

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université KASDI MERBAH -OUARGLA



Faculté des Nouvelles Technologies de L'Informations et de la Communication

Département de Mathématique et Informatique

Filière : Informatique

**Mémoire en vue de l'obtention du diplôme de Master 2 en Administration
et Sécurité des Réseaux**

Thème :

**Reconnaissance de forme et classification automatique :
Application à la prédiction du genre et l'estimation de l'âge**

Présenté par :

LAHLALI Sara TRABELSI Maroua

Devant le jury composé de :

- **Président de Jury :** KHALDI Amine *MCA UKM, Ouargla*
- **Encadreur :** BENKADDOUR Mohammed Kamel *MAA UKM, Ouargla.*
- **Examineur :** BOUKHAMLA Akram *MCA UKM, Ouargla*

Année universitaire 2019-2020

REMERCIEMENTS

Tout d'abord, louange Seigneur "ALLAH" qui nous a dotées de la merveilleuse faculté de raisonnement.

Nos remerciements les plus cordiaux sont adressés particulièrement à notre encadreur Mr **BENKADDOUR** Mohammed Kamel, de nos avoirs soutenus et fait confiance durant notre projet avec une grande patience. Avec son expérience dans la recherche et l'enseignement, avec ses conseils, nous avons pu découvrir le monde de la recherche scientifique dans le domaine d'intelligence artificielle et des techniques de la biométrie.

Nous tenons également à remercier les membres du jury pour avoir accepté d'examiner et d'évaluer ce modeste travail.

Nous souhaitons à présent adresser notre sincère remerciements à toutes les personnes avec qui nous avons eu la chance de travailler pendant notre mémoire, et à tous les enseignants de FNTIC de l'université de OUARGLA.

Dédicace

- Je dédie ce modeste travail :
 - À mes chers parents
 - À ma sœur : Chiffa et son fils Mohammed
 - À mes frères : Abd El Basset, Yakoub, Tarek, Moussaab.
 - À toute ma famille.
 - À tous mes amis, tout particulièrement : khawter, Asma, Chourouk, Amina, Ikram, Safa, Lamia, Nihad.
 - À mes collègues
 - À mon binôme et sa famille Lahlali.
 - À mon encadreur pour son aide.
 - À toute personne, qui m'a aidé à réaliser ce travail, de près ou de loin

Maroua

SOMMAIRE

LISTE DES FIGURES.....	IV
LISTE DES TABLEAUX.....	V
LISTE DES ACRONYMES	IV
INTRODUCTION GENERALE.....	8

CHAPITRE 1 : LA BIOMETRIE

1.1. INTRODUCTION.....	11
1.2. LA BIOMETRIE	11
1.2.1 Définition de la biométrie	11
1.2.2 Les modes opératoires d'un système biométrique	12
1.2.3 Les exigences de la Biométrie.....	12
1.3. LE MODE DE FONCTIONNEMENT D'UN SYSTEME BIOMETRIQUE.....	13
1.3.1 Structure interne d'un système biométrique.....	14
1.4. LES DIFFERENTES MODALITES BIOMETRIQUES	15
1.5. LES TECHNOLOGIES BIOMETRIQUES	16
1.5.1. L'empreinte digitale	16
1.5.2. Géométrie de la main	17
1.5.3. Le visage	17
1.5.4. L'iris.....	17
1.5.5. La rétine.....	18
1.5.6. La dynamique de frappe au clavier	18
1.5.7. La dynamique du tracé de la signature.....	19
1.5.8. La voix.....	19
1.5.9. Analyse de l'ADN.....	19
1.6. ÉVALUATION DES PERFORMANCES DES SYSTEMES BIOMETRIQUES	19
1.6.1. Mesure de la performance d'un système biométrique	22
1.6.2. Les applications des systèmes biométriques	23
1.7. LA RECONNAISSANCE FACIALE.....	23
1.7.1. Principales difficultés de la reconnaissance du visage.....	24
1.7.2. Motivation :la reconnaissance faciale pour la prédiction de genre et l'âge	26
1.8. CONCLUSION.....	27

CHAPITRE 2 : L'ETATS DE L'ART DE PREDICTION DU GENRE ET L'ESTIMATION D'AGE

2.1. INTRODUCTION.....	29
2.2. LA RECONNAISSANCE DE VISAGE	29
2.3. LES TECHNIQUES DE RECONNAISSANCE FACIALE.....	29
2.3.1. Approche globale (holistique).....	30
2.3.2. Approche locale.....	33
2.3.3. Approche hybride.....	35
2.4. RECAPITULATIF DES PRINCIPALES APPROCHES DE RECONNAISSANCE FACIALE	37

2.5.	QUELQUES TRAVAUX SUR LES APPROCHES DE LA RECONNAISSANCE	38
2.6.	LA PREDICTION DU GENRE ET DE L'AGE.....	39
2.6.1.	Les techniques de prédiction de genre	40
2.6.2.	Les techniques d'estimation d'âge	40
2.6.3.	Quelques travaux développés pour prédiction du genre et de l'âge :.....	41
2.7.	CONCLUSION :.....	42

CHAPITRE 3 : APPRENTISSAGE PROFOND LES RESEAUX DE NEURONES CONVOLUTIFS

3.1.	INTRODUCTION.....	44
3.2.	L'INTELLIGENCE ARTIFICIELLE.....	44
3.2.1.	Définition	44
3.2.2.	Domaines de l'AI	45
3.3.	L'APPRENTISSAGE AUTOMATIQUE.....	45
3.3.1.	Les réseaux de neurones artificiels (RNA)	46
3.3.2.	Le fonctionnement des réseaux neurones.....	47
3.3.3.	Les topologies de RNA	48
3.3.4.	Le perceptron multicouche (PMC).....	49
3.4.	L'APPRENTISSAGE PROFOND	50
3.4.1.	Pourquoi l'apprentissage profond ?.....	50
3.4.2.	Les différents types de modèles	51
3.5.	LES RESEAUX DE NEURONES CONVOLUTIFS	53
3.5.1.	L'architecture en couche du réseau CNN	55
3.5.2.	Les modèles CNN célèbre.....	57
3.5.3.	L'entraînement d'un CNN	60
3.5.4.	L'optimisation dans les réseaux profonds.....	61
3.5.5.	Les Algorithmes d'optimisation de la descente de gradient	61
3.6.	L'APPRENTISSAGE PAR TRANSFERT	63
3.7.	CONCLUSION	63

CHAPITRE 4 : IMPLEMENTATION, EXPERIMENTATION ET RESULTATS

4.1.	INTRODUCTION.....	65
4.2.	LES BASE DE DONNEES DE VISAGE.....	65
4.2.1.	FERET.....	65
4.2.2.	IMDB & WIKI.....	65
4.2.3.	Séparation des bases de données	66
4.3.	L'ENVIRONNEMENT DE TRAVAIL	66
4.3.1.	Les outils de programmation.....	66
4.3.2.	Les bibliothèques utilisées :	67
4.4.	LES SYSTEMES DE RECONNAISSANCE VISAGE PAR CNN.....	67
4.5.	SYSTEME DE PREDICTION DE GENRE : PRINCIPES ET EXPERIMENTATIONS	68
4.5.1.	L'effet de l'architecture :.....	69
2.5.1	Effet du nombre d'images d'apprentissage sur le taux de reconnaissance	75
4.6.	SYSTEME D'ESTIMATION D'AGE	77
4.7.	SYSTEME DE RECONNAISSANCE DE VISAGE	78
4.8.	COMPARAISON DES PERFORMANCES.....	79

4.9. PRESENTATION DE L'APPLICATION	80
4.10. CONCLUSION.....	86
CONCLUSION ET PERSECTIVES	87
ANNEXE	88
BIBLIOGRAPHIE	89
RESUME.....	96

LISTE DES FIGURES

Fig 1.1 : Les modes de fonctionnement d'un système biométrique [2].	14
Fig 1.2 : Principaux modules d'un système biométrique et différents modes de fonctionnement [5].	15
Fig 1.3 : Les différentes modalités biométriques	16
Fig 1.4 : Les technologies biométriques de l'empreinte digitale.	16
Fig 1.5 : Dispositif de reconnaissance par géométrie de la main [8].	17
Fig 1.6 : Les technologies biométriques de reconnaissance de visage.	17
Fig 1.7 : Système biométrique basée sur l'iris [8].	18
Fig 1.8 : Exemple deux rétines distinct.	18
Fig 1.9 : Système biométrique base sur la frappe au clavier.	18
Fig 1.10 : Système Biométrique Base Sur La Frappe Au Clavier [2].	19
Fig 1.11 : Illustration du FRR et du FAR.	23
Fig 1.12 : Illustration de l'impact des différents éclairages sur les images des visages.	25
Fig 1.13 : Exemples de variation d'éclairage.	25
Fig 2.1 : Classification des approches principales utilisées en reconnaissance faciale.	30
Fig 2.2 : Conversion d'image (matrice) en vecteur colonne	31
Fig 2.3 : Les projections ACP et LDA d'un ensemble de données [5].	31
Fig 2.4 : (a) Localisation des caractéristiques géométriques utilisées dans [44].	33
Fig 2.5 : Approche proposée par Heisele et al. [47].	33
Fig 2.6 : Modèle Actif d'Apparence [49].	34
Fig 2.7 : Exemple de la méthode MB_LBP pour un visage [51].	34
Fig 2.8 : Approche Local Component Analysis [52].	35
Fig 2.9 : Graphe appliqué aux visages pour l'approche EBGM [53].	36
Fig 2.10 :Création des jets pour l'approche EBGM [53].	36
Fig 3.1 : Les différentes méthodes de ML.	46
Fig 3.2 : Exemple d'un neurone artificiel (perceptron) [108].	47
Fig 3.3 : Fonctions d'activation couramment utilisés	48
Fig 3.4 : Topologie des Réseaux de neurones artificiels [109].	49
Fig 3.5 : Schéma d'un perceptron multicouche [108].	49
Fig 3.6 : La relation entre l'intelligence artificielle, L'apprentissage automatique et l'apprentissage profond.	50
Fig 3.7 : La différence de performance entre l'apprentissage profond et la plupart des algorithmes de l'apprentissage automatique en fonction de la quantité de données [109].	51
Fig 3.8 : Le procède du ML classique comparé à celui du Deep Learning [109]	51
Fig 3.9 : Les modèles de l'apprentissage profond [109].	53
Fig 3.10 : Architecture basique d'un réseau de neurone convolutifs [111]	53
Fig 3.11 : Sparse interactions [119].	54
Fig 3.12 : Exemple de convolution d'une Image 2D De Taille 4×3 Par Un Noyau 2×2 [119].	56
Fig 3.13 : Max Pooling avec un filtre 2×2 et un pas de 2.	56
Fig 3.14 : La couche entièrement connectée.	57
Fig 3.15 : Modules de l'architecture GoogLeNet [120].	59
Fig 3.16 : Blocs résiduels qui permettent aux couches intermédiaires de ResNet [120].	59
Fig 4.1 : Exemples d'images de visages de la base FERET.	65
Fig 4.2 : Exemples d'images de visages de la base IMDB.	66
Fig 4.3 : Un système de reconnaissance de visages.	67
Fig 4.4 : Le processus de prédiction de genre par le modèle CNN.	68

Fig 4.5 : L'architecture du réseau CNN (modèle n°1).	69
Fig 4.6 : La configuration du modèle n°1.	70
Fig 4.7 : L'architecture du modèle CNN (modèle n°2).	71
Fig 4.8 : La configuration du modèle N°2.	71
Fig 4.9 : L'architecture du modèle CNN (modèle n°3).	72
Fig 4.10 : La configuration du modèle CNN N°3.	73
Fig 4.11 : Taux de précision du modèle 1.	74
Fig 4.12 : Taux de précision du modèle 2.	74
Fig 4.13 : Taux de précision du modèle 3.	75
Fig 4.14 : Evolution de la courbe d'apprentissage du modèle 3.	76
Fig 4.15 : L'architecture CNN pour l'estimation d'âge (modèle n°3).	77
Fig 4.16 : Fenêtre d'accueil de l'application.	80
Fig 4.17 : Fenêtre qui illustre le module de vérification.	81
Fig 4.18 : L'emplacement des photos.	81
Fig 4.19 : Cas d'une personne bien identifiée	81
Fig 4.20 : La performance de système de reconnaissance de visage.	82
Fig 4.21 : L'interface de prédiction du genre et d'âge	82
Fig 4.22 : Un exemple de prédiction du genre.	83
Fig 4.23 : Un exemple d'estimation d'âge.	83
Fig 4.24 : Exemple de prédiction du genre et estimation d'âge simultanément.	84
Fig 4.25 : Résultats de classifications sur la base de données wiki et IMDB.	84
Fig 4.26 : Exemple de prédiction du genre et d'âge en temps réel.	85
Fig 4.27 : Le cas d'aucune personne détectée	85

LISTE DES TABLEAUX

Tab 1.1 : Les avantages et inconvénients des différentes technologies biométriques.	21
Tab 2.1 : Les avantages et les limitations des approches de reconnaissance faciale.	38
Tab 2.2 : Comparaison des approches basées sur les caractéristiques locales et globales en fonction des critères de variation [60].	38
Tab 2.3 : Quelques travaux de recherches sur la reconnaissance	39
Tab 2.4 : Synthèse des travaux développés pour la prédiction de genre.	41
Tab 2.5 : Synthèse des travaux développés pour la prédiction d'âge.	42
Tab 3.1 : Les gagnants du concours ImageNet de 2012 à 2015[120].	58
Tab 4.1 : Les bases de données utilisées.	66
Tab 4.2 : Les résultats de reconnaissance de genre obtenus sur IMDB	73
Tab 4.3 : Taux de précision par rapport le nombre d'entraînement	75
Tab 4.4 : Les résultats après l'entraînement du modèle 3.	76
Tab 4.5 : Les taux de classification d'âge sur les bases de données WIKI et IMDB	78
Tab 4.6 : Taux de reconnaissance faciale sur la base de données FERET.	79
Tab 4.7: Comparaison des performances de certaines méthodes de prédiction du genre et d'âge citées dans l'états de l'arts.	79

LISTE DES ACRONYMES

- **2D** : Représentation A 2 Dimensions
- **ACI** : Analyse en Composantes Indépendantes
- **ACP** : Analyse en Composantes Principale
- **ADB** : Analyse Discriminante Bilinéaire
- **AAM** : Active Appearance Models
- **ANN** : Artificial Neural Network
- **AI** : Artificial Intelligence
- **CNN** : Convolutional Neural Network
- **DL** : Deep Learning
- **GPU** : Graphical Processor Unit
- **FAR** : False Acceptance Rate
- **FERET** : Face Recognition Technologie
- **FRR** : False Rejection Rate
- **IBG** : Américaine International Biometric Group
- **IMDB** : Internet Movie Database
- **LBP** : Local Binary Pattern
- **LDA** : Linear Discriminant Analysis
- **LEM** : Line Edge Maps
- **LLE** : Local Linear Embedding
- **LSTM** : Long Short-Term Memory
- **LTSA** : Local Tangent Space Analysis
- **MDS** : Le Multidimensional Scaling
- **MBLBP** Multi-Scale Block Binary Pattern
- **NMF** : Non-Negative Matrix Factorization
- **PMC** : Perceptron Multi-Couche

- **RELU** : Rectified Linear Unit
- **RNA** : Réseau de Neurones Artificiels
- **RNN** : Recurrent Neural Network
- **SVC** : Support Vectors Clustering
- **SVM** : Support Vectors Machines
- **VCD** : Vecteurs Communs Discriminants

INTRODUCTION GENERALE

La biométrie, est la science qui porte sur l'analyse des caractéristiques physiques ou comportementales propres à chaque individu permettant l'authentification de son identité d'une manière fiable, elle offre des avantages significatifs par rapport aux techniques d'authentification classiques, comme les mots de passe, les cartes puce ...etc., notamment en raison des caractéristiques biométrique ne sont pas transférables, unique pour chaque personne et ne sont pas perdues ou volés. Le choix des solutions biométriques dépend de l'acceptation de l'utilisateur, du niveau de sécurité, du coût de précision et du temps de mise en œuvre.

Cette dernière décennie, la reconnaissance faciale est l'un des domaines de recherche les plus actives dans la vision par ordinateur et la reconnaissance des formes, de nombreuses applications font appel à cette technologie car le visage humain est considéré comme une source d'information très riche, au point que certains psychologues en parlent comme d'une « fenêtre de l'âme ». En particulier, le genre et l'âge sont des caractéristiques faciales qui peuvent être très utiles pour une multitude d'applications, par exemple un système automatique de prédiction du genre et de l'âge est utilisé afin de profiler les clients qui sont intéressés par un produit ou pour la publicité cible.

La prédiction du genre et l'estimation d'âges sont des tâches que les humains effectuent naturellement et sans effort dans leurs vies quotidiennes, la dotation d'un ordinateur par cette capacité de raisonnement reste encore un grand challenge.

Afin de donner aux ordinateurs la particularité de reconnaissance de visage et la classification automatique des objets, plusieurs méthodes ont été proposé au cours de ces dernières années. Dans notre travail, nous avons opté pour les réseaux de neurones convolutifs (CNN : convolutional neural network) appartenant à la classe des réseaux de neurones feedforward basés sur un apprentissage profond. Ces réseaux convolutifs comporte deux parties différentes, une par l'extraction des caractéristiques de l'image et une leurs classifications.

L'objectif de ce travail consiste à étudier et à analyser l'architecture des reseaux CNN pour implémenter une application capable de prédire le genre et d'estimer l'âge en temps réel à partir de l'images faciales. Pour ce faire, nous avons structuré notre mémoire en quatre chapitres :

- Le premier chapitre est consacré à des généralités sur la biométrie, le concept des systèmes biométriques et leurs modes de fonctionnement.

- Dans le deuxième chapitre, nous évoquerons des techniques de reconnaissance de visage (globales, locales, hybrides), ainsi qu'un état de l'art sur quelques travaux sur la prédiction du genre et de l'âge.

- Le troisième chapitre est consacré à la description des réseaux de neurones à convolution ainsi que leurs l'intérêt dans le domaine de la classification des images.

- Dans le dernier chapitre nous allons présenter les résultats expérimentaux obtenus par notre système de reconnaissance de visage et classification de genre et d'âge suivi d'une discussion avec interprétation des résultats.

CHAPITRE I

LA BIOMETRIE

1. Chapitre I

1.1. Introduction

Face à la fraude documentaire et au vol d'identité, aux nouvelles menaces telles que le terrorisme ou la cybercriminalité et face à l'évolution logique des réglementations internationales, de nouvelles solutions technologiques sont progressivement mises en œuvre. Parmi ces technologies, la biométrie s'est rapidement distinguée comme la plus pertinente pour identifier et authentifier les personnes, de manière fiable et rapide, en fonction de caractéristiques biologiques unique [1].

Nous introduirons dans ce chapitre quelques définitions de base liées à la biométrie. Puis nous donnerons les motivations et les objectifs de ce thème qui porte principalement sur la modalité faciale pour estimer l'âge et le genre d'un personne.

1.2. La biométrie

Les modes opératoires d'un système biométrique peuvent être divisés en deux catégories l'identification et la vérification.

L'identification des visages est un problème de correspondance un contre plusieurs , car il doit comparer un visage avec tous les visages existants dans la base de données. la vérification des visages est un problème de correspondance (1 contre 1) , car il compare un visage avec un autre, l'identification du visage est considérée comme une tâche complexe parce que elle doit faire correspondre le modèle de visage avec d'autres modèles faciales différents, alors que dans la vérification ,il ne doit être associé qu' un seul modèle de visage.

1.2.1 Définition de la biométrie

La biométrie est un terme d'origine grec, qui se compose en deux parties : « bios » qui veut dire « la vie » et « métron » qui se traduit par « la mesure » autrement dit, c'est la mesure du vivant [2], donc la biométrie est la science qui porte sur l'analyse des caractéristiques physiques ou comportementales propres à chaque individu et permettant l'authentification de son identité. Au sens littéral et de manière plus simplifiée, la biométrie signifie la « mesure du corps humain ». [3]. Une autre définition de la biométrie est donnée par Roethenbagh : « la biométrie s'applique à des particularités ou des caractères humains uniques en leur genre et mesurables, permettant de reconnaître ou de vérifier automatiquement l'identité » [4].

1.2.2 Les modes opératoires d'un système biométrique

On trouve deux grands axes dans la biométrie : l'identification et la vérification. Avec l'identification ou la reconnaissance, le système biométrique pose et essaye de répondre à la question, “ Qui suis-je ? ”. Dans une application d'identification, le dispositif biométrique requiert une information biométrique et la compare avec chaque information stockée dans la base de données, c'est une comparaison un à plusieurs (1 : N). Le but des applications d'identification est d'identifier des criminels et des terroristes en utilisant les données des surveillances [5].

Dans la vérification ou l'authentification, le système biométrique demande à l'utilisateur son identité et essaye de répondre à la question, “ Suis-je bien la personne que je prétends être ? ”. Dans une application de vérification l'utilisateur annonce son identité par l'intermédiaire d'un mot de passe, d'un numéro d'identification, d'un nom d'utilisateur, ou toute combinaison des trois. Le système sollicite également une information biométrique provenant de l'utilisateur, et compare la donnée caractéristique obtenue à partir de l'information entrée, avec la donnée enregistrée correspondante à l'identité prétendue, c'est une comparaison un à un (1 :1). Le système trouvera ou ne trouvera pas d'appariement entre les deux [5] [6].

La vérification est communément employée dans des applications de contrôle d'accès et de paiement par authentification. La biométrie offre beaucoup plus d'avantages que les méthodes existantes d'authentification personnelle telles que les clefs, les numéros d'identification (ID), les mots de passe et les cartes magnétiques. En effet elle fournit encore plus de sûreté et de convenance ce qui engendre d'énormes avantages économiques et elle comble les grandes failles de sécurité des mots de passe [6].

1.2.3 Les exigences de la Biométrie

Dans la théorie tout trait de l'être humain (physiologique ou comportemental) peut être employé comme caractéristique biométrique pour la reconnaissance des personnes aussi longtemps qu'il répond aux conditions suivantes [5] :

- **Universalité** : ceci signifie que chaque personne devrait avoir le trait.
- **Unicité** : ceci indique que deux personnes quelconques devraient être suffisamment différentes en termes de leurs traits biométriques.
- **Permanence** : ceci signifie que le trait devrait être suffisamment invariable dans le temps.

- **Collecte** : ceci indique que le trait peut être mesuré quantitativement.

Dans la pratique, il y a d'autres problèmes importants qui doivent être pris en considération, incluant :

- **Performance** : qui se rapporte à l'exactitude et la vitesse d'identification, les ressources requises pour obtenir l'exactitude d'identification et la vitesse désirées, aussi bien que les facteurs opérationnels et environnementaux qui les affectent.
- **Acceptabilité** : ce qui indique le point auquel les gens sont disposés à accepter l'utilisation d'un trait biométrique particulier dans leurs vies quotidiennes.
- **Vulnérabilité** (Mise en échec) : ce qui reflète à quel point il est facile de tromper le système par des méthodes frauduleuses.

1.3. Le mode de fonctionnement d'un système biométrique

Généralement les systèmes biométriques fonctionnent selon trois modes principaux (Fig1.1) : l'enrôlement, l'authentification et l'identification

- **Enrôlement** : C'est la première phase de tout système biométrique, il s'agit de l'étape pendant laquelle un utilisateur est enregistré dans le système pour la première fois et où une ou plusieurs modalités biométriques sont capturées et enregistrées dans une base de données. [7]
- **Authentification** : Il permet de contrôler l'identité d'un individu en effectuant une comparaison entre les données biométriques acquises avec le modèle biométrique propre stocké dans la base de données.
- **Identification** : Il assure la reconnaissance des individus. Le système biométrique va comparer l'identité d'une personne inconnue avec les modèles de toutes les personnes enregistrées dans la base de données.

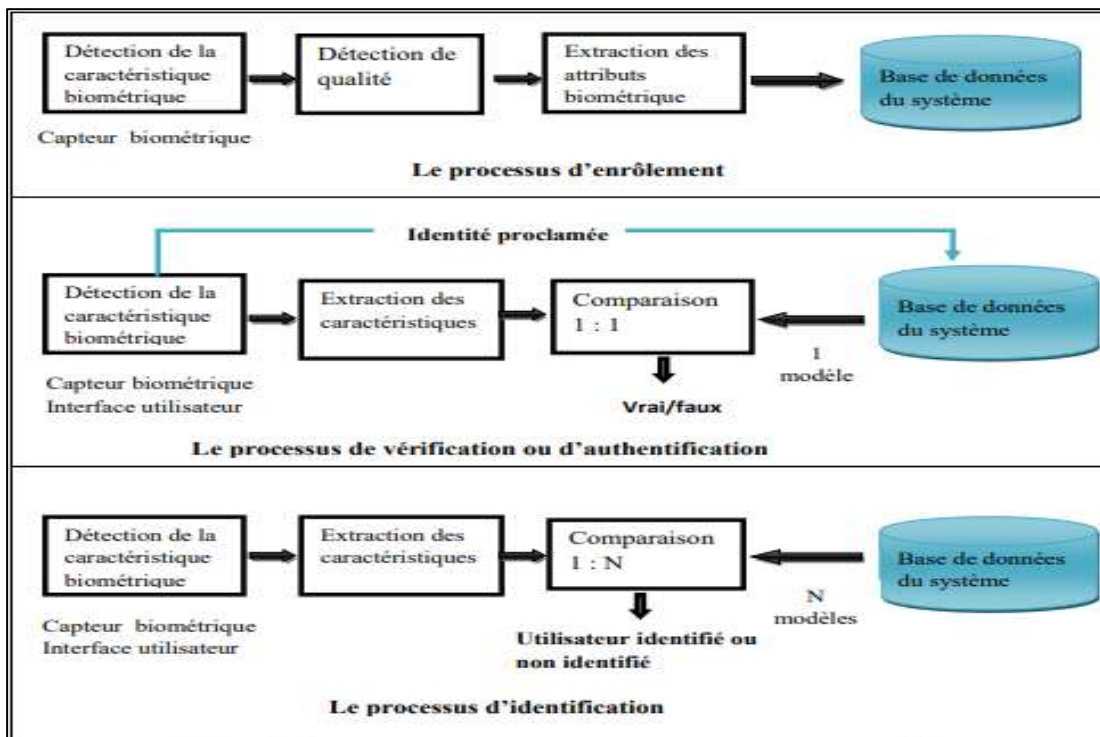


Fig 1.1 : Les modes de fonctionnement d'un système biométrique [2].

1.3.1 Structure interne d'un système biométrique

L'architecture d'un système biométrique peut être représentée par quatre modules principaux. Pour avoir plus de détails, nous allons expliquer le fonctionnement de chaque module comme suit (Fig. 1.2) :

- **Module de capture biométrique :** Ce module est responsable de l'acquisition des données biométriques d'un individu (cela peut être un appareil photo, un lecteur d'empreintes digitales, une caméra de sécurité... etc).
- **Module d'extraction de caractéristiques :** Il prend en entrée les données biométriques acquises par le module de capture et extrait seulement l'information pertinente afin de former une nouvelle représentation des données. Idéalement, cette nouvelle représentation est censée être unique pour chaque personne et relativement invariante aux variations intra-classe.
- **Module de correspondance :** Dans ce cas, le module compare l'ensemble des caractéristiques extraites avec le modèle enregistré dans la base de données du système et détermine le degré de similitude (ou de divergence) entre les deux.

