introduit dans le produit de convolution pour chaque filtre. La forme donnant la taille de l'image à l'entrée de chaque couche du réseau profond est donnée comme suit [92] :

$$M = \frac{W - K + 2P}{S} + 1 \quad (4.1)[92]$$

où :

**W** est la taille de l'image en entrée (224x224x3), **K** représente la taille du filtre (3\*3), **P** est la valeur du padding qui égale à 1 ,et **S** désigne le pas de convolution.

Chaque couche de convolution a pour fonction d'activation une ReLU. Autrement dit, il y a toujours une couche de correction ReLU après une couche de convolution. L'opération de pooling est réalisée avec des cellules de taille 2x2 pixel et un pas de 2 pixels. Les deux premières couches entièrement connectées (fully-connected) calculées sont des vecteurs de taille 4096, et sont chacune suivies d'une couche ReLU. La dernière couche entièrement connectée renvoie le vecteur de probabilités de taille 1000 (le nombre de classes) en appliquant la fonction softmax. De plus, ces

## Chapitre 4 Mise en œuvre du système de reconnaissance et d'authentification faciale

trois couches utilisent un paramètre de biais pour chaque élément du vecteur en sortie.

La fonction softmax est appliquée dans la dernière étape du modèle, elle réalise la fonction de classification sur la base des traits faciaux extraits(caractéristiques faciaux) de CNN.

Le tableau 4.1 détaille le type des couches, la taille des filtres et le pas utilisés dans l'architecture VGG16.

Type de couche	Taille du filtre	Nombre de filtre	pas
Convolution	3x3	64	1
Convolution	3x3	64	1
Max pooling	2x2	-	2
Convolution	3x3	128	1
Convolution	3x3	128	1
Max pooling	2x2	-	2
Convolution	3x3	256	1
Convolution	3x3	256	1
Convolution	3x3	256	1
Max pooling	2x2	-	2
Convolution	3x3	512	1
Convolution	3x3	512	1
Convolution	3x3	512	1
Max pooling	2x2	-	2
Convolution	3x3	512	1
Convolution	3x3	512	1
Convolution	3x3	512	1
Max pooling	2x2	-	2
Full-connection	4096	-	-
Full-connection	4096	-	-

**Tableau 4.1 : Architecture VGG16[93].**



### 4.2.3 La phase de comparaison de reconnaissance et d'authentification

Il s'agit de chercher le visage similaire à partir d'un ensemble de visages par une première extraction de caractéristiques à l'aide du réseau de neurones de convolution VGG-16 préalablement formé, puis par similarité de cosinus pour chaque paire de visage. Pour cela les visages d'entrée et le visage détecté sont présentés sous forme de vecteurs. La décision que deux visages sont identiques ou non est effectuée en comparant ces représentations vectorielles en moyennant la distance cosinus. Celle-ci est donnée par la formule suivante :

$$\text{similarité} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n (A_i B_i)}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (4.2)[94]$$

$$\text{Distance} = 1 - [\text{similarité}] \quad (4.3)[95]$$

## 4.3 Conclusion

Dans ce chapitre, nous avons abordé en détails le modèle VGG FACE 16 incluant le modèle MTCNN adopté dans notre application de détection, alignement et de reconnaissance de visage. Dans le chapitre suivant, nous présenterons les résultats des tests effectués.

# Chapitre 5

## Expérimentation et résultats

### 5.1 Introduction

Nous abordons dans ce chapitre les détails techniques de l'implémentation réalisée en termes de langage de programmation python et tensorflow. Nous présentons l'interface de l'application développée tout en donnant des exemples du travail accompli à travers des captures d'écran.

### 5.2 Environnement du travail

Dans cette section, nous présentons l'environnement matériel utilisé pour réaliser l'application. Les caractéristiques sont les suivantes :

1. Un ordinateur **DELL** :
  - Processeur : Intel(R) Core (TM) i5-5200U CPU @ 2.20 Ghz.
  - Mémoire installée (RAM) : 8.00 Go.
  - Disque Dur : 1 To.
  - système d'exploitation 64 bits.
  - OS : Microsoft Windows 7.

### 5.3 Technologies utilisées

#### 5.3.1 Langage Python

**Python** est un langage de programmation interprété populaire multi-paradigme et multiplateformes, offrant des outils de haut niveau et une syntaxe simple à utiliser. Il a été créé par Guido van Rossum et sorti en 1991. La version majeure la plus récente de Python que nous avons utilisé est Python3 (Figure5.1)



*Figure 5.1 : Logo de Python*

### 5.3. 2 Logiciel PyCharm

**PyCharm**, c'est un logiciel multi-plateforme qui fonctionne sous Windows, Mac OS X et Linux, développé par l'entreprise tchèque JetBrains. Il permet l'analyse de code et contient un débogueur graphique. Il permet également la gestion des tests unitaires, l'intégration de logiciel de gestion de versions, et supporte le développement web avec Django. Trois logiciels sont apparues. La figure 5.2 montre le logo de PyCharm.

