

UNIVERSITE KASDI MERBAH OUARGLA
Faculté des Nouvelles technologies de l'information et de la
communication
Département D'informatique et de technologie de l'information



Mémoire Master Professionnel

Domaine : Mathématique et Informatique

Filière : Informatique

Spécialité : Administration et sécurité des réseaux

Présenté par : Gouamid Asma et Mammeri Abir

***Reconnaissance et classification des traits
caractéristiques biométriques faciale***

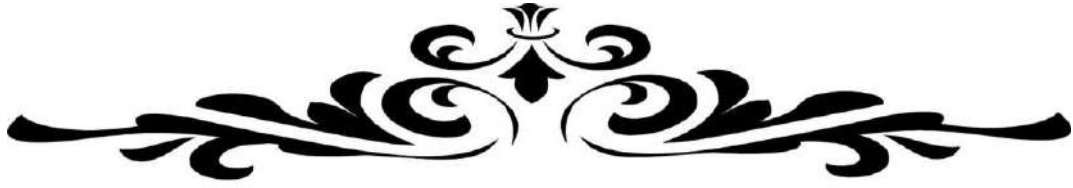
Soutenu publiquement le :

05 /10/2020

Devant le Jury :

Mr. XX xx	M.A Président	U.K.M Ouargla
Mr. BENKADDOUR Mohammed Kamel	M.A Encadreur	U.K.M Ouargla
Mr. XX xx	M.A Examineur	U.K.M Ouargla

Année Universitaire : 2019/2020



Remerciements

Avant tout, nous remercions Dieu tout puissant de nous avoir accordé la force, le courage, et les moyens afin de pouvoir accomplir ce modeste travail.

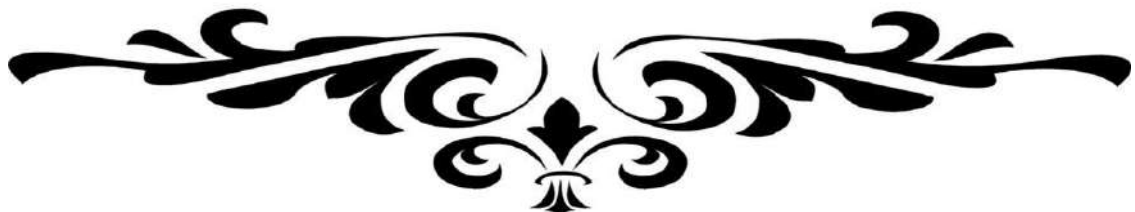
En premier lieu, nous remercions notre Encadreur Monsieur BENKADDOUR Mohamed Kamel, enseignant à l'Université KASDI MERBAH d'Ouargla pour nous avoir proposé ce thème et pour son encadrement. Sa grande connaissance dans le domaine, ainsi que son expérience, ont joué un rôle important dans la conception de ce travail.

Nos remerciements et nos respect vont également à Mr. BOUKHAMLA Akram d'avoir accepté de présider le jury, J'adresse mes remerciements aussi à Mr. KHALDI Amin qui m'ont fait l'honneur d'accepter le jugement de notre travail.

Nous adressons aussi nos sincères reconnaissances à tous les enseignants de département de l'informatique –Ouargla- qui nous ont donnés beaucoup de connaissance.

Un gros merci à nos familles, plus spécifiquement nos parents pour leur soutien.

Enfin, que tous ceux qui ont contribué de près ou de loin à la réalisation de ce mémoire, puissent trouver ici, toute notre reconnaissance.



Dédicace

Je dédie ce modeste travail

À la mémoire de mon père ma gratitude ne suffit pas à exprimer ce qu'elle mérite pour tous tes sacrifices depuis ma naissance, pendant mon enfance et même à l'âge adulte.

À ma chère mère. Merci pour tes conseils, tes sacrifices, ton soutien et tes encouragements.

À mon frère Salah.

À tous mes sœurs : Oum elkhair, Kaltoum, Safia, Sara, Zineb, Hafsa.

À mes grands-pères, à mes grandes mères, à toute ma famille

À tous mes amis, qui m'ont soutenu dans l'accomplissement de cet humble travail

À tous mes professeurs et à tous ceux qui se sont engagés dans ces modestes travaux

Asma

Dédicace

Je dédie ce modeste travail

*A mes parents .Aucun hommage ne pourrait être à la hauteur de l'amour
Dont ils ne cessent de me combler. Que Dieu leur procure bonne santé et
longue vie.*

*A celle que j'aime beaucoup et qui m'a soutenue tout au long de ce projet :
mon mari , et bien sûr à mes sœurs, sans oublier mes belles-sœurs que
j'aime.*

*A toute ma famille, et mes amis pour leurs soutiens et leurs
encouragements.*

*A mon binôme **ASMA** et toute la famille **MAMMERI**.*

*A tous mes professeurs et enseignants que j'ai eu durant tout mon cursus
scolaire et qui m'ont permis de réussir dans mes études.*

*Et à tous ceux qui ont contribué de près ou de loin pour que ce projet soit
possible, je vous dis merci.*

Abir

Sommaire

Liste des figures.....	iv
Liste des tableaux.....	v
Liste des équations.....	viii
Liste des acronymes.....	ix
INTRODUCTION GENERALE.....	1
Chapitre 01 : La biométrie & systèmes de reconnaissance faciale.....	3
1.1 Introduction.....	4
1.2 La biométrie.....	5
1.2.1 Définition.....	5
1.2.2 Pourquoi utiliser la biométrie ?.....	5
1.3 Panorama des différentes techniques biométries.....	5
1.4 Le mode de fonctionnement d'un système biométrique.....	7
1.5 Structure générale d'un système biométrie.....	8
1.6 Evaluation des performances des Systèmes biométriques.....	9
1.7 Applications des systèmes biométriques.....	10
1.8 Mesure de la performance d'un système biométrique.....	10
1.8.1 Performance d'un système d'identification.....	10
1.8.2 Performance d'un système de vérification.....	11
1.9 Reconnaissance de visage.....	13
• Motivation :.....	14
1.9.1 Les classes des techniques de reconnaissance de visage.....	14
1.9.2 Principales difficulté de la reconnaissance de visage.....	16
1.9.3 Avantages et inconvénients de la reconnaissance faciale.....	19
1.10 La reconnaissance de genre et l'estimation d'âge.....	19
1.10.1 La reconnaissance de genre.....	19
1.10.2 L'estimation d'âge.....	19
1.11 Conclusion.....	20
Chapitre 02 : Etat de l'art des techniques de réduction de dimension et classification d'image.....	21
2.1 INTRODUCTION.....	22
2.2 État de l'art des techniques de reconnaissance.....	22

2.3 Les méthodes de réduction de dimension.....	23
2.3.1 Méthodes linéaires de réduction de dimension	23
2.3.2 Méthodes non linéaires de réduction de dimension	29
2.4 Les techniques de classification d'images	32
2.4.1 Les réseaux de neurones artificiels (RNA).....	32
2.4.2 Machine à Vecteur de Support (SVM)	33
2.4.3 Le modèle AAM (Active Appearance Model).....	34
2.5 Conclusion.....	35
Chapitre 03 : Apprentissage Profond : réseaux de neurones convolutifs.....	37
3.1 Introduction.....	38
3.2 Intelligence artificielle.....	38
3.3 Apprentissage automatique (machine learning en anglais).....	39
3.4 Apprentissage profond.....	39
3.4.1 Pourquoi le deep learning ?.....	40
3.4.2 L'apprentissage.....	41
3.5 Réseau de neurone artificiel.....	43
3.5.1 Le neurone.....	43
3.5.2 Architecteurs des réseaux de neurones.....	46
3.5.3 Type des réseaux de neurones.....	47
3.5.4 L'apprentissage profond et Les réseaux de neurones artificiels.....	52
3.6 Les réseaux de neurones convolutifs (CNN)	52
3.6.1 Présentation.....	52
3.6.2 Architecture de réseaux de neurone convolutif.....	53
3.5.4 Quelques réseaux convolutifs célèbres	60
3.6 Conclusion.....	63
Chapitre 04 : Expérimentations, Résultats et Discussions	64
4.1 Introduction.....	65
4.2 Les bases de données de visage.....	65
4.2.1 ORL.....	65
4.2.2 FEI.....	66
4.2.3 Essex	66
4.2.4 Casia WebFace.....	67
4.2.5 Séparation des bases de données.....	67
4.3 Environnement du travail.....	68
4.3.1 Environnement matériel.....	68

4.3.2 Outils de développement	68
4.4 Système proposé : principes et expérimentations.....	68
4.4.1 L'approche de reconnaissance faciale basée sur le réseau CNN	69
4.4.2 L'application des techniques de réduction des dimensions	76
4.4.3 Discussion	77
4.5 Présentation de l'application	79
4.5.1 Interface de Présentation du projet	79
4.5.2 Interface de la reconnaissance faciale	80
4.5.3 Interface des méthodes de réduction des dimensions	81
4.6 Test de l'application sur la base Essex.....	82
4.7 Test des méthodes de réduction des dimensions sur la base FEI.....	84
4.8 Conclusion.....	85
CONCLUSION GENERALE.....	87
Référence bibliographies.....	89

Liste des figures

<i>Figure 1.1 : Caractéristiques biométriques.</i>	6
<i>Figure 1.2 : Principaux modules d'un système biométrique ainsi que les différentes mode.</i>	7
<i>Figure 1.3 : Illustration du FRR et du FAR.</i>	12
<i>Figure 1.4 : La courbe ROC.</i>	12
<i>Figure 1.5 : Classification des algorithmes principaux utilisés en reconnaissance facial</i>	15
<i>Figure 1.6 : Exemples de changement d'illumination.</i>	17
<i>Figure 1.7 : Exemples de variation de pose.</i>	17
<i>Figure 1.8 : Exemples de variation d'expressions.</i>	17
<i>Figure 1.9 : Exemples des Occultations partielles.</i>	18
<i>Figure 1.10 : Exemples de composants structurels.</i>	18
<i>Figure 2.1 : Les projections PCA et LDA d'un ensemble de données.</i>	28
<i>Figure 2.2 : Système de reconnaissance faciale utilisant ACP et LDA.</i>	29
<i>Figure 2.3 : Les étapes de la méthode de t SNE.</i>	30
<i>Figure 2.4 : t-SNE.</i>	30
<i>Figure 2.5 : Le fonctionnement de l'UMAP.</i>	31
<i>Figure 2.6 : UMAP vs t-SNE.</i>	31
<i>Figure 2.7 : Modèle d'un neurone artificiel.</i>	33
<i>Figure 3.1 : La relation entre l'intelligence artificielle, le ML et le Deep Learning.</i>	40
<i>Figure 3.2 : La différence de performance entre le Deep Learning et la plupart des algorithmes de ML en fonction de la quantité de données</i>	41
<i>Figure 3.3 : L'apprentissage profond vs machine learning</i>	41
<i>Figure 3.4 : L'apprentissage Supervisé</i>	42
<i>Figure 3.5 : L'apprentissage non Supervisé</i>	43
<i>Figure 3.6 : Le neurone biologique</i>	44
<i>Figure 3.7 : Le neurone artificiel</i>	44
<i>Figure 3.8 : Les modèles de fonctions d'activation.</i>	45
<i>Figure 3.9 : Correspondance entre le neurone biologique et le neurone formel.</i>	45
<i>Figure 3.10 : Architecture de réseau non bouclé</i>	46
<i>Figure 3.11 : Architecture de réseau bouclé.</i>	47
<i>Figure 3.12 : perceptron monocouche.</i>	48

<i>Figure 3.13 : perceptron multicouche</i>	49
<i>Figure 3.14 : Réseau de Hopfield</i>	49
<i>Figure 3.15 : Structure du modèle Adaline</i>	50
<i>Figure 3.16 : Architecture du modèle Adaline</i>	51
<i>Figure 3.17 : Le modèle de Kohonen</i>	51
<i>Figure 3.18 : L'architecture standard d'un réseau de neurone convolutionnels</i>	52
<i>Figure 3.19 : Schéma du parcours de fenêtre de filtre sur l'image</i>	54
<i>Figure 3.20 : L'opération de convolution</i>	55
<i>Figure 3.21 : Les deux méthodes de Pooling</i>	55
<i>Figure 3.22 : Pooling avec un filtre 2x2 et un pas de 2</i>	56
<i>Figure 3.23 : Fonction relu</i>	56
<i>Figure 3.24 : La Couche de Fully-Connected</i>	57
<i>Figure 3.25 : Schéma de deux réseaux de neurones avec et sans dropout</i>	59
<i>Figure 3.26 : L'architecture LeNet</i>	60
<i>Figure 3.27 : L'architecture AlexNet</i>	60
<i>Figure 3.28 : L'architecture ZFNet</i>	61
<i>Figure 3.29 : L'architecture GoogLeNet/ Inception</i>	61
<i>Figure 3.30 : L'architecture VGGNet</i>	62
<i>Figure 3.31 : L'architecture ResNet</i>	62
<i>Figure 4.1 : Exemples d'images de visages de la base ORL</i>	65
<i>Figure 4.2 : Exemples d'images de visages de la base FEI</i>	66
<i>Figure 4.3 : Exemples d'images de visages de la base Essex</i>	66
<i>Figure 4.4 : Exemples d'images de visages de la base CASIA WebFace</i>	67
<i>Figure 4.5 : l'architecture de notre réseau CNN</i>	71
<i>Figure 4.6 : le résultat de la fonction Softmax</i>	72
<i>Figure 4.7 : le résultat de la fonction Sigmoidé</i>	72
<i>Figure 4.8 : le résultat de classification avec 4 Couche (fonction Sigmoidé)</i>	73
<i>Figure 4.9 : Taux de reconnaissance obtenue sur les bases de données (CASIA Webface, FEI, ORL, Essex)</i>	Erreur ! Signet non défini.
<i>Figure 4.10 : Taux de perte des bases de données utilisées (CASIA Webface, FEI, ORL, Essex)</i>	Erreur ! Signet non défini.
<i>Figure 4.11 : Les résultats de visualisation des bases de données utilisées (CASIA Webface, FEI, ORL, Essex)</i>	Erreur ! Signet non défini.

Figure 4.12 : Les résultats de Regroupement des bases de données utilisées (CASIA Webface, FEI, ORL, Essex).....	Erreur ! Signet non défini.
Figure 4.13 : Interface graphique de notre application (menu principale).....	79
Figure 4.14 : La deuxième interface.....	80
Figure 4.15 : Interface de la reconnaissance faciale.	81
Figure 4.16 : Interface de la Réduction des dimensions.....	82
Figure 4.17 : message de choisir une image.....	82
Figure 4.18 : le choix d'image test.	83
Figure 4.19 : le choix du système de reconnaissance.....	83
Figure 4.20 : Résultat de teste.....	84
Figure 4.21 : Interface des méthodes de Réduction des dimensions.	84
Figure 4.22 : Résultat de teste.....	85

Liste des tableaux

Tableau 1.1 : Avantages et inconvénients de la reconnaissance faciale.....	19
Tableau 3.1 : Analogie entre neurone biologique et le neurone artificiel.....	46
Tableau 4.1 : Architecture détaillée de notre CNN utilisée pour l'extraction des caractéristiques.....	70
Tableau 4.2 : Taux de reconnaissance par rapport de nombre d'époques.....	73
Tableau 4.3 : Taux de reconnaissance selon le nombre d'exemple d'apprentissage.....	74
Tableau 4.4 : Taux de reconnaissance et de perte pour les différentes bases de données utilisées.....	74

Liste des équations

<i>Équation 1.1 : Taux d'indentification correct</i>	11
<i>Équation 1.2 : Taux des faux Rejects</i>	11
<i>Équation 1.3 : Taux des Fausses Acceptations</i>	11
<i>Équation 1.4 : Taux d'erreur égal</i>	13
<i>Équation 1.5 : Demi-taux d'erreur total</i>	13
<i>Équation 1.6 : Taux d'erreur total</i>	13
<i>Équation 2.1 : moyenne de visages</i>	26
<i>Équation 2.2 : Enlever de moyenne d'image</i>	26
<i>Équation 2.3 : la matrice de covariance</i>	26
<i>Équation 2.4 : Représentation des visages dans notre nouvel espace</i>	27
<i>Équation 2.5 : Choix de k vecteurs propres</i>	27
<i>Équation 3.1 : la somme du perceptron</i>	48
<i>Équation 3.2 : La sortie du perceptron</i>	48
<i>Équation 3.3 : Sortie du modèle Adaline</i>	50

Liste des acronymes

- **2D** : *Représentation à 2 Dimensions*
- **3D** : *Représentation à 3 Dimensions*
- **AAM** : *Active Appearance Model*
- **ACI** : *Analyse en Composantes Indépendantes*
- **ACP** : *Analyse en Composantes Principales*
- **ADALINE** : *Adaptive Linear Neuron*
- **ADN** : *Acide Désoxyribo Nucléique*
- **ANN** : *Artificial Neural Network*
- **AI** : *Artificial Intelligence*
- **BDD** : *Base De Donnée*
- **CNN** : *Convolutional Neural Network*
- **DBN** : *Deep Belief Network*
- **DL** : *Deep Learning*
- **DLDA** : *Direct Linear Discriminant Analysis*
- **EER** : *Equal Error Rate*
- **FA** : *Fausse Acceptation*
- **FAR** : *False Acceptance Rate*
- **FR** : *Faux Rejects*
- **FRR** : *False Rejection Rate*
- **FLD** : *Fisher linear Discriminant*
- **GDA** : *General Discriminant Analysis*
- **HTER** : *Half Total Error Rate*
- **IBG** : *International Biométrie Group*
- **KFDA** : *Kernel Fisher Discriminant Analysis*
- **LDA** : *Linear Discriminant Analysis*
- **LSTM** : *Long Short-Term Memory*
- **ML** : *Machine Learning*
- **MLP** : *Multi-Layer Perceptron*
- **ORL** : *Olivetti Research Laboratory*
- **PCA** : *Principal Component Analysis*
- **PMC** : *Perceptron Multi-Couche*
- **RBF** : *Radial Basis Function*

- **ReLU** : *Rectified Linear Unit*
- **RLDA** : *Regression Linear Discriminant Analysis*
- **RNA** : *Réseau de Neurones Artificiels*
- **RNN** : *Recurrent Neural Network*
- **ROC** : *Receiver Operating Characteristic*
- **SAM** : *Supervised Appearance Model*
- **SVC** : *Support Vectors clustering*
- **SVM** : *Support Vectors Machines*
- **TER** : *Total Error Rate*
- **t-SNE** : *t- Stochastic Neighbor Embedding*
- **UMAP** : *Uniform Manifold Approximation and Projection*
- **TIC** : *Taux d'Identificaion Correct*



INTRODUCTION

INTRODUCTION GENERALE

La sécurité des systèmes d'information est devenue un domaine de recherche d'une très grande importance. La conception d'un système d'identification fiable, efficace et robuste est une tâche prioritaire. L'identification de l'individu est essentielle pour assurer la sécurité des systèmes et des organisations. Elle correspond à la recherche de l'identité de la personne qui se présente dans une base de données et peut servir à autoriser l'utilisation des services. Chaque être humain peut dès son plus jeune âge, reconnaître les visages des personnes qui lui sont familières. En fait, comment peut-on de manière automatique par ordinateur reconnaître un individu, d'estimer de son âge et prédire de son genre par la seule prise en compte de l'image de son visage ? La réponse à cette seule question sera le fil directeur de notre travail.

Les visages humains peuvent être considéré comme des données biométriques qui véhiculent une grande quantité d'informations, comme par exemple la pose, l'identité, l'état émotionnel, le genre, l'âge, etc. L'utilisation de ces indices faciaux dans une interaction entre les humains et la machines est rendue possible par la remarquable capacité de ces derniers de reconnaître et interpréter les visages et leurs comportements. De manière schématique, la tâche de la reconnaissance de visage consiste à comparer une image d'entrée avec une base de données contenant un ensemble d'images de visage avec leurs identités, ensuite identifier le sujet auquel appartient le visage d'entrée. Cependant, leur implémentation reste difficile, parce que les visages font preuve de plusieurs variations significatives dans leur apparence, ces variations sont dues aux conditions d'acquisition des images, l'illumination, la position des visages et l'âge.

Plusieurs méthodes ont été développées dans la littérature pour la reconnaissance de visage et la classification de genre et d'âge. Dans notre travail nous avons opté pour une technique basée sur les réseaux de neurone profond, principalement les réseaux de neurone convolutifs (CNN) qui utilisent un apprentissage en profondeur (deep learning). Le CNN est composé de deux parties bien distinctes, une partie d'extraction de caractéristiques qui peut être utilisée pour simplifier un ensemble de données, et une partie de classification qui donne la classe d'appartenance de ces données (caractéristiques).

Nous avons développé dans ce projet, une partie de visualisation, de regroupement (clustering) et réduction de dimension des données (images faciales) mise en œuvre pour la classification de genre et estimation d'âge. Pour cela, différentes approches ont été utilisées afin de mapper les données dans un espace de plus petite dimension, cela permet de réduire la taille des vecteurs des caractéristiques et avoir une meilleure représentation des données.

Nous avons choisi d'articuler notre étude autour de quatre chapitres principaux.

Le premier chapitre est consacré à la présentation générale de la biométrie. Il décrit le principe de fonctionnement des systèmes biométriques puis définit les outils utilisés pour évaluer leurs performances. Ensuite, la place de la reconnaissance faciale parmi les autres techniques biométriques est analysée. A travers ce chapitre, nous voulons positionner le problème de la reconnaissance faciale et estimation de genre et d'âge pour présenter ses enjeux et intérêts par rapport aux autres techniques. Enfin, nous mettons en lumière sur les difficultés rencontrées par les systèmes de reconnaissance de visage.

Dans le deuxième chapitre, nous évoquerons l'état de l'art des techniques de réduction de dimension et classification d'image, où nous décrivons brièvement certaines techniques les plus importantes ou les plus courantes utilisées dans la réduction de dimension et classification automatiques, ainsi que les approches de reconnaissance faciale, d'estimation du genre et d'âge qui ont été proposées dans ce travail.

Dans le troisième 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 le constituant. Une partie de chapitre a été consacrée aux propriétés et paramètres du CNN. Cette étude détaillée du modèle CNN nous permet de mieux comprendre son processus.

Dans le quatrième chapitre, Nous présentons les résultats expérimentaux obtenus par les méthodes utilisées pour la reconnaissance de visage, l'estimation d'âge et la prédiction de genre en analysons leurs performances, suivies d'une discussion avec interprétation des résultats. Enfin, une conclusion générale résumera les résultats obtenus par les différentes approches et donnera quelques perspectives sur les travaux futurs.

*Chapitre 01 : La
biométrie & systèmes
de reconnaissance
faciale*

1.1 Introduction

Les empreintes digitales, les veines de la main, l'iris, le visage, la voix ou l'ADN deviennent des données d'identification et d'authentification comme les autres grâce au progrès de la numérisation. L'empreinte digitale remplaçait déjà le badge d'entreprise, désormais elle se substitue au code PIN. À la pointe de « l'expérience utilisateur », Apple, le premier, en 2013, a équipé son iPhone 5 de la fonction Touch ID permettant le déverrouillage de l'appareil par la reconnaissance de l'empreinte digitale. Puis, en 2017, l'iPhone X sera équipé de la fonction Face ID, permettant l'identification du propriétaire de l'appareil par une image en 3D de son visage, même si l'on ajoute une paire de lunettes ou si l'on change de coupe de cheveux.

Elle Conçue à des fins de sécurité ou de défense, cette technique d'authentification et d'identification par le corps trouve de nouveaux débouchés commerciaux. La reconnaissance digitale ou faciale remplace les mots de passe et les codes, simplifiant le déverrouillage des terminaux, l'accès aux services et les modes de paiement.

L'analyse du visage humain constitue l'une des tâches les plus importantes de la vision par ordinateur, car l'analyse automatique d'un tel objet déformable est d'une grande importance [1]: la caractérisation de l'âge, du sexe, des attributs faciaux, des expressions, du vêtement et même de la personnalité, pour n'en citer que quelques-uns, sont cruciaux dans plusieurs applications, comme l'identification des utilisateurs, l'interaction sociale, le suivi du visage et la reconnaissance des comportements [2, 3]. En ce qui concerne la classification d'âge et de genre, bien que ces deux tâches aient été largement traitées dans le passé, les performances rapportées sont loin d'être optimales [4, 5].

Au cours des dernières années, les réseaux de neurones convolutifs (CNN) [6] sont devenus le principal cheval de bataille pour l'estimation de l'âge et du genre. Les CNN se sont avérés très performants dans une variété de tâches de vision par ordinateur telles que la reconnaissance de l'action humaine [7], la reconnaissance des chiffres manuscrits [8], la vérification du visage [9] ou la détection automatique du visage [10]. En ce qui concerne la tâche de l'analyse de la biométrie douce, les CNN ont récemment été appliqués à la tâche d'estimation de l'âge apparent [11, 12, 13], de la classification du genre et du sourire [14], et de la prévision réelle de l'âge et du genre [15].

1.2 La biométrie

1.2.1 Définition

Le terme biométrie vient des mots grecs Bios (vies) et Metrikos (mesure ou distance), qui veut dire Distance Biophysiques de l'être humain, ou en d'autre terme, c'est l'anthropométrie [16].

1.2.2 Pourquoi utiliser la biométrie ?

La biométrie est un domaine émergent, ou la technologie améliore notre capacité à identifier une personne. La protection des gens contre la fraude ou le vol est l'un des buts de la reconnaissance par la biométrie [16].

L'avantage de la reconnaissance biométrique est que chaque individu a ses propres caractéristiques physiques qui ne peuvent être changées, perdues ou volées. La méthode de reconnaissance biométrique peut aussi être utilisée en complément ou remplacement des mots de passe.

Plusieurs raisons peuvent motiver l'usage de la reconnaissance par la biométrie :

- **Une haute sécurité** : en l'associant à d'autres technologies comme le cryptage.
- **Confort** : en remplaçant juste le mot de passe, exemple pour l'accès à un système d'exploitation, la biométrie permet de respecter les règles de base de la sécurité (ne pas inscrire son mot de passe à côté du PC, ne pas désactiver l'écran de veille pour éviter des saisies de mots de passe fréquentes). Et quand ces règles sont respectées, la biométrie évite aux administrateurs de réseaux d'avoir à répondre aux nombreux appels pour perte de mot de passe (que l'on donne parfois au téléphone, donc sans sécurité).
- **Sécurité/Psychologie** : dans certains cas, particulièrement pour le commerce électronique, l'utilisateur n'a pas de confiance. Il est important pour les acteurs de ce marché de convaincre le consommateur de faire des transactions. Un moyen d'authentification connu comme les empreintes digitales pourrait faire changer le comportement des consommateurs.

1.3 Panorama des différentes techniques biométries

Les systèmes biométriques sont généralement classés en trois grandes catégories : la biométrie morphologique ou physiologique, la biométrie comportementale et la biométrie biologique ou émergente [17].

- **La biométrie morphologique** est basée sur l'identification de traits physiques particuliers qui, pour toute personne, sont uniques et permanents. Cette catégorie regroupe la reconnaissance des empreintes digitales, la forme de la main, la forme du visage, la rétine et l'iris de l'œil et les veines de la main [17].

- **La biométrie comportementale** se base sur l'analyse de certains comportements d'une personne comme le tracé de sa signature, l'empreinte de sa voix, son mouvement de lèvres ou encore sa façon de taper sur un clavier [17,18].

- **L'ADN (Acide désoxyribonucléique)** : la forme des oreilles, l'odeur du corps humain et l'analyse de la démarche sont également étudiés, mais ses caractéristiques sont considérées comme émergentes (stade purement expérimental) [18].

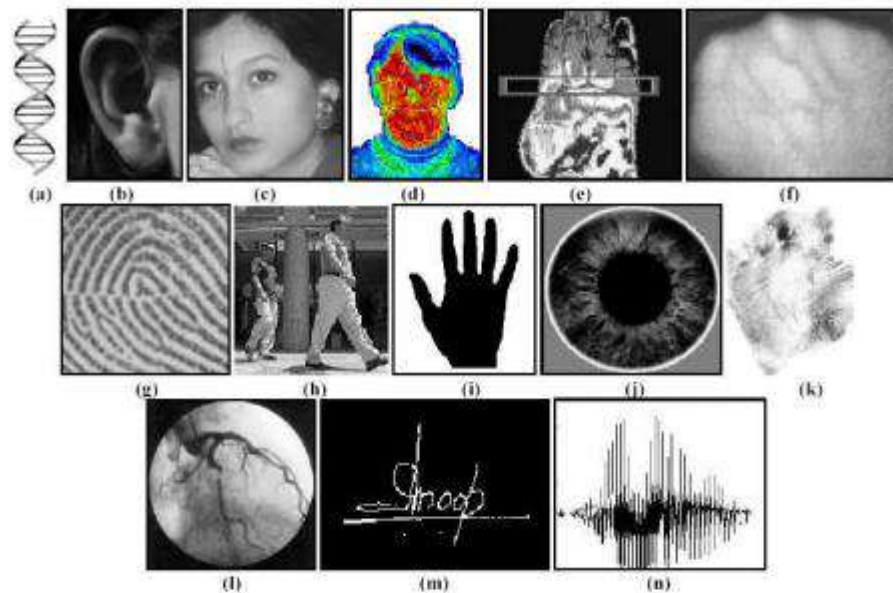


Figure 1.1 : Caractéristiques biométriques : a) ADN, b) Oreille, c) visage, d) visage infrarouge, e) thermogramme main, f) veine main, g) Empreintes digitales, h) marche, i) geste j) iris, k) empreinte de la paume, l) rétine, m) signature, n) voix.

