



Université Mohamed Khider de Biskra
Faculté des Sciences et de la Technologie
Département de Génie Electrique

MÉMOIRE DE MASTER

Sciences et Technologies
Filière : Télécommunication
Option : Réseaux et Télécommunication

Réf. : Entrez la référence du document

Présenté et soutenu par :
Sellami Zineddine

Le : dimanche 7 juillet 2019

Optimisation du CNN par l'algorithme Génétique pour la Reconnaissance de Visage

Jury :

Mme BARKAT Aicha	MAA	Université de Biskra	Présidente
Mme BELAHCENE Mébarka	Pr	Université de Biskra	Encadreur
Mme. MEGHERBI mohamed	MCA	Université de Biskra	Examineur

Année universitaire : 2018 – 2019

Dédicace

Je dédie ce travail à mes très chers parents, vous n'avez cessé de me soutenir et de m'encourager avec votre tendresse et affection tout au long de mon parcours.

À mes chères sœurs *Dhikra & Wissall*.

À ma tante *Khadidja*

À ma grande famille.

Et à tous mes amis.

Remerciements

*Mes sentiments les plus respectueux à mon encadreur: Mme BENATIA née BELAHCEN
Mebarka*

*Ses conseils pertinents grâce à ses compétences scientifiques, amabilité
et patience tout au long de ce mémoire, ont permis à mon travail
d'aboutir et de voir le jour.*

*Je remercie également le président et les membres du Jury qui me font
l'honneur d'accepter de juger mon travail*

*Sans oublier tous les enseignants qui ont contribués par leur savoir et
leurs encouragements long de mon parcours*

Liste des figures

Figure 1.1 système de reconnaissance de visage	8
Figure 1.2 Emergence des sous-domaines de l'IA	11
Figure 1.3 Illustration d'un perceptron	12
Figure 1.4 Réseau de neurones (a) Simple, (b) en profondeur	13
Figure 1.5 Les réseaux de neurones convolutifs.....	13
Figure 1.6 : Illustration du FRR et du FAR en fonction de scores	17
Figure 1.7 Vue d'ensemble d'un système de reconnaissance de visages (identification ou vérification) avec les deux phases de fonctionnement et d'enrôlement	17
Figure 1.8 Taux d'égale erreur pour trois SRV S1, S2 et S3	19
Figure 2.1 Principe du système de retrait de lunettes par apprentissage en profondeur	26
Figure 2.2 Images utilisées	26
Figure 2.3 Matrice de confusion pour la détection d'émotions par DeepLearning	27
Figure 2.4 Exemple d'optimisation un SRV 3D par GA	28
Figure 2.5 Exemple d'image segmentée par le CNGA.....	36
Figure 2.6 La taille de l'image Vs Temps de traitement	37
Figure 2.7 Architecture de l'AG.....	39
Figure 2.8 Structure de la méthode hybride GA-CNN-SVM.....	40
Figure 3.1 Paramètres de neurone	44
Figure 3.2: un exemple Calcul d'une étape de la convolution	47
Figure 3.3 Un exemple de max-pooling.	48
Figure 3.4 Schéma illustrant la couche full connected du CNN.....	48
Figure 3.5 Principe général d'un AG.....	52
Figure 3.6 Les cinq niveaux d'organisation d'un algorithme génétique	53
Figure 3.7 Codage binaire	53
Figure 3.8 Codage réel	54
Figure 3.9 Exemple de structure de réseau utilisée	55
Figure 3.10 Schéma de principe du modèle CNGA étudié.....	56
Figure 3.11 Principe du SVM binaire	58
Figure 4.1 Exemple de fenêtre Matlab	64
Figure 4.2 Différentes captures de la base de données CASIA 2D	65
Figure 4.3 Quelques variantes de la base de données CASIA2DV4	65
Figure 4.4 Exemples de la base de données LFW	66
Figure 4.5 Exemple d'images de la BDD Face	67
Figure 4.6 Schéma du système de reconnaissance de visage CNNAG	67

Figure 4.7 Architecture du CNN μ	69
Figure : Résultat des filtres appris par la première couche convolutiveC1.....	69
Figure 4.8 Sélection des paramètres caractéristiques par l'AG.....	70
Figure 4.9 Courbe d'évaluation de la fonction du fitness.....	72
Figure 4.10 Échantillon de 8 images d'hommes et de femmes pour l'apprentissage ..	74
Figure 4.11Affichage d'un échantillon pour l'apprentissage de la BDD (Homme/Femme)	75
Figure 4.12 Système de reconnaissance du genre basé sur le CNN(BDD CASIA2DV4) .	76
Figure 4.13Temps de reconnaissance du genre pour une image test avec le CNN	76
Figure 4.14 Optimisation du système de reconnaissance du genre basé sur CNNAG	77
Figure 4.15Temps de reconnaissance du genre pour une image test avec le CNNAG	77
Figure4.16 Résultat d'affichage d'un échantillon d'image de la BDD Face	78
Figure 4.17 Système de reconnaissance de visage basé sur le CNN(BDD Face)	79
Figure 4.18 Temps pour la reconnaissance basé sur CNN de l'image test (BDD Face)	79
Figure 4.19 Optimisation du système de reconnaissance de visage basé sur CNNAG (BDD Face)	80
Figure 4.20Temps pour la reconnaissance basé sur CNNAG de l'image test (BDD Face).....	80

Listedes tableaux

Tableau 2.1 Avantages et inconvénients du CNN	24
Tableau 2.2Données sur l'exactitude de cnn	24
Tableau 2.3Taux de reconnaissance en fonction de l'augmentation pose sur ORL.....	30
Tableau 2.4 Taux de reconnaissance sur les BDDs ORL, FERET, Yale, Yale B	30
Tableau 3.1Résumé de la terminologie utilisée en AG	50
Tableau 4.1Protocole utilisé dans la reconnaissance du genre (sans et avec optimisation)...	75
Tableau 4.2 Protocole utilisé pour la reconnaissance (sans et avec optimisation)	78

Liste des abréviations

AG : algorithme génétique

CNN : convolution neural network

BDD : base des données

ANN : Artificiel Neural Network

FAR : False Acceptation Rate

ADN : Acide Désoxyribo Nucléique

SVM : système Vecteur Machine

TER : Total Error Rate

TR : Taux de Reconnaissance de Visage

SRV : système de Reconnaissance de Visage

TAR: True Accept Rate

DL: Deep Learning

PCA: Principal Component Analysis

FRR: False Rejection Rate

FR: Faux Rejet

CASIA:institute of automation Chinese Academy of Sciences Iris Image Database

NN :Neural Network

IA : intelligence artificielle

Résumé

La reconnaissance de visage est un domaine de la vision par ordinateur consistant à reconnaître automatiquement une personne à partir d'une image de son visage. C'est un sujet particulièrement étudié en vision par ordinateur. Avec le développement de l'apprentissage en profondeur, la technologie de reconnaissance des visages basée sur CNN (réseau de neurones convolutionnel) est devenue la principale méthode adoptée dans le domaine de la reconnaissance des visages. Les algorithmes génétiques (AG) sont des méthodes heuristiques de recherche ou d'optimisation. Ces techniques ont été inspirées à l'origine du principe darwinien de l'évolution par sélection (génétique). Une AG est basée sur une forme très abstraite de processus évolutifs pour apporter des solutions à des problèmes complexes. Chaque AG opère sur une population de chromosomes artificiels. Chaque chromosome signifie une solution au problème à résoudre et a une forme physique. Une forme physique du chromosome est une mesure numérique réelle qui représente sa performance en tant que solution du problème spécifique. Dans notre travail, l'AG est utilisée pour optimiser le système de reconnaissance de visage (SRV) basée sur le CNN en sélectionnant les meilleures caractéristiques. Les résultats obtenus sont encourageants et prometteurs. L'approche CNINGA proposée, non seulement classe convenablement les images de visage, mais en plus avec un temps de calcul qui a considérablement baissé de 14.986921 secondes pour le SRV utilisant le CNN à 0.847437 secondes.

Abstract

Face recognition is a field of computer vision that automatically recognizes a person from an image of their face. This is a particularly studied subject in computer vision. With the development of deep learning, face recognition technology based on CNN (convolutional neural network) has become the main method adopted in the field of face recognition. Genetic algorithms (GA) are heuristic methods of research or optimization. These techniques were originally inspired by the Darwinian principle of evolution by selection (genetics). An AG is based on a very abstract form of evolutionary processes to provide solutions to complex problems. Each AG operates on a population of artificial chromosomes. Each chromosome means a solution to the problem to be solved and has a physical form. A physical form of the chromosome is a real digital measure that represents its performance as a solution to the specific problem. In our work, the AG is used to optimize the CNN-based face recognition system (SRV) by selecting the best features. The results obtained are encouraging and promising. The CNNGA approach proposed, not only correctly classifies the face images, but in addition with a calculation time that has significantly decreased from 14.986921 seconds for the SRV using the CNN at 0.847437 seconds.

Tableau des matières

Dédicaces

Remerciements

Listes des figures

Listes des tableau

Listes des abréviations

Résumé

Abstract

ملخص

Introduction générale

Chapitre I Généralité sur l'optimisation pour la reconnaissance de visage

1.1 Introduction	6
1.2 La biométrie et les systèmes de reconnaissance de visage.....	6
1.2.1 L'Identification	6
1.2.2 La Vérification (l'authentification) :	7
1.3 Phase de reconnaissance	7
1.4 Principales difficultés de la reconnaissance de visages	8
1.5 Reconnaissance de visage basée sur le Deep Learning	9
1.5.1 L'intelligence artificielle	9
1.5.2 L'apprentissage automatique	9
1.5.3 L'apprentissage profond	10
1.6 principe générale de l'apprentissage profond.....	11
1.7 Principe général de l'algorithme génétique	14
1.8 Performances d'un système de reconnaissances de visage	15
1.8.1 La mesure de la performance du système biométrique.....	16
1.8.2 Scénarios applicatifs et mesures de performance	17

1.8.3 Mesure des taux de performance.....	18
1.9 Conclusion.....	20

Chapitre II Etat de l'Art Deep Learning et l'Algorithme Génétique pour la RV

2.1 Introduction	22
2.2 Etat de l'Art sur la reconnaissance de visage par le Deep learning.....	23
2.2.1 Couche de convolutionnelle (Convolutional Layer).....	23
2.2.2 Couche de sous-échantillonnage (Subsampling Layer)	23
2.2.3 Couche de connexion complète (Full Connected Layer)	24
2.2.4 Couche de sortie (Output Layer).....	24
2.2.5 Application de retrait de lunettes par Deep Learning	25
2.2.6 Deep Learning et détection d'émotions	26
2.3 Etat de l'Art de l'optimisation par l'algorithme génétique	28
2.3.1 Optimisation du système de reconnaissance de visage 3D par AG.....	28
2.3.2 Segmentation d'image par le CNN et le AG	31
2.3.2.1 Codage de l'image dans un format binaire approprié à l'aide du CNN	31
2.3.2.2 Formation des cordes	32
2.3.2.3 Choix de la population initiale	32
2.3.2.4 Évaluation de la fonction de mise en forme dans l'ordinateur CNN	32
2.3.2.5 -Évolution de l'image codée à l'aide d'opérateurs génétiques.....	34
2.3.2.6 -Résultats de la segmentation des images par le CNGA :	36
2.3.3 Reconnaissance de visage par CNN optimisé par AG	38
2.4 Conclusion.....	40

Chapitre III Optimisation de la reconnaissance de visage par le CNNAG

3.1 Introduction	42
3.1 Optimisation du CNN par l'algorithme génétique.....	42
3.1.1 Apprentissage profond basée sur un réseau de neurones convolutionnel	43
3.1.2 Quelques définitions sur les réseaux de neurones et le CNN.....	43
3.1.2.1 Filtre de convolution	44
3.1.2.2 Fonction d'activation.....	45
3.1.2.3 Le Pooling	45

3.1.2.4 Classification (Fully Connected Layer).....	46
3.2 Structure du réseau convolutionnel	46
3.3 Principe de l'algorithme génétique	49
3.3.1 Les caractéristiques des AGs	50
3.3.2 Les fondements des algorithmes génétiques	50
3.3.3 Fonctionnement des algorithmes génétiques	51
3.3.4 Le codage.....	52
3.3.4.1 Codage binaire	52
3.3.4.2 Codage réel	53
3.3.5 Application d'un algorithme génétique au CNN	54
3.4 Approche d'optimisation du CNN par AG pour la reconnaissance de visage	55
3.4.1 Réduction de données	56
3.4.2 Data Mining	57
3.5 Classification par le Support Vector Machine	57
3.5.1 Principe du SVM binaire	58
3.5.2 Principe du SVM multi-classe.....	58
3.5.2.1 Méthodes de décompositions.....	59
3.6 Conclusion.....	61

Chapitre IV Conception du système et résultat

4.1 Introduction	63
4.2 Présentation des outils de développement.....	63
4.2.1 Matériel.....	63
4.2.2 Logiciel.....	63
4.3 Bases données des visages utilisées	64
4.3.1 Base de données CASIA 2DV4	64
4.3.1.1 Présentation de la base de données CASIA 2DV4.....	64
4.3.2.1 Protocole de la BDD CASIA2D	65
4.3.2 Base de données LFW (Labeled Faces in the Wild).....	65
4.3.3 Base de données Face	66
4.4 Système de reconnaissance de visage 3D basé sur CNN et AG	67

4.5 Extraction des caractéristiques et réduction de données par CNN	68
4.6 Utilisation de l'algorithme génétique	69
4.6.1 Sélection des caractéristiques par AG.....	69
4.6.2 Fonction de performance (fitness).....	69
4.7 Classification	73
4.7.1 Classification par SVM.....	73
4.7.2 Classification par mesure de similarité	73
4.8 Etude expérimentale et Résultats	74
4.9 Conclusion	81
<i>Conclusion générale</i>	82

La Bibliographie

Introduction générale

La reconnaissance faciale est une technologie informatique qui détermine l'emplacement et la taille d'un visage humain dans une image numérique, technologie clé du traitement de l'information faciale. Il a été largement appliqué à la reconnaissance des formes, à l'identité, à l'authentification, aux interfaces ordinateur / homme, à la vidéosurveillance automatisée, au commerce électronique, à la santé et aux finances, etc... Il existe deux applications courantes de reconnaissance faciale: l'identification et la vérification du visage. L'identification signifie que les images du visage peuvent être utilisées pour définir l'identité d'une personne. Pendant la vérification des visages, compte tenu de l'image de visage et de l'estimation de l'identité, le système doit pouvoir déclarer que l'estimation est vraie ou fausse.

1. Contexte

La difficulté de reconnaître les visages par ordinateur varie grandement en fonction des conditions de posture. La reconnaissance des visages est l'un des domaines de recherche les plus intéressants et les plus importants de ces dernières années. La reconnaissance des visages a fait l'objet d'une attention considérable de la part des chercheurs en biométrie, en reconnaissance de formes et en systèmes de sécurité. Ces systèmes sont maintenant utilisés dans divers domaines tels que les robots, la reconnaissance militaire, la télédétection, le traitement de documents et l'automatisation industrielle. Le but de la reconnaissance faciale est de construire des systèmes informatiques adaptés aux personnes. Un système de reconnaissance de visage reconnaît un individu en comparant l'image d'entrée avec les images de tous les utilisateurs stockés dans une base de données et en recherchant la meilleure correspondance en appliquant certains algorithmes. La plupart des approches de reconnaissance des visages sont basées sur l'analyse globale de l'image du visage qui retient les informations globales du visage ou la position ou la forme d'attributs faciaux tels que les yeux, les sourcils, le nez, les lèvres et le menton et leurs relations spatiales ou une approche hybride combinant les deux approches ci-dessus. Ces caractéristiques extraites forment un vecteur de caractéristiques qui est l'élément décisionnel le plus responsable dans la formation d'un réseau de neurones artificiels. La technique de reconnaissance des visages utilise plusieurs autres disciplines telles que le traitement d'images, la vision par ordinateur, la reconnaissance de formes, les réseaux de neurones et la psychologie. Compte tenu de la

situation actuelle en matière de sécurité dans le monde, les gouvernements et les entreprises ont besoin de méthodes fiables pour identifier les individus avec précision, sans empiéter de manière excessive sur les droits au respect de la vie privée ni exiger une conformité significative de la part de l'individu reconnu.

2. Motivation

Durant les vingt dernières années, 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.

La reconnaissance faciale est quelque chose que nous faisons tous naturellement depuis la naissance, mais la reconnaissance automatique du visage est un concept relativement nouveau, où ce dernier utilise une image ou une vidéo de la structure physique du visage d'une personne dans un but particulier. Ce concept a débuté en 1960 par Woodrow Wilson Bledsoe, qui a mis au point le premier système de reconnaissance faciale semi-automatique qui oblige l'administrateur à identifier manuellement les caractéristiques telles que les yeux, les oreilles et le nez sur les photographies. Dans les années 1970, Goldstein, Harmon, and Lesk ont utilisé 21 auto-marqueurs spécifiques tels que la couleur des cheveux et l'épaisseur des lèvres pour identifier le visage, mais les mesures sont encore calculées manuellement. En 1988, la technologie de reconnaissance faciale a été développée par Kirby and Sirovich, indiquant moins de 100 valeurs nécessaires pour faire pivoter et normaliser l'image faciale. Enfin, Turk and Pentland ont découvert la reconnaissance faciale en 1991, en rendant celle-ci réelle et efficace. Depuis lors, les études et la recherche se sont poursuivies jusqu'à ce que la technologie ait atteint sa forme actuelle.

En raison de l'intérêt accru pour les questions de sécurité mondiale, on s'intéresse de plus en plus aux systèmes informatiques pour identifier les visages et augmenter le nombre de systèmes de sécurité dans ce domaine.

3. problématique

Dans les dernières décennies, des informaticiens ont mis au point un certain nombre de produits commerciaux qui ont permis d'améliorer les performances des algorithmes de reconnaissance automatique du visage. Pour un certain nombre de domaines nécessitant des fonctionnalités de reconnaissance du visage, plusieurs questions ont été soulevées pour savoir dans quelle mesure l'algorithme est capable d'identifier un visage approprié et quelles sont les limites des techniques ?

Il existe plusieurs algorithmes de reconnaissance de visage basés sur des techniques tels que l'analyse en composantes principales (ACP), l'analyse multilinéaire en composantes principales (MPCA) et l'analyse linéaire discriminante (ADL) ... Seulement, tous les systèmes de reconnaissance de visage (SRV) conventionnels restent supervisés et nécessitent un ajustement manuel des paramètres mise en jeu et restent à améliorer concernant les performances et la robustesse. Notons que les SRV classiques sont gourmands en temps de calcul et présentent une complexité de calcul.

L'intelligence artificielle ou IA pour faire court est le domaine de faire penser à l'ordinateur comme les humains en créant un cerveau artificiel. Tout ce que l'humain peut faire de façon intelligente doit être déplacé dans des machines. On peut se demander comment l'ordinateur exécute certaines tâches qui peuvent sembler difficiles à première vue, et d'autres qui peuvent sembler impossibles. Il est reconnaissable de reconnaître les visages ainsi que de nombreuses autres applications.

L'apprentissage profond est un sous-domaine de l'apprentissage machine. Conceptuellement, le fonctionnement de l'apprentissage profond se traduit par plusieurs couches de calques auxquels des données sont échangées pour générer un ou plusieurs rendu(s). L'image de la personne est déconstruite en plusieurs parties afin de créer plusieurs "calques" différents les uns-des-autres qui reconstruite donnent de nouveaux visages.

Il s'agit simplement d'une boîte noire qu'on entraîne avec de grandes quantités de données afin que l'ordinateur comprenne de lui-même (de manière abstraite) l'élément sur lequel on l'entraîne.

4. Objectif

Lorsque nous pensons à une personne en particulier, nous imaginons très souvent son visage, son sourire et différentes caractéristiques faciales uniques. Nous savons reconnaître la personne par son visage, comprendre ses émotions, estimer son âge, avec presque 100% de

certitude quant à son sexe. Le système de vision humaine peut très facilement faire cela avec beaucoup d'autres choses. Peut-on faire la même chose avec les algorithmes d'intelligence artificielle modernes, en particulier les réseaux de neurones profonds? Dans ce travail, nous aimerons décrire la plupart des tâches courantes de l'analyse faciale, en fournissant également des références à des ensembles de données déjà existants et à des solutions basées sur l'apprentissage en profondeur.

Dans notre travail, un algorithme génétique (GA) qui fonctionne sur les chromosomes générés aléatoirement est proposé pour optimiser du descripteur obtenu par l'apprentissage automatique.

5. Structure du mémoire

Le mémoire se présente sous quatre chapitres :

Dans le premier chapitre se trouve des généralités sur le système de reconnaissance de visage et les techniques de Deep Learning ainsi qu'un aperçu sur l'optimisation du SRV par l'Algorithme Génétique.

Le deuxième chapitre est un état de l'art sur les travaux récents dans le domaine de la reconnaissance de visage associée au Deep Learning et son optimisation par l'algorithme génétique.

Le troisième chapitre est consacré à la conception du modèle d'optimisation de la reconnaissance de visage par le Deep Learning étudié.

Un quatrième chapitre est dédié à l'implémentation du modèle retenu et sa validation sur des bases de données universelles.

Enfin une conclusion générale est présentée et quelques perspectives pour les pistes à investiguer sont proposées.



Chapitre

1.1 Introduction

La reconnaissance de visage a été abordée par plusieurs chercheurs et chaque année des avancements dans ce domaine ont vu le jour. L'utilisation de plusieurs notions de base a abouti à des meilleurs résultats pour l'identification de visage. Au cours des vingt dernières années, 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 et aussi à ses caractéristiques avantageuses. Plusieurs méthodes de reconnaissance faciale ont été proposées sur deux axes principaux: la reconnaissance d'images statiques et l'identification de séquences d'images (vidéo). Sans aucun doute, les investissements dans la reconnaissance faciale conduisent à une diversité dans les domaines d'application (haute sécurité, télésurveillance, contrôle d'accès ...)

En plus, cette technique est très efficace pour les situations non standards, c'est le 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 être ainsi si on utilise des approches plus efficaces en plus du bon choix des caractéristiques d'identification représentant le visage en question.

Dans un système de reconnaissance de visages, une image suit -depuis son entrée- un processus bien précis pour arriver à déterminer l'identité du porteur de visage.

1.2 La biométrie et les systèmes de reconnaissance de visage

Il existe deux modes en biométrie :

1.2.1 L'Identification :qui 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 ? ».

Au cours de cette phase, on distingue deux types d'identification :

Chapitre I Généralité sur l'optimisation pour la reconnaissance de visage

✓ **Identification en ensemble fermé**

Par exemple on utilise ce type d'identification afin d'enregistrer la présence de personnes dans certaine entreprise. Si l'échantillon possède un certain degré de similitude avec les échantillons dans le system, la personne sera acceptée.

✓ **Identification en ensemble ouvert**

S'il y a une grande similitude entre l'échantillon biométrique testé et tous les modèles préenregistrés et si cette similitude est inférieure (ou supérieure) au seuil de sécurité, cette personne est rejetée. Cela signifie que la personne ne fait pas partie de celles enregistrées par le système.

1.2.2 La Vérification (l'authentification) :

Qui est une comparaison "1 :1", 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[1]

1.3 Phase de reconnaissance

Dans un système de reconnaissance de visage classique une image peut se trouver dans 3 catégories différentes : i) non-visage ; b) visage inconnu ; c) visage connu. Le principe général est de déterminer si l'image test correspond bien à un visage et ensuite à quelle personne, dans le sous-espace de données, elle se rapproche le plus. Dans ce type de système de reconnaissance de visage (SRV), on peut voir les étapes suivantes :

1. Détection de l'image de l'individu dans la scène
2. La segmentation puis l'extraction de caractéristiques de l'image
3. Détermination si le visage est un visage connu de la base ou inconnu par le classifieur
4. Décision de la personne qui correspond le mieux.

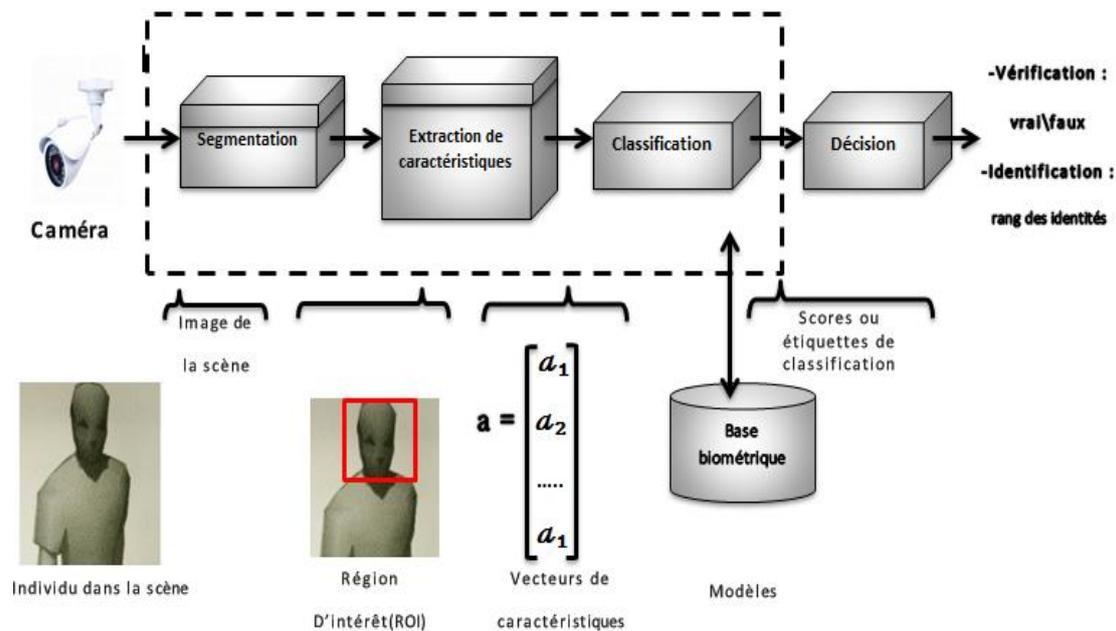


Figure 1.1 système de reconnaissance de visage

1.4 Principales difficultés de la reconnaissance de visages

Pour le cerveau humain, le processus de la reconnaissance de visages est une tâche visuelle de haut niveau. Bien que les êtres humains puissent détecter et identifier des visages dans une scène sans beaucoup de peine, construire un système automatique qui accomplit de telles tâches représente un sérieux défi. Ce défi est d'autant plus grand lorsque les conditions d'acquisition des images sont très variables. Il existe deux types de variations associées aux images de visages : inter et intra sujet. La variation inter-sujets est limitée à cause de la ressemblance physique entre les individus. Par contre la variation intra-sujets est plus vaste. Elle peut être attribuée à plusieurs facteurs [3]:

- **changement d'illumination,**
- **variation de poses,**
- **expressions faciales,**
- **présence ou absence des composants structurels,**
- **les vrais jumeaux....**

1.5 Reconnaissance de visage basée sur le Deep Learning

Avant de présenter la reconnaissance de visage utilisant le Deep Learning (DL), nous passons en revue l'essentiel des différentes définitions dans le domaine de l'intelligence artificielle.