- **PyCharm CE** (la version gratuite Community Edition, diffusé sous licence Apache. : c'est cette version que nous allons utiliser)
- **PyCharm Pro** (la version professionnelle du logiciel, diffusé sous licence propriétaire)
- **PyCharm Edu** (la version éducative du logiciel)



*Figure 5.2 : Logo de PyCharm*

### 5.3.3 SQLite

**SQL** est un langage standard pour stocker, manipuler et récupérer des données dans des bases de données. Créé au début des années 2000 par D. Richard Hipp De plus est une bibliothèque qui implémente un moteur de base de données SQL transactionnel autonome, sans configuration, sans serveur.

SQLite est lié à plusieurs langages de programmation tels que C, C ++, BASIC, C #, Python, Java et Delphi. Il est également disponible dans les systèmes d'exploitation embarqués tels que iOS, Android, Symbian OS, Maemo, Blackberry et WebOS en raison de sa petite taille et de sa facilité d'utilisation.

### 5.3.4 Logiciel DB Browser (SQLite)

**DB Browser pour SQLite** est un outil visuel et open source de haute qualité pour créer, concevoir et éditer des fichiers de base de données compatibles avec SQLite. Il utilise une interface similaire à une feuille de calcul et son utilisation ne nécessite la connaissance des commandes SQL compliquées. La figure 5.3 montre le logo de Python.



**Figure 5.3 : Logo de DB Browser**

## 5.4 Outils utilisés

### 5.4.1 Tkinter (Tool kit interface):

**Tkinter** est la bibliothèque graphique libre d'origine pour le langage Python, elle permet la création d'interfaces graphiques. Elle vient d'une adaptation de la bibliothèque graphique Tk écrite pour Tcl. Il est Installé par: `pip install tkinter`. La figure 5.4 montre le logo de la bibliothèque Tkinter



**Figure 5.4 : Logo de Tkinter**

### 5.4.2 OpenCV (Open Source Computer Vision):

**OpenCV** est une bibliothèque graphique libre, initialement développée par Intel, spécialisée dans le traitement d'images en temps réel. Cette bibliothèque est distribuée sous licence BSD (licence libre utilisée pour la distribution de logiciels). Dans notre travail on utilisons la version 2 (cv2) [103]. Il est Installé par : `pip install opencv-python`. La figure 5.5 montre le logo de Open CV.



**Figure 5.5 : Logo de OpenCV.**

### 5.4.3 haarcascade\_frontalface\_default.xml

Une cascade de Haar est fondamentalement un classificateur utilisé pour détecter des objets particuliers à partir de la source. L'haarcascade\_frontalface\_default.xml est une cascade de haar conçue par OpenCV pour détecter la face frontale. Cette cascade de haar est disponible sur github. Une cascade de Haar fonctionne en formant la cascade sur des milliers d'images négatives avec l'image positive superposée. La cascade de haar est capable de détecter des caractéristiques de la source.

### 5.4.4 OS

Ce module fournit une manière portable d'utiliser les fonctionnalités dépendantes du système d'exploitation (interagir avec votre système d'exploitation ; créer, supprimer, déplacer des dossiers et parfois de modifier le répertoire de travail). Il est Installé par: *pip install os-sys*

### 5.4.5 NumPy

**NumPy** est une extension du langage de programmation Python, bibliothèque logicielle libre et open source, destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux. Installer par : *pip install numpy*. La figure 5.6 montre le logo de NumPy.



**Figure 5.6 : Logo de NumPy**

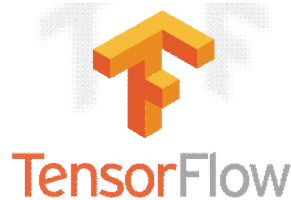
### 5.4.6 TensorFlow

**TensorFlow** est une bibliothèque de logiciels open source pour le calcul numérique haute performance.

Son architecture flexible permet de déployer facilement les calculs sur diverses plates-formes (processeurs, GPU, TPU).

Développé par Google Brain au sein de l'organisation d'intelligence artificielle de Google, il intègre un support puissant en matière d'apprentissage automatique et d'apprentissage approfondi. Il est installé par: *pip install tensorflow*

La Figure 5.7 montre le logo de TensorFlow



**Figure 5.7 - : Logo de TensorFlow**

### 5.4.7 Keras

**Keras** est une API de réseaux de neurones de haut niveau, a été créé pour être convivial, modulaire, facile à étendre et pour travailler avec Python et capable de fonctionner au dessus de TensorFlow . Il a été conçu pour les êtres humains et non pour les machines et suit les meilleures pratiques pour réduire la charge cognitive. Il est installé par : *pip install Keras*. La Figure 5.8 montre le logo de Keras.