En théorie, on dit qu'un critère physiologique ou comportemental est exploitable en reconnaissance d'individus s'il satisfait aux conditions ci-après :

- **Universalité** : toutes les personnes d'une population à identifier doivent la posséder ;
- **Unicité** : deux personnes ne doivent pas posséder exactement la même caractéristique ;
- **Permanence** : cela signifie qu'elle ne doit pas varier au cours du temps ;
- **Mesurabilité** : qui justifie de la possibilité de le quantifier [19].

En pratique, il existe d'autres conditions importantes :

- **Performance** : l'identification doit être précise et rapide ;
- **Acceptation** : elle doit être bien acceptée par les utilisateurs du système ;
- **Circonvension** : qui réfère à la facilité ou la difficulté de le pirater [20].

1.4 Le mode de fonctionnement d'un système biométrique

Un système biométrique est essentiellement un système de reconnaissance de formes qui utilise les données biométriques d'un individu. Il existe deux modules dans un système biométrique, le module de vérification et le module d'identification [22] :

- **Le mode de vérification ou authentification** est une comparaison "un à un", dans lequel le système valide l'identité d'une personne en comparant les données biométriques saisie avec le modèle biométrique de cette personne stockée dans la base de données du système. Dans un tel mode, le système doit alors répondre à la question suivante : «Suis-je réellement la personne que je suis en train de proclamer ?». Actuellement la vérification est réalisée via un numéro d'identification personnel, un nom d'utilisateur, ou bien une carte à puce.

- **Le mode d'identification** est une comparaison "un à N", dans lequel le système reconnaît un individu en l'appariant avec un des modèles de la base de données. La personne peut ne pas être dans la base de données. Ce mode consiste à associer une identité à une personne. En d'autres termes, il répond à des questions du type : « Qui suis-je ? ».

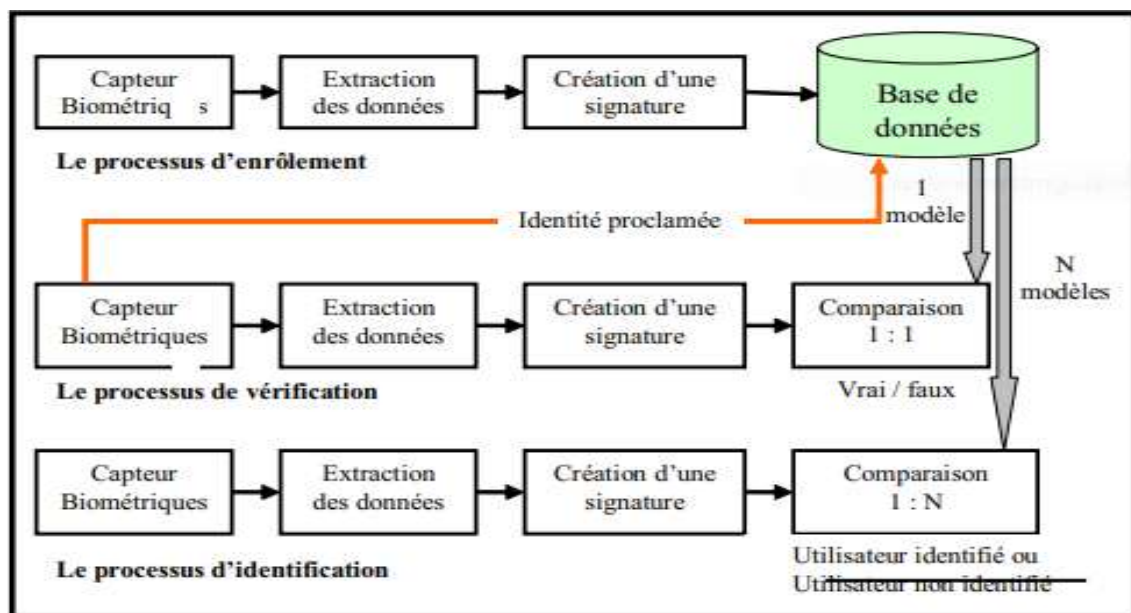


Figure 1.2 : Principaux modules d'un système biométrique ainsi que les différents modes [22].

Les différents modules qui composent un système biométrique sont représentés sur *la figure 1.2* ; leur fonctionnement peut être résumé comme suit :

- **Module capteur biométrique** : correspond à la lecture de certaines caractéristiques physiologiques, comportementales ou biologiques d'une personne, au moyen d'un terminal de capture biométrique (ou capteur biométrique).
- **Module extraction des données** : extrait les informations pertinentes à partir des données biométriques brutes, par exemple des images de visage ou des régions caractéristiques de visage.
- **Module création d'une signature** : crée un modèle numérique afin de représenter la donnée biométrique acquise. Ce modèle, appelé aussi signature, sera conservé sur un support portable (puce ou autre) ou dans une base de données.
- **Module comparaison** : compare les caractéristiques biométriques d'une personne soumise à contrôle (volontairement ou à son insu) avec les « signatures » mémorisées. Ce module fonctionne soit en mode vérification (pour une identité proclamée) ou bien en mode identification (pour une identité recherchée).
- **Module base de données** : stocke les modèles biométriques des utilisateurs enrôlés.

1.5 Structure générale d'un système biométrie

Un système biométrique se compose essentiellement de quatre modules :

- **L'acquisition** : On utilise un système d'acquisition pour acquérir une caractéristique spécifique de l'individu, au moyen d'un capteur.
- **L'extraction** : Après fait l'acquisition d'une image ou d'une voix, on réalise les informations nécessaires et les stocker dans une base de données.
- **La classification** : En examinant les modèles stockés dans la base de données, le système collecte un certain nombre de modèles qui ressemblent le plus à celui de la personne à identifier, et constitue une liste limitée de candidats. Cette classification intervient uniquement dans le cas d'identification car l'authentification ne retient qu'un seul modèle (celui de la personne proclamée) [31].

- **La décision** : Le module de décision doit être capable d'accepter ou refuser une authentification après fait la comparaison entre les nouvelles données de l'utilisateur et les différents modèles stockés dans la base de données afin de déterminer l'identité de l'utilisateur si elle existe.

1.6 Evaluation 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 des systèmes biométriques. La société américaine – l'International Biometric Group [IBG] – à par exemple effectué une étude basée sur quatre critères d'évaluation [22] :

- **intrusivité** : 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 la section suivante.

- **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 non contraignante.

Les résultats de cette étude peuvent être présentés comme suit :

- les techniques les moins "intrusives" aux plus "intrusives" : la voix, la frappe sur le clavier, la signature, la main, le visage, l'empreinte digitale, l'iris et enfin la rétine.

- les techniques les plus fiables aux moins fiables : l'iris, la rétine, l'empreinte digitale, le visage, la main, la voix, et enfin à un niveau équivalent, la frappe sur le clavier et la signature.
- les techniques les moins coûteuses aux plus coûteuses : la frappe sur le clavier, la voix, la signature, l'empreinte digitale, le visage, la main, la rétine et enfin l'iris.
- les techniques les plus faciles d'utilisation aux plus difficiles : le visage, la signature, l'iris, la frappe sur le clavier, la voix, l'empreinte digitale, la main et enfin la rétine.

1.7 Applications des systèmes biométriques

Aujourd'hui, les principales applications sont la production de titres d'identité, le contrôle d'accès à des sites sécurisés, le contrôle des frontières, l'accès aux réseaux, systèmes d'information et stations de travail, le paiement électronique, la signature électronique et même le chiffrement de données. Cette liste n'est pas exhaustive, et de nouvelles applications vont très certainement voir rapidement le jour.

Les techniques biométriques sont appliquées dans plusieurs domaines et leur champ d'application couvre potentiellement tous les domaines de la sécurité où il est nécessaire de connaître l'identité des personnes. Les applications peuvent être divisées en trois groupes principaux [23] :

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

1.8 Mesure de la performance d'un système biométrique

1.8.1 Performance d'un système d'identification

Pour évaluer les performances d'un système d'identification, on calcule le taux de reconnaissance du système. L'erreur commise par ce type de système est d'attribuer à l'individu présenté une identité autre que la sienne. Les performances de ces systèmes sont mesurées à l'aide du taux d'identification correcte [30] :

$$\text{TIC} = \frac{\text{nombre de test correspondant à une identification correcte}}{\text{nombre total de tests}} \dots\dots\dots 1.1$$

1.8.2 Performance d'un système de vérification

Un système de vérification ne peut donner, lors d'une comparaison entre deux échantillons qu'un résultat sous forme de probabilité de coïncidence ou score (S), ce score est comparé à un seuil de décision (T) que l'on détermine grâce à de nombreux tests ; Si S est supérieur à T le sujet en question est accepté et dans le cas contraire le sujet est considéré comme imposteur. Dans ce cas, deux types d'erreurs peuvent être commis [31] :

- **Faux Rejets FR (False-Rejection)** : correspond au cas où le système rejette un client légitime.
- **Fausse Acceptation FA (False-Acceptance)** : correspond au cas où le système accepte un imposteur.

La fiabilité d'un système de reconnaissance biométrique est caractérisée par deux valeurs statistiques :

- **Le Taux de Faux Rejets (FRR)** : il exprime le pourcentage d'utilisateurs légitimes rejetés.

$$\text{FRR} = \frac{\text{nombre des clients rejetés(FR)}}{\text{nombre total d'accès de clients}} \dots\dots\dots 1.2$$

- **Le Taux de Fausses Acceptations (FAR)** : il exprime le pourcentage d'imposteurs acceptés.

$$\text{FAR} = \frac{\text{nombre des imposteurs acceptés(FA)}}{\text{nombre total d'accès imposteurs}} \dots\dots\dots 1.3$$

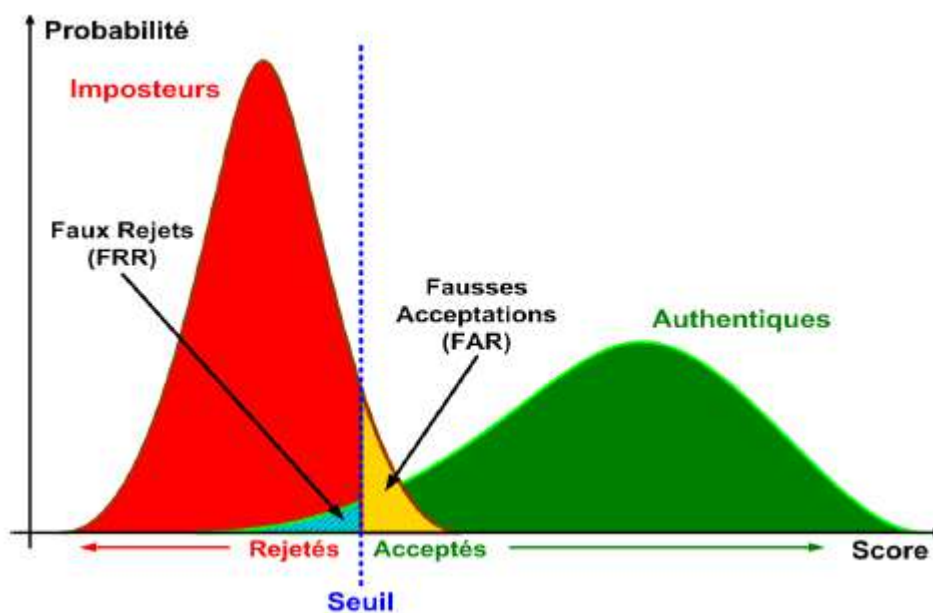


Figure 1.3 : Illustration du FRR et du FAR.

On peut représenter graphiquement les performances d'un système de reconnaissance de visage à l'aide de la courbe **ROC (Receiver Operating Characteristic)**. Les **FRR** sont donnés en fonction des **FAR** pour les différentes valeurs du seuil Θ

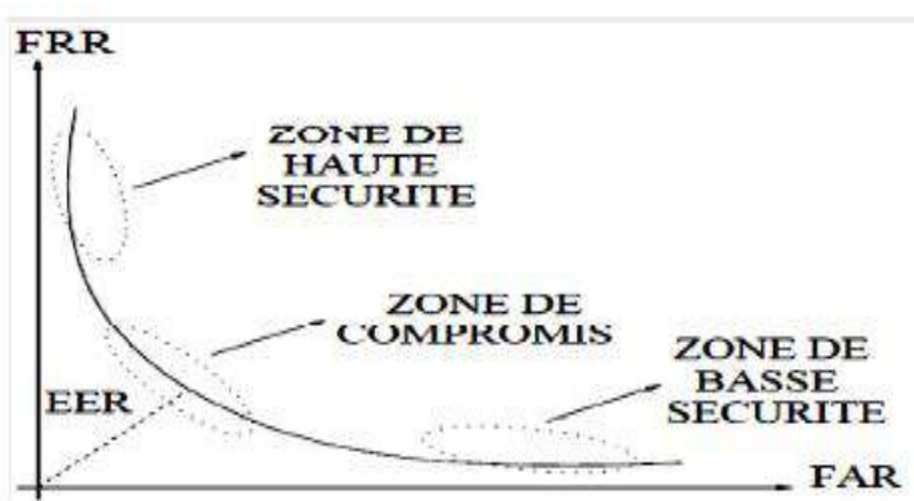


Figure 1.4 : La courbe ROC.

Il existe d'autres critères statistiques pour mesurer les performances d'un système biométrique :

- **L'EER (Equal Error Rate ou EER)** : Il correspond à l'intersection de la courbe ROC avec la première bissectrice :

$$\mathbf{FAR = FRR} \dots\dots\dots 1.4$$

- **Le HTER (Half Total Error Rate)** : Il représente la moyenne de FAR et FRR donc :

$$\mathbf{HTER = \frac{FAR + FRR}{2}} \dots\dots\dots 1.5$$

- **Le TER (Total Error Rate)** : représente le taux d'erreur total.

$$\mathbf{TER = \frac{\text{nombre des clients rejetés} + \text{nombre des imposteurs acceptés (FA)}}{\text{nombre total d'accès}}} \dots\dots\dots 1.6$$

1.9 Reconnaissance de visage

Les empreintes digitales sont les caractéristiques biométriques les plus communément utilisées pour la reconnaissance de criminels. Le premier système automatique d'authentification utilisant les empreintes digitales a été commercialisé au début des années soixante. D'autres parts, plusieurs études ont démontré que l'iris est la caractéristique la plus fiable car la texture de l'iris reste stable au cours de la vie (c.f. section 1.3). Toutefois, ces méthodes présentent l'inconvénient majeur d'être intrusives, ce qui limite énormément leurs domaines d'applications. De plus, une méthode comme l'identification de l'iris reste contraignante pour les utilisateurs qui n'apprécient pas de placer leur œil devant un appareil. A l'inverse, des systèmes d'identification basés sur l'analyse des images de visage ne présentent aucune contrainte pour les utilisateurs.

La reconnaissance faciale est une modalité qui peut être implémentée de manière indépendante des autres modalités biométriques, elle est souvent utilisée dans des applications de surveillance. Elle offre plusieurs avantages : le système de capture (c.à.d les caméras) est facile à installer, il est accepté dans les lieux publics ce qui permet d'avoir des bases de données de plus en plus grandes et ainsi d'améliorer les performances de la reconnaissance [24].

La reconnaissance faciale est une tâche que les humains effectuent naturellement et sans effort dans leurs vies quotidiennes. Et en tant qu'une des technologies biométriques de base, a pris une part de plus en plus importante dans le domaine de la recherche, ceci étant dû aux avancées rapides dans des technologies telles que les appareils photo numériques, Internet et les dispositifs mobiles, le tout associé à des besoins en sécurité sans cesse en augmentation.

Les systèmes de reconnaissance du visage reposent sur des algorithmes d'analyse de l'image, pouvant identifier les personnes qui y sont associées. Ces programmes créent une image du visage, en mesurant ses caractéristiques. Ils en produisent ensuite un fichier individuel, dénommé "Template ou signature". Les Template sont alors comparées avec toutes les images existantes au niveau de la base de données, en ressortant un score de similitude.

- **Motivation :**

- **Pourquoi la reconnaissance de visages ?**

Au cours de 4 dernières décennies, la reconnaissance automatique des visages est devenue un enjeu primordial, notamment dans les domaines de l'indexation de documents multimédias et surtout dans la sécurité, ceci est dû aux besoins du monde actuel mais aussi à ses caractéristiques avantageuses dont on peut citer :

- 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. En plus, cette technique est très efficace pour les situations non standards, c'est les cas où on ne peut avoir la coopération de l'individu à identifier, par exemple lors d'une arrestation des criminels. Certes que la reconnaissance des visages n'est pas la plus fiable comparée aux autres techniques de biométrie, mais elle peut l'être aussi si on utilise des approches plus efficaces en plus du bon choix des caractéristiques d'identification représentant le visage en question [32].

1.9.1 Les classes des techniques de reconnaissance de visage

Les méthodes de reconnaissance de visage peuvent être classées en trois grandes familles sont : méthode globale, locale et hybride

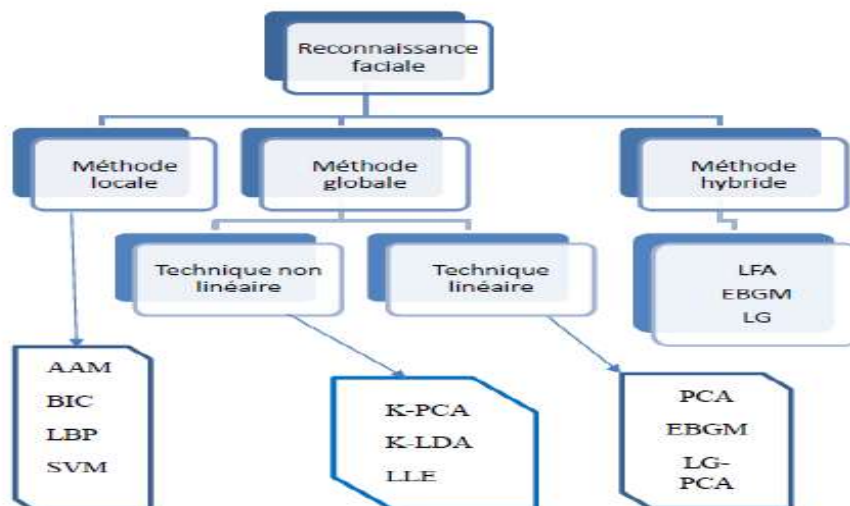


Figure 1.5 : Classification des algorithmes principaux utilisés en reconnaissance faciale.

- **Méthodes globale :**

Ce type d'approches utilisent la segmentation de ses parties, elles se basent principalement sur des informations pixel, ces algorithmes s'appuient sur des techniques d'analyse statique bien connue. Dans ces méthodes, les images de visage sont généralement transformées en vecteur plus faciles à manipuler.

L'avantage principale des méthodes globales sont relativement rapides à mettre en œuvre, mais sont sensibles aux problèmes d'éclairément, de pose et d'expression faciale.

Nous pouvons distinguer deux types de techniques parmi les méthodes globales les techniques linéaires et les techniques non linéaires.

Il y'a plusieurs méthodes globale, mais les plus connues sont : PCA, LDA/FLD,...

- **Méthode locale :**

Les méthodes locales, basées sur des modèles, utilisent des connaissances a priori que l'on possède sur la morphologie du visage et s'appuient en général sur des points caractéristiques de celui-ci. Kanade [25] présenta un des premiers algorithmes de ce type en détectant certains points ou traits caractéristiques d'un visage, puis en les comparants avec des paramètres extraits d'autre visage. Ces méthodes constituent un autre approche pour prendre en compte la non-linéarité en construisant un espace de caractéristiques locale et en utilisant des filtres d'image approprié de manière à ce que les distributions des visages soient moins affectées par divers changements.

L'avantage de ces méthodes est qu'elles prennent en compte la particularité du visage en tant que forme naturelle à reconnaître, en plus elles utilisent un nombre réduit de paramètres. Par rapport aux méthodes globales elles sont plus robustes aux problèmes posés par les variations d'éclairage, de pose et d'expression faciale.

Parmi ces approches on peut citer : filtres de Gabor, HMM...

- **Méthode hybride :**

La méthode globale et méthode locale ont des propriétés différentes, la méthode hybride fait la combinaison entre la détection de caractéristiques structurales et l'extraction de caractéristiques d'apparence locales, pour obtenir des informations complémentaires utiles ce qui permet d'augmenter la performance de reconnaissance lors de changements de pose, d'éclairage et d'expressions faciales [26].

1.9.2 Principales difficultés de la reconnaissance de visage

Pour le cerveau humain, la reconnaissance de visages est un processus visuel de haut niveau. Bien que les humains puissent détecter et identifier des visages et les reconnaître dans la scène sans trop de tracas, la construction d'un système automatisé qui accomplit ces tâches représente un sérieux défi. Ce défi devient encore plus grand lorsque les conditions de prise des images sont très changeantes. Il existe deux types de variations associées aux images de visages : inter et intra sujet. La variation inter-sujet est limitée à cause de la ressemblance physique entre les individus. Par contre la variation intra-sujet est plus vaste. Elle peut être attribuée à plusieurs facteurs que nous analysons ci-dessous [27] :

1.9.2.1 Changement d'illumination

L'apparence du visage sur la photo varie considérablement en fonction de l'illumination de la scène lors de la prise de vue (**voir figure 1.6**).

Les différences d'éclairage rendent la tâche de reconnaissance faciale très difficile. En effet, le changement d'apparence du visage dû à l'éclairage, est parfois plus important que la différence physique entre les individus, et peut conduire à une mauvaise classification des images saisies [27].



Figure 1.6 : Exemples de changement d'illumination.

1.9.2.2 Variation de pose

C'est une variation de la rotation et c'est un gros problème avec les systèmes de reconnaissance faciale. En effet, de nombreux tests ont montré que la restauration de la tête n'entraîne pas de réduction significative des taux de détection à $\pm 25^\circ$. Mais si cette rotation dépasse ce seuil, cela réduira les performances. *La figure 1.7* illustre une personne à laquelle les changements sont faits [28].



Figure 1.7 : Exemples de variation de pose.

1.9.2.3 Expressions faciales

La déformation faciale due aux expressions faciales affecte principalement la partie inférieure du visage. L'information faciale trouvée en haut du visage reste presque constante, ce qui est habituellement suffisant pour mener à bien le processus d'identification. Cependant, puisque l'expression faciale modifie l'apparence du visage, elle entraîne nécessairement une diminution du taux de reconnaissance. L'identification faciale avec l'expression faciale est un



Figure 1.8 : Exemples de variation d'expressions.

problème difficile qui est toujours pertinent et reste non résolu [29], *la figure 1.8* ci-dessous montre quelques expressions faciales.

1.9.2.4 Les Occultations partielles

Les visages peuvent être partiellement masqués par d'autres objets qui couvrent le visage. En effet, dans une image qui contient un groupe de personnes, par exemple, le visage peut masquer partiellement d'autres visages.



Figure 1.9 : Exemples des Occultations partielles.

1.9.2.5 Présence ou absence des composants structurels

Des aspects particuliers tels que la barbe, la moustache et les lunettes, comme le montre *la figure 1.10*, provoquent des changements importants dans les composants structuraux du visage, notamment la forme, la couleur, la taille, etc.



Figure 1.10 : Exemples de composants structurels.

1.9.3 Avantages et inconvénients de la reconnaissance faciale

Le tableau 1.1 ci-dessous présente certains avantages et inconvénients de la technologie de reconnaissance faciale.

Tableau 1.1 : Avantages et inconvénients de la reconnaissance faciale.

Avantages	Inconvénients
<ul style="list-style-type: none"> • Facile à mettre en œuvre et à tester car il ne nécessite pas d'équipement spécial, mais tout ce dont vous avez besoin est d'une caméra haute résolution. • Bien accepté par le public. • Aucune action de l'utilisateur (n'implique pas de coopération de l'utilisateur). • Pas de contact physique. • Technique peu coûteuse (Capteurs très bon marché) 	<ul style="list-style-type: none"> • Sensible à l'environnement d'acquisition (éclairage, position, expression du visage...) • Les vrais jumeaux ne sont pas différenciés. • Sensible aux changements (barbe, moustache, lunettes, piercing, chirurgie...)

1.10 La reconnaissance de genre et l'estimation d'âge

1.10.1 La reconnaissance de genre

La perception et la reconnaissance du genre ont déjà fait l'objet de nombreuses recherches dans les domaines social et la psychologie cognitive travaille dans le contexte de la reconnaissance des visages [33].

1.10.2 L'estimation d'âge

Au cours des dernières décennies, 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 de visages, la reconnaissance de visages, la classification de genre, et la reconnaissance de l'expression du visage) ont attiré une attention

particulière dans les communautés de la vision par ordinateur et de la reconnaissance de formes. En lien avec ces recherches, l'estimation de l'âge d'une personne à partir de l'analyse numérique de son visage est un sujet relativement nouveau. Les solutions proposées ou à développer s'appuient sur les avancées réalisées dans les domaines de technologie de capteurs, traitement du signal et de l'apprentissage machine. L'estimation de l'âge par analyse numérique du visage trouve de nombreuses applications potentielles telles que le développement d'interfaces homme-machine intelligentes et l'amélioration de la sécurité et la protection des mineurs, dans des secteurs divers et variés (transport, médecine, etc.). L'attribut de l'âge pourrait être également exploité dans la vérification du visage et de la récupération des données pour améliorer par exemple les outils utilisés dans les enquêtes policières. De manière générale, l'estimation automatique de l'âge par une machine est utile dans les applications où l'objectif est de déterminer l'âge d'un individu sans l'identifier précisément.

1.11 Conclusion

Dans ce chapitre, nous avons présenté les technologies utilisées dans les systèmes biométriques pour l'identification de personnes. Nous avons aussi donné un aperçu sur les techniques de mesure de leurs performances. Cette étude nous a permis de constater que la reconnaissance de visage suscite de plus en plus l'intérêt de la communauté scientifique ,car elle présente plusieurs challenges et verrous technologiques. Enfin, nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance automatique de visages ,ce qui nous a permis de bien définir les problématiques traitées dans cette thèse ,notamment l'invariance à l'illumination.

*Chapitre 02 : Etat de
l'art des techniques
de réduction de
dimension et
classification d'image*

2.1 INTRODUCTION

Dans ce chapitre, nous décrivons brièvement certaines techniques les plus importantes ou les plus couramment utilisées dans le domaine de la reconnaissance de forme, en commençant par les méthodes basées sur la réduction de dimension qui vise à produire une représentation de faible dimension des données de haute dimension en préservant leurs structures pertinentes (la pertinence dépendant souvent de l'application). Ce type de technique sont importantes en science des données, à la fois pour la visualisation et potentiellement pour le prétraitement des données avant l'apprentissage automatique. Nous allons aussi présenter de nombreuses techniques de classification automatiques utilisées en reconnaissance faciale, estimation du genre et d'âge qui ont été proposées au cours de ces dernières années.

Dans cette description de l'état de l'art vu le grand nombre de méthodes appliquées dans ce domaine, la citation de la totalité des méthodes ne peut être complète. On s'est limité à décrire la plupart des grandes approches classiques et la majorité des techniques qui ont apporté une vraie originalité et une avancée sensible dans le domaine.

2.2 État de l'art des techniques de reconnaissance

Nous commençons la description de l'évolution de la recherche par une description de l'objet à reconnaître. Le visage est un objet à nature tridimensionnel, cet objet peut être sujet à diverses rotations, non seulement planes mais aussi spatiales, et également sujet à des déformations dues aux expressions faciales. La forme et les caractéristiques de cet objet évoluent aussi dans le temps. Les informations de cet objet tridimensionnel sont capturées à l'aide de caméras et à la sortie, nous obtenons une image bidimensionnelle qui le caractérise. Plusieurs méthodes de reconnaissance de visage et d'estimation du genre et d'âges ont été proposées durant les trois dernières décennies. La reconnaissance de formes est devenue un challenge intéressant qui a attiré les chercheurs de différentes spécialités [34] [35] : psychologie, réseaux de neurones, vision par ordinateur, infographie...etc. Suite à cela la littérature en reconnaissance de formes et du visage en particulier est vaste et diverse.

Dans l'état de l'art, deux familles d'algorithmes ressortent. Une première famille, la plus populaire, est basée sur des méthodes de réduction d'espace. Une deuxième famille d'algorithmes est basée sur des approches de classification dite aussi de discrimination [36].