- **Module de décision** : Ce module, vérifie l'identité affirmée par un utilisateur ou détermine l'identité d'une personne, basée sur le degré de similitude entre les caractéristiques extraites et le(s) modèle(s) stocké(s).

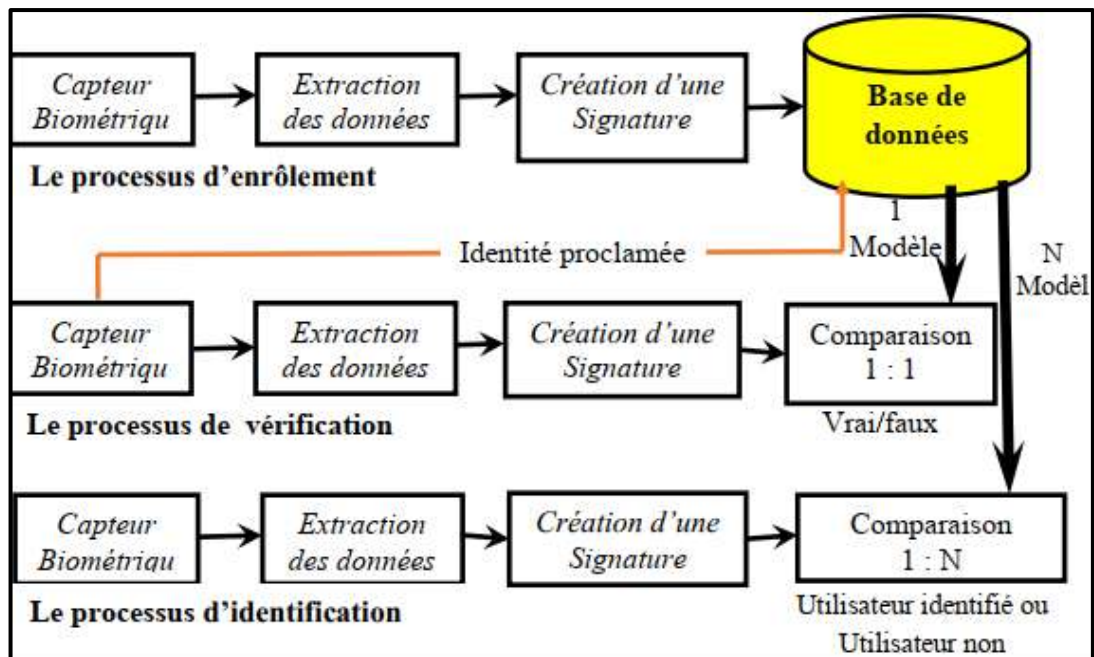


Fig 1.2 : Principaux modules d'un système biométrique et différents modes de fonctionnement [5]

1.4. Les différentes modalités biométriques

La biométrie est la reconnaissance automatique d'une personne en utilisant des traits distinctifs. Une autre définition de la biométrie est "toutes caractéristiques physiques ou traits personnels automatiquement mesurables, robustes et distinctives qui peuvent être utilisées pour identifier un individu ou pour vérifier l'identité prétendue d'un individu ». Les différentes modalités biométriques peuvent être classées selon les catégories suivantes :

- **Biométriques morphologiques (physiologiques)** : Cette catégorie utilise une partie du corps humain comme le visage, l'empreinte digitale, l'iris, la rétine et la géométrie de la main ...etc.
- **Biométriques comportementales** : Ce type est basé sur l'analyse de comportements d'une personne comme la dynamique de la frappe sur un clavier, la signature et la manière de marcher.
- **Biométriques morpho-comportementales** : Ce type combine entre les deux dernières catégories et contient seulement la voix. Puisque la voix est liée à la morphologie des cordes vocales mais aussi elle peut être changée par la personne selon ses états émotionnels, son âge.

- **Biométries biologiques** : Cette catégorie s'appuie sur l'ADN, le sang et la salive.

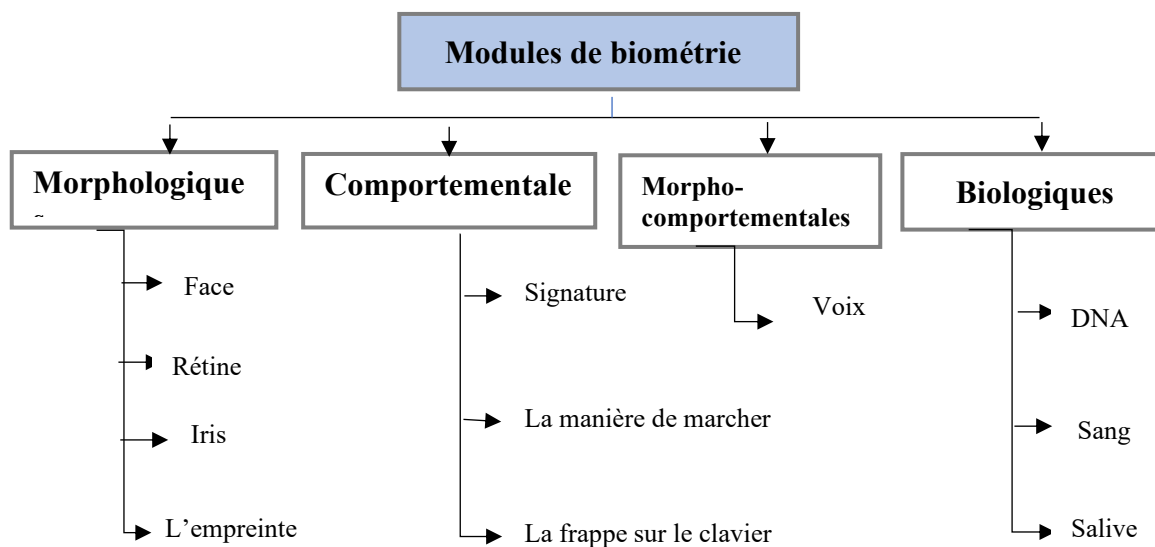


Fig 1.3 : Les différentes modalités biométriques

1.5. Les technologies biométriques

1.5.1. L'empreinte digitale

Une empreinte digitale définie comme une trace produite par les sillons de la peau des doigts ,elle est composée de lignes localement parallèles présentes par des points singuliers « minuties » et constitue un motif unique, peut être classée en deux catégories : les techniques d'appariement sur la détection locale des minuties et celles basées sur la corrélation , les empreintes peuvent être utilisées sur des micro-ordinateurs ou téléphones portables pour la sécurité d'utilisation , Les empreintes digitales sont un outil efficace d'identification parce qu'elles sont uniques pour chaque individu.



Fig 1.4 : Les technologies biométriques de l'empreinte digitale

1.5.2. Géométrie de la main

C'est une technique biométrique basée sur la forme de la main, il s'agit de mesurer plusieurs caractéristiques de la main telle que la longueur, la largeur, l'épaisseur et la surface de la main. Cette technique peut-être utilise afin de contrôles des présences et le temps de retard des employés sur un lieu de travail, donc une méthode de contrôle d'accès et d'identification des personnes. L'acquisition d'une image de la main est obtenue principalement de l'imagerie infrarouge à l'aide d'un scanner spécialisé.

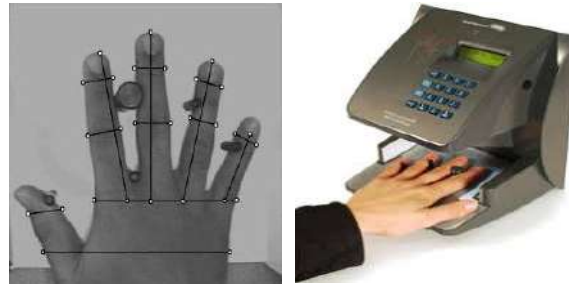


Fig 1.5 : Dispositif de reconnaissance par géométrie de la main [8].

1.5.3. Le visage

C'est une technique assure le contrôle d'accès et la surveillance, il consiste à capter le visage d'une personne à l'aide de caméras et d'extraire certaines parties de visage telles que la forme de visage, le nez, la bouche, les yeux et les joues, ensuite le système analyse et traite géométriquement les données extraites (distance entre différents points, positions, formes...).



Fig 1.6 : Les technologies biométriques de reconnaissance de visage.

1.5.4. L'iris

L'iris est la région, sous forme d'anneau, située entre la pupille et le blanc de l'œil, il a une structure extraordinaire et offre de nombreuses caractéristiques de texture qui sont uniques pour chaque individu. [8]

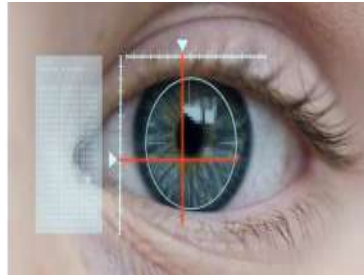


Fig 1.7 : Système biométrique basée sur l'iris [8].

1.5.5. La rétine

C'est la couche sensorielle de l'œil qui permet la vision. La détermination des caractéristiques de la rétine consiste à extraire la distribution géographique des vaisseaux sanguins [24]. Cette technologie est adaptée aux applications de haute sécurité (salles de coffres forts, sites militaires... etc.) [2].

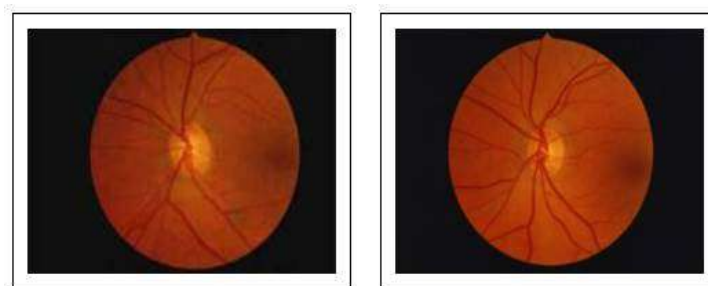


Fig 1.8 : Exemple deux rétines distinct.

1.5.6. La dynamique de frappe au clavier

C'est une propriété comportementale propre à chaque individu. Il s'agit d'une graphologie des temps modernes car on écrit plus souvent avec un clavier qu'avec un stylo. Les éléments principaux analysés par cette modalité sont : la vitesse de frappe, la suite de lettre, le temps de frappe, les pauses, etc [2].

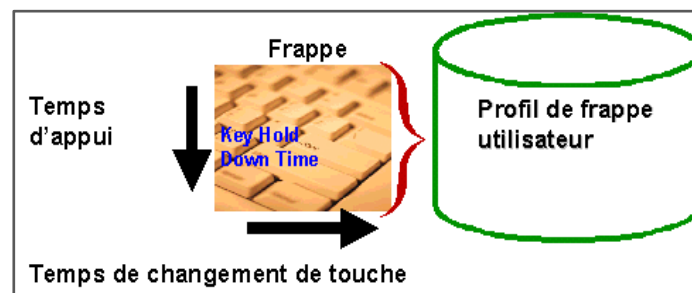


Fig 1.9 : Système biométrique base sur la frappe au clavier.

1.5.7. La dynamique du tracé de la signature

Chaque personne est caractérisée par sa façon d'écriture unique et à partir de sa signature on peut définir un modèle qui pourra être utilisé pour l'identification des personnes [8] Les systèmes de reconnaissance de l'écriture consistent à analyser les caractéristiques spécifiques d'une signature comme la vitesse, la pression sur le crayon, le mouvement, les points et les intervalles de temps où le crayon est levé [2].



Fig 1.10 : Système Biométrique Base Sur La Frappe Au Clavier [2].

1.5.8. La voix

La reconnaissance vocale consiste à identifier les personnes en utilisant leurs caractéristiques vocales. L'empreinte vocale contient plusieurs éléments tels que : l'accent, le rythme et inflexion et des facteurs physiques (la forme, la taille de la personne et les cordes vocales). [9]

1.5.9. Analyse de l'ADN

L'ADN (acide désoxyribonucléase) présent dans les cellules du corps, il est spécifique pour chaque individu. C'est une analyse du patrimoine génétique qui permet une identification à partir des cheveux, fragment de peau, d'une trace de sang et d'une goutte de salive. De plus l'ADN contient plus d'information sur l'identité des personnes.

1.6. Évaluation des performances des systèmes biométriques

Chaque caractéristique (ou modalité) biométrique a ses forces et ses faiblesses, et le choix dépend de l'application visée. On ne s'attend à ce qu'aucune modalité biométrique ne réponde efficacement aux exigences de toutes les applications. En d'autres termes, aucun système biométrique n'est "optimal".

Faire correspondre un système biométrique spécifique à une application dépend du mode opérationnel de l'application et des caractéristiques biométriques choisies. Plusieurs études ont été menées afin d'évaluer les performances de ces systèmes [1]. La société

américaine International Biometric Group [IBG] a par exemple effectué une étude basée sur quatre critères d'évaluation [10].

- **Intrusive** : ce critère permet de classifier les systèmes biométriques en fonction de l'existence d'un contact direct entre le capteur utilisé et l'individu à reconnaître. La reconnaissance faciale est une technique « non intrusive », car il n'existe aucun contact entre le capteur (la caméra) et le sujet, elle est bien acceptée par les utilisateurs à l'inverse d'autres techniques « intrusives » comme l'iris où un contact direct est nécessaire entre le capteur et l'œil.
- **Fiabilité** : dépend de la qualité de l'environnement (éclairage par exemple) dans lequel l'utilisateur se trouve. Ce critère influe sur la reconnaissance de l'utilisateur par le système. Nous verrons ce point en détail dans les sections suivantes.
- **Coût** : doit être modéré à cet égard nous pouvons dire que la reconnaissance faciale ne nécessite pas une technologie coûteuse. En effet, la plupart des systèmes fonctionnent en utilisant un appareil à photo numérique de qualité standard.
- **Effort** : requis par l'utilisateur lors de la saisie de mesures biométriques, et qui doit être réduit le plus possible. La reconnaissance faciale est la technique biométrique la plus facile à utiliser car elle est non contraignante.

Une comparaison (avantages / inconvénients) des principales technologies biométriques (tableau 1.1) en se basant sur la facilité ou l'ergonomie d'utilisation, la vulnérabilité aux attaques et aux contournements et la fiabilité relative à la précision et à l'efficacité de la reconnaissance.

La technologie	Avantages	Inconvénients
L'empreinte digitale	-La technologie la plus connue et éprouvée par le public. -Faible coût, -Traitement rapide -Bon compromis entre le taux de faux rejet et le taux de fausse acceptation	- Difficulté de lecture : sensibilité aux altérations pouvant survenir au cours de la vie (égratignure, cicatrice, vieillissement ou autres) et à certaines variations (humidité, saleté...) - Possibilité d'attaque

Géométrie de la main	<ul style="list-style-type: none"> -Bonne acceptation des usagés, très simple à utiliser. -Le résultat est indépendant de l'humidité et de l'état de propreté des doigts. 	<ul style="list-style-type: none"> -Risque de fausse acceptation pour des jumeaux ou des membres d'une même famille - La forme de la main ou des doigts se modifie avec le vieillissement, ce qui nuit à la mesure à long terme.
Le visage	<ul style="list-style-type: none"> - utilisation facile. -Faible coût -aucun contact physique 	<ul style="list-style-type: none"> - La reconnaissance des visages ne fonctionne pas bien incluent l'éclairage pauvre, les lunettes de soleil...etc. - Problème de distinguer les vrais jumeaux.
L'iris	<ul style="list-style-type: none"> -Les iris sont uniques et différant même pour les vrais jumeaux. -Grande quantité d'information contenue dans l'iris -Iris très difficilement falsifiable -L'iris ne varie pas 	<ul style="list-style-type: none"> -Aspect psychologiquement invasif de la méthode. -Une photo ou une lentille de contact reproduisant l'image de l'iris peut affecter la fiabilité.
La rétine	<ul style="list-style-type: none"> -Haute sécurité - La rétine est différente chez les vrais jumeaux. -La rétine est stable durant la vie d'un individu -Très efficace 	<ul style="list-style-type: none"> - Système intrusif, il faut placer l'œil près du capteur -Mauvaise acceptation du public -Coût plus important que d'autres technologies
La dynamique de frappe au clavier	<ul style="list-style-type: none"> -Identification d'une personne à distance à partir de son ordinateur. - Mise en œuvre rapide pour un grand nombre d'utilisation. - Non intrusif, geste naturel. 	<ul style="list-style-type: none"> Sensibilité à la différence entre les claviers -L'état de santé et la fatigue peuvent altérer la façon de frapper les touches
La signature	<ul style="list-style-type: none"> - Facile à utiliser. -Très acceptée par les usagers 	<ul style="list-style-type: none"> - Besoin d'une tablette graphique. - Non utilisable pour le contrôle d'accès en extérieur
La voix	<ul style="list-style-type: none"> - Technologie biométrique facile à mettre en œuvre. - Permet de sécuriser une conversation téléphonique 	<ul style="list-style-type: none"> - Il est très facile d'enregistrer ou de reproduire la voix. - Nécessite une excellente qualité audio. -Sensible aux bruits ambiants
L'analyse de l'ADN	<ul style="list-style-type: none"> -Une Très grande précision. - Il impossible que le système a fait des erreurs, Il est standardisé 	<ul style="list-style-type: none"> - Coûteux. - Pour avoir les résultats, il faut attendre une longue durée

Tab 1.1 : Les avantages et inconvénients des différentes technologies biométriques.

Donc, le choix de la modalité biométrique dépend essentiellement des exigences et des besoins de chaque application.

1.6.1. Mesure de la performance d'un système biométrique

Pour évaluer les performances d'un système biométrique, il est nécessaire de mesurer les deux taux d'erreurs suivant [11] :

-Le FRR (False Rejet Rate ou Taux de Faux Rejet) : indique la probabilité qu'un utilisateur connu soit rejeté par le système biométrique, ce taux définit en partie le confort d'utilisation du système biométrique

$$FRR = \frac{\text{nombre des clients rejetés}(FR)}{\text{nombre total d'accès de clients}}$$

Telle que FR Le faux rejet correspond au cas où le système rejette un client légitime.

-Le FAR (False Acceptation Rate ou Taux de Fausse Acceptation) : indique la probabilité qu'un utilisateur inconnu soit identifié étant un utilisateur connu, ce taux définit la sécurité de système biométrique.

$$FAR = \frac{\text{nombre des imposteurs acceptés}(FA)}{\text{nombre total d'accès imposteurs}}$$

Le FAR correspond au cas où le système accepte un individu qui a proclamé une identité qui n'est pas la sienne, Le paramétrage d'un système consiste à trouver le bon équilibre entre ces deux taux, le FAR augmentant lorsque le FRR diminue, et inversement. Un contrôle d'accès très sécurisé aura un FAR très bas, pour garantir qu'aucune personne non autorisée n'accède au site, mais, en contrepartie le FRR sera élevé, ce qui signifie que des utilisateurs valides se verront refuser l'accès.

Les autres mesures de performance sont les temps d'encodage de l'empreinte et de mise en correspondance. Ces valeurs peuvent varier considérablement d'une application à une autre. Un troisième paramètre FER (False Equal Rate) mesure le taux d'échec à l'enrôlement. Il traduit la probabilité d'absence d'une caractéristique biométrique pour un individu dans une population, donne un point sur lequel le TFA est égal au TRF [12]. La figure 1.11 illustre le FRR et le FAR à partir de distributions des scores authentiques et l'imposteurs.

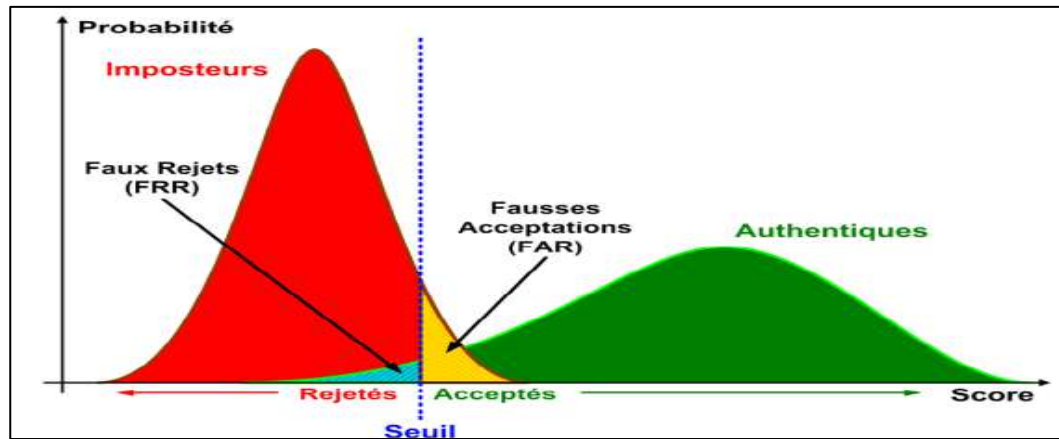


Fig 1.11 : Illustration du FRR et du FAR.

1.6.2. Les applications des systèmes biométriques

Aujourd'hui, il est nécessaire de connaître l'identité des personnes pour contrôler d'accès à des sites sécurisés donc il existe des techniques biométriques sont appliquées dans plusieurs domaines et leur champ d'application couvre potentiellement tous les domaines de la sécurité. Généralement, les applications peuvent être divisées en trois groupes principaux [12] :

- **Applications commerciales** : telles que l'accès au réseau informatique, la sécurité de données électroniques, le commerce électronique, l'accès d'internet, la carte de crédit, le contrôle d'accès physique, le téléphone portable, la gestion des registres médicaux, l'étude de distances ...etc.
- **Applications de gouvernement** : telles que la carte nationale d'identifications, le permis de conduite, la sécurité sociale, le contrôle du passeport...etc.
- **Applications juridiques** : telles que l'identification de cadavre, la recherche criminelle, l'identification de terroriste, les enfants disparus...etc.

1.7. La reconnaissance faciale

La reconnaissance faciale est récemment devenue un domaine de recherche très actif, en raison de l'intérêt accru pour les systèmes de sécurité biométriques en général, mais aussi le progrès récent qui a pris l'état de l'art au-delà des tentatives initiales d'utiliser la comparaison directe de l'image.

Le premier système de reconnaissance du visage a été développé depuis le début de 1970 et en raison de la limitation du calcul, le système ne peut pas être satisfait de l'exigence des utilisateurs, qui est d'identifier la photographie de passeport en temps réel. Au début des années

1990, les techniques sont proposées axées sur la reconnaissance faciale et augmentent le besoin de détection du visage. La reconnaissance faciale a attiré beaucoup d'attention et ses recherches se sont rapidement développées non seulement par les ingénieurs, mais aussi par les neuroscientifiques, puisqu'il y a nombreuses applications potentielles dans la vision par ordinateur et le système de contrôle d'accès automatique.

D'autre part, la détection du visage est une partie importante dans un système de reconnaissance du visage qui est considéré comme la première étape de la reconnaissance automatique. Cependant, la détection du visage n'est pas simple processus parce qu'elle a beaucoup de variations de l'apparence de l'image, telles que la variation de pose, l'occlusion, l'état éclairant et l'expression faciale. Le but de la détection du visage est de détecter les visages dans toutes les images ou vidéos. Le visage est notre principal objectif d'attention dans un système de reconnaissance faciale ainsi que dans les rapports sociaux de notre vie quotidienne et joue un rôle majeur dans la transmission de notre identité.

La reconnaissance faciale peut être utiliser avec d'autres technologie biométriques telles que l'iris, les empreintes digitales afin d'améliorer les performances de reconnaissance d'un système (système multimodale). Néanmoins, le cerveau humain a ses limites dans le nombre total de personnes qu'il peut « se rappeler ou se souvenir » exactement. Un avantage clé d'un système informatique est sa capacité à traiter un grand nombre d'images faciales.

1.7.1. Principales difficultés de la reconnaissance du visage

Construire un système qui accomplit la tâche de la reconnaissance automatique du visage représente un sérieux défi. Ce défi est d'autant plus grand lorsque les conditions d'acquisition des images sont très variables. Ces conditions, notamment la puissance des sources de luminosité et l'angle sous lequel le visage est observé, influent considérablement sur l'apparence d'un visage.

Nous détaillons ici les principales difficultés qui ont un impact sur les performances du système de reconnaissance faciale automatique dans des conditions réelles :

- **Changement d'illumination (éclairage) :** Les changements d'éclairage peuvent entraîner des variations considérables de l'apparence dans d'un visage. Deux types d'éclairage permet d'influencer celle-ci : l'éclairage global et l'éclairage local, alors que l'éclairage global affecte tout le visage de manière uniforme, par contre l'éclairage locale entraîne la création d'ombre et des zones éclairées. Le problème de la reconnaissance faciale sur les changements

d'éclairage est largement reconnu et étant difficile pour les humains ainsi pour les algorithmes. Les difficultés posées par les conditions d'éclairage variables demeurent donc un défi important pour les systèmes de reconnaissance faciale automatique. Il est constaté que la différence entre deux images de la même personne prises sous éclairage variable est plus grande par rapport la différence entre les images de deux personnes différentes sous un même éclairage. La variation de l'éclairage change radicalement l'aspect du visage comme le montre la figure (1.12).



Fig 1.12 : Illustration de l'impact des différents éclairages sur les images des visages

- **Variation de pose :** dans une image est également un sujet de préoccupation dans la reconnaissance faciale. La pose d'un visage change avec l'angle de vue de l'observateur et la rotation en position tête. Ces changements de posture posent un sérieux problème pour l'identification de l'image d'entrée. Un système de reconnaissance faciale peut tolérer les cas avec de petits angles de rotation, mais cela devient un défi lorsque l'angle de rotation augmente et que l'image disponible dans la base de données peut n'avoir que la vue frontale du visage qui peut différer dans la pose avec l'image d'entrée et qui induit le système en erreur entraîne une identification défectueuse ou aucune reconnaissance.



Fig 1.13 : Exemples de variation d'éclairage.

- **Expression faciale :** L'apparence d'un visage varie grandement en présence d'expressions faciales. Les éléments faciaux tels que la bouche ou les yeux peuvent alors subir des déformations importantes, pouvant faire échouer un système de reconnaissance faciale fondée par exemple sur des points d'intérêt (ceux-ci pouvant ainsi subir d'importantes translations). La bouche est en général l'élément facial qui varie le plus, mais l'aspect des sourcils peut par exemple être grandement modifié.

- **La résolution :** Un problème de faible résolution survient dans un système de reconnaissance faciale lorsque la résolution de l'image faciale à reconnaître est inférieure à 16x16. Ce problème se produit dans de nombreuses applications de surveillance, telles que les applications de caméra autonome à petite échelle dans les supermarchés et les banques et dans les rues publiques, etc. lorsque les images prises à partir d'une caméra de surveillance se composent généralement d'une très petite surface faciale et ne peuvent fournir une résolution suffisante de la face pour la reconnaissance. Comme le visage de la personne n'est pas proche de la caméra, la région du visage sera inférieure. Une telle image du visage à faible résolution est constituée d'informations très limitées car la plupart des détails sont perdus. Cela peut réduire considérablement le taux de reconnaissance.
- **Présence ou absence des composants structurelles :** La présence des composants structurels telle que la barbe, la moustache et les lunettes. Il peut cacher énormément les caractéristiques de base d'un visage. Ce problème peut affecter le processus de classification du système de reconnaissance.
- **Les vrais jumeaux :** Les vrais jumeaux peuvent tromper les gens qui ne les connaissent pas. Pour un système de reconnaissance, il est probable que la tâche de vérification ne pourra jamais détecter les différences très proches qui existent entre les vrais jumeaux.