**Figure 5.8 - : Logo de Keras**

## 5.5 Présentation des interfaces de l'application

Ce paragraphe est consacré à la présentation des interfaces de l'application réalisée. L'interface d'accueil contient les différents boutons, tel que (Figure 5.9):

- Bouton pour la détection de visage basé sur la technique « cascade de Haar de Viola & Jones ».
- Bouton de détection et reconnaissance de visage en temps réel. La détection est effectuée par l'emploi du modèle MTCNN et et la

reconnaissance par le modèle VGG-16. Le système ajoute un nouveau visage en appuyant sur le bouton « **ajouter un nouveau employé** ».

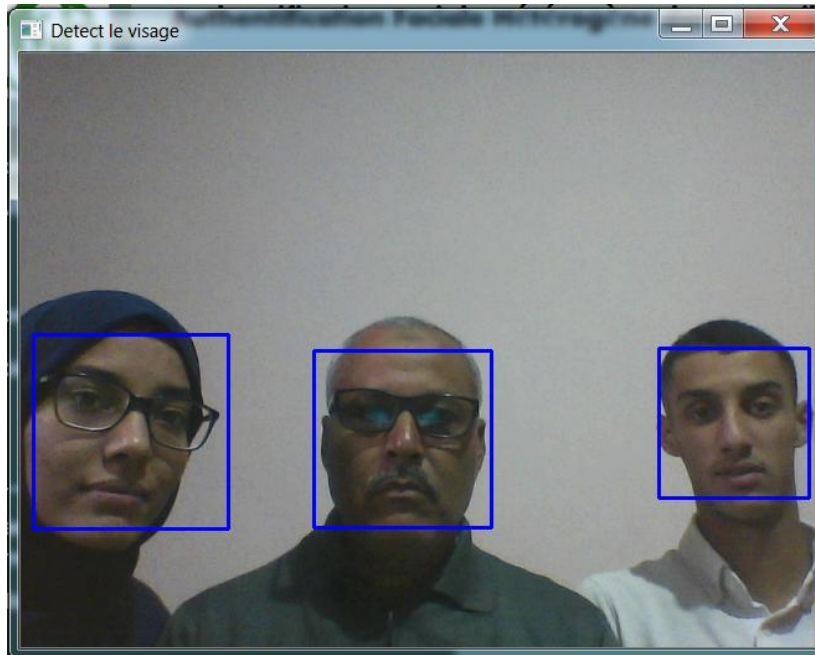
- Afin de visualiser la liste du personnels CDTA, on clique sur "**Afficher la base de données** "



**Figure 5.9 : L'interface principale de l'application.**

### 5.6 Détection standard de visage

L'intérêt de cette application est d'utiliser une méthode standard de détection de visages dans les images. Pour cela l'algorithme de détection de Viola-Jones est utilisé, car il permet une détection robuste et rapide des visages, utilisant le fichier " [haarcascade\\_frontalface\\_default.xml](#)". Figure 5.10 montre les visages détectés dans une scène vidéo.