2.3 Les méthodes de réduction de dimension

Les techniques de classification par réduction d'espace restent parmi les techniques les plus utilisées et les plus populaires dans le domaine de la reconnaissance du visage [37]. Ces méthodes considèrent l'image entière du visage comme un vecteur dans un espace multidimensionnel. L'objectif de ces méthodes est de trouver une base réduite de projections qui optimise un certain critère et l'espace original est ensuite projeté sur cette base de dimension inférieure. De ce fait on réduit considérablement la complexité des calculs tout en conservant l'information discriminante entre les images. La classification est habituellement effectuée selon une mesure de distance simple dans l'espace multidimensionnel. Ainsi, l'utilisation de distances classiques dans l'espace projeté ne permet pas toujours de réaliser une bonne classification entre les classes. Les approches de réduction d'espace peuvent se décomposer en deux types de techniques : les techniques linéaires et les techniques non linéaires. Parmi les approches les plus importantes réunies au sein de cette classe on trouve :

2.3.1 Méthodes linéaires de réduction de dimension

La réduction de dimensionnalité consiste à projeter les données dans un espace de grande dimension, afin de réduire la taille de ces données et avoir une meilleure représentation [38].

La réduction des dimensions est l'une des plus anciennes approches permettant d'apporter des éléments de réponse à ce problème. Son objectif est de sélectionner ou d'extraire un sous-ensemble optimal de caractéristiques pertinentes pour un critère fixé auparavant. La sélection de ce sous-ensemble de caractéristiques permet d'éliminer les informations non-pertinentes et redondantes selon le critère utilisé. Cette sélection/extraction permet donc de réduire la dimension de l'espace des exemples et rendre l'ensemble des données plus représentatif du problème. En effet, les principaux objectifs de la réduction de la dimension sont [39] :

- une meilleure visualisation et compréhension des données.
- réduire l'espace de stockage nécessaire.
- réduire le temps d'apprentissage et de reconnaissance.
- identifier les caractéristiques les plus pertinentes.

2.3.1.1 L'Analyse en Composantes Principales (ACP)

L'Analyse en Composantes Principales est une méthode de la famille de l'analyse des données et plus généralement de la statistique multivariée, qui consiste à transformer des variables liées entre elles (dites "corrélées" en statistique) en nouvelles variables indépendantes les unes des autres (donc "non corrélées"). Ces nouvelles variables sont

nommées "composantes principales". Elle permet au praticien de réduire l'information en un nombre de composantes plus limité que le nombre initial de variables. Il s'agit d'une approche à la fois géométrique (représentation des variables dans un nouvel espace géométrique selon des directions d'inertie maximale) et statistique (recherche d'axes indépendants expliquant au mieux la variabilité - la variance - des données). Lorsqu'on veut alors compresser un ensemble de N variables aléatoires, les N premiers composantes de l'ACP sont un meilleur choix, du point de vue de l'inertie ou la variance expliquée.

L'ACP prend sa source dans un article de Karl Pearson [40], pour ses travaux dans le domaine de la régression et des corrélations entre plusieurs variables. Pearson utilise ces corrélations non plus pour expliquer une variable à partir des autres (comme en régression), mais pour décrire et résumer l'information contenue dans ces variables. Encore connue sous le nom de transformée de Karhunen-Loève ou de transformée de Hotelling, l'ACP a été de nouveau développée et formalisée quelque années après par Harold Hotelling [41]. La puissance mathématique de l'économiste et statisticien américain le conduira aussi à développer l'analyse canonique, généralisation des analyses factorielles dont fait partie l'ACP. Les champs d'application sont aujourd'hui multiples, allant de la biologie à la recherche économique et sociale, et plus récemment le traitement d'images.

L'ACP est l'une des méthodes les plus anciennes et le mieux étudiées pour la représentation des visages dans un espace de dimension réduite, reste la technique dite des Eigenfaces présentée par Turk et Pentland en 1991 [42] basé sur l'analyse en composante principale (ACP). Cette représentation est réalisée sur un ensemble d'images d'apprentissage de visages. Les principaux vecteurs propres résultant de l'ACP définissent le nouvel espace. Les images de visages sont ensuite projetées sur cet espace, et les vecteurs obtenus sont utilisés pour la classification. L'ACP appliquée aux visages permet la réduction de la dimension de représentation du visage [43]. C'est une approche rapide et très facile à mettre en œuvre, néanmoins elle présente quelques inconvénients tels qu'une baisse des performances (temps d'exécution plus long) dans le cas où la base d'image est volumineuse.

La méthode des Eigen Faces repose sur une utilisation des premiers vecteurs propres comme visages propres, d'où le terme Eigen Faces. La base formée par ces vecteurs constitue un espace qu'on utilise pour représenter les images des visages. Les personnes se voient donc attribuer un vecteur d'appartenance pour chacune de leur image. Cela étant dit, la reconnaissance est réalisée en comparant les coefficients de projection d'un visage test avec

ceux appartenant aux visages d'entraînement. Après le choix des visages sur lesquels on va travailler, l'idée de cette méthode est de représenter chaque image sous forme de vecteur, puis regrouper ses derniers pour former une matrice de vecteur qu'on appellera matrice d'image, soit Γ cette matrice. On suppose Γ_i un vecteur de $N^2 \times 1$ correspondant à une image I_i de taille $N \times N$. Le but est de représenter Γ dans un espace de dimension inférieur, ce dernier devra être orthogonale (vecteurs composant sa base sont orthogonaux deux à deux), pour pouvoir discriminer les images. Nous présentons dans ce qui suit les principales étapes de la méthode Eigen Face.

Etape1 :

Cette étape consiste à définir les images des personnes, soit M le nombre d'image allant de I_1, I_2, \dots, I_M . Ces images doivent être centré et de même taille.

Etape2 :

Cette étape consiste à représenter chaque image I_i par un vecteur Γ_i , pour cela on va superposer les colonnes de chaque image.

Exemple :

Soit I_i une image de taille 3×3

$$I_i = \begin{bmatrix} 5 & 3 & 0 \\ 2 & 7 & 9 \\ 1 & 1 & 4 \end{bmatrix} \quad \text{On obtiendra} \quad \Gamma_i = \begin{bmatrix} 5 \\ 3 \\ 0 \\ 2 \\ 7 \\ 9 \\ 1 \\ 1 \\ 4 \end{bmatrix}$$

Etape 3 :

Cette étape consiste à calculer la moyenne des visages et de les représenter sous forme de vecteur Ψ .

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \dots\dots\dots 2.1$$

Etape 4 :

Cette étape consiste à enlever la moyenne de la matrice d'image, en d'autres termes : enlever tous ce qui est commun aux individus. Soit Φ la matrice résultante obtenue comme suit :

$$\Phi_i = \Gamma_i - \Psi \dots\dots\dots 2.2$$

Etape 5 :

Cette étape consiste à construire la matrice de covariance C de la matrice Φ . La matrice de covariance représente l'interaction entre les individus.

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = A \times A^T \quad (N^2 \times N^2) \dots\dots\dots 2.3$$

Où

$$A = [\Phi_1 \Phi_2 \dots \Phi_M] \quad (N^2 \times M)$$

Etape 6 :

Cette étape consiste à calculer les vecteurs propres qui constitueront notre espace d'étude. Ces vecteurs u_i seront tirés de la matrice de covariance $C = AA^T$. Cependant, cette matrice est très grande donc non praticable, à ce niveau, nous allons procéder à une déviation qui nous permettra de détourner le problème.

Considérons la matrice $A^T A (M \times M)$, si on essaye de calculer les valeurs propres de cette dernière, on remarquera que ce sont les mêmes que celles de la matrice C . De plus il existe un lien entre les vecteurs propres de chacune des deux matrices qui est le suivant : $u_i = Av_i$ On notera que pour la matrice AA^T on aura N^2 valeurs et vecteurs propres. Et pour la matrice $A^T A$ on aura M valeurs et vecteurs propres. Ainsi, les M valeurs propres de $A^T A$ correspondent aux M plus grandes valeurs de la matrice AA^T (en correspondance avec leurs vecteurs propres).

Etape 7 :

Cette étape est assez simple à réaliser, elle consiste à prendre K vecteurs propres correspondant aux K plus grandes valeurs propres.

Une fois les vecteurs propres trouvés, il ne reste plus qu'à déterminer la représentation des visages dans notre nouvel espace, pour cela on procède comme suit : Chaque visage (moins la moyenne) sera représenté comme étant une combinaison linéaire des K vecteurs propres choisis.

$$\hat{\phi}_i - mean = \sum_{j=1}^K w_j u_j \dots\dots\dots 2.4$$

Où

$$w_j = u_j^T \phi_i$$

Donc chaque visage d'apprentissage ϕ_i sera représenté dans l'espace comme suit :

$$\Omega_i = \begin{bmatrix} w_1^i \\ w_2^i \\ \cdot \\ \cdot \\ w_K^i \end{bmatrix} \quad i = 1, 2, \dots, M$$

Le problème qui reste à résoudre est le choix de K , la dimension de l'espace, Pour cela on aura besoin d'un seuil (pourcentage) dit de quantité d'information. Le but est de pouvoir représenter une certaine quantité d'information en un minimum de vecteurs base.

$$\frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^N \lambda_i} > SEUIL \dots\dots\dots 2.5$$

2.3.1.2 Analyse Discriminante Linéaire

L'analyse discriminante linéaire (LDA) appelée également "Fisher Linear Discriminant" (FLD), est l'une des méthodes les plus utilisées pour la reconnaissance de visage [44]. Elle est utilisée pour trouver la combinaison linéaire des caractéristiques qui séparent le mieux les classes d'objet ou d'événement. Les combinaisons résultantes peuvent être employées comme

classificateur linéaire, ou généralement dans la réduction de caractéristiques avant la classification postérieure. LDA est étroitement lié à l'ACP, du fait que tous les deux recherchent des combinaisons linéaires des variables qui représentent au mieux les données. Les combinaisons linéaires obtenues en utilisant FLD s'appellent les Fisherfaces, en analogie avec les Eigenfaces [45].

Chaque image d'entrées, qui se compose d'un grand nombre de Pixel, est réduite à un plus petit ensemble de combinaisons linéaires avant la classification. Chacune des nouvelles dimensions est une combinaison linéaire des valeurs de pixel, qui forment un template. Les combinaisons linéaires obtenues en utilisant *FLD* s'appellent les Fisherfaces, en analogie avec les Eigenfaces [46]. *LDA* est une technique qui cherche les directions qui sont efficaces pour la discrimination entre les données.

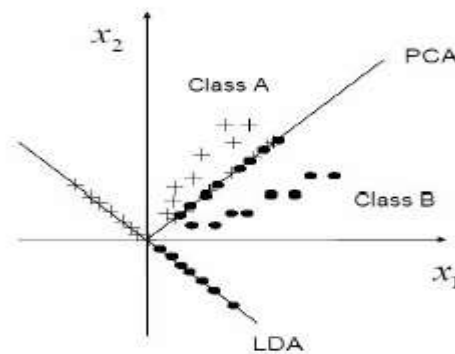


Figure 2.1 : Les projections PCA et LDA d'un ensemble de données.

Récemment, des algorithmes pour l'application directe de la LDA pour l'extraction des informations discriminantes, ont été proposés. Ces solutions ont pour but d'éviter la perte d'information discriminante qui se produit après une réduction par ACP. On cite la DLDA (Direct LDA) [47], la LDA (Regression LDA) [48], GDA (General Discriminant Analysis) [49] et l'analyse non linéaire discriminante de Fisher KFDA (Kernel Fisher Discriminant Analysis) [50].

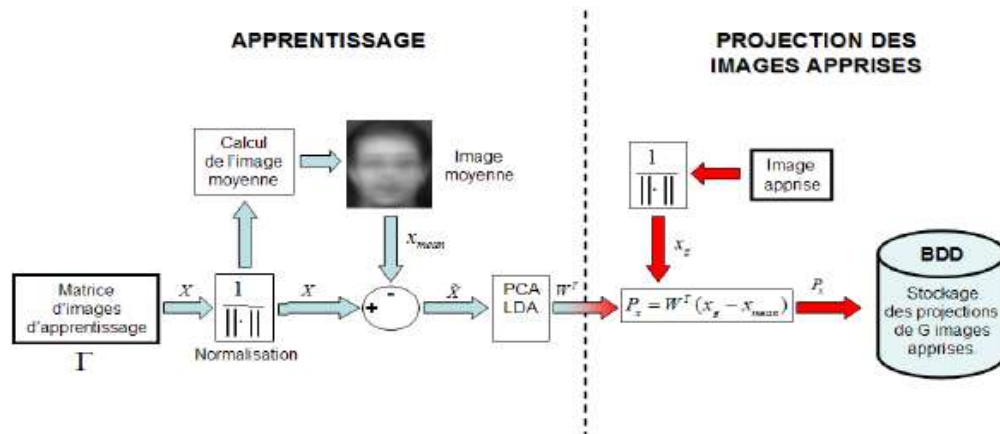


Figure 2.2 : Système de reconnaissance faciale utilisant ACP et LDA [51].

2.3.2 Méthodes non linéaires de réduction de dimension

La taille des données peut être mesurée selon deux dimensions, le nombre de variables et le nombre d'exemples. Ces deux dimensions peuvent prendre des valeurs très élevées, ce qui peut poser un problème lors de l'exploration et l'analyse de ces données. Pour cela, il est fondamental de mettre en place des outils de traitement de données permettant une meilleure représentation de la valeur des caractéristiques disponibles dans ces données.

2.3.2.1 T-SNE (*t-distributed stochastic neighbor embedding*)

T-SNE est une technique de réduction de dimension non linéaire. Elle permet de diminuer fondamentalement la multidimensionnalité à des dimensions 2D ou 3D de telle sorte qu'elle peut être visualisée par les yeux humains et l'analyse de données sera diminuée car il peut montrer divers modèles dans l'ensemble de données [52].

Cette technique est une variation de Stochastic Neighbor Embedding qui est beaucoup plus facile à optimiser et produit des visualisations nettement meilleures en réduisant la tendance à regrouper les points au centre de la carte. t-SNE est meilleur que les techniques existantes pour créer une carte unique qui révèle la structure à de nombreuses échelles différentes. Ceci est particulièrement important pour les données de haute dimension qui se trouvent sur plusieurs variétés différentes, mais liées, de faible dimension, telles que des images d'objets de plusieurs classes vues de plusieurs points [53].

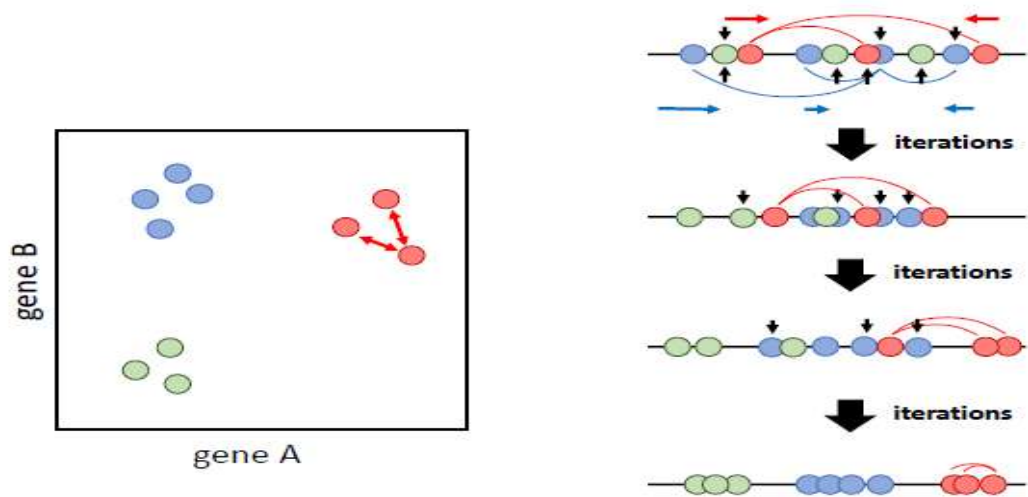


Figure 2.3 : Les étapes de la méthode de t SNE.

Le même concept s'applique à l'intégration dans des images a 2 dimensions

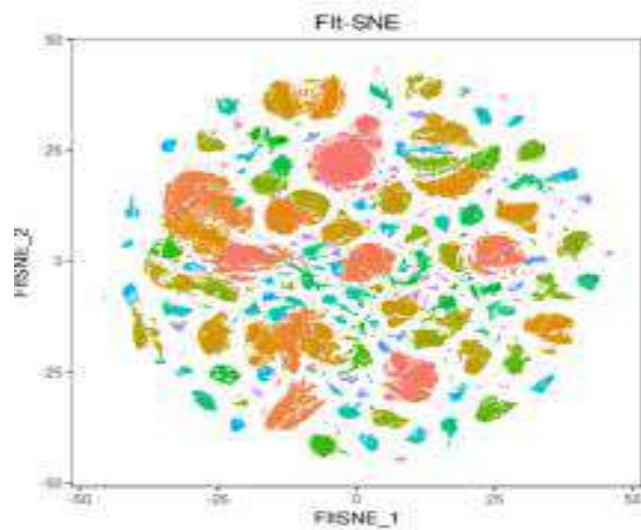


Figure 2.4 : t-SNE

2.3.2.2 UMAP (Uniform Manifold Approximation and Projection)

UMAP est une nouvelle technique d'apprentissage multiple pour la réduction des dimensions. Elle est construite à partir d'un cadre théorique basé sur la géométrie riemannienne et la topologie algébrique. Le résultat est un algorithme évolutif pratique qui s'applique aux données du monde réel. L'algorithme UMAP est compétitif avec t-SNE pour la qualité de la visualisation et préserve sans doute une plus grande partie de la structure globale avec des performances d'exécution supérieures. En outre, UMAP n'a aucune restriction de calcul sur l'intégration de la dimension, ce qui la rend viable comme technique de réduction de dimension à usage général pour l'apprentissage automatique.

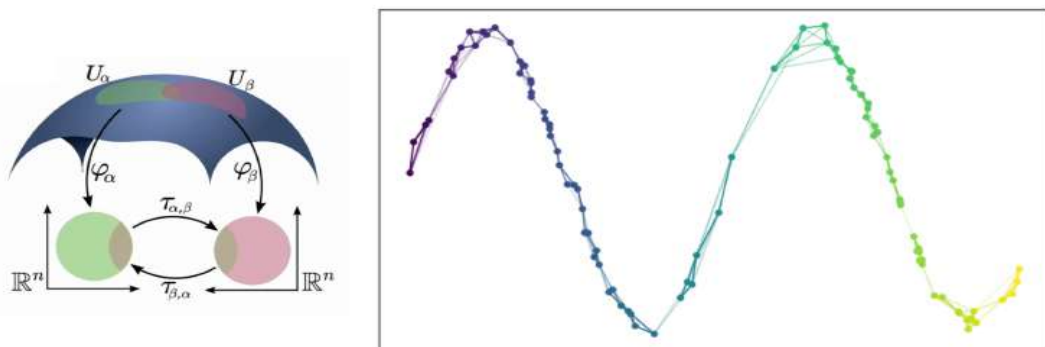
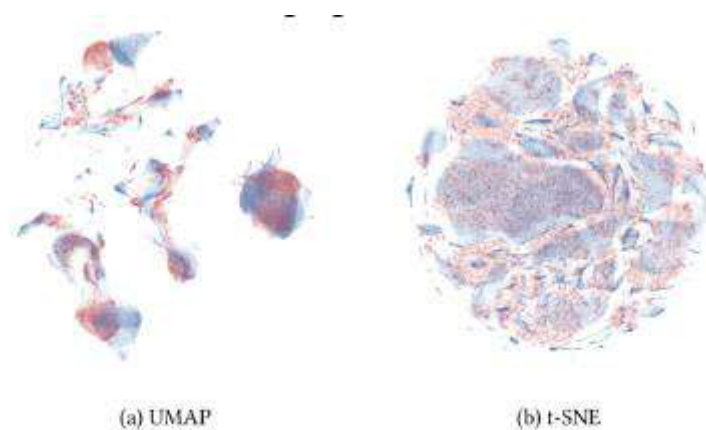


Figure 2.5 : Le fonctionnement de l'UMAP.

La technique UMAP apprend la structure de données globale et est moins dépendant d'initiateurs aléatoires (comme le t-SNE), il peut recréer incorporation dimensionnelle quelle que soit la taille du jeu de données.



(a) UMAP

(b) t-SNE

Figure 2.6 : UMAP vs t-SNE.

2.4 Les techniques de classification d'images

Dans cette section, nous présentons les principales approches basées la classification de caractéristiques extraites de différentes régions des visages.

2.4.1 Les réseaux de neurones artificiels (RNA)

L'application des réseaux de neurones artificiels dans la reconnaissance des images de visages a visé plusieurs problèmes : la classification homme et femme, l'estimation de la pose et la classification des expressions faciales. Les RNA (Réseaux de Neurones Artificiels) ont été utilisés dans nombreuses applications, particulièrement pour la classification de données, la modélisation des processus complexes et le traitement non-linéaire des signaux. Ces recherches ont engendré une panoplie d'architectures de réseaux dont chacune répond parfaitement à une application donnée. On trouve plusieurs types de réseaux de neurones : On note en particulier l'architecture MLP (Multi-Layer Perceptron, Perceptron Multi-Couches : PMC), l'architecture RBF (Radial Basis Function, Fonctions à base radiale) [54].

Les réseaux de neurones fonctionnent en répartissant les valeurs des variables dans des automates (les neurones). Ces unités sont chargées de combiner entre elles leurs informations pour déterminer la valeur du paramètre de discrimination. C'est de la connexion de ces unités entre elles qu'émerge la capacité de discrimination du RNA. Chaque neurone reçoit des informations numériques en provenance de neurones voisins, à chacune de ces valeurs est associé un poids représentatif de la force de la connexion. Chaque neurone effectue localement un calcul dont le résultat est transmis ensuite aux neurones avals [55].

Un neurone formel (ou simplement neurone artificiel) est une fonction algébrique non linéaire et bornée, dont la valeur dépend des paramètres appelés coefficients ou poids. Les variables de cette fonction sont habituellement appelées "entrées" du neurone, et la valeur de la fonction est appelée sa sortie [56]. Un neurone est donc avant tout un opérateur mathématique, dont on peut calculer la valeur numérique par quelques lignes de logiciel.

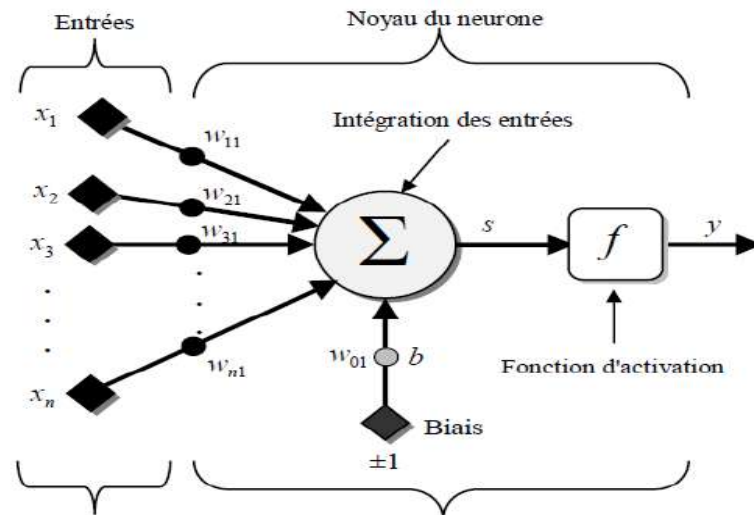


Figure 2.7 : Modèle d'un neurone artificiel.

Aujourd'hui, les réseaux de neurones artificiels ont de nombreuses applications dans des secteurs très variés :

- *Traitement d'images* : reconnaissance de caractères et de signatures, compression d'images, reconnaissance de forme, cryptage, classification, etc.
- *Traitement du signal* : filtrage, classification, identification de source, traitement de la parole...etc.
- *Contrôle* : commande de processus, diagnostic, contrôle qualité, asservissement des robots, systèmes de guidage automatique des automobiles et des avions...etc.
- *Défense* : guidage des missiles, suivi de cible, reconnaissance du visage, radar, sonar, lidar, compression de données, suppression du bruit...etc.
- *Optimisation* : planification, allocation de ressource, gestion et finances, etc.
- *Simulation* : simulation du vol, simulation de boîte noire, prévision météorologique, recopie de modèle...etc.

2.4.2 Machine à Vecteur de Support (SVM)

Une machine à vecteurs de support (SVM) est une technique qui a été proposée par V.Vapnik en 1995 [57], elle est utilisée dans plusieurs domaines (classement, statistiques, régression, fusion,...). Depuis son introduction dans le domaine de la reconnaissance de formes, plusieurs travaux ont pu montrer l'efficacité de cette technique principalement en traitement d'image. L'idée essentielle de cette approche consiste à projeter les données de l'espace d'entrée (appartenant à des classes différentes) non linéairement séparables, dans un

espace de plus grande dimension appelé espace de caractéristiques, de façon à ce que les données deviennent linéairement séparables [57] [58].

Le SVM est une méthode de discrimination de données basé sur un apprentissage supervisé, utilisée pour la classification et la régression. Elle consiste à séparer deux ou plusieurs ensembles de points par un hyperplan. Selon les cas et la configuration des points, la performance de la machine à vecteurs de support peut être supérieure à celle d'un réseau de neurones ou d'un modèle de mixture gaussienne [59]. Le but du SVM est donc de trouver une séparatrice qui minimise l'erreur de classification sur l'ensemble d'apprentissage mais qui sera également performante en généralisation sur des données non utilisées en apprentissage. Pour cela le concept utilisé est celui de marge (d'où le nom de séparateurs à vaste marge). La marge est la distance quadratique moyenne entre la séparatrice et les éléments d'apprentissage les plus proches de celle-ci appelés vecteurs de support (*figure 2.3*). Ces éléments sont appelés vecteurs de support car c'est uniquement sur ces éléments de l'ensemble d'apprentissage qu'est optimisée la séparatrice. Dans cette section, nous présenterons quelques notions de base qui sont Hyperplan, marge et support vecteur. Nous mettrons également en évidence l'utilisation des SVM dans les deux cas où les données sont (linéairement ou non linéairement) séparables [60].

Ramin Azarmehr et all [61], ont proposé un cadre complet pour une classification en temps réel et précise de l'âge et du genre sur les systèmes intégrés dans des environnements sans contrainte. Plusieurs améliorations ont été présentées pour l'alignement de la taille, la normalisation de l'éclairage et l'extraction des caractéristiques à l'aide d'une méthode de modèle binaire multi-résolution. Pour surmonter les limites des systèmes embarqués, ils ont introduit une technique de réduction de la dimensionnalité segmentaire et mis en place un classificateur SVM+RBF ainsi qu'une stratégie de classification démographique discriminante pour améliorer la performance.

2.4.3 Le modèle AAM (Active Appearance Model)

Un modèle d'apparence active (AAM) est un algorithme de vision par ordinateur pour faire correspondre un modèle statistique de forme et d'apparence d'objet à une nouvelle image. Ils sont construits au cours d'une phase de formation. Un ensemble d'images, ainsi que les coordonnées des points de repère qui apparaissent dans toutes les images, est fourni au superviseur de la formation.

Le modèle a d'abord été introduit par Edwards, Cootes et Taylor dans le contexte de l'analyse du visage lors de la 3e Conférence internationale sur la reconnaissance du visage et des gestes, 1998. [62] Cootes, Edwards et Taylor ont en outre décrit l'approche comme une méthode générale dans la vision par ordinateur à la Conférence européenne sur la vision par ordinateur la même année. [63][64] L'approche est largement utilisée pour l'appariement et le suivi des visages et pour l'interprétation des images médicales. L'algorithme utilise la différence entre l'estimation actuelle de l'apparence et l'image cible pour piloter un processus d'optimisation. En tirant parti des techniques des moindres carrés, il peut correspondre à de nouvelles images très rapidement.

- **Avantages** : Capture les informations de forme et de texture, et l'image peut être reconstruite à partir des entités extraites.⁵⁴ Par conséquent, elle convient à la modélisation d'objets déformables.

- **Inconvénients** : Il ne prend pas en compte les étiquettes de classe. Il est de nature linéaire et ne fonctionne donc pas lorsque les objets présentent une variation non linéaire.

L'AAM est l'une des techniques d'extraction de caractéristiques les plus utilisées. Cependant, elle souffre de certains inconvénients, surtout lorsqu'il est utilisé pour la classification. Cela s'explique principalement par le fait que l'analyse des composantes principales (ACP), qui est au cœur du modèle, fonctionne de façon non supervisée, c'est-à-dire que la réduction de la dimension de l'ACP ne tient pas compte de la façon dont les variables prédictives se rapportent à la réponse (étiquettes de classe). Il explore plutôt seulement la structure sous-jacente des variables prédictives, et il n'est donc pas surprenant que l'ACP écarte des parties précieuses des données qui représentent des caractéristiques discriminatoires. À cette fin, Ali Maina Bukar et al [65], proposent un modèle d'apparence supervisée (SAM) qui améliore l'AAM en remplaçant l'ACP par une régression partielle des moindres carrés. Cette technique d'extraction des caractéristiques est ensuite utilisée pour les problèmes de classification selon l'âge et le sexe. Leurs expériences montrent que SAM a une meilleure puissance prédictive que l'AAM classique.