1.7.2. Motivation : la reconnaissance faciale pour la prédiction de genre et l'âge

De nos jours, la reconnaissance faciale est une technologie de plus en plus répandue, elle combine les traits biométriques et les techniques de l'intelligence artificielle pour comparer et analyser le visage d'une personne afin de l'identifier. Cette technologie est présente plusieurs avantages par rapport aux autres technologies dont on peut citer [5] :

- ✓ La disponibilité des équipements d'acquisition, leur simplicité et leurs coûts faibles.
- ✓ Passivité du système : un système de reconnaissance de visages ne nécessite aucune coopération de l'individu, du genre : mettre le doigt ou la main sur un dispositif spécifique ou parler dans un microphone. En effet, la personne n'a qu'à rester ou marcher devant une caméra pour qu'elle puisse être identifiée par le système.

La reconnaissance du visage est de plus appliquée à des fins de surveillance et de sécurité, notamment par les gouvernements et les autorités qui tendent à l'incorporer à des systèmes de vidéosurveillance. D'autre part, la reconnaissance faciale est utilisée sur les services à vendre pour confirmer le paiement des achats en ligne. On appelle cette nouvelle méthode

d'authentification le « selfie pay », On peut trouver aussi dans les grands aéroports, utilisant les panneaux publicitaires intelligents à fin d'identifier le genre et l'âge des passants et de le proposer des publicités ciblées. En outre, un système d'estimation d'âge peut être employé pour contrôler les droits d'accès des personnes, par exemple les parents pourraient permettre à leurs enfants de regarder uniquement certains programmes de télévision ou sites web, alors que la télévision ou l'ordinateur reconnaissent automatiquement les personnes qui se trouvent devant elle.

La prédiction de genre et l'âge est un cas particulier de reconnaissance faciale qui est un problème bien étudié dans la vision par ordinateur. De ce sens, notre objectif dans ce travail est de créer un modèle performant pour la prédiction du genre, l'estimation d'âge et la reconnaissance du visage basée sur l'utilisation des réseaux CNN pour l'extraction des caractéristique pertinentes et la classification automatique des images faciales.

1.8. Conclusion

Dans ce chapitre, nous avons présenté les différentes technologies utilisées dans les systèmes biométriques pour l'identification de personnes, puis nous avons décrit le mode de fonctionnement d'un système biométrique, ainsi que les technologies biométriques avec leurs avantages et inconvénients. Enfin nous avons cité les critères de performance d'un système biométrique et nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance automatique de visages. Les techniques utilisées aux différentes étapes de reconnaissance des formes et la reconnaissance de visage en particulier seront détaillées dans le chapitre suivant.

CHAPITRE II

ETAT DE L'ART DE LA RECONNAISSANCE DE VISAGE ET LA PREDICTION DU GENRE ET L'AGE

2. Chapitre II

2.1. Introduction

Au cours de ces dernières années et avec la nécessité croissante d'automatiser les systèmes de reconnaissance et de surveillance, les recherches sur le traitement et l'analyse numérique de visages humains (y compris la détection et la reconnaissance de visages, la prédiction du genre et de l'âge) ont attiré une attention particulière dans les communautés de la vision par ordinateur et de la reconnaissance de forme. La reconnaissance c'est une étape qui suit l'étape de détection de visage afin d'identifier le genre et l'âge des personnes. Plusieurs approches ont été développées afin d'améliorer les systèmes de la reconnaissance des formes qui sont subdivisées en trois grandes approches : globales, locales et hybrides.

Dans ce deuxième chapitre, nous présentons un état de l'art sur les méthodes les plus populaires utilisées dans le domaine de la reconnaissance faciale et la prédiction du genre et de l'âge. Ensuite, nous décrivons quelques travaux développés pour ces domaines.

2.2. La reconnaissance de visage

La reconnaissance de visage est un domaine de la vision par ordinateur consistant à identifier automatiquement des personnes à partir de son visage [13], C'est un sujet particulièrement étudié en vision par ordinateur, avec de très nombreuses publications et brevets, et des conférences spécialisées. Mais elle définit un nombre considérable de problèmes complexes et qui ne peuvent être évités, et qui affectent négativement sur le résultat final. Andrew W. Senior et al [14] a catégorisé les problèmes de la reconnaissance automatique du visage en trois catégories, ce sont : Changements physique : les expressions faciales, vieillissement, Changements de géométrie d'acquisition : changement d'échelle, localisation et d'orientations, Changements d'éclairage : la variation de l'éclairage, les variations de la caméra.

2.3. Les techniques de reconnaissance faciale

Dans cette partie, On va citer en revue brièvement quelques techniques les plus importantes et les plus utilisées courantes dans ce domaine. Selon Tana et al [15] on peut diviser ou classer les approches de reconnaissance faciale en trois grandes familles qui sont : les approches globales, les approches locales ainsi que les approches hybrides. Les principales méthodes de ces approches sont décrites comme suit.

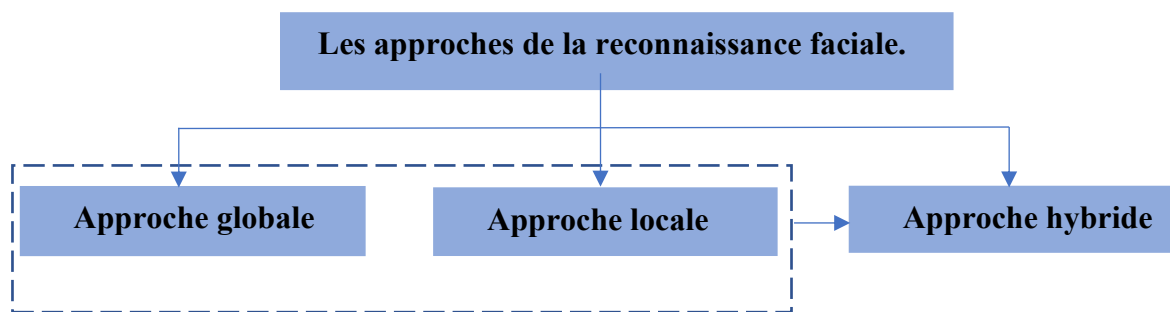


Fig 2.1 : Classification des approches principales utilisées en reconnaissance faciale.

2.3.1. Approche globale (holistique)

Le principe de cette approche selon O'Toole et al [16] est de représenter l'image du visage par un seul vecteur de grande dimension $n \times m$, en concaténant les niveaux de gris de tous les pixels du visage [17], il n'est pas nécessaire de repérer certains points caractéristiques du visage locales comme les yeux, la bouche et le nez.

Xiao Guang Lu [18] a distingué deux types de techniques parmi les méthodes globales, les techniques linéaires et les techniques non linéaires

a. Techniques linéaires :

Les techniques linéaires réalisent une projection linéaire des visages (espace dont la dimension est égale à la dimension des images, donc grande) sur un espace de plus faible dimension. Cependant, ces techniques linéaires sont sensibles aux conditions de luminosité notamment, et plus généralement aux variations non convexes. Ainsi, l'utilisation de distances classiques dans l'espace projeté ne permet pas toujours de réaliser une bonne classification entre les classes visages et non visages. Les techniques les plus populaires de l'indentification de visage de cette approche ACP, LDA :

1. Analyse en composants principales (ACP) :

Dite aussi Eigenfaces une méthode très populaire dans le domaine de reconnaissance proposé par Turk et Pentland [19]. Une méthode mathématique utilisée pour simplifier et réduire les dimensions d'un ensemble de donnée pour représenter des images de visage qui peuvent être reconstruite à partir d'un visage standard et un ensemble de points. Il s'agit de trouver l'ensemble des composants principaux d'un visage dans un ensemble d'image, chaque exemple de visage décrit par une combinaison linéaire des vecteurs propres (voir la figure2.2), il consiste à faire trois étapes suivantes :

- Transformation des visages en vecteurs.

- Détermination de la matrice de covariance.
- Détermination des vecteurs propres de la matrice de covariance formée par l'ensemble des images exemple. Chaque élément dans le vecteur correspond à l'intensité lumineuse d'un pixel.

L'ACP est une technique simple et rapide, propose des bons résultats dans les systèmes d'identification, ainsi que la projection optimale pour la reconstruction d'une base de dimension réduite. Les inconvénients de cette technique sont la sensibilité aux problèmes d'éclairage, expression faciale et la pose.

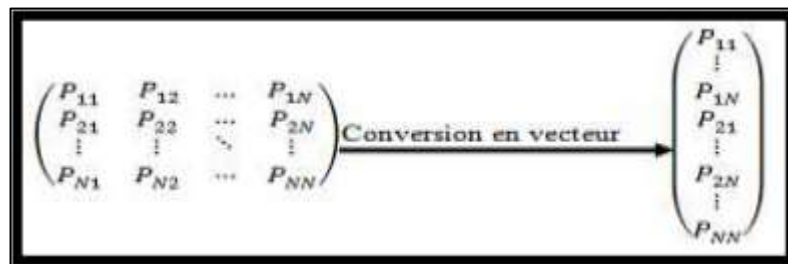


Fig 2.2 : Conversion d'image (matrice) en vecteur colonne

2. Analyse discriminante linéaire (LDA) :

Connu aussi sous le nom « Fisherfaces » Belhumeur et al [20] sont les premiers qui ont introduit cet algorithme en 1997, il effectue une séparation de classes et pour pouvoir l'utiliser il est nécessaire d'organiser une base d'apprentissage d'images en plusieurs classes, une classe par personne et plusieurs images par classe La LDA détermine les directions de projection, et pour cela elle maximise les variations entre les images des individus différents « inter-classe » avec la minimisation des variations entre les images d'un même individu « intra-classe ». Si le nombre d'individus à traiter est plus faible que la résolution de l'image cela rend les performances de LDA faible par rapport à Eigenface [21] pour résoudre ce problème beaucoup d'autre méthode basée sur LDA ont été développée tels que : U_LDA, O_LDA, N_LDA [17].

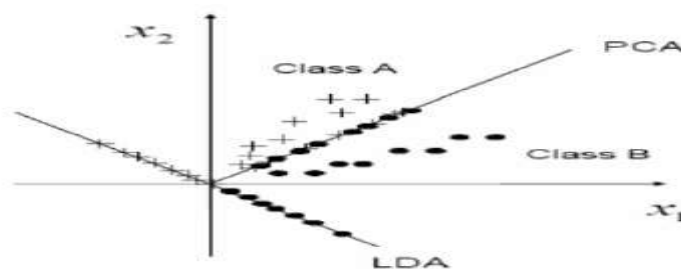


Fig 2.3 : Les projections ACP et LDA d'un ensemble de données [5].

D'autres techniques linéaires ont également été utilisées pour le calcul de vecteurs caractéristiques :

- L'analyse en composantes indépendantes (ICA) dans [22].
- La Factorisation de matrices non négatives (NMF) dans [23] ou [24].
- L'analyse discriminante bilinéaire (BDA) dans [25].
- La technique dite de Vecteurs communs discriminants (DCV) dans [26].

Bien que ces méthodes linéaires soient assez efficaces, elles manquent de précision dès lors que les images de visages subissent des transformations non linéaires. Une simple modification de la luminosité transforme celui-ci de façon non linéaire étant donné la complexité de la forme.

b. Techniques non linéaires :

Des techniques globales non linéaires ont été développées, souvent à partir des techniques linéaires. Ainsi l'Analyse en Composantes Principales à Noyaux (ou Kernel-PCA) [27], [28] et l'Analyse Discriminante Linéaire à Noyaux (ou Kernel-LDA) [29] utilisent la notion mathématique des noyaux pour étendre les techniques linéaires que sont l'ACP et la LDA. D'autres techniques non linéaires ont également été utilisées dans le contexte de la reconnaissance faciale :

- Le MultiDimensional Scaling (MDS) dans [30] ou [31].
- L'Isomap dans [32].
- Les Diffusion Maps dans [33].
- Le Local Linear Embedding (LLE) dans [34] ou [35].
- Les Laplacian EigenMaps dans [36], [37] ou [38].
- Le Hessian LLE dans [30].
- Le Local Tangent Space Analysis (LTSA) dans [39].
- Les approches neuronales dans [40] ou [41] (auto-encodeurs), dans [42] (cartes de Kohonen), et dans [43] (réseaux de neurones convolutionnels).

L'utilisation de ces méthodes de projection de l'espace des images sur l'espace de caractéristiques est non linéaire permet ainsi dans une certaine mesure de réduire la dimension des images de meilleure façon. Cependant, bien que ces méthodes permettent souvent l'amélioration des taux de reconnaissance sur des jeux de tests donnés, elles sont trop flexibles pour être robustes à de nouvelles données, contrairement aux méthodes linéaires.

2.3.2. Approche locale

C'est une approche qui se base sur les caractéristiques locales de visage pour la reconnaissance faciale, tels que : le nez, la bouche et les yeux, dans cette approche contrairement à l'approche globale le visage est représenté par un ensemble de vecteurs de caractéristiques de faible dimension. (L'approche globale utilise un vecteur de très grande dimension) [17].

La plupart du temps, cela implique la détection/extraction de caractéristiques faciales locales. Brunelli et Poggio [44] proposent une technique qui extrait automatiquement un ensemble de 35 caractéristiques géométriques d'une image de visage (voir la figure 2.4 (a)). Ces ensembles de caractéristiques sont ensuite comparés deux à deux via la distance de Mahalanobis pour réaliser la reconnaissance.

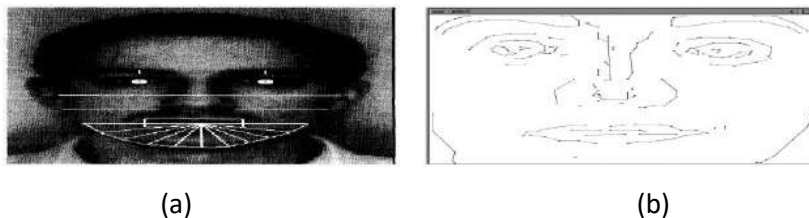


Fig 2.4 : (a) Localisation des caractéristiques géométriques utilisées dans [44].
(b) carte de contours utilisées dans [45].

Une autre approche géométrique a été proposée par Takács [46]. Des cartes binaires de contour sont extraites des images de visage via un filtre de Sobel. La similarité entre deux contours est ensuite calculée en utilisant une variante de la distance de Hausdorff. Cette approche a été étendue par Gao et al [45] qui ont transformé les cartes de contours en cartes de lignes de contours (ou LEM pour Line Edge Maps) contenant des listes de segments (voir la figure 2.4 (b)). La distance utilisée pour mesurer la similarité est la même que celle de Takács.

L'approche de Heisele et al [47] commence par détecter la région contenant le visage dont dix points caractéristiques sont extraits. Les zones autour de ces points sont ensuite extraites (voir la figure 2.5), et concaténées pour former le vecteur caractéristique du visage. La classification est finalement réalisée grâce à l'utilisation d'une machine à vecteurs de support (SVM).

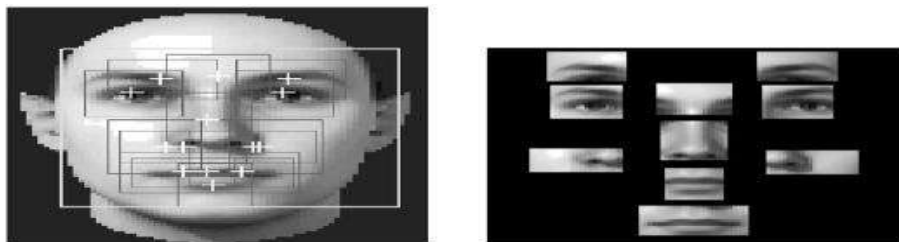


Fig 2.5 : Approche proposée par Heisele et al. [47].

Samaria et al [48] présentent une approche basée sur les chaînes de Markov cachées (HMM pour Hidden Markov Models), le visage est segmenté en sous-bandes partiellement recouvrantes, ces sous-bandes étant ensuite concaténées en un vecteur de grande taille ou compressées par DCT. Puis, pour chaque classe (individu), un HMM est créé modélisant la distribution probabiliste des sous-bandes. Les images de visages sont finalement classées en appliquant l'algorithme de Viterbi pour comparer la séquence des sous-bandes de l'image avec les modèles appris.

Les Modèles Actifs d'Apparence (ou AAM pour Active Appearance Models) sont présentés par Cootes et al dans [49] ils consistent en la création d'un modèle statistique d'un visage (figure 2.6). Le modèle est ensuite déformé pour « coller » au plus près des traits du visage. La reconnaissance est effectuée sur le résidu calculé correspondant à l'erreur de prédiction du modèle.



Fig 2.6 : Modèle Actif d'Apparence [49].

Les LBP (pour Local Binary Patterns) ont également été utilisés pour la reconnaissance faciale, notamment dans [50]. Le visage est subdivisé en sous-régions carrées de taille égale sur lesquelles sont calculées les caractéristiques LBP. Les vecteurs obtenus sont ensuite concaténés pour obtenir le vecteur de caractéristiques final. Des extensions des LBP comme les MB-LBP (pour « Multi-Scale Block Binary Pattern ») ont été proposées et appliquées aux visages par Liao et al [51] (fig2.7).

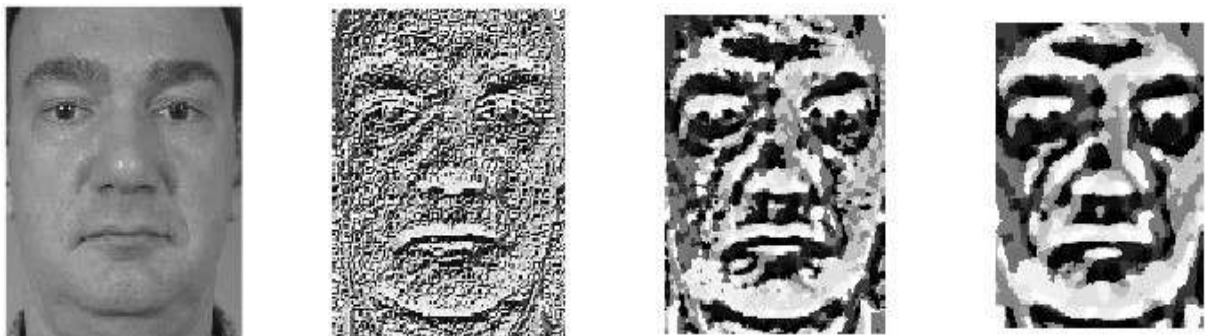


Fig 2.7 : Exemple de la méthode MB_LBP pour un visage [51].

2.3.3. Approche hybride

Les méthodes hybrides résultent de l'association des méthodes locales et des méthodes globales. Elles combinent la détection de caractéristiques locales avec l'extraction de caractéristiques globales. Ces techniques essaient finalement de tirer parti des avantages des deux types de méthodes citées plus haut.

L'approche appelée Analyse en Composantes Locales (LCA pour Local Component Analysis) a été proposée par Penev et Atick [52]. Plusieurs ACP (Analyses en Composantes Principales) sont réalisés pour extraire différentes caractéristiques locales (figure 2.8). Celles-ci sont ensuite combinées et une procédure minimisant l'erreur de reconstruction avec une contrainte parcimonieuse permet de réaliser la reconnaissance.

L'approche dite de l'Elastic Bunch Graph Matching (EBGM) a été proposée par Wiskott et al [53]. Les visages sont représentés par des Face Bunch Graph (FBG), Où chaque nœud du graphe correspond à une certaine caractéristique faciale (l'œil droit ou gauche par exemple), (voir la figure 2.9).

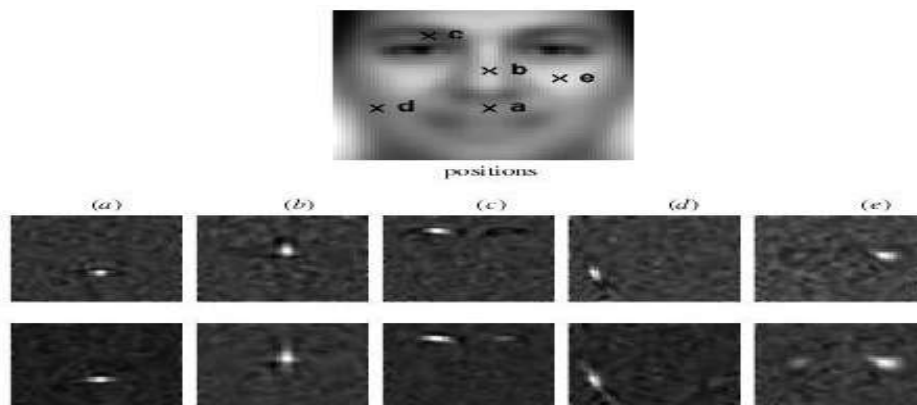


Fig 2.8 : Approche Local Component Analysis [52].

A chaque nœud du graphe est associée l'apparence du voisinage de la caractéristique via un jet. Les jets représentent l'ensemble des 40 coefficients issus de la convolution du voisinage de la caractéristique par un filtre de Gabor spécifique (voir la figure 2.10). Les arêtes du graphe sont pondérées par la distance relative des caractéristiques adjacentes. Une fois que le graphe est créé pour chaque personne de la base d'apprentissage, un algorithme spécifique de mise en correspondance permet d'identifier une personne inconnue. L'algorithme essaie itérativement de faire correspondre le graphe créé à chaque graphe de la base d'apprentissage en minimisant

une fonction de coût prenant en compte à la fois une mesure de similarité géométrique ainsi qu'une mesure de similarité de l'apparence modélisée via les jets.

Perlibakas présente plus récemment l'algorithme LogGabor PCA dans [54]. Une convolution par des ondelettes de Gabor orientées est réalisée autour de certains points caractéristiques du visage. Les vecteurs ainsi créés contiennent à la fois la localisation ainsi que les amplitudes des énergies locales. Une Analyse en Composantes Principales est ensuite réalisée afin de réduire la dimension de ces vecteurs.

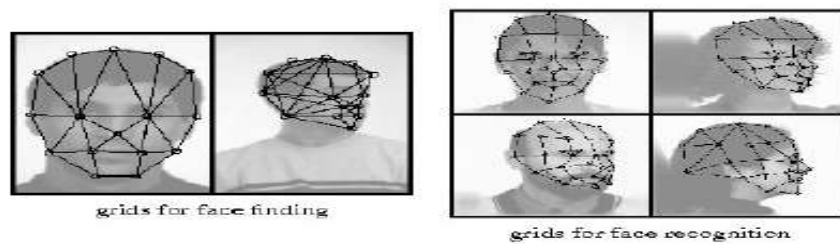


Fig 2.9 : Graphe appliqué aux visages pour l'approche EBG [53].

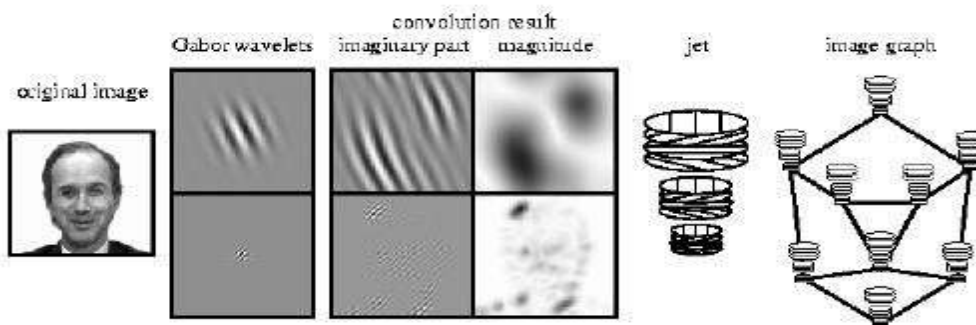


Fig 2.10 : Création des jets pour l'approche EBG [53].

Pentland et al dans [55] présentent l'approche dite des espaces propres modulaires (Modular Eigenspaces) Cette technique réalise une Analyse en Composantes Principales et une classification sur des régions distinctes du visage, comme les yeux, le nez, la bouche ou encore le visage entier. La zone de la bouche subit de grosses déformations dues aux expressions faciales, ainsi l'ajout de cette région au processus entier fait décroître les taux de reconnaissance.

L'approche proposée par Cootes et al dans [56] met en œuvre les Modèles Actifs d'Apparence. Cette méthode modélise indépendamment la forme et la texture d'un visage en appliquant une Analyse en Composantes Principales. Les vecteurs obtenus pour la forme et la texture sont ensuite utilisés pour la reconnaissance. Un nouveau visage qui doit être identifié est adapté au modèle par un processus d'optimisation itératif. Les paramètres de forme et de

texture obtenus sont ensuite comparés à ceux de la base. Latinis et al dans [57] appliquent cette méthode pour la première fois à la reconnaissance de visages. Edwards et al dans [58] proposent des méthodes basées sur les AAMs pour la reconnaissance faciale.

2.4. Récapitulatif des principales approches de reconnaissance faciale

Nous avons présenté les principales approches utilisées dans la littérature pour la reconnaissance faciale. Les méthodes peuvent principalement se classer en deux catégories : les méthodes globales et les méthodes locales (les méthodes hybrides essayant de faire un lien entre ces types d'approches), en effet les caractéristiques locales et globales sont complètement différentes, chacune de ces méthodes à ses inconvénients, mais l'une peut être complémentaire de l'autre pour le but d'améliorer la classification [59].

Approche	Avantage	Inconvénient
Globale	<ul style="list-style-type: none"> -Le problème de la reconnaissance faciale automatique est transformé en un problème d'analyse de sous-espaces de visages. Pour lequel de nombreuses Méthodes statistiques existent. -Les méthodes globales sont souvent applicables à des images basses résolutions ou de mauvaises qualités. -Méthode simple et facile à appliquer. 	<ul style="list-style-type: none"> -Il est nécessaire de disposer de suffisamment de données représentatives des visages. -Il n'y a pas d'a priori sur le physique d'un visage. - Ces méthodes ne sont robustes qu'à des variations limitées (pose, illumination, expression).
Local	<ul style="list-style-type: none"> -Le modèle créé possède des relations intrinsèques bien définies avec les visages réels. -Les modèles créés peuvent prendre en compte explicitement les variations telles que la pose, l'illumination ou les expressions. La reconnaissance est plus efficace dans le cas de fortes variations. 	<ul style="list-style-type: none"> -La construction du modèle, reposant souvent sur la détection de points caractéristiques faciaux, peut être laborieuse. -L'extraction des points caractéristiques peut être difficile dans le cas de variations de pose, d'illumination, d'occlusion . . . -Les images doivent être relativement de bonne qualité ou

	-La connaissance a priori sur les visages peut être intégrée aux modèles afin d'améliorer leur efficacité.	être de résolution suffisante afin de pouvoir extraire les points caractéristiques.
--	--	---

Tab 2.1 : Les avantages et les limitations des approches de reconnaissance faciale.

Facteur de variation	Caractéristiques locales	Caractéristiques globale
Illumination	Très sensible	Sensible
Expression	Pas sensible	Sensible
Pose	Sensible	Très sensible
Bruit	Très sensible	Très sensible

Tab 2.2 : Comparaison des approches basées sur les caractéristiques locales et globales en fonction des critères de variation [60].