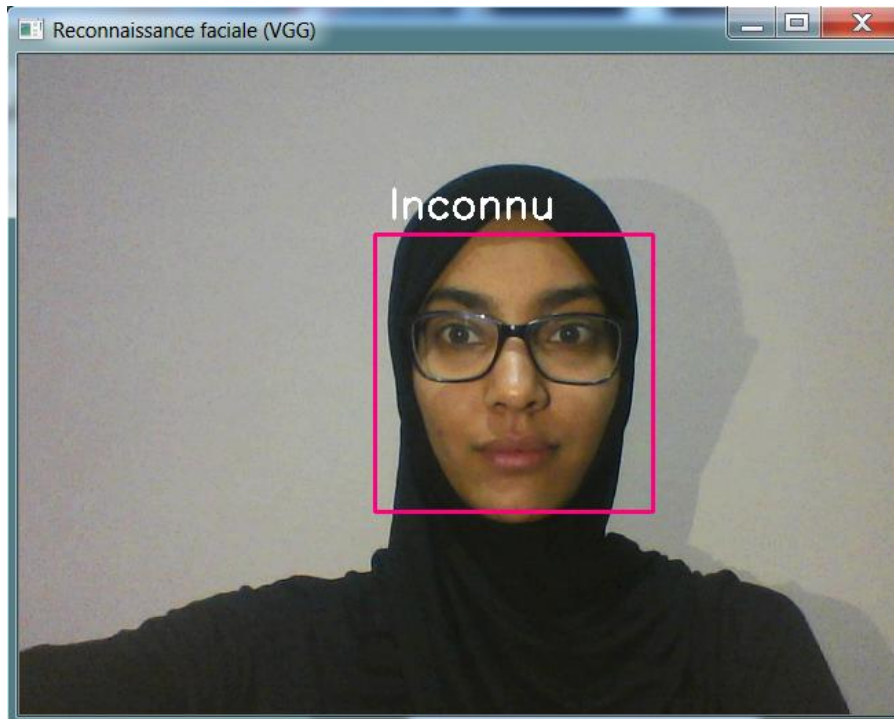


**Figure 5.10 : Détection des visages par Haar cascade.**

### **5.6.1 Détection et reconnaissance de visage en utilisant les modèles MTCNN et VGG-16**

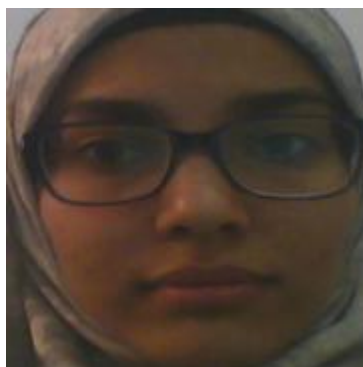
Suite à la détection, le système utilise le modèle VGG-16 pour faire la reconnaissance de personne, l'image est redimensionnée à 224x224 pixels et est pré-normalisée en enlevant la valeur moyenne des pixels de toutes les images de la base d'entraînement. Ainsi, tous les régions d'intérêt (ROI: region of interest) capturés par Viola-Jones sont redimensionnés pour combler les requis des entrées du CNN. Figure 5.11 montre le cas où la personne n'est pas reconnue par le système.





***Figure 5.11 : Personne non reconnu par le système.***

Pour que le système reconnaisse la personne (Figure5.11), il faut l'acquérir et la stocker dans la base de données. La Figure 5.12 montre une image stockée dans la base de donnée d'une personne reconnu par le système.



***Figure 5.12 : L'image alignée et stockée dans la base de données des images.***



**Figure 5.13** : La fenêtre montrant l'ajout d' une nouvelle personne.

Dès que on click sur la bouton " **Ajouter un nouveau personne**", l'interface affiche la fenêtre illustrée sur la Figure 5.13.

- **(1)Ajouter** : pour ajouter et enregistrer les informations saisies dans la base de donnée.(Figure5.14).
- **(2)Acquisition** : Après l'enregistrement des données, le bouton « **acquisition** » exécute la fonction d'acquérir un visage d'une nouvelle personne.
- **(3) Alignement** : cette fonction permet d'aligner le visage capturé. Elle est assuré par le modèle MTCNN.
- **Afficher DB:** Affiche la liste des personnes stockés.(Figure 5.15)
- **Ouvrir DB:** ouvrir la base de données dans une autre interface.(Figure 5.14)

La Base de Donnée des Personnes

--- Les Personnes enregistrés dans la Base de Donnée ---

ID	Nom	Prénom
1	AITSAADI	Karima
2	AZOUZ	Sarah
3	AMEUR	Kaouther
5	MEBARKI	Latifa
6	BELKACEM	Abderrahime
8	AMEUR	Moussa
9	AMEUR	Selma

Terminer

Figure 5.14 : La personne ajoutée dans la base de données.

Ajouter une nouvelle personne

### Ajouter une nouvelle personne

Nom :

Prénom :

5 MEBARKI Latifa

3 AMEUR Kaouther

2 AZOUZ Sarah

1 AITSAADI Karima

( Ils sont classés du plus récent )