2.5 Conclusion

Dans ce chapitre, on a passé en revue les principales techniques de reconnaissance, de réduction de dimension et classification les plus utilisées dans le domaine de la reconnaissance de formes. Cet engouement pour les systèmes de reconnaissance basée sur les images faciales est justifié par les nombreux avantages de cette approche. En effet cette technologie est peu

couteuse, peu encombrante, elle est de surcroit peu contraignant pour les usagers. Ce chapitre n'a pas pour finalité de décrire tous les algorithmes cité dans l'état de l'art, mais nous nous sommes contentés de présenter les algorithmes qui ont introduit une nette évolution dans le domaine de biométrie, permettant ainsi une véritable amélioration des performances.

Chapitre 03 :
Apprentissage
Profond : réseaux de
neurones convolutifs

3.1 Introduction

Depuis plusieurs années, l'intelligence artificielle (IA) connaît une très forte croissance de popularité aussi bien dans le milieu de la recherche scientifique qu'auprès des grandes compagnies des technologies de l'information comme Google, Amazon, Microsoft ou Facebook. Cette technologie s'invite aujourd'hui dans de nombreux domaines tels que les moteurs de recherche sur le Web, les assistants virtuels, la reconnaissance d'images ou encore les voitures autonomes.

Cette popularité est due aux récents développements d'un domaine de l'intelligence artificielle, l'apprentissage profond (Deep Learning en anglais). Où des avancées remarquables ont été faites dans la reconnaissance de parole et d'images en surpassant la précision des algorithmes utilisés jusqu'alors.

Dans ce chapitre, nous présentons une étude détaillée sur les réseaux de neurones CNN et ses différentes composantes. Ce chapitre vise à fournir une présentation au concept de réseaux de neurones convolutifs, pour ce faire, il est nécessaire de comprendre le concept du réseau neuronal artificiel, où une partie du chapitre lui est consacrée.

3.2 Intelligence artificielle

L'intelligence artificielle (IA) fait partie intégrante de la digitalisation qui a modifiée durablement notre société. Ce qui était il y a quelques années encore de l'ordre de la science-fiction est désormais réalité. Nous parlons avec des ordinateurs, nos téléphones nous orientent et nous indiquent le chemin le plus court, nos montres savent si nous avons suffisamment bougé dans la journée. La technique est de plus en plus intelligente, et les scientifiques, ingénieurs et programmeurs deviennent des enseignants, ils « entraînent » les ordinateurs à apprendre de façon autonome.

L'intelligence artificielle est donc à la fois la théorie et le développement concret de machines/systèmes/logiciels qui imitent l'intelligence humaine pour accomplir des tâches très évoluées [66].

L'IA tente de reproduire une partie de l'intelligence humaine à travers une application, un système ou un processus. La reconnaissance de la parole, la perception visuelle et la traduction linguistique sont des exemples de systèmes d'intelligence artificielle.

La machine learning (apprentissage automatique) et le deep learning sont des sous-ensembles de l'intelligence artificielle qui rendent l'IA plus efficace et plus accessible.

3.3 Apprentissage automatique (machine learning en anglais)

Aujourd'hui, l'apprentissage automatique est omniprésent, lorsque nous interagissons avec les banques, magasinons en ligne ou utilisons les médias sociaux, les algorithmes d'apprentissage automatique entrent en jeu pour rendre notre expérience fluide, efficace et sécuritaire. L'apprentissage automatique et les technologies afférentes se développent rapidement, et nous commençons à peine à entrevoir leurs capacités.

C'est un champ d'étude de l'intelligence artificielle, concerne la conception, l'analyse, le développement et l'implémentation de méthodes permettant à une machine (au sens large) d'évoluer par un processus basé sur les données plutôt que par des algorithmes déterministiques classiques [67]. Le rapport à l'intelligence vient de la capacité de ces modèles à généraliser, c'est-à-dire à extraire l'information pertinente de données étudiées au fil d'un processus de mise(s)-à-jour appelé entraînement, et de savoir la réutiliser avec efficacité sur de nouvelles données jamais rencontrées auparavant, ce qui rend l'apprentissage machine vraiment utile.

3.4 Apprentissage profond

Le terme Deep Learning ou Apprentissage profond, a été introduit pour la première fois au ML par Dechter en 1986, et aux réseaux neuronaux artificiels par Aizenberg et al en 2000 [68].

Le Deep Learning est un nouveau domaine de recherche du ML, qui a été introduit dans le but de le rapprocher de son objectif principal l'intelligence artificielle, il concerne les algorithmes inspirés par la structure et le fonctionnement du cerveau [69]. Autrement dit, c'est un ensemble de méthodes d'apprentissage automatique utilisables pour modéliser à un haut niveau d'abstraction des données et avec lesquelles la machine apprend par elle-même à réaliser une tâche donnée. Par exemple, lorsque l'on parle de reconnaissance d'images, le système apprend automatiquement à classifier les images en différentes catégories selon le jeu de données qui lui est fourni pour reconnaître une automobile ou une motocyclette sur une image.

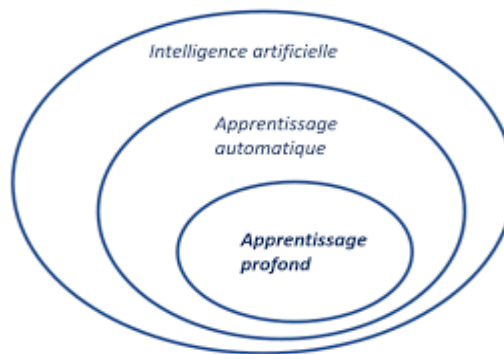


Figure 3.1 : La relation entre l'intelligence artificielle, le ML et le Deep Learning [70].

Le Deep Learning est basé sur l'idée des réseaux de neurones artificiels et il est taillé pour gérer de larges quantités de données en ajoutant des couches au réseau. Un modèle de deep learning a la capacité d'extraire des caractéristiques à partir des données brutes grâce aux multiples couches de traitement composé de multiples transformations linéaires et non linéaires et apprendre sur ces caractéristiques petit à petit à travers chaque couche avec une intervention humaine minimale [71].

3.4.1 Pourquoi le deep learning ?

Les algorithmes de ML décrits dans la première partie fonctionnent bien pour une grande variété de problèmes. Cependant ils ont échoué à résoudre quelques problèmes majeurs de l'IA telle que la reconnaissance vocale et la reconnaissance de visage. Le développement du deep learning fut motivé en partie par l'échec des algorithmes traditionnels dans de telle tâche de l'IA. Mais ce n'est qu'après que de plus grandes quantités de données ne soit disponibles grâce notamment au Big Data et aux objets connectés et que les machines de calcul soient devenues plus puissantes qu'on a pu comprendre le potentiel réel du Deep Learning.

Une des grandes différences entre le Deep Learning et les algorithmes de ML traditionnelles c'est qu'il s'adapte bien, plus la quantité de données fournie est grande plus les performances d'un algorithme de Deep Learning sont meilleures. Contrairement à plusieurs algorithmes de ML classiques qui possèdent une borne supérieure à la quantité de données qu'ils peuvent recevoir des fois appelée "plateau de performance", les modèles de Deep Learning 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.

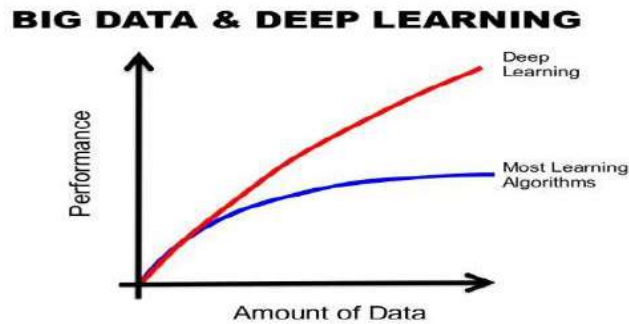


Figure 3.2 : La différence de performance entre le Deep Learning et la plupart des algorithmes de ML en fonction de la quantité de données [71].

Autre différence entre les algorithmes de ML traditionnelles et les algorithmes de Deep Learning c'est l'étape de l'extraction de caractéristiques. Dans les algorithmes de ML traditionnelles l'extraction de caractéristiques est faite manuellement, c'est une étape difficile et coûteuse en temps et requiert un spécialiste en la matière alors qu'en Deep Learning cette étape est exécutée automatiquement par l'algorithme.

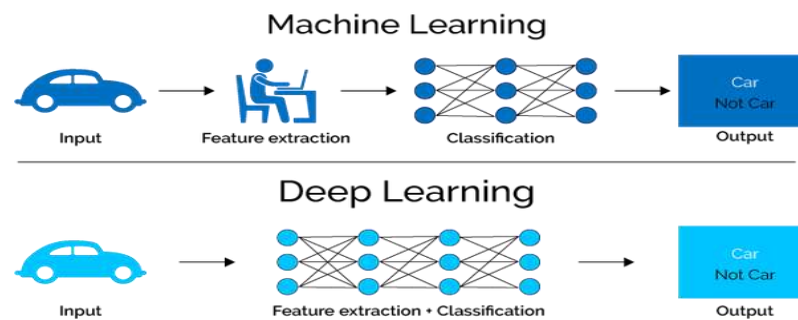


Figure 3.3 : L'apprentissage profond vs machine learning [72].

3.4.2 L'apprentissage

L'apprentissage est la propriété la plus intéressante des réseaux neuronaux. C'est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré.

L'apprentissage peut être *supervisé* ou *non supervisé* selon la présence ou l'absence de la réponse désirée.

- **L'apprentissage Supervisé** : est le type d'apprentissage le plus utilisé. Il consiste en des variables d'entrée (x) et une variable de sortie (Y), où l'algorithme d'apprentissage supervisé va trouver une fonction de mapping entre les variables prédictives en entrée X et la variable à prédire Y . la fonction de mapping décrivant la relation entre X et Y s'appelle un modèle de prédiction.

Il s'appelle l'apprentissage supervisé, car le processus d'un algorithme tiré de l'ensemble de données de formation (training set) peut être considéré comme un enseignant supervisant le processus d'apprentissage. Nous connaissons les réponses correctes, l'algorithme effectue des prédictions itératives sur les données d'apprentissage et est corrigé par l'enseignant. L'apprentissage s'arrête lorsque l'algorithme atteint un niveau de performance acceptable.

L'apprentissage supervisé est généralement effectué dans le contexte de la classification et de la régression.

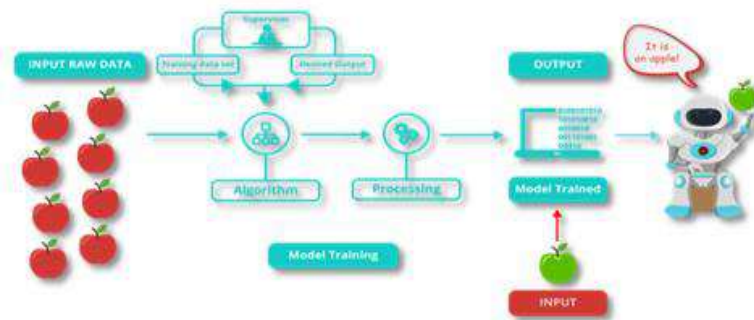


Figure 3.4: L'apprentissage Supervisé [73].

- **L'apprentissage non Supervisé** : consiste à ne disposer que de données d'entrée (X) et pas de variables de sortie correspondantes. Son objectif est de modéliser la structure ou la distribution sous-jacente dans les données afin d'en apprendre davantage sur les données.

On l'appelle apprentissage non supervisé car, contrairement à l'apprentissage supervisé, il n'y a pas de réponse correcte ni d'enseignant. Les algorithmes sont laissés à leurs propres mécanismes pour découvrir et présenter la structure intéressante des données. L'apprentissage non supervisé comprend deux catégories d'algorithmes: Algorithmes de regroupement, d'association et de réduction de la dimensionnalité.

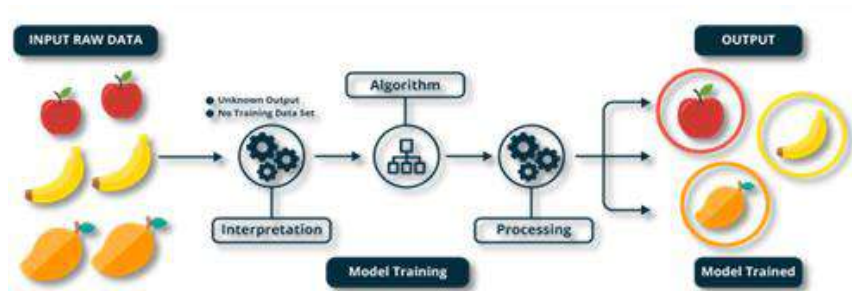


Figure 3.5 : L'apprentissage non Supervisé [74].

3.5 Réseau de neurone artificiel

Le réseau de neurone artificiel est un système de traitement de l'information né il y a une cinquantaine d'années et sont toujours en cours de développement. Ce système est inspiré du fonctionnement de neurones humain. Ils se composent d'un grand nombre de processeurs élémentaires reliés (neurone) travaillent ensemble, chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit, pour exécuter une tâche de classification donnée.

3.5.1 Le neurone

- **Modèle biologique**

L'élément principal du système nerveux central est le neurone. Le cerveau se compose d'environ mille milliards de neurones, avec 1000 à 10000 synapses (connexions) par neurone. Le neurone biologique est une cellule vivante spécialisée dans le traitement des signaux électriques. Les neurones sont reliés entre eux par des liaisons appelées axones, ces axones vont eux-mêmes jouer un rôle important dans le comportement logique de l'ensemble. Ces axones conduisent les signaux électriques de la sortie d'un neurone vers l'entrée (synapse) d'un autre neurone. Les neurones font une sommation des signaux reçus en entrée et en fonction du résultat obtenu vont fournir un courant en sortie [75]. (Figure 2.13)

La structure d'un neurone se compose de trois parties :

- ❖ **La somma** : ou cellule d'activité nerveuse, au centre du neuro
- ❖ **L'axone** : attaché au somma qui est électriquement actif, ce dernier conduit l'impulsion conduite par le neurone.
- ❖ **Dendrites** : électriquement passives, elles reçoivent les impulsions d'autres neurones.

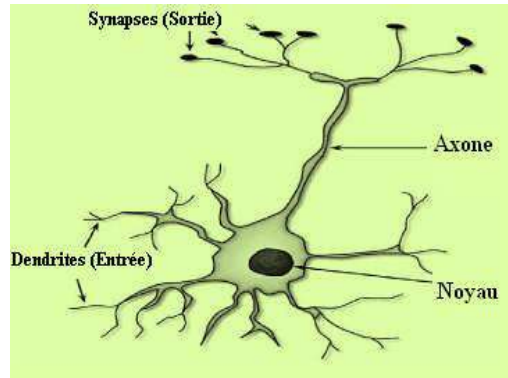


Figure 3.6 : Le neurone biologique [76].

- **Modèle formel (artificiel)**

Le neurone artificiel est une fonction algébrique non linéaire et bornée, dont la valeur dépend des paramètres appelés coefficients ou poids. Les variables de cette fonction sont habituellement appelées entrées du neurone, et la valeur de la fonction est appelée sa sortie [12]. Le neurone Artificiel est donc un modèle mathématique simplifié du neurone biologique.

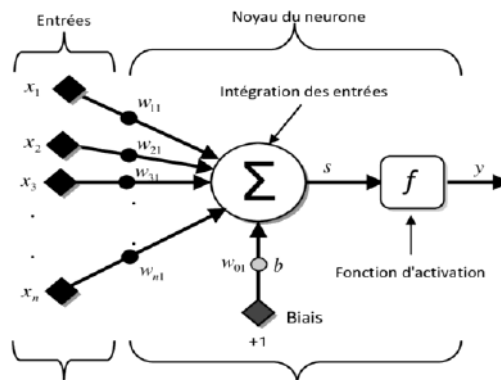


Figure 3.7 : Le neurone artificiel [78].

- **Les X_i** représentent les vecteurs d'entrées
- **Les W_{ij}** sont les poids synaptiques du neurone j .
- **Biais W_0** : entrée prend souvent les valeurs -1 ou $+1$ qui permet d'ajouter de la flexibilité
- **Noyau** : intègre toutes les entrées et le biais et calcul la sortie du neurone.
- **La fonction d'activation f** est la fonction d'activation du neurone. Il existe de nombreuses formes possibles pour la fonction d'activation, les plus courantes sont

présentées sur le tableau suivant. On remarquera qu'à la différence des neurones biologiques dont l'état est binaire, la plupart des fonctions d'activations sont des fonctions continues, offrant une infinité de valeurs possibles comprises dans l'intervalle $[0, +1]$ (ou $[-1, +1]$).










Nom de la fonction	Relation d'entrée/sortie	Icône
seuil	$a = 0$ si $n < 0$ $a = 1$ si $n \geq 0$	
seuil symétrique	$a = -1$ si $n < 0$ $a = 1$ si $n \geq 0$	
linéaire	$a = n$	
linéaire saturée	$a = 0$ si $n < 0$ $a = n$ si $0 \leq n \leq 1$ $a = 1$ si $n > 1$	
linéaire saturée symétrique	$a = -1$ si $n < -1$ $a = n$ si $-1 \leq n \leq 1$ $a = 1$ si $n > 1$	
linéaire positive	$a = 0$ si $n < 0$ $a = n$ si $n \geq 0$	
sigmoïde	$a = \frac{1}{1+\exp^{-n}}$	
tangente hyperbolique	$a = \frac{e^n - e^{-n}}{e^n + e^{-n}}$	
compétitive	$a = 1$ si n maximum $a = 0$ autrement	

Figure 3.8 : Les modèles de fonctions d'activation [79].

Donc le calcul de la sortie se fait en deux étapes :

1. Une combinaison linéaire des entrées : $S = \sum_{i=1}^n (W_i X_i) \pm b$
2. Application d'une fonction d'activation : $f(x) = f(S)$

• **Analogie entre le neurone biologique et le neurone formel**

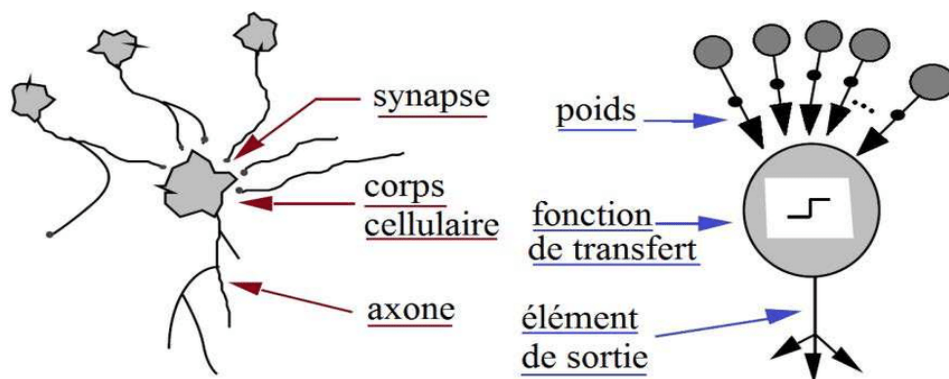


Figure 3.9 : Correspondance entre le neurone biologique et le neurone formel [80].

Le tableau suivant montre la transition entre le neurone biologique et le neurone artificiel.

Tableau 3.1 : Analogie entre neurone biologique et le neurone artificiel.

Neurone biologique :	Neurone artificiel :
Synapses	Poids de connexions
Axones	Signal de sortie
Dendrite	Signal d'entrée
Somma	Fonction d'activation

3.5.2 Architecteurs des réseaux de neurones

Les réseaux à couches sont les modèles connexionnistes les plus couramment utilisés. Leur architecture est organisée en couches successives, comprend une couche d'entrée, une couche de sortie et une ou plusieurs couches intermédiaires appelées couches cachées car elles ne sont pas vues de l'extérieur, chaque couche est composée d'un certain nombre de neurones. Les connexions sont établies entre les neurones appartenant à des couches successives mais les neurones d'une même couche ne peuvent pas communiquer entre eux dans le cas des réseaux à couches [81].

On distingue deux types de RNA : Les réseaux non bouclés et les réseaux bouclés :

- **Réseaux de neurones non bouclés**

Dans un réseau de neurones non bouclée, l'information circulant des entrées vers les sorties sans "retour en arrière" ; si l'on représente le réseau graphiquement, le graphe d'un réseau non bouclé est acyclique : Si l'on se déplace dans le réseau, à partir d'un neurone quelconque, en suivant les connexions, on ne peut pas revenir au neurone de départ [81].

La majorité des réseaux de neurones non bouclés sont mise en œuvre pour les tache de classification automatique sont organisé en plusieurs couches dont certaines sont cachées.

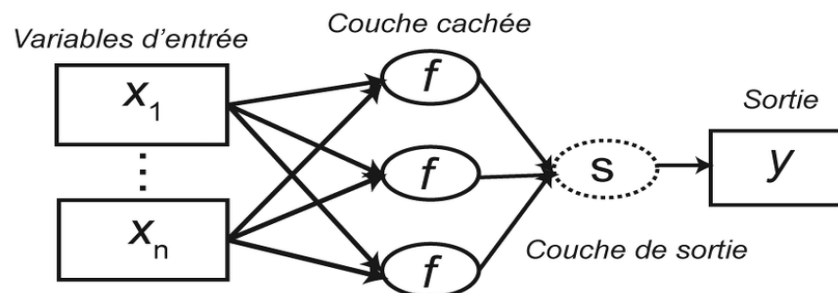


Figure 3.10 : Architecture de réseau non bouclé [82].

- **Réseaux de neurones bouclés**

Un réseau de neurones bouclé ou bien à récurrentes (RNN) signifie qu'une ou plusieurs sorties de neurones d'une couche aval sont connectées aux entrées des neurones de la couche amont ou de la même couche. Ces connexions récurrentes ramènent l'information en arrière par rapport au sens propagation défini dans un réseau non bouclé [81].

Contrairement aux réseaux de neurones non bouclés, le graphe de connexions des réseaux de neurones bouclés est cyclique : lorsqu'on se déplace dans le réseau, en suivant le sens des connexions, il est possible de trouver au moins un chemin qui revient à son point de départ (un tel chemin est désigné sous le terme de "Cycle").

Au cours des dernières années, un type de RNN est devenu la norme grâce à ses excellentes performances sur des tâches aussi nombreuses que variées : les réseaux de neurones à base de cellules Long Short-Term Memory (LSTM).

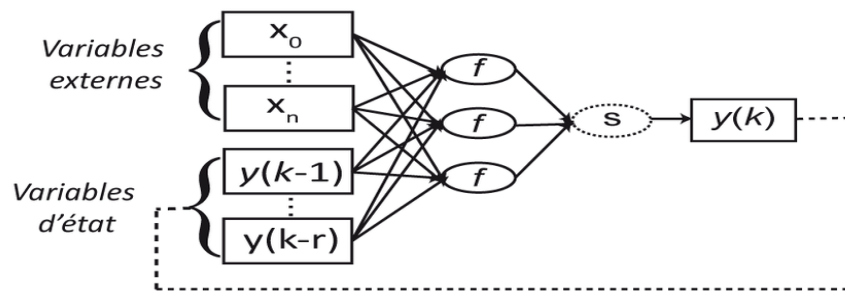


Figure 3.11 : Architecture de réseau bouclé [83].

3.5.3 Type des réseaux de neurones

- **Le perceptron monocouche**

Le perceptron est un réseau de neurones très simple qui a été proposé par le psychologue FRANK ROSENBLATT en 1958 et inspiré du système visuel. Il est linéaire et monocouche ayant juste deux couches, une couche représente les entrées de système et l'autre pour les sorties, acceptant uniquement des valeurs d'entrées et de sorties binaires (0 ou 1). Les connexions entre ces deux couches sont modifiables et bidirectionnelles.

Le procédé d'apprentissage est supervisé et le réseau est capable de résoudre des opérations logiques simples comme les opérations "ET logique" ou "OU logique" [84].

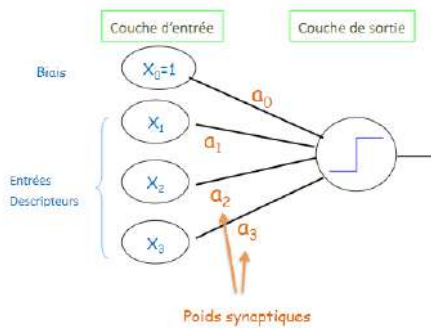


Figure 3.12 : perceptron monocouche [85].

La sortie du perceptron dépend de la somme des composantes X_i du vecteur d'entrée, pondérées par des poids réels W_i . Conformément aux notations utilisées dans ce travail, nous appellerons un potentiel, la somme pondérée notée par net.

$$net_k = \sum_{i=0}^n W_{ki} X_i \dots \dots \dots 3.1$$

La sortie du réseau sera désignée par 'O'. Elle sera de la forme suivante :

$$O_k = \begin{cases} 1 & \text{si } net_k \geq \theta_k \\ 0 & \text{si } net_k \leq \theta_k \end{cases} \dots \dots \dots 3.2$$

W_{ji} : Coefficients synaptiques de j vers i, X_i : sortie (ou état) de j, et θ_k : seuil

• **Le perceptron multicouche (Multi Layer Perceptron MLP)**

Les Perceptrons Multicouches sont des réseaux de neurones bouclés comportant au moins trois couches de neurones, en particulier une couche d'entrée, au moins une couche cachée et une seule couche de sortie. Les fonctions d'activation de la couche d'entrée suivent généralement la fonction identité, tandis que les fonctions d'activation des neurones dans les couches cachées et dans la couche de sortie suivent d'une manière générale une fonction mathématique avec des dérivées continues premières et secondes. Citons comme exemple de fonctions d'activation les fonctions identité, logistique sigmoïde, tangente hyperbolique, exponentielle et sinus [86].

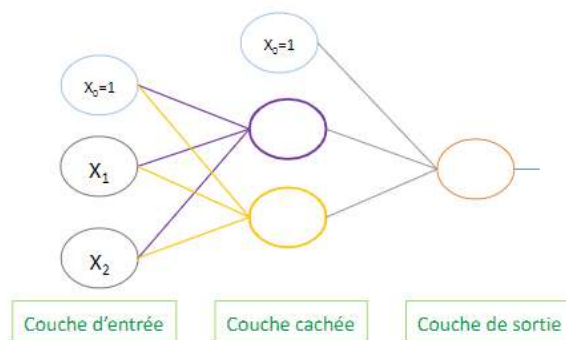


Figure 3.13 : perceptron multicouche [85].

• Le Modèle de Hopfield

Le modèle de Hopfield fut présenté en 1982. Ce modèle très simple est basé sur le principe des mémoires associatives. C'est d'ailleurs la raison pour laquelle ce type de réseau est dit associatif (par analogie avec le pointeur qui permet de récupérer le contenu d'une case mémoire). Le modèle de Hopfield utilise l'architecture des réseaux entièrement connectés et récurrents (dont les connexions sont non orientées et où chaque neurone n'agit pas sur lui-même). Les sorties sont en fonction des entrées et du dernier état pris par le réseau [75].

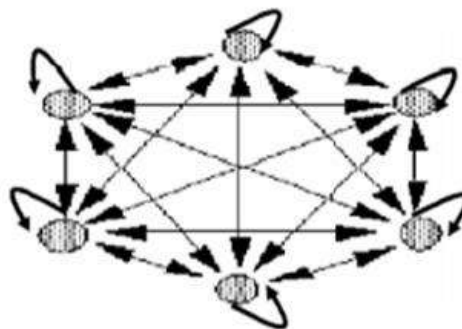


Figure 3.14 : Réseau de Hopfield [75].

• Le Modèle d'Adaline

ADALINE (Adaptive Linear Neuron ou plus tard Adaptive Linear Element) est un réseau de neurones artificiels simple-couche. Il a été développé par le professeur Bernard Widrow et un de ses étudiants, Ted et Hoff de l'université Stanford en 1960. Ce modèle est similaire au modèle de perceptron, il s'agit en fait d'un Perceptron sans saturation des sorties, et la fonction de transfert change, mais reste toujours linéaire c'est-à-dire : $F(x) = x$.

La règle d'apprentissage de ce réseau consiste à minimiser l'erreur quadratique en sortie du réseau de neurone. La règle d'apprentissage est identique à celle du Perceptron, à la différence près que ce sont les entrées non-saturées qui sont prises en compte [86].

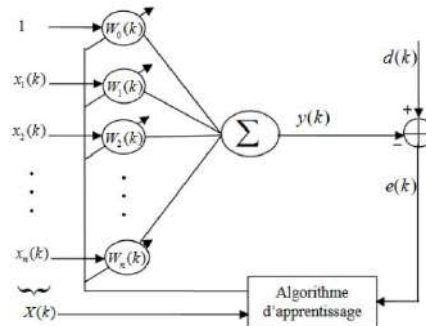


Figure 3.15 : Structure du modèle Adaline [86].

La sortie estimée $y(k)$ du signal de référence $d(k)$ sera composée par la relation linéaire suivante :

$$Y(k) = W_T(k) X(k) \dots \dots \dots 3.3$$

Avec $W_T(k) = [W_0(k) \ W_1(k) \ W_2(k) \ \dots \ W_n(k)] \dots$

Et $X_T(k) = [1 \ x_1(k) \ x_2(k) \ \dots \ x_n(k)] \dots$

- **Le réseau à fonction de base radiale**

Les réseaux de neurones à base radiale RBFs (Radial Basis Function) sont une classe particulière des réseaux de neurones multicouches. Ces réseaux sont une architecture Feedforward puissante, à trois couches: une couche d'entrée, une couche cachée composée de fonctions noyaux et une couche de sortie, dont les neurones sont généralement animés par une fonction d'activation linéaire. Ce type de réseaux a été introduit pour la première fois par Hardy, et la théorie correspondante a été développée par Powell, ensuite, ces réseaux ont pris le terme de réseaux de neurones grâce à Broomhead et Lowe.

Les réseaux de neurones sont principalement utilisés pour résoudre des problèmes d'approximation de fonctions dans des espaces de grandes dimensions. Ils sont lus adaptés, en raison d'apprentissage local.

Un réseau RBF est plus facile à construire qu'un PMC, et apprend beaucoup plus vite, mais sur les problèmes difficiles les réseaux RBF ne peuvent pas lutter avec les réseaux PMC, en terme de qualité des résultats [77].

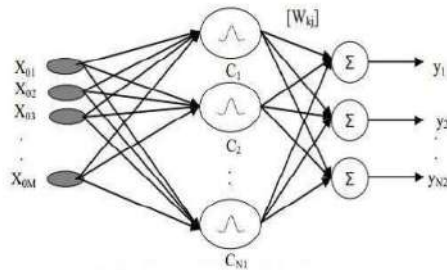


Figure 3.16 : Architecture du modèle Adaline [81].

La sortie du réseau est simplement une combinaison linéaire des sorties des neurones RBF multipliés par le poids de leur connexion respective.

- **Le Modèle de Kohonen**

Réseau de Kohonen, ou bien carte auto-organisatrice de Kohonen. Il s'agit d'un réseau non supervisé avec un apprentissage compétitif ou l'on apprend non seulement à modéliser l'espace des entrées avec des prototypes, mais également à construire une carte à une ou deux dimensions permettant de structurer cet espace. Les neurones de ce réseau sont constitués d'un vecteur de poids dans l'espace des entrées. La carte des neurones définit quant à elle des relations de voisinage entre les neurones. Les cartes de Kohonen peuvent être utilisées dans le cadre de la projection de données multi-variées, d'approximation de densité ou de classification. Elles ont été utilisées avec succès en reconnaissance de la parole, classification et traitement d'images.

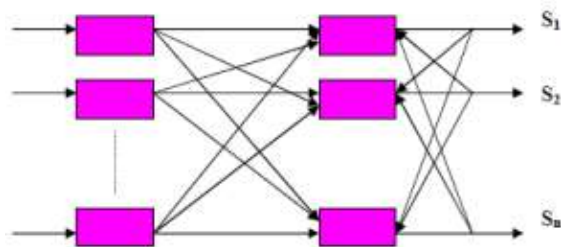


Figure 3.17 : Le modèle de Kohonen [81].

Les cartes de Kohonen sont réalisées à partir d'un réseau à deux couches, une en entrée et une en sortie. Notons que les neurones de la couche d'entrée sont entièrement connectés à la couche de sortie [81].

3.5.4 L'apprentissage profond et Les réseaux de neurones artificiels

La plupart des algorithmes d'apprentissage profond reposent sur des modèles appelés réseaux de neurones artificiels. Il existe de nombreuses de ces réseaux qui ont été adaptées pour des problématiques spécifiques, l'un type de modèles qui sont très utilisés pour la classification d'images sont : les réseaux de neurones convolutifs.

3.6 Les réseaux de neurones convolutifs (CNN)

Le réseau de neurone convolutif ou convolutifs est proposé initialement par le chercheur français Yann LeCun [81], c'est est un type particulier de réseau de neurone qui se base sur l'opération de convolution, son but principal est d'extraire des caractéristiques d'une image donnée.

Les réseaux convolutifs sont les structures les plus performantes pour faire la classification ou la reconnaissance des images [69]. Un réseau est dit convolutif quand chaque neurone reçoit ses informations non pas de toute la couche précédente, mais seulement des neurones situés dans son champ réceptif.

3.6.1 Présentation

Les réseaux de neurones convolutifs se représente en général sous la forme suivante :

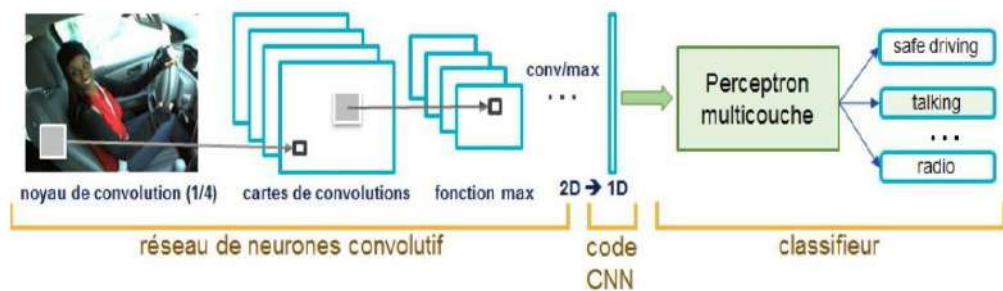


Figure 3.18 : L'architecture standard d'un réseau de neurone convolutionnels [68].

Les réseaux de neurone convolutifs (CNN en anglais) comportent deux parties bien distinctes. En entrée, une image est fournie sous la forme d'une matrice de pixels, 2 dimensions pour une image en niveaux de gris. La couleur est représentée par une troisième dimension, de profondeur 3 pour représenter les couleurs fondamentales [Rouge, Vert, Bleu].

La première partie d'un CNN est la partie convolutive à proprement parler. Elle 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 [68].

Ce code CNN en sortie de la partie convolutive est ensuite 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 une dernière couche comportant un neurone par catégorie. Les valeurs numériques obtenues sont généralement normalisées entre 0 et 1, de somme 1, pour produire une distribution de probabilité sur les catégories.

3.6.2 Architecture de réseaux de neurone convolutif

- **La Couche de Convolution**

Quand on lui présente une nouvelle image, le CNN ne sait pas exactement si les caractéristiques seront présentes sur le bord ou le centre de l'image, il cherche donc à les trouver dans toute l'image et dans n'importe quelle position. En calculant dans toute l'image si une caractéristique est présente, nous faisons un filtrage. Les mathématiques que nous utilisons pour réaliser cette opération sont appelés une convolution, de laquelle les réseaux de neurones à convolution tiennent leur nom [86].

Les couches convolutives constituent le cœur des réseaux convolutifs. Elles se composent d'une grille rectangulaire de neurones ayant un petit champ réceptif étendu à travers toute la profondeur du volume d'entrée.

Elle fonctionne comme un extracteur de caractéristiques des images par l'opération de filtrage sur l'image. Pour effectuer cette opération sur toute l'image, on définit une fenêtre de voisinage de taille ($s*s$) qui va se déplacer à travers toute l'image. Au tout début de la convolution, la fenêtre sera positionnée tout en haut à gauche de l'image puis elle va se décaler d'un certain nombre de cases (appelé le pas) vers la droite et lorsqu'elle arrivera au bout de l'image, elle se décalera d'un pas vers le bas ainsi de suite jusqu'à ce que le filtre est parcourue la totalité de l'image.

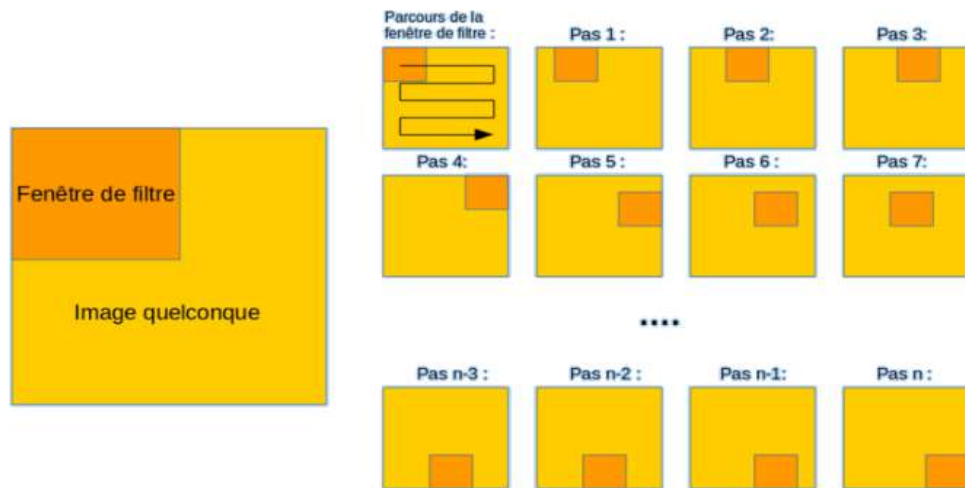


Figure 3.19 : Schéma du parcours de fenêtré de filtre sur l'image [68].

Trois hyper paramètres permettent de dimensionner le volume de la couche de convolution :

1. **Profondeur de la couche K :** nombre de noyaux de convolution
2. **Le pas S :** contrôle le chevauchement des champs récepteurs
3. **La marge (à 0) ou zero padding P :** Cette marge permet de contrôler la dimension spatiale du volume de sortie.

Pour chaque image de taille $W \times H \times D$ en entrée, où W est sa largeur en pixels, H sa hauteur en pixels et D le nombre de canaux (1 pour une image en noir et blanc, 3 pour une image en couleurs) la couche de convolution renvoie une matrice de dimensions $WC \times HC \times DC$, où

$$WC = \frac{w-F+2P}{s} + 1 \quad \text{et} \quad HC = \frac{w-F+2P}{s} \quad \text{et} \quad DC = k$$

Où F est la taille des filtres et le S c'est le pas avec lequel on fait glisser la fenêtré, le P correspond au zéro padding, et K c'est le nombre des filtres.

Choisir $P = \frac{F-1}{2}$ et $S=1$ permet ainsi d'obtenir des features maps de même largeur et hauteur que celles reçues en entrée, Où F est la taille du filtre, chaque filtre est de dimensions $F \times F \times D$ pixels

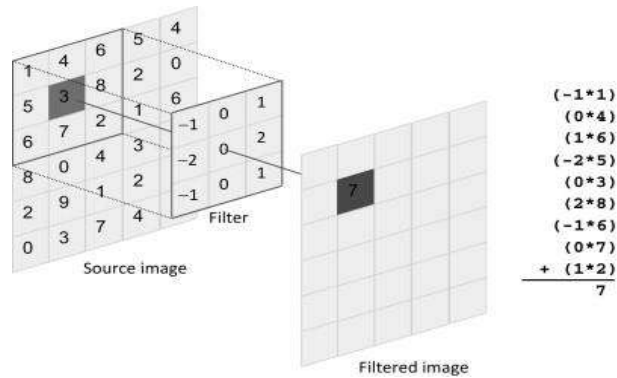


Figure 3.20 : L'opération de convolution [87].

• **La Couche de sous-échantillonnage (Pooling)**

Un autre outil très puissant utilisé par les CNNs s'appelle le Pooling. Qui est une méthode permettant de prendre une large image et d'en réduire la taille tout en préservant les informations les plus importantes qu'elle contient [86]. Ce qui est une forme de sous-échantillonnage de l'image [88].

Ce type de couche est souvent placé entre deux couches de convolution, elle reçoit en entrée plusieurs features maps et applique à chacune d'entre elles l'opération de pooling.

Les deux méthodes les plus utilisées pour appliquer cette opération sont les suivantes, soit on fait la moyenne des valeurs de la zone (pooling average), soit on extrait uniquement la valeur la plus élevée (pooling max). Nous utilisons dans ce travail celle qui applique l'opération max au résultat de chaque filtre [86].

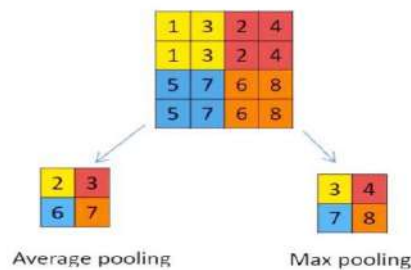


Figure 3.21 : Les deux méthodes de Pooling [69].

Pour cela, on découpe l'image en cellules régulières, puis on garde au sein de chaque cellule la valeur maximale. En pratique, on utilise souvent des cellules carrées de petite taille pour ne pas perdre trop d'informations.

La forme la plus courante est une couche de mise en commun avec des tuiles de taille 2×2 (largeur/hauteur) et comme valeur de sortie la valeur maximale en entrée son utilité consiste à réduire la quantité de poids à apprendre, ce qui réduit le temps de calcul ainsi que la probabilité de sur apprentissage, on parle dans ce cas de « Max-Pool 2x2 ».

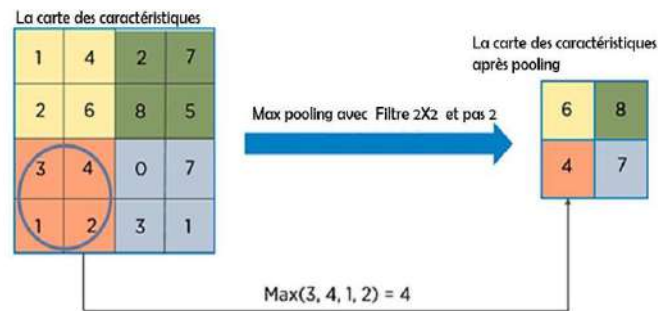


Figure 3.22 : Pooling avec un filtre 2x2 et un pas de 2 [69].

- Couches de correction (Relu)

Souvent, il est possible d'améliorer l'efficacité du traitement en intercalant entre les couches de traitement une couche qui va opérer une fonction mathématique (fonction d'activation) sur les signaux de sortie. La couche de correction Relu (abréviation de Unité Linéaire Rectifiée) désigne la fonction réelle non- linéaire définie par $\text{ReLU}(x) = \max(0, x)$.

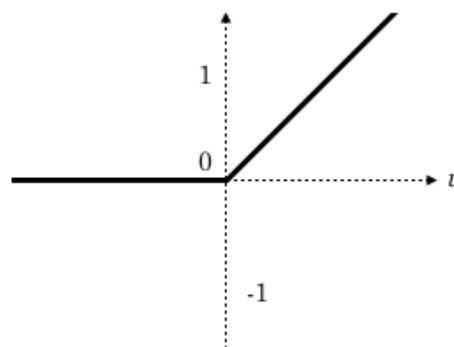


Figure 3.23 : Fonction relu [86].

Cette fonction, appelée aussi « fonction d'activation non saturante », augmente les propriétés non linéaires de la fonction de décision et de l'ensemble du réseau sans affecter les champs récepteurs de la couche de convolution [88].

La couche de correction ReLU remplace donc toutes les valeurs négatives reçues en entrées par des zéros. Elle joue le rôle de fonction d'activation souvent, la correction Relu est préférable, mais il existe d'autre forme [68] :

La correction par **tangente hyperbolique** : $f(x) = \tanh(x)$

La correction par la **tangente hyperbolique saturante** : $f(x) = |\tanh(x)|$

La correction par la **fonction sigmoïde** : $f(x) = (1 + e^{-x})^{-1}$

- **La Couche entièrement connectée (Fully-Connected)**

La couche fully-connected est un perceptron ou bien un MLP (multi layer perceptron) qui constitue toujours la dernière couche d'un réseau de neurones, convolutif ou non – elle n'est donc pas caractéristique d'un CNN [68].

Après l'extraction des caractéristiques des entrées à travers plusieurs couches de convolution et de max-pooling en un vecteur 1D, le perceptron prend comme entrée les caractéristiques extraites et produit un vecteur de taille N, où N est le nombre de classes dans notre problème de classification d'images. Chaque élément du vecteur indique la probabilité pour l'image en entrée d'appartenir à une classe, les probabilités d'appartenance à chaque classe sont calculées à l'aide de la fonction d'activation de type sigmoïde ou softmax.

La couche FC consiste à multiplier le vecteur en entrée par la matrice contenant les poids. Le fait que chaque valeur en entrée soit connectée avec toutes les valeurs en sortie explique le terme fully-connected [86].

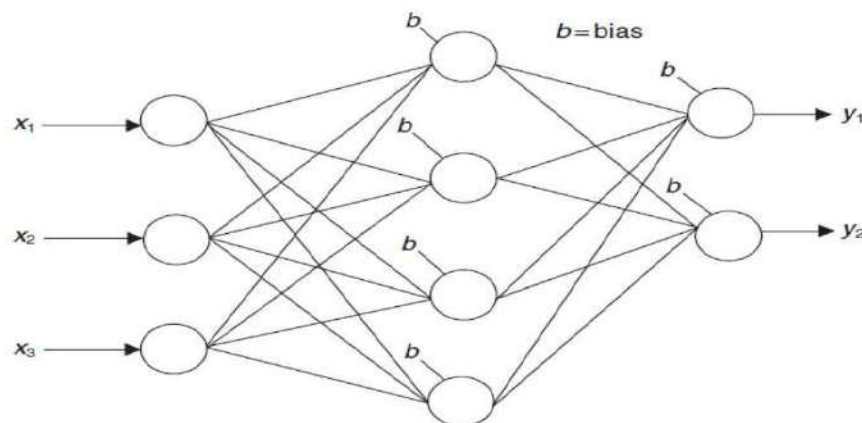


Figure 3.24 : La Couche de Fully-Connected [89].

3.6.3 Outils d'optimisation des réseaux convolutifs

Les réseaux de neurones convolutifs utilisent différents paramètres d'optimisation. Ces méthodes permettant une meilleure minimisation et optimisation associée au réseau CNN. La plupart de ces méthodes sont utilisées pour le développement et l'apprentissage des réseaux de neurones convolutifs

- **La Batch Normalization**

La technique de Batch Normalization a été introduite récemment en 2015 afin d'apprendre les CNN de manière plus rapide et efficace. Son idée est normaliser les valeurs entre 0 et 1 par exemple. En toute rigueur, le Batch Normalization normalise les données en soustrayant la moyenne du batch de données et en divisant par sa moyenne. Cette opération va permettre de rendre le réseau plus stable et aussi va permettre d'accélérer sa phase d'apprentissage. D'un point de vue mathématique, on ne perd pas d'information en passant de valeurs entre -10 et 10 qu'entre -1 et 1. Cela permet en outre de réduire les écarts grossiers qui pourraient avoir un trop gros impact sur la mise à jour des poids (lors de la descente du gradient). Par ailleurs elle permet aussi de le rendre plus générique en terme de prédiction. En effet supposons qu'on ait entraîné un réseau à détecter des chats dans des centaines d'images mais que les chats soient majoritairement d'une couleur sombre (noir, gris). Sans Batch Normalisation, cette couleur aura mettons pour valeur 5. Si maintenant le réseau est confronté à des chats dont la robe est de couleur plus vive comme l'orange ou le blanc, ce champ de valeur pour ces couleurs allant de 50 à 100, le réseau serait perdu et reconnaîtrait moins les chats de cette couleur. Cet exemple montre l'importance d'une couche de Batch Normalisation, et ce après chaque couche de convolution si possible [92].

- **Les fonctions de perte (Loss)**

Une fonction de perte représente une certaine mesure de la différence entre les valeurs observées des données et les valeurs calculées. C'est la fonction qui est minimisée dans la procédure d'optimisation d'un modèle [91]. Ce type de fonction est dépendant de la tâche que le réseau doit effectuer (classification, régression...).

La fonction (ou couche) de perte gère les ajustements de poids sur le réseau. Avant que l'entraînement du réseau ne commence, les poids des couches convolution et entièrement connectées reçoivent des valeurs aléatoires. Ensuite, pendant l'entraînement, la couche de perte vérifie en permanence les suppositions de la couche entièrement connectée par

rapport aux valeurs réelles dans le but de minimiser autant que possible la différence entre l'estimation et la valeur réelle. La couche de perte effectue cela en ajustant les poids à la fois dans la couche de convolution et dans les couches entièrement connectées [86].

- **Méthodes de régularisation**

Les méthodes de régularisation sont des variations introduites dans l'apprentissage d'un réseau pour limiter le surapprentissage et à contrôler l'erreur de type variance pour aboutir à de meilleures performances, autrement dit à diminuer l'erreur sur les nouvelles données si elles suivent les mêmes lois que les données d'apprentissage. La régularisation est introduite dans l'algorithme pour réduire l'erreur de généralisation sans perturber l'erreur d'apprentissage. Différents types de régularisation ont été introduit peuvent être envisagées .

- **le Dropout (Srivastava et al.,2014)**, est une méthode qui consiste à retirer de l'apprentissage une partie des neurones du réseau (en général 15 à 50% par couche) lors de l'entraînement comme montré par *la Figure 3.24*. Le Dropout s'attaque au problème de coadaptation des neurones dans le réseau. La coadaptation est une situation où plusieurs neurones d'une même couche sont utilisés pour modéliser une information [93]. D'après les travaux de (Srivastava et al.,2014) les coadaptations dites complexes apprises par un réseau ne sont pas toujours nécessaires et introduisent deux problèmes :

1. la diminution des capacités de modélisation du réseau. Si plusieurs neurones modélisent la même information, alors ils sont perdus pour en modéliser de nouvelles.
2. une tendance au surapprentissage. En effet des coadaptations simples généralisent mieux que les coadaptations complexes.

Le Dropout est donc une solution peu couteuse en calcul qui va permettre d'améliorer les capacités de modélisation et de généralisation du réseau. En contre partie, le nombre d'itérations nécessaires à l'apprentissage est augmenté.

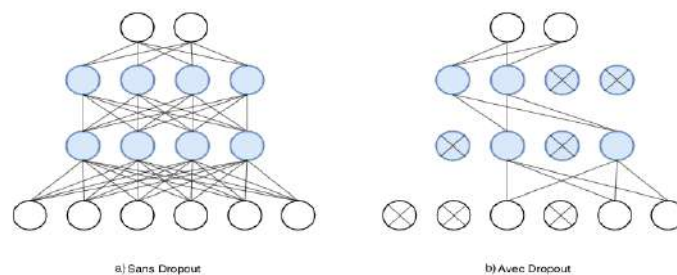


Figure 3.25 : Schéma de deux réseaux de neurones avec et sans dropout [93].

- Pooling Stochastique

Le pooling stochastique est une méthode de régularisation inspirée par le dropout. Au lieu de toujours capturer la plus forte activité au sein de chaque région de pooling comme le fait le MaxPooling, le pooling stochastique sélectionne les activations de manière aléatoire en fonction d'une distribution multinomiale [86].

3.5.4 Quelques réseaux convolutifs célèbres

- **LeNet (1990)**

Les premières applications réussies des réseaux convolutifs ont été développées par Yann LeCun dans les années 1990. Parmi ceux-ci, le plus connu est l'architecture **LeNet** utilisée pour lire les codes postaux, les chiffres, etc [71].

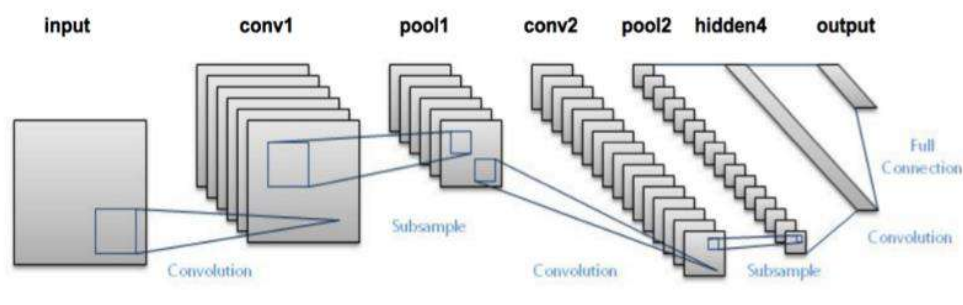


Figure 3.26 : L'architecture LeNet [90].

- **AlexNet (2012)**

Le premier travail qui a popularisé les réseaux convolutifs dans la vision par ordinateur était AlexNet, développé par Alex Krizhevsky, Ilya Sutskever et Geoff Hinton. Ce CNN été soumis au défi de la base ImageNet en 2012 et a nettement surpassé ses concurrents. Le réseau avait une architecture très similaire à LeNet, mais était plus profond, plus grand et comportait des couches convolutives empilées les unes sur les autres (auparavant, il était commun de ne disposer que d'une seule couche convolutifs toujours immédiatement suivie d'une couche de pooling) [71].

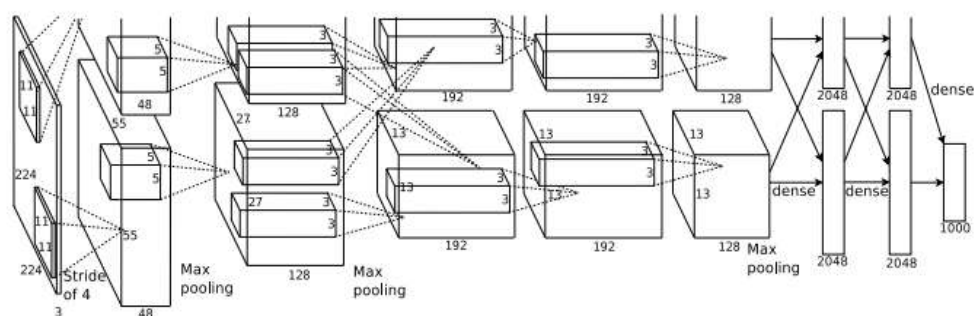


Figure 3.27 : L'architecture AlexNet [90].

- **ZFNet (2013)**

C'était une amélioration d'AlexNet en ajustant les hyper-paramètres de l'architecture, en particulier en élargissant la taille des couches convolutifs et en réduisant la taille du noyau sur la première couche [69].

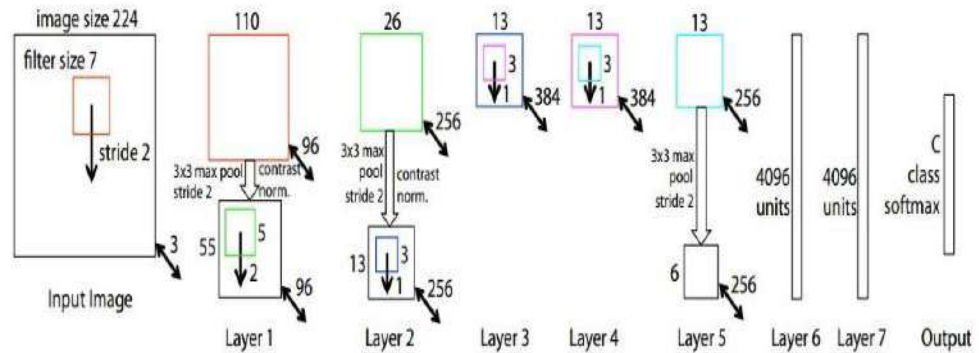


Figure 3.28 : L'architecture ZFNet [90].

- **GoogLeNet (2014)**

C'est un modèle de Google. Sa principale contribution a été le développement d'un module inception qui a considérablement réduit le nombre de paramètres dans le réseau (4M, par rapport à AlexNet avec 60M). En outre, ce module utilise le global Average pooling ce qui élimine une grande quantité de paramètres. Il existe également plusieurs versions de GoogLeNet, parmi elles, **Inception-v4** et **Xception** ce dernier est l'un des modèles auxquels notre système s'inspire, plus de détails dans le chapitre de conception [69].

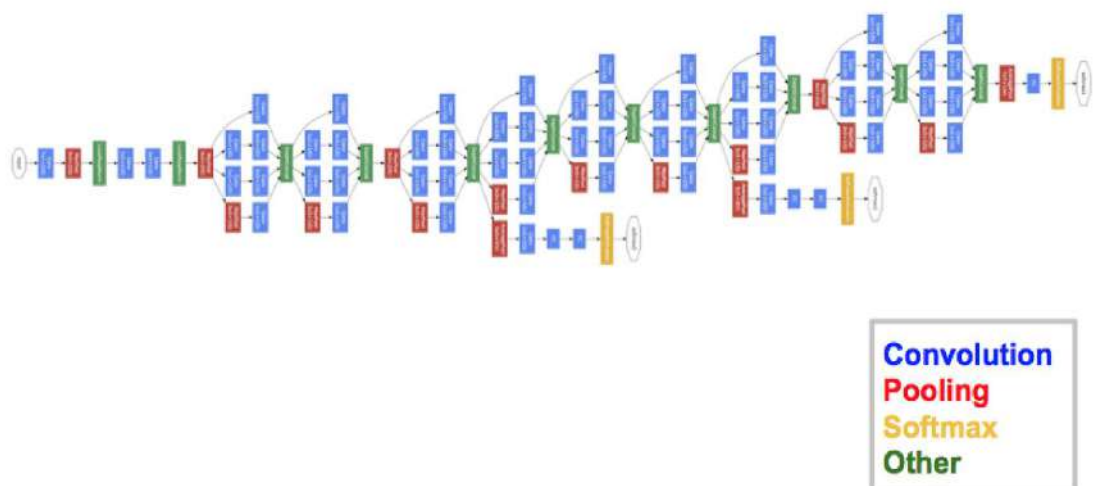


Figure 3.29 : L'architecture GoogLeNet/ Inception [90].