2.5. Quelques travaux sur les approches de la reconnaissance

Après avoir cité les trois approches de la reconnaissance (locale, globale, hybride) , plusieurs techniques de reconnaissance de visage ont est développées , pour les méthodes locale ,on trouve l'analyse des composants principale PCA , la technique des faces propres (eigenfaces) qui emploi un classificateur du voisin le plus proche , la technique des faces Fisher qui basé sur l'analyse linéaire discriminante ,aussi les méthodes SVM qui utilisent comme un classificateur ,la plupart des méthodes précédentes de cette approche utilisent la largeur de la tête, les distances et les angles entre les coins des yeux, les extremis...etc. Ainsi l'approche globale contient sur plusieurs méthodes, HMM utilisent les bandes de pixels qui couvrent le front, les yeux, le nez, la bouche, et le menton et la méthode géométrique pure. Un des systèmes les plus réussis dans cette catégorie les réseaux de neurones de convolution qui ont été développe par Lawrence et al.

Dans l'approche des méthodes hybrides, on passera en revue brièvement la méthode des faces propres modulaires, une représentation hybride fondé sur PCA et l'analyse des traits locaux et d'autres méthodes basées sur les modèles d'apparence et sur les composantes.

Le tableau (2.3) au-dessous récapitule la catégorisation des techniques de reconnaissance de visage à partir d'images.

Approche	Travail représentatif	
Globale		
ACP	L. Sirovich, M. Kirby	[61]
Faces propres	Craw et Cameron	[62]
	Kirby et sirovich	[61]
	Turk et Pentland	[19]
Faces propres probabilistes	Moghaddam et pentland	[63]
SVM	Phillips	[64]
LDA	Etemand et chellapa	[65]
LDA	Zhao et AL	[66]
fisherfaces	Wang et AL	[67]
Locale		
Méthode géométrique pure	Kanade	[68]
	Kelly	[69]
	Cox et autre	[70]
	Manjunath et Al	[71]
Architecture à liens dynamique	Okada et AL	[72]
	Wiskott et Al	[73]
Modèles cachés Markov	Nefian et Hayes	[74]
	Samaria	[75]
	Samaria et Young	[76]
Réseaux de neurones de convolution	Lawrence et AL	[77]
Hybride		
Faces propres modulaires	Pentland et AL	[78]
A formes normalisées AAM	Lanitis et Al	[79]
LFA Hyride	Penev et Atck	[80]
Basée sur les composantes	Huang et Al	[81]

Tab 2.3 : Quelques travaux de recherches sur la reconnaissance

2.6. La prédiction du genre et de l'âge

L'intérêt pour la prédiction automatique du genre et de l'âge des personnes à partir de l'image du visage n'a cessé de croître ces dernières années. Un classement automatique par genre et par âge sont également requis chaque fois qu'ils sont nécessaires d'identifier le genre

et l'âge d'un public, par exemple pour analyser l'efficacité de la publicité. Une telle prédiction peut être également utile pour créer des interfaces homme-ordinateur où le comportement du système est ajusté à un utilisateur spécifique en fonction d'un certain nombre de facteurs, y compris son genre et son âge. Dans cette partie, nous faisons un aperçu des techniques et des travaux existants sur la prédiction du genre et de l'âge à partir d'images de visage.

2.6.1. Les techniques de prédiction de genre

Un certain nombre des techniques bien étudiées qui ont été appliquées pour la prédiction de genre à partir d'images de visage, cela avant l'arrivée massive des CNN.

Hui-Cheng Lian et al [82] ont proposé la classification du genre en tenant compte à la fois des informations sur la forme et la texture pour représenter les images faciales. La zone du visage est divisée en petites régions, à partir desquelles les histogrammes binaires locaux (LBP) sont extraits et concaténés en un seul vecteur représentant efficacement l'image faciale.

En outre, La machine à vecteur de soutien (SVM) est appliquée pour la prédiction de genre. Jing Wu et coll [83] ont proposé une classification en utilisant la forme à partir de l'ombrage (SFS). L'analyse Linéaire discriminante est utilisée en fonction du géodésique principal, les paramètres d'analyse pour distinguer le genre féminins et masculins des visages de test et La technique SFS est utilisée pour améliorer l'analyse des performances de la classification des images en échelle de gris. Kazuya Ueki et coll [84] ont présenté la classification à l'aide d'images faciales dans diverses conditions d'éclairage. Aussi, les MLP ont été appliquées pour ce domaine par Golomb et al. [85] et Khan et al [86] en 2000 qui ont appliqué dans leur travaux le renforcement des classificateurs (et en particulier l'Adaboost) pour la prédiction de genre pendant cette période.

2.6.2. Les techniques d'estimation d'âge

Le problème de prédiction d'âge est généralement abordé comme un problème de classification. Parmi les études sur la prédiction d'âge, Yamaguchi et coll [87] ont confirmé que les différences entre les caractéristiques du visage d'un adulte et le visage d'un enfant comprennent la longueur du visage et le rapport de chaque partie. Burt et Perrett [88] ont étudié l'estimation de l'âge d'après l'utilisation de visages moyens de personnes entre 25 à 60 ans, ils ont employé une méthode qui généraliser la texture du visage et la forme, aussi Ueki et Coll [89] ont signalé une méthode de classification des groupes d'âge par analyse discriminante linéaire (LDA). Ueki et al [90] ont formé 11 modèles gaussiens correspondant à 11 catégories d'âge en utilisant l'algorithme de maximisation des attentes (EM). À l'étape du test, chaque

image est adaptée aux 11 Gaussiens, et la classe d'âge qui correspond au modèle avec le maximum de vraisemblance choisi. Bien que le SVM a été testé pour classification par âge à plusieurs reprises [91]. D'autre part, Kanno et al [92] montrent qu'une personne peut être identifiée par des réseaux neuronaux représentant quatre groupes d'âges (12 ans, 15 ans, 18 ans et 22 ans). Kwon et Lobo [93] ont indiqué une méthode pour classer les images d'entrée dans l'un des trois groupes d'âge suivants : enfant, jeunes et personnes âgées en utilisant les informations sur la texture. Cependant, presque toutes les études précédentes ont été basées sur la méthode de développement cranio-facial et l'analyse des rides de la peau.

Ainsi, le travail de Lanitis et Al [94] montrant la supériorité de MLP avec le classificateur du voisin le plus proche (NN) et H-F Yang and Chen C-S [95] ont proposé un modèle CNN pour estimation de l'âge, les résultats expérimentaux rapportés dans sont très prometteurs, car ils démontrent que les étiquettes distribuées introduisent la notion de continuité d'âge dans l'entraînement de la classification pure. (Nous présentons les CNN en détail dans le chapitre 3)

2.6.3. Quelques travaux développés pour prédiction du genre et de l'âge :

Ici, nous citons au tableau (2.4, 2.5) les méthodes récentes pour un système d'estimation de l'âge et du genre selon des caractéristiques du visage.

Approche	Travail représentatif	Année
LPB + AdaBoost	Shan [96]	2012
AAM + Bayesian	Shih [97]	2013
Multiscale LBP+SVM	Tapia and Perez [98]	2013
Appearance based +LDA	Bekios-Calfa et al. [99]	2014
CNN	Levi and Hassner [100]	2015

Tab 2.4 : Synthèse des travaux développés pour la prédiction de genre.

Approche	Travail représentatif	Année
AAM + SVR	Khoa and Ching [101]	2009
BIF + OLPP	Guodong and Guowang [102]	2010
CAM + SVR	Khoa And Ching [103]	2011
BIF + KCCA	Guodong and Guowang [104]	2014
CNN	H-F Yang and Chen C-S [105]	2015

Tab 2.5 : Synthèse des travaux développés pour la prédiction d'âge.

2.7. Conclusion :

Dans ce chapitre, on s'intéresse à la reconnaissance de visage, on a énuméré les différentes approches de ce domaine avec les avantages et les inconvénients de chacune. Ensuite, on a présenté quelques travaux qui ont été développés pour la reconnaissance faciale et la prédiction du genre et de l'âge.

Dans cette étude nous avons classifié les techniques de ces domaines en trois catégories : les approches globales, locales et hybrides. On a souligné que les réseaux neurones à convolution sont appliqués récemment pour la prédiction du genre et l'estimation d'âge, nous détaillons cette technique dans le chapitre prochain.

CHAPITRE III

APPRENTISSAGE PROFOND : LES RESEAUX NEURONES CONVOLUTIFS

3. Chapitre III

3.1. Introduction

Nous allons décomposer ce chapitre en deux parties : nous présentons dans la première partie une vue générale sur l'intelligence artificielle et ses domaines, ensuite nous introduisons l'apprentissage automatique comme une sous-discipline de l'intelligence artificiel ainsi nous déterminons ses différents paradigmes d'apprentissage.

Dans la deuxième partie, nous allons décrire en détail les notions de l'apprentissage profond basées essentiellement sur les réseaux neurones artificielles, ensuite nous allons étudier les réseaux de neurones convolutifs pour la reconnaissance de visage de genre et estimation d'âge, enfin nous donnerons un aperçu sur les différentes architectures des réseaux convolutifs.

3.2. L'intelligence artificielle

L'intelligence artificielle (IA) [106] est une science qui date d'une trentaine d'années. Son objet est de reconstituer à l'aide de moyens artificiels (presque toujours des ordinateurs) des raisonnements et des actions intelligentes. Les difficultés sont a priori de deux types :

- Pour la plupart de nos activités nous ne savons pas nous-mêmes comment nous nous y prenons. Nous ne connaissons pas de méthode précise, pas d'algorithme qui guide aujourd'hui les informaticiens pour comprendre un texte écrit ou reconnaître un visage, de montrer un théorème, établir un plan d'action, résoudre un problème, apprendre...
- Les ordinateurs sont a priori très loin d'un tel niveau de compétence. Il faut les programmer depuis le tout début. Les langages de programmation ne permettent, en effet, que d'exprimer des notions très élémentaires.

3.2.1. Définition

On peut définir l'IA comme tout problème pour lequel aucune solution algorithmique n'est connue, relève a priori de l'intelligence artificielle.

Par algorithme, il faut entendre ici toute suite d'opérations ordonnées, bien définies, exécutables sur un ordinateur actuel et qui permet d'arriver à la solution en un temps raisonnable (de l'ordre de la minute ou de l'heure). Ainsi, on ne connaît pas d'algorithme pour jouer aux échecs, bien que ce jeu ne comporte qu'un nombre fini de situations, mais les étudier

toutes demanderaient des millénaires. De même, il n'existe pas d'algorithme pour effectuer un diagnostic médical, résumer un texte ou le traduire dans une langue étrangère.

3.2.2. Domaines de l'AI

L'intelligence artificielle regorge en nombreuses domaines et de tâches possibles comme :

- La perception et la reconnaissance des formes.
- Les mathématiques et la démonstration automatique de théorèmes.
- Les jeux.
- Compréhension du langage naturel (LN).
- Systèmes experts.
- La robotique.
- La résolution de problèmes.

L'essor remarquable de l'intelligence artificielle est notamment lié aux progrès de l'une de ses composantes : l'apprentissage automatique (Machine Learning, en anglais).

3.3. L'apprentissage automatique

L'apprentissage automatique (ML) est une discipline de l'IA qui fait référence au développement, à l'analyse et à l'implémentation de méthodes qui permettent à une machine d'évoluer grâce à un processus d'apprentissage, et ainsi de remplir des tâches qu'il est difficile ou impossible de remplir par des moyens algorithmiques plus classique, objectif est d'extraire et exploiter automatiquement l'information présente dans un jeu de données. On distingue généralement trois types d'apprentissage [107] :

- 1. L'apprentissage supervisé :** Un ensemble d'apprentissage et les résultats souhaités correspondants de la fonction pour apprendre sont bonne fonctionnement. Ainsi, lors de l'apprentissage de l'algorithme présente itérativement des exemples au système et adapte ses paramètres en fonction de la distance entre les sorties produites et les sorties souhaitées.
- 2. Apprentissage non supervisé :** La structure sous-jacente des données d'expérience, c'est-à-dire le résultat souhaité, est inconnu et doit être déterminée par l'entraînement d'Algorithme. Par exemple, pour une méthode de classification, cela signifie que les informations sur les classes ne sont pas disponibles et doivent être approximées par groupe des exemples d'apprentissage utilisant une certaine mesure de distance, une technique appelée le regroupement (*clustering*).

3. L'apprentissage par renforcement : le résultat exact de la fonction à apprendre n'est pas prédéfini, et l'apprentissage consiste en ajustant des paramètres basés seulement sur deux concepts, la récompense et la pénalité. Autrement dit, si le système ne fonctionne pas bien (assez) il est « pénalisé » et les paramètres sont adaptés en conséquence. Sinon, il est « récompensé », c.-à-d. un certain renforcement positif prend lieu.

Il existe d'autre type d'apprentissage on peut le citer aussi l'apprentissage semi-supervisé, il combine la technique d'apprentissage supervisé et non supervisé. L'apprentissage automatique contient nombreuses méthodes basées sur les différents paradigmes d'apprentissage, la figure (3.1) au-dessous montre quelques algorithmes largement utilisé en ML.

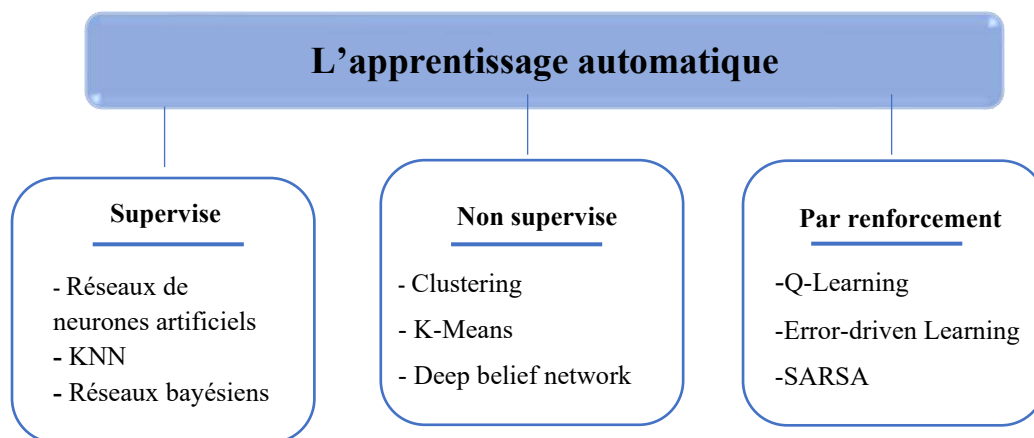


Fig 3.1 : Les différentes méthodes de ML

Permet les méthodes efficaces de ML pour la classification des données sont ; les réseaux de neurone artificiel (RNA), qui sont en effet capables d'approximer des fonctions non-linéaires complexes afin de traiter des données de grande dimension. Dans cette phase, on va présenter quelque notion de RNA.

3.3.1. Les réseaux de neurones artificiels (RNA)

Les réseaux de neurones sont des modèles de calcul très schématiquement inspirés du fonctionnement de la biologie du cerveau humain [108] [109]. Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire (neurone artificiel) calcule une sortie unique sur la base des informations qu'il reçoit.

-Fondements Biologiques : L'influx nerveux est assimilable à un signal électrique se propageant comme ceci : Les dendrites reçoivent l'influx nerveux d'autres neurones, ensuite le neurone évalue l'ensemble de la stimulation reçue, si elle est suffisante, il est excité et transmet un signal (0/1) le long de l'axone. L'excitation est propagée jusqu'aux autres neurones qui y sont connectés via les synapses.

3.3.2. Le fonctionnement des réseaux neurones

Les réseaux de neurones sont constitués d'un ensemble de nœuds qui représentent des neurones artificiels. Les nœuds sont reliés par des bords dirigés. Ces bords représentent les synapses d'un réseau neuronal biologique [108]. Chaque neurone a une fonction activation l_j . La notation pour le nœud où le bord commence est i et le nœud où le bord se termine est j . Chaque arête est associée à un poids w_{ij} , qui correspond à l'arête dirigée entre deux nœuds. La valeur du neurone est calculée par la somme pondérée des valeurs des nœuds d'entrée, on décrit comme l'activation entrante, notée a_j et calculée comme dans l'équation (3.1). La fonction d'activation sur cette somme pondérée est calculée comme dans l'équation (3.2). Un exemple d'un neurone artificiel simple, également connu sous le nom de perceptron, est illustré dans la figure 3.2.

$$a_j = \sum_i w_i^j x_i \quad (3.1)$$

$$\hat{y}_j = l_j(a_j) \quad (3.2)$$

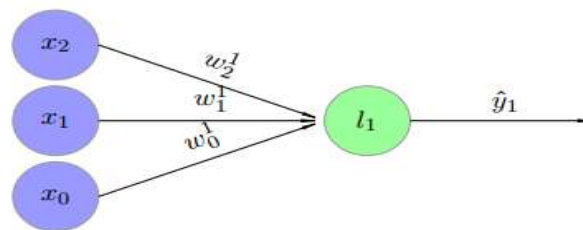


Fig 3.2 : Exemple d'un neurone artificiel (perceptron) [108]

(Où x_0 , x_1 et x_2 sont les entrées, w_0^1 , w_1^1 et w_2^1 constituent le vecteur de poids du neurone et l_1 est la fonction d'activation. La sortie est définie comme \hat{y}_1)

Les fonctions d'activation établissent des corrections sur les signaux de sortie pour les neurones. Les réseaux neuronaux peuvent utiliser de nombreuses fonctions d'activation différentes. Le choix d'une fonction d'activation pour un réseau neuronal est une étude importante car elle peut affecter la façon dont vous devez entraîner les données d'entrée. Parmi les différentes fonctions d'activation qui peuvent être utilisées, on trouve :

- **La fonction sigmoïde :** La fonction d'activation sigmoïde est un choix très courant pour les réseaux neurones à action anticipée qui ne doivent produire que des nombres positifs et comprise entre 0 et 1, la sigmoïde $\sigma(a)$ est calculée par la fonction suivante :

$$\sigma(a) = \frac{1}{(1+e^{-a})}, \text{ où } a \text{ est l'entrée de la fonction} \quad (3.3)$$

- **La Fonction tangente hyperbolique :** la fonction est très courante pour les réseaux neuronaux qui doivent produire des valeurs comprises entre -1 et 1, la fonction $\tanh \varphi(a)$ est calculée comme suit :

$$\varphi(a) = \frac{(e^a - e^{-a})}{(e^a + e^{-a})} \quad (3.4)$$

- **La fonction Relu :** la plupart des recherches actuelles recommandent la fonction ReLU en raison de meilleurs résultats lors de l'entraînement du réseau, Il est calculé par l'équation suivante

$$f(x) = (0, x) \quad (3.5)$$

- **La fonction softmax :** Softmax se trouve généralement dans la couche de sortie d'un réseau neuronal. La fonction softmax est utilisée sur un réseau neurone de classification. Le neurone qui a la plus haute valeur revendique l'entrée en tant que membre de sa classe. La fonction d'activation softmax force la sortie du réseau neurone à représenter la probabilité de l'entrée associées à chacune des classes. la fonction softmax comme suite :

$$f(a_k) = \frac{e^{a_k}}{\sum_{k'=1}^k e^{a_{k'}}}, k \text{ nombre de classes} \quad (3.6)$$

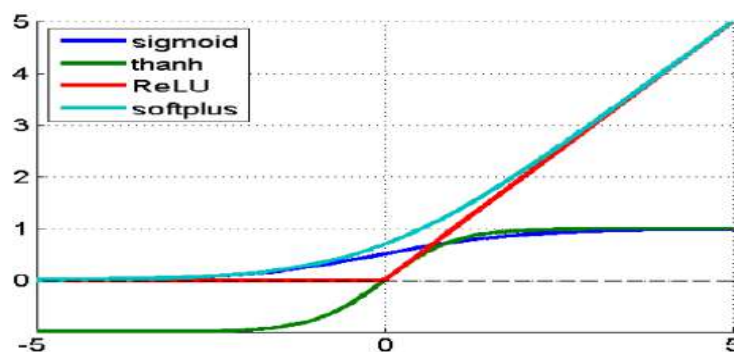


Fig 3.3 : Fonctions d'activation couramment utilisés

3.3.3. Les topologies de RNA

On distingue généralement deux topologies de RNA selon le type de connexions des neurones, la figure suivante montre les deux familles de RNA : (a) non bouclés (feedforward) et (b) bouclés (feedback).

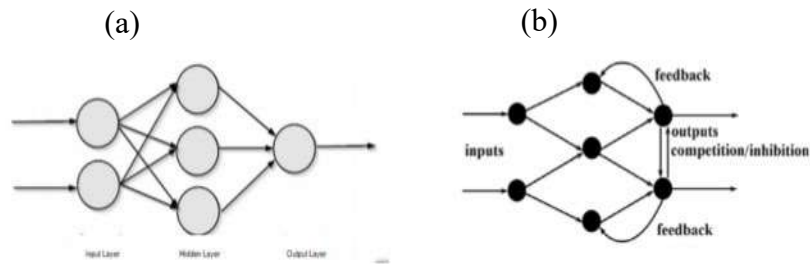


Fig 3.4 : Topologie des Réseaux de neurones artificiels [109].

3.3.4. Le perceptron multicouche (PMC)

Le perceptron multicouche (Multi Layer Perceptron : MLP) est un classifieur de type réseaux de neurones feedforward organisés en couches. Les perceptrons multicouches sont capables de traiter des données qui ne sont pas linéairement séparables. Avec l'arrivée des algorithmes de rétro propagation, ils deviennent le type de réseaux de neurones le plus utilisé. Les PMC sont généralement organisés en trois couches, la couche d'entrée, la couche intermédiaire (dite couche cachée) et la couche de sortie. La figure 3.5 illustre la structure d'un PMC présentant trois neurones en entrée, deux couches caches avec quatre neurones et un neurone en sortie. Lorsque tous les neurones d'une couche sont connectés aux neurones de la couche suivante, on parle alors de couches complètement connectées [110][111].

L'apprentissage des réseaux PMC consiste à présenter au réseau les données d'entrées, on lui demande après de modifier sa pondération de façon à retrouver la sortie correspondante. D'abord on propage vers l'avant les entrées jusqu'à obtenir une sortie calculée par le réseau puis on compare cette sortie avec la sortie désirée, enfin on modifie les poids de telle sorte qu'à la prochaine itération l'erreur commise entre les sorties réelles et les sorties désirées soit minimisée, on répète ce processus jusqu'à ce qu'on obtienne une erreur de sortie négligeable [111].

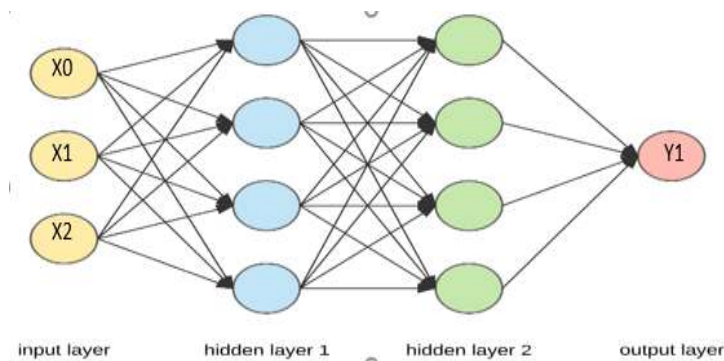


Fig 3.5 : Schéma d'un perceptron multicouche [108].

3.4. L'apprentissage profond

L'apprentissage profond (Deep Learning en anglais) est un nouveau domaine de recherche de ML qui a été introduit dans le but de rapprocher le ML de son objectif principal à savoir : l'intelligence artificielle. Ce domaine a été appliquée avec succès dans nombreuses applications du monde réel. Il concerne les algorithmes inspirés par la structure et le fonctionnement du cerveau. Ils peuvent apprendre plusieurs niveaux de représentation dans le but de modéliser des relations complexes entre les données différentes.

Les modèles de l'apprentissage profond sont bâtis sur le même modèle que les perceptrons multicouches précédemment décrits. Cependant, il convient de souligner que les différentes couches intermédiaires sont plus nombreuses. Chacune des couches intermédiaires va être subdivisée en sous partie, traitant un sous problème, plus simple et fournissant le résultat à la couche suivante, et ainsi de suite jusqu'à la couche de classification [119].

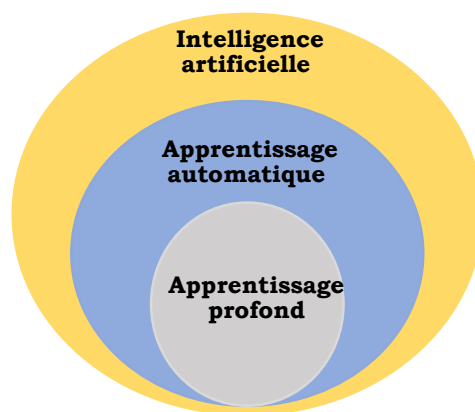


Fig 3.6 : La relation entre l'intelligence artificielle, L'apprentissage automatique et l'apprentissage profond

3.4.1. Pourquoi l'apprentissage profond ?

Le développement des algorithmes du l'apprentissage profond a pour but de pallier l'échec des algorithmes traditionnels de l'apprentissage automatique qui tente à résoudre une grande variété de problèmes de l'intelligence artificielle (l'IA).

Parmi les grandes différences entre le l'apprentissage profond et les algorithmes de l'apprentissage automatique , c'est qu'il s'adapte lorsque la quantité de données fournie est grande plus les performances d'un algorithme du apprentissage profond sont meilleures, tandis que plusieurs algorithmes de l'apprentissage automatique traditionnel possèdent une borne supérieure à la quantité de données qu' ils peuvent recevoir des fois appelée "plateau de

performance", les modèles d'apprentissage profond n'ont pas de telles limitations (théoriquement) et ils sont même allés jusqu'à dépasser la performance humaine dans des domaines comme l'image processing et la reconnaissance de formes.

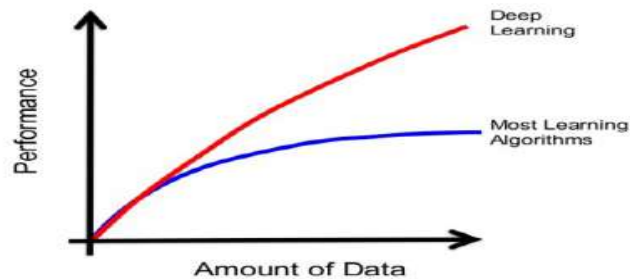


Fig 3.7 : La différence de performance entre l'apprentissage profond et la plupart des algorithmes de l'apprentissage automatique en fonction de la quantité de données [109].

Autre différence entre les algorithmes de ML classique et les algorithmes de Deep Learning, réside dans le processus d'extraction des caractéristiques de l'ensemble de données, les algorithmes de ML nécessitent une étape d'extraction des caractéristiques qui est faite manuellement chose qui est difficile à faire et coûteuse en temps, contrairement en Deep Learning ce processus est exécuté automatiquement par l'algorithme.

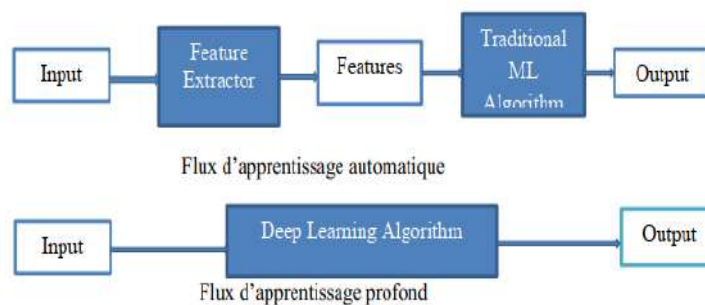


Fig 3.8 : Le procédé du ML classique comparé à celui du Deep Learning [109]

3.4.2. Les différents types de modèles

- 1. Réseaux de neurones récurrents (RNN)** ont été introduit en deep Learning afin de pouvoir traiter des données séquentielles. Dans un réseau neuronal conventionnel, nous supposons que toutes les entrées et les sorties sont indépendantes les unes des autres. Mais pour de nombreuses tâches, c'est une très mauvaise idée. Si on veut prédire le prochain mot dans une phrase, il faut connaître les mots qui sont venus avant. Les RNN sont appelés récurrents car ils exécutent la même tâche pour chaque élément d'une séquence, la sortie étant dépendante des calculs précédents. Les RNN ont une mémoire qui capture un état caché

qui résume leur historique, elle utilise les mêmes paramètres pour chaque entrée car elle effectue la même tâche sur toutes les entrées ou couches cachées pour produire la sortie. Cela réduit la complexité des paramètres, contrairement aux autres réseaux de neurones. Les RNNs sont utilisés pour la traduction automatique, la reconnaissance vocale et la modélisation de la langue [111] [112].

2. Les réseaux de longue mémoire à court terme (LSTM) sont une extension pour les réseaux neuronaux récurrents, qui ont la capacité d'apprendre et de sauvegarder de longues séquences de données d'entrée grâce à l'utilisation de portes qui régulent le flux d'informations du réseau. Les modèles LSTM tiennent en compte les trois limites des RNN. La mémoire par l'élimination des informations des étapes précédentes lors du passage à des étapes ultérieures, il est possible de perdre des informations importantes. La disparition de gradient, car le gradient est la valeur utilisée pour mettre à jour le poids utilisé dans un réseau neuronal, si une valeur de gradient devient extrêmement faible, elle ne contribue pas à l'apprentissage. Dans le problème de la disparition du gradient, le gradient se rétrécit en se propageant dans le temps, et enfin le gradient explosant qui se produit lorsque le réseau attribue une importance déraisonnablement élevée aux poids [113][114].

3. Les réseaux convolutifs ou convolutionnels (CNN), sont basés sur l'architecture d'un perceptron multicouche, inspirés du comportement du cortex visuel des vertébrés, Ils utilisent des données qui ont une topologie en forme de grille. Exemples de données de type grille sont des données de séries temporelles en 1D et les images en 2D de pixels. Le nom réseau de neurones convolutif indique que le réseau applique une opération mathématique appelée convolution. Il existe quatre (4) opérations principales de base dans un réseau CNN à savoir : la couche de convolution, la couche Rectified Linear Unit (fonction d'activation), la couche pooling et la couche entièrement connectée.

4. Modèle génératif permet à la fois de générer de nouveaux exemples à partir des données d'entraînement et d'évaluer la probabilité qu'un nouvel exemple provienne ou ait été généré à partir des données d'entraînement. Voici quelques exemples de modèle génératif :

- Boltzmann Machines [115].
- Restricted Boltzmann Machines [116].
- Deep Belief Networks [117].
- Deep Boltzmann Machines [118].

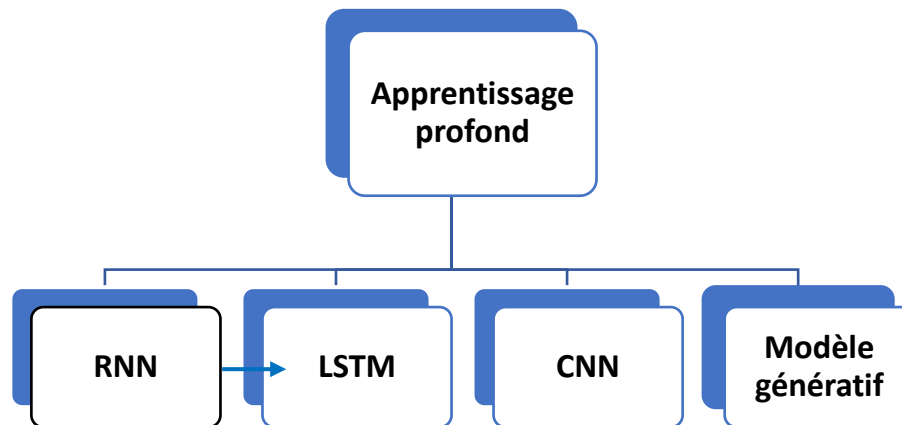


Fig 3.9 : Les modèles de l'apprentissage profond [109].

3.5. Les réseaux de neurones convolutifs

Les réseaux de neurones convolutifs sont à ce jour les modèles les plus performants pour classer des images. L'architecture basique de CNN comporte deux parties fondamentales. La première partie d'un CNN est la partie de convolution qui fonctionne comme un extracteur de caractéristiques des images. Une image est passée à travers une succession de filtres, ou noyaux de convolution, créant de nouvelles images appelées cartes de convolutions. Certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum local. Au final, les cartes de convolutions sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN. Ce code CNN est résultat de sortie de la partie convolutive, ensuite il est branché en entrée d'une deuxième partie, constituée de couches entièrement connectées (perceptron multicouche). Le rôle de cette partie est de combiner les caractéristiques du code CNN pour classer l'image. La sortie est la dernière couche comportant un neurone par catégorie. Les valeurs numériques obtenues sont généralement normalisées entre 0 et 1, pour produire une distribution de probabilité sur les catégories [110].

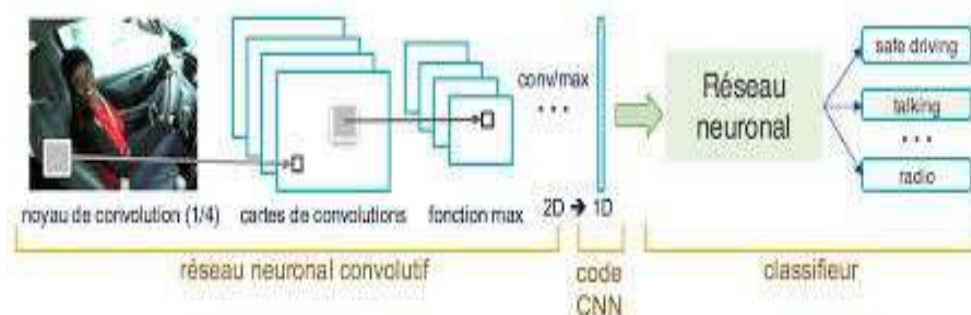


Fig 3.10 : Architecture basique d'un réseau de neurone convolutifs [111]

Les réseaux de neurones convolutifs s'appuient sur trois aspects importants qui peuvent aider à améliorer un système d'apprentissage profond [119] : les interactions clairsemées, le partage des paramètres et les représentations invariantes.

1. Les interactions clairsemées (*Sparse interactions*) : l'image d'entrée peut avoir des millions de pixels, mais on peut détecter de petites caractéristiques significatives telles que les bords avec des noyaux qui n'occupent que des dizaines ou des centaines de pixels. Cela signifie que nous devons stocker moins de paramètres, ce qui réduit à la fois les besoins de mémoire du modèle et améliore son efficacité statistique. Donc, le calcul de la sortie nécessite moins d'opérations. Ces améliorations de l'efficacité sont généralement assez importantes.

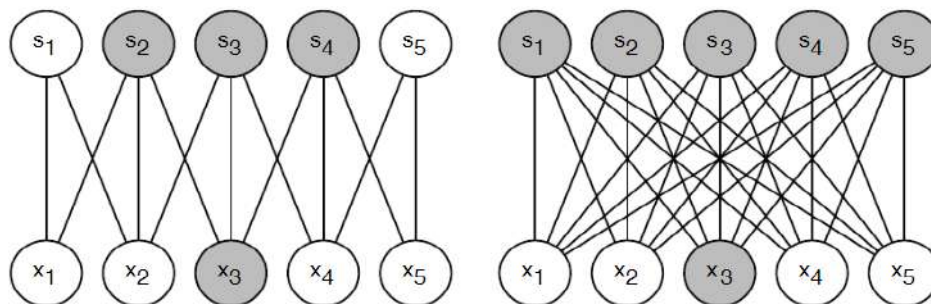


Fig 3.11 : *Sparse interactions* [119].

On met l'unité d'entrée, x3, et les unités de sortie qui sont affectées par cette unité. (À gauche) Lorsque s'est formé par convolution avec un noyau de largeur 3, seules trois sorties sont affectées par x (à droite) Lorsque s'est formé par une multiplication matricielle, toutes les sorties sont atteintes par x3.

- 2. Le partage des paramètres** : il est appliqué pendant l'opération de convolution, signifie qu'au lieu d'apprendre un ensemble de paramètres séparé pour chaque emplacement, nous apprenons un seul ensemble. Cela réduit encore les exigences de stockage du modèle en k paramètres.
- 3. Les représentations invariantes** : dans la convolution, le type particulier de partage de paramètres fait que la couche a une propriété appelée représentations invariantes. Dire qu'une fonction est équivariante signifie que si l'entrée change, la sortie change de même contexte, spécifiquement une fonction $f(x)$ est équivalente à une fonction g si $f(g(x)) = g(f(x))$. Par exemple, dans le cas d'une image, Il est utile de détecter les arêtes dans la première couche d'un réseau convolutif. Les mêmes arêtes apparaissent plus ou moins partout dans l'image, donc il est pratique de partager des paramètres sur l'image entière.

Une architecture CNN est formée par un empilement de couches de traitement indépendantes :

- La couche de convolution qui traite les données d'un champ récepteur.
- La couche de pooling qui permet de compresser l'information en réduisant la taille de l'image intermédiaire (souvent par sous-échantillonnage).
- La couche de correction, souvent appelée par abus ReLU en référence à la fonction d'activation (Unité de rectification linéaire).
- La couche entièrement connectée, qui est une couche de type perceptron.

Pour avoir comment l'image traite au niveau de CNN, on va expliquer dans cette phase le fonctionnement de chaque couche dans de réseaux neurones.

3.5.1. L'architecture en couche du réseau CNN

1. La couche convolution

Elle est la composante principale des réseaux de neurones convolutifs, son objectif de construire une carte de caractéristique de l'image d'entrée. Elle prend une image X de taille $l_i \times h_i$ en entrée et la convolue par un filtre de noyau K de taille $l_k \times h_k$, les dimensions de l'image de sortie sont :

$$l_0 = l_i - \lfloor \frac{l_k}{2} \rfloor \quad \text{Largeur de l'image X} \quad (3.7)$$

$$h_0 = h_i - \lfloor \frac{h_k}{2} \rfloor \quad \text{Hauteur de l'image X} \quad (3.8)$$

Alors, le potentiel est une image Y (aussi appelée carte de convolution) dont les valeurs par propagation avant (ou forward propagation) sont calculées ainsi :

$$Y(i, j) = \sum_{(u,v) \in K} K(u, v) X(i + u, j + v), \text{ Où } 0 \leq u \leq l_k \text{ et } 0 \leq v \leq h_k \quad (3.9)$$

Les principaux paramètres à prendre en compte dans les couches de convolution sont :

- **Profondeur de la couche** : nombre de noyaux de convolution (ou nombre de neurones associés à un même champ récepteur).
- **Le pas (*stride*)** : contrôle le chevauchement des champs récepteurs. Plus le pas est petit, plus les champs récepteurs se chevauchent et plus le volume de sortie sera grand.
- **Le remplissage (*padding*)** : Le remplissage est simplement un processus d'ajout de couches de zéros à l'images d'entrée afin d'aider le noyau de couvrir tout l'image, il est parfois souhaitable de conserver la même surface que celle du volume d'entrée

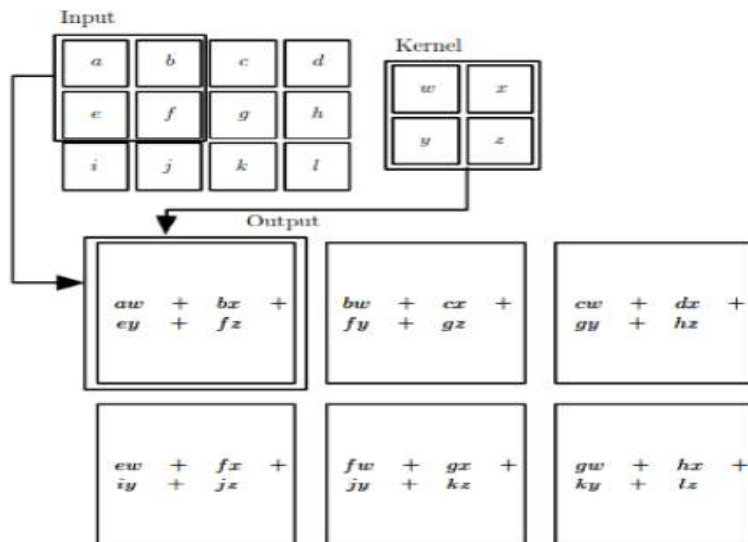


Fig 3.12 : Exemple de convolution d'une Image 2D De Taille 4×3 Par Un Noyau 2×2 [119].

2. Couche de pooling

Il est utile d'insérer une couche pooling entre les couches convolutifs successives dans la structure de CNN, pour réduire progressivement la taille spatiale de l'image donc réduire les paramètres de traitements dans le réseau. La couche pooling prend en entrée une image X' de taille $l_i \times h_i$ et un filtre de taille $l_k \times h_k$ ce qui charge à diminuer la taille de l'image par un facteur l_k en largeur et h_k en hauteur. L'opération revient à convoluer l'image par un filtre, Le résultat de chaque pas de convolution est présenté comme suit :

$$y(i, j) = \sum_{(u, v \in k)} X(h_k \times i + u, l_k \times j + v) \quad (3.10)$$

Il existe plusieurs fonctions de pooling : Le max pooling (le filtre retient à prendre la valeur maximale de la sélection). L'average pooling (la moyenne des pixels de la sélection, on calcule la somme de toutes les valeurs et on divise par le nombre de valeurs.) et la somme pooling (la moyenne sans avoir divisé par le nombre de valeurs)

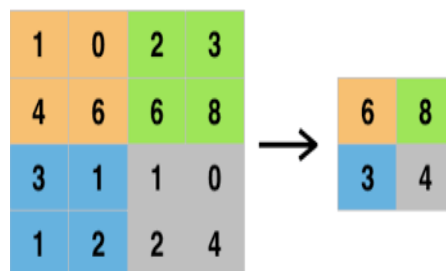


Fig 3.13 : Max Pooling avec un filtre 2×2 et un pas de 2.

3. La couche de correction (fonction d'activation)

Souvent, il est possible d'améliorer l'efficacité du traitement en intercalant entre les couches de traitement une couche qui va opérer comme une fonction mathématique (fonction d'activation) sur les signaux de sortie. La fonction ReLU (abréviation de Unités Rectifié linéaires) reste la fonction la plus utilisée ($F(x)=\max(0, x)$). Cette fonction force les neurones à retourner des valeurs positives.

4. La couche entièrement connectée (fully connected) :

C'est la dernière couche dans le réseau neurone convolutif, l'objectif de cette couche est de prendre les résultats du processus convolution/ sous-échantillonnage et de les utiliser pour la prédiction de la classe de l'image. La sortie de convolution et de pooling est aplatie en un seul vecteur de valeurs, chacune représentant une probabilité qu'une certaine caractéristique appartient à une classe. La figure 3.14 ci-dessous montre les flux des valeurs d'entrée dans la première couche de neurones, ils sont multipliés par des poids et passent par une fonction d'activation (généralement Relu) tout comme dans réseau neurone classique, ils circulent ensuite à la couche de sortie, dans laquelle chaque neurone représente une étiquette de classification.

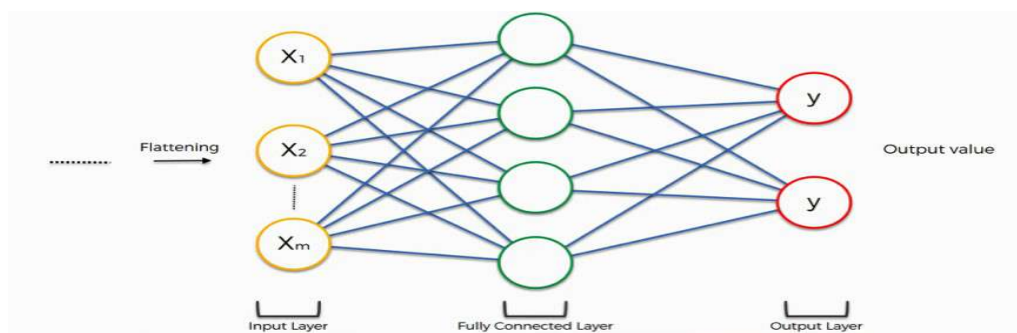


Fig 3.14 : La couche entièrement connectée.

3.5.2. Les modèles CNN célèbre

Il existe nombreux modèles de CNN utilisés dans plusieurs disciplines notamment la reconnaissance de forme. Récemment un concours annuel est organisé depuis 2010, le Imagenet Large Scale Visual Recognition Challenge (ILSVRC) organisé par ImageNet, le but du concours est de groupe de recherches s'affrontent a coup d'algorithmes et de modèles de réseau,

afin de faire avancer la recherche dans le domaine de la vision par ordinateur, le gagnant du concours sera considéré comme le meilleur modèle de réseau CNN de l'année (tableau 3.1).

Le nom de CNN	Année	Nombre des couches	Erreur (%) de classification (top-5)
ALEXNET	2012	8	16.4
ZFNET	2013	8	14.8
VGG	2014	16/19	6.8
GoogleNet	2014	22	6.7
ResNET	2015	152	3.6

Tab 3.1 : Les gagnants du concours ImageNet de 2012 à 2015[120].

Dans cette partie, on va présenter les architectures de CNN les plus répandues dans le domaine de reconnaissance de forme et vision par ordinateur :

- LeNet-5 : c'est le réseau convolutifs le plus ancien, développé par Y.Lecun et al en 1998, il a permis de classer les chiffres manuscrit , il a été utilisé par plusieurs banques pour reconnaître les nombres manuscrits sur les chèques. L'architecture de LeNet-5 est très simple, il prend en entrée une image de taille 32X32 pixel, il contient cinq couches, parmi lesquelles deux couches de convolutions suivies par un average pooling et trois couches de fully connected [19].
- ALXNET : a été le premier modèle qui a propagé les réseaux convolutifs dans la vision par ordinateur, développé par Alex .Krizhervsky, Ilya. Sutskever et Geoff. Hinton. Le modèle ALXNET a remporté le concours ILSVRC en 2012 et a nettement surpassé ses concurrents (15,3% VS 26,2% de taux d'erreur), Le réseau avait une architecture très similaire à LeNet-5 mais était plus profond, qui vient d'empiler quelques couches supplémentaires. Ce modèle contient huit couches, cinq couches de convolution et trois couches de fully connected, il est capable d'entraîner plus d'un million d'images et de les classer en 1000 catégories d'objets, ce qui lui permet d'apprendre une grande variété de représentations des caractéristiques pour un large éventail d'images [122].
- VGGNET-16 : le modèle a été développé par Simonyan et Zisserman. En effet, la méthode la plus efficace d'améliorer les performances des réseaux de neurones profonds est d'augmenter leur taille, donc l'architecture de VGGNET est basée sur 16 couches, parmi lesquelles 13 couches de convolutions avec des filtres uniquement de taille 3x3 et trois

couches de fully connected, VGGNET contient plus de 138 millions de paramètres, ce qui rend un peu difficile pour entraîner. Ils ont conçu un autre modèle plus tard plus profond, le VGG-19 [123].

- GoogleNet (*Inception v1*) : Le vainqueur de ILSVRC challenge 2014 était un réseau convolutif de Szeged et al. L'idée principale de l'architecture Inception est de découvrir comment une structure locale clairsemée optimale dans un réseau CNN peut être approchée et couverte par des composants denses facilement disponibles [124]. Pour cela le modèle Inception est basé sur des modules (fig3.15) qui démunie une grande quantité de paramètres. Cette architecture à 22 couches avec 5 millions de paramètres, Il existe également plusieurs versions de GoogleNet, parmi elles, Inception-v3.

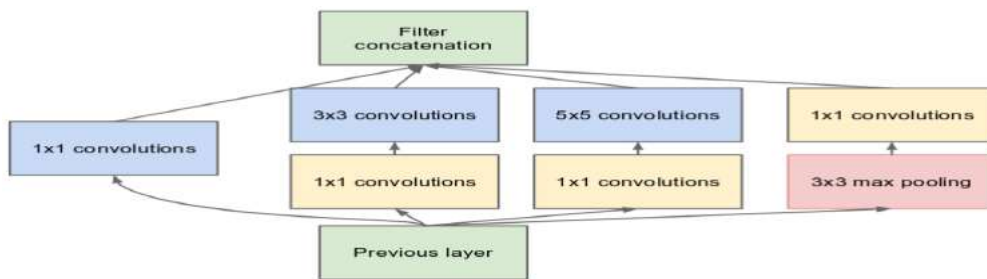


Fig 3.15 : Modules de l'architecture GoogleNet [120].

- RESNET : le modèle développé par Kaiming He et al, a été le vainqueur de ILSVRC en 2015. Il utilise des blocs résiduels qui sont la principale nouveauté dans ce modèle, et qui ont également donné le nom a cette architecture CNN. Ce modèle contient 152 couches cachées. L'idée clé de ResNet est très simple, consiste à ajouter de connexions d'identité explicites de l'entrée à la sortie d'un bloc résiduel permettant de contourner les connexions pondérées. Mathématiquement, cela signifie qu'au lieu d'apprendre un mappage original $H(x)$, une couche ResNet apprend le mappage résiduel $F(x) = H(x) - x$ [120].

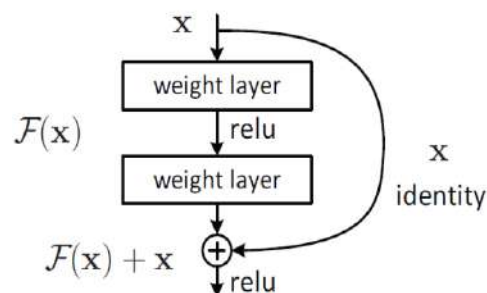


Fig 3.16 : Blocs résiduels qui permettent aux couches intermédiaires de ResNet [120].

Après avoir présenté les modèles CNN, Il est nécessaire de préciser que ces architectures CNN sont évaluées et comparées entre elles sur la base de données ImageNet, la classification générale des images n'est pas la seule application des CNN. Ainsi, l'apprentissage par transfert permet d'adapter les modèles de CNN a d'autres problèmes visuels, même avec des données d'entraînements limites, de manière plus générale, les CNN sont actuellement utilisés dans presque tous les domaines de la vision par ordinateur, la détection, la localisation et le domaine de reconnaissance d'objets.

3.5.3. L'entraînement d'un CNN

Après la création d'un modèle de CNN ou tout simplement quand on utilise un modèle déjà existant, il est utile d'entraîner ce modèle avec des données afin qu'il puisse apprendre et d'être capable de reconnaître un objet, pour cela on utilise une base de données contenant des images de grandes quantités. Une partie de cette base de données servira à entraîner le modèle CNN, c'est ce qu'on appelle la base d'apprentissage, et la partie de base de données restante sert pour évaluer et tester le modèle, on l'appelle la base de données de test. Il est nécessaire que toutes les images doivent être préalablement étiquetées, c'est à dire que chacune des images que nous utilisons doit appartenir à une classe spécifique, c'est l'apprentissage supervisé.

L'entraînement d'un CNN consiste à déterminer et à calculer empiriquement la valeur de chacun de ses poids. Le principe est le suivant, le CNN traite une image (de la base de données d'entraînement) et en sortie il fait une prédiction, c'est-à-dire qu'il dit à quelle classe il pense que cette image appartient. Sachant qu'on connaît préalablement la classe de chacune des images d'entraînement, on peut vérifier si ce résultat est correct. En fonction de ce résultat, on met à jour tous les poids du CNN selon un algorithme qui s'appelle la rétropropagation du gradient de l'erreur. Lors de la phase d'entraînement du modèle, le processus expliqué ci-dessus est répété plusieurs fois et avec la totalité des images de la base de données d'entraînement. Le but étant que le modèle classifie au mieux ces données. Lorsque le modèle a fini de mettre à jour ses poids, on évalue le modèle en lui présentant la base de données de test. On classe toutes ces images et on calcule son taux de bonne classification, c'est ce qu'on appelle la précision du modèle (accuracy). A la fin du processus d'apprentissage trois cas de figure peuvent se présenter :

- Le modèle est aussi performant sur les données d'entraînement (images sur lesquelles il s'entraîne) que sur les données de test, Cela est le cas de figure idéale, ça signifie que le

modèle a très bien fait son travail et qu'il reconnaît aussi bien les images qu'il connaît que celles qu'il n'a jamais vues.

- Le modèle reconnaît très bien les images d'entraînement et moins bien celles de test. Le modèle aura une faible capacité prédictive, il n'arrive pas à généraliser. On parle alors de sur-apprentissage. Dans ce cas-là on peut ajouter davantage d'images pour pallier ce problème.
- Le modèle ne reconnaît pas très bien les images d'entraînements et les images de test. On parle alors de sous-apprentissage. Dans ce cas-là ajouter plus d'images ne servira à rien, c'est généralement le modèle choisi qui ne convient pas, il faudrait utiliser un modèle plus complexe.

3.5.4. L'optimisation dans les réseaux profonds

La méthode de descente de gradient a fait ses preuves pour l'optimisation des paramètres d'un réseau de neurones. Elle est ainsi l'une des approches les plus efficaces et les plus utilisées. Nous rappelons brièvement quelques notions sur cette méthode. Soit θ le vecteur contenant tous les paramètres de réseau CNN [124].

Soit $J(\theta, g(x), y)$ la fonction de coût représentant l'erreur entre la fonction $g(x)$ (liée à la donnée d'entraînement x) et la valeur prédite y estimée par le réseau en utilisant les paramètres θ . Un pas de l'itération de la descente de gradient est donné par :

$$\theta_{t+1} \leftarrow \theta_t - \eta \nabla_{\theta} J(\theta, g(x), y) \quad (3.11).$$

Où η est une constante positive appelée taux d'apprentissage. La méthode s'applique pour tout espace de dimension.

3.5.5. Les Algorithmes d'optimisation de la descente de gradient

Dans ce qui va suivre, on présentera quelque algorithme largement utilisé en Deep Learning a fin de d'optimiser le réseau neurone en fonction de la quantité de données, on offre un compromis entre la précision de la mise à jour des paramètres et le temps d'exécution de cette mise à jour [125] [126].

-Momentum : L'objectif principal de la méthode dite de Momentum est d'accélérer le processus de descente de gradient, et ceci en ajoutant un vecteur de vélocité à l'expression initiale :

$$v_{t+1} \leftarrow \gamma v_t - \eta \nabla J(\theta_t) \quad (3.12)$$

$$, \theta_{t+1} \leftarrow v_{t+1} + \theta_t$$

Le vecteur v_{t+1} est calculé au début de chaque itération et représente la mise à jour de la vitesse d'une "balle dévalant une pente". La vitesse s'accumule à chaque itération d'où l'introduction d'un hyper-paramètre γ permettant d'amortir la vitesse quand on atteint une surface plate. Une bonne stratégie peut être de modifier γ en fonction du niveau d'apprentissage.

-Adagrad : Le principe de cette méthode est de faire que le taux d'apprentissage s'adapte aux paramètres, faisant de sorte qu'il s'ajuste automatiquement, en fonction de "l'éparité" des paramètres. Adagrad abaisse progressivement le taux d'apprentissage mais pas de la même manière pour tous les paramètres : les dimensions à pente plus prononcée voient leur taux abaissé plus rapidement que celles à pente douce. Plus formellement, le pas est décrit par :

$$\forall i, (\theta_{t+1})_i \leftarrow (\theta_t)_i - \alpha \frac{(\nabla J(\theta_t))_i}{\sqrt{\sum_{u=1}^t (\nabla J(\theta_u))_i^2}}, \alpha > 0. \quad (3.13)$$

-RMS Prop : L'algorithme ajuste automatiquement le taux d'apprentissage à chaque paramètre, comme Adagrad. Cependant, il ne cumule que les gradients issus des itérations récentes. Pour cela, il utilise une moyenne glissante

$$\forall i, (\nabla_t)_{t+1} \leftarrow \partial(\nabla_t)_i + (1 - \partial)(\nabla J(\theta_t))_i^2 \quad (3.14)$$

$$\forall i, (\theta_{t+1})_i \leftarrow (\theta_t)_i - \alpha \frac{(\nabla J(\theta_t))_i}{\sqrt{(\nabla_t)_{t+1} }}, \alpha > 0. \quad (3.15)$$

Où $\partial(\nabla_t)_i$ est la moyenne quadratique glissante du gradient. La division du gradient de la fonction objective par la racine de la moyenne quadratique glissante (c'est à dire l'amplitude) améliore la convergence.

-Adam est l'un des algorithmes les plus récents et les plus efficaces pour l'optimisation par descente de gradient. Le principe est le même que pour Adagrad et RMSPro : il adapte automatiquement le taux d'apprentissage pour chaque paramètre. Sa particularité est de calculer (m_t, v_t) des "estimations adaptatives des moments". Il peut donc être vu comme une généralisation de l'algorithme Adagrad :

$$\forall i, (m_{t+1})_i \leftarrow \beta_1 \cdot (m_t)_i + (1 - \beta_1) \cdot (\nabla J(\theta_t))_i \quad (3.16)$$

$$\forall i, (v_{t+1})_i \leftarrow \beta_2 \cdot (v_t)_i + (1 - \beta_2) \cdot (\nabla J(\theta_t))_i^2, \quad (3.17)$$

$$\forall i, (\theta_{t+1})_i \leftarrow (\theta_t)_i - \alpha \frac{\sqrt{1-\beta_2}}{1-\beta_1} \frac{(m_t)_i}{\sqrt{(v_t)_i + \epsilon}}, \quad (3.18)$$

$\alpha, \epsilon > 0$ et $\beta_1, \beta_2 \in]0,1[$.

Ici m_t est le premier moment du gradient (la moyenne) et v_t est son second moment (variance non-centrée). ϵ Est un paramètre de précision. Les paramètres β_1 et β_2 sont utilisés

pour réaliser des moyennes d'exécution sur les moments m_t et v_t respectivement, Les valeurs par défaut sont : 0.9 pour β_1 , 0.999 pour β_2 et 10^{-8} pour ϵ .

3.6. L'apprentissage par transfert

Le développement d'un nouveau réseau neurone convolutifs pour une nouvelle tâche est un processus difficile qui ajoute une surcharge au processus d'entraînement, ce dernier nécessite des milliers, voire des millions d'images. Récemment l'unité graphique de traitement (GPU) a permis de réduire considérablement le temps de traitement. Face au problème de la collecte suffisamment de données d'entraînement et la complexité d'apprentissage pour construire les modèles, l'apprentissage par transfert vise à utiliser un modèle pré-entraîné pour extraire automatiquement des caractéristiques à partir d'un nouveau jeu de données, pour cela on va transférer les connaissances d'un grand ensemble de données appelé domaine source vers un petit ensemble de données domaine cible [126].

L'apprentissage par transfert est une solution pratique pour appliquer un apprentissage profond sans avoir à recourir à un très grand jeu de données ni procéder à des calculs et un apprentissage très long [127]. Afin d'éviter toute la complexité de création de CNN, il est utile d'adapter des réseaux pré-entraînés disponibles publiquement exemple ResNet, VGG ...etc. Cette technique exploite la connaissance acquise sur un problème de classification général pour l'appliquer sur un nouveau problème particulier et similaire.

L'utilisation de réseaux CNN pré-entraînés sur d'immenses bases de données pour ensuite les adapter à une tâche spécifique. C'est ce que l'on appelle le fine-tuning [128]. Cette pratique permet d'apprendre des réseaux profonds en utilisant une initialisation des poids et des biais déjà très pertinente et générique. L'adaptation de ces paramètres est ensuite réalisée durant la phase d'apprentissage de la tâche spécifique que l'on souhaite réaliser. Cela a pour incidence une vitesse d'apprentissage bien plus rapide et une convergence quasiment garantie [129].

3.7. Conclusion

Dans ce chapitre nous avons passé en revue les fondements essentiels des réseaux de neurones convolutifs. Nous avons défini l'architecture détaillée de ce type de réseau ainsi que les différentes couches qui le constitue. Une partie a été consacrée aux propriétés et paramètres du CNN. Cette étude détaillée du modèle CNN nous permet de mieux comprendre son processus et faire une conception pour une tâche de reconnaissance du visage et d'estimation de genre et d'âge qui fera objet du chapitre suivant.

CHAPITRE IV

IMPLEMENTATION, EXPERIMENTATIONS ET RESULTATS

4. Chapitre IV

4.1. Introduction

Après avoir abordé l'aspect théorique dans les chapitres précédents, nous passons à la conception et l'implémentation de notre système élaborer lors de travail pour la prédiction du genre, l'estimation d'âges et la reconnaissance faciale des individus basé principalement sur les réseaux de neurones convolutifs.

4.2. Les bases de données de visage

4.2.1. FERET

Le programme de technologie d'identification de visage FERET (The Face Recognition Technology) est une base de données volumineuse pour l'évaluation du système de reconnaissance faciale. La base de données FERET présente 1199 personnes pour un total de plus de 14000 images couleur et en niveau de gris de taille 512×768 (dans la nouvelle version de la base). La base présente des variations d'expressions faciales, de pose, d'illumination ainsi qu'un certain délai entre différentes captures d'une même personne [130].



Fig 4.1 : Exemples d'images de visages de la base FERET.

4.2.2. IMDB & WIKI

Ils sont les plus grosses bases de données collaboratives du monde, avec des informations sur 3 250 000 œuvres visuelles pour un totale égale à 460 723 images de visages d'où on a 20 284 visages en IMDB et 62 328 de Wiki, les deux bases données contiennent des informations de l'âge et pour quelques images ont été collectées à des situations différentes, avec des variations dans les conditions d'éclairage, les expressions faciales et occultations partielles.



Fig 4.2 : Exemples d'images de visages de la base IMDB.

4.2.3. Séparation des bases de données

Pour implémenter notre application, il est nécessaire de diviser les bases de données en deux parties, la première pour l'apprentissage afin d'entraîner les modèles CNN et la deuxième pour le test pour valoriser les taux de classification du système, le tableau dessous montre comment on a décomposé la base d'images.

Système	La base de données	Les images d'apprentissages	Les images de tests
La reconnaissance faciale	FERET	1000	250
La prédiction du genre + l'estimation d'âge	IMDB	2000	500
	WIKI	2000	500

Tab 4.1 : les bases de données utilisées.

4.3. L'environnement de travail

Afin de mener à bien ce projet, nous avons utilisé un ensemble de matériels dont les principales caractéristiques sont les suivantes :

- Un PC portable HP i3 CPU 2.00 GHZ.
- Carte graphique Intel(R) HD Graphics 520.
- RAM de taille 4 GO.
- Disque dur de taille 309 GO.
- Système d'exploitation Windows 10 Pro, 64 bits.

4.3.1. Les outils de programmation

- **Python** : est un langage de programmation open source de haut niveau. Il est très sollicité par une large communauté de développeurs et de programmeurs. Python prend en charge de multiples paradigmes de programmation, y compris orientés objet, impératifs, fonctionnels

et procédurales, et dispose d'une bibliothèque standard vaste et complète. Les interpréteurs de Python sont disponibles pour de nombreux systèmes d'exploitation [131].

4.3.2. Les bibliothèques utilisées

- **Keras** : est une bibliothèque de réseaux neuronaux de haut niveau, écrite en Python et capable de s'exécuter sur TensorFlow ou Theano. Il a été développé dans le but de permettre une expérimentation rapide, pour en mesure de passer de l'idée au résultat le plus rapidement possible. Nous avons utilisé keras car il permet un prototypage facile et rapide (grâce à la modularité totale), il prend en charge les réseaux de neurones convolutionnels et fonctionne de façon transparente sur CPU et GPU [132].
- **Open CV** : est une bibliothèque qui propose plus de 2500 algorithmes de vision par ordinateur spécialisé dans le traitement d'images, accessible au travers d'API pour les langages C, C++, et Python. Elle est distribuée sous une licence BSD (libre) pour les plateformes Windows, GNU/Linux, Android et MacOS [132], nous avons utilisé cette bibliothèque pour la détection du visage à partir des images faciales.
- **Numpy** : La bibliothèque Numpy permet d'effectuer des calculs numériques avec Python. Elle introduit une gestion facilitée des tableaux de N-dimensions [132].

4.4. Les systèmes de reconnaissance visage par CNN

Les systèmes de reconnaissance faciale sont généralement sous divisés en deux étapes importantes. Dans un premier temps la détection des visages et le prétraitement de base d'images afin d'extraire une ou plusieurs zones d'intérêt, la seconde est la reconnaissance de visages qui contient habituellement deux parties : l'extraction des caractéristique et la classification des caractéristiques afin d'identifier les visages d'individus sous forme d'une probabilité de décision.

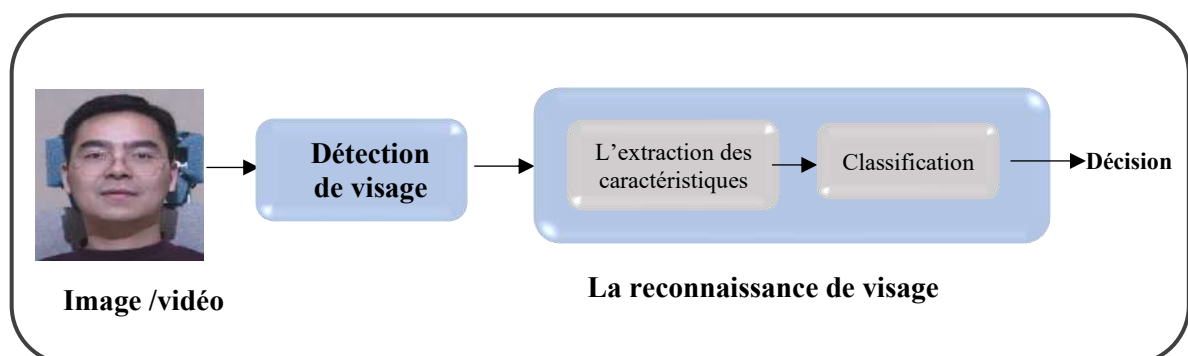


Fig 4.3 : Un système de reconnaissance de visages.

Pour la détection de visage, Il existe un très grand nombre de méthodes permettant d'accomplir la première étape d'un système de reconnaissance faciale, c'est-à-dire la détection de la position de régions d'intérêt des visages sur les images ou vidéos. La technique la plus répandue est sans doute l'algorithme proposé par les auteurs Viola et Jones. La détection par la technique de Haar cascade consiste à appliquer plusieurs faibles classificateurs en cascade qui ont été entraînés progressivement et valider avec de plus en plus de rigueur la présence de caractéristiques bien précises et qui représentent la réponse typique d'un visage. Ainsi pour l'étape de reconnaissance faciale, un traitement d'image sera effectué avant d'être classifié par les réseaux CNN. Une image de taille réduite sera l'entrées de la première couche de la structure du réseau, ensuite l'information se diffuse au niveau des autres couches du réseau, où à chaque couche un filtre numérique est glissé afin d'obtenir des caractéristiques des données observées.

L'opération fournit un niveau d'invariance au déplacement, à l'échelle et à la rotation lorsque le champ réceptif local permet au neurone ou à l'unité de traitement d'accéder à des éléments élémentaires tels que des arêtes ou des coins orientés. Après l'image d'entre est convoluer par un nombre de filtres, pour former une carte de caractéristiques qui sont le résultat des processus de convolution. Généralement, on applique la fonction max pooling pour le sous-échantillonnage. La carte des caractéristiques sous-échantillonnées est aplatie en un seul vecteur de valeurs, chacune représentant une probabilité qu'une certaine caractéristique appartient à une classe.

4.5. Système de prédiction de genre : principes et expérimentations

La prédiction de genre consiste à classer une personne selon deux catégories homme ou femme, le processus de classification par les réseaux CNN est illustre dans la figure (4.4).

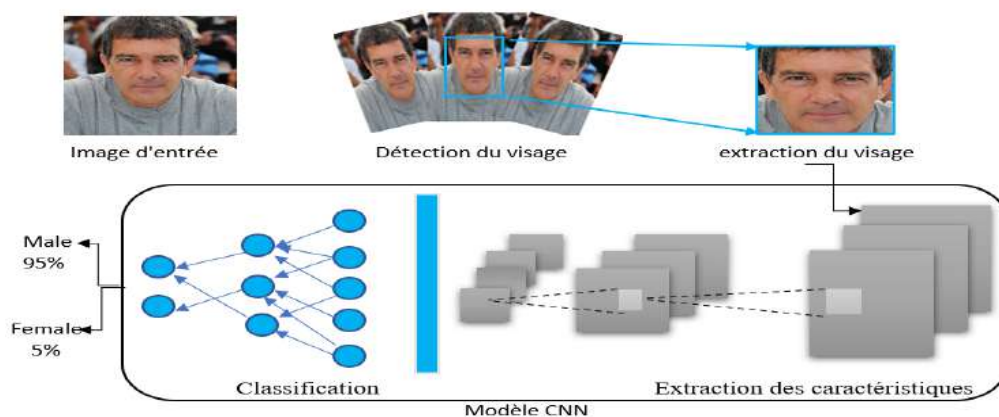


Fig 4.4 :Le processus de prédiction de genre par le modèle CNN.

4.5.1. L'effet de l'architecture

Au cours de nos expérimentations, nous avons créé trois modèles de configuration du réseaux convolutifs CNN avec différentes architectures entraînés sur la base d'image IMDB.

-Architecture du modèle N°1

Le premier modèle que nous présentons (figure 4.5) est composé de deux couches de convolution, deux couches de max-pooling et deux couches de fully connected (FC). L'image en entrée du réseau est de taille 28*28 en niveaux de gris, l'image passe d'abord à la première couche de convolution, cette dernière est composée de 16 filtres de taille 3*3, chacune des couches de convolution est suivie par une fonction d'activation ReLU qui force les neurones à retourner des valeurs positives. Les 16 carte de caractéristiques (feature maps) issues de l'étape de convolution, sont ensuite sous-échantillonnée par l'opérateurs max-pooling. Nous répétons ce processus avec une autre couche de convolution avec 32 filtres de taille 3*3 suivie par max-pooling.

Après ces deux couches de convolution et sous-échantillonnage, nous utilisons deux couches entièrement connectées (FC) avec l'appliquons de la fonction d'activation Relu dans la première couche de FC, et pour la deuxième couche une fonction sigmoid qui permet de calculer la distribution la probabilité des 2 classes (homme, femme).

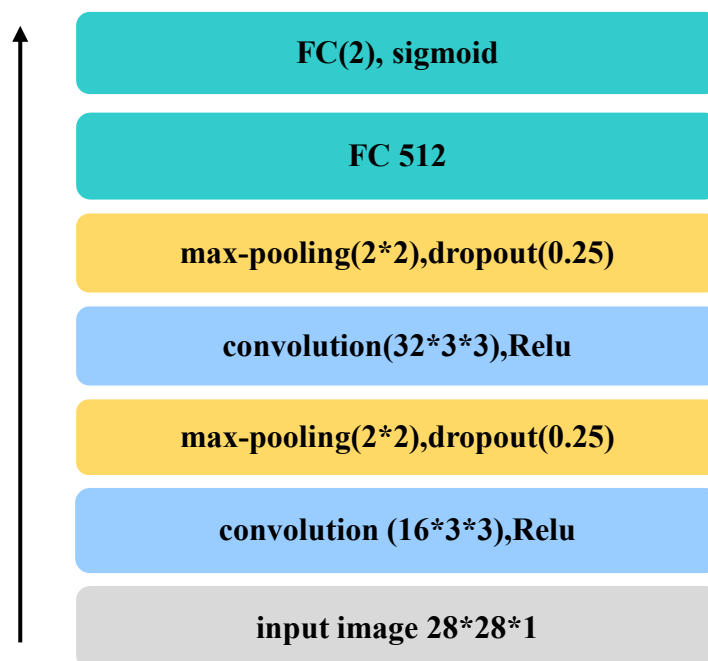


Fig 4.5 : L'architecture du réseau CNN (modèle n°1).

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 16, 28, 28)	160
max_pooling2d_1 (MaxPooling2)	(None, 16, 14, 14)	0
dropout_1 (Dropout)	(None, 16, 14, 14)	0
conv2d_2 (Conv2D)	(None, 32, 14, 14)	4640
max_pooling2d_2 (MaxPooling2)	(None, 32, 7, 7)	0
dropout_2 (Dropout)	(None, 32, 7, 7)	0
flatten_1 (Flatten)	(None, 1568)	0
dense_1 (Dense)	(None, 512)	803328
dropout_3 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 2)	1026
activation_1 (Activation)	(None, 2)	0

Fig 4.6 : La configuration du modèle n°1.

-Architecture du modèle N° 2

Le deuxième modèle que nous présentons (figure 4.7) est similaire au modèle 1 mais nous avons augmenté le nombre des filtres, l'architecture est composée de deux couches de convolution, deux couches de max-pooling et deux couches de fully connected. L'image en entrée est de taille 28*28 en niveau de gris, l'image passe d'abord à la première couche de convolution. Cette couche est composée de 32 filtres de taille 3*3, la fonction d'activation ReLU est utilisée, après cette convolution 32 de feature maps de taille 28*28 seront créés. Ensuite, les 32 feature maps qui sont obtenus, passent par une étape de max-pooling, puis on refait le processus de convolution et max-pooling avec un 64 filtres de convolution. À la sortie de la dernière couche, nous obtenons 64 feature maps de taille 7*7. Le vecteur de caractéristiques issu des convolutions a une dimension de 3136, après nous utilisons un réseau de neurones composé deux couches entièrement connectées (FC).

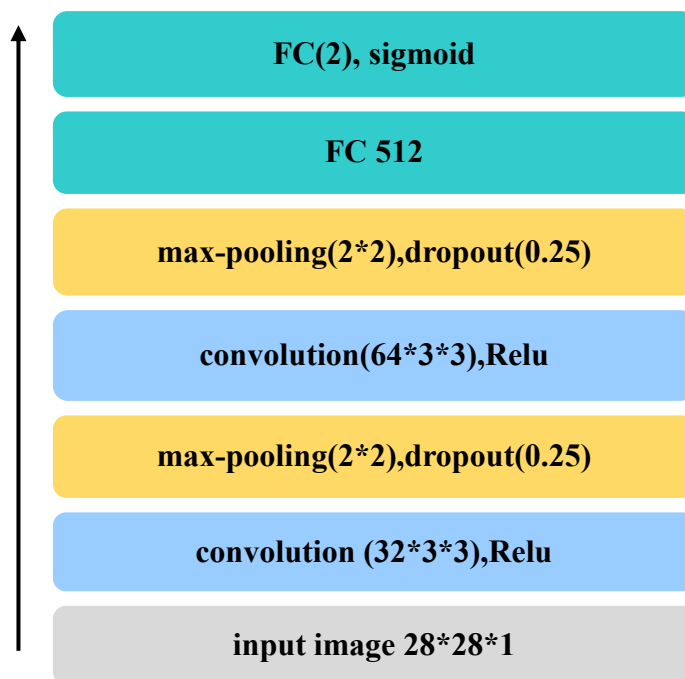


Fig 4.7 : L'architecture du modèle CNN (modèle n°2).

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 32, 28, 28)	320
max_pooling2d_1 (MaxPooling2)	(None, 32, 14, 14)	0
dropout_1 (Dropout)	(None, 32, 14, 14)	0
conv2d_2 (Conv2D)	(None, 64, 14, 14)	18496
max_pooling2d_2 (MaxPooling2)	(None, 64, 7, 7)	0
dropout_2 (Dropout)	(None, 64, 7, 7)	0
flatten_1 (Flatten)	(None, 3136)	0
dense_1 (Dense)	(None, 512)	1606144
dropout_3 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 2)	1026
activation_1 (Activation)	(None, 2)	0

Fig 4.8 : La configuration du modèle N°2.

-Architecture du modèle N°3

Le troisième modèle que nous présentons dans la figure (4.9) est plus profond par rapport aux modèles 1 et 2, il est composé de trois couches de convolution, trois couches de max-pooling et deux couches de fully connected.

L'image en entrée est de taille 28*28, l'image passe d'abord à la première couche de convolution. Cette couche est composée de 32 filtres de taille 3*3 ensuite une étape de sous-échantillonnage de max-pooling. On répète ce même processus avec les couches de convolutions deux et trois. À la sortie de la dernière couche max-pooling, nous aurons 64 feature maps (carte de caractéristiques) de taille 7*7.

Les vecteurs de caractéristiques issus des convolutions a une dimension de 3136. Après ces trois couches de convolution, nous utilisons un réseau de neurones composé deux couches fully connected (FC), la dernière couche FC contient 512 neurones où nous appliquons une fonction sigmoïde afin de classer l'image en deux classe femme ou homme.

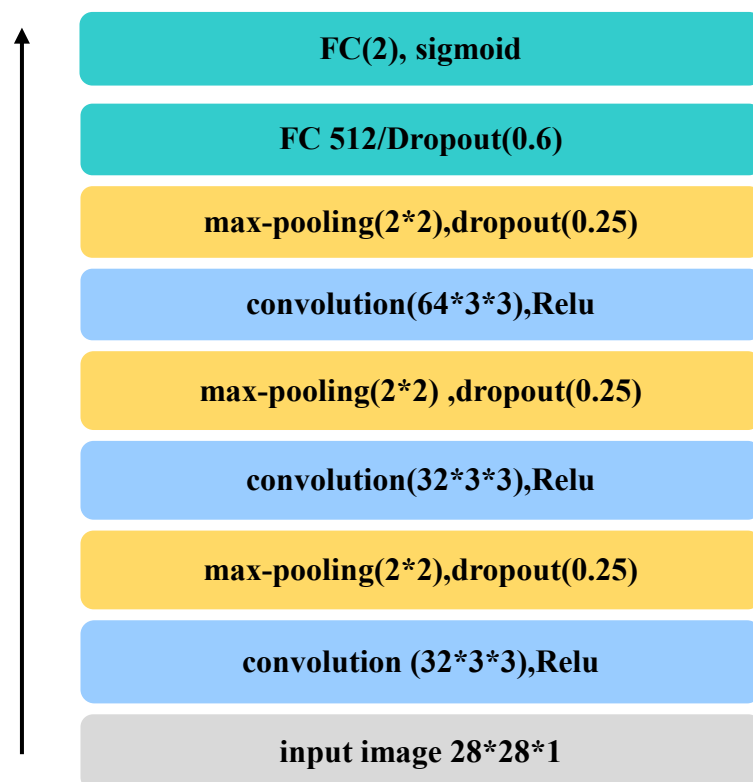


Fig 4.9 :L'architecture du modèle CNN (modèle n°3).

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 32, 28, 28)	320
max_pooling2d_1 (MaxPooling2)	(None, 32, 28, 28)	0
conv2d_2 (Conv2D)	(None, 32, 28, 28)	9248
max_pooling2d_2 (MaxPooling2)	(None, 32, 14, 14)	0
dropout_1 (Dropout)	(None, 32, 14, 14)	0
conv2d_3 (Conv2D)	(None, 64, 14, 14)	18496
max_pooling2d_3 (MaxPooling2)	(None, 64, 7, 7)	0
dropout_2 (Dropout)	(None, 64, 7, 7)	0
flatten_1 (Flatten)	(None, 3136)	0
dense_1 (Dense)	(None, 512)	1606144
dropout_3 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 2)	1026
activation_1 (Activation)	(None, 2)	0

Fig 4.10 : La configuration du modèle CNN N°3.

Les résultats obtenus par les trois architectures CNN proposées sur la base de données IMDB sont présentés sur le tableau suivant (Tab.4.2).

Modèle	Nbr data train	Nbr data test	Précision obtenue sur la base d'apprentissage	Précision obtenue sur la base de test	Valeur de perte d'apprentissage	Valeur de perte test
N°1	2000	500	100%	87.90%	0.006	0.992
N°2	2000	500	100%	91.80%	0.005	0.224
N°3	2000	500	100%	93.20%	0.0021	0.209

Tab 4.2 : les résultats de reconnaissance de genre obtenus sur IMDB

Le tableau 4.2 illustre les taux de classification obtenus sur les trois modèles utilisés, le nombre de données d'apprentissage de la base IMDB est 2500 échantillons réparties entre test et apprentissage, les résultats obtenus sont exprimés en termes de taux de précision en phase d'apprentissage et de test ainsi que la valeur de perte.

Le modèle N°3 a présenté les meilleurs résultats cela est due à la profondeur du réseau et le nombre de couches de convolution utiliser dans ce modèle par rapport aux (modèle N°1 et N°2). Ainsi, le modèle N°2 présente des résultats acceptables par rapport au modèle N°1, la

différence repose sur le nombre des filtres utilise. Dans le premier modèle, on a employé 16 filtres à la première convolution et au niveau de modèle 2 on a augmenté le nombre de filtres jusqu'à 32, on remarque que la performance du réseau CNN augmente lorsque le nombre des filtres est plus volumineux pour obtenir plus de caractéristiques afin de faciliter la tâche de classification.

La valeur de perte est aussi un élément critique dans les réseaux neurones, ce qu'il nécessite de minimiser cette valeur afin d'éliminer de nombre des caractéristiques perdues c'est pourquoi on a inséré des dropout avec des valeurs variée.

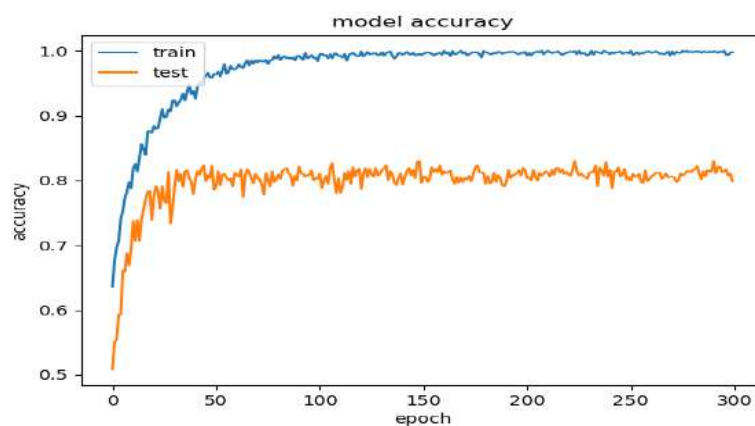


Fig 4.11 : Taux de précision du modèle 1.

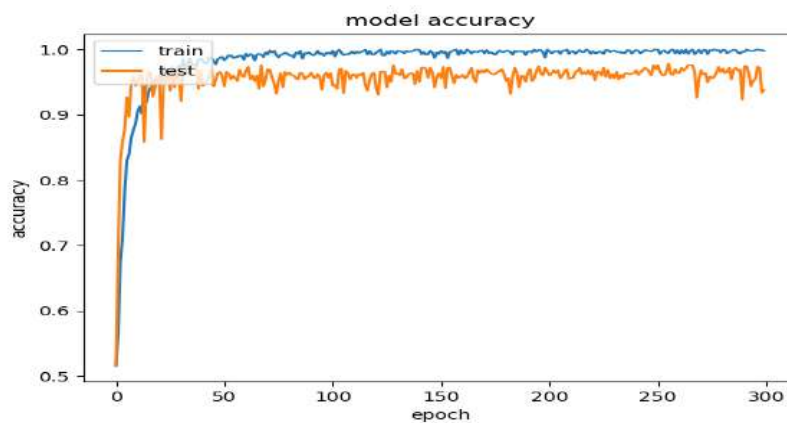


Fig 4.12 : Taux de précision du modèle 2.

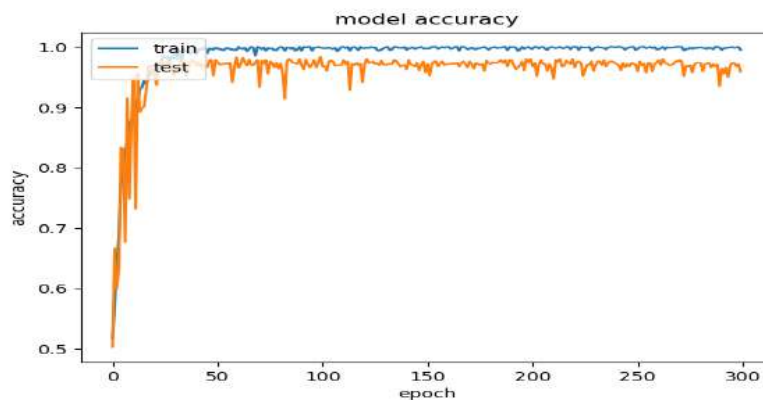


Fig 4.13 : Taux de précision du modèle 3.

Après l'analyse des résultats obtenus (les figures 4.11, 4.12, 4.13), on constate que :

- La précision d'apprentissage et test augmente avec le nombre d'époque, ceci reflète qu'à chaque époque le modèle apprend plus d'informations.
- Le modèle N°3 est plus performant avec un taux de précision en phase de test égale à 93.20%, suivie par le modèle N°2 et le modèle N°1 le moins performant avec 87.90%.

2.5.1 Effet du nombre d'images d'apprentissage sur le taux de reconnaissance

Dans cette expérimentation, nous avons changé le nombre d'images d'apprentissage et de test afin de voir l'effet de la quantité d'image d'apprentissage sur le taux de classification. Dans cette partie nous reprenons l'architecture du modèle n°3 qui sera entraîné sur la base IMDB avec différentes galeries comprenant une variation dans le nombre d'image d'apprentissage et de test, le tableau 4.3 résume nos résultats.

Galleries	Images d'apprentissage	Images de test	La précision de test
Galerie 1	20	80	65%
Galerie 2	400	100	81.2%
Galerie 3	800	200	91.1%
Galerie 4	2000	500	93.2%

Tab 4.3 : Taux de précision par rapport le nombre d'entraînement

D'après les résultats obtenus dans le tableau 4.3, nous constatons que la précision de reconnaissance est positivement proportionnel au nombre d'images d'apprentissage et de test (c'est-à-dire que le taux de reconnaissance augmente avec l'augmentation de nombre d'images d'apprentissage et de test) jusqu'à atteindre une amélioration de précision significatif. Cette augmentation est représentée par la courbe de la figure 4.14.

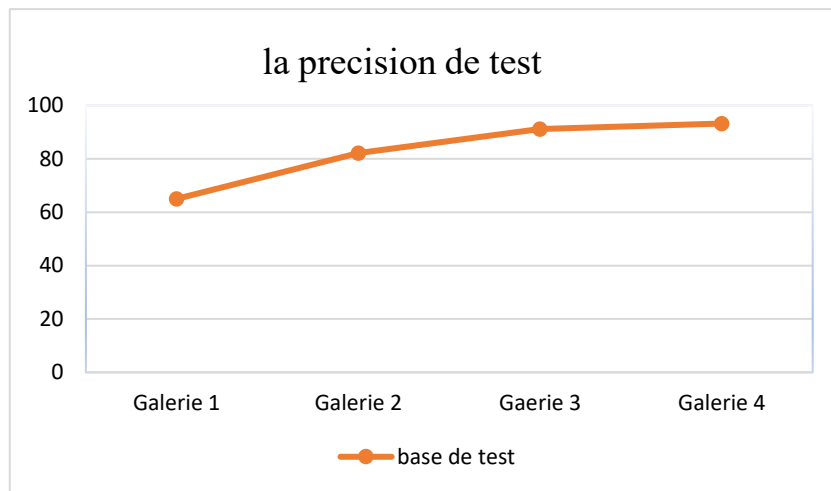


Fig 4.14: Evolution de la courbe d'apprentissage du modèle 3.

Alors d'après ces résultats obtenues (fig. 4.14) on peut dire que la base d'entraînement (l'apprentissage + test) est un élément déterminant dans les réseaux de neurones convolutifs, il faut avoir une base d'entraînement de grande taille (volume) pour aboutir à des meilleurs résultats.

A la suite de nos expérimentations et d'après les résultats obtenus précédemment, nous avons mentionné que le modèle N°3 présente les meilleurs taux de précision, afin de valider les performances de ce modèle, nous avons entraîné notre réseau CNN (modèle n3) sur les bases de données WIKI et IMDB avec un nombre d'échantillons égale à 1500 images, les résultats obtenus sont représentés dans le tableau ci-dessous qui présente la précision du modèle suivant le nombre d'images testées.

La base de données	Nombre des images teste	Nombre des images bien classe	Nombre des images mal classe	La précision
IMDB	500	472	28	94.49%
WIKI	500	467	33	93.56%

Tab 4.4 : Les résultats après l'entraînement du modèle 3.

D'après les résultats obtenue (Tab 4.4) on constate que le taux de reconnaissance du réseau CNN appliqué sur les bases WIKI et IMDB sont presque similaires et le nombre d'images mal classées sont des images faciales qui possèdent des éléments cachant le visage (occultation) tels que : les lunettes, variation des pose ...etc.

4.6. Système d'estimation d'âge

L'estimation automatique d'âge est un sujet d'actualité, mais reste une tâche relativement difficile en raison de la variation du processus de vieillissement. Différentes personnes ayant le même âge peuvent avoir une apparence nettement différente due au processus de vieillissement qui diffère d'une personne à une autre. En ce sens, nous avons appliqué le modèle N°3 afin d'estimer l'âge d'une personne selon trois plages d'âge : Young [20 -39ans], Middle [40-59 ans], old [60-100ans]. La figure 4.15 ci-dessous illustre l'architecture du modèle N°3 pour l'entraînement sur les bases WIKI et IMDB

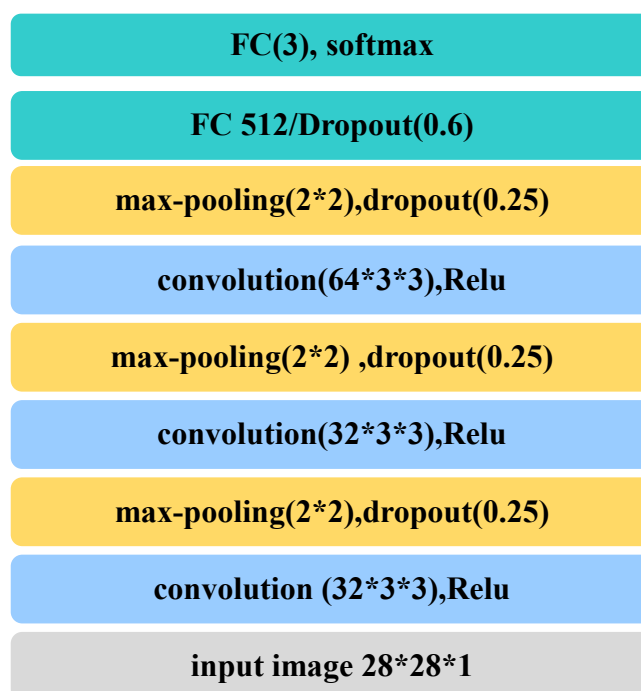


Fig 4.15 : L'architecture CNN pour l'estimation d'âge (modèle n°3).

La précision du modèle (taux de classification) après la phase d'apprentissage sur la base WIKI et IMDB est présentée dans le tableau 4.5.

Base de données	Image d'apprentissage	Image test	La précision d'apprentissage	La précision de test	Les images test bien classés	Les images test mal classés
WIKI	2000	500	100%	83.97 %	419	81
IMDB	2000	500	100%	86.20%	431	69

Tab 4.5 : Les taux de classification d'âge sur les bases de données WIKI et IMDB

Dans cette expérimentation on a appliqué l'architecture du modèle décrit dans la figure 4.15 pour l'estimation d'âge avec l'utilisation d'une fonction softmax dans la couche de sortie (couche de décision) , après l'entraînement du réseau CNN sur les bases de données WIKI et IMBD nous avons remarqué que les performances obtenues sont relativement acceptables avec un nombre d'échantillons totales égal à 3000 images, en tenant compte de la taille réduite d'images a 28×28 qui comporte un nombre limité des caractéristiques du visage. On a constaté aussi que la précision du modèle sur la base données IMDB est plus performante par rapport à la base de données wiki.

4.7. Système de reconnaissance de visage

Dans cette partie nous présentons la mise en œuvre d'un système de reconnaissance de visages humains basé sur les réseaux de neurones convolutifs. Dans ce type d'application on distingue deux modes opératoires d'un système de reconnaissance du visage.

- Mode vérification : ce mode consiste à comparer une image à autre stocké dans la base de donnes.
- Mode d'identification : ce mode consiste à comparer une image avec toutes les images de la base de données.

Dans ce travail, nous nous somme intéresser au mode vérification (authentification), pour cette raison on a choisi l'architecture CNN modèle n°3 (figure 4.9) sur une partie de la base de données FERET, avec une séparation de la base de donnes suivant le protocole suivant :

- Images apprentissages : Les 6 premières images servent pour la phase d'apprentissage.
- Images tests : images de chaque individu nous ont servies pour la réalisation des différents tests.

Images totales d'apprentissage	Images totales test	La précision (%) d'apprentissage	La précision (%) image de test	La valeur de perte d'apprentissage	La valeur de perte de test
1000	250	100 %	93.27 %	0.002	0.42

Tab 4.6 : Taux de reconnaissance faciale dur la base de données FERET.

Nous constatons (tableau 4.6) que le résultat de l'utilisation d'un réseau de neurones convolutifs avec une grande quantité de données dans la phase d'apprentissage (1000 échantillons) donne des performances satisfaisantes, en termes de taux de reconnaissance et la vitesse de convergence.

4.8. Comparaison des performances

Dans cette partie, nous présentons une comparaison entre les résultats obtenus par notre système de prédiction du genre et l'estimation d'âge avec d'autres techniques comme le classificateur SVM ou des modèles par réseau convolutifs pré-entraîné comme ResNet. Le tableau 4.7 montre un résumé de certains résultats de prédiction du genre et d'âge sur certaine base de données faciale, en fonction de leurs précisions et la base d'apprentissage utilisés, on constate que notre modèle donne les meilleurs résultats en termes de taux de classification.

Méthode /Modèle	La base de données	La précision (Accuracy%) du genre	La précision (Accuracy %) d'âge
SVM [133]	ADIENCE	79.3 %	45.1 %
4cef-CNN [134]	ADIENCE	86.8 %	50.7 %
VGG-16[135]	ADIENCE	88.36 %	54.6 %
ResNET-34[135]	WIKI+IMDB	93.17 %	66.42 %
RoR-34[135]	WIKI+IMDN	93.24 %	66.74 %
<u>L'approche proposé</u>	WIKI	93.56 %	83.97 %
	IMBD	94.49 %	86.20 %

Tab 4.7: Comparaison des performances de certaines méthodes de prédiction du genre et d'âge citées dans l'états de l'arts.

4.9. Présentation de l'application

Nous avons tenu à réaliser une application avec les approches proposées dans ce travail avec une interface graphique afin de pouvoir visualiser les performances de notre système de reconnaissance facilement sans être obligé de passer à chaque fois par les lignes de codes. Cette interface présente le menu principal de notre application, une sorte de passerelle pour accéder aux systèmes à partir duquel nous choisissons le type de reconnaissance à utiliser :

La fenêtre principale contient quatre buttons :

1. « **Face Recognition** » : système de vérification des individus.
2. « **Test Image** » : système de prédiction genre et l'estimation d'âge.
3. « **Real Time** » : système de prédiction du genre et l'estimation d'âge en temps réel.
4. « **Close** » : pour fermer l'application.



Fig 4.16 : Fenêtre d'accueil de l'application.

Dès qu'on clique sur « **face recognition** » on aura le menu présenté dans la figure 4.17 :

- Le Button « **Image test** » permet de charger deux images à partir d'un répertoire et affiche le chemin de répertoire dans lesquels elle a été stockée (figure 4.18).
- Le Button « **Diagram** » présente la performance obtenue sur la base FERET (figure4.20).
- « **Return** » et « **close** ».



Fig 4.17 : Fenêtre qui illustre le module de vérification.

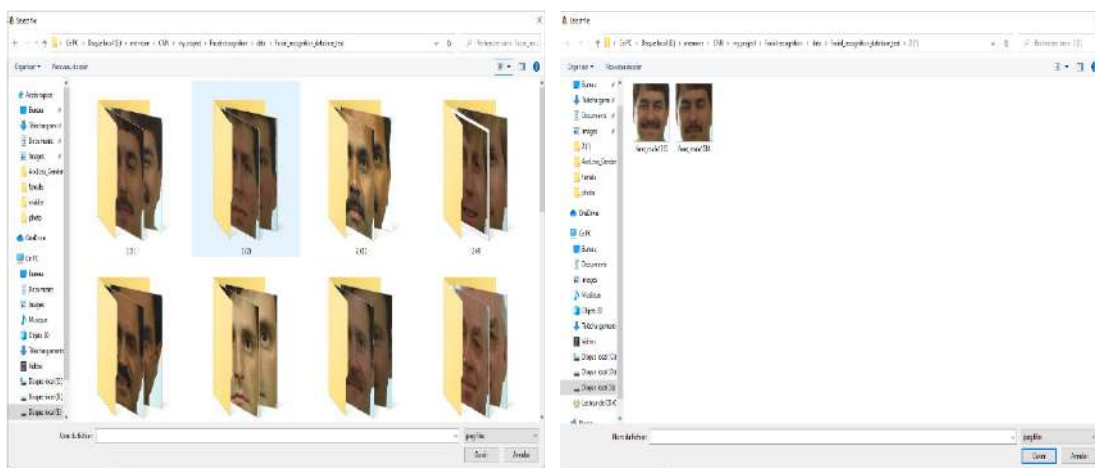


Fig 4.18 : L'emplacement des photos.

Après le choix des photos, nous aurons par la suite comme résultats d'identification soit :

- « *same individual* » personne bien vérifiée.
- « *different individual* » personne non vérifiée.



Fig 4.19 : Cas d'une personne bien identifiée

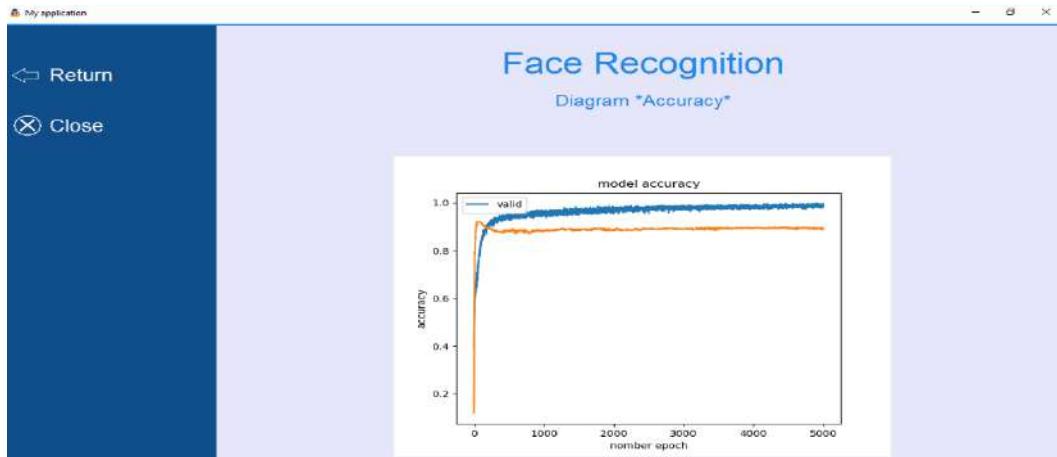


Fig 4.20 : La performance de système de reconnaissance de visage

Après avoir terminé la visualisation de la partie reconnaissance faciale, le bouton « *close* » ferme la fenêtre de système de reconnaissance de visage pour revenir à la fenêtre d'accueil.

Lorsque on clique sur « image test » dans la fenêtre d'accueil on passera au système de prédiction du genre et d'âge à partir d'une image de visage et on aura une interface qui possède quatre buttons :

- « **Gender** » un système de prédiction du genre.
- « **Age** » un système d'estimation d'âge.
- « **AgeGender** » un système qui combine les résultats de prédiction du genre et l'estimation d'âge.
- « **Diagram** » la précision du système sur la base WIKI et IMDB du genre et d'âge.

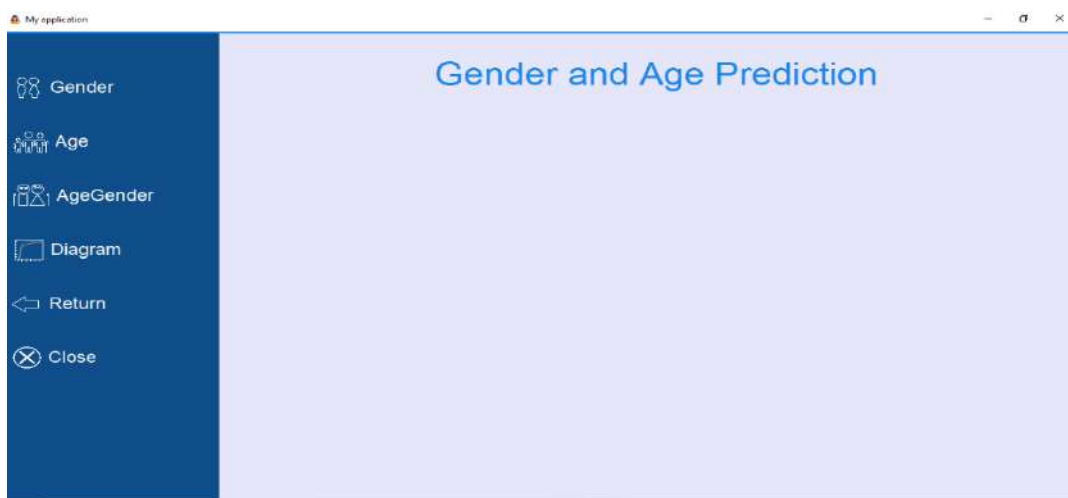


Fig 4.21 : L'interface de prédiction du genre et d'âge

Dès qu'on appuie sur le bouton « **Gender** », on passera à l'emplacement de la base WIKI, après on choisit une image afin de prédire le genre d'une personne et on aura comme résultat de classification « **female** » ou « **male** », un exemple est illustre dans la figure 4.22.



Fig 4.22 : Un exemple de prédiction du genre.

La deuxième interface du système « **image test** » sert à l'estimation d'âge qu'on a réalisé selon trois catégories « **Young [20-40] ans** », « **middle [41-60] ans** » et « **old plus de 60ans** », en cliquant sur « **Age** » (Fig 4.23) l'emplacement de la base wiki sera chargé afin de tester une image.



Fig 4.23 : Un exemple d'estimation d'âge.

Si vous voulez voir la prédiction du genre et l'estimation d'âge simultanément, il suffit de cliquer sur « **Age Gender** » dans la figure 4.21.

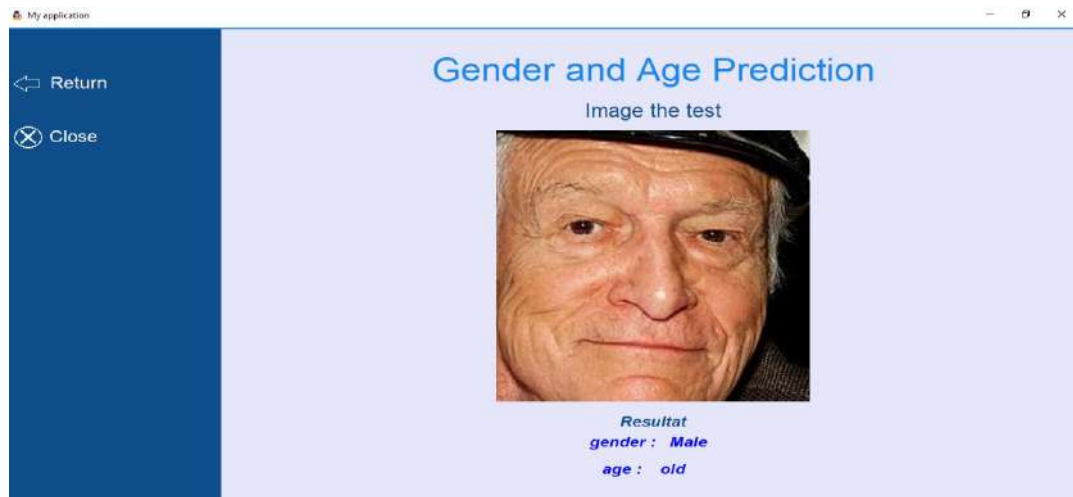


Fig 4.24 : Exemple de prédiction du genre et estimation d'âge simultanément.

Afin de visualiser la précision de système de prédiction du genre et l'estimation d'âge après l'entraînement sur la base de données WIKI et IMDB, il suffit de cliquer sur « *Diagram* » dans la fenêtre de figure 4.21.

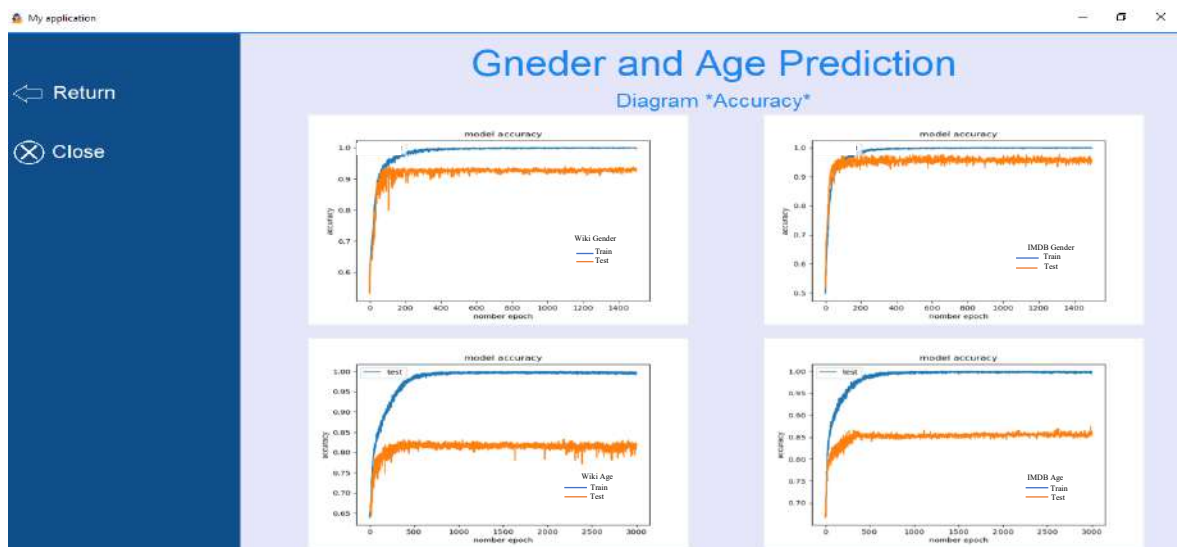


Fig 4.25 : Résultats de classifications sur la base de données wiki et IMDB.

Une autre partie de l'application est consacré à la prédiction du genre et estimation d'âge en temps réel, en cliquant sur « *real time* » dans la fenêtre d'accueil (figure 4.16).

Dès qu'on appuyé sur « *real time* » la caméra de PC va être activé pour détecter la présence d'un visage. Si le visage a été détecté, le programme va dessiner la zone d'intérêt dans un rectangle en bleu « l'emplacement du visage » ensuite va prédire le genre et l'âge du visage et affichera le nombre des visages détectés, la figure 4.26 ci-dessous montre un exemple.



Fig 4.26 : Exemple de prédiction du genre et d'âge en temps réel.

Si aucun visage n'a été détecté, le système affiche « no faces » (fig 4.27).

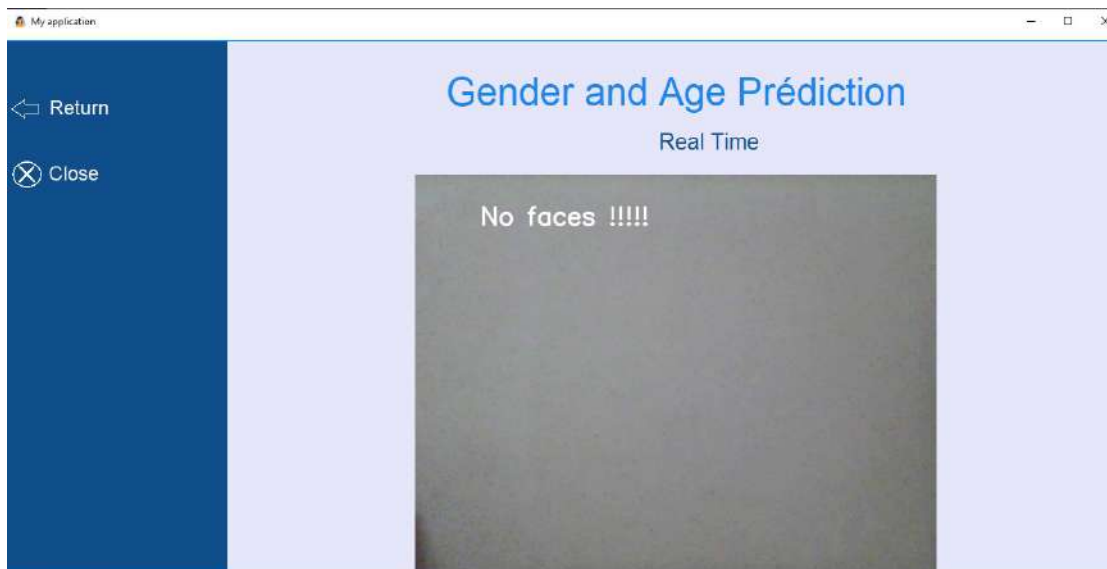


Fig 4.27: Le cas d'aucune personne détectée

À la lumière des résultats expérimentaux obtenus dans ce projet qui visent la reconnaissance de visage ainsi que la prédiction de genre et estimation d'âge, nous constatons que le taux de reconnaissance du réseau de neurones convolutif introduit dans ce travail, atteint une performance de classification très satisfaisante.

4.10. Conclusion

Nous avons étudié dans ce dernier chapitre une approche de classification automatique basée sur les réseaux de neurones à convolution, dans ce sens nous avons réalisé trois modèles avec différentes architectures pour plusieurs applications, de reconnaissance faciale, estimation d'âge et perdition du genre. Les différents résultats obtenus en termes de précision, comparés avec ceux cités dans l'état de l'art, ont montré que la profondeur des réseaux convolutifs est un facteur important pour aboutir à de meilleures précisions qui ont été validés sur les bases de données FERET, IMDB, et WIKI. Cette performance est due à la fois à la projection combinant le pouvoir représentatif et discriminant des réseaux CNN.

CONCLUSION ET PERSPECTIVES

Les systèmes biométriques jouent un rôle de plus en plus important dans la société moderne. La reconnaissance faciale en tant qu'une technologie biométrique de base a suscité un intérêt particulier, car elle fournit un moyen discret et non intrusif de détection, d'identification et de vérification sans avoir besoin de la connaissance ou de consentement du sujet.

L'objectif de ce travail est de concevoir et d'implémenter une application de reconnaissance faciale, capable en temps réel de prédire le genre et l'âge d'une personne qui peut être utilisée dans les réseaux sociaux et les panneaux publicitaires en répondant à des exigences de rapidité et robustesse.

Cette étude a été divisée en trois parties : la première partie consiste à étudier les réseaux de neurones à convolution afin de proposer un modèle de prédiction du genre, la deuxième partie est consacré d'appliquer le modèle plus performant pour l'estimation d'âge et la reconnaissance de visage à partir d'une image faciale, et la troisième partie elle a pour but d'intégrer un système de prédiction du genre et d'âge en temps réel, nous avons validé les systèmes développer dans ce travail sur plusieurs bases de données WIKI, IMDB et FERET.

L'exploitation de ce travail nous a permis de découvrir plus profondément plusieurs aspects du développement d'une application complexe. Il nous a fallu d'abord nous renseigner sur le domaine de la biométrie, la reconnaissance de forme, et plus généralement de la vision par ordinateur, qui est un domaine en pleine expansion.

A cet effet, si la biométrie est un enjeu important au niveau économique, la recherche, en particulier dans le domaine de la reconnaissance des visages offre encore un champ d'investigations très ouvert pour l'amélioration de la performance des systèmes de détection et d'élargissement de domaine d'application de ces applications.

Enfin en guise de perspective, une extension de ce travail peut être envisagé par réalisation d'un système de détection et de reconnaissance du visage basée sur les CNN comme un extracteur des caractéristiques et le support vecteur machine comme un classificateur, une autre perspective serait le teste de notre approche sur d'autres bases de données faciales présentant de fortes variations dans l'éclairage et de la pose.

ANNEXE

Généralités sur le traitement d'images :

Définition de l'image : L'image numérique est l'image dont la surface est divisée en éléments de tailles fixes appelés cellules ou pixels, ayant chacun comme caractéristique un niveau de gris ou de couleurs prélevé à l'emplacement correspondant dans l'image réelle, ou calculé à partir d'une description interne de la scène à représenter.

Caractéristiques d'une image numérique :

L'image est un ensemble structuré d'informations caractérisé par les paramètres suivants :

- **Pixel :** Le pixel représente le plus petit point distinguable dans une image. Chaque pixel possède une teinte, c'est la juxtaposition des différents pixels qui produit une image.
- **Dimension :** C'est la taille de l'image. Cette dernière se présente sous forme de matrice dont les éléments sont des valeurs numériques représentatives des intensités lumineuses (pixels). Le nombre de lignes de cette matrice multipliée par le nombre de colonnes nous donne le nombre total de pixels dans une image
- **Résolution** est mesurée en pixels par pouce (ppp). Elle décrit la clarté ou finesse de détail atteinte par un moniteur ou une imprimante dans la production d'images.
- **Définition :** Le nombre de pixels (de points) constituant l'image, c'est-à-dire le nombre de colonnes de l'image que multiplie son nombre de ligne.
- **Un bruit** dans une image est considéré comme un phénomène de brusque variation de l'intensité d'un pixel par rapport à ses voisins, il provient de l'éclairage des dispositifs optiques et électroniques du capteur.
- **Image en niveau de gris :** Les images réalisées en niveau de gris utilisent les teintes grises allant du blanc au noir. Le nombre de niveaux de gris dépend du nombre de bits utilisés pour décrire la couleur de chaque pixel de l'image. Par exemple pour 256 niveaux de gris compris entre 0 et 255, chaque pixel sera codé sur 8 bits, 0 représente le noir et 255 le blanc.
- **Luminance :** C'est le degré de luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface
- **Contraste :** C'est l'opposition marquée entre deux régions d'une image. Plus précisément, il quantifie la différence de luminosité entre les parties claires et sombres d'une image.

BIBLIOGRAPHIE

- [1] A. Join, A. Ross «introduction to biometrics in: handbook of biometrics», Springer, 2008.
- [2] S. Idriguen, W. Bahloul « Réalisation d'un Système de Reconnaissance Biométrique Multimodal », mémoire master, Université Akli Mohand Oulhadj de Bouira,2018.
- [3] <https://www.gemalto.com/france/gouv/inspiration/biometrie>. Consulte le 07/02/2020.
- [4] G. Roethenbaugh «An Introduction to Biometrics and General History Biometrics Explained » Section 1, 1998.
- [5] A. Bettahar, F. Saber « Extraction des caractéristiques pour l'analyse biométrique d'un visage », mémoire master université kasdi Merbah de Ouargla, 2014.
- [6] P. Verlinde, A Contribution to Multi-Modal Identity Verification Using Decision Fusion. Department of Signal and image Processing, U- Paris, 17 Septembre 1999.
- [7] N. Hezil « Méthode hybride en biométrie : Application à la paume de la main & l'Oreille » thèse de doctorat université 08 Mai 1945 de Guelma,2017.
- [8] Z. Salah « Méthode d'extraction des caractéristiques des images biométriques » mémoire master Université Abdelhamid Ibn Badis de Mostaganem,2017.
- [9] L.ZOUBIDA « Reconnaissance des Formes Appliquée à la Biométrie » thèse de doctoratUniversité Djilali Liabes de Sidi Bel Abbes,2018.
- [10] Technique de Contrôle d'Accès par Biométrie, http://www.clusif.fr_dossier_technique, CLUSIF, consulté le 20/01/2020.
- [11] <https://www.biometrie-online.net/biometrie> , consulté 24/02/2020.
- [12] I.Benchennane « Etude et mise au point d'un procédé biométrique multimodale pour la reconnaissance des individus » thèse de doctorat Université Mohamed Boudiaf d'Oran,2016
- [13] S. Yessaadi « Un modèle basé Template Matching/ Réseau de Neurones pour la Reconnaissance des visages », université Badji Mokhtar Annaba, 2006.
- [14] W. Andrew et al« Face recognition and its applications » IBM T.J. Watson Research Center- USA, 2003
- [15] X. Tana et al « Face recognition from a single image per person: A survey. Pattern Recognition» 2006.
- [16] A.J. O'Toole et al «Low-dimensional representation of faces in higher dimensions of the face space», 1993.
- [17] S. GUERFI « Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D » thèse Doctorat, l'Université Evry Val d'Essonne ,2008
- [18] X. Lu « Image Analysis for Face Recognition » université Michigan USA ,2003
- [19] M. Turk, al «Eigenfaces for Recognition» J. Cognitive Neuroscience, vol. 3, 1991
- [20] P. Belhumeur et al «fisherfaces: recognition using class specific linear projection», IEEE, 1997
- [21] A. Martinez, al « PCA versus LDA » IEEE Trans, 2001.
- [22] M. S. Bartlett, J. R. Movellan, et T. J. Sejnowski « Face recognition by independent component analysis. Transactions on Neural Networks » ,2002
- [23] I. Buciu, I. Pitas « Application of non-negative and local non negative matrix factorization to facial expression recognition » en International Conférence on Pattern Recognition, 2004

- [24] Y. Wang, Y. Jia, C. Hu et M. Turk « Non-negative matrix factorization Framework for face recognition » International Journal of Pattern Recognition and Artificial Intelligence, 2005.
- [25] M. Visani, C. Garcia et J. M. Jolion « Normalized radial basis function networks and bilinear discriminant analysis for face recognition». In IEEE ,2005.
- [26] H. Cevikalp, M. Neamtu, M. Wilkes et A. Barkana «Discriminative common vectors for face recognition»IEEE, 2005.
- [27] B. Scholkopf, A. Smola et K. Muller «Nonlinear component analysis as a kernel eigenvalue problem»Neural Computation, pages 1299–1319, 1998.
- [28] K. Hotta «View independent face recognition based on kernel principal component analysis of local parts» International Conference on Image Processing, pages III: 760–763, 2005
- [29] S. Mika et J. Weston « Fisher discriminant analysis with kernels Neural Networks for Signal Processing», 1999.
- [30] H. Kim, H. Park et H. Zha «Distance preserving dimension reduction for manifold learning » International Conference on Data Mining SIAM,2007.
- [31] S. Biswas, K. W. Bowyer, and Patrick J. Flynn. « Multidimensional scaling for matching low-resolution facial images. In Biometrics : Theory, Applications and Systems », 2010
- [32] M. H. Yang. « Face recognition using extended isomap» International Conference on Image Processing, pages 117–120, 2002.
- [33] G. Hagen, T. Smith, A. Banasuk, R.R. Coifman et I. Mezić «Validation of low-dimensional models using diffusion maps and harmonic averaging» IEEE Conference on Decision and Control, 2007.
- [34] J. Wang, C. Zhang, and Z. Kou «An analytical mapping for LLE and its application in multi-pose face synthesis» British Machine Vision Conference, 2003.
- [35] S. Pang & N. Kasabov « Investigating LLE eigenface on pose and face identification » International Symposium on Neural Networks, 2006
- [36] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang « Face recognition using laplacianface » IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005.
- [37] Y. Pang, Z. Liu et Y. Sun « Subspace learning based on laplacian eigenmaps and LDA for face recognition » International Journal on Information Acquisition,2006.
- [38] B. Raducanu et F. Dornaika « Dynamic facial expression recognition using laplacian eigenmaps-based manifold learning»International Conference on Robotics and Automation IEEE, 2010.
- [39] Q. Wang et J. Li « Combining local and global information for nonlinear dimensionality reduction. Neurocomputing », 2009.
- [40] M. N. Teli « Dimensionality reduction using neural networks. Technical report »,2008.
- [41] M. K. Fleming et G. W. Cottrell «Categorization of faces using unsupervised feature extraction ». International Joint Conference on Neural Networks IEEE, 1990.
- [42] S. Lawrence, C. L. Giles, A. C. Tsoi, et A. D. Back « Face recognition: A convolutional neural-network approach»Transactions on Neural Networks IEEE,1997
- [43] S. Duffner et C. Garcia «Face recognition using non-linear image reconstruction. In i-LIDS: Bag and Vehicle Detection Challenge», pages 459–464,2007.

- [44] R. Brunelli et T. Poggio « Face recognition : Features versus templates». Transactions on Pattern Analysis and Machine Intelligence IEEE, 1993
- [45] Y. Gao et M. K. H. Leung «Face recognition using line edge map» Transactions on Pattern Analysis and Machine Intelligence IEEE, 2002
- [46] B. Takács «Comparing face images using the modified hausdorff distance in Pattern Recognition», 1998
- [47] B. Heisele, P. Ho, J. Wu et T. Poggio. «Face recognition: component-based versus global approaches. Computer Vision and Image Understanding », 2003.
- [48] F. S. Samaria and A. C. Harter «Parameterisation of a stochastic model for human face identification » Proceedings of the 2nd IEEE workshop on Applications of Computer Vision, 1994.
- [49] T. F. Cootes, G. J. Edwards et C. J Taylor « Active appearance models » Transactions on Pattern Analysis and Machine Intelligence IEEE, 2001.
- [50] T. Ahonen, A. Hadid, and M. Pietikainen «Face description with local binary patterns: Application to face recognition » Transactions on Pattern Analysis and Machine Intelligence IEEE, 2006.
- [51] S. Liao, X. Zhu, Z. Lei, L. Zhang et S. Z. Li « Learning multi-scale block local binary patterns for face recognition » International Conference on Biometrics, 2007
- [52] P. S Penev et J. J Atick «Local feature analysis: a general statistical theory for object representation » Network : Computation in Neural System, 1996.
- [53] L. Wiskott, J.-M. Fellous, N. Kruger et C. von der Malsburg «Face recognition by elastic bunch graph matching» Transactions on Pattern Analysis and Machine Intelligence IEEE.
- [54] V. Perlibakas «Face recognition using principal component analysis and loggabor filters» CoRR, 2006.
- [55] A. Pentland, B. Moghaddam et T. Starner «View-based and modular eigenspaces for face recognition» Conference on Computer Vision and Pattern Recognition IEEE , 1994
- [56] T. F. Cootes et C. J. Taylor «Constrained active appearance models » International Conference on Computer Vision, 2001.
- [57] A. Lanitis, C. J. Taylor et T. F. Cootes «Automatic face identification system using flexible appearance models» Image and Vision Computing, 13(5) :393– 401, June 1995.
- [58] G. J. Edwards et C. J. Taylor and T. F. Cootes « Interpreting face images using active appearance models In Automatic Face and Gesture Recognition», pages 300–305, 1998
- [59] B. Dhikra, « Reconnaissance de visage basée sur une approche triangulaire», mémoire Master Université de 8 Mai 1945 - Guelma, 2019
- [60] H. Zedadra et al « Réalisation d'un système de reconnaissance de visage basée sur les caractéristiques biométriques (SkinMap, EyeMap, MouthMap) : application de la vidéosurveillance » Mémoire Master, Université 8 mai 1945-Guelma, 2016
- [61] M. Kirby et al «Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces» Transactions on Pattern Analysis and Machine Intelligence IEEE, 1990
- [62] I. Craw et al « Face recognition by computer » Proceedings British Machine Vision Conference 1996.
- [63] M. Baback et al « Probabilistic Visual Learning for Object Representation» Transactions on Pattern Analysis and Machine Intelligence IEEE, 1997
- [64] I. Biederman et al « psychophysical analysis of object and face recognition. In Face Recognition: From Theory to Applications». 1998.

- [65] K.Etemad et al « Discriminant analysis for recognition of human face images » 1997.
- [66] W.Zhao et al «Robust image based 3d face recognition », University of Maryland at College Park, College Park, 1999 .
- [67] K.JieWang et al « Pattern Recognition Letters Selecting discriminant eigenfaces for face recognition » 2005.
- [68] T.Kanade «Computer recognition of human faces»,1973
- [69] M. Kelly «Visual identification of people by computer»Tech. Rep. AI-130, Stanford Aiproject,1970
- [70] J. Cox et al «Feature-Based Face Recognition Using Mixture-Distance», Proceedings Conference on Computer Vision and Pattern Recognition,1996
- [71] B. S Manjunath et al «feature based approach to face recognition» Conference on Computer Vision and Pattern Recognition IEEE,1992.
- [72] K.Okada et al «The Bochum/USC Face Recognition System and how it fared in the FERET Phase III Test In Face Recognition From Theory to Applications»,1998
- [73] L. Wiskott et al «Face Recognition by Elastic Bunch Graph Matching» Transactions on Pattern Analysis and Machine Intelligence IEEE, 1997.
- [74] A. V.Nefian et al «Markov models for face recognition» International Conference on Acoustics, Speech and Signal Processing,1998.
- [75] F.Samaria, « Face recognition using hidden markov models » thèse de doctorat University of Cambridge-U.K1994.
- [76] F.Samaria et al «HMM based architecture for face identification »Image Vis Comput,1994.
- [77] S.Lawrence et al«Face recognition: A convolutional neural-network approach». IEEE Trans Neural Netw,1997
- [78] A.Pentland et al «View-based and modular eigenspaces for face recognition» Conference on Computer Vision and Pattern Recognition IEEE. 1994.
- [79] A. Lanitis et al « Automatic face identification system using flexible appearance models» Image Vis,1995.
- [80] P.Penev et al « Local feature analysis: A general statistical theory for object representation»Netw Computat Neural Syst,1996
- [81] J.Huang et al « Component-based face recognition with 3D morphable models » International Conference on Audio- and Video-Based Person Authentication,2003.
- [82] Hui-Cheng Lain et Bao-Liang Lu « Multi-View Gender Classification using Local Binary Patterns and Support Vector Machines» International Conference on Neural Networks, 2006.
- [83] Jing Wu, W. A. P. Smith et E. R. Hancock« Gender Classification using Shape from Shading» International Conference on Image Analysis and Recognition, 2008.
- [84] K.Ueki, T.Hayashida et T.Kobayashi « Subspace- based Age-group Classification using Facial Images under Various Lighting Conditions » International Conference on Automatic Face and Gesture Recognition, 2006.
- [85] A.Golomb, T.Lawrence et T.Sejnowski « A Neural Network Identifies Sex From Human Faces» in Neural Information Processing Systems, USA, 1990.

- [86] A.Khan, A.Majid et A.Mirza « Combination and optimization of classifiers in gender classification using genetic programming» International Journal of Knowledge-based and Intelligent Engineering Systems,2005
- [87] M. K. Yamaguchi, T. Kato « Age and sex information »IEICE Trans, 1996.
- [88] D. M. Burt and D. I. Perrett « Preception of age in adult caucasian male faces: computer graphic manipulation of shape and colour information in Perception», 1995.
- [89] K. Ueki, T. Hayashida, et T. Kobayashi « Subspace-based age-group classification using facial images under various lighting conditions» IEEE Int. Conf. on Automatic Face and Gesture.
- [90] Kazuya Ueki, Teruhide Hayashida and Tetsunori Kobayashi. “Subspace-based age-group classification using facial images under various lighting conditions”. In: Proceedings of Automatic Face and Gesture Recognition. Malaga, Spain, 2006.
- [91] Huiyu Zhou, Paul C Miller and Jianguo Zhang. “Age classification using Radon transform and entropy based scaling SVM.” In: Proceedings of British Mashine Vision Conference. Dundee, UK, 2011.
- [92] T. Kanno, M. Akiba et Y. Teramachi «Classification of age group based on facial images of young males by using neural networks » IEICE Trans,2001
- [93] Y. H. Kwon et N. Lobo «Age classification from facial images» 1994.
- [96] C.Shan «Learning local binary patterns for gender classification on realworld face images. Pattern Recognition » 2012.
- [97] H.Shih « Robust gender classification using a precise patch histogram. Pattern Recognition »2013
- [98] J.Tapia, et A.Perez « Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of lbp, intensity, and shape. » IEEE Transactions on Information Forensics and Security, 2013.
- [99] C.Bekios, J. Buenaposada et J.M., Baumela « Robust gender recognition by exploiting facial attributes dependencies» Pattern Recognition ,2014.
- [100] G.Levi, et T.Hassner «Age and gender classification using convolutional neural networks» Workshop on Analysis and Modeling of Faces and Gestures at CVPR, IEEE, 2015
- [101] L.Khoa et K. Ricanek ,T.Bui «Age estimation using active appearance models and support vector machine regression» Biometrics : Theory, Applications and Systems USA, 2009.
- [102] G.Guodong et M.Guowang «Human age estimation: What is the influence across race and gender » Computer Vision and Pattern Recognition Workshops2010.
- [103] L.Khoa , K.Seshadri et M.Savvides «Contourlet appearance model for facial age estimation » International Joint Conference on Biometrics,2011.
- [104] G.Guodong et G.Guowang Mu «A framework for joint estimation of age, gender and ethnicity on a large database” Image and Vision Computing,2014.
- [105] H.Yang, B.Lin ,K. Chang et C.Chen «Automatic Age Estimation from Face Images via Deep Ranking” British Mashine Vision Conference,2015
- [106] Jean-Louis Lauriere « Intelligence Artificielle Résolution de problèmes par l’Homme et la machine »
- [107] S.Duffner « Face Image Analysis With Convolutional Neural Networks », 2007
- [108] Jeff Heaton et AIFH « Deep Learning and Neural Network » Vol :3ISBN 978 1505714340, p10-14, 2015.
- [109] N. IALLO et A. Dian « la reconnaissance par l’expression faciale » mémoire de Master p16,2019.

- [110] Mokri Mohammed Zakaria « Classification des images avec les réseaux de neurones convolutionnels », p17,2017.
- [111] D.Moualek « Deep Learning pour la classification des images » mémoire master Université Abou Bakr Belkaid - Tlemcen2017
- [112] T.Stérin « Réseaux de neurones récurrents et mémoire : application à la musique »,2016
- [113] <https://medium.com/datadriveninvestor/a-high-level-introduction-to-lstms-34f81bfa262d> consulte le 01/05/2020
- [114] S.Hochreiter et F.at « long short-term memory neural computation »
- [115] D. H. Ackley, G. E. Hinton et T. J. Sejnowski « A Learning algorithm for boltzmann machines » Cognitive science, vol. 9, no. 1, 1985.
- [116] R. Salakhutdinov, A. Mnih et G. Hinton « Restricted boltzmann machines for collaborative filtering » international conférence on Machine Learning, ACM,2007.
- [117] G. E. Hinton, S. Osindero et Y.-W. Teh « A fast learning algorithm for deep belief nets »Neural computation, vol. 18, no. 7, 2006.
- [118] R. Salakhutdinov et G. Hinton « Deep boltzmann machines » in Artificial Intelligence andStatistics, pp. 448–455, 2009
- [119] <https://book.huihoo.com/deep-learning/version-30-03-2015/> consulté 05/04/2020
- [120] Grigory Antipov « Deep learning for semantic description of visual human traits. Neural and Evolutionary Computing » ,2017
- [121] Y. Lecun, L. Bottu, Y. Bengio et P.Haffener « Gradient based learning applied to document recognition »,1998
- [122] M. Zahir, N. Fazira, Zaidah Ibrahim and Nurbaity Sabri « Evaluation of Pre-Trained Convolutional Neural Network Models for Object Recognition »,2018
- [123] K. Simonyan, A. Zisserman et V.Geometry Group « very deep convolutional networks for large-scale image recognition »
- [124] C.Szegedy, W. Liu, Ch. Hill, P. Srmanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke et A Rabinovich « Going deeper with convolutions »
- [125] L. V. Fausett «Fundamentals of Neural Networks Architectures Algorithms and Applications ». Prentice-Hall, 1994.
- [126] M.K. Benkaddour, A. Bounoua. « Feature extraction and classification using deep convolutional neural networks, PCA and SVC for face recognition », Traitement du Signal, Vol. 34, No. 1-2, pp. 77-91, 2017.
- [127] Neurones à convolution. Article disponible sur :<https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html> Consulté le 14/06/2020.
- [128] F. Chabot « Analyse fine 2D/3D de véhicules par réseaux de neurones profonds » Thèse de doctorat université Clermont auvergne, France ,2018.
- [129] K. Simonyan et A. Zisserman « Very Deep Convolutional Networks for Large-Scale Image Recognition » ICLR, 2014.
- [130] <https://www.nist.gov/itl/products-and-services/color-feret-database> ,Consulté le 04/06/2020
- [131] [https://fr.wikipedia.org/wiki/Python_\(langage\)](https://fr.wikipedia.org/wiki/Python_(langage)) , Consulté le 04/06/2020

- [132] <https://medium.com/@diegoagher/what-can-you-do-with-keras-numpy-opencv-5309d01d8dab> Consulté le 04/06/2020
- [133] E. Eiding, R. Enbar et T. Hassner « Age and Gender Estimation of Unfiltered Faces » IEEE Trans. Inf. Forensics Security, 2014
- [134] G. Levi et T. Hassner « Age and Gender Classification Using Convolutional Neural networks » CVPR Workshop, 2015.
- [135] K. Zhang, Ce. Gao, L. Guo, M. Sun², X. Yuan², Tony X. Han², Zhenbing Zhao¹ Et Baogang Li¹ « Age Group and Gender Estimation in the Wild With Deep RoR Architecture », IEEE, 2017

RESUME

La reconnaissance des formes et la classification automatique sont des domaines de recherche très actifs, leurs principales objectives est de développer des systèmes intelligents capables d'apprendre et de reconnaître des objets. Une section essentielle de ces applications est attachée à la biométrie, cette dernière est employée à des fins de sécurité en général. La modalité faciale en tant qu'une technologie biométrique de base a pris une part de plus en plus importante dans le domaine de recherche.

L'objectif de ce travail est de développer un système de reconnaissance de visage, de prédiction du genre et d'estimation d'âge basé sur les réseaux de neurones convolutifs qui s'appuie sur une image du visage ou une vidéo en temps réel. Dans ce projet, on a réalisé trois modèles de réseau CNN avec différentes architectures (le nombre des filtres, le nombre des couches convolution...) valider sur plusieurs bases de données FERET, IMDB et WIKI, Les résultats obtenus ont montré que les réseaux CNN améliorent considérablement les performances du système ainsi que la précision de la reconnaissance.

Mots clés : *des systèmes intelligents, la biométrie, la reconnaissance de visage, les réseaux neurones convolutifs, la prédiction du genre, l'estimation d'âge.*

ABSTRACT

Pattern recognition and automatic classification are very active research areas, their main objectives are to develop intelligent systems capable of learning and recognizing objects. An essential section of these applications is attached to biometrics, which is used for security purposes in general. Facial modality as a basic biometric technology has become increasingly important in the field of research.

The goal of this work is the development of a system of facial recognition, gender prediction and age estimation based on convolutional neural networks based on a face image or real-time video. In this project, we realized three CNN models with different architectures (the number of filters, the number of convolution layers...) validate on several databases FERET, IMDB and WIKI, the results obtained showed that the CNN significantly improve the performance of system as well the accuracy of the recognition.

Keywords: *intelligent systems, biometrics, facial recognition, convolutional neural networks, gender prediction, age estimation.*

ملخص

إن التعرف على الأشكال والتصنيف التلقائي هما مجالان من مجالات البحث النشطة للغاية، وتتمثل أهدافهما الأساسية في تطوير الأنظمة الذكية القادرة على تعلم الأجسام والتعرف عليها. كما يتم إرفاق القسم الأساسي من هذه التطبيقات بالقياسات الحيوية، الذي يُستخدم لأغراض الأمان بشكل عام. وقد أصبحت طريقة التعرف على الوجه كتقنية حيوية أساسية ذات أهمية متزايدة في مجال البحث.

والهدف من هذا العمل تطوير نظام التعرف على الوجه والتنبؤ بنوع الجنس وتقدير العمر استناداً إلى الشبكات العصبية الالتفافية التي تعتمد على صورة الوجه أو فيديو في الوقت الحقيقي. في هذا المشروع حققنا ثلاثة نماذج لشبكة CNN مع هياكل مختلفة (عدد عوامل التصفية، عدد طبقات الالتفاف...) وذلك بالتحقق من صحة العديد من قواعد البيانات IMDB, WIKI و FERET أظهرت النتائج التي تم الحصول عليها أن شبكات CNN تحسن بشكل كبير الأداء نظام وكذلك دقة التعرف.

الكلمات مفتاحية: *الأنظمة الذكية، القياسات الحيوية، التعرف على الوجه، الشبكات العصبية الالتفافية، التنبؤ بنوع الجنس، تقدير العمر.*