**X Effacé**

**(1) Ajouter**

**(2) Acquisition**

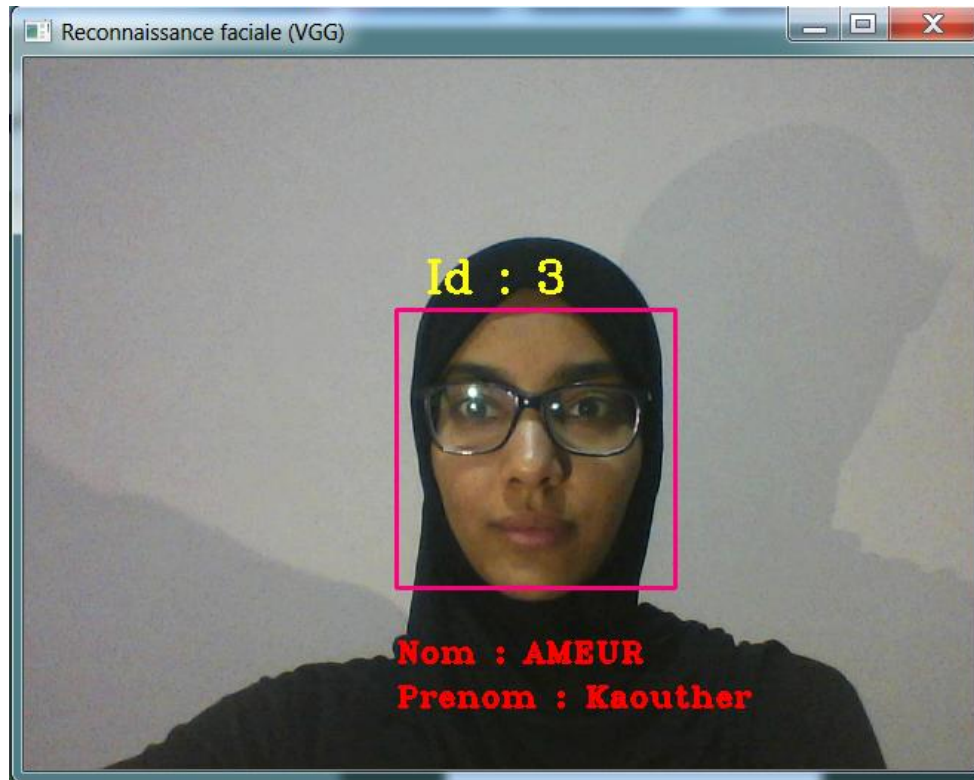
**(3) Alignement**

**Afficher BDD**

**Ouvrire BDD**

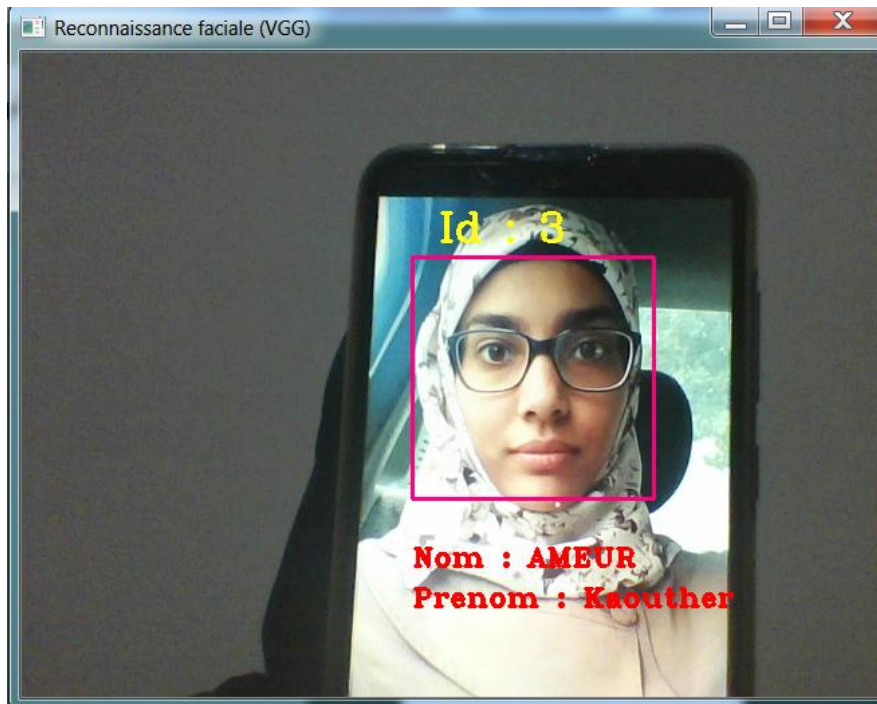
Figure 5.15 : Affichage de la liste de personnes stockées.

Figure 5.16 illustre la reconnaissance d'un visage stocké dans la base de données et son authentification en extrayant les informations correspondant de la base de donnée.



**Figure 5.16 : Authentification de la personne.**

Dans la Figure 5.18. Le visage est détecté , reconnu et authentifié à partir d'un téléphone portable.



**Figure 5.17 : Image sur un téléphone portable.**

## **5.7 Conclusion**

A travers ce chapitre, nous avons présenté l'environnement de développement et les outils utilisés, plus la réalisation de notre système d'authentification faciale ainsi que l'interface graphique permettant superviser les différents traitements du système.

## Conclusion générale et perspectives

L'objectif de ce mémoire est d'implémenter une application de détection, de reconnaissance et d'authentification faciale en temps réel qui répond à des besoins de gestion et de contrôle d'accès des employés du Centre de Développement des Technologies Avancées (CDTA). L'application en question doit répondre à des exigences de rapidité et de robustesse des résultats.

L'application implémentée est basée sur le modèle MTCNN pour la détection des visages et leurs alignement et le modèle VGG-16 pour la reconnaissance et l'authentification. Les raisons de ce choix est double. D'une part, les modèles exploités dans notre application sont imposés par les encadreurs, et d'autre part, les articles dans la littérature ont bien démontré l'avantage de ce modèle par rapport à d'autres modèles. La combinaison des modèles MTCNN et VGG-16 reprend les avantages de réseaux classiques tout en améliorant grandement sa vitesse.

L'application ainsi développée répond aux objectifs du projet et ses performances en termes de taux de reconnaissance reste à améliorer

Ce mémoire de master 2 nous a permet de :

- ❖ Approfondir nos connaissances théoriques et pratique déjà acquises, maîtriser les nouvelles techniques et compléter notre initiale pour atteindre ainsi un niveau de perfection supérieur et de pouvoir apprendre d'autres nouveautés dans les différents domaines de la science en générale et de l'informatique en particulier.
- ❖ Construire des savoirs et des savoir-faire dans le domaine de traitement d'images et de vidéo particulièrement, la détection, la reconnaissance.
- ❖ Comprendre les différentes étapes pour créer des algorithmes de deep-learning en python.