• **VGGNet (2014)**

Il s'agit d'une structure du Visual Geometry Group d'Oxford réalisée par Andrea Vedaldi et Andrew Zisserman, Sa principale contribution a été de montrer que la profondeur du réseau est un composant essentiel pour de bonnes performances. Son inconvénient est qu'il est plus coûteux à évaluer et utilise beaucoup plus de mémoire et de paramètres (140M).

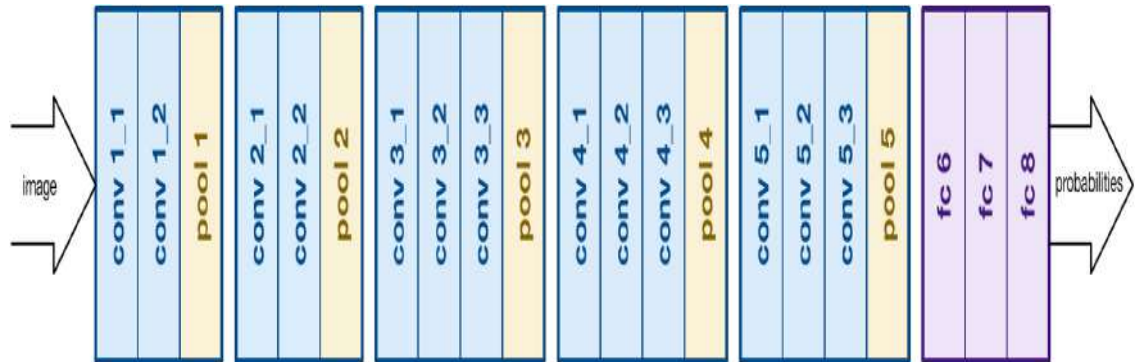


Figure 3.30 : L'architecture VGGNet [90].

• **ResNet (2015)**

Residual network développé par Kaiming He et al, il a été le vainqueur de ILSVRC 2015. Il présente des sauts de connexion et une forte utilisation de la batch normalisation. Il utilise aussi le global AVG pooling au lieu du PMC à la fin [71].

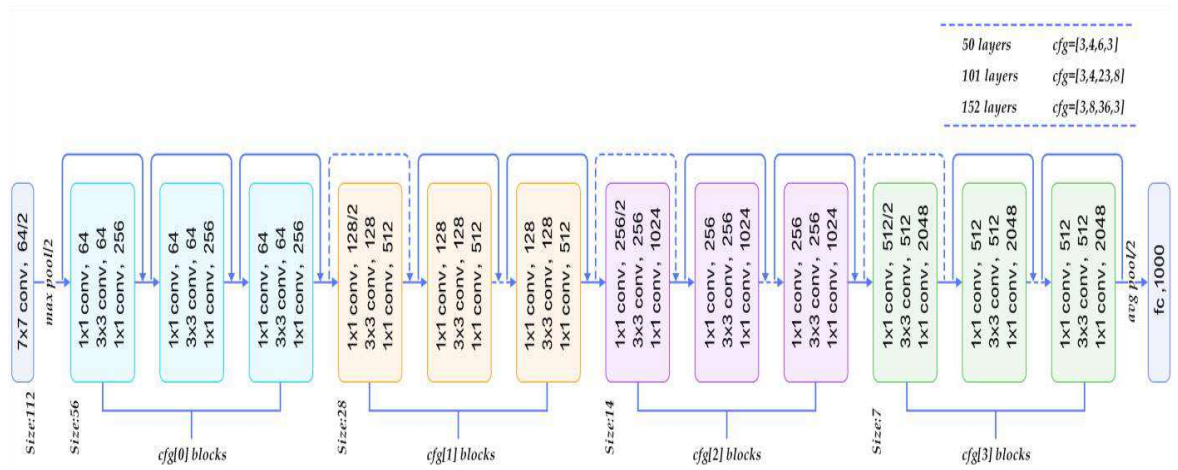


Figure 3.31 : L'architecture ResNet [90].

3.6 Conclusion

Les réseaux de neurones artificiels ont réussi à se faire une place de choix dans des domaines tels que le traitement de l'image. Des résultats remarquables ont été obtenus par exemple en détection d'objets et la reconnaissance de forme, L'état de l'art et les possibilités de techniques utilisés en Deep Learning sont conséquents, et représentatifs ce qui nous a motivé pour l'utilisation de cette technique dans ce travail. Le réseau de neurone convolutif est l'un des techniques de deep learning les plus utilisées, puisqu'il intègre les deux étapes de reconnaissance de forme au sein d'un même réseau.

Dans ce chapitre nous avons présenté les concepts liés à l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond et la relation qu'ils ont avec, après nous avons présenté les réseaux de neurones et leurs différents types ensuite on s'est focalisé notre attention sur les réseaux de neurones convolutifs CNN, leur structure, et ses différentes couches, nous avons présenté enfin quelques exemples d'architectures. Cette étude détaillée du modèle CNN nous permet de mieux comprendre son processus et de proposer notre modèle CNN qui fera l'objet du chapitre 4.

Chapitre 04 :
Expérimentations,
Résultats et
Discussions

4.1 Introduction

Pour évaluer l'efficacité des méthodes décrites dans les chapitres précédant, nous avons opté pour une validation sur des bases de données standard « ORL », « Essex », « Casia-WebFace » et « FEI ». Notre travail consiste à concevoir un système d'identification biométrique de personnes, de prédiction de genre et estimation d'âge par reconnaissance faciale se basant sur un réseau de neurone convolutifs. Plusieurs étapes sont nécessaires, en commençant par l'étape d'extraction des caractéristiques qui est la plus importante car les performances du système en dépendent (résultats et robustesse). Nous évaluerons dans ce chapitre les résultats obtenus sur les bases de données utilisées sous différentes conditions et attributs des méthodes.

4.2 Les bases de données de visage

4.2.1 ORL

Les images faciales de la base de données ORL ont été prises entre avril 1992 et avril 1994 par le laboratoire AT&T à l'université de Cambridge en Angleterre. Il y a dix images différentes de chacun des 40 sujets distincts. Pour certains sujets, les images ont été prises à des moments différents, variant l'éclairage, les expressions faciales (yeux ouverts / fermés, souriant / ne souriant pas) et les détails du visage (lunettes / pas de lunettes). Toutes les images ont été prises sur un fond sombre et homogène avec les sujets en position frontale verticale (avec une tolérance pour certains mouvements latéraux). Les fichiers sont au format PGM et la taille de chaque image est de 92x112 pixels, avec 256 niveaux de gris par pixel [94].



Figure 4.1 : Exemples d'images de visages de la base ORL.

4.2.2 FEI

La base de données de visage FEI est une base brésilienne qui contient un ensemble d'images de visage prises entre juin 2005 et mars 2006 au laboratoire d'intelligence artificielle FEI de São Bernardo font Campo, São Paulo, au Brésil. Elle contient des images de visage de 200 personnes, avec 14 images pour chacune d'entre elles, un total de 2800 images. Toutes les images sont en couleur et prises sur un fond blanc homogène en position frontale avec divers angles de pose de chaque personne, s'étendant du profil aux vues frontales. La variation d'éclairage est de 10% et la résolution de l'image est de 640x480 Pixel [95].



Figure 4.2 : Exemples d'images de visages de la base FEI.

4.2.3 Essex

Les images de la base ESSEX sont capturées à une distance fixe avec une orientation buccale différente sous différentes expressions faciales. La base de données se compose de 153 images d'individus de résolution 180 par 200 pixels repartit entre femmes et d'hommes (20 images chacun) avec un fond vert uni sans échelle mais présente une variation très mineure du tour de tête, d'inclinaison et d'éclairage [94].



Figure 4.3 : Exemples d'images de visages de la base Essex.

4.2.4 Casia WebFace

L'ensemble de données CASIA WebFace a été collecté à des fins de reconnaissance faciale par Yi et al en 2014. Cette base de données contient des photos d'acteurs et d'actrices nés entre 1940 et 2014 [94], incluent des variations aléatoires de poses, d'illuminations, d'expressions faciales et de résolutions d'images. Au total, il y a 494 414 images de visage de 10 575 sujets [96]. À notre connaissance, CASIA WebFace est le plus grand ensemble de données de visage accessible au public aujourd'hui, et c'est pourquoi nous l'avons utilisé pour l'apprentissage du réseau CNN dans ce travail. Les auteurs de CASIA WebFace fournissent les noms de 10 575 sujets mais pas leurs sexes. Nous avons annoté les sexes en utilisant les métadonnées fournies par IMDb et également par annotation manuelle.



Figure 4.4 : Exemples d'images de visages de la base CASIA WebFace.

4.2.5 Séparation des bases de données

Afin de développer une application de reconnaissance de visage, il est nécessaire de disposer de deux bases de données : une base pour effectuer l'apprentissage et une autre pour tester les techniques et déterminer leurs performances, mais Il n'y a pas de règle pour déterminer ce partage de manière quantitatif. Il résulte souvent d'un compromis tenant compte du nombre de données dont on dispose et du temps pour effectuer l'apprentissage. Dans ce travail nous avons utilisé 5 bases de données faciales pour l'évaluation de notre système de reconnaissance de visage.

- **Base ORL:** contient 20 personnes chaque personne possède 10 des images différentes.
- **Base FEI:** contient 20 personnes chaque personne possède 10 des images différentes.
- **Casia WebFace :** contient 20 personnes chaque personne possède 10 des images différentes.
- **Essex :** contient 33 personnes chaque personnes possède 10 des images différentes.
- **Notre base :** contient 20 personnes chaque personne possède 10 des images différentes.

Dans la série de tests que nous avons effectuée, la base de données était divisée comme suit :

- **Images apprentissages** : Les 8 premières images servent pour la phase d'apprentissage.
- **Images Tests** : Les 2 dernières images de chaque individu nous ont servies pour la réalisation des différents tests.

4.3 Environnement du travail

Dans cette section, nous présenterons les environnements matériels et logiciel utilisés pour l'élaboration de notre travail.

4.3.1 Environnement matériel

Afin de mettre en œuvre ce projet, nous avons besoin d'un ensemble de matériel dont les caractéristiques sont les suivantes :

Un ordinateur DELL avec les caractéristiques suivantes :

- Processeur : Intel(R) Core (TM) i5-7200U CPU @ 2.50 Ghz 2.70 Ghz.
- Mémoire installée (RAM) : 8.00 Go.
- Disque Dur : 500 Go.
- Type du système : système d'exploitation 64 bits.
- OS : Microsoft Windows 10.

4.3.2 Outils de développement

Nous avons eu recours lors de l'élaboration de notre système au langage python, qui est un langage de programmation interprété de haut niveau et à usage général (open sources). Créée par Guido van Rossum et publiée pour la première fois en 1991, la philosophie de conception de Python met l'accent sur la lisibilité du code avec son utilisation notable, ses constructions de langage et son approche orientée objet visent à aider les programmeurs à écrire un code clair et logique pour des projets à petite et à grande échelle. Python prend en charge plusieurs paradigmes de programmation, y compris la programmation structurée (en particulier, procédurale), orientée objet et fonctionnelle. Python est souvent décrit comme un langage "batteries incluses" en raison de sa bibliothèque standard complète [97].

4.4 Système proposé : principes et expérimentations

Le problème de la reconnaissance faciale est défini tel qu'à partir de l'image du visage, la personne correspondante doit être identifiée. Pour ce faire, il est nécessaire d'obtenir une image de référence (images d'apprentissage) sous la forme d'une base de données de tous les visages connus du système. Chaque image est associée à un vecteur de propriétés qui ne sont pas fixes pour la même personne et varient d'une personne à l'autre. La reconnaissance consiste alors à comparer le vecteur caractéristique du visage à reconnaître avec chaque vecteur de la base

d'apprentissage. Autrement dit, trouver la personne dont le visage est le plus similaire à celui qu'on cherche à identifier.

4.4.1 L'approche de reconnaissance faciale basée sur le réseau CNN

Les réseaux de neurones convolutifs sont un type spécial de réseaux de neurones multicouches, ils suivent le chemin de son prédécesseur en neocognitron dans sa forme, sa structure et sa philosophie d'apprentissage [98]. Traditionnellement, les réseaux de neurones convertissent les données d'entrée en un vecteur unidimensionnel [99], ils ont la capacité d'effectuer à la fois d'extraction de caractéristiques et la classification. La couche d'entrée reçoit des images normalisées avec les mêmes tailles (dimensions) [100]. Chaque image d'entrée passera par une série de couches de convolution avec filtres (Kernels) suivit d'une fonction d'activation pour introduire une non linéarité, regroupement (pooling), et une couches entièrement connectées (FC). *La figure 4.5* ci-dessous illustre la configuration du réseau CNN utilisé pour traiter une image d'entrée et classe les objets dans leurs classes d'appartenance [101].

L'architecture de notre CNN, comme la montre le tableau 4.1, est entraînée sur une base de données pour classer les visages, l'âge et les sexes de chaque individu. Dans cette section, nous présentons notre modèle CNN et expliquons les différentes expérimentations et résultats obtenues.

Tableau 4.1 : Architecture détaillée de notre CNN utilisée pour l'extraction des caractéristiques.

Nom	Type de couches	Description de la taille de sortie
Couche d'entrer	Image	3x92x112
Conv1	Convolution + ReLu	16 filtres 3x3x3
S1	Max Pooling	16x2x2 stride 1
Conv2	Convolution + ReLu	32 filtres 16x3x3
S2	Max Pooling	32x2x2 stride 1
Conv3	Convolution + ReLu	64 filtres 32x3x3
S3	Max Pooling	64x2x2 stride 1
Conv4	Convolution + ReLu	128 filtres 64x3x3
S4	Max Pooling	128x2x2 stride 1
FC1	Entièrement connectée + fonction de perte	5376
FC2	Entièrement connectée + fonction de perte	512
FC3	Entièrement connectée + fonction de perte	128

- **L'entrée :** est une image sous forme RGB (couleur) de taille 3x92x112 pixels.
- **Première couche de convolution :** composée de 16 filtres de taille 3x3x3.
- **Couche de sous-échantillonnage:** composée de 16 filtres de taille 16x2x2.
- **Deuxième couche de convolution :** composée de 32 filtres de taille 16x3x3.
- **Couche de sous-échantillonnage :** composée de 32 filtres de taille 32x2x2.
- **Troisième couche de convolution :** composée de 64 filtres de taille 32x3x3.
- **Couche de sous-échantillonnage :** composée de 64 filtres de taille 64x2x2.
- **Quatrième couche de convolution :** composée de 128 filtres de taille 64x3x3.
- **Couche de sous-échantillonnage :** composée de 128 filtres de taille 128x2x2.
- **Première couche entièrement connectée (FC1):** contient 5376 neurones.
- **Deuxième couche entièrement connectée (FC1) :** contient 512 neurones.
- **Troisième couche entièrement connectée (FC1) :** contient 128 neurones.

L'entrée de notre CNN est une image sous forme RGB (couleur) de taille 3x92x112 pixels. Elle passe à travers quatre couches de convolution qui forment le réseau dont chacune est suivie d'une fonction ReLU et des couches de regroupement (pooling) avec conv1 composée de 16 filtres, conv2 composée de 32 filtres, conv3 composée de 64 filtres et conv4 composée de 128 filtres, puis quatre couches de regroupement sont utilisées, couche S1, S2, S3 et S4 respectivement, sont des couches de sous-échantillonnage (subsampling).

Après les couches de prétraitement, arrive trois couches entièrement connectées (FC), chacune d'elles est suivie d'une fonction ReLU et d'une fonction de perte, FC1 avec 5376 neurones, FC2 avec 512 neurones et FC3 avec 128 neurones, ce qui donne respectivement les scores de classe non normalisés pour le genre ou l'âge. Enfin, pour la classification une fonction softmax est utilisée pour donner les probabilités de perte et de classification finale pour l'âge, ainsi que la fonction sigmoïde est utilisée sur la couche FC3 pour donner les probabilités de perte et de classification finale pour la reconnaissance de genre et la reconnaissance faciale.

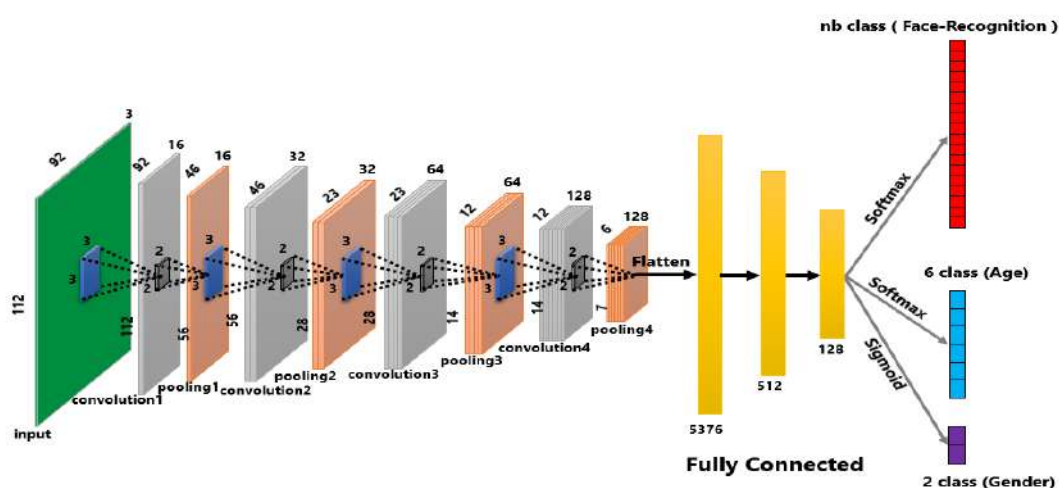


Figure 4.5 : l'architecture de notre réseau CNN.

- **La première expérience** : Dans une première expérience on a voulu montrer l'influence de la fonction de classification utilisée sur le résultat d'identification finale de la base de données Essex pour la tâche de reconnaissance de genre. Dans un premier temps, on a utilisé la fonction de classification softmax pour la reconnaissance de genre (*figure 4.6*), ainsi que la fonction sigmoïde (*figure 4.7*) pour comparer leurs résultats.

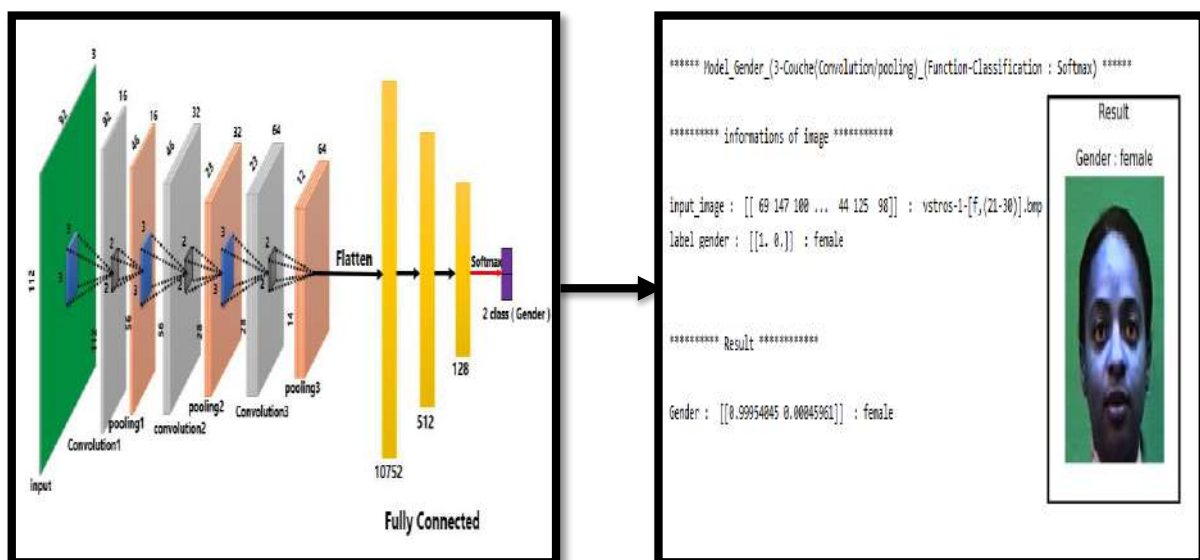


Figure 4.6 : le résultat de la fonction Softmax.

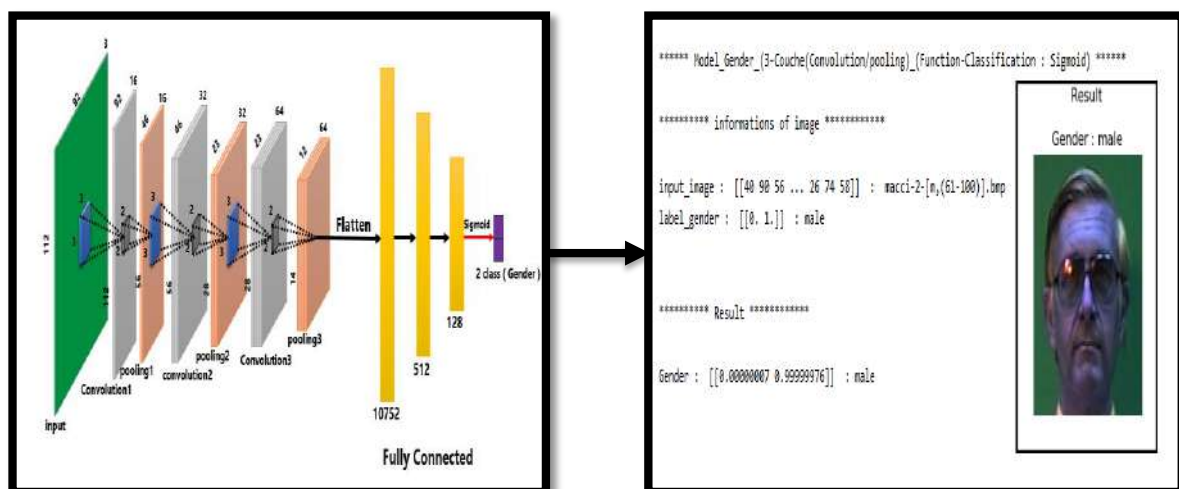


Figure 4.7 : le résultat de la fonction Sigmoïde.

D’après ces résultats obtenus, on remarque que la fonction sigmoïde a permis une précision de [0.00000007, 0.99999976], qui est meilleurs que le résultat de la fonction softmax [0.99954045,0.00045961].

A partir de cette expérience et les résultats obtenus, on conclut que la fonction sigmoïde est meilleure l’orsqu’il s’agit d’une tâche de classification binaire (problème a deux classes), alors

que la fonction softmax est plus efficace pour une tâche de multi classification (plus de 2 classes).

- **La deuxième expérience :** Dans une deuxième expérimentation, on a augmenté le nombre des couches convolution et sous-échantillonnage (pooling) à 4 couches chacune, afin de voir l'effet de nombre de couche sur le taux de reconnaissance.

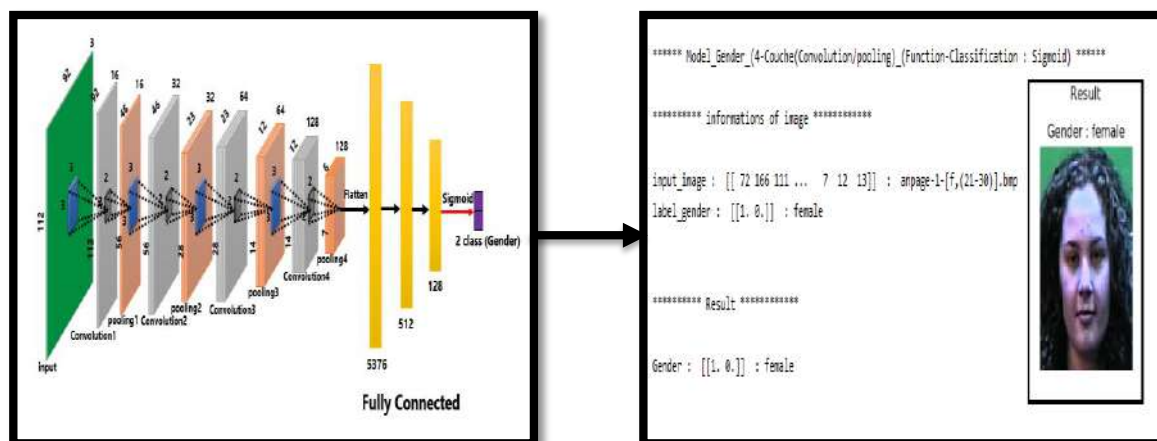


Figure 4.8 : le résultat de classification avec 4 Couche (fonction Sigmoidé).

On remarque que les résultats de classification ce sont nettement améliorer avec l'intégration de cette quatrième couche supplémentaire.

- **La troisième expérience :** lors de cette expérience, nous avons augmenté le nombre d'époque d'apprentissage (itérations) jusqu'à atteindre un taux de reconnaissance le meilleurs possible.

Tableau 4.2 : Taux de reconnaissance par rapport de nombre d'époques

Nombre d'époques	Taux de Reconnaissance
3000	50%
10000	69%
30000	78%
50000	87%
70000	95%
80000	99%

Nous avons aussi voulu illustrer l'influence du nombre d'exemples d'apprentissage sur les performances de la reconnaissance. Dans cette expérimentation, on a augmenté le nombre de visages d'apprentissage pour chaque personne.

Tableau 4.3 : Taux de reconnaissance selon le nombre d'exemple d'apprentissage

Nombre d'exemples d'apprentissage	Taux de Reconnaissance
4	80%
6	88%
8	99%

Après les expériences précédentes, nous avons adopté l'architecture de notre réseau neuronal convolutifs illustrer dans *la figure 4.4*, pour appliquer sur les quatre bases de données (CASIA Webface, FEI, ORL, Essex) pour la tâche de reconnaissance faciale de classification de genre et l'estimation d'âge.

Tableau 4.4 : Taux de reconnaissance et de perte pour les différentes bases de données utilisées.

Base de données	Taux de reconnaissance + perte (Genre)	Taux de reconnaissance + perte (Age)	Taux de reconnaissance + perte (Reconnaissance de visage)	Taux de reconnaissance + perte (Totale)
FEI	0.9995+0.001	/	0.9948+0.016	0.9974+0.016
ORL	0.9995+0.001	/	0.9977+0.008	0.9987+0.008
Casia-WebFace	0.9989+0.002	/	0.9953+0.015	0.9974+0.016
Essex	0.9983+0.005	0.9983+0.0	0.9977+0.010	0.9985+0.005

D'après les résultats obtenus dans les tableaux (4.2-4.3-4.4), nous constatons que la précision de reconnaissance est positivement proportionnelle 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 ainsi que le nombre d'itérations est un facteur très important pour minimiser le taux de perte. *Les figures 4.9 et 4.10* ci-dessous illustre les performances du modèle CNN utilisé sur les différentes bases de données.

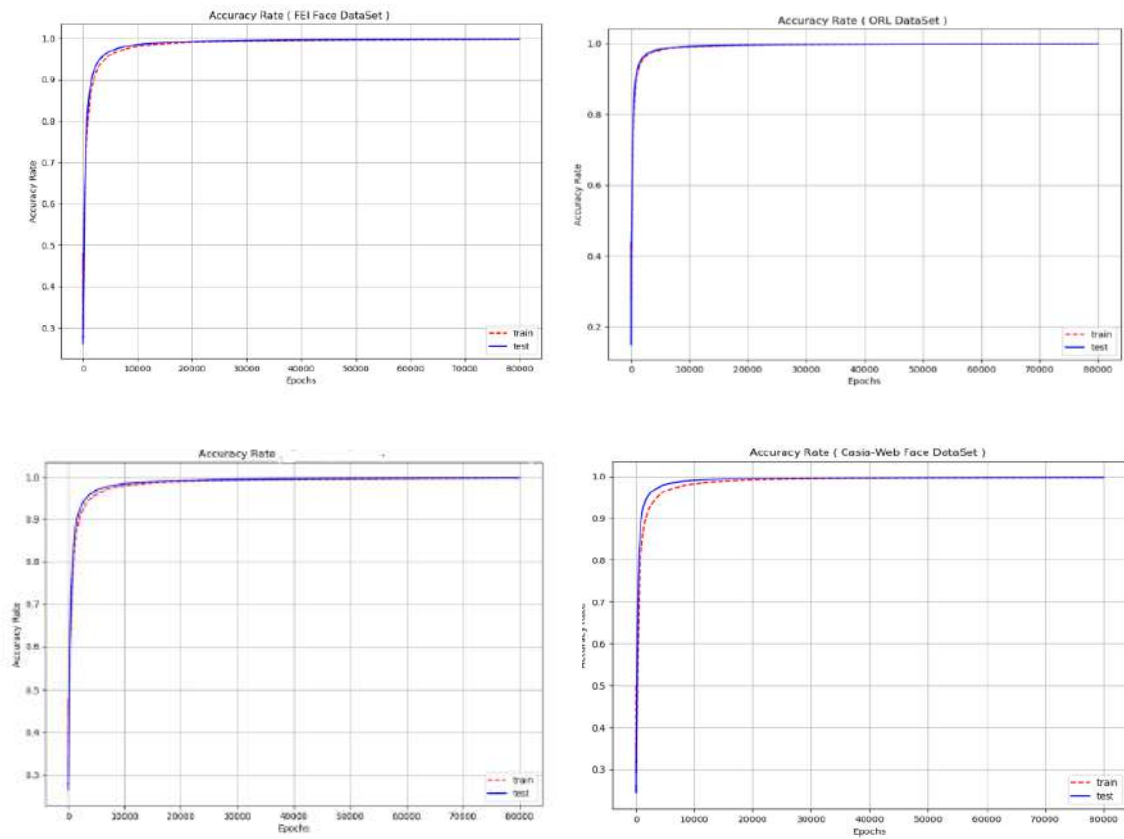


Figure 4. 10 : Taux de reconnaissance obtenue sur les bases de données (CASIA Webface, FEI, ORL, Essex).