1.5.1 L'intelligence artificielle

L'intelligence artificielle (IA) est un domaine de recherche visant à reproduire par des systèmes artificiels les diverses capacités cognitives de l'être humain (et plus largement les comportements "intelligents" du vivant), en particulier en ce qui concerne leur capacité à résoudre des problèmes complexes. Sa définition est assez floue; elle est aussi large que la définition que vous pouvez avoir de l'intelligence en général. Dans la pratique cela peut recouvrir actuellement des choses aussi variées qu'identifier des images d'objets dans des vidéos, traduire un texte en russe vers le mandarin, permettre à un robot de naviguer de façon autonome dans son environnement, ou poser un diagnostic médical à partir de diverses informations sur un patient. Il y a toutes sortes de problèmes d'IA, des grands, des petits, des ouverts, d'autres résolus, beaucoup de mal posés, bref c'est un champ fertile et très ouvert sur le reste du monde scientifique et technologique.

1.5.2 L'apprentissage automatique

L'apprentissage automatique (ou apprentissage machine, *machine learning* en anglais) est un sous-domaine de l'IA qui s'intéresse en particulier aux capacités d'apprentissage. Le principe est de reproduire un comportement non pas en le programmant "à la main" dans un ordinateur, mais en concevant un système plus général capable d'apprendre à partir d'exemples à résoudre le problème.

Pour reprendre l'exemple du système capable de reconnaître des images dans des vidéos, on pourrait imaginer écrire un programme dans lequel on intègre notre connaissance de ce qu'est l'objet (exemple un chat (il a quatre pattes, des moustaches, une queue, il est généralement noir, blanc ou roux, etc.)), on utilise des systèmes de traitement d'image pour reconnaître ces divers éléments, on indique ce qui différencie un chat d'autres choses (un chien par exemple, etc.) et on essaye ainsi de créer un système à la main avec nos connaissances. (On parle parfois pour ce type de méthode d'approche ingénieuse", car on a considéré le problème comme un problème d'ingénierie classique, que l'on résout en le découpant en sous-

Chapitre I Généralité sur l'optimisation pour la reconnaissance de visage

problèmes bien définis que l'on sait résoudre avec notre expertise technique.) Ou bien on peut au contraire créer un système qui n'a aucune idée au départ de ce que c'est qu'un chat, mais qui peut être "entraîné", en lui montrant un grand nombre d'images en lui disant lesquelles sont des chats et n'en sont pas, de façon à au final pouvoir reconnaître des chats. C'est une approche basée sur l'apprentissage automatique (ici un apprentissage supervisé).

L'apprentissage automatique ne représente qu'un domaine et une approche particulière au sein de l'IA; il est incontournable et est l'un des sous-domaines les plus prometteurs actuellement en ce qui concerne les applications industrielles notamment. Ses avantages sont souvent la performance (c'est typiquement le cas sur les tâches de traitement d'image par exemple) et une certaine généralité (on peut reprendre le même système et lui apprendre à reconnaître autre chose, par exemple des chiens, en lui donnant un jeu de données d'apprentissage différent). Ses inconvénients sont la nécessité d'un jeu de données qui peut parfois devoir être très grand, et souvent des gros besoins techniques

1.5.3 L'apprentissage profond

L'apprentissage profond (*Deep Learning*) est un terme abrégé pour "apprentissage dans les réseaux de neurones profonds". Il s'agit des méthodes d'apprentissage automatique utilisant les réseaux de neurones profonds; c'est donc un sous-domaine de l'apprentissage automatique (et un sous-sous-domaine de l'IA en général).(figure 1.2)

Les sont un outil utilisé en apprentissage automatique populaire depuis un moment (années 80). On qualifie généralement un réseau de "profond" dès lors qu'il a plus d'une couche de neurones cachée (d'après la définition Wikipédia). Longtemps, il n'était pas possible d'entraîner facilement ces réseaux "profonds" pour résoudre des problèmes intéressants, en raisons de problèmes inhérents à leur structure, aux algorithmes d'apprentissage utilisés et aux ressources disponibles (computationnelles et jeux de données). Depuis le début des années 2010 environ, des astuces techniques et algorithmiques découvertes ainsi que le progrès considérable du matériel disponible (calcul sur GPU en particulier) ont permis de surmonter ces problèmes, et d'utiliser ces réseaux profonds pour améliorer énormément les performances dans de nombreux domaines de l'IA. Ils sont encore en développement très rapide et connaissent déjà de nombreuses applications industrielles (dans les startups comme dans les grands groupes).

Pour autant, cela reste une technique parmi d'autres et elle a ses soucis: en particulier les jeux de données d'apprentissage doivent souvent être gigantesques; l'entraînement peut demander

Chapitre I Généralité sur l'optimisation pour la reconnaissance de visage

de très grosses ressources computationnelles (parfois des semaines voire des mois sur de gros systèmes distribués) et beaucoup de temps ce qui rend les techniques peu adaptées à l'apprentissage en ligne en temps réel (pour un robot apprenant à interagir avec son environnement, par exemple); la définition de la structure des réseaux et le réglage des hyperparamètres (les réglages qui contrôlent comment le réseau apprend) relèvent toujours davantage de l'artisanat que de la science, et les systèmes obtenus sont assez opaques (on constate que le système a appris à résoudre une tâche, mais il est difficile de l'analyser pour comprendre *comment*). Il n'en reste pas moins que ce type de système permet des applications très impressionnantes.

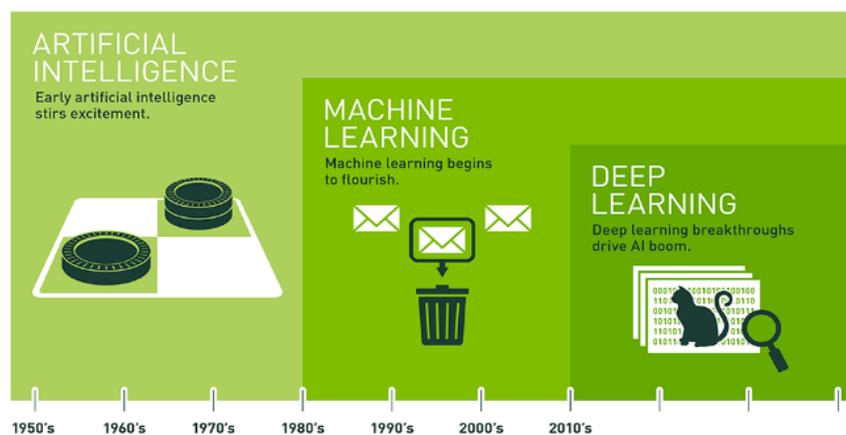


Figure 1.2 Emergence des sous-domaines de l'IA[4]

1.6 Principe général de l'apprentissage profond

L'apprentissage profond (Deep Learning (DL)) est un ensemble de techniques d'apprentissage automatique qui a permis des avancées importantes en intelligence artificielle dans les dernières années.

Dans l'apprentissage automatique, un programme analyse un ensemble de données afin de tirer des règles qui permettront de tirer des conclusions sur de nouvelles données. L'apprentissage profond est basé sur ce qui a été appelé, par analogie, des « réseaux de neurones artificiels », composés de milliers d'unités (les « neurones ») (figure 1.4) qui effectuent chacune de petites opérations simples. Les résultats d'une première couche de « neurones » servent d'entrée aux calculs d'une deuxième couche et ainsi de suite. Par exemple, pour la reconnaissance visuelle, des premières couches d'unités identifient des lignes, des courbes, des angles... des couches supérieures identifient des formes, des combinaisons de

Chapitre I Généralité sur l'optimisation pour la reconnaissance de visage

formes, des objets, des contextes... Les progrès de l'apprentissage profond ont été possibles notamment

grâce à l'augmentation de la puissance des ordinateurs et au développement de grandes bases de données (« big data »). L'apprentissage profond peut par exemple permettre à un ordinateur de mieux reconnaître des objets hautement déformables et/ou analyser par exemple les émotions révélées par un visage photographié ou filmé, ou analyser les mouvements et position des doigts d'une main, ce qui peut être utile pour traduire le langage des signes, améliorer le positionnement automatique d'une caméra, etc... Elles sont utilisées pour certaines formes d'aide au diagnostic médical (ex. : reconnaissance automatique d'un cancer en imagerie médicale), ou de prospective ou de prédiction (ex. : prédiction des propriétés d'un sol filmé par un robot).

Il existe un grand nombre de variantes d'architectures profondes. La plupart d'entre eux sont dérivés de certaines architectures originales.

Les réseaux de neurones à convolution sont à ce jour les modèles les plus performants pour classer des images. Désignés par l'acronyme CNN, de l'anglais convolutional Neural Network, ils comportent deux parties bien distinctes.

Les réseaux de neurones sont parmi les méthodes de Machine Learning les plus populaires. Leur unité de base c'est le **neurone**(figure 1.3), une fonction mathématique qui va analyser les données entrantes et produire un signal plus ou moins important selon que le neurone est sensible ou pas aux informations qu'ils traitent.

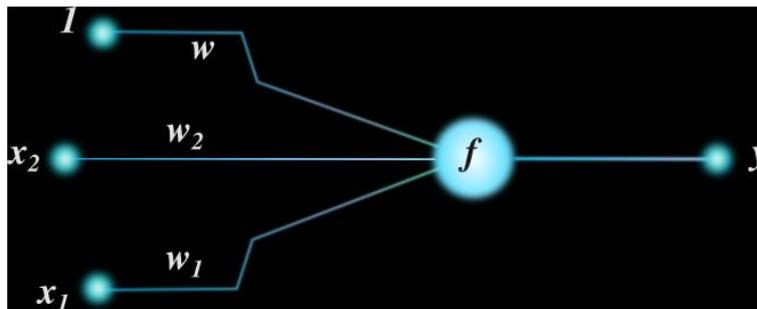


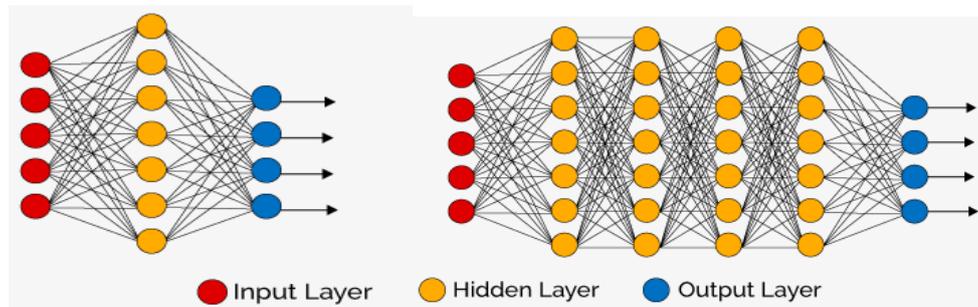
Figure 1.3 Illustration d'un neurone [5]

Un neurone artificiel ou formel (perceptron est défini par : $f(w+w_1x_1+w_2x_2) = y$

Ce sont les paramètres de la fonction qui permettent cette activation. Durant l'étape d'apprentissage le réseau analyse des données pour lesquelles la solution recherchée est connue et trouve pour chaque neurone les paramètres qui lui permettront d'être pertinent.

Généralement les neurones sont organisés en couches qui transforment petit à petit les données jusqu'à en extraire toute l'information nécessaire.

Deux couches sont particulières : la couche d'entrée qui reçoit les données brutes et la couche de sortie qui produit le résultat sous la forme souhaitée. Lorsque, entre ces deux couches, le réseau de neurones possède plusieurs couches intermédiaires, nous rentrons dans le domaine du **DeepLearning**.



(a) Simple neural Network

(b) Deep Learning neural Network

Figure 1.4 Réseau de neurones (a) Simple, (b) en profondeur [5]

Les réseaux de neurones involutifs sont à ce jour les modèles les plus performants pour classer des images. Désignés par l'acronyme CNN, de l'anglais convolutionel Neural Network, ils comportent deux parties bien distinctes.

En entrée, une image est fournie sous la forme d'une matrice de pixels. Elle a deux dimensions pour une image aux 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 d'une succession de filtres, ou noyaux de convolution, créant de nouvelles images appelées cartes de convolutions ci-dessous **figure 1.5** Certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum (ou moyenne) local. Enfin, les cartes de convolutions sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN.

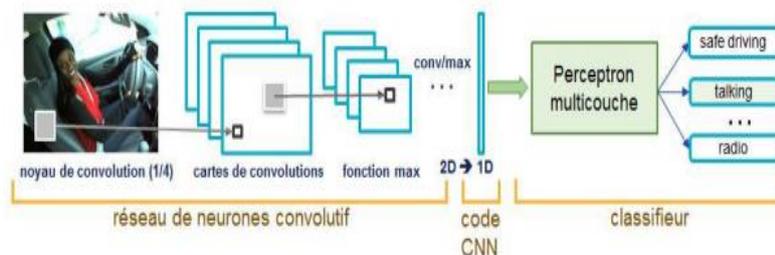


Figure1.5 Les réseaux de neurones convolutifs.[6]

Chapitre I Généralité sur l'optimisation pour la reconnaissance de visage

Ce code CNN en sortie de la partie convolutive est ensuite branché en entrée d'une deuxième partie, constituée de couches entièrement connectées (perceptron multicouches). 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.[6]

1.7 Principe général de l'algorithme génétique

Les algorithmes génétiques (GA), initié dans les années 1970 par John Holland ; sont des algorithmes d'optimisation basés sur les dérivées des techniques de la génétique et les mécanismes de l'évolution de la nature. Les travaux récents dans le domaine de la reconnaissance faciale utilisent des algorithmes évolutifs pour la sélection des fonctionnalités. Les algorithmes génétiques sont des méthodes de recherche stochastiques inspirées de la théorie Darwinienne de l'évolution naturelle, à savoir le principe de "survie du plus apte" et des mécanismes de la génétique. Certaines des principales caractéristiques attrayantes de cette approche sont la simplicité des opérations et la possibilité d'utiliser parallèlement. Dans l'espace de recherche, les chromosomes caractérisent des individus représentant des solutions de segmentation d'image. Basé sur la sélection naturelle et l'évolution, l'algorithme maintient une population d'individus considérés génétiquement supérieurs en évaluant une fonction de fitness appropriée.

Les GA sont des algorithmes évolutifs qui peuvent être utilisés pour résoudre ce genre de problème. Pour utiliser les GA, nous commençons avec une population initiale choisie arbitrairement chromosomes et évaluons l'aptitude relative de chaque chromosome. Un AG est un algorithme itératif de recherche optimum qui manipule une population de taille constante. La taille constante de la population provoque un phénomène de concurrence entre les chromosomes. Chaque chromosome représente le codage d'une solution potentielle au problème à résoudre, et se compose d'un ensemble d'éléments appelés gènes, qui peuvent prendre plusieurs valeurs appartenant à un alphabet qui n'est pas nécessairement numérique.

Comme nous l'avons déjà cité, les algorithmes génétiques sont des méthodes de recherche stochastiques inspirées de la théorie Darwinienne de l'évolution naturelle, en particulier du

Comme nous l'avons déjà cité, les algorithmes génétiques sont des méthodes de recherche stochastiques inspirées de la théorie Darwinienne de l'évolution naturelle, en particulier du

Chapitre I Généralité sur l'optimisation pour la reconnaissance de visage

principe de la "survie du plus apte" et des mécanismes de la génétique. Certaines des principales caractéristiques attrayantes de cette approche sont la simplicité des opérations et la possibilité de l'utiliser parallèlement. Dans l'espace de recherche, les chromosomes caractérisent des individus représentant des solutions de segmentation d'image. Basé sur la sélection naturelle et l'évolution, l'algorithme contient une population d'individus considérés génétiquement supérieurs en évaluant une condition physique appropriée. Ces chromosomes sont transmis aux sources, ce qui a pour résultat d'améliorer le caractère génétique des individus dans les générations à venir. Sur de longues périodes, de nouvelles espèces mieux adaptées à des conditions écologiques particulières vont émerger. La solution finale se présente sous la forme d'un chromosome de l'individu le mieux adapté qui peut être décodé en une image segmentée.

À chaque itération, appelée génération, une nouvelle population avec le même nombre de chromosomes est créée. Cette génération se compose de chromosomes mieux «adaptés» à leur environnement tel que représenté par la fonction sélective. Comme les générations progressent, les chromosomes tendent vers l'optimum de la fonction sélective. La création d'une nouvelle population de la précédente se fait par l'application des opérateurs génétiques qui sont: sélection, croisement et mutation. [7]

Bien que les AG aient connu un succès pratique considérable, les tentatives visant à établir un compte-rendu théorique de leur fonctionnement précis se sont révélées plus difficiles. Il y a deux objectifs pour une théorie satisfaisante des AG. La première consiste à expliquer à quelles classes de AG les problèmes sont particulièrement adaptés et pourquoi. La seconde consiste à fournir des techniques et des approches pour une conception et une mise en œuvre optimales des AG, car de nombreux choix de structure et de paramètres doivent être faits.

1.8 Performances d'un système de reconnaissances de visage

Les performances d'un système de reconnaissance dépendent de plusieurs facteurs qui interviennent à plusieurs niveaux et qui peuvent limiter le degré de précision. Cependant, il serait judicieux de s'intéresser à ces facteurs avant de mesurer la performance d'un système de reconnaissance. Nous citons ici les principaux facteurs :

- L'environnement au moment de l'acquisition.
- Les différentes positions des capteurs.
- La qualité des capteurs.

- La mauvaise interaction entre l'utilisateur et les capteurs.

1.8.1 La mesure de la performance du système biométrique

Tout d'abord, afin de comprendre comment déterminer la performance d'un système biométrique, il nous faut définir clairement trois critères principaux.

1. -Le premier critère s'appelle le taux de faux rejet ("False Reject Rate" ou FRR). Ce taux représente le pourcentage de personnes censées être reconnues mais qui sont rejetées par le système.

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

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

2. Le deuxième critère est le taux de fausse acceptation ("False Accepter Rate" ou FAR).Ce taux représente le pourcentage de personnes censées ne pas être reconnues mais qui sont tout de même acceptées par le système.[8]

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

Tel que FA correspond au cas où le système accepte un individu qui a proclamé une identité qui n'est pas la sienne.

3. Le troisième critère est connu sous le nom de taux d'égale erreur ("EqualErrorRate" ou EER). Ce taux est calculé à partir des deux premiers critères et constitue un point de mesure de performance courant. Ce point

correspond à l'endroit où $FRR = FAR$, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations

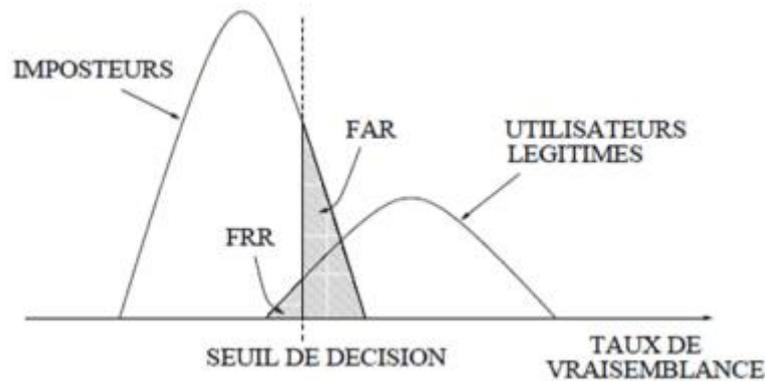


Figure 1.6 : Illustration du FRR et du FAR en fonction de scores [9]

1.8.2 Scénarios applicatifs et mesures de performance

Dans un système biométrique, on distingue généralement deux principaux scénarios ou deux formes de fonctionnement : l'identification ou la vérification (ou l'authentification). **La figure 1.7** montre une vue d'ensemble d'un système de reconnaissance de visages avec ses deux formes.

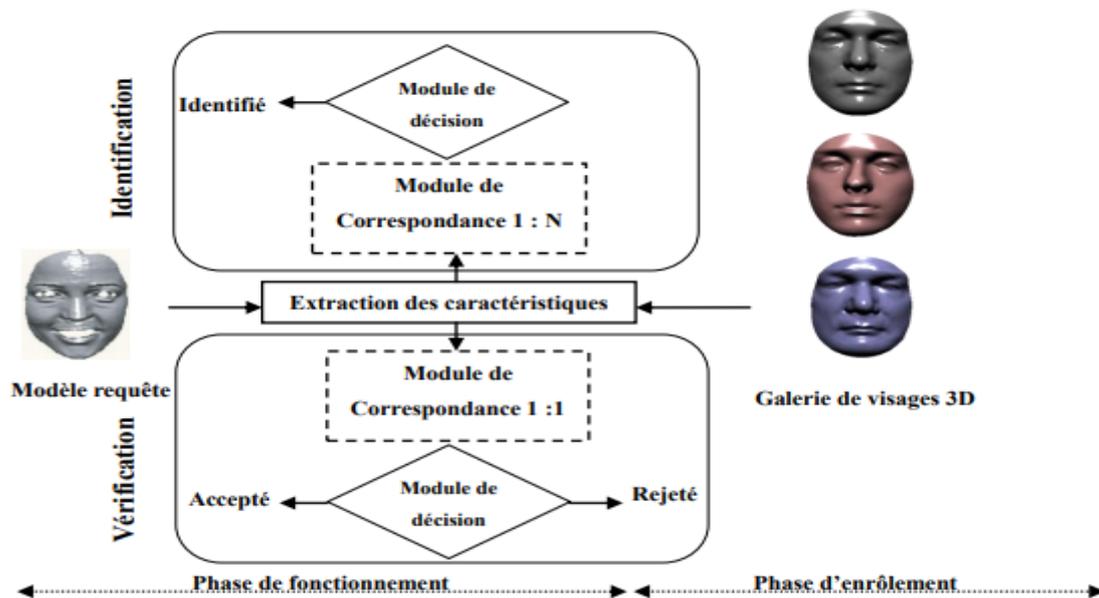


Figure 1.7 Vue d'ensemble d'un système de reconnaissance de visages (identification ou vérification) avec les deux phases de fonctionnement et d'enrôlement .[10]

Il s'agit essentiellement de représenter un utilisateur dans le système tout en capturant et enregistrant ses informations biométriques dans une base de données. Pour la phase de fonctionnement, elle est répartie suivant trois modules: un premier module pour l'extraction des caractéristiques, un deuxième module pour la mise en correspondance et un module final pour la prise de décision.[10]

Chapitre I Généralité sur l'optimisation pour la reconnaissance de visage

Lorsque le système biométrique opère suivant une application d'authentification ou de vérification, l'utilisateur affirme son identité et le système vérifie si cette affirmation est valide ou non. On est donc dans le cas de la mise en correspondance (1 : 1) dans lequel une mesure de similitude est effectuée entre l'entrée biométrique de l'utilisateur et le modèle enregistré dans la base qui est réclamé. Ainsi, si le degré de similitude, entre l'identité affirmée et le modèle enregistré est élevé, l'affirmation de la validation de l'identité est rejetée et l'utilisateur est considéré comme étant un imposteur. Dans le cas d'une application d'identification, l'utilisateur ne dévoile pas explicitement son identité. Cependant, l'affirmation implicite faite par l'utilisateur est qu'il s'agit d'une personne déjà représentée dans le système. Ainsi, L'échantillon biométrique de l'individu est comparé avec les modèles de toutes les personnes de la base de données.

On parle alors de la mise en correspondance (1 : N). La sortie du système biométrique est constituée par l'identité de la personne dont le modèle possède le degré de la similitude le plus élevé avec l'échantillon et tous les modèles est inférieure à un seuil de sécurité minimum fixé, la personne est rejetée, ce qui implique que l'utilisateur n'était pas une des personnes représentées dans le système. Dans le contraire, la personne est acceptée.

1.8.3 Mesure des taux de performance

Pour les deux scénarios applicatifs présentés dans la section 1.6.1, la performance des systèmes biométrique est traditionnellement évaluée selon quatre critères de performances suivants. Le taux d'identification représente le premier critère. Il définit le pourcentage de visage correctement identifiés parmi les visages présentés au système.

$$\text{Taux d'identification} = \frac{\text{Nombre de visage correctement identifiés}}{\text{Nombre total de visage}} * 100 \quad (1.1)$$

Le deuxième critère s'appelle le taux de faux rejet (False RejectRate ou FRR ou TFR). Ce taux représente le pourcentage de personnes censées être reconnues mais qui sont rejetées par le système. Le troisième critère est le taux de fausse acceptation

(False Accept Rate ou FAR ou TFA). Ce taux représente le pourcentage de personnes censées ne pas être reconnues mais qui sont tout de même acceptées par le système. Le critère final est connu sous le nom de taux d'égal erreur (EqualErrorRate ou EER ou TEE). Ce taux est calculé

Chapitre I Généralité sur l'optimisation pour la reconnaissance de visage

à partir des deux critères (FRR, FAR) et constitue un point de mesure de performance courant. Ce point correspond à l'endroit où

$$\mathbf{FRR = FAR} \quad (1.2)$$

Ce taux représente le meilleur compromis entre les faux rejets et les fausses acceptations. La **figure 1.8** permet d'illustrer les taux d'égal erreur pour trois systèmes de reconnaissance de visages S1, S2 et S3.

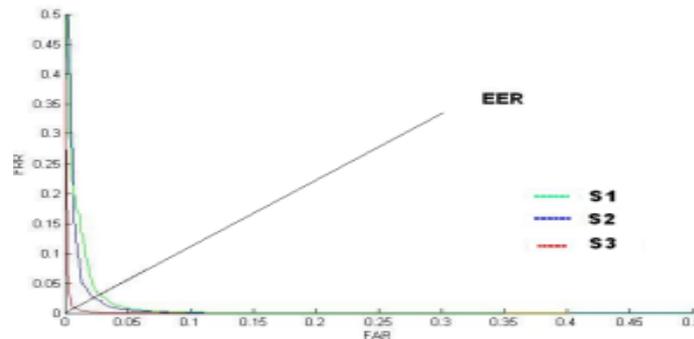


Figure 1.8 Taux d'égal erreur pour trois SRV S1, S2 et S3[2]

Dans un scénario de vérification, on présente la performance d'un système biométrique sous forme d'une courbe **ROC Receiver Operating Characteristic**. Cette courbe trace les deux premiers critères à savoir le taux de faux rejet en fonction du taux de fausse acceptation. Dans de nombreux travaux de l'état de l'art ; le taux de vérification est représenté par le taux de bonne acceptation (True Accept Rate ou TAR) avec $TAR = 1 - FRR$ ou la valeur du taux de fausse acceptation (FAR) est fixée à 0.1. Ce taux est un point spécifique de la courbe ROC. En revanche, dans un scénario d'identification, nous utilisons ce que l'on appelle une courbe **CMC (Cumulative Match Characteristic)**. Cette courbe donne le taux d'identification en fonction d'une variable que l'on appelle le rang R. On dit qu'un système identifie au rang R lorsqu'il choisit l'image la plus proche, parmi les R premières images, comme résultat de la reconnaissance. Nous pouvons donc dire que plus le rang augmente, plus le taux de reconnaissance correspondant est lié à un niveau de sécurité faible [2].

1.9 Conclusion

Dans ce chapitre, nous avons présenté les différentes étapes d'un processus de reconnaissance et les deux technologies utilisées dans l'optimisation du système de reconnaissance du visage. 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. Nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance automatique de visages. Enfin, nous avons aussi donné un aperçu sur les techniques et la mesure de leurs performances.



Chapitre II

2.1 Introduction

Actuellement, le système de reconnaissance des visages se développe de manière durable sur une plus grande échelle. Il y a quelques années, la reconnaissance faciale était utilisée comme une identification personnelle d'une portée limitée. Cette technologie s'est développée dans le domaine de la sécurité, en termes de prévention des fraudeurs, des criminels et des terroristes. En outre, la reconnaissance des visages est également utilisée pour détecter l'état de fatigue d'un conducteur, réduire le nombre d'accidents de la route, ainsi que dans les domaines du marketing, de la publicité, de la santé et autres. Le terme Deep Learning (en français : apprentissage profond) est très en vogue ces derniers temps. C'est bien simple, lorsque l'on parle d'intelligence artificielle, on parle presque systématiquement de Deep Learning. A tel point que dans l'esprit de beaucoup, ces deux termes sont synonymes. C'est pourtant inexact, le Deep Learning (DL) est une sorte d'apprentissage machine et de formation permettant de construire un modèle éduqué à partir d'une grande quantité de données. Ce type d'algorithme (DL) est conçu pour apprendre les caractéristiques d'apprentissage d'entités sans avoir à les spécifier à l'avance. En outre, il s'agit de l'un des meilleurs algorithmes permettant à la machine d'apprendre différents niveaux de propriétés de données (images). À mon avis, l'apprentissage en profondeur se distingue par la création de nouvelles caractéristiques pouvant être apprises à différents niveaux, ce qui poussera les chercheurs de l'avenir à se concentrer sur cet aspect très important. Les caractéristiques sont le premier facteur du succès de tout algorithme d'apprentissage automatique intelligent. La capacité à extraire et / ou à sélectionner correctement les propriétés et à représenter et préparer les données pour l'apprentissage constitue le point de division entre le succès et l'échec de l'algorithme[11].

De nombreuses méthodes sont développées pour donner la meilleure précision dans la reconnaissance faciale. Les méthodes d'apprentissage en profondeur deviennent une tendance dans ce domaine en raison de résultats époustouflants et d'un calcul rapide. Cependant, le problème de l'exactitude, de la complexité et de l'évolutivité devient un défi pour la reconnaissance des visages. Notre travail se concentre sur la reconnaissance de l'importance de cette technologie et sur la manière d'atteindre une grande précision avec une complexité réduite. Pour cela, nous essayons d'approcher la résolution de ces problèmes par l'optimisation du softmax du CNN par l'algorithme génétique.

2.2 Etat de l'Art sur la reconnaissance de visage par le Deep learning

La recherche a également indiqué que l'apprentissage en profondeur peut résoudre des problèmes complexes en utilisant des architectures multicouches. Le processus de résolution de problèmes est donc plus court et les résultats plus précis. Multilayer est une implémentation du processus de sous-échantillonnage trouvé dans une architecture d'apprentissage en profondeur. Cela rend l'apprentissage en profondeur très efficace pour résoudre des problèmes complexes. L'apprentissage en profondeur est très efficace pour prédire des données connues ou inconnues. L'apprentissage en profondeur fonctionne bien sur un large éventail de grands ensembles de données. L'apprentissage en profondeur a été largement appliqué dans les scènes visuelles, la reconnaissance de la parole, la reconnaissance des visages, la reconnaissance des empreintes digitales, la reconnaissance de l'iris, etc. .

La reconnaissance des visages peut être développée avec une approche d'apprentissage en profondeur. L'un d'eux utilise le réseau de neurones à convolution ou est couramment abrégé en CNN. L'approche CNN est une méthode d'un réseau de neurones capable d'identifier les caractéristiques de la nature de l'image en entrée. Cette méthode permet d'extraire l'image brute en une image classée. Le réseau de neurones à convolution (CNN) est une variante de la distorsion et de la transformation géométrique. La couche servant à extraire l'image est appelée couche convolutionnelle (neuron convolution). Le réseau (CNN) possède des architectures à quatre couches pour vérifier les degrés de décalage, d'échelle et de distorsion. Les quatre couches sont :

2.2.1 Couche de convolutionnelle (Convolutional Layer)

Cette couche est la couche principale qui sous-tend le processus CNN. Le processus de convolution est le processus d'application répétée d'une fonction à la sortie d'une autre fonction. Dans les techniques de traitement d'image, le but de la convolution est d'extraire des caractéristiques de l'image insérée. Le résultat du processus d'extraction est la transformation linéaire des données.

2.2.2 Couche de sous-échantillonnage (Subsampling Layer)

Le sous-échantillonnage est le processus de réduction de la taille d'une donnée d'image. CNN, la méthode de sous-échantillonnage largement utilisée est le pooling maximum. Cette technique de pooling maximum fonctionne en divisant la sortie de la couche de convolution en plusieurs petites grilles. Ensuite, la valeur la plus élevée de chaque grille est organisée dans une matrice.[12]

2.2.3 Couche de connexion complète(Full Connected Layer)

La couche de connexion complète est une couche qui sert à effectuer la transformation sur la dimension de données afin que les données puissent être classées de manière linéaire. Chaque neurone de la couche de convolution doit d'abord être transformé pour que les informations contenues ne soient pas perdues.

2.2.4 Couche de sortie (Output Layer)

La couche en sortie est la dernière couche résultant du processus CNN (Convolution Neural Network). Vous trouverez ci-dessous un organigramme du réseau de neurones à convolution. Le réseau de neurones à convolution (CNN) présente également plusieurs avantages et inconvénients.

Tableau 2.1 Avantages et inconvénients du CNN[12]

Avantages	Inconvénient
<i>1 . Peut être implémenté dans diverses résolutions d'image.</i>	<i>1.Ne convient pas pour être simple</i>
<i>2. L'informatique est tellement détaillée que le taux d'erreur est probablement faible.</i>	<i>2. Traiter assez longtemps</i>
<i>3. Le réseau de neurones de convolution (CNN) est capable de résoudre des problèmes complexes ayant de nombreux paramètres à calculer.</i>	<i>3. L'informatique est très complexe, directement proportionnelle à la complexité des problèmes rencontrés.</i>
<i>4. Peut classer la forme du visage de données connues et inconnues.</i>	<i>4. Je ne peux pas décrire le visage avec une certaine position.</i>

Le niveau de précision du réseau de neurones à convolution (CNN) est assez élevé. Certains chercheurs mesurent le taux de reconnaissance dans ses recherches sur la reconnaissance des visages en utilisant l'approche du réseau de neurones à convolution (CNN).

Tableau 2.2 Données sur l'exactitude de cnn[12]

No	Auteurs	Taux Reconnaissance
1	Manisha(2016)	85.1%
2	Hurieh(2016)	83.62%
3	Xiaozheng(2016)	70%

D'après les résultats du **tableau 2.2**, que l'apprentissage en profondeur n'est pas nécessairement la méthode la plus optimale pour résoudre le problème car le taux de reconnaissance n'est pas trop précis. Ceci est fortement influencé par l'ensemble de données utilisé. Certains chercheurs utilisent couramment un ensemble de données concernant la reconnaissance faciale. La recherche sur différents jeux de données peut entraîner un taux de

reconnaissance différent. Par conséquent, un développement supplémentaire est nécessaire pour obtenir une méthode précise.

2.2.5 Application de retrait de lunettes par Deep Learning

Combien de temps faut-il habituellement pour choisir une nouvelle paire de lunettes au magasin? 10 minutes? 30minutes? Lorsqu'on est laissé sans surveillance, certes en moyenne une heure(inclinaison de la tête, demi sourire, coup de côté, paire suivante...). C'est un gros problème, car il est scientifiquement établi que le type de lunettes que vous portez a une incidence sur la perception de votre intelligence, de votre succès et de votre attrait. En 2016, il y avait une sorte de technologie qui a résolu ce problème. Les technologies DITTO ont développé un miroir virtuel qui permet aux clients d'essayer des centaines de produits dans le confort de leur maison : centrer le visage, tourner à gauche, tourner à droite et c'est terminé. On peut s'asseoir et évaluer quelles lunettes associent le visage au style et au caractère.

Il y a cependant une petite mise en garde pour ceux d'entre nous qui ont des difficultés optiques: on doit retirer les lunettes pour utiliser le miroir virtuel.

Retirer les lunettes des visages n'est pas une tâche nouvelle, et de loin. Une quantité considérable de littérature scientifique documente divers algorithmes de traitement d'images destinés à retirer les lunettes, souvent dans le but d'améliorer les technologies de reconnaissance faciale. En utilisant certaines mathématiques réfléchies avec des caractéristiques telles que le contraste, les bords et la congruence, ces techniques détectent et soustraient les pixels d'image contenant les lunettes, puis synthétisent la région faciale par lissage ou inférence. Malgré l'ingéniosité, ces algorithmes peuvent ne pas aboutir à la reconnaissance des lunettes et / ou à la reconstruction du visage. Ils peuvent également avoir beaucoup de difficulté à généraliser à travers différents tons de peau et à corriger les ombres, le grossissement et les reflets causés par les montures et les lentilles.

Alors pourquoi pourrait-on penser qu'en 4 semaines, ils pourraient améliorer quelque chose que des groupes de spécialistes ont consacré des années à développer? La réponse est qu'il y a eu une recrudescence récente d'algorithmes, de code et d'outils open source et d'informatique GPU qui ont ouvert la voie à l'application d'un apprentissage en profondeur à un ensemble infini de données de grande dimension problèmes. La force de l'apprentissage en profondeur réside dans le fait qu'il n'est pas nécessaire de concevoir et d'optimiser un algorithme en

fonction des fonctionnalités que l'on juge importantes pour la tâche; il suffit de fournir des exemples et le réseau de neurones identifie et pondère les caractéristiques pertinentes. Plutôt que de concevoir un algorithme pour identifier les lunettes, de les supprimer, puis de reconstruire les visages, on peut simplement former un réseau de neurones pour qu'il puisse effectuer toutes les opérations décrites ci-dessus. Non seulement cela nécessite beaucoup moins de temps et de connaissances de domaine, mais le réseau résultant peut être généralisé sur une gamme d'entrées beaucoup plus large.

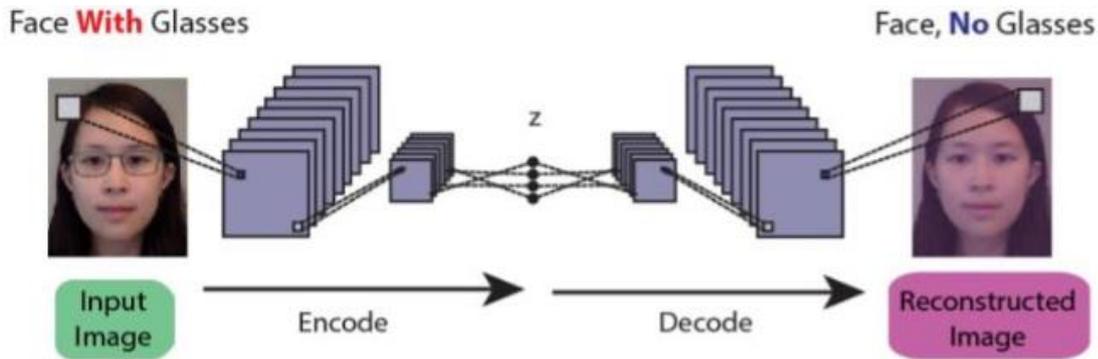


Figure 2.1 Principe du système de retrait de lunettes par apprentissage en profondeur[13]

2.2.6 Deep Learning et détection d'émotions

La détection d'émotion consiste à analyser des données ([textuelles](#), visuelles, audio, etc.) puis à déterminer les émotions qui y sont exprimées. La classification la plus simple de nos émotions comprend six catégories : la joie, la tristesse, la peur, la colère, le dégoût, la surprise. Généralement il faut ajouter une dernière étiquette : le **neutre** pour le cas où aucune émotion n'est présente. Le problème sur lequel nous avons travaillé consiste à identifier une émotion à partir de photographies d'un visage issues du data-set [fer2013](#), de petites images, avec un gros plan sur le visage et un seul visage à la fois.



Figure 2.2 Images utilisées [5]

Le résultat obtenu 53% de réussite, en soi, ça ne veut pas dire grand-chose. Pour comprendre ce score, il est intéressant de regarder les résultats classe par classe grâce à une matrice de confusion. Celle-ci indique pour chaque classe quel est la proportion de données correctement prédites et à quelles autres classes le reste des données ont été attribuées.

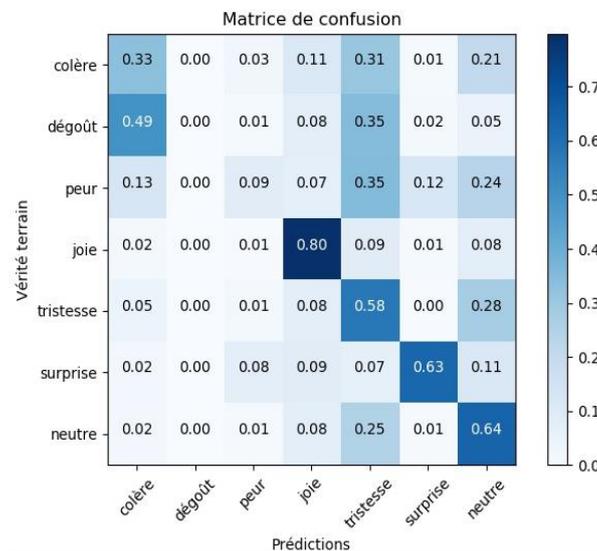


Figure 2.3 Matrice de confusion pour la détection d'émotions par DeepLearning [5]

L'analyse de la matrice de confusion pour ce réseau de neurones nous amène à être beaucoup moins sévères. Le réseau reconnaît très bien les visages joyeux et ne se débrouille pas trop mal pour la surprise ou l'absence d'émotion. Là où ça se gâte, c'est sur les émotions négatives : tristesse, colère, peur et dégoût : tout cela semble bien étrange pour cet algorithme. Ainsi pour la joie qui est la classe la mieux détectée, quelques 2872 images, soit au moins 900 images de plus que les autres classes. De même, avec seulement 176 photographies, il n'est pas surprenant que le réseau de neurone ne sache pas bien identifier le dégoût. En effet l'une des particularités des réseaux de neurones est de nécessiter de très nombreuses données d'apprentissage. À titre de comparaison, la célèbre base de données ImageNet contient environ 3000 images pour chacune des classes. Ce manque de données bloque notre réseau de deux manières différentes : i) en ne lui donnant pas assez d'information pour apprendre à différencier certaines classes ; ii) en l'amenant à se spécialiser sur les données d'apprentissage (over-fitting) et donc à être incapable de fournir une réponse correcte pour les données de tests.[5]

2.3 Etat de l'Art de l'optimisation par l'algorithme génétique

Les algorithmes génétiques sont des algorithmes d'optimisation heuristiques inspirés de la génétique et des mécanismes d'évolution de la nature : croisement, mutation et sélection en se basant sur l'idée que seuls les plus forts subsistent dans une population.[14] Ces algorithmes fonctionnent sur une population composée de chromosomes[15], tous différents, pouvant être des solutions potentielles du problème à résoudre. Dans notre problème de reconnaissance de visage 3D, les chromosomes représentent les paramètres caractéristiques. Chaque chromosome est évalué, cette évaluation permet de juger la pertinence des solutions (chromosomes) par rapport au problème considéré. Ceci conduit à éliminer les solutions jugées inutiles ou très mauvaises. Une fois cette élimination effectuée, les gènes des solutions sélectionnées sont combinés pour obtenir une nouvelle population qui doit être mieux adaptée à la résolution du problème que la population précédente. La population initiale donne ainsi naissance à des générations successives, sous la forme de chromosomes fils, mutés et hybridés à partir des chromosomes initiaux dits chromosomes parents. Ce processus est réitéré jusqu'à ce que l'on atteigne le nombre de génération prédéfini à l'avance

2.3.1 Optimisation du système de reconnaissance de visage 3D par AG

Un exemple d'application à l'optimisation des SRV3D est illustré dans la figure suivante :

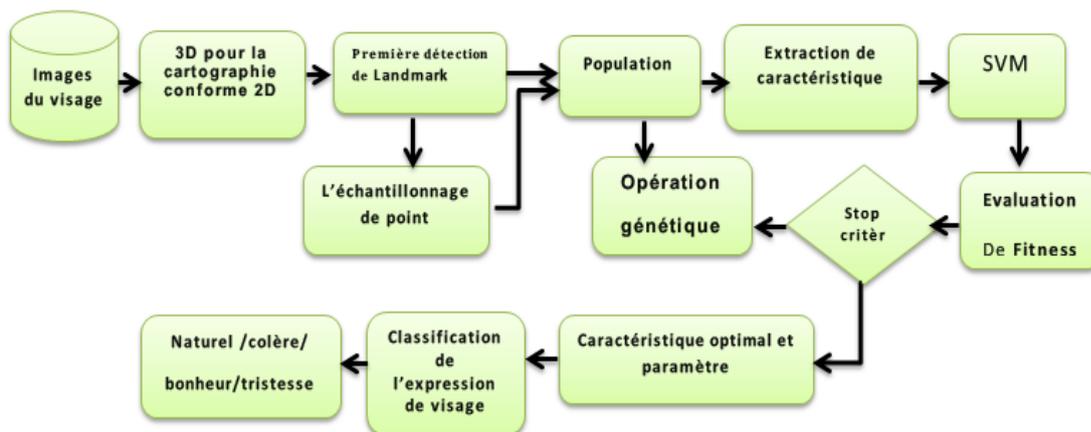
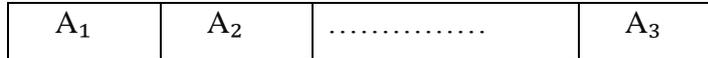


Figure 2.4 Exemple d'optimisation un SRV 3D par GA[7]

Chapitre II Etat de l'Art Deep Learning et l'Algorithme Génétique pour la RV

Un schéma de codage binaire est utilisé pour représenter la présence ou l'absence d'un vecteur propre particulier de la dilatation linéaire des images d'apprentissage .

La représentation des chromosomes est comme montré ci-dessous 1 bit pour chaque caractéristique.

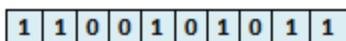


$A_i=0$: la caractéristique n'est pas sélectionnée,

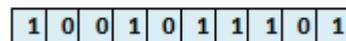
$A_i=1$: la fonction est sélectionnée.

Chaque chromosome est vérifié pour voir comment il est bon à résoudre le problème à la main et attribuer un score de remise en forme en conséquence. Le score fitness est une mesure de la façon dont ce chromosome est bon à résoudre le problème. Ici on a utilisé le taux de reconnaissance comme le score de fitness du chromosome. Pour trouver le taux de reconnaissance pour le score fitness , on utilise les ensembles distincts d'apprentissage et de test. L'ensemble d'apprentissage dans ce travail [7] se compose d'une image frontale de chacune des 20 personnes et 20 autres images des memes 20 personnes sont utilisées pour les test. Deux chromosomes parents sont choisis dans la population actuelle . La chance d'être sélectionné est proportionnelle au fitness des chromosomes. La méthode de roulette est utilisée pour sélectionner chromosome parents. Le score de fitness totale de la population est représenté par un graphique circulaire, ou la roue de roulette. Pour chaque membre de la population une tranche de la roue est affectée. La taille de la tranche est proportionnelle à la note individuelle de fitness . le membre est choisi en faisant tourner le ballon . Selon le taux de croisement , les bits de chaque chromosome choisi sont croisé à un point choisi au hasard . Seul deux points croisement et croisement uniforme peuvent être utilisés. Supposons que :

PARENT 1 :

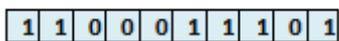


PARENT 2 :

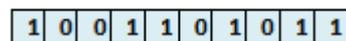


Le point de croisement est sélectionné après le 4 ème bit , puis les chromosomes du fils seront :

FILS 1 :



FILS 2 :



L'étape à travers laquelle on choisit les bits de chromosomes et bascule en fonction du taux de mutation. C'est la probabilité qu'un bit dans un chromosome sera retourné. C'est généralement une valeur très faible pour gènes codés binaires 0,001.

Chapitre II Etat de l'Art Deep Learning et l'Algorithme Génétique pour la RV

Il faut trouver les deux pires individus et les remplacer par les fils. Ce processus est : (i) la sélection de deux parents (ii) la permission de créer deux fils (iii) le remplacement immédiat des deux pires individus dans la population avec le fils.

Ce processus est répété jusqu'à ce qu'un critère d'arrêt est atteint. (on utilise le comptage égale à 15 pour répéter processus).

Enfin la sélection de caractéristiques dans le score le plus élevé de fitness de chromosomes est utilisée pour le traitement ultérieure de l'APC (Analyse en Composante Principale ACP ou PCA)) et ensuite pour la classification. Les ondelettes et le PCA pour l'extraction de caractéristiques sont utilisées. Les images sont décomposées jusqu'à quatre niveaux en utilisant la transformée en ondelettes (Daubechies 4) DB4. La bande de fréquence basse (LL) au 4^{ème} niveau est utilisée pour le calcul de la PCA. Un algorithme génétique est utilisé pour la sélection des caractéristiques. Le SVM et classificateur de distance sont utilisés pour la classification. Les résultats sont testés sur les bases de données ORL, FERET, Yale, Yale B.

Tableau 2.3 Taux de reconnaissance en fonction de l'augmentation pose sur ORL[7].

Technique	Nombre d'images d'apprentissage par personne / nombre d'images test							
	4 / 240				6 / 160			
	% vecteurs propres sélectionnés pour vecteurs propres totales par GA							
	10	15	20	25	10	15	20	25
Wavelet-PCA-SVM(Poly-Linear)	90	95.8	96.6	97.5	91.8	95	98.7	99.3
Wavelet-PCA-SVM(Poly-Quad)	89.1	94.1	95	96.2	91.2	94.3	96.8	97.5
Wavelet-PCA-SVM(Radial Basis)	89.1	95	95	95.8	91.2	94.6	96.8	97.5
Wavelet-PCA-ND	89.1	95	95	95.8	91.2	93.1	96.8	97.5

Tableau 2.4 Taux de reconnaissance sur les BDDs ORL, FERET, Yale, Yale B [7].

Algorithme	Taux de reconnaissance %			
	ORL	FERET	Yale	Yale B
PCA-ND	88	86	88.3	95
Wavelet-ND	82	83.6	90	92.5
Wavelet-PCA-ND	90	92	91.6	96.3
Wavelet-PCA-SVM	93.7	93	93.3	96.3
Wavelet-PCA-GA-ND	94.3	94.1	93.3	97.5
Wavelet-PCA-GA-SVM	97.3	98	100	100

Pour les visages normaux, la plupart des algorithmes peuvent atteindre un taux de reconnaissance élevé. Tous les algorithmes pris en considération pour les bases de données de Yale et de Yale B ont atteint un taux de reconnaissance élevé parce que seulement 15 et 10 sujets différents sont utilisés respectivement.

2.3.2 Segmentation d'image par le CNN et le AG

La mise en œuvre de la segmentation d'images en utilisant un algorithme génétique (AG) est divisée en deux parties:

- Codage de l'image dans un format binaire approprié ;
- Évolution de l'image codée à l'aide d'opérateurs génétiques.

Une image discrétisée est typiquement modélisée comme un tableau d'intensité de pixels :

$p(n, m)$ $n = 1, \dots, N$, $m = 1, \dots, M$.

Les images considérées sont composées d'objets d'intensité constante R_o intégrés dans l'arrière-plan d'intensité R_b et corrompu par le blanc additif bruit gaussien avec moyenne nulle et variance 'a'. Le bruit la variance signifie la quantité de bruit gaussien présent dans l'image. Ceci est principalement dû au bruit de la caméra et apparaît fréquemment sous forme de grains dans l'image. Le filtre de Weiner est utilisé pour supprimer le bruit gaussien dans l'image. La variance de l'image de différence résultante donne la variance réelle du bruit de l'image d'origine. En raison de la nature non stationnaire des images, traiter une image en tant que processus stationnaire peut ne pas conduire à une restauration optimale. L'approche de la restauration adaptative peut aussi être une solution et l'approche la plus appropriée. Théoriquement, l'image brute peut être de n'importe quelle taille. Par exemple, pour une image de taille 64 x 64 pixels, la recherche sur un hyperspace de $2^{64} \times 64$ dimensions est nécessaire. Pour limiter la dimensionnalité de l'espace de recherche, l'image est divisée en sous-images avant d'effectuer l'image segmentation. Dans ce travail [16], l'algorithme de segmentation est réalisée sur 16 x 16 sous-images ne se chevauchant pas, qui sont ensuite combinés pour obtenir l'ensemble de l'image segmentée.

2.3.2.1 Codage de l'image dans un format binaire approprié à l'aide du CNN

Puisque GA traite les valeurs 1 (R_o) ou 0 (R_b), l'image grise est convertie en image binaire. Ceci est possible en utilisant un modèle CNN présenté ci-dessous :

$$\mathbf{A} = \begin{array}{|c|c|c|} \hline 0 & 0 & 0 \\ \hline 0 & 2 & 0 \\ \hline 0 & 0 & 0 \\ \hline \end{array} \qquad \mathbf{B} = \begin{array}{|c|c|c|} \hline 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline \end{array}$$

$$\mathbf{Z} = -\mathbf{Z}^* = -1 < \mathbf{Z}^* < 1$$

Z est le seuil de la cellule.

Ce modèle convertit l'image de niveaux de gris en image binaire. Z est le seuil de la cellule.

2.3.2.2 Formation des cordes

Dans l'AG, un schéma de représentation chromosomique détermine la structure de la solution, ainsi que les opérateurs génétiques utilisés. Ici, une sous-image à deux dimensions 16 x 16 p (n, m) est convertie en utilisant une technique de balayage dans un vecteur $[X_1, X_2, \dots, X_{256}]$ où $X_i = p(n, m)$

$$n = 1, \dots, 16, \quad m = 1, \dots, 16 \text{ et } i = (n-1) \times 16 + m.$$

Les chromosomes correspondants sont des chaînes de 256 bits constituées de gènes représentant soit les intensités de l'objet (R_o), soit l'arrière-plan (R_b).

2.3.2.3 Choix de la population initiale

La population individuelle est un ensemble d'individus de première génération dans l'espace de recherche représenté par

$$\{Y_i^k\} \quad i = 1, 2, \dots, 256 \quad k = 1, 2, \dots, N.$$

Où $\{Y_i^k\}$ est choisi au hasard pour être R_o ou R_b et N est la taille de la population, le nombre qui joue un rôle important dans la décision du temps d'exécution de l'algorithme. Ces solutions d'initialisation aléatoire sur l'ensemble espace de recherche ne dépend pas de la connaissance de variance de bruit. Afin de s'assurer que l'AG converge pour la solution globale, la taille de la population doit être grande. Malheureusement, cela se traduit par une augmentation des calculs temps.

2.3.2.4 Évaluation de la fonction de mise en forme dans l'ordinateur CNN

Une fonction de fitness est le lien entre l'AG et le problème d'application à résoudre. Il retourne la valeur associée à chaque individu en indiquant l'adéquation du membre dans un environnement donné. La condition physique mesure la similarité de l'individu par rapport

Chapitre II Etat de l'Art Deep Learning et l'Algorithme Génétique pour la RV

à l'image bruyante d'origine. Laisser l'image originale être représentée comme ça : $[\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{256}]$

Et la population initiale par :

$$\begin{bmatrix} Y_1^1, Y_2^1, \dots, Y_{256}^1 \\ Y_1^2, Y_2^2, \dots, Y_{256}^2 \\ \dots \\ Y_1^N, Y_2^N, \dots, Y_{256}^N \end{bmatrix}$$

Le modèle CNN, utilisé pour calculer la forme physique, est présenté ci-dessous.

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} f(Y^k) & f(Y^k) & f(Y^k) \\ f(Y^k) & f(Y^k) & f(Y^k) \\ f(Y^k) & f(Y^k) & f(Y^k) \end{bmatrix}$$

La condition physique de chaque individu $f(Y^k)$ de la population est définie comme ça :

$$f(Y^k) = E(Y^k) + \alpha T(Y^k); K = 1, 2 \dots N \quad \dots \textcircled{3}$$

$$E(Y^k) = \frac{1}{\sum_{i=1}^{256} |Y_i - z_i|} \quad \dots \textcircled{4}$$

C'est une mesure de similitude entre l'individu et l'image bruitée d'origine. Le second terme, $T(Y^k)$, est la fonction de pénalité définie par l'inverse du nombre de transitions dans les directions horizontale et verticale. Ce terme est introduit dans la fonction de mise en forme afin de réduire le bruit haute fréquence dans les résultats segmentés. Est le facteur de pondération qui assure la normalisation des deux quantités. Ici, B est un modèle dynamique dont les valeurs sont obtenues à l'aide de la formation et cela dépend de l'image.

2.3.2.5 -Évolution de l'image codée à l'aide d'opérateurs génétiques

Une fois que l'évaluation de la condition physique des individus a été effectuée, les individus les plus aptes doivent être choisis comme parents pour la production de printemps, qui serviront de population à la génération suivante. Dans cette étude, les individus ayant une aptitude supérieure au seuil θ sont sélectionnés,

$$\theta = \frac{\text{Max}(f) + \text{Min}(f)}{2} \quad \dots(5)$$

Où $\text{Max}(f)$ et $\text{Min}(f)$ sont les valeurs maximales et minimales de la condition Physique de la population. Le seuil est abaissé de manière appropriée si le nombre d'individus qualifiés K est inférieur au nombre minimal de parents N_p requis dans le processus de reproduction.

- La reproduction

La reproduction est une étape importante de l'AG qui code l'adaptation du processus d'évolution à un environnement. Cette étape génère la population d'individus qualifiés de la génération suivante. Dans cet article, la phase de reproduction comprend deux opérateurs, à savoir cross-over et mutation. L'implémentation de ces opérateurs est expliquée ensuite.

- Cross-over

L'opérateur de croisement génère deux nouveaux ressorts à partir d'une paire de parents en échangeant des parties de chromosomes parents. Traditionnellement, les parents sont accouplés de manière aléatoire et les deux nouvelles sources créées par les parents. Les deux nouvelles sources créées par les parents (Y_1, Y_j) à l'aide d'un croisement à deux points sont données par

$$\{Y_1, Y_2^i, \dots, Y_{p_1-1}, Y_{p_1}, \dots, Y_{p_2}, Y_{p_2+1}, \dots, Y_{256}\}$$

Et

$$\{Y_1^j, Y_2^j, \dots, Y_{p_1-1}^j, Y_{p_1}^j, \dots, Y_{p_2}^j, Y_{p_2+1}^j, \dots, Y_{256}^j\}$$

Où p_1 et p_2 sont deux points de croisement choisis au hasard. Les parents sélectionnés dans une population de taille N engendreront une progéniture de K ($K - 1$), ce qui donnera un ensemble d'individus $N + K$ ($K - 1$) dans le bassin d'accouplement.

- Mutation

L'opérateur de mutation est introduit dans l'AG afin de développer les traits de chaque individu influencés par son environnement. L'opérateur ne produit pas de nouveaux ressorts mais améliore la condition physique de l'individu. Un filtre médian attribue à chaque gène la valeur que possède la majorité de ses voisins comme suit:

$$R_o \quad \text{Si} \quad N_o > N^b \quad Y_i^1 = \begin{cases} R^b & \text{Si} \quad N^b > N_o \end{cases}$$

Où N^b et N_o sont les nombres de pixels d'arrière-plan et d'objet dans un voisinage sélectionné de Y , respectivement. Le filtre médian est mis en œuvre avec CNN, les modèles étant utilisés comme suit:

$$A = 1 \quad , \quad B = \begin{bmatrix} \dot{\alpha} & d & d \\ d & 0 & d \\ \dot{\alpha} & d & d \end{bmatrix}$$

$$\dot{\alpha} = Q_u \text{sigv}_1(\Delta v_x^k), I_f = 0$$

Où $Q^k = Q = 1$. Chaque cellule de CNN fonctionnera dans la région caractéristique linéaire. A ce stade, seuls les individus les plus aptes sont sélectionnés pour former la population de la

prochaine génération. Pour maintenir la diversité des solutions, s'il y a des individus identiques dans la population, un seul sera conservé.

- Critère de résiliation

Il existe deux choix de critères de terminaison, qui indiquent la convergence de l'AG.

1. La forme physique moyenne ou la valeur de seuil ne peut pas changer de manière significative d'une génération à suivante.[17]
2. Le nombre de générations dépasse le nombre souhaité et la forme physique maximale et minimale de la population actuelle se situe dans la plage souhaitée.

2.3.2.6-Résultats de la segmentation des images par le CNGA :

La procédure de segmentation décrite a été implémentée sur des images synthétiques. Une étude des effets des paramètres GA a été effectuée sur des images 64x 64.



(a) Image originale (b) Image segmentée

Figure 2.5 Exemple d'image segmentée par le CNGA [d'après le net]

L'image segmentée montre diverses frontières de régions et fond segmentés avec une bonne clarté. Le processus pourrait discriminer les objets du fond et aussi les contours des régions.

- **Analyse du temps de traitement** : L'image est divisée en blocs de 16x16, pour une image de 64x64, 16 blocs sont pris. Chacun de ces blocs est acheminé vers le CNGA et traité de manière parallèle. Pour un bloc considéré le modèle de conversion CNN 3x3 est utilisé pour obtenir une image binaire. On a besoin de 456 modèles CNN pour traiter un bloc 16x16 de manière parallèle. Un processus similaire est effectué pour calculer la fonction

- de mise en forme et la mutation. Par conséquent, pour les blocs 16x16, si la segmentation d'images basée sur l'AG prend 1 milliseconde; lorsqu'il est implémenté par CNN dans un ordinateur il prend 0,7 microsecondes. Tableau 2.5 présente la comparaison entre la segmentation d'images basée sur GA et la segmentation basée sur GA mise en œuvre à l'aide d'un ordinateur CNN pour une génération. L'analyse est effectuée pour une génération seulement si la génération de nombres est augmentée, la convergence de la segmentation d'images basée sur l'AG prend plus de temps et ce problème peut être efficacement résolu s'il est implémenté dans un ordinateur CNN. La figure 2.6 montre le temps de traitement pour différentes tailles d'image. Les résultats montrent que des tailles d'image plus grandes telles que 512x512 et 1024x1024 peuvent être facilement segmentées dans un ordinateur CNN de manière efficace en temps et en vitesse.

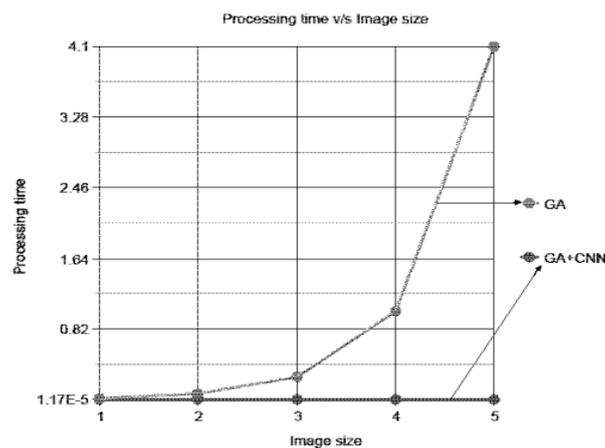


Figure 2.6 La taille de l'image Vs Temps de traitement [16]

Tableau 2.5 Temps de traitement pour différentes tailles d'image [16]

Serial no	Image Size	Processing Time GA (Seconds)	Processing Time for GA+CNN (Seconds)
1	64x64	0.016	1.172e-05
2	128x128	0.064	4.688e-05
3	256x256	0.256	1.852e-04
4	512x512	1.024	7.500e-04
5	1024x1024	4.096	3.750e-03

2.3.3 Reconnaissance de visage par CNN optimisé par AG

En tant que méthode performante, l'apprentissage en profondeur présente encore certaines faiblesses. L'un d'entre eux est la grande complexité. Les autres problèmes qui apparaissent toujours lors de la reconnaissance des visages sont la pose variée du visage, l'illumination et les expressions faciales à reconnaître. Sur la base de ces problèmes ça nécessite une nouvelle méthode. L'une des méthodes développées dans le but d'optimisation peut être appelée système de reconnaissance de visage hybride, qui combine les capacités d'un réseau de neurones convolutionnels(CNN), la machine à vecteurs de support (SVM) et l'algorithme génétique (GA). Le but de cette méthode d'optimisation est d'obtenir des résultats de haute précision et de faible complexité. Dans cette méthode d'optimisation, l'algorithme génétique est utilisé pour trouver la structure optimale de CNN, puis la dernière couche de la structure CNN est remplacée par SVM. Voici la phase d'optimisation du système de reconnaissance de visage hybride utilisant GA, CNN et SVM:

Détermination des paramètres sur CNN par l'algorithme génétique (GA)

L'algorithme génétique (GA) est une technique de recherche méta-heuristique basée sur les mécanismes de sélection naturelle, de génétique et d'évolution. Les algorithmes génétiques sont couramment utilisés pour produire des solutions de haute qualité à des fins de dépannage nécessitant une optimisation.

Dans cette étude, un algorithme génétique est utilisé pour déterminer la structure sur CNN. Le CNN utilise généralement une recherche exhaustive. Lorsqu'un réseau a peu de paramètres (le nombre de neurones dans la couche cachée), une recherche exhaustive permet de l'utiliser. Mais quand un réseau a de nombreux paramètres qui rendent les structures CNN complexes, il n'est pas possible d'utiliser une recherche exhaustive, comme certaines réalisées qui utilise des centaines de millions de paramètres qui nécessitent de grandes bases de données et une puissance de calcul importante. Donc, cette limitation est résolue avec l'application de l'algorithme génétique qui remplace la recherche exhaustive.

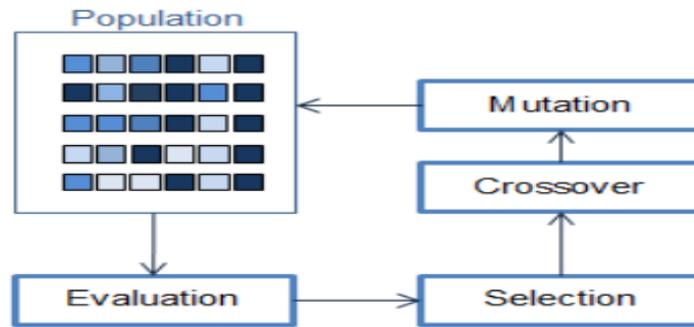


Figure 2.7 Architecture de l'algorithme génétique[12]

Sur la base du diagramme ci-dessus, trois étapes principales sont utilisées pour déterminer la structure de CNN, à savoir:

1) La sélection est un opérateur utilisé pour déterminer quel individu convient au système. Cet opérateur fonctionne comme un système biologique humain, dans lequel il évaluera chaque individu. Les individus malsains seront éliminés.

2) Le crossover, un opérateur utilisé pour échanger des informations entre individus. Cet opérateur fonctionne comme un système reproducteur humain, dans lequel des individus sont combinés pour produire une nouvelle génération.

3) la mutation, opération utilisée pour modifier la caractéristique de l'individu afin de former un nouvel individu.

Les individus des étapes ci-dessus sont représentés sous forme de paramètre existant sur CNN. Basé sur trois étapes au-dessus du nombre de paramètres que beaucoup ont ensuite simplifié pour passer à l'étape suivante. L'optimisation de ce paramètre inclut la taille des champs récepteurs sur chaque couche, le nombre de champs récepteurs dans chaque couche, les relations entre les couches en séquence.

- **Optimisation de CNN et SVM**

Au stade précédent, la méthode proposée produisait des paramètres qui permettaient de construire des structures CNN. Il faut choisir combien de couches de la structure CNN. Une fois la structure de CNN créée, le processus sera lancé avec le processus sur la première couche. Il y a quelques processus dans l'architecture CNN. Premièrement, le premier calque est utilisé comme emplacement comme entrée sous la forme de pixels d'image bruts. Deuxièmement, les deuxième et quatrième couches du réseau sont des couches de convolution d'alternateur avec des couches de sous-échantillonnage, qui sont utilisées pour extraire la carte combinée en entrée. Cela entraîne le système à extraire des fonctionnalités

Chapitre II Etat de l'Art Deep Learning et l'Algorithme Génétique pour la RV

avec des fonctionnalités nombreuses et variées. Ils fonctionnent comme une couche cachée dans l'architecture CNN [32].

Une fois que la couche masquée de CNN a obtenu le résultat, ce résultat est distribué dans la dernière couche. La dernière couche est remplacée par la classification SMV. Le principal objectif de SVM est de trouver la limite de décision la plus appropriée et de maximiser la distance entre les classes de données. Les résultats de la couche masquée sont pris par SVM en tant que vecteur de caractéristiques pour le processus de formation.

Cette étape de la formation se fait en continu jusqu'à ce que l'on soit en bonne condition. Ainsi, les tests dans le jeu de données peuvent être effectués par classifieur SVM avec des fonctionnalités extraites automatiquement. Voici la structure de l'optimisation du modèle hybride GA-CNN-SVM.

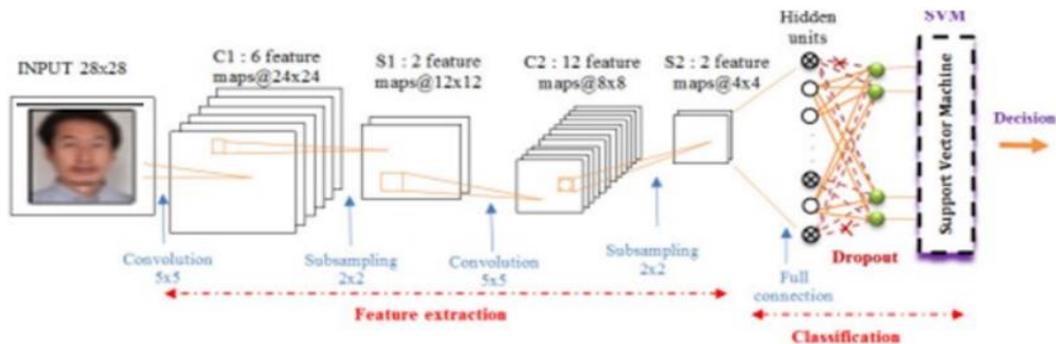
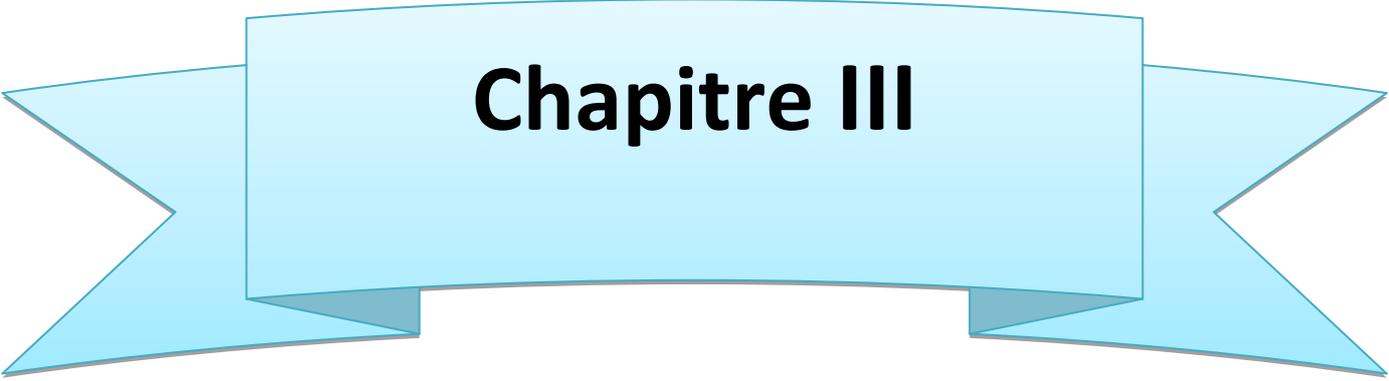


Figure 2.8 Structure de la méthode hybride GA-CNN-SVM[12]

2.4 Conclusion

Dans ce chapitre nous avons passé en revue les approches récentes d'optimisation de SRV utilisant CNN et GA. Cette étude nous permis de constater que la reconnaissance du visage présente plusieurs défis aussi nous avons présenté les étapes de base de chaque méthode ainsi que leur applications dans des différents domaines.



Chapitre III

3.1 Introduction

L'apprentissage en profondeur (deep Learning), qui utilise comme modèle les réseaux de neurones profonds, a montré de bonnes performances dans de nombreuses tâches difficiles d'intelligence artificielle et d'apprentissage automatique, telles que la reconnaissance d'images.[18]

Ces réseaux profonds ont donné des résultats impressionnants et parfois concurrentiels. L'architecture profonde de CNN peut être divisée en deux parties principales. La première partie, basée sur les couches convolutives CNN, offre la possibilité d'extraire des caractéristiques et d'encoder l'image d'entrée. Tandis que le second est un classifieur de réseau neuronal entièrement connecté dont le rôle est de générer un modèle de prédiction pour la tâche de classification. Un modèle CNN est décrit par de nombreux hyper-paramètres, notamment le nombre de couches convolutives, le nombre de filtres et leurs tailles respectives, etc[19]

Un algorithme génétique (GA) peut être utilisé pour optimiser les nombreux paramètres d'un réseau de neurones à convolution (CNN) qui contrôlent la structure du réseau. Les CNN sont utilisés dans les problèmes de classification d'images dans lesquels il est nécessaire de générer des descripteurs de caractéristiques pour distinguer les classes d'images[20]. Dans notre travail, nous intéressons à l'optimisation du CNN par l'algorithme génétique afin d'améliorer les performances du système de reconnaissance et surtout améliorer le temps d'exécution qui est le rôle approprié du CNN.

3.1 Optimisation du CNN par l'algorithme génétique

Dans la méthode proposée, une GA est utilisée pour optimiser les paramètres d'un CNN. Le but de cette méthode est d'optimiser les performances de CNN lors de la reconnaissance faciale. La reconnaissance des visages est un type de problème de classification des images dans lequel les méthodes classiques d'analyse d'images ont échoué en raison de diverses difficultés, notamment les modifications de l'éclairage, le flou de fond des images. Les CNN ont montré qu'ils sont capables d'atteindre une grande précision lors de la reconnaissance faciale, mais le choix d'une structure CNN performante prend du temps en raison du nombre de paramètres qui affectent la structure du réseau. Pour cette raison, une AG est appliquée pour rechercher l'espace des structures de réseau possibles. Dans cette section, un examen de la classification générale des images est fourni.

3.1.1 Apprentissage profond basée sur un réseau de neurones convolutionnel

La reconnaissance des visages est un exemple de tâche de classification d'images dans le monde réel. Les travaux modernes ont tendance à montrer une grande précision dans la reconnaissance des visages, divers modèles et techniques fonctionnent de mieux ou moins bien en fonction de l'environnement de test, du système de caméra, de l'éclairage, de la présence d'autres objets dans l'image,

La classification des images est utilisée dans de nombreuses applications, notamment la reconnaissance de visage, la détection d'objet et la reconnaissance de scène. En règle générale, le processus de classification des images implique la modélisation d'un ensemble de données d'apprentissage afin qu'il soit possible de distinguer deux classes ou plus d'objets ou d'images à l'aide d'un classificateur. Ensuite, une fois les données modélisées, le classificateur peut être appliqué à un nouvel ensemble de données dans un paramètre d'application. [20]

CNN est utilisé comme extracteur de fonctionnalités pour acquérir automatiquement des fonctionnalités remarquables. Nous pré-entraînons d'abord notre CNN à l'aide de données auxiliaires pour obtenir les pondérations mises à jour, puis nous entraînons le CNN en fonction du jeu de données cible pour extraire davantage de traits masqués. Enfin, nous utilisons SVM comme classificateur au lieu de CNN pour reconnaître toutes les classes. Grâce à la saisie des caractéristiques faciales extraites de CNN et optimisée par l'AG, SVM reconnaîtra plus précisément les images de visage [21]

3.1.2 Quelques définitions sur les réseaux de neurones et le CNN

Un réseau de neurones est un modèle mathématique dont la conception est inspirée des neurones biologiques. Cette architecture de réseau est divisée en couches. Chaque couche est un ensemble de neurones. La première couche d'un réseau de neurones est la couche d'entrée dans laquelle nous injectons les données à analyser. La dernière couche est la couche de sortie. Dans un problème de classification, il retourne un nombre de classes. Les couches au milieu sont les couches cachées du réseau de neurones.

Comme le montre la **figure 3.1**, dans un réseau de neurones, un seul neurone a plusieurs entrées. Chaque connexion d'entrée est caractérisée par un poids W_{ij} . Lors de l'activation du neurone artificiel, il calcule son état S_{jen} additionnant toutes les entrées multipliées par leurs poids de connexion correspondants. Pour que le neurone soit activé même lorsque toutes les

entrées sont nulles, une entrée supplémentaire, appelée polarisation b , est ajoutée. Cette entrée supplémentaire est toujours égale à 1 et possède sa propre connexion de poids.

Pour normaliser son résultat Y_i (normalement entre 0 et 1), le neurone le transmet à travers sa fonction d'activation

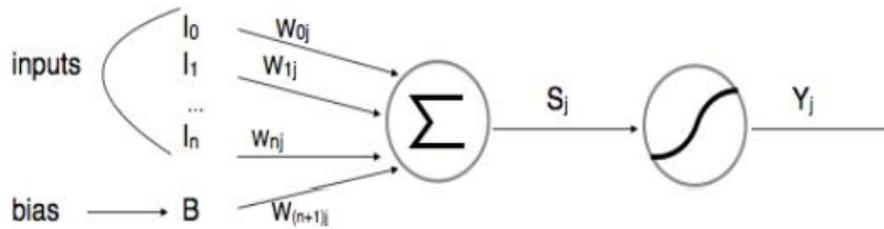


Figure3.1 Paramètres du neurone [19]

Il y a quatre opérations principales dans les blocs de chaque CNN : [19]

1. La convolution
2. La fonction d'activation (ReLU)
3. Le Pooling or Sub Sampling
4. La classification (Fully Connected Layer)

3.1.2.1 Filtre de convolution

La couche de convolution tire son nom de l'opérateur de convolution. Le but de cette couche est l'extraction des caractéristiques de l'image. La convolution conserve la relation spatiale entre les pixels en apprenant les caractéristiques de l'image à l'aide de petits carrés de données d'entrée. Chaque couche de convolution utilise divers filtres pour la détection et l'extraction de caractéristiques, telles que détection des contours, netteté, flou, etc. Ces filtres sont également appelés «noyaux» ou détecteurs de fonctionnalités ». Après avoir glissé le filtre sur l'image, nous obtenons une matrice appelée carte des fonctionnalités.

Dans la première couche de convolution, la convolution se situe entre l'image d'entrée et ses filtres. Les valeurs de filtre sont les poids des neurones (voir l'équation (1)). Dans les couches profondes du réseau, l'image résultante des convolutions est la somme des k^{l-1} convolutions, avec k^{l-1} le nombre de sorties de couche $l - 1$ (voir l'équation (2)).

$$\sigma_k^{(l)} = x_k^{(l)} \cdot w_k^{(l)} + b_k^{(l)} \quad (1)$$

$$\sigma_k^{(l)} = \sum_{i=1}^{k^{l-1}} x_{k,i}^{(l)} * w_{k,i}^{(l)} + b_k^{(l)} \quad (2)$$

Avec $\sigma_k^{(l)}$ la valeur du neurone k de la couche l , $x_k^{(l)}$ le vecteur d'entrée du neurone k , $w_k^{(l)}$ le vecteur de poids et $b_k^{(l)}$ le biais. En pratique, un CNN apprend seul les valeurs de

ces filtres au cours du processus de formation. Les paramètres tels que le nombre de filtres, la taille du filtre et l'architecture du réseau sont spécifiés par le scientifique avant le lancement du processus de formation. Plus nous avons de filtres, plus les caractéristiques d'image sont extraites et plus le réseau devient performant lors de l'extraction des caractéristiques et de la classification des images.

La taille de la carte de caractéristiques est contrôlée par trois paramètres qui sont:

- *Depth* : correspond au nombre de filtres utilisés pour l'opération de convolution.
- *Stride* : c'est le nombre de pixels par lequel nous glissons notre matrice de filtre sur la matrice d'entrée. Avoir une foulée (stride) plus grande produira des cartes de caractéristiques plus petites.
- *Zero-padding* : Pour appliquer le filtre aux éléments adjacents à l'image d'entrée, il est utile de compléter la matrice d'entrée avec des zéros autour de la bordure.

3.1.2.2 Fonction d'activation

Une fois la convolution terminée, une fonction d'activation est appliquée à toutes les valeurs de l'image filtrée pour extraire les entités non linéaires. Il existe de nombreuses fonctions d'activation telles que :

- le ReLU, défini comme suite :

$$f(x) = \max(0, \sigma)$$

- la fonction $\tanh()$
- ou la fonction sigmoïde

La valeur de sortie $s_k^{(l)}$ d'un neurone k de couche l dépend de sa fonction d'activation et se

définit comme suit (voir l'équation (3.3)) : $s_k^{(l)} = f(\sigma_k^{(l)})$ (3.3)

Avec la fonction d'activation $\sigma_k^{(l)}$ et la valeur de neurone k de couche l .

Le choix de la fonction d'activation peut dépendre du problème. La fonction ReLU remplace toutes les valeurs de pixel négatives dans la carte de fonctions par zéro. Le but de ReLU est d'introduire la non-linéarité dans le CNN, car l'opérateur de convolution est une opération linéaire et la plupart des données d'entrée du CNN seraient non linéaires. Résultat de convolution et l'opération ReLU s'appelle une carte de caractéristiques rectifiée.

3.1.2.3 Le Pooling

L'étape de Pooling est aussi appelé sous-échantillonnage. Il vise à réduire la dimensionnalité de chaque carte de caractéristiques rectifiée et conserve les informations les plus importantes. Les deux méthodes les plus utilisées dans cette opération sont le

poolingmoyen ou maximal. Après cette étape de sous-échantillonnage, nous obtenons une carte de caractéristique définie dans (4).

$$I_k^{(l)} = \text{pool}(s_k^{(l)}) \quad (4)$$

avec $I_k^{(l)}$ la carte de caractéristiques de la couche l , le Pool des opérations et $s_k^{(l)}$ la valeur de sortie du neurone k de la couche l .

Les avantages de la fonction de pooling sont:

- .les représentations en entrée (dimension d'entité) sont plus petites et plus maniables.
- .réduit le nombre de paramètres et de calculs dans le réseau, donc, contrôler l'overfitting.
- .rend le réseau invariant aux petites transformations, distorsions... dans l'image d'entrée. En fait, une petite distorsion en entrée ne changera pas la sortie de la mise en commun prise depuis la valeur maximale / moyenne dans un voisinage local.
- .aide à obtenir une représentation presque invariante de l'image d'entrée. Ceci est très puissant puisque nous pouvons détecter des objets dans une image, peu importe où ils se trouvent situé.

3.1.2.4 Classification (FullyConnected Layer)

La sortie des couches de convolution et de regroupement d'un CNN est le vecteur des caractéristiques d'image. L'objectif de la couche entièrement connectée est d'utiliser ces vecteurs d'entités pour classer les images d'entrée en plusieurs classes en fonction d'un jeu de données d'apprentissage étiqueté. La couche entièrement connectée est composée de deux parties. La première partie est constituée de couches dites entièrement connectées où toutes ses neurones sont connectées à toutes les neurones des couches précédentes et suivantes. La deuxième partie est basée sur une fonction objective. En fait, les CNN cherchent à optimiser une fonction objective, en particulier la fonction de perte. La fonction de perte la plus utilisée est la fonction Softmax. Elle normalise les résultats et produit une distribution de probabilité entre les différentes classes (chaque classe aura une valeur comprise dans l'intervalle [0, 1]). L'ajout d'une couche entièrement connectée permet l'apprentissage de combinaisons non linéaires d'entités extraites, ce qui pourrait être encore meilleur pour la tâche de classification. [19]

3.2 Structure du réseau convolutionnel

Plusieurs types de couches composent la structure des CNN, notamment: les couches de convolution, les couches de regroupement et les couches entièrement connectées. Les couches de convolution effectuent une multiplication entre l'image d'entrée ou un tableau à deux dimensions et un filtre qui est également un tableau à deux dimensions **Figure 3.2**

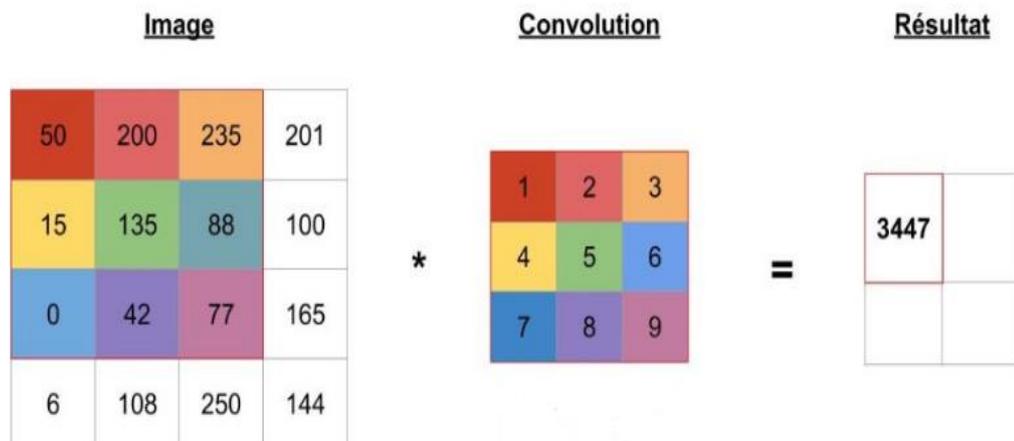


Figure 3.2 - : un exemple Calcul d'une étape de la convolution

Cette multiplication se produit à chaque emplacement du tableau d'entrée et le résultat est ensuite ajouté. Ce processus produit un autre tableau bidimensionnel et peut être décrit mathématiquement comme suit :

$$f[x, y] * g[x, y] = \sum_{i=-\frac{S}{2}}^{\frac{S}{2}} \sum_{j=-\frac{S}{2}}^{\frac{S}{2}} f[i, j] \cdot g[x - i, y - j], \quad (1)$$

f : image d'entrée ; g : filtre de convolution; S : taille du filtre (filtre carré) ; x et y : sont des emplacements dans les tableaux d'entrée.

Cependant, l'équation (1) apparaît généralement plusieurs fois dans chaque couche de convolution, mais avec des filtres différents utilisés pour g . Ainsi, il peut y avoir un ensemble de filtres utilisés dans chaque couche de convolution appelée g_1, g_2, \dots, g_f , où f est le nombre de filtres dans chaque couche de convolution. avec une taille de filtre de S et f couches de convolution dans le réseau, il faut au moins optimiser les pondérations $S^2 \cdot f$. Pendant le processus de formation, ce qui peut rapidement créer un problème qui n'est pas réalisable en fonction des besoins de l'application.

Le deuxième type majeur de couche dans un CNN est une couche de regroupement. Il existe plusieurs types de couches de mise en commun, mais celle qui est utilisée le plus souvent en raison de son efficacité est la couche max-pooling, qui effectue un sous-échantillonnage de son entrée en sélectionnant la valeur maximale dans chaque région et seulement en passant cette valeur.

Un exemple de ce processus peut être vu à la **figure 3.3** La mise en pool maximum est effectuée pour réduire la complexité de la sortie dans l'ensemble du CNN, car elle

sous-échantillonne et produit une sortie de dimension inférieure. Cependant, les informations ne sont pas perdues ici car les valeurs maximales sont capables de préserver les informations importantes tout au long du processus de convolution

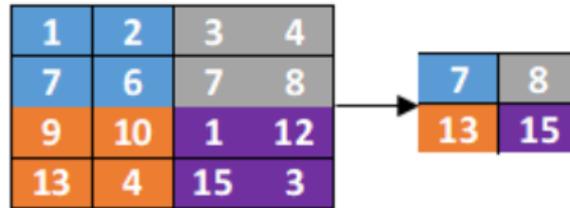


Figure 3.3 Un exemple de max-pooling.[20]

Dans la **Figure 3.4** l'entrée (à gauche) est divisée en quatre parties et chaque partie est recherchée pour la valeur locale maximale. La sortie (à droite) contient les valeurs maximales trouvées dans la recherche.

Le dernier type de couche souvent utilisé dans les CNN est une couche entièrement connectée. Ce type de couche fonctionne de manière similaire à un réseau multicouche de perceptrons, qui est un type classique de réseau neuronal. **Figure 3.4**

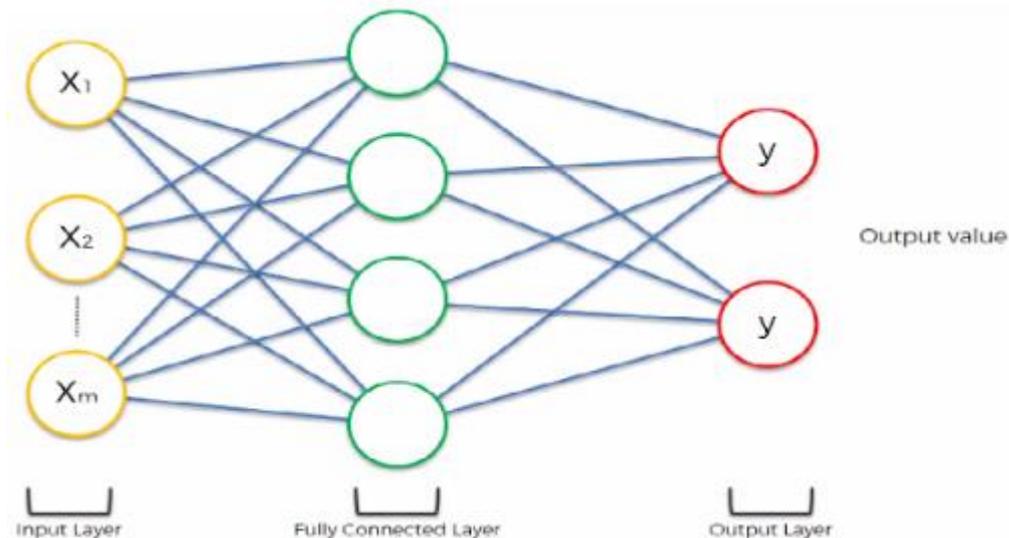


Figure 3.4 Schéma illustrant la couche full connected du CNN

Dans les couches entièrement connectées, chaque nœud d'entrée est connecté directement à chaque nœud de sortie et un poids est associé à chaque connexion. Ces poids sont optimisés à l'aide du processus de rétro-propagation, qui ne sera pas traité ici pour des raisons d'espace. Cependant, un nœud de sortie dans la couche entièrement connectée additionne sa valeur comme suit :

$$\text{Output} = \omega_1 V_1 + \dots + \omega_N V_N, \quad (2)$$

où il y a N nœuds connectés au nœud de sortie dans l'équation (2), et chaque nœud a un poids associé, ω_i , et une valeur, V_i , où $i = 1, \dots, N$. Les couches entièrement connectées sont généralement connectées à la fin d'un CNN comme dernier moyen d'introduire la non-linéarité dans le réseau. Cependant, les grandes couches entièrement connectées doivent être utilisées avec précaution car elles peuvent rapidement optimiser l'optimisation de nombreux poids et augmenter considérablement le temps de formation requis pour le réseau. Cela résulte des poids N^2 qui doivent être optimisés pour chaque couche entièrement connectée.[20]

3.3 Principe de l'algorithme génétique

Les GA sont des méthodes heuristiques de recherche ou d'optimisation. Ces techniques ont été inspirées à l'origine du principe darwinien de l'évolution par sélection (génétique). Une AG est basée sur une forme très abstraite de processus évolutifs pour apporter des solutions à des problèmes complexes. Chaque AG opère sur une population de chromosomes artificiels. Chaque chromosome signifie une solution au problème à résoudre et a une forme physique. Une forme physique du chromosome est une mesure numérique réelle qui représente sa performance en tant que solution du problème spécifique.

La méthode GA commence par une population de chromosomes générée aléatoirement. Il effectue ensuite un processus de sélection et de recombinaison en fonction de la forme de chaque chromosome. Les matériels génétiques parentaux sont recombinaisonnés pour générer des chromosomes enfants, produisant une nouvelle génération. Ce processus est itéré jusqu'à ce qu'un critère d'arrêt soit atteint. De cette manière, une AG développe la meilleure solution à un problème donné.

John McCall a tout d'abord proposé les AG comme méthode pour trouver les meilleures solutions aux problèmes autrement difficiles à résoudre du point de vue informatique. Le théorème de McCall, ainsi que l'hypothèse de bloc de construction associée, ont fourni une base théorique pour la conception d'AG efficaces. Le développement et le succès des AG ont considérablement contribué à leur adoption dans de nombreuses approches informatiques basées sur des phénomènes naturels. L'AG est désormais une partie importante du domaine plus vaste de l'intelligence informatique, comme les réseaux de neurones, l'optimisation de colonies de fourmis, etc[19]

Les algorithmes génétiques étant basée sur des phénomènes biologiques, il convient de rappeler au préalable quelques termes de génétique et leur définition dans le domaine informatique.[22]

Tableau 3.1 Résumé de la terminologie utilisée en AG [22]

Nature	définition
Individu	Solution à un problème
Population	Ensemble de solutions
Adéquation au milieu	Qualité de la solution
Gène	Caractéristique, partie de représentation d'une solution
Chromosome	Chaîne, représentation codée d'une solution
Locus	Position dans la chaîne
Allèle	Valeur de caractéristique
Génotype	structure
Phénotype	Ensemble de paramètres ou une structure décodé
Croissance	Décodage d'une solution depuis sa représentation
Croisement (Cross-over)	Opérateur d'exploration de l'espace de recherche
Mutation	Modification d'une solution
Sélection naturelle	Réutilisation d'une bonne (sous) solution
Fonction de fitness	La fonction-positive-que nous cherchons à maximiser

3.3.1 Les caractéristiques des AGs

LermanetNgouenet[33] et distinguent 4 principaux caractéristiques qui font la différence fondamentale entre les algorithmes génétiques et les autres méthodes d'optimisation :

1. Les algorithmes génétiques utilisent un codage des paramètres et non les paramètres eux-mêmes
2. Les AGs travaillent sur une population de points au lieu d'un point unique
3. Les algorithmes génétiques n'utilisent que les valeurs de la fonction étudiée, pas sa dérivée ou une autre connaissance auxiliaire.
4. Les AGs utilisent des règles de transition probabiliste et non déterministes.

3.3.2 Les fondements des algorithmes génétiques

Indépendamment de la problématique traitée, les AGs sont fondés sur :

- 1) Comment choisir un codage des solutions.
- 2) Une méthode pour générer une population initiale de solutions.
- 3) Une fonction d'évaluation (fitness) permettant d'évaluer une solution et la comparer aux autres
- 4) Un mécanisme de sélection pour déterminer quels individus sont plus aptes à obtenir les meilleurs résultats

5) Des opérateurs génétiques de reproduction qui font évoluer la population de solutions de manière progressive :

a) *Opérateur de croisement* :

Lors de cette opération, deux chromosomes s'échangent des parties de leurs chaînes afin de produire des individus meilleurs ou différents. Cet opérateur est effectué selon une probabilité P_c comprise entre 0 et 1 strictement.

b) *Opérateur de mutation* :

De façon aléatoire, un gène peut au sein d'un chromosome être substitué à un autre, cet opérateur évite d'établir des populations uniformes incapables d'évoluer. De la même manière que pour le croisement on définit un taux de mutation qui est généralement compris entre 0,001 et 0,01.

6) Un mécanisme d'insertion ou de remplacement pour décider ce qui doit rester et ce qui doit disparaître dans la nouvelle génération.

3.3.3 Fonctionnement des algorithmes génétiques

L'algorithme génétique présenté dans la figure 3.5 débute avec une population de base qui se compose le plus souvent de chaînes de caractères correspondant chacune à un chromosome. Le contenu de cette population initiale est généré aléatoirement on attribue à chacune des solutions une note qui correspond à son adaptation au problème, ensuite on effectue une sélection au sein de cette population. Lorsque deux chromosomes ont été sélectionnés, on réalise un croisement avec un opérateur de croisement choisie selon une probabilité P_c . On effectue ensuite des mutations sur une faible proportion d'individus avec une probabilité de mutation P_m . Ce processus nous fournit une nouvelle population. Les individus issus de ces opérateurs génétiques seront insérés par une méthode d'insertion dans la nouvelle population et un test d'arrêt sera effectué pour vérifier la qualité des individus obtenus. Si ce test est vérifié alors l'algorithme s'arrête avec une solution optimale sinon on réitère le processus un grand nombre de fois de manière à initier le principe d'évolution.

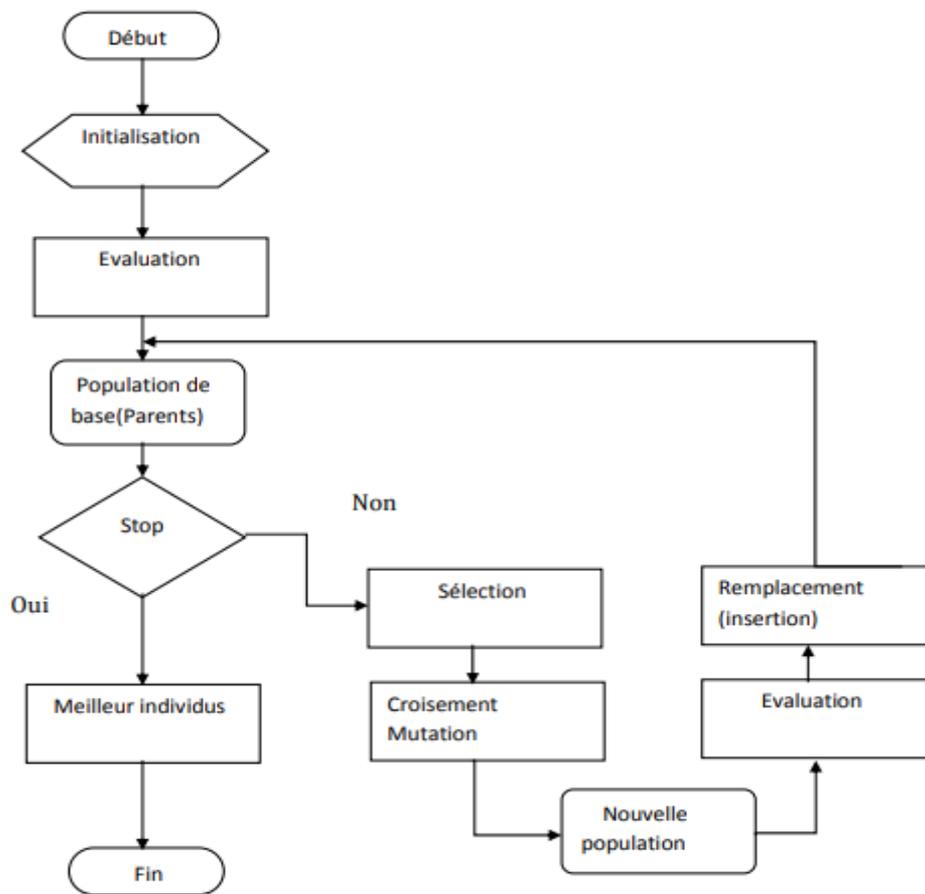


Figure 3.5 Principe générale d'un algorithme génétique [22]

3.3.4 Le codage

Le codage est une modélisation d'une solution d'un problème donné sous forme d'une séquence de caractères appelée chromosome où chaque caractère, dit aussi gène, représente une variable ou une partie du problème. La tâche principale consiste à choisir le contenu des gènes qui facilite la description du problème et respecte ses contraintes. Il utilise principalement deux types de codage : le codage binaire, le codage réel :

3.3.4.1 Codage binaire

Dans ce type de codage le chromosome représente simplement une suite de 0 et de 1. Le codage binaire est également indépendant des opérateurs génétiques (croisement et mutation) du moment où ces derniers ne nécessitent aucune spécification. En effet, toute manipulation d'un chromosome donne naissance à un nouveau chromosome valide. Cependant, ce type de codage n'est pas toujours bon, il est parfois très difficile ou très lourd de coder des solutions de cette manière. En outre, dans certain cas la taille mémoire requise peut devenir prohibitive. [23]

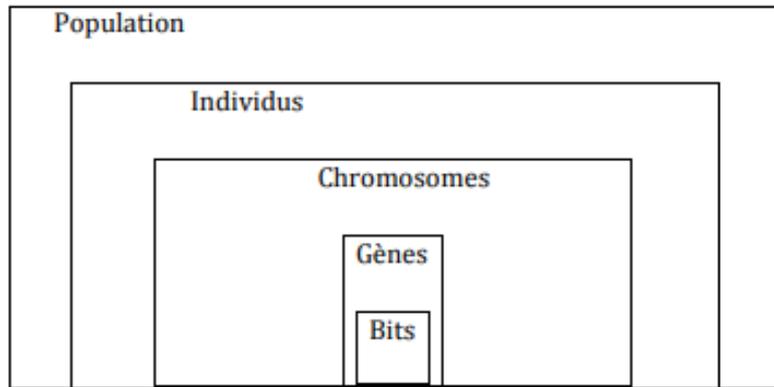


Figure 3.6 Les cinq niveaux d'organisation d'un algorithme génétique [22]

Il existe cependant [22] des limites à ce type de codage parmi eux :

- Ce codage est souvent peu naturel par rapport à un problème donné.
- Deux éléments voisins en termes de distance de Hamming ne codent pas nécessairement deux éléments proches dans l'espace de recherche.
- Pour des problèmes d'optimisation dans des espaces de grande dimension le codage binaire peut rapidement devenir mauvais.

Exemple :

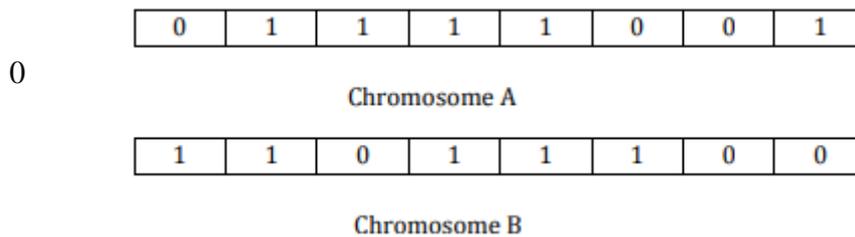


Figure 3.7 Codage binaire [22]

3.3.4.2 Codage réel

Dans ce type de codage, chaque chromosome est un vecteur dont les composantes sont les paramètres du processus d'optimisation. Par exemple, si on recherche l'optimum d'une fonction de n variables $f(x_1, x_2, \dots, x_{a-1}, x_a)$, on peut utiliser tout simplement un chromosome contenant les n variables : $f(x_1, x_2, \dots, x_{a-1}, x_a)$ avec ce type de codage, la procédure d'évaluation des chromosomes est plus rapide vu l'absence des étapes de codage et de transcoding (du réel vers le binaire et vice-versa) [23]

Exemple :

gène 1	gène 2	gène 3
10010011	11101011	00011010
$x_1 = 3,256$	$x_2 = 0,658$	$x_3 = 10,26$

Figure3.8 Codage réel [22]

3.3.5 Application d'un algorithme génétique au CNN

Pour optimiser les performances, les CNN nécessitent un grand nombre de paramètres afin de fonctionner de manière optimale, ce qui peut entraîner de longues périodes d'entraînement pour les réseaux et nécessite souvent une compréhension approfondie de l'application et des données pour une utilisation efficace. La liste des paramètres à prendre en compte pour construire un CNN comprend: le nombre de couches du réseau, les fonctions d'activation utilisées dans le réseau, la taille des opérations sur les couches (pour la convolution et le pooling / sous-échantillonnage), le nombre de couches utilisé dans chaque convolution, les paramètres d'entraînement de l'optimiseur, etc. Il en résulte un nombre infini de structures de réseau possibles, et il est souvent trop fastidieux d'obtenir de bons résultats pour les problèmes de classification d'images en raison de la complexité de la construction de réseaux d'apprentissage en profondeur. Pour illustrer ce point, nous donnons comme exemple, la structure de réseau utilisée dans [20] est illustrée à la figure 3.9 Cette exemple d'application va nous permettre de mieux comprendre et présenter notre modèle. La figure 3.10 comprend trois couches de convolution, deux couches de regroupement / sous-échantillonnage et deux couches d'activation. Dans ce cas, les paramètres à déterminer pour construire ce réseau comprennent les tailles de couche de convolution, les tailles de couche de regroupement, les types de fonction d'activation, la profondeur du réseau, le type de fonction d'optimisation et le taux d'apprentissage de la fonction d'optimisation. Le but de la méthode proposée est d'utiliser une AG pour obtenir un CNN extrêmement précis en optimisant les paramètres du réseau, ce qui contribuera à éliminer certaines des difficultés liées à la construction d'un réseau précis. En utilisant le réseau dans [20]comme base, il sera possible de valider les performances des réseaux nouvellement construits par rapport à la méthode de l'état de la technique.

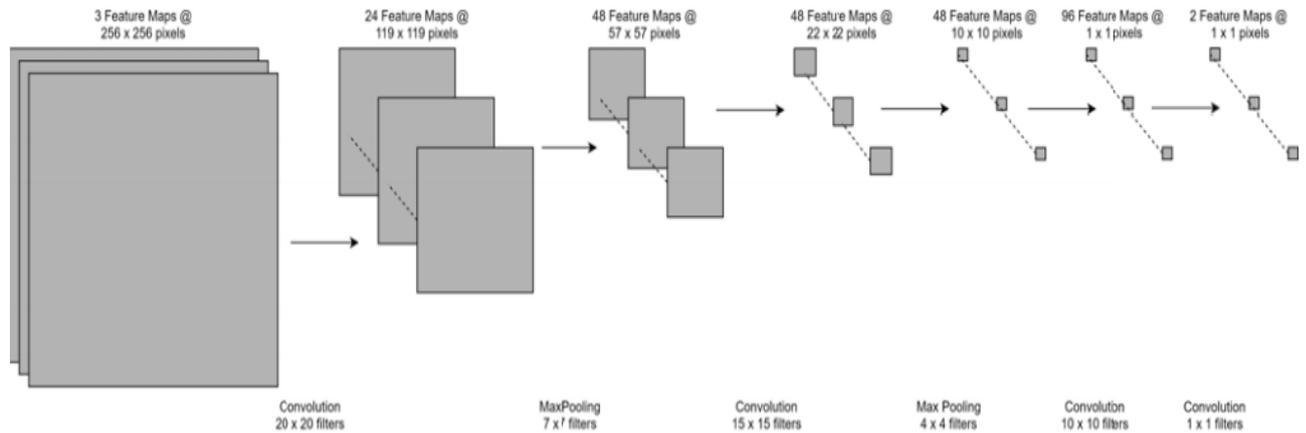


Figure 3.9 Exemple de structure de réseau utilisée [1].

La fonction d'activation ReLU est utilisée dans toutes les couches de ce réseau[20]

3.4 Approche d'optimisation du CNN par AG pour la reconnaissance de visage

En utilisant les systèmes conventionnels pour la reconnaissance de visage, la recherche est difficile. Identifier et/ou vérifier un visage dans une image est un problème extrêmement difficile. De nombreux éléments doivent être pris en compte à propos du visage en question, car il peut sembler très différent des données d'entraînement. La pose, l'occlusion, les expressions, l'éclairage et le vieillissement sont parmi les problèmes les plus courants qui empêchent une reconnaissance réussie. De plus, le maquillage ou la chirurgie esthétique peut modifier considérablement ses propriétés. Cela nécessite une grande flexibilité dans la définition du visage, impossible à obtenir dans des réseaux de neurones simples, à couche unique. Ce type de problème nécessite un grand nombre de couches, pour stocker toutes les informations sur le visage. Les réseaux d'apprentissage en profondeur multicouches peuvent être utilisés ici pour représenter des visages dans l'image (dans la détection de visage par exemple).

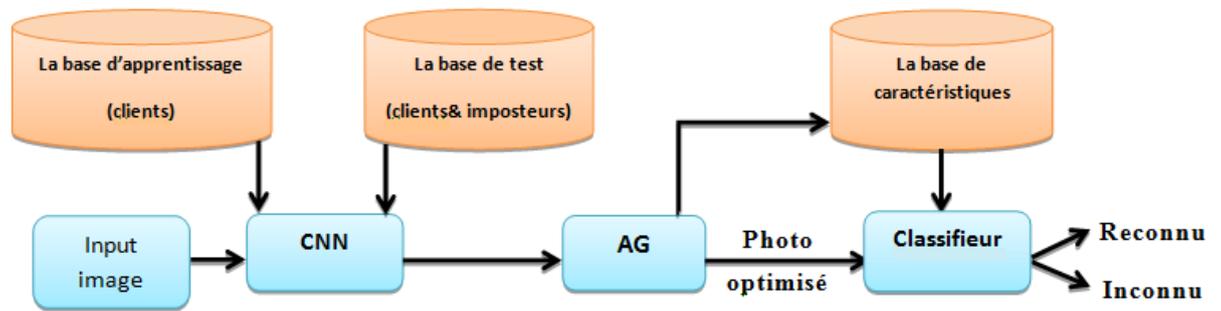


Figure 3.10 Schéma de principe du modèle CNNGA étudié

Les premières couches de notre modèle se basent sur les filtres conventionnels, le pooling et la couche ReLu que nous avons définis en début de ce chapitre 3. Le réseau d'apprentissage en profondeur assure la réduction des caractéristiques extraites par les premières couches du réseau. Dans ce qui suit nous présentons un aperçu sur les méthodes de réduction.

3.4.1 Réduction de données

Les réseaux d'apprentissage en profondeur peuvent recréer des approximations proches de leurs objets d'origine à partir d'un formulaire compressé. Ignorant le coût du réseau lui-même, qui représente un coût unique substantiel, cet algorithme effectuera une compression sur n'importe quel objet qui lui est attribué. Si l'on suppose une matrice parfaitement entraînée, de sorte que chaque sortie soit recréée aussi parfaitement que possible, il est facile d'en déduire que l'algorithme devrait fonctionner presque au niveau d'entropie optimal, comme indiqué dans le théorème de Shannon. S'il n'y a qu'un seul objet, il sera parfaitement recréé sans données, car l'algorithme peut simplement enregistrer toutes les données dans le terme de biais. S'il existe deux objets, les deux algorithmes ont besoin de deux lignes non nulles, ce qui nécessite le stockage d'un bit de données. Il peut, s'il est parfaitement entraîné, enregistrer un 0 dans le cas du premier objet et un 1 dans le cas du second. La ligne de biais contient le premier objet, tandis que la ligne inverse contient le deuxième objet moins le premier. La reconstruction est donc parfaite. Repousser les limites au-delà du théorème de Shannon dans cette expérience de pensée aboutit clairement à une approximation avec pertes, où le résultat optimal est le point médian entre les deux objets, de sorte que les deux objets ont perdu la moitié de leurs données. Si chaque exemplaire d'entraînement reçoit un poids égal pendant l'entraînement et si toutes les données à stocker sont parfaitement entraînées dans l'algorithme, le résultat doit être que chacune d'entre elles se

voit attribuer un pourcentage égal de l'entropie nécessaire pour créer une reconstruction parfaite. On peut tester si cet ensemble de formation peut être réalisé pendant la formation du réseau. Choisir une valeur minimale, v , où v est le plafond de l'entropie de l'ensemble de tests, car le nombre de valeurs dans l'état intermédiaire donné à une matrice parfaitement entraînée entraînera la recréation parfaite de toutes les données de test, avec un coût de stockage par entité de formation égale au plafond de la quantité minimale théorique de données à stocker.

3.4.2 Data Mining

Un réseau d'apprentissage en profondeur est conçu pour créer un sous-ensemble de fonctionnalités de l'objet d'origine. Comme la formation d'un réseau d'apprentissage en profondeur repose sur le maintien d'un maximum d'informations sur les différences entre les modèles d'objets de ces catégories, on peut supposer que la représentation interne des objets conserve autant que possible la variance entre eux. Ainsi, ce formulaire contient toute la variance nécessaire pour effectuer des tâches liées au tri de ces objets, telles que l'analyse des composants primaires et les techniques d'exploration de données. Ces modèles ont nettement moins de dimensionnalité que les images d'entrée, ce qui permet un partitionnement plus rapide et plus précis de l'espace de données.

Les algorithmes Modern Data Mining utilisent des outils simples de réduction de dimensionnalité pour réduire la taille de l'espace de données à rechercher tout en maximisant la distance entre les points. En réduisant la quantité de données à comparer, ces algorithmes réduisent le temps d'exécution des comparaisons et réduisent les risques de confusion, débouchant sur de meilleures solutions. En extrayant automatiquement les caractéristiques qui sont importantes pour recréer l'image, un réseau Deep Learning peut effectuer à peu près la même tâche, en réduisant la dimensionnalité globale des données, sans réduire l'importance des données pour la classification de l'image.[24]

3.5 Classification par le Support Vector Machine

Le classifieur SVM (Support Vector Machines) est un algorithme ou une méthode d'apprentissage statistique [25][26] qui a été développée par V.Vapnik. . Son principe général est le découpage du problème en deux sous-problèmes

1. Transformation non-linéaire des entrées.
2. Choix d'une séparation linéaire 'optimale' [27]

Initialement, SVM a été conçu pour traiter le problème de la classification binaire supervisé et très vite les SVM se sont étendus pour des problèmes de natures diverses. Parmi leurs avantages, nous citerons la possibilité de traiter des problèmes dont les exemples sont non

homogènes, la convergence vers une solution unique (pas de problème de minimum local comme le cas des réseaux de neurones). De nombreux travaux ont été achevés dans le domaine de la reconnaissance de la parole et image en appliquant les SVM et plusieurs noyaux sont développés pour cet objectif[18].Elles reposent sur une théorie mathématique solide à l'inverse des méthodes de réseaux de neurones. Elles ont été développées au sens inverse du développement des réseaux de neurones. Les SVMs sont dans leur origine utilisées pour la classification binaire et la régression. Aujourd'hui, elles sont utilisées dans différents domaines de recherche et d'ingénierie tel que le diagnostic médical, le marketing, la biologie, la reconnaissance de caractères manuscrits et de visages humains[25]

L'objectif de SVM est de trouver l'hyperplan optimal $f(\mathbf{w}, \mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$ pour séparer deux classes dans un ensemble de base de données donné avec $X \dots$ [26]

3.5.1 Principe du SVM binaire

Le cas le plus simple est celui où les données d'entraînement viennent uniquement de deux classes différentes (+1 ou -1), on parle alors de classification binaire. L'idée des SVMs est de rechercher un hyperplan (droite dans le cas de deux dimensions) qui sépare le mieux ces deux classes(figure3.11)

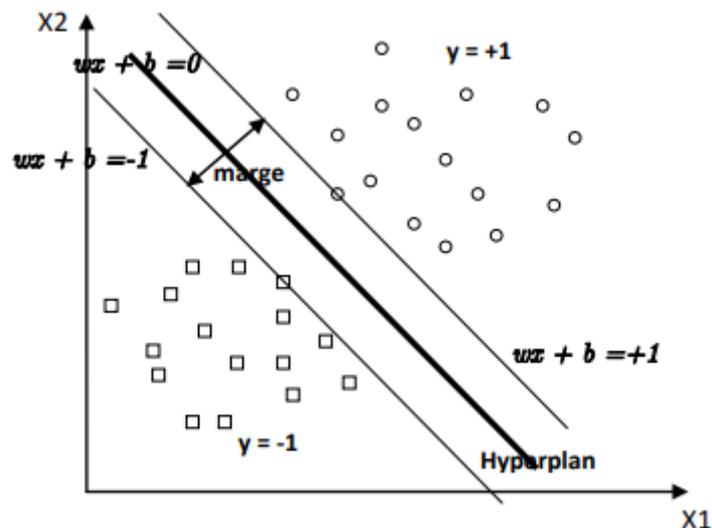


Figure3.11Principe du SVM binaire

Si un tel hyperplan existe, c'est-à-dire si les données sont linéairement séparables, on parle d'une machine à vecteur support à marge dure.

3.5.2 Principe du SVM multi-classe

SVM est utilisé à l'origine pour résoudre des problèmes avec seulement deux classes et ne peut pas être appliqué à la classification multi-classes [29]

Les machines à vecteur support sont dans leur origine binaire. Cependant, les problèmes du monde réel sont dans la plupart des cas multi-classe, l'exemple le plus simple est la reconnaissance des caractères optiques (OCR). Dans tels cas, on ne cherche pas à affecter un nouvel exemple à l'une de deux classes mais à l'une parmi plusieurs, que la décision n'est plus binaire et un seul hyperplan ne suffit plus. Les méthodes des machines à vecteur support multi-classe, réduisent le problème multi-classe à une composition de plusieurs hyperplans classes permettant de tracer les frontières de décision entre les différentes classes. Ces méthodes décomposent l'ensemble d'exemples en plusieurs sous-ensembles représentant chacun un problème de classification binaire. Pour chaque problème un hyperplan de séparation est déterminé par la méthode SVM binaire. On construit lors de la classification une hiérarchie des hyperplans binaires qui est parcourue de la racine jusqu'à une feuille pour décider de la classe d'un nouvel exemple. On trouve dans la littérature plusieurs méthodes de décomposition: [25]

3.5.2.1 Méthodes de décomposition

A. Approches "un contre tous"

[28]L'approche "un contre tous" est la plus simple et la plus ancienne des méthodes de décomposition. Elle consiste à utiliser un classifieur binaire (à valeurs réelles) par catégorie. Le \mathcal{K} -ième classifieur est destiné à distinguer la catégorie d'indice \mathcal{K} de toutes les autres. Pour affecter un exemple, on le présente donc à Q classifieurs, et la décision s'obtient en application du principe "winner-takes-all" : l'étiquette retenue est celle associée au classifieur ayant renvoyé la valeur la plus élevée. On cite ordinairement comme plus anciens travaux évoquant l'emploi de cette stratégie avec des SVM], les auteurs soutiennent la thèse selon laquelle cette approche, aussi simple soit-elle, lorsqu'elle est mise en œuvre avec des SVM correctement paramétrées, obtient des performances qui ne sont pas significativement inférieures à celles des autres méthodes de décomposition et des SVM multi-classes actuelles. Il convient cependant de souligner qu'elle implique d'effectuer des apprentissages aux répartitions entre catégories très déséquilibrées, ce qui soulève souvent des difficultés pratiques.

B. Approches "un contre un"

Une autre méthode de décomposition très naturelle est la méthode "un contre un" elle consiste à utiliser un classifieur par couple de catégories. Le classifieur indicé par le couple $(\mathcal{K}, \mathfrak{u})$ (avec $1 \leq \mathcal{K} < \mathfrak{u} \leq Q$), est destiné à distinguer la catégorie d'indice k de celle d'indice \mathfrak{u} . Pour affecter un exemple, on le présente donc à $\binom{Q}{2}$ classifieurs, et la décision s'obtient habituellement en effectuant un vote majoritaire

("max-wins-voting"). La voix de chaque classifieur peut-être pondérée par une fonction de la valeur de la sortie calculée. L'auteur y dérive l'approche un contre un dans le cadre de l'estimation du classifieur de Bayes. Sous l'hypothèse que la frontière séparant

une catégorie d'une autre peut être moins complexe que celle séparant cette même catégorie de toutes les autres, il y voit un moyen d'obtenir des estimateurs présentant un biais plus faible qu'avec l'approche un contre tous. Naturellement, le prix à payer est un possible accroissement de la variance de ces estimateurs, compte tenu du fait que les bases d'apprentissage de chacun des classifieurs sont plus petites que l'échantillon initial.

C. Utilisation de codes correcteurs d'erreurs

L'introduction de l'emploi de codes correcteurs d'erreurs (**ECOC**) en apprentissage est ordinairement attribuée à Duda et ses co-auteurs.

Le principe consiste à représenter les catégories par des mots binaires de même taille, les "mots codes". En notant N la taille de ces mots, on obtient ainsi une matrice $M = (m_{kl})$ de $M_{Q,N}(\{0, 1\})$ (ou $M_{Q,N}(\{-1, 1\})$) dont les lignes $(M_{k\cdot})_{1 \leq k \leq Q}$ sont les mots codes et dont les colonnes $(M_{\cdot t})_{1 \leq t \leq N}$ spécifient N dichotomies (partitions en deux super-catégories de l'ensemble des catégories). Chaque dichotomie est calculée par un classifieur. Ainsi, chaque exemple est associé à un vecteur de $\{0, 1\}^N$ ou $\{-1, 1\}^N$ (suivant l'espace auquel appartient M). Il est affecté à la catégorie correspondant au mot code (vecteur $M_{k\cdot}$) le plus proche de ce vecteur au sens de la distance de Hamming. Naturellement, cette approche est d'autant plus efficace que les mots codes sont plus distants les uns des autres (toujours au sens de la distance de Hamming). C'est là qu'interviennent les **ECOC**. En fait, il convient de maximiser non seulement la séparation des lignes de la matrice, mais aussi la séparation de ses colonnes. Si cette dernière séparation n'est pas assurée, alors les classifieurs associés à des colonnes proches risquent d'effectuer des erreurs similaires (corrélées). Or, l'utilisation des **ECOC** n'est efficace que si les erreurs effectuées sur les différents bits sont relativement peu corrélées, de manière que la probabilité d'observer des erreurs simultanées sur un grand nombre de bits soit faible. Leur conclusion est que la dernière solution donne les meilleures performances, même dans le cas des petits échantillons, ceci en dépit du fait qu'elle conduise à la construction de fonctions de décision relativement plus complexes. Le gain n'est pas fonction de la manière dont les mots codes sont associés aux différentes catégories. [28]

3.6 Conclusion

Dans ce chapitre nous avons décrit les démarches pour réaliser notre application. Nous avons commencé par sa description globale. Ensuite nous avons mis en relief la description des méthodes proposées pour la reconnaissance des visages. Une bonne partie du chapitre est consacrée à la partie d'optimisation du système par la sélection des caractéristiques en utilisant l'algorithme génétique et le CNN.

Le chapitre suivant sera consacré essentiellement à l'implémentation de ces méthodes et le test de la performance de notre application.



Chapitre IV

Chapitre IV Conception du système et résultat

4.1 Introduction

Dans le chapitre précédent nous avons présenté l'étude et la conception du système de reconnaissance de visage basé sur le CNN et l'algorithme génétique (AG). Nous consacrons ce chapitre 4 pour présenter l'implémentation de la conception de l'approche conçue utilisant l'apprentissage en profondeur optimal par l'algorithme génétique, ainsi que son application à la reconnaissance de visage et aussi la reconnaissance du genre. Nous présenterons également certains résultats sur des échantillons de trois bases de données de données. Des comparaisons de quelques expériences seront présentées afin de vérifier l'efficacité et l'optimalité de cette méthode. Pour valider notre travail nous avons testé notre approche sur des images frontales de notre base de données en milieu contrôlé CASIA-2DV4 (avec plusieurs variantes), Face V4 (visage frontale avec plusieurs variantes) et milieu réel LFW.

4.2 Présentation des outils de développement

4.2.1 Matériel

Le matériel utilisé est un PC personnel HP avec une capacité mémoire 4GB, et un processeur Intel® core™ i3 CPU M370 @ 2.40GHZ avec Windows 7 professionnel service pack 1 64 bit type système.

4.2.2 Logiciel

MATLAB pour « Matrix Laboratory », est une application qui a été conçue afin de fournir un environnement de calcul matriciel simple, efficace, interactif et portable, permettant la mise en œuvre des algorithmes développés dans le cadre des projets linpack et eispack. MATLAB est constitué d'un noyau relativement réduit, capable d'interpréter puis d'évaluer les expressions numériques matricielles qui lui sont adressées : — soit directement au clavier depuis une fenêtre de commande ; — soit sous forme de séquences d'expressions ou scripts enregistrées dans des fichiers-texte appelés m-files (ou fichiers .m) et exécutées depuis la fenêtre de commande ; — soit plus rarement sous forme de fichiers binaires appelés mex-files (ou fichiers .mex) générés à partir d'un compilateur C ou fortran. Ce noyau est complété par une bibliothèque de fonctions prédéfinies, très souvent sous forme de fichiers m-files, et regroupés en paquetages ou toolboxes. A côté des toolboxes requis locaux et matlab, il est possible d'ajouter des toolboxes spécifiques à tel ou tel problème mathématique, Optimisation Toolbox, Signal Processing Toolbox par exemple, ou encore des toolboxes créées par

l'utilisateur lui-même. Un système de chemin d'accès ou path permet de préciser la liste des répertoires dans lesquels MATLAB trouvera les différents fichiers m-files utilisés[1].

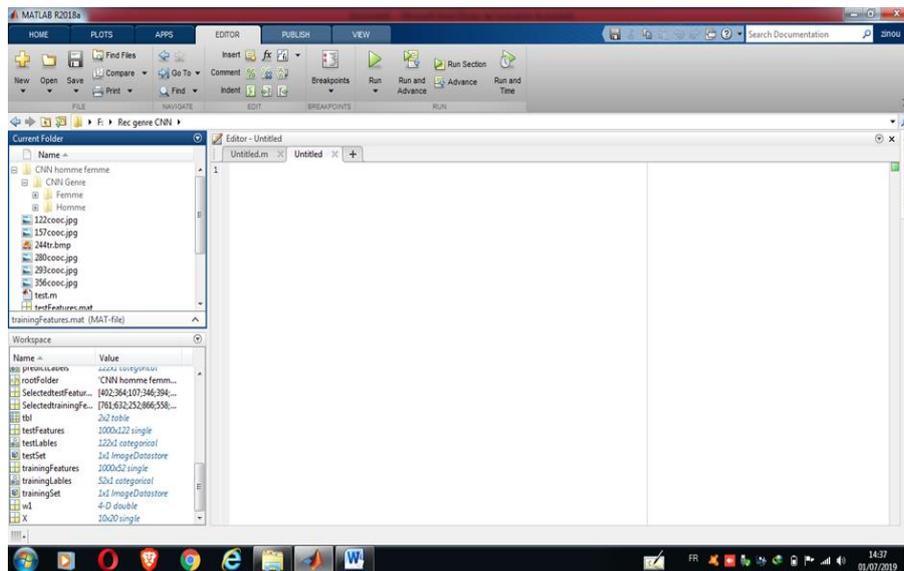


Figure 4.1 Exemple de fenêtre Matlab

4.3 Bases données des visages utilisées

4.3.1 Base de données CASIA 2DV4

4.3.1.1 Présentation de la base de données CASIA 2DV4

Pour évaluer notre système nous avons utilisé la base de données de visage **CASIA2DV4**. Elle a été développée par l'académie Chinoise de l'Institut des Sciences de l'automatisation (Chinese Academy Sciences Institut of Automation (**CASIA**)) et les données qu'elle regroupe ont été acquises avec un scanner **MINOLTA VIVID 910**, qui fournit non seulement des informations de **Texture** (image 2D), mais aussi des fichiers VRML. Correspondant à un **maillage triangulaire** d'environ 1000000 points (image 3D). C'est ainsi une base multimodale universelle.

Cette base contient 123 personnes et dont chacune est représentée par un ensemble de 37 ou 38 captures avec différentes poses sous différentes conditions d'éclairage et d'expression faciales ainsi que certains mouvements aléatoires. Et ce dans le but de couvrir toutes les

Chapitre IV Conception du système et résultat

variantes possibles. Dans tous les cas, un seul paramètre a été modifié entre deux captures. C'est l'un des principaux avantages de cette base de données. La base de données est divisée en trois ensembles : **Apprentissage**, **Évaluation** et **Test**.



Figure 4.2 Différentes captures de la base de données CASIA 2D [30]

4.3.1.2 Protocole de la BDD CASIA 2D

Dans notre travail, nous utilisons seulement 20 modèles à chaque individu. (1-5) scans avec variation de l'illumination **Figure 4.3**. (6-10) scans avec variation des expressions (rire, sourire, la colère, la surprise, œil fermé) (a) scans avec sous des variations d'éclairage.

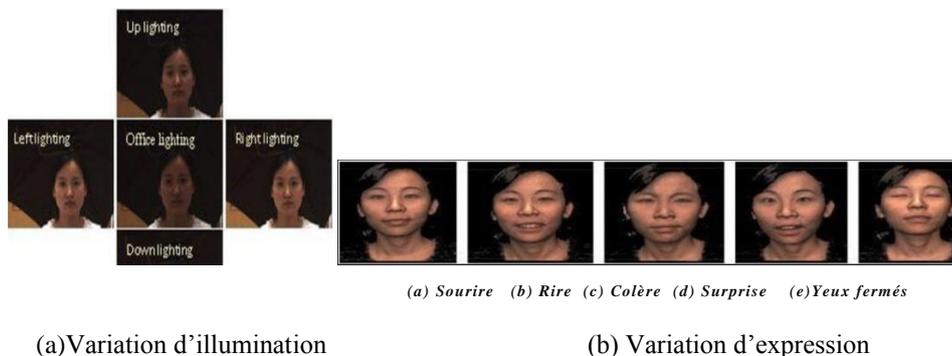


Figure 4.3 Quelques variantes de la base de données CASIA2DV4 [30]

4.3.2 Base de données LFW (Labeled Faces in the Wild)

La principale motivation de la base de données, qui est examinée plus en détail ci-dessous, est de fournir un grand ensemble d'images de visage relativement peu contraignantes. Par non contraint, nous entendons des visages qui montrent une large gamme de variations observées dans la vie quotidienne. Cela comprend les variations de pose, d'éclairage, d'expression, d'arrière-plan, de race, d'origine ethnique, d'âge, de sexe, de vêtements, de coiffures, de qualité de l'appareil photo, de saturation des couleurs, de mise au point et d'autres paramètres. **La figure 4.4** montre quelques exemples d'images de la base de données. La raison pour

Chapitre IV Conception du système et résultat

laquelle nous nous intéressons à la variation naturelle est que la BDD s'intéresse au problème des images dont nous n'avons aucun contrôle sur la composition. Nous considérons ce problème sans contrainte comme l'un des problèmes de reconnaissance de visage les plus généraux et les plus fondamentaux. Avant de passer aux détails de la base de données, nous présentons quelques statistiques sommaires et propriétés de la base de données.

La base de données contient 13 233 images de visage cibles. Certaines images contiennent plus d'un visage, mais c'est le visage qui contient le pixel central de l'image qui est considéré comme le visage qui définit l'image. Les visages autres que le visage cible doivent être ignorés comme «arrière-plan». Le nom de la personne représentée au centre de l'image est donné. Chaque personne se voit attribuer un nom unique; aucun nom ne doit donc correspondre à plus d'une personne; chaque personne ne doit apparaître que sous un seul nom (sauf s'il existe des erreurs inconnues dans la base de données).

La base de données contient des images de 5749 individus différents. 1680 personnes ont deux ou plusieurs images dans la base de données. Les 4069 personnes restantes n'ont qu'une seule image dans la base de données.

Les images sont disponibles sous forme d'images JPEG 250 x 250 pixels. La plupart des images sont en couleur, bien que quelques-unes ne soient en niveaux de gris. Toutes les images sont le résultat de détections effectuées par le détecteur de visage Viola-Jones, mais ont été redimensionnées et rognées à une taille fixe. Après avoir utilisé le détecteur Viola-Jones sur une grande base de données d'images, les détections de faux positifs de visage ont été éliminées manuellement, ainsi que les images pour lesquelles le nom de l'individu n'a pas pu être identifié. Deux «vues» de la base de données sont définies, une pour le développement d'algorithmes et une pour les rapports sur les performances.



Figure 4.4 Exemples de la base de données LFW

Chapitre IV Conception du système et résultat

4.3.3 Base de données Face

C'est une base de données qui contient 140 images d'apprentissage pour 10 personnes et 50 images test.



Figure 4.5 Exemple d'images de la BDD Face

4.4 Système de reconnaissance de visage 3D basé sur CNN et AG

Le schéma bloc de notre optimisation par l'algorithme génétique du système de reconnaissance de visage basée sur le CNN est montré par la figure 4.6.

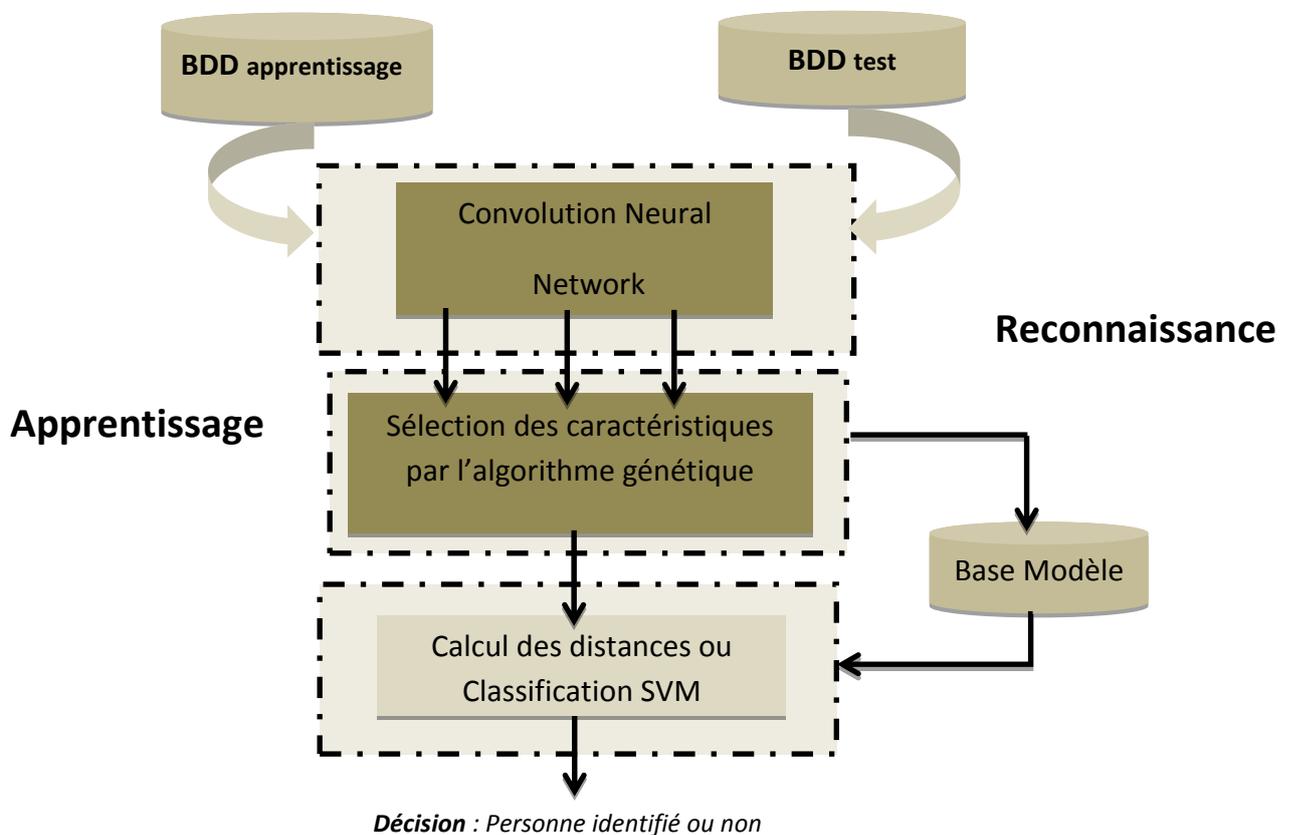


Figure 4.6 Schéma du système de reconnaissance de visage CNNAG

Chapitre IV Conception du système et résultat

Comme le montre la **figure 4.6**, le système de reconnaissance CNNAG est composé de trois modules essentiels :

1. Module d'extraction de caractéristiques et réduction par l'architecture CNN ;
2. Module d'optimisation des paramètres caractéristiques ;
3. Module de classification.

Ces trois modules sont détaillés dans ce qui suit.

4.5 Extraction des caractéristiques et réduction de données par CNN

La couche de convolution constitue la phase fondamentale d'un réseau de convergence qui s'occupe de la plupart des tâches de calcul lourdes. L'objectif principal de la couche de convolution est d'extraire des entités des données d'entrée qui sont une image. La convolution préserve la relation spatiale entre les pixels en apprenant les caractéristiques de l'image à l'aide de petits carrés d'image d'entrée. L'extraction de caractéristiques de l'image d'entrée est compliquée, dans ce cas on utilise un ensemble de neurones pouvant être appris. Cela produit une carte de caractéristiques ou une carte d'activation dans l'image de sortie, puis les cartes de caractéristiques sont transmises en tant que données d'entrée à la couche convolutive suivante.

Un réseau de neurones à convolution, capable d'apprendre les caractéristiques locales à partir de données d'entrée, est utilisé pour discriminer les images faciales. Un classifieur CNN typique comprend un CNN avec une alternance de couches de convolution et de sous-échantillonnage pour l'extraction de caractéristiques et un réseau de neurones dans la dernière couche pour la classification. L'architecture du classifieur CNN considérée dans notre travail est représentée à la **figure 4.7**. La taille du modèle prise en compte dans les couches de convolution (C_1, C_2) et les couches de sous-échantillonnage (S_1, S_2) est respectivement de 5×5 et 2×2 . Les images faciales sont sous-échantillonnées pour obtenir des images de taille 28×32 et données en entrée dans CNN. Le nombre de cartes de caractéristiques prises en compte dans le premier ensemble de cartes de caractéristiques (F_1) et le troisième ensemble de cartes de caractéristiques (F_3) est respectivement de 6 et 12. Le deuxième ensemble de cartes de caractéristiques (F_2) et le quatrième ensemble de cartes de caractéristiques (F_4) auront respectivement le même nombre de cartes de caractéristiques que F_1 et F_3 . La sortie du CNN est traversée dans une rangée majeure afin d'obtenir un vecteur de colonne (I) de dimension 240×1 utilisé par le réseau de neurones pour classer les images faciales en entrée dans l'une

Chapitre IV Conception du système et résultat

des 30 classes de sortie (O).[32] Les deux dernières couches du classifieur CNN, à savoir F_4 et le réseau de neurones sont entièrement connectés. Le classifieur CNN est formé en utilisant un algorithme de rétro-propagation en mode batch, apprendre les masques de convolution utilisés dans C_1 , C_2 et les poids de connexion entre les deux dernières couches du classificateur. Le résultat des filtres de convolution de la couche C_1 est présenté à la **figure 4.8**.

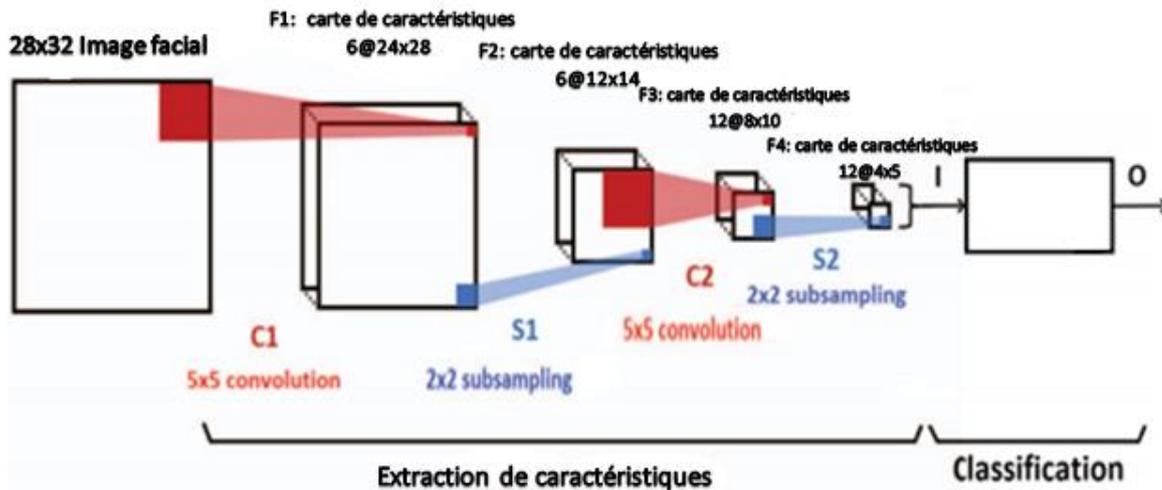


Figure 4.7 Architecture du CNN μ

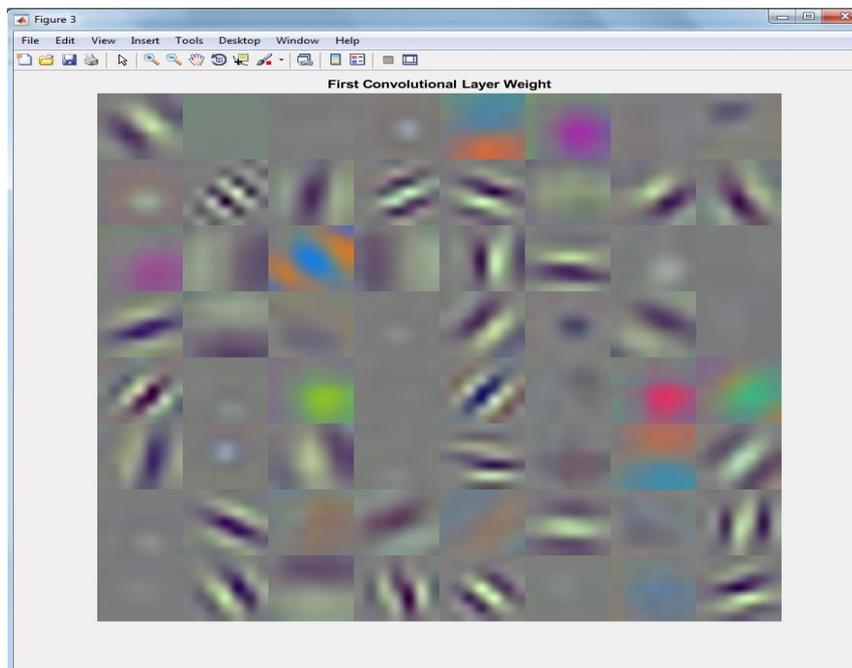


Figure : Résultat des filtres appris par la première couche convolutive C

4.6 Utilisation de l'algorithme génétique

4.6.1 Sélection des caractéristiques par AG

Après avoir inséré l'image dans le CNN elle sort en forme de vecteur, puis celui-ci est soumis à l'algorithme génétique pour la sélection des caractéristiques optimales.

Comme le montre la **figure 4.8**, l'algorithme génétique utilise son fitness (définir la fonction fitness utilisée) et les fonctions cross-over et mutation afin d'obtenir la sélection des meilleurs caractéristiques pour un vecteur optimisé.

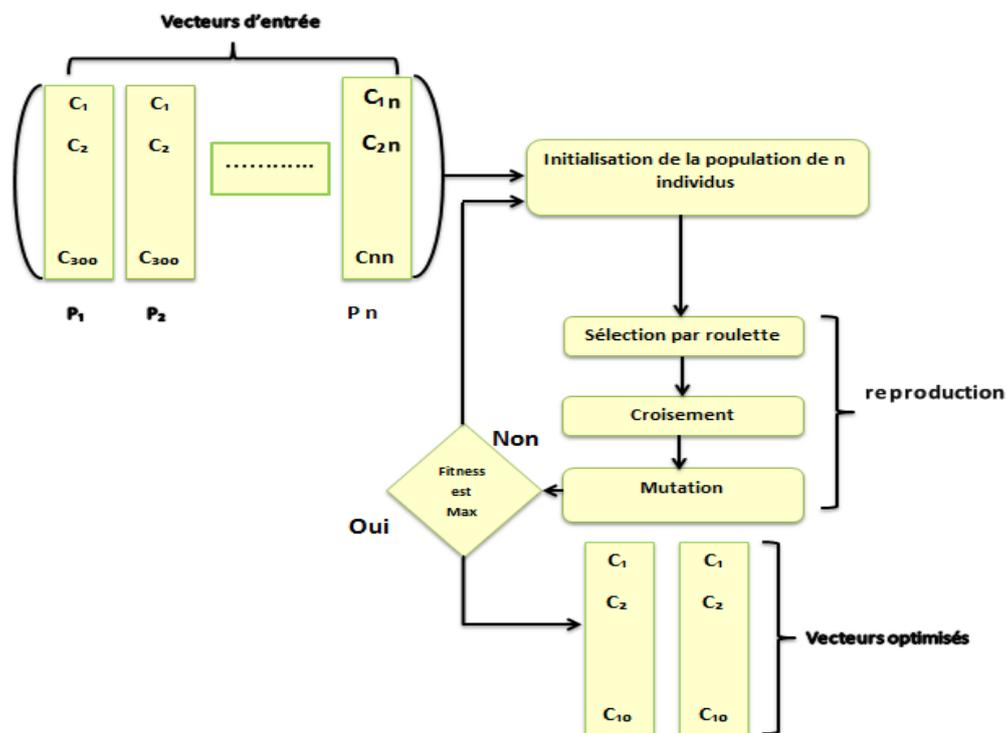


Figure 4.8 Sélection des paramètres caractéristiques par l'AG[7]

4.6.2 Fonction de performance (fitness)

Dans les algorithmes génétiques, chaque solution est généralement représentée par une chaîne de nombres binaires, appelée chromosome. Nous devons tester ces solutions et trouver le meilleur ensemble de solutions pour résoudre un problème donné. Chaque solution doit donc se voir attribuer une note indiquant son degré de satisfaction par rapport aux spécifications globales de la solution souhaitée. Ce score est généré en appliquant la fonction de condition physique au test ou les résultats obtenus à partir de la solution testée.

Chaque problème a sa propre fonction de fitness. La fonction de fitness à utiliser dépend du problème posé. Venir avec une fonction de fitness pour le problème donné est la partie la plus

Chapitre IV Conception du système et résultat

difficile quand il s'agit de formuler un problème en utilisant des algorithmes génétiques. Il n'y a pas de règle absolue selon laquelle une fonction particulière doit être utilisée dans un problème particulier. Cependant, certaines données ont été adoptées par des experts en données concernant certains types de problèmes.

En règle générale, pour les tâches de classification où l'apprentissage supervisé est utilisé, les mesures d'erreur telles que la distance euclidienne et la distance de Manhattan ont été largement utilisées en tant que fonction de mise en forme. Pour les problèmes d'optimisation, les fonctions de base telles que la somme d'un ensemble de paramètres calculés liés au domaine du problème peuvent être utilisées comme fonction de fitness. Passons en revue quelques exemples de problèmes et leurs fonctions de mise en forme associées.

L'évaluation de l'aptitude consiste à vérifier la valeur de la solution de la fonction objective soumise à des contraintes en utilisant une procédure de décodage. En général, la fonction objective fournit le mécanisme d'évaluation. Cependant, sa gamme de valeurs varie d'un problème à l'autre. Pour maintenir l'uniformité dans divers domaines problématiques, nous pouvons utiliser la fonction d'aptitude pour normaliser la fonction objective dans une plage de 0 à 1. La valeur normalisée de la fonction objective est l'aptitude de l'individu, et le mécanisme de sélection l'utilise pour évaluer les individus de la population. Lorsque les GA sont utilisés pour effectuer des recherches, la population subit une évolution avec fitness et forme une nouvelle population. A cette époque, à chaque génération, des solutions relativement bonnes sont reproduites et des solutions relativement mauvaises sont éliminées, de sorte que la progéniture composée des bonnes solutions est reproduite. Pour distinguer les solutions, une fonction d'évaluation (également appelée fonction de fitness) joue un rôle important dans l'environnement, et des mécanismes de mise à l'échelle doivent également être appliqués dans une fonction objective pour être une fonction de fitness.

L'algorithme suivant permet de calculer la fonction du fitness en utilisant la fonction matlab "GA_feature_selector".

```
[Selected]=GA_feature_selector(feat_num,X,y)
```

Où feat_num : nombre de caractéristiques, X: caractéristiques et y : étiquettes

La **figure 4.9** ci-dessous nous montre l'évaluation de la fonction fitness en fonction du nombre de génération.

Chapitre IV Conception du système et résultat

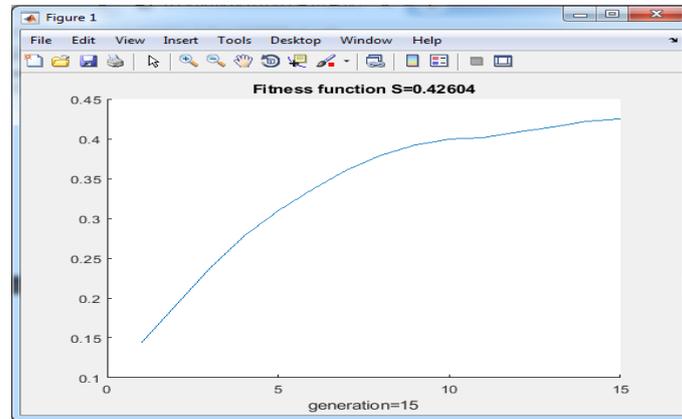


Figure 4.9 Courbe d'évaluation de la fonction du fitness

4.7 Classification

4.7.1 Classification par SVM

4.7.1.1 SVM (Apprentissage)

La classification par **SVM** est une méthode de classification qui a fait ses preuves dans la mesure de similarité, aussi nous tentons de l'utiliser dans le cadre de notre travail et tester ses performances, nous avons utilisé le **LS-SVMlab** qui a chaque toolbox contient les noms des algorithmes correspondants. Les noms de fonctions avec l'extension « **lssvm** » sont spécifiques des méthodes **LS-SVM**.

Pour l'apprentissage de **SVM** on utilise la fonction **trainlssvm** contenue dans **LS-SVMlab** qui a pour paramètres la matrice d'apprentissage, le groupe de chaque image contenue dans la matrice, les deux paramètres **Gamma**, et **Sigma**, le type de **SVM** classification ou régression, et qui a pour sortie les deux paramètres **alpha** et **béta**.

Pour représenter le résultat de l'apprentissage on utilise la fonction **plotlssvm** qui représente les deux classes séparé par l'hyperplan calculé.

4.7.1.2 SVM (Vérification)

Pour la vérification (classification) des images tests clients ou imposteurs on utilise la fonction **simlssvm()** qui prend pour paramètres tous les paramètres pris pour l'apprentissage plus la matrice des images qu'on veut classer. La sortie de cette fonction est un vecteur du nombre d'images à classer chaque élément 'i' de ce vecteur est la classe de l'image 'i' de la matrice.

Pour calculer le taux d'erreur dans la classification on utilise la fonction **missclass()** en sortie de cette fonction on trouve trois variables : i) Taux d'erreur du système ; ii) Nombre de fausse classification ; iii) Rang des images mal classifiées.

4.7.2 Classification par mesure de similarité

Lorsqu'on souhaite comparer deux vecteurs de caractéristiques réduit issus du module de réduction de données d'un système biométrique, on peut soit effectuer une mesure de similarité (ressemblance), soit une mesure de distance (divergence).

Bien que la distance euclidienne soit optimale dans la théorie, les diverses expériences ont trouvé que la distance euclidienne est surpassée par d'autres distances. Une d'elles est la **corrélation normalisée** (normalized correlation) qui est définie par :

$$S(A, B) = \frac{A^T B}{\|A\| \|B\|}$$

Cette fonction calcule simplement le cosinus de l'angle entre les deux vecteurs caractéristiques A et B. Une valeur élevée de corrélation normalisée à une bonne similarité entre les deux vecteurs.

4.8 Etude expérimentale et Résultats

1. Expérience 01 Reconnaissance du genre sans optimisation (CNN uniquement)

Dans cette expérience de classification notre base d'apprentissage est un échantillon représenté par différentes images de visage du genre homme et femme. La **figure 4.10** montre quelques images de la base de données d'apprentissage. Cette dernière contient un nombre d'image 182 de 89 hommes et 93 femmes différents.



Figure 4.10 Échantillon de 8 images d'hommes et de femmes pour l'apprentissage

Afin de tester notre système de reconnaissance nous avons respecté le protocole présenté dans le **tableau 4.1** Les personnes sont séparées en deux classes homme et femme. La base de données est partagée en une BDD d'apprentissage contenant que des clients et une

BDD de test contenant des clients et des imposteurs. Nous avons au total 177 images comme images d'apprentissage dans la base de données.

La BDD d'apprentissage contient 6 personnes où chaque personne possède presque 30 images, et une BDD de test qui contient 6 clients et 2 imposteurs.

Tableau 4.1 Protocole utilisé dans la reconnaissance du genre (sans et avec optimisation)

Tableau 4.1 Protocole utilisé dans la reconnaissance du genre (sans et avec optimisation)

BDD	Femme	Homme
Apprentissage	89 images	93 images
Test	3 images	3 images

Un exemple d'affichage d'images d'apprentissage Homme/Femme est présenté par la figure 4.11.



Figure 4.11 Affichage d'un échantillon pour l'apprentissage de la BDD (Homme/Femme)

La figure 4.12 nous montre le principe fonctionnel de la phase d'extraction de caractéristiques du visage dans les deux phases d'apprentissage et du test par le CNN avec les résultats obtenus par notre première expérience.

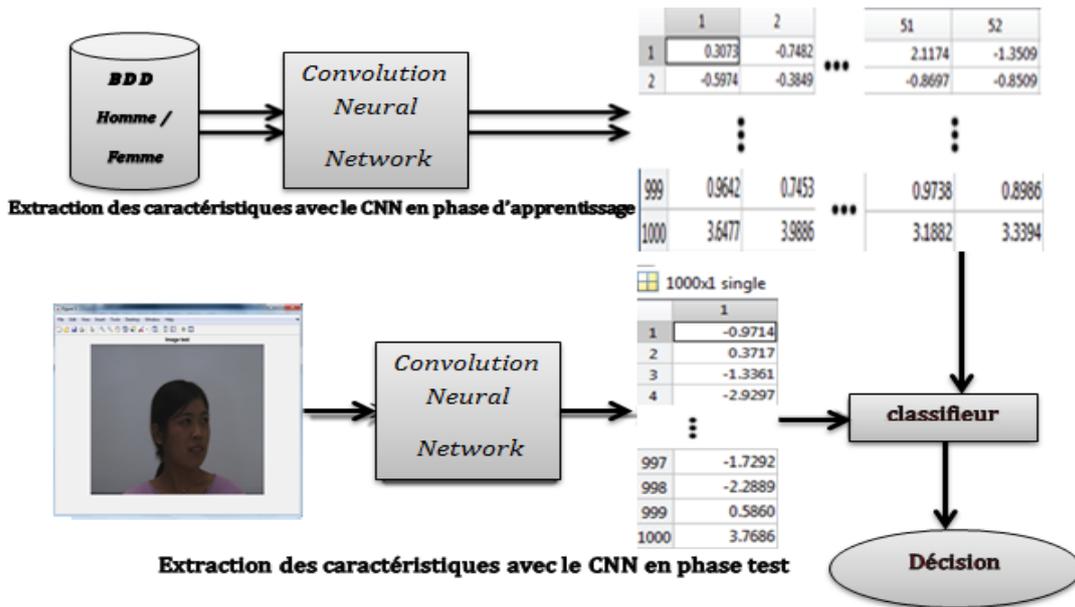


Figure 4.12 Système de reconnaissance du genre basé sur le CNN (BDD CASIA2DV4)

Chaque fois que nous voulons reconnaître une image test, nous la soumettons simplement au système proposé qui va connaître sa classe d'une façon automatique. Nous procédons comme suit:

- Extraction de caractéristiques de l'image d'entrée après son insertion dans l'architecture CNN à la sortie nous récupérons un vecteur de caractéristiques réduit comme le montre la phase de test de la figure 4.12.
- Calcul du temps total du processus de reconnaissance d'une image test de visage

Le résultat concernant le temps calculé obtenu est présenté par la ci-dessous.

```

Command Window
New to MATLAB? See resources for Getting Started.
>> Untitled
ans =
    'Image test chargée appartient à Femme classe'
Elapsed time is 14.986921 seconds.
/;> >>
    
```

Figure 4.13 Temps de reconnaissance du genre pour une image test avec le CNN

2. Expérience 02 Reconnaissance du genre avec optimisation (CNNAG)

Nous gardons les mêmes bases de données (images d'apprentissage et test) ainsi que le choix des paramètres du CNN que dans l'expérience. Nous ajoutons simplement un bloc

d'optimisation par algorithme génétique qui va nous permettre la sélection des meilleurs paramètres caractéristiques. Le principe de l'approche proposée est illustré sur la **figure 4.14**.

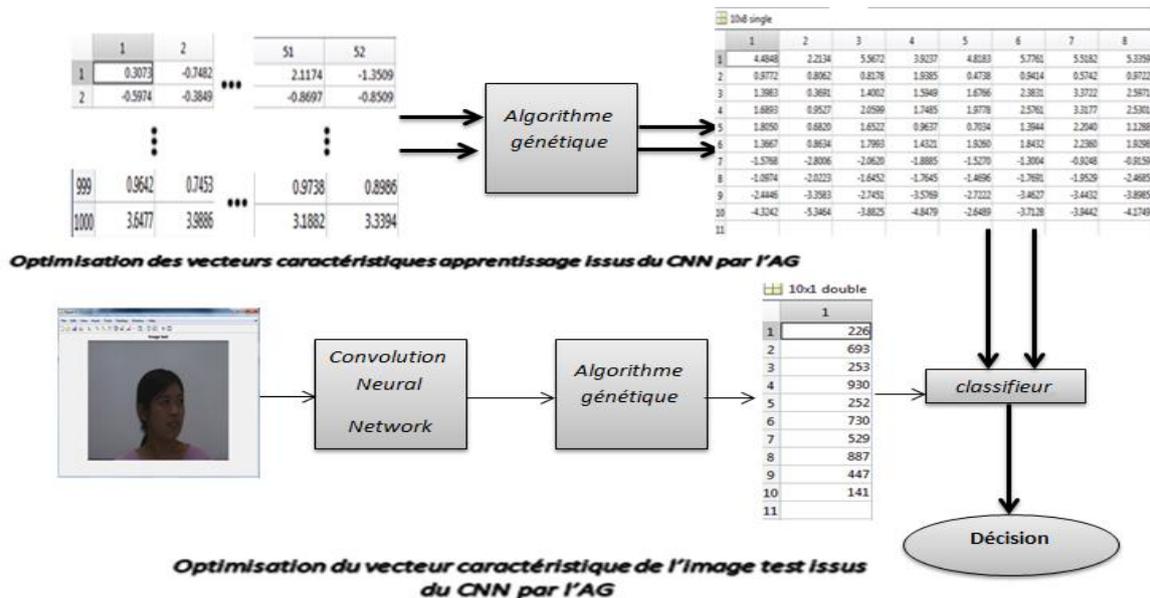


Figure 4.14 Optimisation du système de reconnaissance du genre basé sur CNNAG (BDD CASIA2DV4)

Le temps nécessaire pour classer l'image avec CNNAG est calculé et se trouve dans la **figure 4.15**.

```

Command Window
New to MATLAB? See resources for Getting Started.
484
829
778
680

ans =

'image test chargée appartient à Femme classe'

Elapsed time is 0.847437 seconds.
fx >>
    
```

Figure 4.15 Temps de reconnaissance du genre pour une image test avec le CNNAG

D'après les figures 4.13 et 4.15, nous pouvons dire que l'algorithme a considérablement amélioré le temps du processus de classification de visage. L'approche CNNAG proposée, non seulement classe convenablement les images de visage, mais en plus avec un temps de calcul qui a considérablement baissé de **14.986921 secondes** à **0.847437 secondes**

3. Expérience 03Reconnaissance de Visage sans optimisation (CNN)

La base de données d'apprentissage dans cette troisième expérience est la base Face. Nous disposons pour cette expérience de reconnaissance de visage de 154 images(14 images/personne) comme images d'apprentissage partagées par 11 personnes, et 14 images pour 14 personnes pour le test. Dans cette expérience, l'apprentissage est effectué sur la BDD Face et le test utilise une BDD où se trouvent différentes images des BDDs LFW et CASIA2DV4.

Tableau 4.2 Protocole utilisé pour la reconnaissance (sans et avec optimisation)

BDD	Clients	Imposteurs
Apprentissage	181 (14images/personne)	0 image
Test	11 images (1image/personne)	3 images(1image/personne)

L'affichage d'un échantillon de la base de données d'apprentissage est présenté par la **figure4.16**.

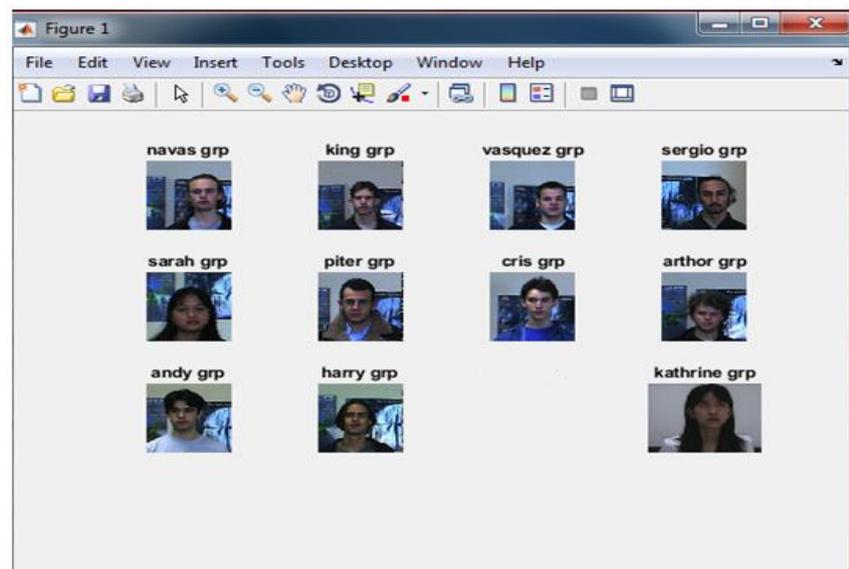


Figure4.16 Résultat d'affichage d'un échantillon d'image de la BDD Face

Dans cette expérience le système conçu reste le même, nous changeons simplement le protocole de la BDD où chaque personne dans ce cas représente une classe à lui seul.

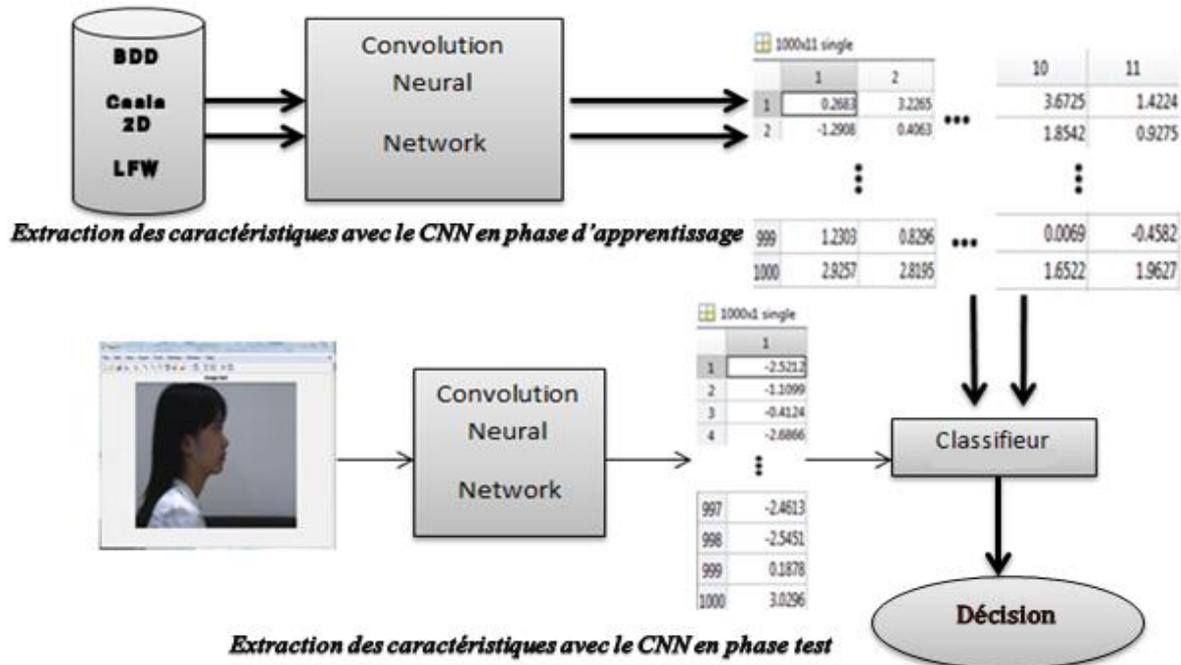


Figure 4.17 Système de reconnaissance de visage basé sur le CNN(BDD Face)

Toutes les images à l'entrée du système de reconnaissance de visage sont soumises au CNN et une classification a lieu entre le vecteur caractéristique de l'image test et les vecteurs caractéristiques de la BDD modèle.

Nous calculons le temps de calcul nécessaire pour classer l'image test est la **figure 4.18** ci-dessous nous indique les résultats concernant la reconnaissance et ce temps.

```

Command Window
New to MATLAB? See resources for Getting Started.
>> Untitled3
ans =
'image de test appartient à kathrine grp'
Elapsed time is 13.274166 seconds.
fx >>
    
```

Figure 4.18 Temps pour la reconnaissance basé sur CNN de l'image test (BDD Face)

Les résultats de la **figure 4.18** nous montre l'efficacité de la méthode CNN qui assure la reconnaissance test en la plaçant dans la classe de la personne katherin grp ce qui est juste. Et ceci avec un temps de reconnaissance **t = 13.274166 secondes**. Ce temps est appréciable, mais peut paraître lent pour des bases de données gigantesque. Dans l'expérience suivante, nous essayons d'optimiser notre système par l'algorithme génétique.

4. Expérience 04Reconnaissance de Visage avec optimisation (CNNAG)

Dans cette expérience, nous utilisons les mêmes bases de données d'apprentissage et test ainsi que les mêmes paramètres du CNN. Les caractéristiques dans ce cas sont sélectionnées par l'algorithme génétique. Nous voulons dans cette expérience réaliser l'optimisation par AG du système de reconnaissance basé sur le CNN ainsi conçu et implémenté dans l'expérience 03. La figure montre l'implémentation de ce système.

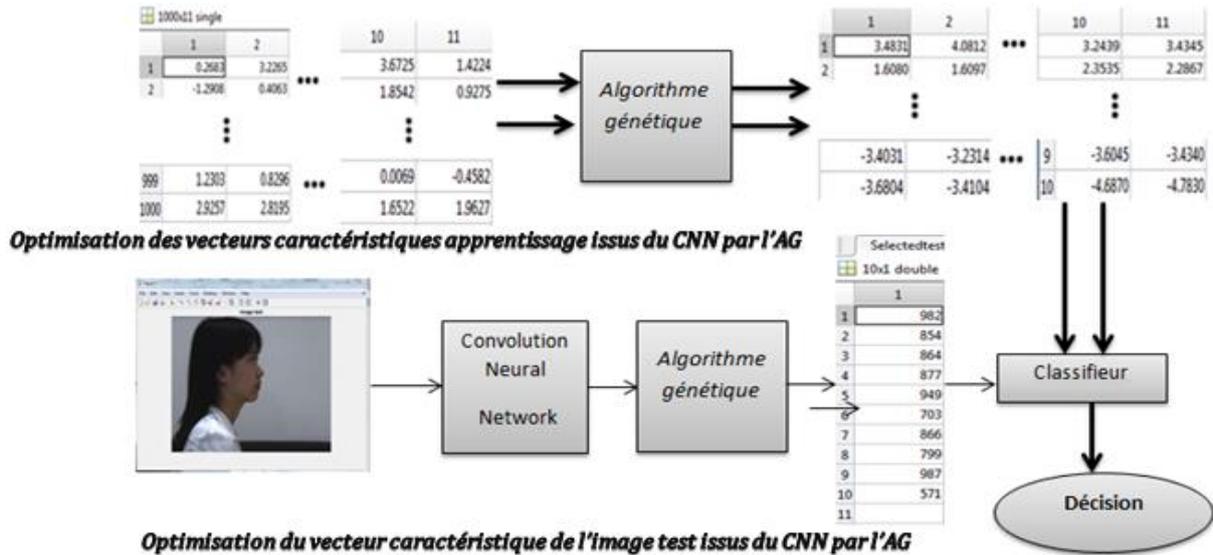


Figure 4.19 Optimisation du système de reconnaissance de visage basé sur CNNAG (BDD Face)

Le temps nécessaire pour classer l'image test est donné par la figure

```

Command Window
New to MATLAB? See resources for Getting Started.
655
544
249
485

ans =

'image de test appartient à kathrine grp'

Elapsed time is 0.722380 seconds.
fx >>
    
```

Figure 4.20 Temps pour la reconnaissance basé sur CNNAG de l'image test (BDD Face)

Nous remarquons que le temps nécessaire pour classer l'image test en utilisant le CNNAG est meilleur en comparaison avec celui utilisant le CNN.

Dans le cas CNN le temps calculé est $t = 13.274166$ secondes, alors que pour le cas utilisant le CNNAG, $t = 0.722380$ secondes. D'après les résultats obtenus, nous pouvons dire que l'optimisation par l'algorithme génétique est confirmée pour les deux systèmes de reconnaissance du genre et du visage.

4.9 Conclusion

La méthode proposée utilise un l'apprentissage en profondeur CNN pour l'extraction des caractéristiques et leur réduction. Cette méthode nécessite plusieurs paramètres qui dictent la structure d'un CNN, notamment le nombre de couches de convolution, la taille des filtres de convolution et le nombre de filtres de convolution utilisés dans chaque couche. Les structures CNN évolutives produisent des réseaux haute performance offrant une précision de classification supérieure à celle du réseau à la pointe de la technologie lorsqu'ils sont testés sur des images de reconnaissance de visage. Seulement, le temps de calcul reste lent, ce qui nous a poussés à optimiser les paramètres caractéristiques par un algorithme génétique.

Deux applications ont été prises en considération : i) la reconnaissance du genre et ii) la reconnaissance de visage.

Plusieurs expériences sont menées sur des échantillons de trois base de données universelles. D'après les résultats obtenus, nous pouvons affirmer que le système basé sur l'optimisation par algorithme génétique que nous proposons pour la reconnaissance de visage utilisant le CNN améliore nettement les performances pour les deux applications qui ont fait objet de nos expériences et qui sont la reconnaissance du genre et la reconnaissance du visage. La rapidité du système de reconnaissance est extraordinaire avec la méthode CNNAG proposé

Conclusion générale

La biométrie est un domaine à la fois passionnant et complexe. Elle tente, par des outils mathématiques souvent très évolués, de faire la distinction entre des individus, nous obligeant à travailler dans un contexte de très grande diversité. Cette diversité se retrouve également dans le nombre considérable d'algorithmes qui ont été proposés en reconnaissance de visage. Cette dernière est une application logicielle biométrique capable d'identifier ou de vérifier de manière unique une personne en comparant et en analysant des schémas basés sur les contours de la face. La reconnaissance de visage est principalement utilisée à des fins de sécurité, même si d'autres domaines d'utilisation suscitent un intérêt croissant. En fait, la technologie de reconnaissance de visage a fait l'objet d'une attention particulière dans la mesure où elle offre un large éventail d'applications liées à l'application de la loi et à d'autres entreprises.

Dans ce mémoire, nous nous sommes intéressés au problème de reconnaissance faciale. Notre travail consiste à la mise au point d'un algorithme robuste destiné à reconnaître un individu par son visage en utilisant deux méthodes parmi les méthodes les plus récentes dans ce domaine est la méthode d'apprentissage en profondeur CNN est c'est la première technique que nous utilisons pour l'extraction des caractéristiques d'une façon automatique et la deuxième c'est une méthode d'optimisation par les algorithmes génétiques pour la sélection des meilleures qui sont soumises au classifieur SVM.

L'apprentissage en profondeur a déjà été intégré à la plupart des pipelines de reconnaissance faciale à la pointe de la technologie. Cette évolution a entraîné une augmentation massive de la précision des systèmes de reconnaissance faciale et a entraîné la saturation du critère de référence «standard» actuel pour la reconnaissance faciale. En outre, les besoins en données pour les réseaux profonds mettent en évidence la nécessité d'un nouveau jeu de données public à très grande échelle (des dizaines de millions d'images) destiné à la reconnaissance faciale. [15]

Un algorithme génétique est un algorithme itératif qui s'exécute dans un cycle de population. Un cycle représente le passage d'une population à la génération suivante. L'évolution génétique d'une population procède en sélectionnant des individus sur la base de leurs pertinences à résoudre le problème relatives à l'espace de solution mesuré par la fonction fitness. Les éléments les plus pertinents ont la chance à se reproduire par l'intervention d'un croisement entre deux chromosomes ou bien une mutation au niveau des

Conclusion générale

gènes. La succession des cycles conduit à explorer les états possibles et augmente la performance à résoudre le problème posé[32]

Les algorithmes génétiques manipulent une population d'individus de taille constante, cette population de taille constante est soumise à une compétition entre les individus. Chaque individu est donné sous forme d'une seule chaîne de caractères appelée chromosome et qui représente un point de l'espace de recherche. L'algorithme génétique fait évoluer cette population d'individus au cours des générations à l'aide d'opérateurs de sélection, de croisement et de mutation qui s'inspirent des phénomènes naturels. Le croisement et la mutation sont chargés d'explorer l'espace de recherche en construisant de nouveaux individus à partir de la génération précédente, alors que la sélection favorise les individus qui possèdent une adaptation élevée.

Dans tous les cas, il nous était nécessaire de définir une fonction de fitness qui mesure la qualité. Il n'existe pourtant pas de mesure communément admise, et des mesures distinctes peuvent conduire à des classifications distinctes. De plus, la fonction de fitness peut elle-même dépendre de paramètres qu'il faudra ajuster. Il est également nécessaire de déterminer les valeurs des paramètres qui régissent le comportement de ces algorithmes et l'état actuel de la théorie des AGs ne permet pas de disposer de règles automatiques pour fixer les valeurs de ses paramètres, et il est souvent nécessaire de procéder à un ajustement empirique. Bien que certains systèmes fiables aient été introduits dans des conditions relativement contrôlées, leur taux de reconnaissance n'est pas satisfaisant dans les conditions générales. Cela est particulièrement vrai lorsqu'il y a des variations dans la pose, l'éclairage et l'expression du visage. Pour remédier à ces problèmes, un système hybride de reconnaissance des visages a été proposé, qui tire parti de la supériorité du réseau de neurones à convolution (CNN) et de la machine à vecteurs de support (SVM). À cette fin, un algorithme génétique (AG) est tout d'abord utilisé pour trouver la structure optimale de CNN. Ensuite, les performances du système sont améliorées en remplaçant la dernière couche de CNN par un ensemble de SVM. Enfin, en utilisant les concepts de correction d'erreur, la décision est prise. Le potentiel de CNN en tant qu'extracteur d'entraînements constitue un système de reconnaissance flexible capable de reconnaître les visages présentant des variations de pose et d'éclairage. Les résultats de la simulation montrent que le système atteint un bon taux de reconnaissance et est robuste contre les variations en termes de des expressions faciales, de l'occlusion, du bruit et des illuminations.

Conclusion générale

Une population d'individus (vecteur caractéristique) issue du CNN avec variantes d'architectures est initialisée de manière aléatoire en premier par l'algorithme génétique qui va sélectionner les meilleures caractéristiques, puis l'aptitude de chaque individu est évaluée. L'aptitude est mesurée par une fonction déterministe, appelée fonction d'aptitude, basée sur le contexte du problème à optimiser (c'est-à-dire la performance des CNN dans des tâches de classification d'images spécifiques). L'entrée de la fonction est la variable de décision encodée dans les individus. Après cela, les caractéristiques qui ont un meilleur apport en information seront choisies par l'opération de sélection afin de générer des caractéristiques optimales.

Pour mettre en œuvre avec succès un algorithme génétique simple, il faut choisir avec soin les trois paramètres suivants : Probabilité de croisement P_c . Probabilité de mutation P_m ; taille de la population I . Par ailleurs, il faut choisir entre divers mécanismes de sélection, de recombinaison et de mutation. Enfin, on peut être amené à prendre une décision sur la stratégie de remplacement des individus d'une génération à l'autre. Une difficulté supplémentaire réside dans l'interdépendance de ces divers paramètres et diverses études existent, apportant une aide dans ces choix délicats, des essais se révèlent indispensables pour optimiser les valeurs retenues en fonction du problème à résoudre. En répétant une série de ces opérations, on s'attend à ce que la population trouve la solution optimale à la fin de l'AG.

Chaque fois que nous voulons reconnaître une image test, nous la soumettons simplement au système proposé qui va connaître sa classe d'une façon automatique. Nous procédons comme suit:

- Extraction de caractéristiques de l'image d'entrées après son insertion dans l'architecture CNN à la sortie nous récupérons un vecteur de caractéristiques réduit ;
- Optimisation du vecteur caractéristique réduit par l'algorithme génétique ;
- Classification par SVM pour la reconnaissance de visage ;
- Calcul du temps total du processus de reconnaissance d'une image test de visage.

L'approche proposée est validée sur trois bases de données universelles et les résultats sont satisfaisants. Le temps total du processus de reconnaissance d'une image test de visage a été nettement amélioré par l'algorithme génétique.

Conclusion générale

D'après les résultats obtenus, nous pouvons affirmer que le système basé sur l'optimisation par algorithme génétique que nous proposons pour la reconnaissance de visage est

- utilisant le CNN améliore les performances pour les deux applications qui ont fait objet de nos expériences et qui sont la reconnaissance du genre et la reconnaissance du visage
- La rapidité du système de reconnaissance est extraordinaire avec la méthode CNNAG proposée.

Dans les travaux futurs, nous proposons :

- Une meilleure détection basée sur le Deep Learning ;
- Une amélioration de l'extraction des caractéristiques par d'autres techniques d'apprentissage automatique ;
- Une optimisation de la structure du CNN.

Bibliographie

[1]E.Hadjaidji « Modélisation d’empreinte biométrique par un modèle flou de Sugeno optimisé »
Master Université KasdiMerbah– Ouargla | 2017)

[2]Mourad MOUSSA1 *, MahaHMILA2 , Ali DOUIK2 “ A Novel Face Recognition Approach Based on Genetic Algorithm Optimization” 1 Department of Informatics, Sciences Faculty of Gafsa, University of Gafsa, Tunisia 2 Department of Industrial Informatics, National engineering school of Sousse University of Sousse, Tunisia

[2]CHAIB ISHAK optimisation d’un système de reconnaissance visage 3D par l’algorithme génétique
Mémoire master 2015 Biskra

[3]BENHAMZA NAJIB “ fusion 2D et 3D pour la reconnaissance de visage ”, master université
Mohamed khider Biskra 2012.

[4]Alex Coninx. , Doctorat Informatique, Université de Grenoble Alpes (2012) Publié sur [Quora](#) en français sur Facebook

[5] Bérengère Mathieu doctorat image information et hypermédia
« Deep Learning et détection d’émotions » 2018

[6]Boughaba Mohammed etBoukhrisBrahim. L’apprentissage profond (Deep Learning) pour la classification et la recherche d’images par le contenu. Master. UNIVERSITE KASDI MERBAH OUARGLA
Département d'Informatique et des Technologies de l'information 16-17

...[2]CHAIB ISHAK optimisation d’un système de reconnaissance visage 3D par l’algorithme génétique
Mémoire master 2015 Biskra

[8] (Q) Reconnaissance de visages par Analyse Discriminante Linéaire(LDA) MERAMRIA
nabilaunivbadjimokhtarannaba 2016))

[9]TALEB Imène. Le système biométrique : détection et reconnaissance de visage Université des sciences et de la technologie d'Oran - Mohamed-Boudiaf

[10]Wael Ben Soltana Optimisation de stratégies de fusion pour la reconnaissance de visages 3D.
Thèse de doctorat France Mars 2013

[11]Salem bin Al-Olayan's blog 2018 Doctorat de l'Arizona State University, États-Unis. assistant professor at the Faculty of Computer Science and supervisor of the Artificial Intelligence Unit and General Supervisor of Information Technology at King Khalid University.

[12]SulisSetiowati ,Zulfanahri , EkaLegyaFranita, IgiArdiyanto. A Review of Optimization Method in Face Recognition: Comparison Deep Learning and Non-Deep Learning Methods2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE), Phuket, Thailand

[13]Melissa Runfeldt Senior Data Scientist, Einstein AI at SalesforceSan Francisco, California Logicielsinformatiques "

[14]J. McCall. Genetic algorithms for modelling and optimization.Journal of Computational and AppliedMathematics (2005), Volume 184, pp. 205-222

[15]Melle Asma BENDAHMENE une approche basée algorithme génétique pour l'apprentissage chez un agent Magister en Informatique2011Université Mohamed Khider Biskra Faculté des sciences exactes, de la science naturelle et de la vie

[16]SanthoshKumar S, "A FAST TIME SCALE GENETIC ALGORITHM BASED IMAGE SEGMENTATION USING CELLULAR NEURAL NETWORKS (CNN)" 2007Final year students of Department of Electronics and Communication Engineering Sri Venkateswara College of engineering, Sriperumbudur India

[17]Jihoon Yang and Vasant Honavar FEATURE SUBSET SELECTION USING A GENETIC ALGORITHMAI Research Group, Department of Computer Science 226 Atanasoff Hall, Iowa State University

[18] Adjani fatima 'optimisation des SVM multi-classes par des méthodes évolutionnaires (PSO-SVM) mémoire de magister Université des Sciences et de la Technologie d'Oran – mohamedboudiaf Faculté des Sciences Département d'Informatique

[19] Sehla Loussaief, Afef Abdelkrim "Convolutional Neural Network Hyper-Parameters Optimization based on Genetic Algorithms" National Engineering School of Tunis (ENIT), University of Tunis El Manar 2018

[20] Spencer Gibb, Hung Manh La, and Sushil Louis "A Genetic Algorithm for Convolutional Network Structure Optimization for Concrete Crack Detection" 2018

[21] Shanshan Guo, Shiyu Chen and Yanjie Li "Face Recognition Based on Convolutional Neural Network and Support Vector Machine" Department of Mechatronics Engineering and Automation Harbin Institute of Technology Shenzhen Graduate School Shenzhen, Guangdong Province, China 2016

[22] AYADI Abdelghafar "APPLICATION D'UNE VARIANTE D'ALGORITHMES GENETIQUES A LA SEGMENTATION D'IMAGES" UNIVERSITE DE M'SILA FACULTE DES MATHEMATIQUES ET DE L'INFORMATIQUE Département d'Informatique magister

[23] Cours Dr. Mohamed Assaad HAMIDA Introduction aux algorithmes génétiques

[24] Joshua Lamos-Sweeney "Deep learning using genetic algorithms" 2012 Rochester Institute of Technology RIT Scholar Works

[25] Abdelhamid DJEFFAL "utilisations des méthodes support vecteur machine (SVM) dans l'analyse des bases de données département d'informatique, laboratoire LESIA université de biskra 2012

[26] Abien Fred M. Agarap "An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image" 2019

[27] Olivier Bousquet "Introduction aux \Support Vector Machines" (SVM) Centre de Mathématiques Appliquées Ecole Polytechnique, Palaiseau 15 Novembre 2001

[28] Yann Guermeur "SVMMulti-classes, Théorie et Applications" Département de formation doctorale en informatique 28 novembre 2007

Chapitre IV Conception du système et résultat

[29] Zhao Lihong , Song Ying , Zhu Yushi , Zhang Cheng , Zheng Yi ‘‘Face Recognition based on multi-class SVM’’ College of Information Science and Engineering, Northeastern University, Shenyang

[30]<http://www.cbsr.ia.ac.cn/english/3DFace%20Databases.asp>

[31] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments

[32] Alex Krizhevsky - Ilya Sutskever - Geoffrey E. Hinton ‘‘ ImageNet Classification with Deep Convolutional Neural Networks’’