Un prolongement de ce travail pourrait s'axer sur plusieurs points à savoir :

## **Conclusion générale et perspectives**

- ❖ Améliorer l'architecture afin de la rendre plus légère ;
- ❖ Effectuer le modèle de reconnaissance,
- ❖ Enfin, concevoir une plateforme de reconnaissance de visages hétérogènes ou non contrôlée, dans lequel les images faciales présentent plus de variabilités sur les éclairages. En effet, Aujourd'hui, les techniques conventionnelles de reconnaissance faciale ont atteint une performance quasi-parfaite dans un environnement fortement contraint où la pose, l'éclairage, l'expression faciale et d'autres sources de variation sont contrôlées. Cependant, ces méthodes sont souvent destinées aux domaines d'application limités parce que les environnements de prise des visages ne sont pas idéals dans les cas pratiques

# Références

- [1]: Cha.Zhang,Zhengyou Zhang,Morgan& Claypool, « Boosting-Based Face Detection adaptation», Microsoft research,microsoft corporation,(2010).  
<http://www.research.microsoft.com>.
- [2]:yi. quing wang, ' An Analyse of the viola and jones Face detection Algorithm',(2010).
- [3]: Zhao, Wenyi and Chellappa, Rama and Phillips, P Jonathon and Rosenfeld, Azriel, ' Face recognition: A literature survey ', ACM computing surveys (CSUR), 35, (2003).
- [4]: Jain, Anil K and Li, Stan Z, ' Handbook of face recognition', Springer,(2011).
- [5]: Khefif, Bouchra, ' Mise au point d'une application de reconnaissance faciale' , (2013).
- [6]: Adel Oulefki, ' Defense of accreditation to supervise research 'Summary of research and pedagogical activities' ,(2018),  
[https://www.researchgate.net/publication/324991823\\_Defense\\_of\\_accreditation\\_to\\_supervise\\_research\\_Summary\\_of\\_research\\_and\\_pedagogical\\_activities](https://www.researchgate.net/publication/324991823_Defense_of_accreditation_to_supervise_research_Summary_of_research_and_pedagogical_activities).
- [7]: Rouse,M 'access control', (2018)
- [8] : Barki, Hicham, Lgaudan, L, 'La biométrie dans les aéroports et ses enjeux'. 'DETECTION ET RECONNAISSANCE DE VISAGE' ,PhD THE-SIS (2018).
- [9]: Delvau,A, 'Dictionnaire de langue verte : argots parisiens comparés'. E,Dentu (1867).
- [10]: Faudez-zanuy.M, Bionmetricsécurité technology,IEEE, 'Aerospace and Electronic Systems Magazine', 21,6, (2006).
- [11]: Oulefki.A, ' Defense of accreditation to supervise research, summary of research and pedagogical activities ' .
- [12]: Lgaudan L, La biométrie dans les aéroports et ses enjeux.
- [13]: Oulefki A, 'Defense op accreditation to supervise research', 'summary of research and pedagogical activities',(2018).
- [14]: Franrenet S, Dispositifs biométriques, dispositifs éthiques.
- [15]: Oulefki A, Defense op accreditation to supervise research 'summary of research and pedagogical activities',(2018).
- [16]: Golfareli, M , Maio and Malton D, 'On the error-reject trade-off in biometric verification systems', IEEE Transactions on Pattern Analysis and Machine Intelligence 19,7(1991).
- [17]: Golfareli M, Maio D, and Malton D, On the error-reject trade-off in biometric verification systems, IEEE Transactions on pattern Analysis and Machine intelligence 19, 7(1997), 786-796.
- [18]: Ababsa S G, Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2d/3d, Evrv-Val d'Essonne,(2008).
- [19]: ouamane A,Reconnaissance Biométrique par Fusion Multimodale du Visage 2d et 3d,In workshop on faces in 'Real-life' Image : Detection, University Mohamed kheider, Alignementan recognition(Biskra,Algerie,2015,Français,Oct2015).
- [20]: Shi, Yichun et Jain, Anil K,' DocFace +: ID Document à Selfie Matching ', 'Transactions IEEE sur la biométrie, le comportement et la science de l'identité', 1, IEEE ,(2019).