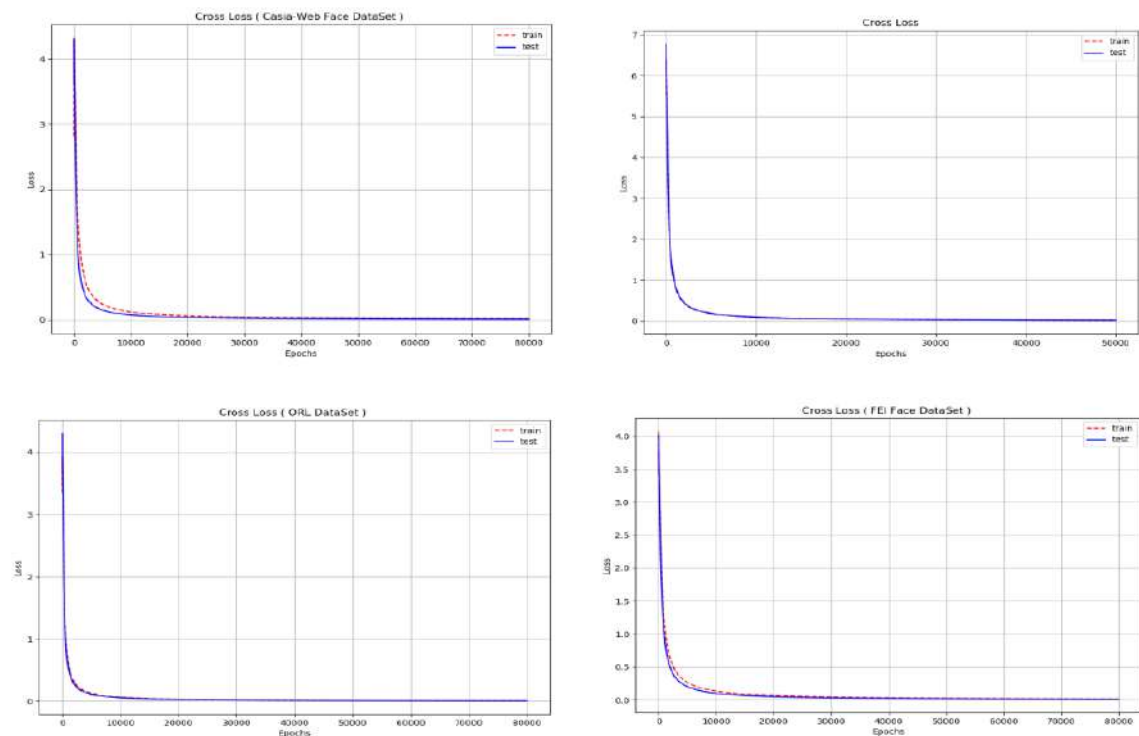


Figure 4. 9 : Taux de perte des bases de données utilisées (CASIA Webface, FEI, ORL, Essex).

4.4.2 L'application des techniques de réduction des dimensions

Les techniques de classification par réduction d'espace restent parmi les techniques les plus utilisées et les plus populaires dans le domaine de la reconnaissance du visage [102]. Ces méthodes considèrent l'image entière du visage comme un vecteur dans un espace multidimensionnel. L'objectif de ces méthodes est de trouver une base réduite de projections qui optimise un certain critère, l'espace original est ensuite projeté sur cette base de dimension inférieure. Dans ce travail on a appliqué trois méthodes très robustes et connue dans ce domaine en l'occurrence de PCA, t-SNE et UMAP respectivement sur les quatre bases de données utilisées dans notre projet (CASIA Webface, FEI, ORL, Essex), pour la visualisation de la distribution des données et leurs regroupements (clustering).

La visualisation des données après classification par le réseau CNN est illustrée sur *la figure 4.11*.

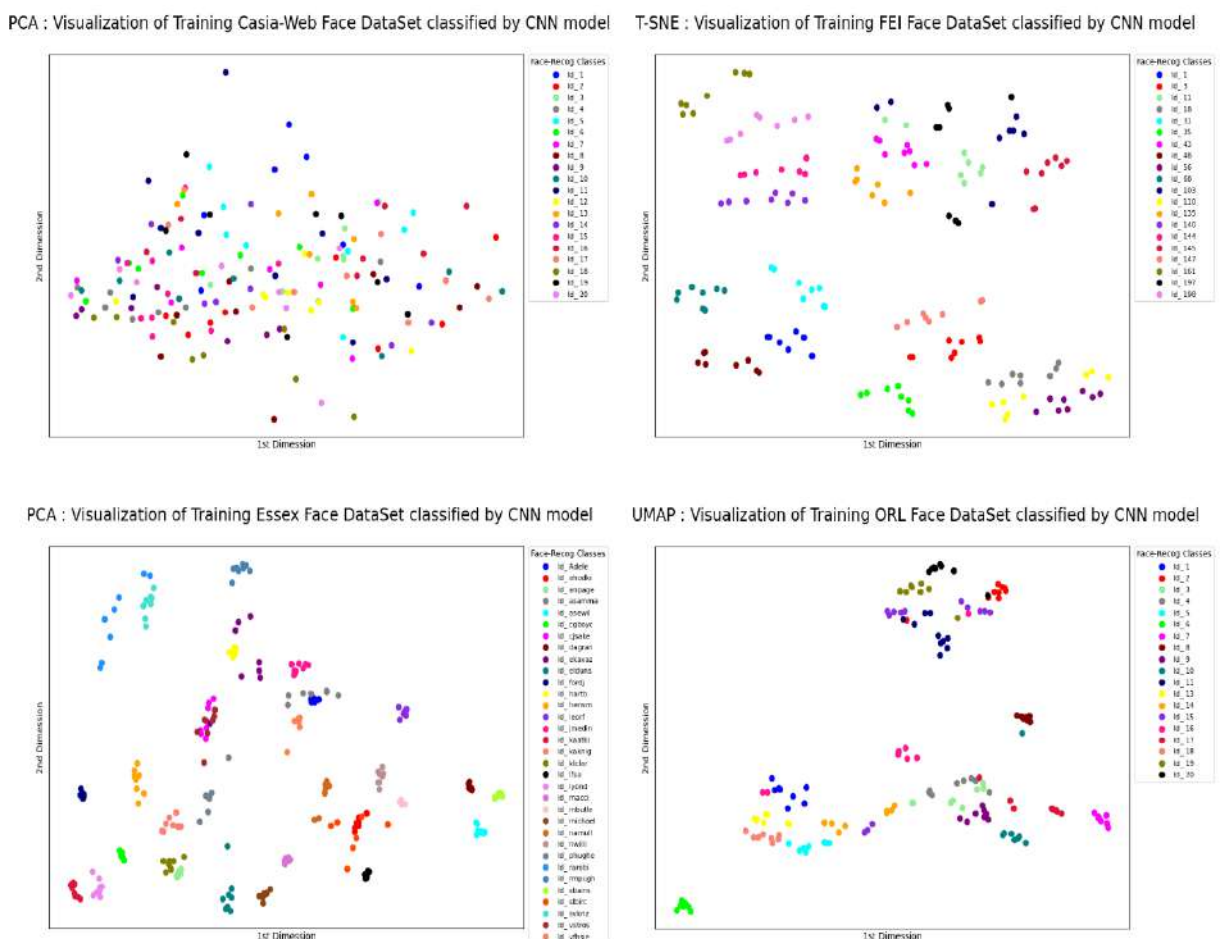


Figure 4. 11 : Les résultats de visualisation des bases de données utilisées (CASIA Webface, FEI, ORL, Essex)

Nous avons ensuite utilisé ces méthodes pour une tâche de regroupement (clustering) sur l'ensemble des bases de données employées dans ce travail.

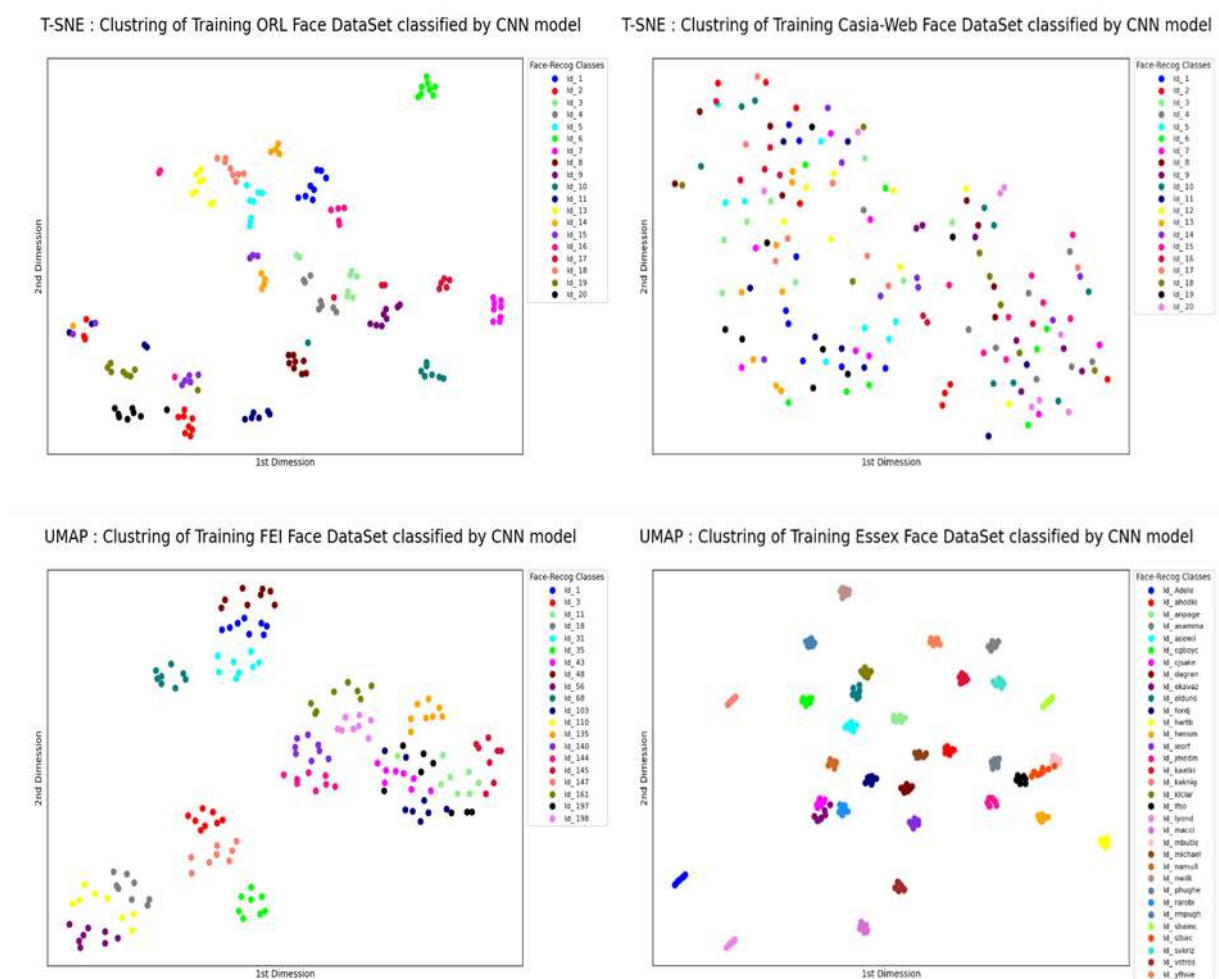


Figure 4. 12 : Les résultats de Regroupement des bases de données utilisées (CASIA Webface, FEI, ORL, Essex)

4.4.3 Discussion

Dans ce travail, nous nous sommes intéressés à développer une application pour la reconnaissance faciale, la classification de genre et l'estimation d'âge testée sur quatre bases de données CASIA WebFace, Essex, ORL et FEI. Au cours de cette étude on s'est basé principalement sur l'utilisation des réseaux de neurones convolutifs (CNN) pour l'extraction des caractéristiques et la classification des traits faciaux.

L'architecture du réseau CNN employé dans ce travail est constitué d'une succession de blocs de traitement, des couches de convolution, des couches de pooling (sous-échantillonnage) et des couches entièrement connectées. Notre réseau neuronal profond (CNN) prend l'image en entrée sans extraction préalable des caractéristiques, ce qui évite grandement la reconstruction massive de données et l'extraction complexe de caractéristiques dans les algorithmes de reconnaissance traditionnels. Le CNN peut effectuer à la fois une extraction et une classification des caractéristiques. La couche d'entrée reçoit des images normalisées de mêmes tailles, ensuite, l'image est convertie sous plusieurs noyaux de filtre à l'aide de poids partagés. Un CNN est composé d'une pile de couches convolutives, qui est paramétrée par le nombre de cartes. On applique un certain nombre de filtres aux données (pixels brutes) d'une image pour extraire les caractéristiques, que le modèle peut ensuite utiliser pour la classification.

Après l'extraction des caractéristiques, le résultat est transmis à trois couches entièrement connectées, qui effectuent la classification des entités extraites. Ces couches sont similaires aux couches de Multilayer Perceptron (MLP). Enfin la sortie de la dernière couche entièrement connectée est classée par la fonction softmax sur un nombre de classes pour la reconnaissance faciale et pour six classes d'âge, en revanche pour la prédiction du genre on a utilisé la fonction sigmoïde. Selon les résultats obtenus le réseau proposé offre une précision de reconnaissance faciale, de classification de genre et l'estimation d'âge très satisfaisante, mais nécessite un temps d'apprentissage très long pour mettre en œuvre une prédiction correcte.

De nombreuses méthodes de réduction de dimensions ont été employées dans ce travail afin de visualiser et regrouper les résultats issus du réseau CNN, y compris celles qui effectuent une transformation linéaire des données par l'analyse en composantes principales (PCA). Cependant, l'ACP est une méthode linéaire et ne fonctionne pas bien dans les cas où les relations ne sont pas linéaires. L'utilisation des méthodes T-SNE qui est considéré comme une méthode focalisée localement et la technique d'apprentissage multiple UMAP ont montré de meilleures performances pour préserver à la fois la structure locale et globale.

Nous avons découvert qu'UMAP fournit les meilleures projections pour tous les modèles évalués, tandis que t-SNE parvient également à regrouper assez bien les attributs. Dans la plupart des cas, certains axes PCA sont capables de fournir des attributs aussi bien représentatifs lorsque les méthodes non linéaires ne sont pas efficaces. Le mappage des données de grande dimension contient généralement de nombreuses observations redondantes de caractéristiques très informatives, ils sont une étape utile dans l'analyse des données, car elle permet de générer des visualisations de données simples telles que des histogrammes ou des nuages de points. L'utilisation de ces méthodes de réduction de dimension permet

considérablement améliorer les performances des algorithmes étudiés, à la fois en termes de précision de clustering et de temps.

4.5 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. On peut définir notre système de reconnaissance et son rôle en deux parties :

➤ Mode vérification

Ce mode consiste à comparer une image à une autre stocké dans la base de données.

- Détecter un visage via la camera de pc.
- Comparer cette image avec ce stocké dans la base de données.

➤ Mode identification

Dans cette phase on compare une image avec toutes les images de la base de données d'apprentissage.

- Importer une image de la base de test.
- Comparer cette image avec ceux stocké dans la base de données

On présente dans cette section les différents aspects de l'application de notre système de reconnaissance.

4.5.1 Interface de Présentation du projet

C'est la première interface dite menu principale, elle est simple et permet aux personnes autorisées d'accéder au reste de l'application :



Figure 4.13 : Interface graphique de notre application (menu principale).

La première interface contient deux boutons :

- **‘Lancer’** : pour démarrer le programme et lancer la caméra qui va vérifier la personne autorisée ou non à accéder à l’application.
- **‘Fermer’** : pour quitter le programme.

Quand vous êtes autorisé à entrer, la deuxième interface vous apparaît.



Figure 4.14: La deuxième interface.

La deuxième interface contient un menu et deux boutons :

- Menu pour choisir l'application voulu 'la reconnaissance' ou bien 'la réduction des dimensions'.
- **‘Lancer’** : pour démarrer l’application.
- **‘Fermer’** : pour quitter le programme.

Pour lancer l’application on clique sur le bouton **‘Lancer’**.

4.5.2 Interface de la reconnaissance faciale

Comme on voit dans *la figure 4.15*, il y a trois classes :

- **'L'objet de la reconnaissance'** : il contient deux choix caméra ou bien image.
- **'Paramètres de la reconnaissance'** : il contient deux menus et trois choix :
 - Premier menu pour choisir la base de données.
 - Deuxième menu pour choisir la méthode de détection (spécifique à la caméra).

- Le troisième menu pour choisir le type de système de reconnaissance souhaitée.
- **'Résultat'** : affiche le résultat des paramètres entré au menu.

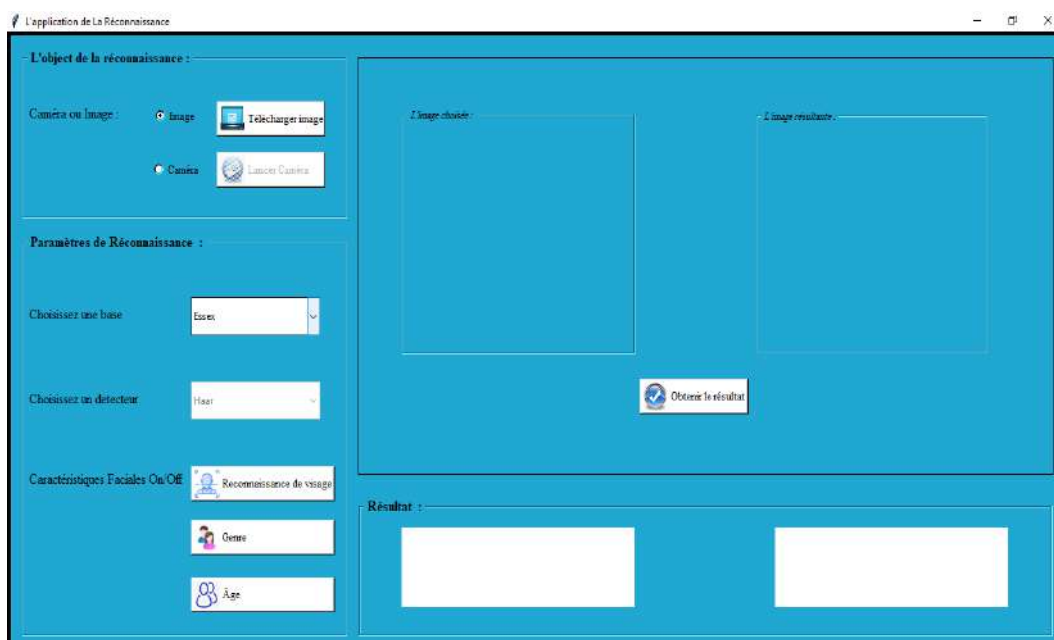


Figure 4.15 : Interface de la reconnaissance faciale.

4.5.3 Interface des méthodes de réduction des dimensions

Comme c'est illustré sur *la figure 4.16*, il se constitue de trois classes

- **'Paramètres des techniques'** : il contient deux choix et deux menus.
 - Les choix pour sélectionner visualisation ou bien regroupement (clustering).
 - Premier menu pour choisir les techniques (PCA , t-SNE ou bien UMAP).
 - Deuxième menu pour choisir le système de reconnaissance voulu (reconnaissance de visage, l'âge ou bien le genre).
- **'Paramètres des bases'** : il contient un menu et deux boutons :
 - Le menu pour choisir la base de données.
 - Les deux boutons sont pour afficher le résultat.
 - Le premier pour afficher le résultat de la base d'entrainement et le deuxième pour afficher le résultat de la base de test.
- **'Résultat'** : affiche le résultat.

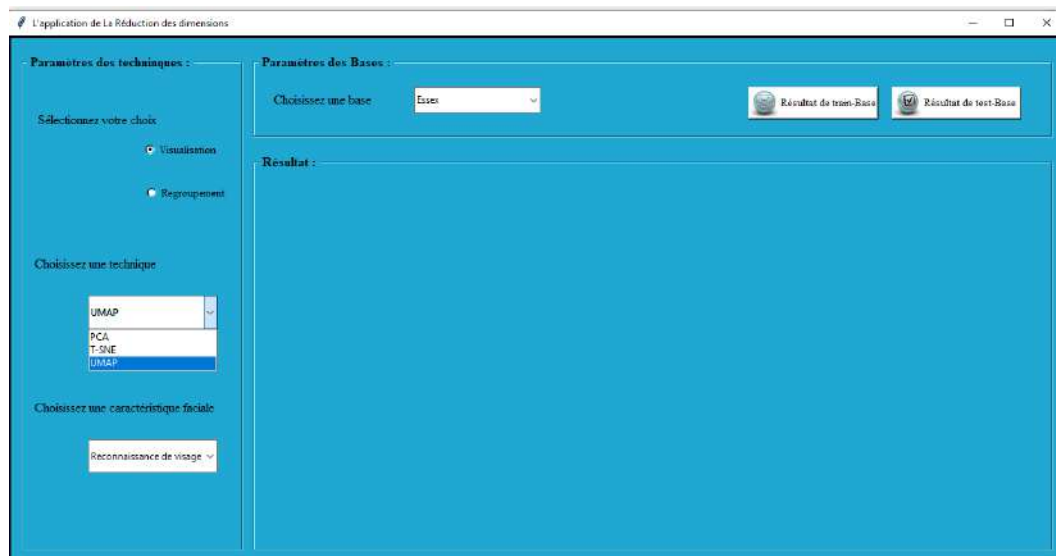


Figure 4.16 : Interface de la Réduction des dimensions.

4.6 Test de l'application sur la base Essex

On choisit la base de données Essex dans le menu et on click sur le bouton 'Télécharger image', on reçoit un message box pour choisir une image de la base Essex puis on clique sur 'OK' pour entrer le chemine d'image de test.

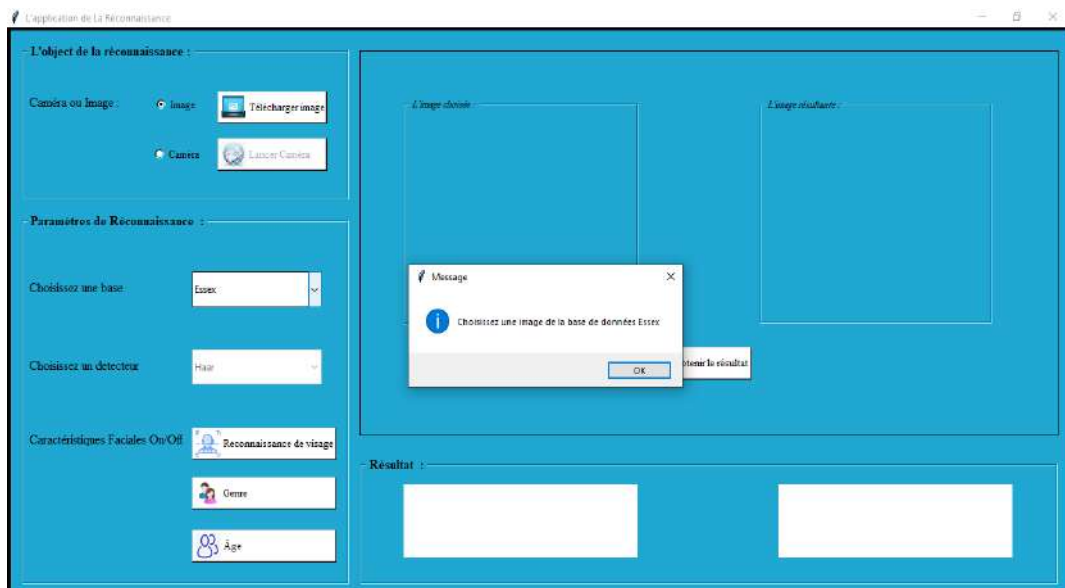


Figure 4.17 : message de choisir une image.

Ensuite on sélectionne une image et on clique le bouton ouvrir

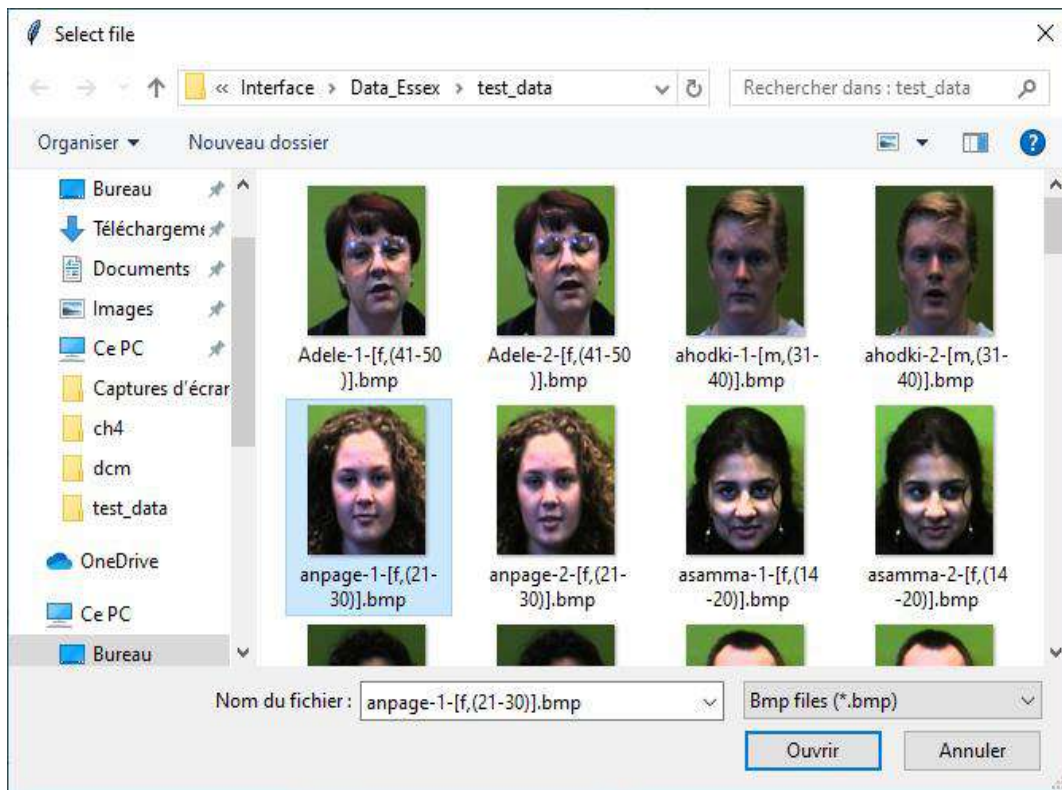


Figure 4.18 : le choix d'image test.

Après on va choisir le systèmes de reconnaissance voulu et on clique sur le bouton 'Obtenir le résultat' fig (4.19-4.20).

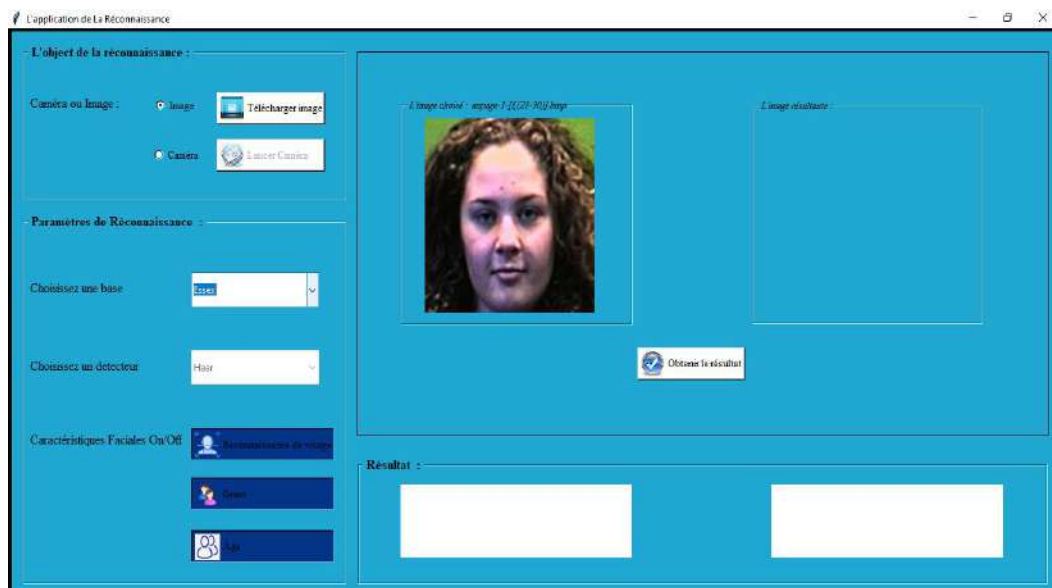


Figure 4.19 : le choix du système de reconnaissance.

Dans ce cas, le programme compare l'image d'entrée de test avec les images d'apprentissage.



Figure 4.20 : Résultat de teste.

4.7 Test des méthodes de réduction des dimensions sur la base FEI

On sélectionne notre choix de visualisation ou bien regroupement, puis on choisit la base de données, la technique de réduction de dimension et le système voulu.

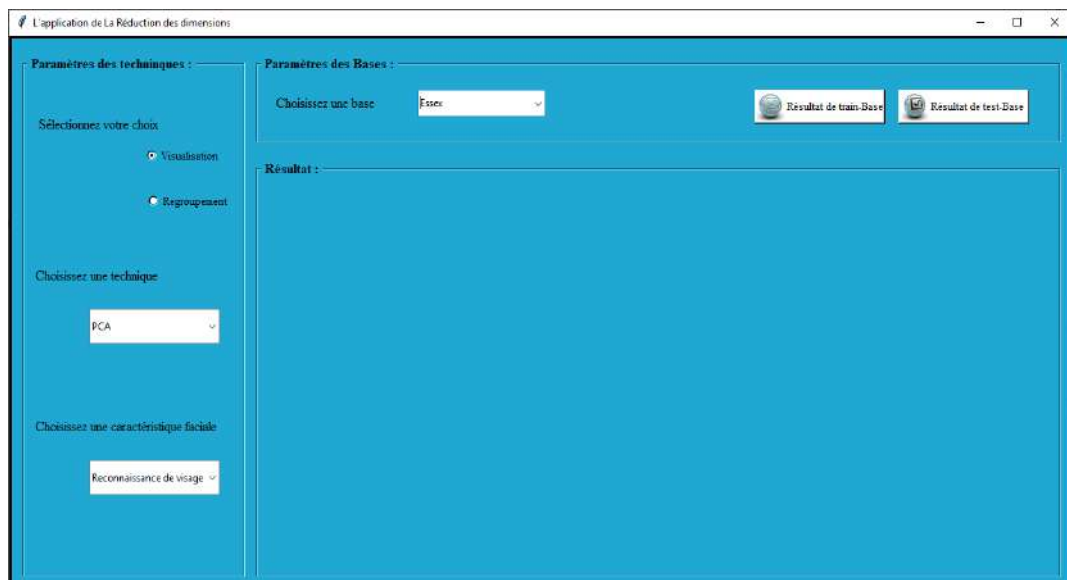


Figure 4.21 : Interface des méthodes de Réduction des dimensions.

Après on clique sur le bouton afficher le résultat

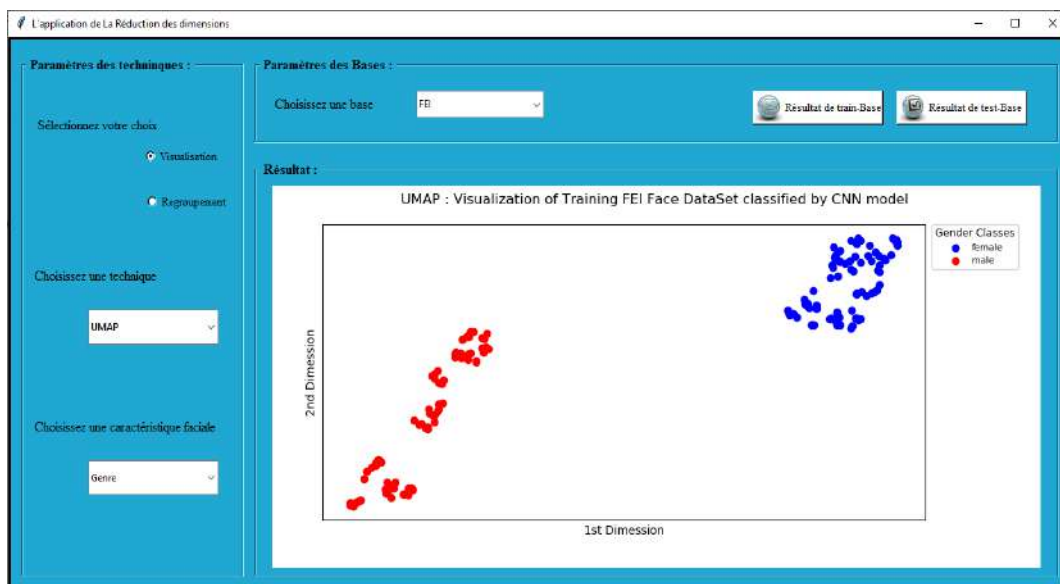


Figure 4.22 : Résultat de teste.

4.8 Conclusion

Dans ce chapitre, nous avons présenté une application de reconnaissance de visage, d'estimation du genre et d'âges basée sur les réseaux de neurones profond de type CNN. Notre système est appliqué et testé sur quatre bases de données de visages ORL, FEI, Casia-WebFace et Essex, on a présenté aussi les différents résultats obtenus pour chaque bases de donnés par le réseau de neurones convolutifs. Après on a appliqué des méthodes de réduction de dimension en l'occurrence de l'ACP, t-SNE et UMAP. Les résultats obtenus montrent que le système proposé pour la prédiction du sexe, d'âge et la reconnaissance faciale donne une plus grande précision que les techniques existantes. D'après les performances de la méthode proposée, nous avons remarqué que le réseau CNN était plus précis avec les classificateurs softmax et sigmoïdes. En général, nous estimons que la performance globale de notre système proposé est très efficace.



CONCLUSION

CONCLUSION GENERALE

Ces dernières années les systèmes de sécurité utilisant des technologies de reconnaissance de forme sont en plein expansion. 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 uniques. La biométrie offre un large éventail de techniques et peut servir dans des domaines très variés, allant de la sécurité d'états et des entreprises, les secteurs de l'identification judiciaire, l'administration et le contrôle d'accès, que ce soit dans les établissements privés ou publics.

Dans ce contexte, la reconnaissance de visages a émergé comme un secteur actif de recherches durant ses dernières années et de nombreuses techniques ont été proposées. La plupart de ces techniques passe par des différentes étapes (étape d'extraction des caractéristiques et autre pour la classification) et parfois nécessite une étape de prétraitement, ce qui rend le système de reconnaissance complexe et augmente le temps l'apprentissage. Pour ces raisons nous avons optés pour l'utilisation des réseaux du neurone convolutifs (CNN) qui combinent les deux étapes dans une seule technique, qui reste l'une des approches les plus fiables pour la reconnaissance de visage.

L'histoire de la conception des réseaux de neurones convolutifs a commencé avec des modèles simples qui empiles des couches de convolutions pour les opérations d'extraction de caractéristiques et de pooling maximal pour le sous-échantillonnage spatial. Ces idées ont été affinées par la suite dans d'autres architectures, sur lesquelles les opérations de convolution étaient répétées plusieurs fois intercaler par des fonctions d'activation et des opérations de pooling, permettant ainsi au réseau d'apprendre des caractéristiques plus riches à toutes les échelles spatiales. Nous nous sommes inspiré de différentes architectures qui ont été proposées ces dernières années, pour proposer notre propre modèle CNN et le tester sur les bases de données d'image faciale Essex, FEI, ORL et Casia Web.

En termes plus techniques, Nous avons développé une application basée sur les réseaux CNN qui utilisent un apprentissage en profondeur (deep learning) pour la reconnaissance faciale dans diverses conditions, telles que la pose, l'éclairage, et mauvais alignement. De plus, une extension de ce travail étaient réaliser pour fonctionner en temps réel dans des conditions

non contrôlées. Nous avons amélioré cette technologie pour l'adopter sur des problématiques l'estimation de l'âge et la classification du genre.

Dans cet esprit, nous avons développé une deuxième partie de notre application, pour la visualisation et le regroupement (clustering) des données (images faciales), basé sur des techniques de réduction des dimensions et mises en œuvre pour la classification de genre et l'estimation de l'âge, ce qui a permis d'obtenir des meilleurs résultats avec un taux de reconnaissance élevé, et d'atteindre d'excellentes performances pour les tâches de classification et de réduction de dimensions. Les expériences effectuées sur les bases de données Essex, FEI, ORL et Casia Web, confirment l'avantage de notre système.

En guise de perspective, dans un premier temps une extension de ce travail peut être envisagée par l'étude et la réalisation de détection et de localisation de visage avec des performances assez hautes. Nous souhaitons aussi tester notre modèle sur d'autres bases des données plus volumineuses présentant de fortes variations d'éclairage et de pose. Dans un travail futur, nous planifions de généraliser les conclusions de ce travail à d'autre challenge de la reconnaissance de visage comme les modèles des images 3D acquises par des caméras de profondeur et l'expression faciale ce qui permettra plus de précision pour indiquer l'état émotionnel des individus.

Référence bibliographies

- [1]. A. Dantcheva, P. Elia, A. Ross, What else does your biometric data reveal? a survey on soft biometrics. *Information Forensics and Security, IEEE Transactions on* 11 (2015): 441-467.
- [2]. L. Best-Rowden, H. Han, C. Otto, B. F. Klare, A. K. Jain, Uncon-strained face recognition: Identifying a person of interest from a media collection. *Information Forensics and Security, IEEE Transactions on* 9 (2014): 2144-2157.
- [3]. J. Orozco, O. Rudovic, J. Gonz_alez, M. Pantic, Hierarchical on-line appearance-based tracking for 3d head pose. eyebrows, lips, eyelids and irises, *Image and Vision Computing* 32 (2014): 14-26.
- [4]. H. Han, C. Otto, A. K. Jain, Age estimation from face images: Human vs. machine performance, in: *Biometrics (ICB). 2013 International Conference on, IEEE: 1-8, 2013.*
- [5]. J. E. Tapia, C. A. Perez, Gender classi_cation based on fusion of di_erent spatial scale features selected by mutual information from histogram of lbp. intensity, and shape, *Information Forensics and Security, IEEE Transactions on* 8 (2013) : 488-499.
- [6]. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hub-bard, L. D. Jackel, Backpropagation applied to handwritten zip code recognition. *Neural computation* 1 (1989) : 541-551.
- [7]. S. Ji, W. Xu, M. Yang, K. Yu, 3D convolutional neural networks for human action recognition, *TPAMI* 35 (2013): 221-231.
- [8]. D. Cirezan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classi_cation. in: *CVPR, IEEE: 3642-3649, 2012.*
- [9]. Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface : Closing the gap to human-level performance in face veri_cation. in: *CVPR, IEEE: 1701-1708, 2014.*
- [10]. S. S. Farfade, M. Saberian, L.-J. Li, Multi-view face detection using deep convolutional neural networks. *arXiv preprint arXiv:1502.02766 (2015).*
- [11]. R. Rothe, R. Timofte, L. Van Gool, Dex: Deep expectation of apparent age from a single image. in: *ICCV Workshops: 10-15, 2015.*
- [12]. X. Liu, S. Li, M. Kan, J. Zhang, S. Wu, W. Liu, H. Han, S. Shan, X. Chen, Agenet: Deeply learned regressor and classi_er for robust apparent age estimation. in: *ICCV Workshops: 16-24, 2015.*

- [13]. Z. Kuang, C. Huang, W. Zhang, Deeply learned rich coding for cross dataset facial age estimation, in: ICCV Workshops, 2015, pp. 96{101.26
- [14]. K. Zhang, L. Tan, Z. Li, Y. Qiao, Gender and smile classification using deep convolutional neural networks, in: CVPR Workshops: 34-38, 2016.
- [15]. G. Levi, T. Hassner, Age and gender classification using convolutional neural networks, in: CVPR Workshops: 34-42, 2015.
- [16]. [En ligne]. Available <http://infoscience.over-blog.com/bioinformatique-56396352.html>. [Accès le 17 03 2020].
- [17]. M. Chassé, La biométrie au Québec : Les enjeux. Commission d'accès à l'information au Québec, Juillet 2002.
- [18]. C.L. Tisse. " Contribution à la vérification biométrique de personnes par reconnaissance de l'iris ". Thèse de doctorat, Université de Montpellier II, 28 octobre 2003.
- [19]. A. Ouamane. " Reconnaissance Biométrique par Fusion Multimodale du Visage 2D et 3D". Thèse de doctorat en sciences en Electronique, Université de Mohamed Khider Biskra, 11 Juin 2015.
- [20]. A. Jain, L. Hong, S. Pankanti ET R. Bolle. "An Identity Authentication System Using Fingerprints ". Proceedings of the IEEE ISSN 0018-9219, volume 85, Numéro 9. 1997.
- [21]. F. Perronnin et J. Dugelay. " Introduction à la Biométrie Authentification des Individus par Traitement Audio-Vidéo ". Revue Traitement du Signal, Volume 19, numéro 4, 2002.
- [22]. S. G. Ababsa, Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D. Thèse de doctorat, Université Evry Val d'Essonne. 03 octobre 2008.
- [23]. BENCHENNANE Ibtissam, Etude et mise au point d'un procédé biométrique multimodale pour la reconnaissance des individus. Thèse de doctorat, Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf
- [24]. Souhila Guerfi. Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D. Traitement du signal et de l'image [eess.SP]. Université d'Evry-Val d'Essonne, 2008. Français. tel-00623243
- [25]. T. Kanade, Picture Processing by Computer Complex and Recognition of Human Faces. these de doctorat université Kyoto , 1973.
- [26]. V. Perlibakas, Face recognition using principal component analysis and log-gabor filters. *arXiv preprint cs/0605025* , 2006
- [27]. S. G. ABABSA, «Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D,» Université d'evry val d'Essonne, France, 2008.

- [28]. GUERROUDJ BENCHERKI BRAHMIN MOUSTAFA, Implémentation d'un système de reconnaissance de visages à base de PCA. Université Djilali Bounaama Khemis Miliana
- [29]. R.Gross, J.shi, j.Cohn, Quo Vadis face Recognition. Third Workshop on Empirical Evaluation Methods in computer Vision, December 2001.
- [30]. LOUIBA Fadia et HADJ ALI Ryma, Système de contrôle d'accès physique basé sur le visage. pp.19 ; 2010.
- [31]. *OUAMANE Hanane*, Identification de reconnaissance faciale avec des expressions. Université Mohamed Khider Biskra Faculté des sciences et de la Technologie 2012
- [32]. Benmansour Radjaa. Benmansour Zineb, Mise au point d'une application de reconnaissance faciale. Université Abou Bakr Belkaid– Tlemcen 2015.
- [33]. Antitza Dantcheva, Carmelo Velardo, Angela D'Angelo , Jean-Luc Dugelay, Bag of Soft Biometrics for Person Identification New trends and challenges. Receiver : 01.08.2010 / Accepté : 11.10.2010.
- [34]. M. *Belahcen*, Authentification et Identification en Biométrie. Université Mohamed Khider Biskra ,2013.
- [35]. M. T. Laskri and D.Chefrou, Who_Is : système d'identification des visages humains, A R I M A ,Volume 1 ,pages 39 à 61,2002.
- [36]. W. Zhao,R. Chellappa, P. J. Phillips and A. Rosenfeld, Face Recognition: A Literature Survey, ACM Computing Surveys, Vol. 35, No. 4, pp.399–458, December 2003.
- [37]. X. xu, Image Based Face Recognition using Global Features, Research Centre for Integrated Microsystems Electrical and Computer Engineering University of Windsor, May13, 2005.
- [38]. Christian Gagné, « [Réduction de la dimensionnalité](#) » [Université de Laval](#)
- [39]. Y. Bennani, S. Guérif, E. Viennet "Réduction des dimensions des données en apprentissage artificiel" Université Paris
- [40]. (en) Pearson, K., « On Lines and Planes of Closest Fit to Systems of Points in Space », dans Philosophical Magazine, vol. 2, no 6, 1901, p. 559–572 [[pdf] texte intégral (<http://stat.smmu.edu.cn/history/pearson1901.pdf>)]
- [41]. Analysis of a Complex of Statistical Variables with Principal Components",1933, Journal of Educational Psychology
- [42]. A.Lemieux Systèmes d'identification de personnes par vision numérique ,Université Laval, Québec, Décembre 2003.
- [43]. D. Blackburn, M. Bone, P. J Phillips. Face recognition vendor test 2000, 2001.

- [44]. A.Mellakh, Reconnaissance des visages en conditions dégradées, l'Institut National des Télécommunications de l'université d'Evry-Val d'Essonne, Avril 2009.
- [45]. H. [Abdi, L.J. Williams](#). Principal component analysis; Wiley Interdisciplinary Reviews: Computational Statistics, pp 433-459, 2010.
- [46]. [en ligne].available : http://en.wikipedia.org/wiki/Linear_discriminant_analysis. [Accès le 16 04 2020].
- [47]. K. Kyungnam, Face Recognition using Principle Component Analysis, Computer Vision and Pattern Recognition, University of Maryland, USA, 1996.
- [48]. M.Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-12(1) :103–108, January 1990.
- [49]. M. Martínez and A. C. Kak. PCA versus LDA. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol (2), pp228–233, 2001.
- [50]. B.Scholkopf, A. Smola, and KR Muller. Nonlinear component analysis as a kernel eigenvalue problem. Technical Report No 44, December 1996.
- [51]. J. Lu, K. N. Plataniotis, and A.N. Venetsanopoulos, Face Recognition Using LDA-Based Algorithms, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 14, NO. 1, 2003.
- [52]. Ravi Ranjan, Data Visualization and Dimensionality Reduction using t-SNE June 20 2019
- [53]. Geoffrey Hinton " Visualizing Data using t-SNE " University of Toronto Department of Computer Science
- [54]. H. Yu and J. Yang. A direct lda algorithm for high-dimensional data - with application to face recognition. Pattern Recognition, 34(10) :2067 2070, 2001.
- [55]. C. Touzet, «Les réseaux de neurones artificiels, introduction au connexionnisme,» cours,exercices et travaux pratiques, Collection de l'EERIE, 1992.
- [56]. Youcef Djeriri " Les Réseaux de Neurones Artificiels " September 2017
- [57]. J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition. PRL, 26(2) :181191, January 2005.
- [58]. S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Mullers. Fisher discriminant analysis with kernels. Neural Networks for Signal Processing IX, Proceedings of the IEEE Signal Processing Society Workshop, pages 41-48, Aug 1999.

- [59]. S. Mika, G. Rätsch, J. Weston, B. Schölkopf, A. Smola, and K.R. Müller. Invariant feature extraction and classification in kernel spaces, 2000.
- [60]. K. Delac, M. Grgic, S. Grgic, "Independent Comparative Study of PCA, ICA, and LDA on the FERET Data Set", Technical Report, University of Zagreb, FER, 2004.
- [61]. Azarmehr, R., Laganier, R., Lee, W-S., Xu, C., Laroche, D. (2015). Real-time Embedded Age and Gender Classification in Unconstrained Video. Conference of IEEE Computer Vision and Pattern Recognition workshop, 56-64.
- [62]. Edwards, G. J.; Taylor, C. J.; Cootes, T. F. (1998). "Interpreting face images using active appearance models". Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. P . 300
- [63]. Cootes, T. F.; Edwards, G. J.; Taylor, C. J. (1998). "Active appearance models". Computer Vision — ECCV'98. Lecture Notes in Computer Science. 1407. p. 484. [CiteSeerX 10.1.1.374.7954](#). [doi:10.1007/BFb0054760](#). ISBN 978-3-540-64613-6.
- [64]. Cootes, T. F.; Edwards, G. J.; Taylor, C. J. (2001). "Active appearance models". IEEE Transactions on Pattern Analysis and Machine Intelligence. 23 (6): 681. [CiteSeerX 10.1.1.128.4967](#). [doi:10.1109/34.927467](#).
- [65]. Bukar, A.M., Ugail, H., Connah, D. (2016). Automatic age and gender classification using supervised appearance model. Journal of Electronic Imaging, 25(6): 061605-061605.
- [66]. [En ligne]. Available : <https://fr.talend.com/resources/ai-vs-machine-learning-vs-deep-learning/>. [Accès le 28 02 2020].
- [67]. [En ligne]. Available : <https://www.centralesupelec.fr/fr/lancement-dun-nouveau-mooc-dedie-au-machine-learning-avec-openclassrooms>. [Accès le 29 02 2020].
- [68]. F. Daoud, F. Louali «La reconnaissance des caractères arabes manuscrits par les réseaux des neurones convolutifs» Mémoire de master, Université Saad Dahlab, Blida-2018
- [69]. F. Nacer «Reconnaissance d'expression faciale à partir d'un visage réel» Mémoire de master, Université de 8/Mai/19450, Guelma- 2019
- [70]. P. Cornioley «Intégration d'un module d'apprentissage profond dans l'architecture logicielle d'un SIG Web» Mémoire de magister, Université Sherbrooke, Mai 2018
- [71]. Dj. Y. Moualek «Deep Learning pour la classification des images» Mémoire de master, Université Abou Bakr Belkaid, Tlemcen-2017
- [72]. [En ligne]. Available : <https://mc.ai/machine-learning-vs-deep-learning-5/>. [Accès le 29 02 2020].

- [73]. [En ligne]. Available : <https://www.i2tutorials.com/top-machine-learning-interview-questions-and-answers/what-are-the-differences-between-supervised-machine-learning-and-unsupervised-machine-learning/>. [Accès le 29 02 2020].
- [74]. [En ligne]. Available : <https://medium.com/datadriveninvestor/exploring-machine-learning-f1dc6f3ec90>. [Accès le 29 02 2020].
- [75]. Ch. Berkani, M. Gatrani « Détection et Reconnaissance de Visage » Mémoire de master, Université Larbi Ben M'hidi, Oum El Bouaghi-2012
- [76]. [En ligne]. Available : https://www.researchgate.net/figure/Structure-dun-neurone-biologique-Graupe-2007_fig22_311680657. [Accès le 15 03 2020].
- [77]. M. T. Chikh «Amélioration des images par un modèle de réseau de neurones» Mémoire de magister, Université Abou Bakr Belkaid, Tlemcen- Septembre 2011
- [78]. [En ligne]. Available : https://www.researchgate.net/figure/Modele-dun-neurone-artificiel_fig30_324929383. [Accès le 15 03 2020].
- [79]. [En ligne]. <http://informatique.coursgratuits.net/methodes-numeriques/reseaux-de-neurones-formels.php>. [Accès le 15 03 2020].
- [80]. [En ligne]. https://www.researchgate.net/figure/Comparaison-entre-neurone-biologique-et-neurone-artificiel_fig6_281015872. [Accès le 15 03 2020].
- [81]. S. Beggari Kh. Khamra «Système de reconnaissance de visage par un réseau de neurone convolutif» Mémoire de master, Université Kasdi Merbah, Ouargla- 2017
- [82]. [En ligne]. https://www.researchgate.net/figure/Exemple-de-reseau-de-neurones-non-boucle-f-est-la-fonction-dactivation-des-neurones_fig9_277064567. [Accès le 15 03 2020].
- [83]. [En ligne]. https://www.researchgate.net/figure/Exemple-de-reseau-de-neurones-boucle-ou-recurrent-f-est-la-fonction-dactivation-des_fig10_277064567. [Accès le 15 03 2020].
- [84]. M. Lemmouchi «Identification des Visages Humains par Réseaux de Neurones» Mémoire de magister, Université De Batna, Batna - 2013
- [85]. [En ligne]. http://eric.univ-lyon2.fr/~ricco/cours/slides/reseaux_neurones_perceptron.pdf. [Accès le 15 03 2020].
- [86]. B. Ghennam, S. Smara. «Les réseaux de neurone convolutif (CNN) pour la classification des images associées aux places de stationnement d'un parc de véhicule» Mémoire de master, Université Djilali Bounaama, Khemis Miliana- 2017
- [87]. [En ligne]. <https://www.sciencedirect.com/topics/computer-science/convolution-filter>. [Accès le 15 03 2020].

- [88]. H. Tolgui. «Deep Learning pour Reconnaissance du Visage» Mémoire de master, Université Mohamed Khider, Biskra-2018
- [89]. [En ligne]. https://www.researchgate.net/figure/Fully-connected-three-layers-feed-forward-network_fig1_228720372. [Accès le 15 03 2020].
- [90]. [En ligne]. <https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5>. [Accès le 15 03 2020].
- [91]. S. Nouas. «Analyse sémantique des publications dans les réseaux sociaux par apprentissage profond» Mémoire de master, Université Saad Dahlab, Blida-2019
- [92]. R. Maxence. «Etude et mise en place du Deep Learning sur des données satellites du CNES» Rapport de Stage,
- [93]. K. Janod. « La représentation des documents par réseaux de neurones pour la compréhension de documents parlés» , Mémoire de Doctorat ,Université d'Avignon, 2017
- [94]. S. Ravi, Ph.D , Sadique Nayeem " A Study on Face Recognition Technique based on Eigenface " Department of Computer Science, School of Engineering and Technology Pondicherry University, Pondicherry, INDIA .
- [95]. [En ligne]. <http://cswww.essex.ac.uk/mv/otherprojects.html>. [Accès le 09 07 2020].
- [96]. Antipova, G., Berrania, S.A., Dugelay, J.L. (2016). Minimalistic CNN-based ensemble model for gender prediction from face images. Pattern Recognition Letters, 70: 59- 65.
- [97]. [En ligne]. [https://en.wikipedia.org/wiki/Python_\(programming_language\)](https://en.wikipedia.org/wiki/Python_(programming_language)) . [Accès le 01 07 2020].
- [98]. Fukushima, K. (1980) Neocognitron: Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. Biological Cybernetics, 36, 193-202.
- [99]. Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE, 86, 2278-2324.
- [100]. Mohammed Kamel Benkaddour, Abdennacer Bounoua , " Feature extraction and classification using deep convolutional neural networks, PCA and SVC for face recognition " International Information and Engineering Technology Association '(IIETA) (2017).
- [101]. [En ligne]. <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>. [Accès le 10 07 2020].
- [102]. X. xu, Image Based Face Recognition using Global Features, Research Centre for Integrated Microsystems Electrical and Computer Engineering University of Windsor, May13, 2005.

Résumé

De nos jours, les systèmes de reconnaissance de formes ne cessent de susciter l'intérêt de la communauté scientifique. Dans ce contexte, la biométrie se présente comme une technologie de base potentiellement puissante et en pleine expansion qui se généralise dans un grand nombre d'applications de vie courantes et recherches scientifiques.

Le travail présenté dans ce mémoire consiste à l'étude et la réalisation d'un système de reconnaissance de visage et de classification du genre et d'âge, basée sur les réseaux de neurone convolutifs (CNN), suivit d'une étape de visualisation et de regroupement (clustering) des données (images faciales) en utilisant des techniques de réduction des dimensions. Les résultats expérimentaux obtenus montrent l'efficacité de notre système proposé avec une bonne performance en termes de taux de reconnaissance et représentation des données.

Mots-clés : Apprentissage profonds, Biométrie, Réseau de neurone convolutif, Reconnaissance de visages, Réduction des dimensions.

Abstract

Nowadays, the systems of recognition of forms constantly arouse the interest of the scientific community. In this context, biometrics presents itself as a potentially powerful and rapidly expanding core technology that is becoming widespread in a large number of current life applications and scientific research.

The work presented in this brief consists in the study and realization of a system of facial recognition and classification of gender and age, based on the networks of the convulsive neuron (CNN), followed by a step of visualization and grouping (clustering) of the data (facial images) using dimensional reduction techniques. The experimental results obtained show the effectiveness of our proposed system with good performance in terms of data recognition and representation.

Keywords: Deep learning, Biometrics, Convolutional neural network, Face recognition, Dimensional reduction.

ملخص

في الوقت الحاضر، تثير أنظمة الاعتراف بالأشكال اهتمام الأوساط العلمية باستمرار. وفي هذا السياق، تقدم المقاييس الحيوية نفسها باعتبارها تكنولوجيا أساسية قوية وسريعة التوسع، والتي أصبحت منتشرة على نطاق واسع في عدد كبير من تطبيقات الحياة الحالية والبحوث العلمية.

يتألف العمل المقدم في هذا الموجز من دراسة وإدراك نظام التعرف على الوجه وتصنيف الجنس والعمر، استناداً إلى شبكات الخلية العصبية الاختلاجة (CNN)، ثم خطوة في التصور والتجميع (تجميع) للبيانات (صور الوجه) استخدام تقنيات تقليل الأبعاد. وتبين النتائج التجريبية التي تم الحصول عليها مدى فعالية نظامنا المقترح مع الأداء الجيد من حيث التعرف على البيانات وتمثيلها.

كلمات مفتاحية: التعلم العميق، المقاييس الحيوية، الشبكة العصبية الالتفافية، التعرف على الوجه، تقليل الأبعاد.