- [21]: Soltane, Mohamed et Bakhti, Mimen, 'Authentications biométriques multimodales: problèmes de concepts et stratégies d'applications', 'Journal international des sciences et technologies avancées',48, Citeseer,(2018)
- [22]: False Acceptance Ratio (FAR), '<https://www.techopedia.com/definition/27569/false-acceptance-ratio-far>'
- [23] : Kodratoff, Yves,L'apprentissage symbolique automatique, Cepadues-Editions(1986).
- [24] : N'Guessan, A and Essai, A and Langrand, and C, 'Estimation multidimensionnelle des controles et de l'effet moyen d'une mesure de sécurité routière', 'Revue de statistique appliqué', 49, 2,(2001)
- [25]: réault, Jeanny and Jutten, Christian and Ans, Bernard, 'tection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé',10 Colloque sur le traitement du signal et des images, FRA, (1985), GRETSI, Groupe d'Etudes du Traitement du Signal et des Images
- [26] : Assude, Teresa,'Ecologie de l'objet', 'racine carré et analyse du curriculum),(1993)
- [27] : Assens, Christophe, Du modèle de bureaucratique au modèle d'organique L'organisation en réseau, 12,23,(1991),GR 903
- [28] : Renders, Jean-Michel,'Algorithmes génétiques et réseaux de neurone', Renders, Jean-Michel, Hermès Paris(1994).
- [29]: Noureddine, Ali , 'Verification faciale de la parenté par un réseau de neurone à convolution, <http://cs231n.github.io/neural-networks-1/>.
- [31]: Noureddine, Ali , 'Verification faciale de la parenté par un réseau de neurone à convolution,
- [32]: OD datamining, ,(2018),<https://od-datamining.com/knownbase/les-reseaux-de-neurones-expliques-a-ma-fille/>
- [33] : Ha, Thanh-Le et Niehues, Jan et Waibel, Alexander , Vers une traduction neuronale multilingue avec codeur et décodeur universels,'arXiv preprint arXiv: 1611.04798',(2016).
- [34]: Rachdi Simo, 'DevOps pour le machine Learning (ML, IA)', (2019), <http://datasciencetoday.net/index.php/fr/machine-learning>.
- [38] : LeCun, Yann et Bengio, Yoshua et Hinton, Geoffrey,'apprentissage en profondeur',nature,521, 'Nature Publishing Group',(2015).
- [38]: Russell, Stuart J et Norvig, Peter, 'Intelligence artificielle: une approche moderne', Malaysia; Pearson Education Limited, (2016).
- [41]: Kodratoff, Yves,L'apprentissage symbolique automatique, Cepadues-Editions ,(1986).
- [42]: François, Olivier and Leray, Philippe, 'Etude comparative d'algorithmes d'apprentissage de structure dans les réseaux bayésiens' , 'Rencontres des Jeunes Chercheurs en IA',(2003).
- [43]: Lynn Heidmann AI vs. Machine Learning vs. Deep Learning.(2019), , <https://blog.dataiku.com/ai-vs.-machine-learning-vs.-deep-learning>
- [45]: Kohli, Marc et Prevedello, Luciano M et Filice, Ross W et Geis, J Raymond,'Implémentation de l'apprentissage automatique dans la pratique et la recherche en radiologie', 'journal américain de radiologie',208, Am Roentgen Ray Soc,(2017).
- [46]: Micheal A, 'Comment fonctionne l'algorithme de rétropropagation' ,(2019), <http://neuralnetworksanddeeplearning.com/chap2.html>.

- [47]: Gillot, Pierre and Zemmari, Akka and Benois-Pineau, Jenny and Nesterov, Yurii, 'Algorithmes de Descente de Gradient Stochastique avec le filtrage des paramètres pour l'entraînement des réseaux à convolution profonds'.
- [48]: Gillot, Pierre and Zemmari, Akka and Benois-Pineau, Jenny and Nesterov, Yurii, 'Algorithmes de Descente de Gradient Stochastique avec le filtrage des paramètres pour l'entraînement des réseaux à convolution profonds', (2018).
- [49]: Lguensat, R and Rjiba, S and Fablet, R and Mason, E 'EddyNet: Réseau de neurones convolutifs pour la détection et la classification des tourbillons océaniques', (2018).
- [50]: Zemouri, Ryad, 'Contribution à la surveillance des systèmes de production à l'aide des réseaux de neurones dynamiques: Application à la e-maintenance', Université de Franche-Comté, (2003).
- [51]: Lambert R, 'le Réseau de Neurones Convolutifs', (2019), <http://penseeartificielle.fr/focus-reseau-neurones-convolutifs/>.
- [52]: Romain Heral, clement chatelain, 'Découvrez le fonctionnement des réseaux de neurones récurrent', (2019), <https://openclassrooms.com/fr/courses/5801891-initiez-vous-au-deep-learning/5814651-decouvrez-le-fonctionnement-des-reseaux-de-neurones-recurrents>.
- [53]: Rachdi Simo, 'DevOps pour le machine Learning (ML, IA)', (2019),
- [54]: Chloé Agat, 'Entraînez un réseau de neurones simple', (2019).
- [55]: Chloé Agat, 'Entraînez un réseau de neurones simple', (2019).
- [55]: Ashish Patel, 'Chapter-7 Under-fitting, over-fitting and its solution', (2018).
- [56]: Viola, Paul et Jones, Michael, 'Détection rapide d'objets utilisant une cascade optimisée de fonctionnalités simples', CVPR, 1, (2001).
- [57]: Mathieu Van Wambeke, 'Reconnaissance et suivi de visages et implémentation en robotique temps-réel', Université Catholique de Louvain Ecole Polytechnique de Louvain, (2010).
- [58]: Maartin C, Richard O, and SAINT MARTIN-France, m, 'Algorithme de vol de vol de travail appliqué au déploiement d'application parallèles, Soumis RenPar's 15, (2003).
- [59]: Dabhi M, and Pancholi B, 'Face detection system based on viola and jones algorithm, international', journal of science and research, (2016).
- [65]: Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. 'Face description with local binary patterns: Application to face recognition', IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006.
- [66]: Zisserman, MM. Andrew et Matas, MM. Jiri et Crowley, MM. James et Jurie, MM. Freric et Triggs, M. Bill, 'Méthodes d'apprentissage machine pour la détection d'objets visuels', (2012).
- [67]: Zhao, Jiali et Wang, Haitao et Ren, Haibing et Kee, Seok-Cheol, 'Analyse discriminante LBP pour la vérification du visage', Conférence 2005 de l'IEEE Computer Society sur la vision par ordinateur et la reconnaissance des formes', IEEE, (2005).
- [68]: Zhao, Jiali et Wang, Haitao et Ren, Haibing et Kee, Seok-Cheol, 'Analyse discriminante LBP pour la vérification du visage', Conférence 2005 de l'IEEE Computer Society sur la vision par ordinateur et la reconnaissance des formes', IEEE, (2005).
- [69]: Schroff, Florian et Kalenichenko, Dmitry et Philbin, James 'Facenet: une intégration unifiée pour la reconnaissance des visages et la mise en cluster', Actes de la conférence de l'IEEE sur la vision par ordinateur et la reconnaissance des formes', (2015).

- [70]: Pietra FT Madio, 'A FaceNet-Style Approach to Facial Recognition on the Google Coral Development board', (2019) <https://towardsdatascience.com/a-facenet-style-approach-to-facial-recognition-dc0944efe8d1>
- [71] : Milind Deore, 'Architecture FaceNet. Partie 1: Architecture et exécution d'un exemple de base sur Google Colab', <https://medium.com/analytics-vidhya/facenet-architecture-part-1-a062d5d918a1>.
- [72] : Ahmed Taha, 'FaceNet: une incorporation unifiée pour la reconnaissance faciale et le regroupement', (2018), <https://medium.com/@ahmdtaha/facenet-a-unified-embedding-for-face-recognition-and-clustering-7d34abde9>
- [75]: Al-Jawfi, Rashad 'Handwriting Arabic character recognition LeNet using neural network', *Int. Arab J. Inf. Technol.* 6, 2009
- [77]: Max Pechyon Kin, 'key deep learning architecture LeNet-5', (2018).
- [78] : Insaf Selma, 'Detection d'objet', Mémoire université bab zouar, (2019).
- [79]: R. Khandelwal. « Deep learning –deep belief network » (2018) : <https://medium.com/datadriveninvestor/deep-learning-deep-belief-network-dbn-ab715b5b8afc>
- [80] : S. Das « CNN Architectures : LeNet, AlexNet, VGG, GoogLeNet, ResNet and more... » (2017) : <https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5>.
- [81] : Iandola, Forrest N and Han, Song and Moskewicz, Matthew W and Ashraf, Khalid and Dally, William J and Keutzer, Kurt, ' AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size', arXiv preprint arXiv:1602.07360, (2016).
- [82]: Zhong, Zhuoyao and Jin, Lianwen and Xie, Zecheng, 'High performance offline handwritten chinese character recognition using googlenet and directional feature maps', '13th International Conference on Document Analysis and Recognition (ICDAR)', IEEE', (2015).
- [83]: S. Das « CNN Architectures : LeNet, AlexNet, VGG, GoogLeNet, ResNet and more... » (2017) : <https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5>.
- [85]: Handa, Ankur and Bloesch, Michael, Viorica and Stent, Simon and McCormac, John and Davison, Andrew, 'gvnn: Neural network library for geometric computer vision', European Conference on Computer Vision', (2016).
- [87] : Xiang, Jia et Zhu, Gengming, 'Détection des visages et reconnaissance de l'expression du visage avec MTCNN', '2017 4ème Conférence internationale sur la science de l'information et l'ingénierie de commande (ICISCE)', IEEE, (2017).
- [88] : Chi Feng Wang, 'What Does A Face Detection Neural Network Look Like', (2018).
- [89] : Van Wambeke, 'Reconnaissance et suivi de visages et implémentation en robotique temps-réel', Université Catholique de Louvain Ecole Polytechnique de Louvain, (2010).
- [92]: Dang Ha le Hien, 'A guide to receptive field arithmetic for Convolutional Neural Networks', (2017), <https://medium.com/mlreview/a-guide-to-receptive-field-arithmetic-for-convolutional-neural-networks-e0f514068807>.
- [93]: R. Khandelwal. « Deep learning –deep belief network » (2018) : <https://medium.com/datadriveninvestor/deep-learning-deep-belief-network-dbn-ab715b5b8afc>.
- [108] : <https://www.infoworld.com/article/3336192/what-is-keras-the-deep-neural-network-api-explained.html>
- [16] : Kaipeng Z, Zhanpeng Z, Zhifeng L, and Yu Qiao. *IEEE*, "Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks.