



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider – BISKRA

Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie

Département d'informatique

N° d'ordre :IVA23./M2/2021

Mémoire

Présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : **Image et Vie Artificielle (IVA)**

Reconnaissance Facial dans une séquence d'images

Par :

SOLTANE KAOUTHAR

Soutenu le 28/06/2022 devant le jury composé de :

BOUCETTA MEBARK

MCB

Président

AKROUR DJOUHER

MCB

Rapporteur

BENAMEUR SABRINA

MCB

Examineur



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider – BISKRA

Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie

Département d'informatique

N° d'ordre :IVA23./M2/2021

Mémoire

Présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : Image et Vie Artificielle (IVA)

Reconnaissance Facial dans une séquence d'images

Par :

SOLTANE KAOUTHAR

Soutenu le 28/06/2022 devant le jury composé de :

BOUCETTA MEBARK

MCB

Président

AKROUR DJOUHER

MCB

Rapporteur

BENAMEUR SABRINA

MCB

Examineur

Dédicace

Je dédie ce modeste travail

A mes très chers parents pour leur soutien et
encouragement

durant toutes mes années d'études et sans
lesquels je n'aurais jamais réussi,

A mes sœurs sara et meriem et à toute ma
famille.

A tous mes amis ainsi qu'à toutes les personnes
que j'ai connues,

qui m'ont aidées, soutenues et encouragées.

A tous mes enseignants durant mes années
d'études avec lesquels
j'ai beaucoup appris

Soltane kaouthar

Remerciement

Tout d'abord, je remercie Dieu Tout-Puissant de m'avoir donné la force et la patience pour y arriver.

Je remercie mes parents pour tout le soutien matériel et moral jusqu'à ce que j'atteigne ce que j'ai aujourd'hui

Que Dieu vous accorde une bonne santé

Je tiens à remercier ma superviseuse, Sabrina Benameur, pour ses précieux conseils.

Je remercie également Messerhi Djalal pour tous les efforts de ma direction

Des remerciements particuliers seront adressés aux membres du jury pour leurs efforts.

Résumé

La reconnaissance faciale est une technologie de plus en plus répandue, basée sur l'intelligence artificielle, permettant d'identifier une personne sur une photo ou une vidéo en comparant son visage avec ceux sauvegardés dans une base de données. Cette technologie est utilisée dans plusieurs domaines tel que le contrôle dans les aéroports et aux frontières, le déverrouillage des téléphones, la recherche des personnes disparues, la réduction de la criminalité dans les magasins, ...

Notre travail consiste à concevoir et implémenter un système permettant de faire la détection et la reconnaissance faciale dans une séquence d'image basé sur la technique CNN. La détection du visage dans chaque frame de la vidéo nous a permis de faire le suivi de la personne reconnu. Nous avons pu ajouter aussi un module de détection des expressions faciales. Les résultats obtenus par notre application sont satisfaisant.

Mots clés: Reconnaissance faciale, OpenCV, Apprentissage profond, CNN, Reconnaissance de visage, Détection de visage, extraction des caractéristiques de visage, deep face, python

TABLE DES MATIÈRES

Introduction générale.....	1
chapitre 1 La reconnaissance faciale: Concepts de base	3
1.1 Introduction	3
1.2 Un système biométrique :.....	3
1.2.1 Les phases d'un système biométrique	3
1.2.2 Modules de systèmes biométriques	4
1.3 La biométrie :	5
1.4 Les caractéristiques de biométrie :.....	5
1.5 La biométrie du visage	6
1.6 La reconnaissance faciale.....	6
1.6.1 Définition.....	6
1.6.2 Système de reconnaissance faciale	7
1.7 La détection de visage :.....	8
1.8 Les difficultés liées à la détection de visage :	9
1.9 Les techniques de détection et reconnaissance de visage.....	10
1.9.1 Les méthodes de détection de visage :	10
1.9.2 Techniques de reconnaissance de visage.....	11
1.9.3 Principales difficultés de la reconnaissance de visage :	13
1.9.3.2 Influence des variations de la pose.....	14
1.9.3.3 Influence des expressions faciales	14
1.9.3.4 Influence des occultations partielles	15
1.10 Conclusion	15
chapitre 2 : L'apprentissage profond et la reconnaissance faciale	16
2.1 Introduction	16
2.2 Définitions	16
2.3 Le principe de l'apprentissage automatique	17
2.4 L'apprentissage profond (deep learning)	19
2.5 Pourquoi le choix deep Learning	19
2.6 Les différentes Architectures du Deep Learning :	20
2.6.1 Réseaux de neurones convolutifs	20
2.6.2 Réseaux de neurones récurrents	21

2.6.3	Modèle génératif.....	21
2.7	L'architecture de CNN	22
2.7.1	Couche de convolution (CONV)	23
2.7.2	Couche de pooling (POOL)	24
2.7.3	Couche de « entièrement connectée » (FC).....	25
2.7.4	Couche de perte (LOSS).....	26
2.8	Domaine d'application de Deep Learning.....	26
2.9	La reconnaissance faciale (Deep Face Recognition) :	26
Un	système de reconnaissance faciale est composé de trois modules :	26
2.9.1	Composantes de la reconnaissance faciale	26
2.9.2	Traitement du visage	27
2.9.3	Extraction en profondeur	27
2.9.4	Correspondance des visages par des traits profonds	28
2.9.5	Les travaux existant	28
2.10	Système ECNN	29
2.10.1	Présentation	29
2.10.2	Architecteur du système ECNN	30
2.11	Conclusion	31
chapitre 3	: Conception et implémentation.....	32
3.1	Introduction	32
3.2	Conception générale de notre système :	32
3.3	Conception détaillé.....	33
Dans ce paragraphe n.....		33
3.3.1	L'apprentissage du modèle	34
3.3.2	Importer des bibliothèques	34
3.3.3	Importer les ensembles de données	34
3.3.4	Entrainement du model (Training)	34
3.3.5	La détection de visage	36
3.3.6	La reconnaissance de visage.....	36
3.3.7	Extraction les caractéristique :	36
3.3.8	Formation du modèle avec keras	37
3.3.9	Présentation l'architecture Xception	37
3.4	Définition de la Convolution séparable.....	37

3.5	La convolution séparable en profondeur dans Xception.....	38
3.6	Implémentation.....	38
3.6.1	Environnements et Outils de développement.....	39
3.6.2	Environnement de développement	39
3.6.3	La base de données	42
3.7	Algorithmes	42
3.8	Présentation de système.....	43
3.8.1	Interface de notre système	44
3.8.2	Résultats et discussion.....	44
3.8.3	Suivi du visage en temps réel	44
3.8.4	Suivi du visage en temps passive	45
3.9	Résultat.....	47
3.10	Conclusion.....	50
	Conclusion général.....	51
	Référence.....	52

Liste des Figures

Figure 1-1 Architecture d'un système biométrique.[1]	4
Figure 1-2 Les différents modules d'un système biométrique.[1]	4
Figure 2-1 la relation entre IA,ML,DL [15].....	17
Figure 2-2 L'apprentissage supervisé et non supervisé.[10].....	18
Figure 2-3 comparaison entre machine learning et deep learning [10]	20
Figure 2-4 Les différentes Architectures du Deep Learning [10].....	20
Figure 2-5 L'architecture d'un réseau de neurone convolutionnels[10].....	22
Figure 2-6 Représentation d'un padding. [10].....	23
Figure 2-7 Principe de la convolution.[10]	24
Figure 2-8 Principe de pooling.	25
Figure 2-9 Exemple des couches entièrement connectées	25
Figure 3-1 Cycle de la vie d'entraînement	36
Figure 3-2 l'architecture Xception[39].....	37
. Figure 3-3 Environnement de développement.	39

Liste des Tableaux

Tableau 1 Les travaux réalisés par l'apprentissage en profondeur [6]	29
--	----

Liste des schémas

schéma 1 Conception général de système.....	32
Schéma 2 Conception détaillée de système.....	33

Introduction générale

La sécurité est un critère primordial dans notre vie quotidienne dans divers secteurs. Le contrôle d'accès se base essentiellement sur deux méthodes traditionnelle d'identification, la première basée sur les connaissances a priori de l'utilisateur « **knowledge-based** » telle que le code PIN (Mot de passe), mais cette technique n'est pas fiable a cent pourcent parce que le mot de passe peut-être oublier par l'utilisateur. La seconde technique est basée sur la possession d'un objet "**token-based**", tels que les clefs, carte bancaire, badge, etc. les faiblesses de cette technique est que les clefs ou les cartes peuvent être volé ou perdu. En effet le contrôle d'accès par les méthodes traditionnelles d'identification n'est pas fiable et il n'assure pas une hautesécurité

Aujourd'hui des nombreuses recherches se sont développées dans le domaine de la biométrie pour trouver des solutions alternatives aux deux méthodes d'identification précédente. La technique basée sur les caractéristiques biométriques est plus puissante et fiable, dont l'avantage est que les caractéristiques biométriques sont universelles, unique et permanente. Le visage est l'un des caractéristiques biométriques utilisé pour l'identification. La reconnaissance faciale est une méthode d'identification elle est devenu ces dernières années l'une des branches de la vision par ordinateur qui remporte un grand succès et qui est en développement continu.

Un système de reconnaissance de visage permet de contrôler l'accès à partir d'une vérification d'identité des individus. Mais avant de vérifier l'identité il faut d'abord détecter le visage et extraire les composantes faciales nécessaire pour la procédure de reconnaissance.

Plusieurs méthodes ont été développées dans la littérature pour la reconnaissance de visage. Dans notre travail nous avons opté pour une technique basée sur les réseaux de neurone appelé réseaux de neurone convolutionnels (CNN) qui est un type de réseau de neurone avec un apprentissage profond, ou Deep Neural Network. Ce dernier comporte plusieurs couches cachées. Le CNN est composé de deux parties bien distinctes, partie d'extraction qui peut être utilisée pour simplifier un ensemble de données, en réduisant sa dimension, et partie de classification qui classifie ces données.

Nous nous sommes intéressés à la détection et la reconnaissance faciale. Ce travail a donc pour but de construire et d'implémenter un modèle de détection et de reconnaissance de visage en utilisant un apprentissage profond. Nous avons donc essayé de réaliser un système basé sur un réseau de neurones convolutif, notre système permet de détecter des visages et les identifier, de plus il permet de faire le suivi de la personne identifier. Nous avons ajouté aussi à notre système un module de reconnaissance des expressions faciales.

En plus de l'introduction générale, notre mémoire est organisé en 3 chapitres, Nous avons vu qu'il est primordial de comprendre les systèmes de reconnaissance faciale et leur principe. Le premier chapitre est ainsi consacré à ce système. Dont nous commençons par la définition des systèmes biométrique, par la suite nous présentons la reconnaissance faciale avec les différentes méthodes de détection et de reconnaissance de visage. A la fin du chapitre nous citons les difficultés liées à ces systèmes.

Dans le chapitre 2, nous définissons l'apprentissage profond (deep learning) et ces différentes architectures, nous présentons ensuite les réseaux de neurones convolutifs tout en expliquant ses différentes couches. Enfin, nous introduisons les systèmes de reconnaissance faciale avec l'exploitation de l'apprentissage profond, et comme exemple nous expliquons le système ECNN qui est un système de reconnaissance d'expression faciale basé sur le CNN.

Le troisième chapitre est réservé à la conception et l'implémentation de notre système. Dont nous commençons par la présentation de la conception générale et détaillée de notre application ensuite nous détaillons les étapes d'implémentation. Les résultats obtenus ainsi qu'une évaluation sont illustrés à la fin du chapitre.

Une conclusion ainsi que des perspectives à ce travail sont présentées à la fin de ce mémoire.

chapitre 1 La reconnaissance faciale: Concepts de base

1.1 Introduction

Le visage humain constitue un stimulus visuel de classe à part. En effet, il suffit d'un clin d'œil porté sur le visage d'un individu pour en distinguer le sexe, l'état émotionnel ou l'identité. Cette grande capacité à identifier les visages a poussé les chercheurs à tenter de rapprocher le cerveau humain dans sa rapidité, son exactitude et sa fiabilité par des systèmes de reconnaissance de visage.

Dès lors, la reconnaissance des visages a connu un fort développement et reste un domaine qui suscite toujours des interrogations par les chercheurs.

Dans ce chapitre, nous commençons par la définition des systèmes biométriques, nous détaillons ensuite la reconnaissance de faciale, tout en présentant ces systèmes et en citant les différentes méthodes de détection et de reconnaissance faciale. Enfin nous présentons les difficultés liées à ces systèmes.

1.2 Un système biométrique :

Un système biométrique est essentiellement un système de reconnaissance de formes qui fonctionne en acquérant des données biométriques à partir d'un individu extrayant un ensemble de caractéristiques à partir des données acquises, et comparant ces caractéristiques contre la signature dans la base de données. [1]

Il sert à vérifier l'identité d'une personne à l'aide d'une ou plusieurs modalités qui lui sont propres (voix, iris, empreintes digitales, visage ...). On peut dire qu'un système de contrôle biométrique est un système automatique de mesure basé sur la reconnaissance de caractéristiques propres à l'individu.[1]

1.2.1 Les phases d'un système biométrique

Chaque système biométrique comporte deux phases d'exécution :

- La phase d'apprentissage
- La phase de test.

Chacune contient trois modules de base plus un module propre à la phase de test

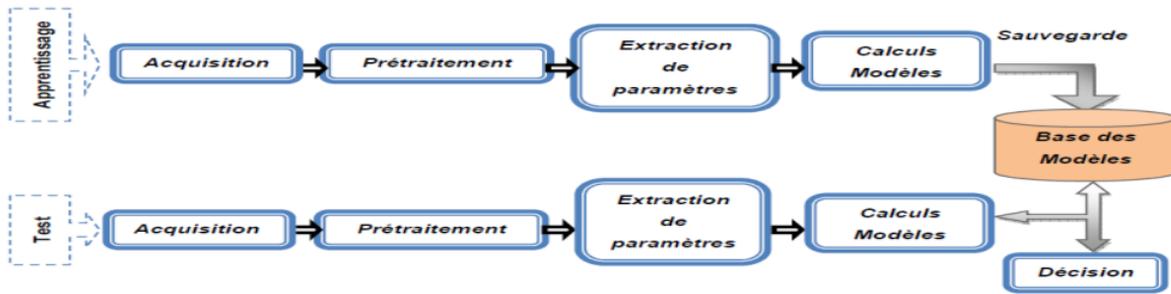


Figure 1-1 Architecture d'un système biométrique.[1]

1.2.2 Modules de systèmes biométriques

Un système biométrique typique peut être représenté par quatre modules principaux :

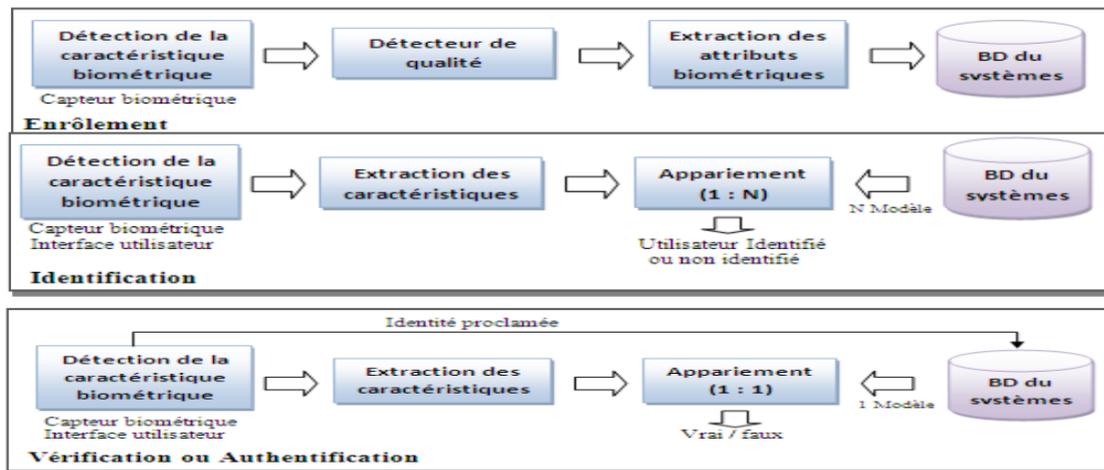


Figure 1-2 Les différents modules d'un système biométrique.[1]

- ✓ **Module de capture** : responsable de l'acquisition des données biométriques d'un individu (cela peut être un appareil photo, un lecteur d'empreintes digitales, une caméra de sécurité, etc.).[1]
- ✓ **Module d'extraction de caractéristiques** : Qui prend en entrée les données biométriques acquises par le module de capture et extrait seulement l'information pertinente afin de former une nouvelle représentation des données. Idéalement, cette nouvelle représentation est censée être unique pour chaque personne et relativement invariante aux variations intra-classe.[1]

- ✓ **Module de correspondance** : Il compare l'ensemble des caractéristiques extraites avec le modèle enregistré dans la base de données du système et détermine le degré de similitude (ou de divergence) entre les deux.
- ✓ **Module de décision** : vérifie l'identité affirmée par un utilisateur ou détermine l'identité d'une personne basée sur le degré de similitude entre les caractéristiques extraites et le(s) modèle(s) stocké(s).[1]

1.3 La biométrie :

La biométrie est la mesure et l'analyse statistique des caractéristiques physiques et comportementales des individus. Cette technologie est principalement utilisée à des fins d'identification et de contrôle d'accès. Elle permet aussi de tracer des individus sous surveillance. Selon le principe de base de l'authentification biométrique, chaque individu est unique et peut être identifié par ses caractères physiques et ses traits de comportement intrinsèques. Il existe deux principales classes d'identificateurs biométriques : [2]

- Les caractéristiques morphologiques : forme ou composition du corps.
- Les caractéristiques comportementales : la façon dont se comporte une personne

Au niveau morphologique, par exemple, les facteurs les plus utilisés pour l'authentification biométrique sont les empreintes digitales, l'ADN, la reconnaissance faciale, la morphologie de la main, l'analyse de la rétine ou de l'oreille et l'odeur. L'analyse comportementale porte sur des éléments propres au comportement d'une personne, tels que la dynamique des frappes au clavier, la démarche, les gestes et la reconnaissance vocale.[2]

1.4 Les caractéristiques de biométrie :

- **Universalité** : toute personne doit la posséder pour être identifiée [2].
- **Unicité** : les informations doivent être aussi différentes que possible entre différentes personnes [2].
- **Permanence** : les informations recueillies doivent être présentes tout au long de la vie Célibataire[2].
- **Collectabilité** : les informations doivent être collectables et mesurables pour être utilisées pour les comparaisons[2].
- **Acceptation** : le système doit respecter certains critères (facilité d'acquisition, vitesse, etc.) pour l'utiliser.[2]

1.5 La biométrie du visage

Rien n'est plus naturel qu'utiliser le visage pour identifier une personne. Les images faciales sont probablement la caractéristique biométrique la plus communément employée par l'homme pour effectuer une identification personnelle. L'utilisation d'une caméra permet de capter la forme du visage d'un individu et d'en dégager certaines particularités.

Selon le système utilisé, l'individu doit être positionné devant l'appareil ou peut être en mouvement à une certaine distance. Les données biométriques qui sont obtenues sont par la suite comparées au fichier référence.

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.

1.6 La reconnaissance faciale

1.6.1 Définition

La reconnaissance faciale est une technologie de plus en plus répandue, basée sur l'intelligence artificielle permettant d'identifier un individu d'après une photo ou vidéo ou d'après un modèle 3D créé à partir de plusieurs images de la personne. En comparant son visage avec ceux sauvegardés dans une base de données. Ce programme est courant dans le domaine de la sécurité, afin d'identifier des criminels ou des disparus et autoriser l'accès à certains lieux publics. Les logiciels de reconnaissance faciale sont aussi créés pour aider à l'utilisation d'applications telles que les réseaux sociaux qui reconnaissent des visages sur les photos mises en ligne ou le déverrouillage d'appareils. Analysant les particularités d'un visage, l'algorithme compare ses résultats à une base de données établie et peut être utilisé en marketing pour analyser et cibler une clientèle selon son genre ou son âge.[3]

 La reconnaissance faciale permet :

- **d'authentifier une personne** :c'est-à-dire vérifier qu'une personne est bien celle qu'elle prétend être (pour un contrôle d'accès par exemple) ;
- **d'identifier une personne** :c'est-à-dire de retrouver une personne au sein d'un groupe d'individus, dans un lieu, une image ou une base de données.

1.6.2 Système de reconnaissance faciale

Un système de reconnaissance faciale est une application informatique pour identifier ou vérifier automatiquement une personne à partir d'une image numérique ou une frame d'une source vidéo[4]. Une des façons de le faire est de comparer les traits du visage sélectionnés de l'image et une base de données des visages. Les systèmes de reconnaissance faciale fonctionnent différemment, mais leur principes sont presque les mêmes.[3]

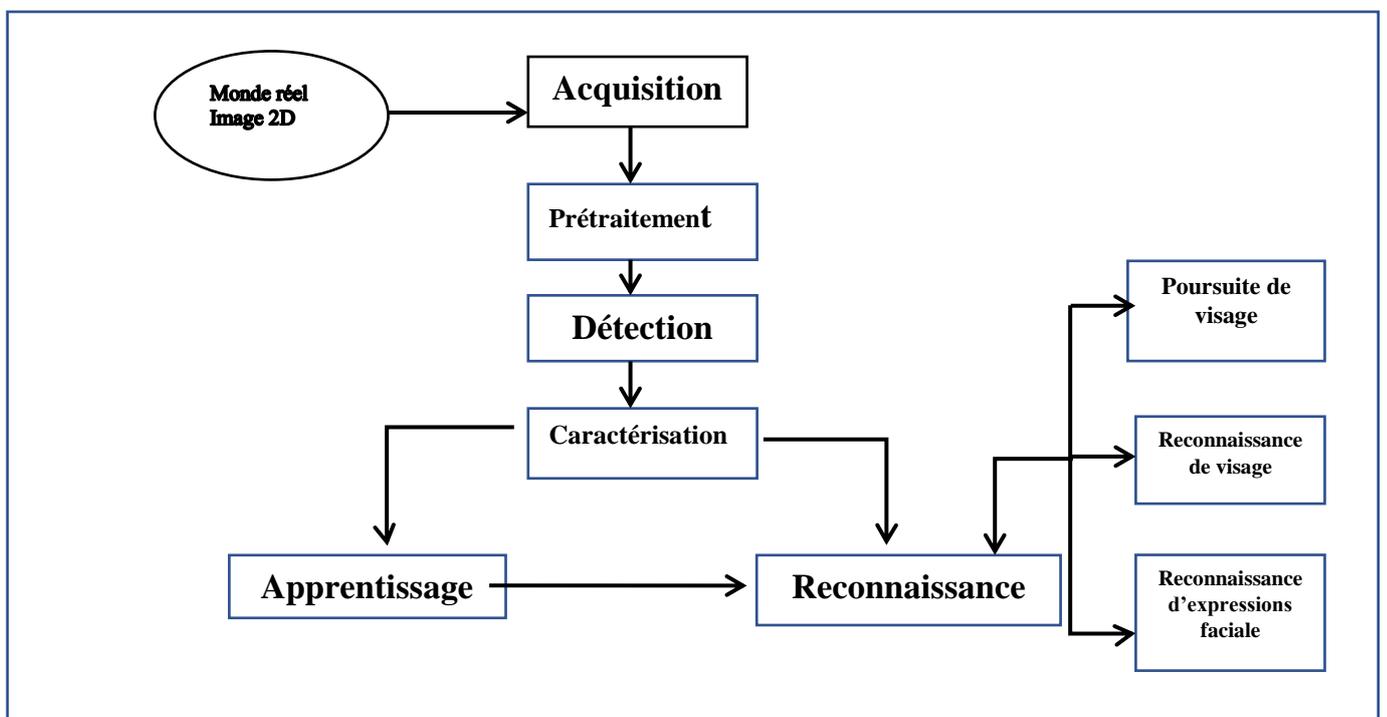


Schéma général de système de reconnaissance faciale

1.6.2.1 Acquisition de l'image :

L'acquisition de l'image c'est la première étape dans un système de reconnaissance, elle est faite en capturant l'image de l'individu du monde extérieur grâce aux appareils comme une caméra ou un appareil photo.

1.6.2.2 Prétraitements

Le rôle de cette étape est d'éliminer les parasites causés par la qualité de l'appareil optique ou électronique lors de l'acquisition de l'image d'entrée, dans le but

de ne garder que les informations de base et ainsi préparer l'image pour l'étape suivante.

1.6.2.3 La détection du visage

Le plus souvent la détection de visage inclut sa localisation, elle vérifie la présence d'un ou de plusieurs visages dans une image et retourne leurs positions.

1.6.2.4 Caractérisation

L'étape de l'extraction des caractéristiques représente le cœur du système de reconnaissance, elle consiste à effectuer le traitement de l'image dans un autre espace de travail plus simple et qui assure une meilleure exploitation de données, et donc permettre l'utilisation, seulement, des informations utiles, discriminantes et non redondantes

1.6.2.5 Apprentissage

C'est l'étape où on fait apprendre les individus au système, elle consiste à mémoriser les paramètres, après extraction et classification, dans une base de données bien ordonnées pour faciliter la phase de reconnaissance et la prise d'une décision, elle est en quelque sorte la mémoire du système.

1.6.2.6 La reconnaissance et/ou vérification du visage :

Elle retourne l'identité de la personne dans le cas de la reconnaissance et le droit d'accès dans le cas de la vérification.

1.6.2.7 La poursuite du visage :

Elle essaye de suivre un visage dans une séquence d'images

1.6.2.8 Expressions faciale :

On peut décrire l'état du visage, par exemple, dans un état ou anxiété, normal...

1.7 La détection de visage :

La détection de visage est un terme technique informatique utilisé lorsqu'un logiciel est utilisé pour déterminer la présence, l'emplacement et la taille d'un visage humain dans une image donnée. Le logiciel est suffisamment intelligent pour détecter les traits du visage, tout en ignorant d'autres choses comme les arbres, les bâtiments et les objets.

Un système de détection des visages est conçu pour répondre à la question : cette image contient-elle un visage ? Un système de détection des visages détermine la présence, l'emplacement, les dimensions et (éventuellement) l'orientation de n'importe quel visage présent dans une image fixe ou une trame vidéo



Figure 1-3 Image de détection d'image[12]

1.8 Les difficultés liées à la détection de visage :

Les difficultés liées à la détection de visage peuvent être attribuées aux facteurs suivants :

- **Posture.** L'image d'un visage change en raison de la position relative caméra-visage (de face, de profil ou dans une position intermédiaire), et certains attributs faciaux tels que les yeux ou le nez peuvent être partiellement ou complètement occultés.[8]
- **Présence ou absence de composantes structurelles** Les attributs faciaux tels que la barbe, la moustache et ou des lunettes peuvent être présents ou pas et cela avec une forte variabilité. De plus, ces attributs peuvent eux-mêmes revêtir des formes très différentes d'un individu à l'autre : géométrie, couleur, taille, *etc.*[8]
- **Expression faciale** L'expression faciale d'une personne affecte directement l'aspect de son visage. [8]
- **Occultation** :Des visages peuvent être partiellement cachés par d'autres objets. Dans une image avec un groupe de personnes, certains visages peuvent partiellement ou entièrement en cacher d'autres. [8]
- **Orientation de l'image** :Les images de visage changent directement pour différentes rotations autour de l'axe optique de la caméra.[8]

- **Conditions de prise de vue :** Des facteurs tels que l'éclairage (distribution, orientation et intensité de la source) et les caractéristiques de la caméra (capteur, optique) affectent l'aspect d'un visage dans l'image.[8]

1.9 Les techniques de détection et reconnaissance de visage

1.9.1 Les méthodes de détection de visage :

1.9.1.1 Les méthodes basées sur la connaissance :

Ces méthodes sont développées à partir des règles de base dérivées de la connaissance des visages humains, ce qui permet de proposer des règles simples pour décrire les caractéristiques du visage et leurs relations, Par exemple, un visage de face apparaît dans une image avec un nez, une bouche et deux yeux symétriques. Les relations entre les caractéristiques peuvent être représentées par leurs positions et leurs distances relatives. Dans ces méthodes il faut d'abord extraire les caractéristiques faciales dans l'image d'entrée et en se basant sur les règles prédéfinies à l'avance les visages candidats sont identifiés. [8]

Le problème essentiel de ces méthodes est la difficulté de traduire la connaissance humaine en règles bien définies. En effet, si celles-ci sont trop strictes, elles peuvent ne pas détecter les visages qui ne satisfont pas toutes les règles. A l'inverse, si elles sont trop « permissives », elles peuvent prendre à tort divers objets de la scène qui les respectent pour des visages. En effet ces méthodes peuvent bien fonctionner sur des scènes simples et où la vue des visages est plus au moins frontale.[8]

1.9.1.2 Les méthodes basées sur des caractéristiques invariantes

Afin de corriger les faiblesses de l'approche précédente, ces méthodes trouvent les caractéristiques invariantes qui existent quels que soit le point de vue ou l'éclairage. Ensuite elles utilisent ces caractéristiques pour la détection des visages. Cependant, du fait de la nature même de l'information de départ (une image), il demeure particulièrement difficile de s'affranchir des variations d'éclairage et des phénomènes d'occultation à l'aide des primitives précédentes. En effet, les contours de celles-ci peuvent être disparus, alors que les ombres peuvent créer de nombreux artefacts qui compliquent considérablement les tâches de ces algorithmes. Toutefois l'utilisation des propriétés colorimétriques de la peau ainsi que sa texture ont largement été utilisées. Celles-ci peuvent présenter une invariance suffisante malgré la posture du visage, son orientation ou son expression.[8]

1.9.1.3 Les méthodes d'appariement de modèles

Cette approche contient de nombreux modèles standards du visage qui sont appris et stockés afin de décrire le visage dans son ensemble ou avec quelques caractéristiques spécifiques. Pour une image d'entrée les valeurs de corrélation avec les modèles appris sont calculées indépendamment du contour de visage, des yeux, du nez et de la bouche. Les valeurs de corrélation permettent de savoir s'il y a un visage dans l'image ou non. Cette approche est facile à mettre en œuvre, mais elle est très peu robuste aux variations d'échelle, de forme, de la taille et de la position, de plus elle est coûteuse en calcul. [8]

1.9.1.4 Les méthodes basées sur l'apprentissage de modèles

A la différence des méthodes d'appariement précédentes, dont les modèles utilisés sont prédéfinis par des experts. Ces méthodes exploitent des modèles appris à partir des exemples d'images elles sont basées sur des techniques de l'analyse statistique et d'apprentissage pour trouver les caractéristiques des images de visage et de non-visage. Les caractéristiques apprises sont exprimées sous forme de modèles de distribution qui sont utilisés pour détecter la présence d'un visage. Les méthodes de cette catégorie obtiennent généralement les meilleures performances (taux de détection) mais elles ont l'inconvénient d'être coûteuses en calcul et elles sont assez difficiles à mettre en œuvre.[8]

1.9.2 Techniques de reconnaissance de visage

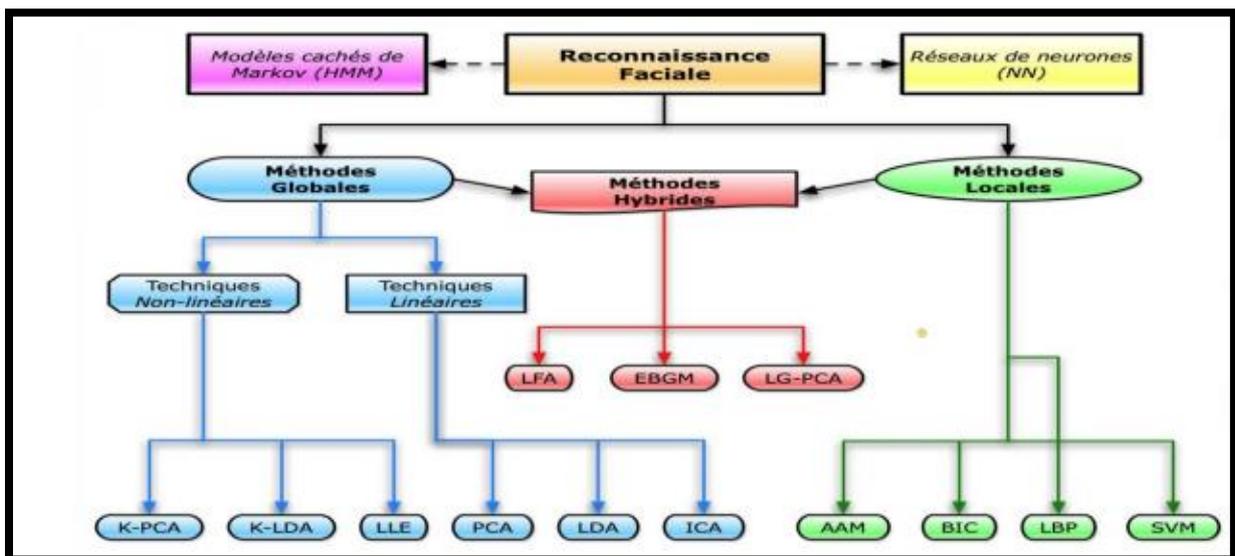


Figure 1-4 Classification des algorithmes principaux utilisés en reconnaissance faciale.

1.9.2.1 Les méthodes globales

Les méthodes globales utilisent toute la surface du visage pour obtenir des informations sans tenir compte des caractéristiques locales telles que la bouche, le nez et les yeux.... L'ACP est l'une des méthodes les plus couramment utilisées pour représenter le visage. Les algorithmes globaux adoptent des propriétés statistiques et une algèbre linéaire, elles ne sont pas coûteuses en temps de calcul mais sensibles aux différences de méthodes dans l'expression faciale en termes d'éclairage. Les méthodes globales sont divisées en deux classes : linéaire et non linéaire[10]

1.9.2.1.1 Techniques linéaires

Ces techniques consistent à réaliser une projection linéaire de l'image d'une grande dimension dans un espace de petite dimension. Cependant, une telle projection ne peut pas conserver les variations de visage non convexes, qui permettent de distinguer différents individus et la distance euclidienne ne peut pas être utilisée pour comparer les pixels car la classification visage, non visage ne sera pas effective et conduit à une détection et reconnaissance insatisfaisantes.[11]

1.9.2.1.2 Techniques non-linéaires

Des techniques globales non linéaires ont été développées, souvent à partir de techniques linéaires en utilisant la fonction « kernel » pour étendre les techniques linéaires.[10]

1.9.2.2 Approches locales

Les approches locales de la reconnaissance de visage sont basées sur des modèles et reposent sur un traitement séparé appliqué sur les différentes régions de l'image contenant un visage. Ce processus conduit à un vecteur caractéristique pour chaque région du visage. Ces approches peuvent être classifiées en deux catégories : Méthodes basées sur l'apparence locale et méthodes basées sur les points d'intérêts.[10]

1.9.2.3 Les méthodes hybrides :

De nombreuses méthodes ont été proposées pour la reconnaissance faciale, mais elles ne peuvent pas surmonter les changements environnementaux tels que l'éclairage, l'expression faciale, etc. ... L'efficacité du système de reconnaissance faciale peut être augmentée en intégrant de nombreuses méthodes. Il est par ailleurs possible d'utiliser une combinaison de classificateurs basés sur des techniques variées dans le but d'unir les avantages de chacun et ainsi pallier à leurs faiblesses. Les techniques hybrides combinent des méthodes des deux classes précédentes pour une meilleure caractérisation des images de visages.[10]

✚ La nouvelle tendance de la reconnaissance faciale (Deep Face Recognition)

L'apprentissage en profondeur (Deep Face Recognition) applique plusieurs couches de traitement pour apprendre des représentations de données avec plusieurs niveaux d'extraction de caractéristiques. Cette technique émergente a remodelé le paysage de la recherche en reconnaissance faciale depuis son apparition en 2014, lancée par les percées de la méthode Deepface. Depuis lors, la technique RF profonde, qui s'appuie sur une architecture hiérarchique pour assembler les pixels en une [6]représentation invariante des visages, a considérablement amélioré les performances de pointe et favorisé le succès des applications du monde réel.

Il y a trois modules nécessaires pour l'ensemble du système :

- Tout d'abord un détecteur de visages est utilisé pour localiser les visages dans les images ou les vidéos.

- Deuxièmement, avec le détecteur de repères faciaux, les visages sont alignés sur des coordonnées canoniques normalisées.

- Troisièmement, le module RF est implémenté avec ces images de visage alignées[6]

1.9.3 Principales difficultés de la reconnaissance de visage :

1.9.3.1 Influence des changements d'éclairage:

L'intensité et la direction d'éclairage lors de la prise de vue influent énormément sur l'apparence du visage dans l'image. En effet, dans la plupart des applications courantes, des changements dans les conditions d'éclairage sont inévitables, notamment lorsque les vues sont collectées à des heures différentes, en intérieur ou en extérieur. Etant donnée la forme spécifique d'un visage humain, ces variations d'éclairage peuvent y faire apparaître des ombres accentuant ou masquant certaines caractéristiques faciales[8].



Figure 1-5 Exemple de variation d'éclairage

1.9.3.2 Influence des variations de la pose

Les changements d'orientation et les changements de l'angle d'inclinaison du visage engendrent de nombreuses modifications d'apparence dans les images collectées. Une phase préliminaire de normalisation de l'image du visage permet de corriger d'éventuelles rotations dans le plan de celle-ci. Les rotations en profondeur engendrent l'occultation de certaines parties du visage comme pour les vues de trois-quarts. D'autre part, elles amènent des différences de profondeur qui, projetées sur le plan 2D de l'image, provoquent des déformations qui font varier la forme globale du visage. Ces déformations qui correspondent à l'étirement de certaines parties du visage et la compression d'autres régions font varier aussi les distances entre les caractéristiques faciales.[8]



Figure 1-6 Exemples de variation de poses

1.9.3.3 Influence des expressions faciales

Les visages sont des éléments non rigides. Les expressions faciales véhiculant des émotions, combinées avec les déformations induites par la parole, peuvent produire des changements d'apparence importants, et le nombre de configurations possibles est trop important pour que celles-ci soient décrites *in extenso* de façon réaliste. L'influence de l'expression faciale sur la reconnaissance est donc difficile à évaluer avec précision. Cependant, du fait que ce facteur affecte la forme géométrique et les positions des caractéristiques faciales, les techniques globales ou hybrides y sont généralement plus robustes que la plupart des techniques géométriques[8]



Figure 1-7 Exemples de variation d'expressions

1.9.3.4 Influence des occultations partielles

Le visage peut être partiellement masqué par des objets ou par le port d'accessoires tels que des lunettes, un chapeau, une écharpe, *etc.* Les occultations peuvent être intentionnelles ou non. Dans le contexte de la vidéosurveillance, il peut s'agir d'une volonté délibérée d'empêcher la reconnaissance. Il est clair que la reconnaissance sera d'autant plus difficile que peu d'éléments discriminants seront simultanément visibles.[8]

1.10 Conclusion

Dans ce chapitre, nous avons introduit le concept de système biométrique, leur définition et ces caractéristiques. Nous avons défini dans la deuxième partie du chapitre la définition de la reconnaissance faciale et leurs systèmes de reconnaissance faciale ainsi ces techniques locale, global et hybride.

Enfin nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance de visages, notamment l'invariance à l'illumination, les changements de pose et les expressions faciales. Les techniques(CNN) étant utilisées aux différentes étapes de la reconnaissance faciale seront détaillées dans le chapitre suivant.

chapitre 2 : L'apprentissage profond et la reconnaissance faciale

2.1 Introduction

Le deep learning est un type d'intelligence artificielle, dérivé de la machine learning (apprentissage automatique) où la machine est capable d'apprendre par elle-même, contrairement à la programmation, où elle se contente d'exécuter à la lettre des règles prédéterminées.

Dans ce chapitre nous définissons le deep learning (l'apprentissage profond) dont nous présentons ces différentes architectures, nous détaillons ensuite les réseaux de neurones convolutifs tout en expliquant les différentes couches. Par la suite nous citons les domaines d'application de deep learning, enfin, nous expliquons les systèmes de reconnaissance faciale et le système ECNN qui est un système de reconnaissance d'expression faciale.

2.2 Définitions

- **Le réseau neuronal** se situe au croisement de l'informatique et de la biologie. Il est calqué sur le paradigme du cerveau humain dont il démultiplie la puissance, sans lui ressembler tout à fait puisqu'il est dépourvu d'émotion.
- **L'intelligence artificielle** est un champ de recherche qui regroupe l'ensemble des techniques et méthodes qui tendent à comprendre et reproduire le fonctionnement d'un cerveau humain.
- **La machine learning** est un ensemble de techniques donnant la capacité aux machines d'apprendre automatiquement un ensemble de règles à partir de données. contrairement à la programmation qui consiste en l'exécution de règles prédéterminées.
- **Le deep learning ou apprentissage profond** est une technique de machine learning reposant sur le modèle des réseaux de neurones, des dizaines voire des centaines de couches de neurones sont empilées pour apporter une plus grande complexité à l'établissement des règles.

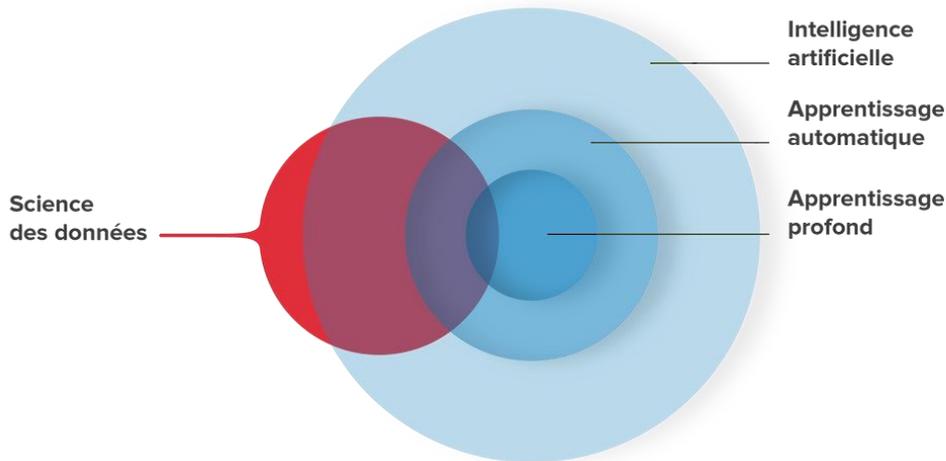


Figure 2-1 la relation entre IA,ML,DL [15].

2.3 Le principe de l'apprentissage automatique

Le principe de l'apprentissage automatique peut être défini de la manière suivante: un programme peut apprendre d'une expérience E à effectuer une tâche T, si sa mesure de performance P, mesuré par la tâche T, s'améliore avec l'expérience E. en d'autre terme le programme peut imiter et généraliser un comportement pour des nouvelles situations similaire à l'expérience E.[10]

La tâche T représente un processus qui regroupe une ou plusieurs fonctions, à savoir la:

- **Régression** : La régression peut être définie comme une méthode ou un algorithme dans l'apprentissage automatique qui modélise une valeur cible basée sur des prédicteurs indépendants. Il s'agit essentiellement d'un outil statistique utilisé pour découvrir la relation entre une variable dépendante et une variable indépendante.[10]
- **Classification** : La classification est un processus de catégorisation d'un ensemble de données en classes. Elle peut être effectuée sur des données structurées ou non structurées. Le processus commence par la prévision de la classe des points de données. Les classes sont souvent appelées cible, étiquette ou catégories (en anglais : Label).[10]
- **Segmentation** : La segmentation est le processus de séparation des données en groupes distincts. L'un des plus grands défis de la création de segments est de déterminer combien de segments existent réellement dans les données[10].

L'expérience E représente l'algorithme de l'apprentissage. Les algorithmes d'apprentissage peuvent être classés en trois catégories principales :

- **L'apprentissage supervisé:** un apprentissage est dit supervisé si les classes sont prédéterminées et les exemples connus dont le système apprend à classer selon un modèle de classification ou de classement. Le processus se passe en deux phases, la première phase est hors ligne c.-à-d. les données sont statiques pendant l'apprentissage. Il s'agit de déterminer un modèle à partir des données étiquetées. La deuxième phase qui est en ligne c.-à-d. les données sont présentés les uns après les autres au fur et à mesure de leur disponibilité, cette phase consiste à prédire une étiquette d'une nouvelle donnée, cette phase est généralement connue sous le nom 'test'[10].
- **L'apprentissage non supervisé:** un apprentissage est dit non supervisé si le système ne dispose d'aucun étiquetage préalable des données, et que le nombre de classes et leur nature n'ont pas été prédéterminés. Aucun expert n'est requis. L'algorithme doit découvrir par lui-même la structure plus ou moins cachée des données. Le partitionnement des données (en anglais : data clustering) est un exemple d'algorithme de l'apprentissage non supervisé[10].
- **L'apprentissage par renforcement:** c'est une méthode d'apprentissage où la machine se comporte comme un agent qui apprend de son environnement d'une manière interactive jusqu'à ce qu'il découvre les comportements qui produisent des récompenses. La figure montre la différence entre l'apprentissage supervisé et non supervisé[10].

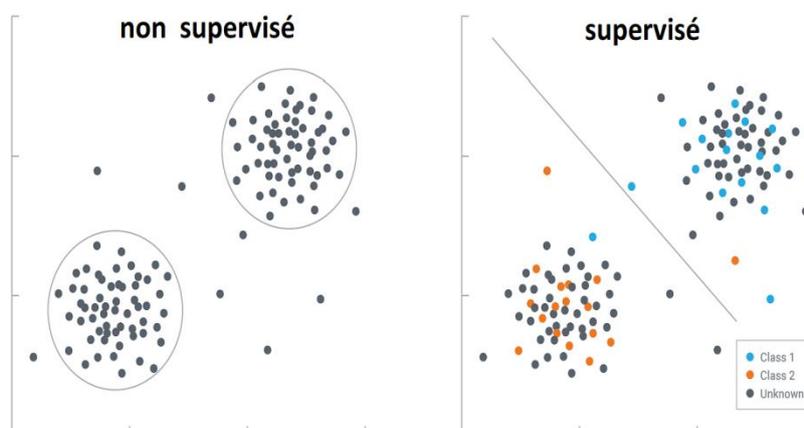


Figure 2-2 L'apprentissage supervisé et non supervisé.[10]

2.4 L'apprentissage profond (deep learning)

L'apprentissage profond (Le deep learning) est un sous-ensemble des techniques de machine learning à base de réseaux de neurones qui s'appuient sur des réseaux de neurones à plusieurs couches dites cachées. Celles-ci permettent par exemple de décomposer de manière hiérarchique le contenu d'une donnée complexe comme de la voix ou une image pour la classifier ensuite : identifier des mots pour la voix ou associer des tags descriptifs à des images.

C'est le principe de l'une des grandes catégories de réseaux de neurones de deep learning, les réseaux convolutionnels ou convolutifs. Un réseau peut être profond mais aussi large si le nombre de neurones est élevé dans chaque couche. Le deep learning remplace les méthodes antérieures de la machine learning dites à base de « hand-craft features » qui consistaient à définir à la main les éléments à rechercher dans les objets (formes dans les images, etc).

Dans le deep learning, notamment pour la détection d'images, le réseau de neurones découvre tout seul ces composantes avec des niveaux d'abstraction évoluant de bas en haut et de couche en couche. Le deep learning sert le plus souvent au traitement du langage, de la parole, du bruit, de l'écriture et des images. Il a d'autres usages dans les outils d'aide à la décision.

2.5 Pour quoi le choix deep Learning

L'apprentissage profond ou le deep learning est généralement choisie afin :

- D'améliorer le développement des algorithmes traditionnels dans de telles tâches de l'IA.
- De développer une grande quantité de données telle que les big data.
- De s'adapter à n'importe quel type de problème
- D'extraire les caractéristiques de façon automatique[5]

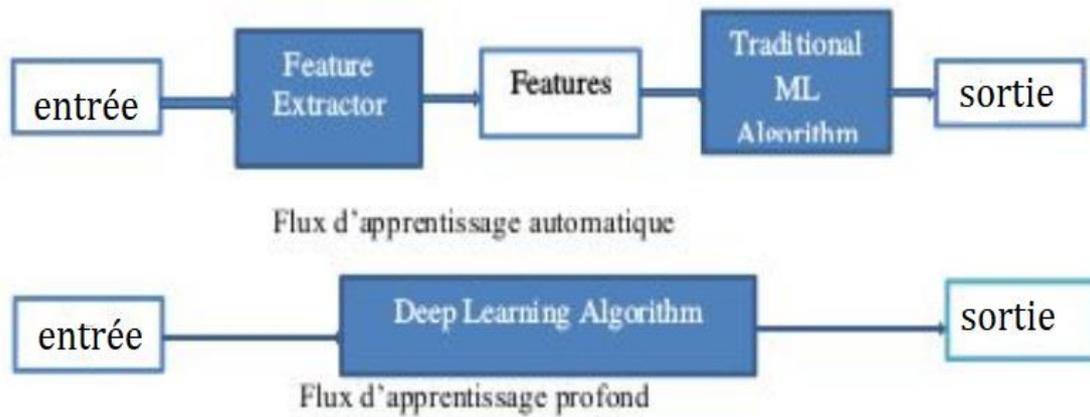


Figure 2-3 comparaison entre machine learning et deep learning [10]

2.6 Les différentes Architectures du Deep Learning :

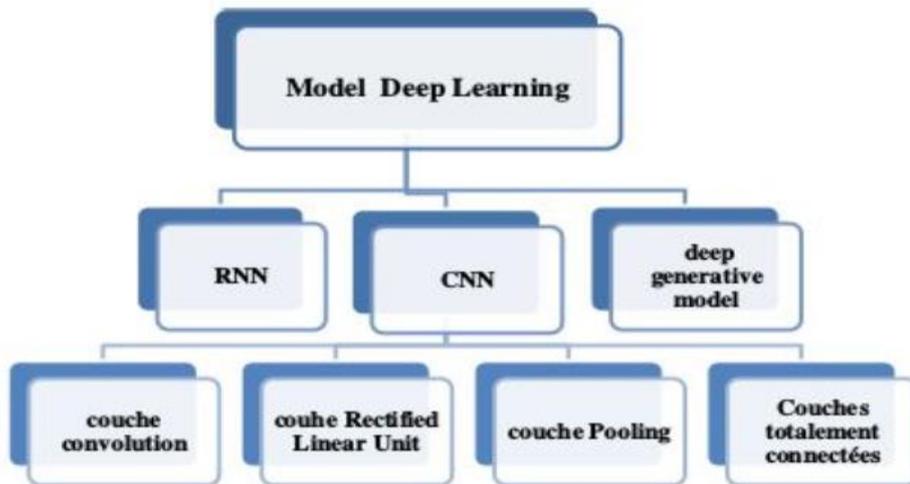


Figure 2-4 Les différentes Architectures du Deep Learning [10]

2.6.1 Réseaux de neurones convolutifs

Un CNN est un réseau neuronal multicouche qui a été biologiquement inspiré par le cortex visuel animal. L'architecture est particulièrement utile dans les applications de traitement d'images. Le premier CNN a été créé par Yann LeCun ; à l'époque, l'architecture se concentrait sur la reconnaissance de caractères manuscrits, comme l'interprétation du code postal. En tant que réseau profond, les premières couches reconnaissent les caractéristiques (telles que les bords) et les couches ultérieures recombinaient ces caractéristiques en attributs de niveau supérieur de l'entrée.

L'architecture CNN est composée de plusieurs couches qui implémentent l'extraction de caractéristiques puis la classification. L'image est divisée en champs récepteurs qui alimentent une couche convolutive, qui extrait ensuite les caractéristiques de l'image d'entrée. L'étape suivante est la mise en commun, qui réduit la dimensionnalité des caractéristiques extraites (par sous-échantillonnage) tout en conservant les informations les plus importantes (généralement, grâce à la mise en commun maximale). Une autre étape de convolution et de mise en commun est ensuite effectuée qui alimente un perceptron multicouche entièrement connecté. La dernière couche de sortie de ce réseau est un ensemble de nœuds qui identifient les caractéristiques de l'image (dans ce cas, un nœud par numéro identifié). Le réseau est formé en utilisant la rétropropagation.

L'utilisation de couches profondes de traitement, de convolutions, de mise en commun et d'une couche de classification entièrement connectée a ouvert la porte à diverses nouvelles applications de réseaux de neurones d'apprentissage en profondeur. En plus du traitement d'image, le CNN a été appliqué avec succès à la reconnaissance vidéo et à diverses tâches dans le cadre du traitement du langage naturel.[13]

2.6.2 Réseaux de neurones récurrents

Le RNN est l'une des architectures réseau fondamentales à partir de laquelle d'autres architectures d'apprentissage en profondeur sont construites. La principale différence entre un réseau multicouche typique et un réseau récurrent est que, plutôt que des connexions entièrement en aval, un réseau récurrent peut avoir des connexions qui renvoient vers les couches précédentes (ou vers la même couche). Cette rétroaction permet aux RNN de conserver la mémoire des entrées passées et de modéliser les problèmes dans le temps.[13]

Les RNN consistent en un riche ensemble. Le différenciateur clé est la rétroaction au sein du réseau, qui peut se manifester à partir d'une couche cachée, de la couche de sortie ou d'une combinaison de celles-ci. Les RNN peuvent être dépliés dans le temps et entraînés avec une rétropropagation standard ou en utilisant une variante de rétropropagation appelée rétropropagation dans le temps (BPTT). [13]

2.6.3 Modèle génératif

Les modèles discriminatifs comme (CNN, RNN) sont utilisés pour prédire les données du label et de l'entrée, tandis que le modèle génératif décrit comment générer les données, il apprend et fait des prédictions en utilisant la loi de Bayes. Les modèles génératifs sont capables de bien plus que la simple classification comme par exemple générer de nouvelles observations[10]

Voici quelques exemples de modèle génératif :

- Boltzmann Machines
- Restricted Boltzmann Machines
- Deep Belief
- Deep Boltzmann Machines
- Generative Adversarial Networks
- Generative Stochastic Networks
- Adversarial autoencoders

2.7 L'architecture de CNN

L'architecture CNN repose sur plusieurs réseaux de neurones profonds consistant en une succession de couches de convolution et d'agrégation (pooling) est dédié à l'extraction automatique de caractéristiques, tandis que la seconde partie, composée de couches de neurones complètement connectés, est dédiée à la classification. Chaque cellule des couches de convolution est connectée à un ensemble de cellules regroupées dans un voisinage rectangulaire sur la couche précédente. Les champs récepteurs locaux permettent d'extraire des caractéristiques basiques.

Les couches sont dites « à convolution » car les poids sont partagés et chaque cellule de la couche réalise la même combinaison linéaire (avant d'appliquer la fonction sigmoïde) qui peut être vue comme une simple convolution. Ces caractéristiques sont alors combinées à la couche suivante afin de détecter des caractéristiques de plus haut niveau. Entre deux phases d'extraction de caractéristiques, le réseau réduit la résolution de la carte des caractéristiques par un moyen de sous-échantillonnage. Cette réduction se justifie à deux titres : diminuer la taille de la couche et apporter de la robustesse par rapport aux faibles distorsions.

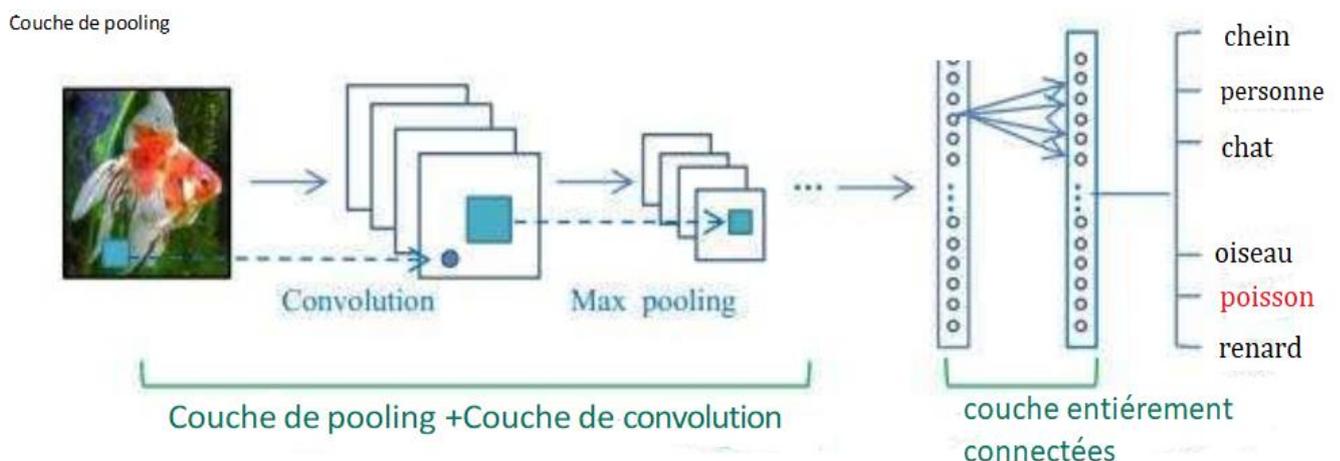


Figure 2-5 L'architecture d'un réseau de neurone convolutionnels[10]

2.7.1 Couche de convolution (CONV)

La couche de convolution est la composante principale des réseaux de neurones convolutifs. Son but est de repérer la présence d'un ensemble des caractéristiques dans les images reçues en entrée. Pour cela, on réalise un filtrage par convolution: le principe est de faire « glisser » une fenêtre représentant la caractéristique sur l'image, et de calculer le produit de convolution entre la caractéristique et chaque portion de l'image balayée et produire une carte de caractéristiques (features map) à sortie.[10]

Avant l'étape de la convolution, il faut déterminer quelques paramètres, ces paramètres font partie du calcul de la taille des images à la sortie de chaque couche (la carte des caractéristiques), Ces paramètres sont :[10]

- **Le pas (Stride) :** Stride est le nombre de pixels décalés sur la matrice d'entrée. Lorsque le pas est de 1, nous déplaçons les filtres à 1 pixel à la fois. Lorsque le pas est de 2, nous déplaçons les filtres à 2 pixels à la fois et ainsi de suite. [10]
- **Le padding :** le padding est le rajout des pixels aux bords de l'image pour pouvoir appliquer la convolution sur les bordures de l'image Le nombre de lignes et de colonnes rajoutés dépend de la taille du filtre de la convolution

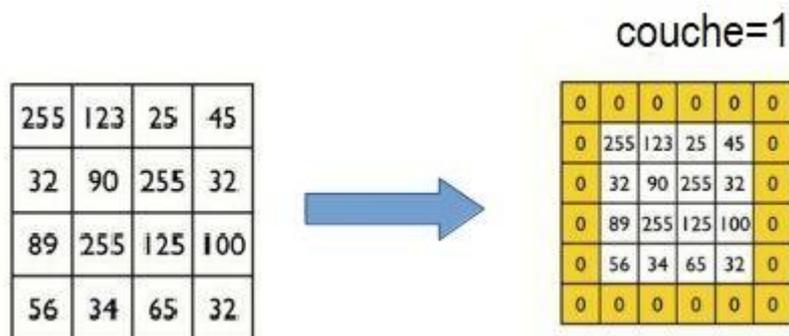


Figure 2-6 Représentation d'un padding. [10]

Le principe de la convolution est illustré sur la figure suivante :

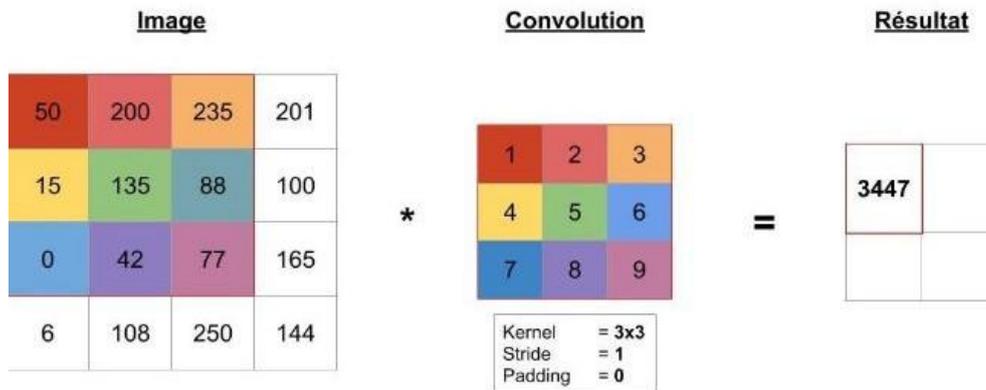


Figure 2-7 Principe de la convolution.[10]

Selon les valeurs présentées dans la figure, le calcul de la convolution est effectué comme suit :

$$50*1+200*2+235*3+15*4+135*5+88*6+0*7+42*8+77*9=3447.$$

A la sortie de chaque couche de convolution la taille de l'image est réduite avec la formule suivante on peut savoir la taille de l'image en sortie d'une convolution à partir de la taille de l'image (W), la taille du filtre (F), le padding (P) et le stride (S):[10]

$$Taille\ sortie\ convolution = 1 + \frac{W - F + 2P}{S}$$

2.7.2 Couche de pooling (POOL)

C'est une méthode mathématique permettant de réduire la taille de l'image sans perdre les informations les plus importantes dans l'image. Son principe consiste à diviser l'image en petites matrices par exemple 2x2 et choisir la valeur maximale de chaque matrice, pour obtenir une image réduite.[10]

La figure suivante montre un exemple de l'application d'un sous échantillonnage 2x2 avec un pas égal à 2.

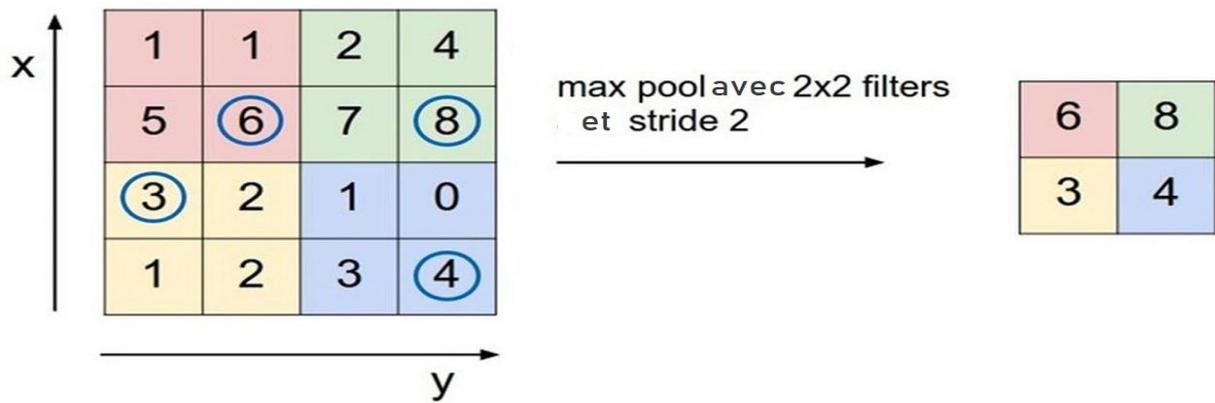


Figure 2-8 Principe de pooling.

- Principe de pooling.
- Couche de correction (ReLU)

Souvent, il est possible d'améliorer l'efficacité du traitement en intercalant entre les couches de traitement une couche qui va opérer une fonction mathématique (fonction d'activation) sur les signaux de sortie. La couche de correction ReLU remplace donc toutes les valeurs négatives reçues en entrées par des zéros[10]

2.7.3 Couche de « entièrement connectée » (FC)

La couche entièrement connectée est un traditionnel perceptron multicouche (Multi Layer Perceptron). Le terme «entièrement connecté » implique que chaque neurone dans la couche précédente est connecté à chaque neurone sur la couche suivante. La représente un exemple d'une couche entièrement connectée [10]

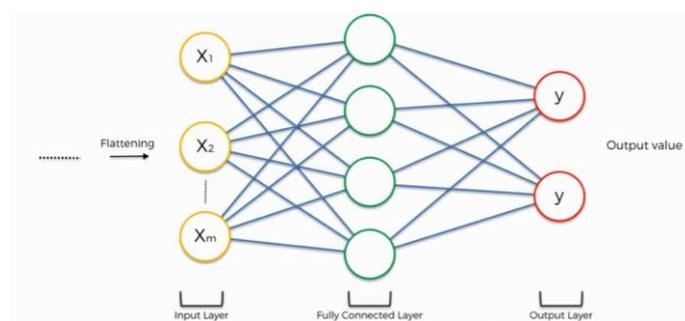


Figure 2-9 Exemple des couches entièrement connectées

La sortie des couches de convolution et de Pooling représente les fonctions de haut niveau de l'image d'entrée. Le but de la couche entièrement connectée est de pouvoir utiliser ces fonctions pour classer l'image d'entrée dans différentes classes en fonction de l'ensemble de données d'apprentissage[10].

2.7.4 Couche de perte (LOSS)

La couche de perte spécifie comment l'entraînement du réseau pénalise l'écart entre le signal prévu et réel. Elle est normalement la dernière couche dans le réseau. Diverses fonctions de perte adaptées à différentes tâches peuvent y être utilisées. La fonction «Softmax» permet de calculer la distribution de probabilités sur les classes de sortie pour la correction des erreurs [10]

2.8 Domaine d'application de Deep Learning

L'apprentissage profond investit progressivement notre quotidien :

- La reconnaissance vocale.
- Le tagging automatique de morceaux de musique.
- La synthèse vocale avancée.
- Deep Boltzmann Machines
- L'étiquetage automatique d'image.
- La conception de nouvelles molécules pharmaceutiques.

Toutes ces applications mettent aujourd'hui en œuvre des techniques de deep learning[5].

2.9 La reconnaissance faciale (Deep Face Recognition) :

Un système de reconnaissance faciale est composé de trois modules :

- Tout d'abord un détecteur de visages est utilisé pour localiser les visages dans les images ou les vidéos.
- Deuxièmement, avec le détecteur de repères faciaux, les visages sont alignés sur des coordonnées canoniques normalisées.
- Troisièmement, le module reconnaissance faciale est implémenté avec ces images de visage alignées.[6]
-

2.9.1 Composantes de la reconnaissance faciale

Un module de reconnaissance de visage se compose d'un traitement de visage, d'une extraction en profondeur des caractéristiques et d'une adaptation de face, et il peut être décrit comme suit:

$M [F(P_i (I_i)), F(P_j (I_j))$ Où :

I_i et I_j : sont deux images de visages, respectivement;

P : représente le traitement des visages pour traiter les variations intra-personnelles, telles que les poses, les illuminations, les expressions et les occlusions ;

F : représente l'extraction de caractéristiques, qui code les informations d'identité.

M : représente un algorithme de correspondance des visages utilisé pour calculer les scores de similarité.[6]

2.9.2 Traitement du visage

Bien que les approches basées sur l'apprentissage en profondeur aient été largement utilisées en raison de leur représentation puissante, ont prouvés que diverses conditions, telles que les poses, les illuminations, les expressions et les occlusions, affectent toujours la performance de la RF en profondeur et que le traitement du visage y bénéficie, particulièrement pour les poses. Puisque la variation de la pose est largement considérée comme un défi majeur dans les applications automatiques, D'autres variations peuvent être résolues par des méthodes similaires. Les méthodes de traitement du visage sont classées en deux catégories: « Augmentation d'un à plusieurs » et « Normalisation de plusieurs à un » [6]

- Augmentation d'un à plusieurs « augmentation one to many »: génère de nombreux patchs ou images de la variabilité de la pose à partir d'une seule image pour permettre aux réseaux profonds d'apprendre des représentations invariantes de la pose.
- Normalisation de plusieurs à un « normalisation of many to one » : récupération de la vue canonique d'images de visage à partir d'une ou plusieurs images d'une vue non frontale ; ensuite, la RF peut être réalisée comme si elle l'était dans des conditions contrôlées[6].

2.9.3 Extraction en profondeur

Les architectures peuvent être classées en réseaux fédérateurs et réseaux assemblés, Inspirées par l'extraordinaire succès remporté par le défi ImageNet, les architectures typiques de CNN, telles que AlexNet, VGGNet, GoogleNet, ResNet et SENet, sont présentées et largement utilisées comme modèle de base en RF directement ou légèrement modifié). En plus du courant dominant, il existe encore quelques architectures nouvelles conçues pour la RF afin d'améliorer les performances. De plus, lorsqu'on adopte les réseaux de base comme blocs de base, les méthodes de RF forment souvent des réseaux assemblés avec des entrées multiples ou des tâches multiples. Un réseau est destiné à un type d'entrée ou à un type de tâche. montrent

qu'il permet d'augmenter les performances après accumulation des résultats des réseaux assemblés.[6]

2.9.4 Correspondance des visages par des traits profonds

Une fois que les réseaux profonds sont formés avec des données massives et une fonction de perte appropriée, chacune des images de test est passée à travers les réseaux pour obtenir une représentation en profondeur des caractéristiques. Une fois les caractéristiques profondes extraites, la plupart des méthodes calculent directement la similarité entre deux caractéristiques à l'aide de la distance cosinus, ensuite le plus proche voisin (PPV) et la comparaison de seuil sont utilisés pour les tâches d'identification et de vérification.

2.9.5 Les travaux existant

Plusieurs travaux ont exploités l'apprentissage profond, nous citons dans cette table quelque uns.

Auteur	Année	Méthode	architecture	nombre de réseaux
Y. Taigman et al [18]	2014	DeepFace	Alexnet	3
Y. Sun et al [19]	2014	DeepID2	Alexnet	25
Y. Sun et al [20]	2015	DeepID3	VGGNet-10	50
F. Schroff et al [21]	2015	FaceNet	GoogleNet-24	1
J. Liu et al [22]	2015	Baidu	CNN-9	10
M. Parkhi et al [23]	2015	VGGface	VGGNet-16	1
X. Wu et al [24]	2015	light-CNN	light CNN	1
Y. Wen et al [25]	2016	Center Loss	Lenet+-7	1
W. Liu et al [26]	2016	L-softmax	VGGNet-18	1
X. Zhang et al [27]	2016	Range Loss	VGGNet-16	1
R. Ranjan et al [28]	2017	L2-softmax	ResNet-101	1
F. Wang et al [29]	2017	Normface	ResNet-28	1
Y. Liu et al [30]	2017	CoCo loss	-	1
M. Hasnat et al [31]	2017	vMF loss	ResNet-27	1
J. Deng et al [32]	2017	Marginal Loss	ResNet-27	1
W. Liu et al [33]	2017	SphereFace	ResNet-64	1
X. Qi et L. Zhang [34]	2018	CCL	ResNet-27	1
F. Wang et al [35]	2018	AMS loss	ResNet-20	1
H. Wang et al [36]	2018	Cosface	ResNet-64	1
J. Deng, J. Guo, et S. Zafeiriou	2018	Arcface	ResNet-100	1
Y. Zheng et al	2018	Ring loss	ResNet-64	1

Ac

Tableau 1 Les travaux réalisés par l'apprentissage en profondeur [6]

2.10 Système ECNN

2.10.1 Présentation

C'est un système de reconnaissance d'expression faciale à partir d'un visage en temps réels qui utilise les réseaux de neurones CNN, il permet de détecter un visage d'une personne à partir d'une image ou via une caméra pour connaître l'expression avec un taux de précision associé au sept expressions universel à savoir la joie, le dégoût, la peur, la colère, la tristesse, la surprise et le neutre.

2.10.2 Architecteur du système ECNN

Le système de reconnaissance des expressions faciales est effectué en trois étapes principales :

- Détection du visage et des parties du visage.
- Extraction des caractéristiques.
- Classification des expressions.

Dans ce qui suit, nous détaillerons chacune des étapes du système ECNN[5].

2.10.2.1 La détection de visage

L'efficacité de système de reconnaissance faciale basée sur l'authentification faciale dépend essentiellement de la méthode utilisée pour localiser le visage dans l'image. L'algorithme Viola-Jones est utilisé pour détecter diverses parties du visage humain telles que la bouche, les yeux, le nez, les narines, les sourcils, la bouche, les lèvres, les oreilles, etc. Cet Algorithme explore les caractéristiques de type Haar via le classificateur Cascade, qui peut efficacement combiner de nombreuses fonctionnalités et déterminer les différents filtres sur un classificateur résultant .[5]

La méthode de Viola et Jones, la première méthode capable de détecter des visages en temps-réel. La méthode devient standard et est reprise et améliorée par de nombreux chercheurs. C'est également à partir des années 2000 que les méthodes basées apparence, qui apprennent des modèles de visage à partir d'un ensemble d'images d'apprentissage, se révèlent généralement supérieures aux autres types d'approche.[14]

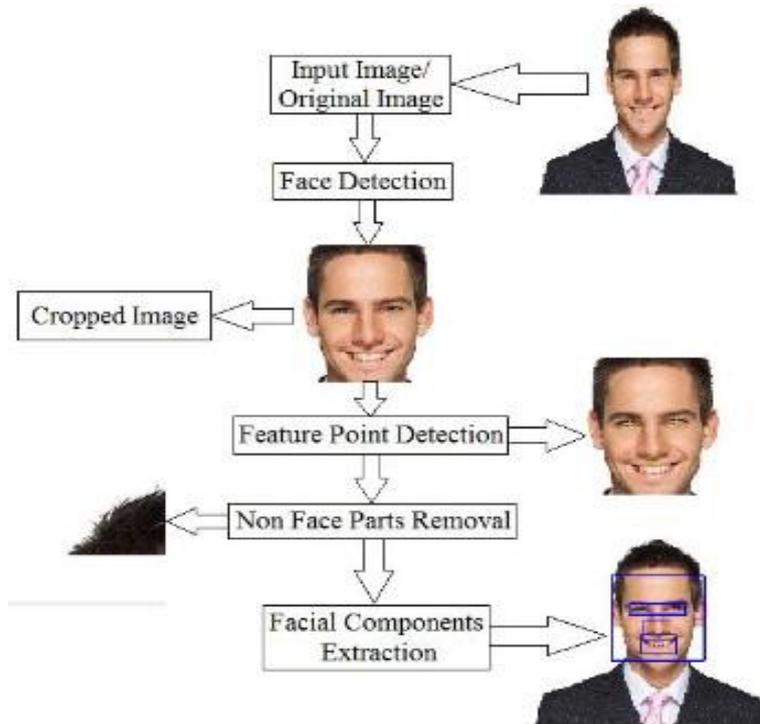


Figure 2-10 principe Viola et Jones [5]

2.10.2.2 Extraction de caractéristiques faciales

Si le visage est détecté dans l'image, le système lance le processus d'extraction des caractéristiques qui va convertir les données des pixels à des représentations et configuration plus réduite et optimal pour que la représentation extraite soit utilisé dans le processus de la classification.[5]

2.10.2.3 Classification des expressions

L'objectif de l'étape de classification est la reconnaissance de l'expression faciale en fonction des caractéristiques extraites. Cette étape est basée sur l'utilisation de la technologie de réseau de neurones convolutionnel (CNN).[5]

2.11 Conclusion

Dans ce chapitre, nous avons défini le deep learning, ainsi que ses différentes architecteurs. Nous avons focalisé notre attention sur les réseaux de neurones convolutifs CNN et leurs structures, et ses différentes couches, nous avons justifié le choix de deep learning puis quelques domaines de l'utilisation du deep learning. Nous avons ensuite présenté les systèmes de reconnaissance faciale.

Dans le chapitre suivant nous allons expliquer les étapes de conception et d'implémentation de notre système de détection et de reconnaissance faciale dans une séquence d'image.

chapitre 3 : Conception et implémentation

3.1 Introduction

L'un des principaux objectifs de la détection et de la reconnaissance automatiques des visages est de résoudre le festin de problèmes, par exemple la perte d'enfants ou la détection de criminels, en particulier dans les opérations de vol. Il est nécessaire d'identifier les visages à partir de photographies prises par des appareils de surveillance. Il faut donc extraire les caractéristiques de l'image afin qu'elles puissent être utilisées par d'autres parties du système,

Dans ce chapitre, nous allons présenter les différentes étapes de la conception et la mise en œuvre de notre système qui concerne la détection et la reconnaissance des visages. La dernière partie du chapitre sera consacrée à l'illustration des résultats obtenus.

3.2 Conception générale de notre système :

L'architecture globale de notre système de reconnaissance faciale est présentée dans le schéma suivant:

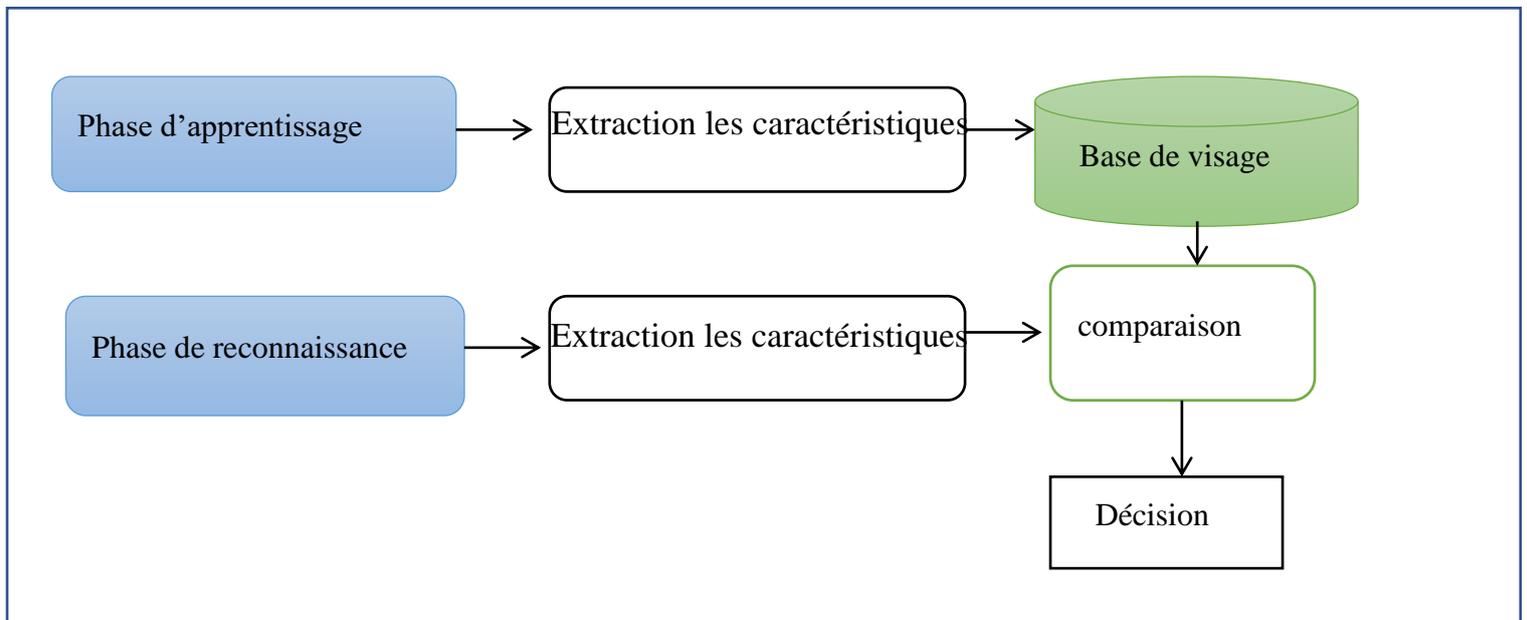


schéma 1 Conception général de système

3.3 Conception détaillé

Dans ce paragraphe nous allons présenter la conception détaillée de notre système, dont nous allons expliquer les composants principaux du système à savoir l'apprentissage, la détection et la reconnaissance de visage ainsi le suivi de ce dernier.

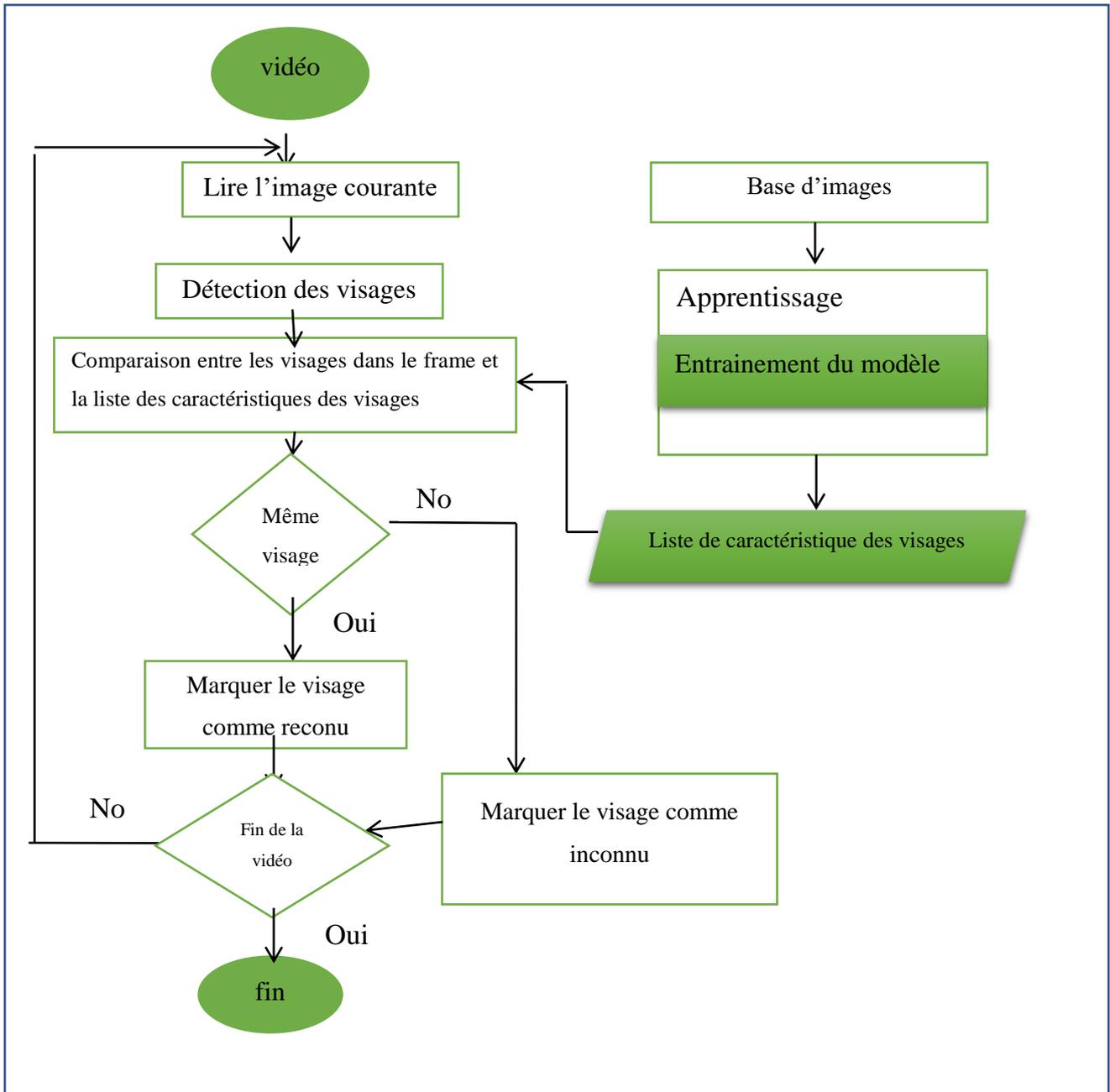


Schéma 2 Conception détaillée de système

3.3.1 L'apprentissage du modèle

Dans ce paragraphe nous allons détailler les différentes étapes d'apprentissage, dans lesquelles le système va faire un entraînement pour qu'il apprenne à détecter et reconnaître des visages. Pour ce faire nous allons utiliser un modèle de réseaux de neurones convolutifs. L'entrée de ce modèle est un ensemble d'image contenant des visages et la sortie est un vecteur de caractéristiques de chaque visage.

3.3.2 Importer des bibliothèques

La première étape consiste à construire le programme. La bibliothèque contient un ensemble de fonctions dont nous avons besoin pour résoudre le problème. Le simple fait de l'écrire dans le programme signifie appeler ses fonctions.

3.3.3 Importer les ensembles de données

De nombreux ensembles de données sont présentés dans différents formats. Nous aurons besoin de données telles que la vidéo et un ensemble d'images pour effectuer le traitement avec l'emplacement de chaque fichier pour pouvoir y accéder.

3.3.4 Entraînement du model (Training)

Il s'agit de l'étape la plus importante de notre système, nous utiliserons un modèle de réseaux de neurones à convolution avec une architecture spécifique pour obtenir le meilleur résultat de précision et moins de perte de données, et ce qui suit peut être utilisé pour enregistrer l'image de reconnaissance faciale, la formation étape peut être divisée en deux étapes

```
face_recognition.api.compare_faces( codages_face_connus , codage_face_à_vérifier , tolérance=0.6 )
```

Comparez une liste d'encodages de visage à un encodage candidat pour voir s'ils correspondent.

Paramètres:

- **known_face_encodings** – Une liste d'encodages de visage connus
- **face_encoding_to_check** - Un seul encodage de visage à comparer à la liste
- **tolérance** - Quelle est la distance entre les faces pour la considérer comme une correspondance. Plus bas est plus strict. 0,6 est la meilleure

performance typique.

Retour: Une liste de valeurs Vrai/Faux indiquant quels codages_de_visage_connus correspondent au codage de visage à vérifier

```
face_recognition.api.face_encodings( face_image , known_face_locations=None , num_jitters=1 , model='small' )\[la source\]
```

Étant donné une image, renvoie le codage de visage à 128 dimensions pour chaque visage de l'image.

- Paramètres:**
- **face_image** – L'image qui contient un ou plusieurs visages
 - **Known_face_locations** – Facultatif - les boîtes englobantes de chaque visage si vous les connaissez déjà.
 - **num_jitters** - Combien de fois rééchantillonner le visage lors du calcul de l'encodage. Plus haut est plus précis, mais plus lent (c'est-à-dire 100 est 100 fois plus lent)
 - **model** – Facultatif - quel modèle utiliser. "grand" ou "petit" (par défaut) qui ne renvoie que 5 points mais est plus rapide.

Retour: Une liste d'encodages de visage à 128 dimensions (un pour chaque visage de l'image)

```
face_recognition.api.load_image_file( fichier , mode='RGB' )\[la source\]
```

Charge un fichier image (.jpg, .png, etc.) dans un tableau numpy

- Paramètres:**
- **file** – nom du fichier image ou objet fichier à charger
 - **mode** – format vers lequel convertir l'image. Seuls 'RGB' (RVB 8 bits, 3 canaux) et 'L' (noir et blanc) sont pris en charge.

Retour: contenu de l'image sous forme de tableau numpy

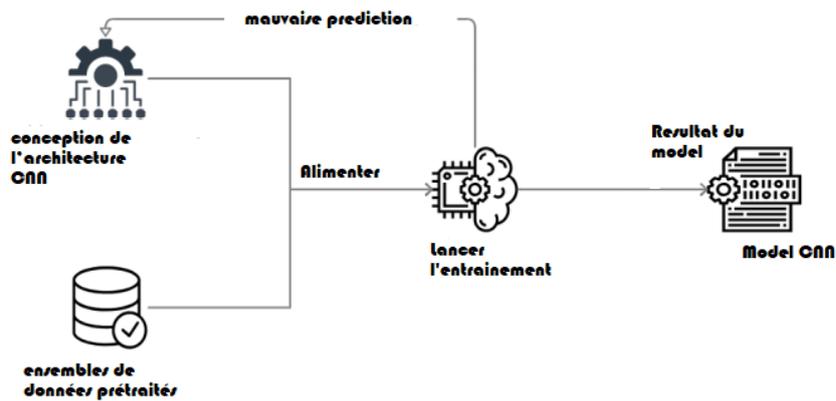


Figure 3-1 Cycle de la vie d'entraînement

3.3.5 La détection de visage

La détection consiste à sélectionner dans chaque frame l'ensemble des visages qu'elle contient ceci va nous permettre de faciliter l'étape de reconnaissance. L'entrée donc de ce module est une image ou frame et la sortie est l'ensemble de visages sélectionnés.

3.3.6 La reconnaissance de visage

Dans ce module nous faisons une comparaison entre chaque visage sélectionné et les caractéristiques des visages de la base. Si le visage est reconnu alors il sera marqué avec le nom de la personne sinon il sera marqué comme inconnu.

En répétant ces étapes pour chaque frame la personne reconnue sera suivi dans toute la séquence d'image (la vidéo).

3.3.7 Extraction les caractéristique :

L'apprentissage de classificateur d'images de réseau neuronal convolutionnel avec Keras Après la préparation des données, on a procéder à l'apprentissage en utilisant l'architecture CNN simple de xception, et on a utilisé une classe avec le nom `cnn.py` qui implémente l'architecture xception du CNN. Nous commençons à former le classifieur d'images en utilisant l'apprentissage profond, avec Keras et Tensorflow. Nous créons un modèle Séquentiel en transmettant une liste d'instances de couche au constructeur. Nous effectuons ensuite un partage d'apprentissage et tests sur les données en utilisant 75%des images pour l'apprentissage et 25% pour les tests.[39]

3.3.8 Formation du modèle avec keras

1. Importer les bibliothèques installées
2. Importer le type de modèle séquentiel de Keras qui il s'agit d'un empilement linéaire de couches de réseaux de neurones à travers la commande : `from keras.models import Sequential`.
3. Importer les couches "principales" de Keras qui sont utilisées dans presque tous les réseaux de neurones : `From keras.layers import Dense, Dropout, Activation, Flatten`
4. Importer les couches CNN de Keras. Ce sont les couches convolutives qui nous aideront à former notre modèle : `from keras.layers import Convolution2D, MaxPooling2D`. [39]

3.3.9 Présentation l'architecture Xception

Xception est une structure moderne proposé par François Chollet lui-même, créateur et responsable de la maintenance de la bibliothèque Keras de Google. Xception est une extension de l'architecture Inception [11] qui remplace les modules Inception standard par des convolutions séparables en profondeur ce qui permet de diminuer la taille de l'architecture jusqu'au 91Mo. [39]

3.4 Définition de la Convolution séparable

C'est une solution au problème de la complexité des calculs au niveau de

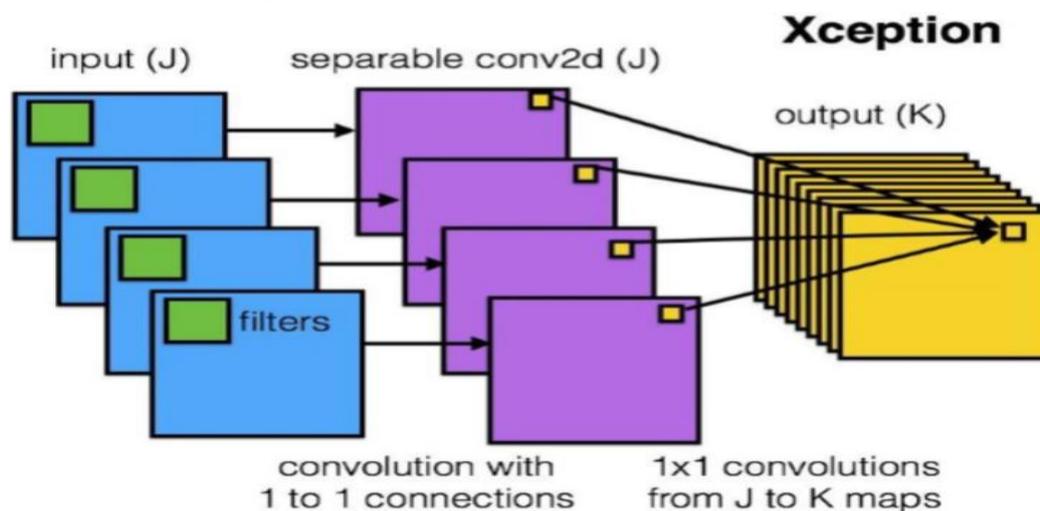


Figure 3-2 l'architecture Xception [39]

convolution qui rend le fonctionnement de réseau un peu long, avec moins de calculs, le réseau est en mesure de traiter davantage de données en un temps réduit. Il existe deux types de convolutions séparables. [39]

– La convolution séparable spatiale :

C'est la solution la plus facile parce qu'elle consiste à diviser la couche de convolution en deux, elle traite principalement des dimensions spatiales d'une image et du noyau : la largeur et la hauteur. Malheureusement elle est très limitée par conséquent elle n'est pas très utilisée[39] dans l'apprentissage en profondeur.

– La convolution séparable en profondeur :

C'est la solution qui nous intéresse dans notre étude car elle traite non seulement des dimensions spatiales, mais également de la dimension en profondeur c'est-à-dire le nombre de canaux. Une image d'entrée peut avoir 3 canaux : RVB. Après quelques convolutions, une image peut avoir plusieurs canaux. Semblable à la convolution spatiale séparable, une convolution séparable en profondeur divise un noyau en 2 noyaux distincts qui effectuent deux convolutions : La convolution en profondeur et la convolution ponctuelle (taille de 1x1).

3.5 La convolution séparable en profondeur dans Xception

La convolution modifiée dans la structure Xception est la convolution ponctuelle suivie d'une convolution en profondeur. Selon lequel, une convolution 1×1 est effectuée avant toutes les convolutions spatiales de taille $n \times n$. Ainsi, c'est un peu différent de l'original. Le modèle Xception atteint une précision de 94,5% sur la base de données ImageNet. Quelques modèles ont réduit l'ensemble des paramètres dans leurs dernières couches en incluant une fonction Globale de pooling moyenne « Opération Global de pooling » comme dans Xception, ce qui réduit l'ensemble des cartes de caractéristiques en une valeur scalaire en prenant la moyenne de tous les éléments de la carte. Nous proposons d'éliminer les couches entièrement connectées dans tous les CNN de notre système et nous incluons la fonction Globale de pooling moyen de tel sorte, nous obtenons à la sortie des réseaux un nombre des cartes de caractéristiques égale au nombre de classe (07 ; les six expressions et l'état neutre). Ensuite, nous appliquons la fonction d'activation softmax, qui est le même principe du modèle Xception.[39]

3.6 Implémentation

Nous allons décrire l'implémentation des différentes étapes de notre système conçu dans le paragraphe précédent. Nous commençons par la justification de

l'environnement et les outils de développement utilisés, ensuite nous présentons l'algorithme principal nécessaire à l'implémentation de notre système.

3.6.1 Environnements et Outils de développement

Pour développer notre système, nous avons utilisé différents environnements et outils pour le langage de programmation y compris des API, bibliothèques, langage de balisage etc...

3.6.2 Environnement de développement

Notre système est développé sous l'environnement :

Ordinateur portable :

Processeur Intel(R) Core(TM) CPU 7200U @ 2.50 GHz 2.70 GHz .

RAM (4Go), HDD (500Go), Carte graphique (1Go).

Système d'exploitation : Microsoft Windows 10 (64 bits).

Spécifications de l'appareil

Nom de l'appareil	DESKTOP-3I0VVBI
Processeur	Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz 2.70 GHz
Mémoire RAM installée	4,00 Go (3,79 Go utilisable)
ID de périphérique	A40F6F56-E78B-43D5-9639- C986A75BB0B1
ID de produit	00329-00000-00003-AA764
Type du système	Système d'exploitation 64 bits, processeur x64
Stylet et fonction tactile	La fonctionnalité d'entrée tactile ou avec un stylet n'est pas disponible sur cet écran

Figure 3-3 Environnement de développement.

3.6.2.1 Python

Python est le langage de programmation open source le plus employé par les informaticiens. Ce langage s'est propulsé en tête de la gestion d'infrastructure, d'analyse de données ou dans le domaine du développement de logiciels. En effet, parmi ses qualités, Python permet notamment aux développeurs de se concentrer sur ce qu'ils font plutôt que sur la manière dont ils le font. Il a libéré les développeurs des contraintes de formes qui occupaient leur temps avec les langages plus anciens. Ainsi, développer du code avec Python est plus rapide qu'avec d'autres langages.[14]

Il reste aussi accessible pour les débutants, à condition de lui consacrer un peu de temps pour la prise en main. De nombreux tutoriels sont d'ailleurs disponibles pour l'étudier sur des sites Internet spécialisés ou sur des comptes Youtube. Sur les forums d'informatique, il est toujours possible de trouver des réponses à ses questions, puisque beaucoup de professionnels l'utilisent.

3.6.2.2 Open cv

OpenCV (Open Source Computer Vision Library) est une bibliothèque de logiciels open source de vision par ordinateur et d'apprentissage automatique. OpenCV a été conçu pour fournir une infrastructure commune pour les applications de vision par ordinateur et pour accélérer l'utilisation de la perception par machine dans les produits commerciaux. En tant que produit sous licence BSD, OpenCV permet aux entreprises d'utiliser et de modifier facilement le code.

La bibliothèque contient plus de 2500 algorithmes optimisés, qui comprennent un ensemble complet d'algorithmes de vision par ordinateur et d'apprentissage automatique classiques et à la pointe de la technologie. Ces algorithmes peuvent être utilisés pour détecter et reconnaître des visages, identifier des objets, classer des actions humaines dans des vidéos, suivre des mouvements de caméra, suivre des objets en mouvement, extraire des modèles 3D d'objets, produire des nuages de points 3D à partir de caméras stéréo, assembler des images pour produire une haute résolution image d'une scène entière, trouver des images similaires dans une base de données d'images, supprimer les yeux rouges des images prises au flash, suivre les mouvements des yeux, reconnaître le paysage et établir des marqueurs pour le superposer avec la réalité augmentée, etc. OpenCV compte plus de 47 000 utilisateurs communauté et nombre estimé de téléchargements dépassant 18 millions. La bibliothèque est largement utilisée dans les entreprises, les groupes de recherche et les organismes gouvernementaux.

3.6.2.3 Face recognition

La bibliothèque `face_recognition` est construite sur des techniques d'apprentissage en profondeur et n'utilise qu'une seule image d'entraînement.

Comment fonctionne la reconnaissance faciale ?

Passez la photo d'une personne avec son nom au modèle.

Le modèle prend chaque image et après les avoir convertis en codage numérique, les stocke dans une liste avec les étiquettes dans une autre liste.

- Lors de la réalisation d'une prédiction, le modèle convertit à nouveau l'image d'entrée en un codage.
- Cette image encodée est utilisée pour trouver des encodages similaires en fonction de la distance. Les encodages les moins éloignés seront ceux qui correspondent le mieux à la personne.
- Une fois que la correspondance la plus proche est trouvée, le nom de la personne détectée finale est produit.

3.6.2.4 Os :

C'est un module fournit par Python dont le but d'interagir avec le système d'exploitation, il permet ainsi de gérer l'arborescence des fichiers, de fournir des informations sur le système d'exploitation processus, variables systèmes, ainsi que de nombreuses fonctionnalités du système.

3.6.2.5 TensorFlow :

Est une bibliothèque open source de Machine Learning, permettant de développer et d'exécuter des applications de Machine Learning et de Deep Learning, Créé par l'équipe Google Brain en 2011, sous la forme d'un système propriétaire dédié aux réseaux de neurones de Deep Learning. En 2015, il a été renommé TensorFlow et Google l'a rendu open source. Depuis lors, il a subi plus de 21000 modifications par la communication et est passé en version 1.0 en février 2017 .

TensorFlow est une bibliothèque de Machine Learning, il s'agit d'une boîte à outils permettant de résoudre des problèmes mathématiques extrêmement complexes avec aisance. Elle permet aux chercheurs de développer des architectures d'apprentissages expérimentaux et de les transformer en logiciels .

TensorFlow regroupe un grand nombre de modèles et d'algorithmes de Machine Learning et de Deep Learning. Son API front-end de développement

d'applications repose sur le langage de programmation Python, tandis que l'exécution de ces applications s'effectue en C++ haute-performance .

Cette bibliothèque permet notamment d'entraîner et d'exécuter des réseaux de neurones pour la classification de chiffres écrits à la main, la reconnaissance d'image, les plongements de mots, les réseaux de neurones récurrents, les modèles séquence-to-séquence pour la traduction automatique, ou encore le traitement naturel du langage .

3.6.2.6 Keras

Est une librairie Python qui encapsule l'accès aux fonctions proposées par plusieurs librairies de machine learning, en particulier Tensorflow. De fait, Keras n'implémente pas nativement les méthodes. Elle sert d'interface avec Tensorflow simplement [43].

Keras est le cadre d'apprentissage profond le plus utilisé parmi les 5 meilleures équipes gagnantes sur Kaggle. Parce que Keras rend plus facile de mener de nouvelles expériences.

3.6.2.7 La base de données

Pour faire l'apprentissage nous avons besoin d'une base d'image contenant un ensemble de visages, notre base de données est constituée d'un ensemble de photos prises avec l'appareil photo ou télécharger par internet.

3.7 Algorithmes

Dans cette section nous allons présenter le principe de l'algorithme de détection, reconnaissance et suivi de visage :

Algorithme principale

Début

Pour chaque image dans la base de données faire

 Enregistrer les caractéristiques du visage

Fin pour

Lire vidéo

Pour chaque frame dans la vidéo

 Détecter le visage de chaque personne dans le frame

 Calculer la distance entre les caractéristiques du visage détecté et les caractéristiques des visages de la base

Si (la distance est trop petite)

 Reconnaitre le visage et écrire le nom de la personne (non fichier)

sinon

 Ecrire inconnu sur le visage détecté

fin si

fin pour

fin

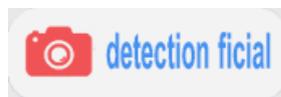
3.8 Présentation de système

On présente dans cette section les différents aspects du notre système de reconnaissance.

3.8.1 Interface de notre système



Figure 3-4 Interface graphique de notre application



Ce bouton permet de faire la reconnaissance et le suivi de visage en temps réel dans ce cas la séquence d'image est captée à l'aide d'une caméra



Ce bouton permet de faire la reconnaissance et le suivi de visage en temps passif, dans ce cas la séquence d'image est une vidéo enregistrée d'avance



Pour extraire les caractéristiques en utilise ce bouton



Supprimer les données

3.8.2 Résultats et discussion

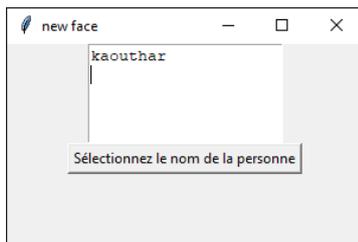
Dans ce qui suit nous allons présenter le fonctionnement du système et illustrer les résultats obtenus.

3.8.3 Suivi du visage en temps réel

Dans ce cas il faut activer la camera,

Si la base de données est vide il faut la remplir en ajoutant des images de personnes

➤ Ajouter le nom de personne



Si nous cliquons sur le bouton sélectionnez le nom de la personne alors la caméra prend des photos de la personne.

```
25 def printInput():
26     inp = inputtxt.get(1.0, "end-1c")
27     #lbl.config(text = "Provided Input: "+inp)
28
29     foldername = 'app/facecam/' + inp
30
31     if not os.path.exists(foldername):
32         os.makedirs(foldername)
33
34         original = r'app/facecam/person1.jpg'
35         imgnamenew = foldername + '/person1.jpg'
36         target = r'' + imgnamenew
37
38         shutil.move(original, target)
39
40
41
42 # TextBox Creation
43 def createface():
44
45     # Button Creation
46     printButton = tk.Button(uiwind, text = "Sélectionnez le nom de la personne", command = printInput)
47     printButton.pack()
48
```

➤ Faire la reconnaissance et le suivi en temps réel (voir le paragraphe suivant).

3.8.4 Suivi du visage en temps passive

- Lancer la vidéo sur laquelle on souhaite détecter un visage
- Faire la reconnaissance de ce visage dans chaque frame de la vidéo (voir le paragraphe suivant).

3.8.4.1 Explication du code

La reconnaissance et le suivi dans notre système sont réalisés par les étapes suivantes :

- **L'apprentissage par deep learning et la bibliothèque face recognition**

```
known_faces = []
known_names = []
known_faces_paths = []

registered_faces_path = 'app/facecam/'
for name in os.listdir(registered_faces_path):
    images_mask = '%s%s/*.jpg' % (registered_faces_path, name)
    images_paths = glob.glob(images_mask)
    known_faces_paths += images_paths
    known_names += [name for x in images_paths]

def get_encodings(img_path):
    image = face_recognition.load_image_file(img_path)
    encoding = face_recognition.face_encodings(image)
    return encoding[0]

known_faces = [get_encodings(img_path) for img_path in known_faces_paths]
```

- **Lire le vidéo frame par frame**

```
while True:
    ret, frame = vc.read()
    ret2, frameOriginal = vc.read()
    if not ret:
        break
    frame_rgb = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)
    faces = face_recognition.face_locations(frame_rgb)
    for face in faces: # top, right, bottom, left
        top, right, bottom, left = face
        cv2.rectangle(frame, (left, top), (right, bottom), (0, 0, 255), 2)
        face_code = face_recognition.face_encodings(frame_rgb, [face])[0]

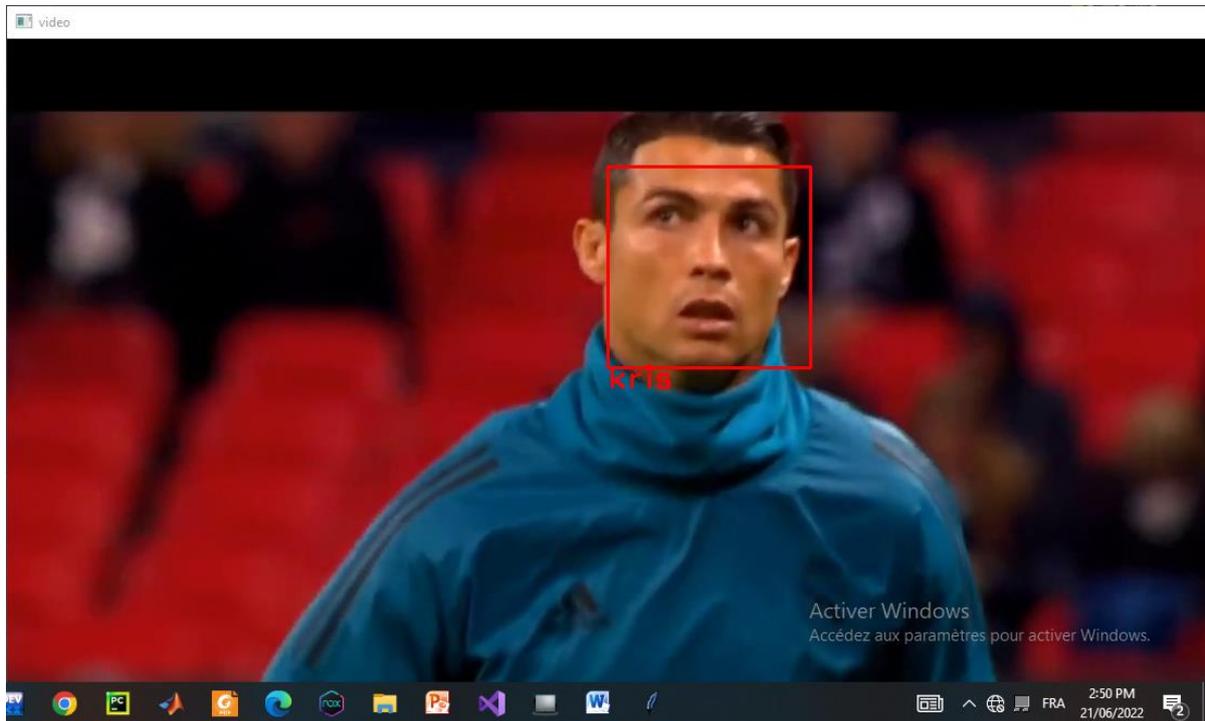
        results = face_recognition.compare_faces(known_faces, face_code, tolerance=0.6)
        if any(results):
            name = known_names[results.index(True)]
        else:
            #name = 'unknown'
            foldername = 'app/facecam/'
            imgname = foldername + 'person1.jpg'
            cv2.imwrite(imgname, frameOriginal)
```

- **Détecter toute les visages dans chaque frame et faire la comparaison**

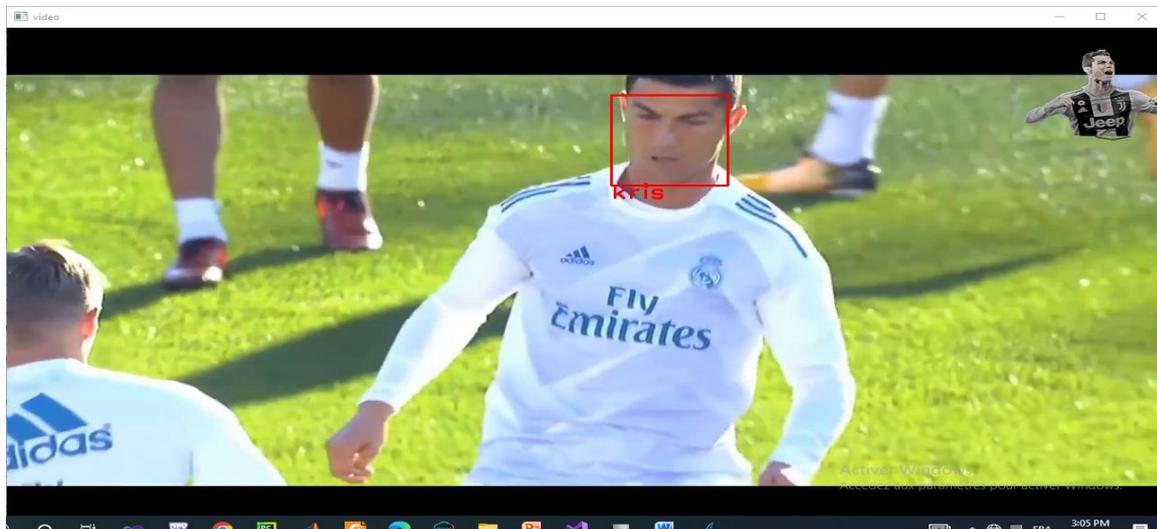
Dans chaque frame les visages sont sélectionnés et une comparaison est faite entre le visage et les visages de la base de données, s'ils sont identiques, on écrit son nom et on suit le visage. S'ils ne sont pas identique on écrit inconnu.

3.9 Résultat

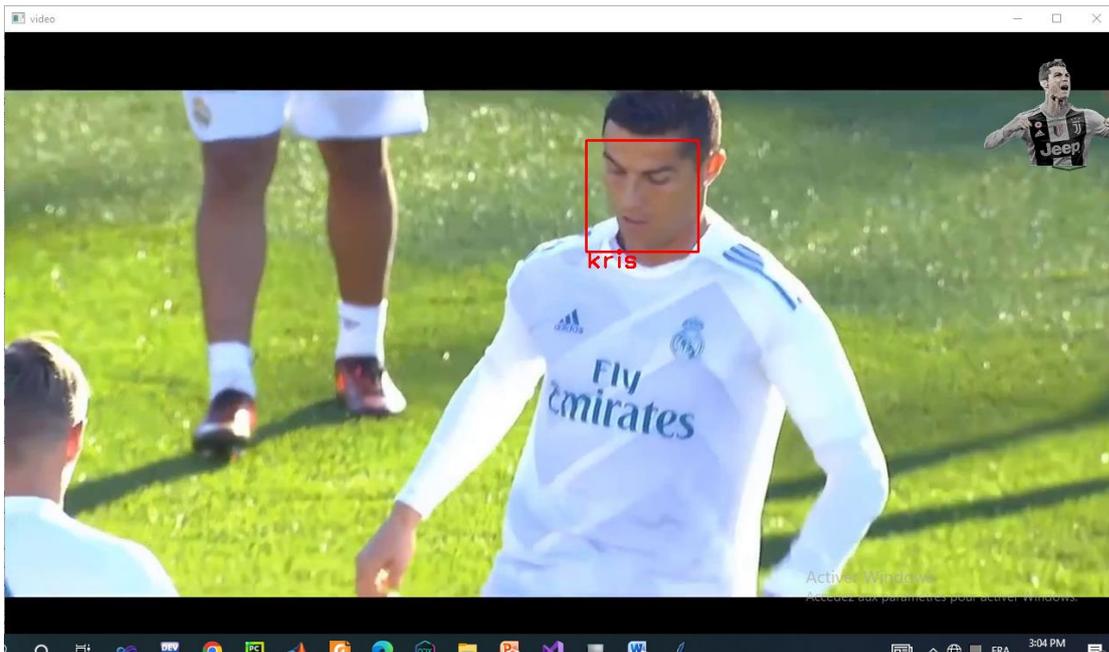
Le résultat Pour le premier frame :



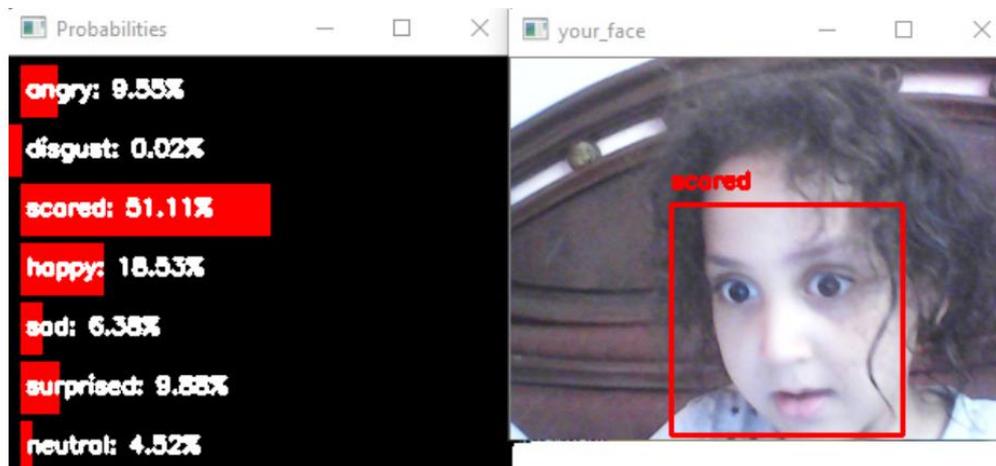
Dans le deuxième frame :

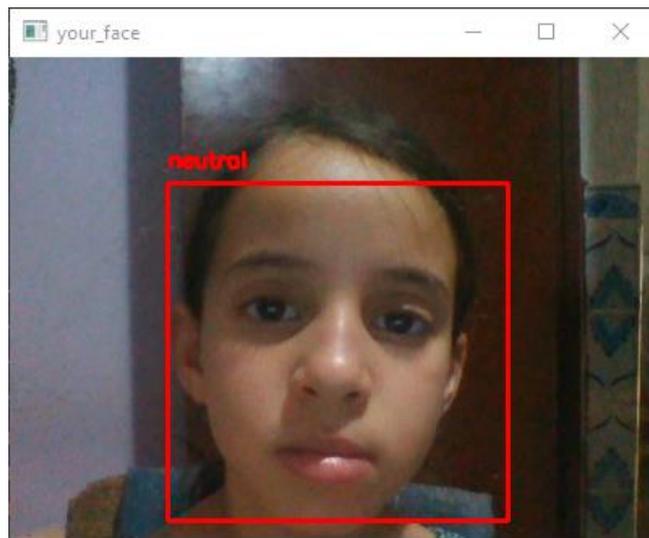


Pour le frame suivant :

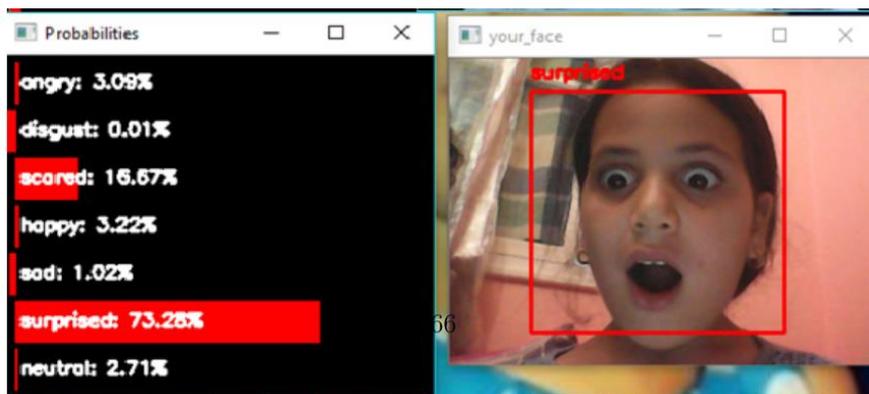


Pour les expressions faciales :





Naural



surprised

- Dans les résultats : détecter les visages et créer les émotions chaque photos contenu une émotion

3.10 Conclusion

Ce chapitre présente la méthodologie qui a été suivie lors du développement de ce projet, et explique en détail comment développer le système de reconnaissance des visages, tout en décrivant en détail la mise en place de chacune des différentes étapes du système, qui commence par un ensemble de images pour apprendre, puis nous insérons une vidéo et reconnaissons les personnes

Conclusion générale

La biométrie est un domaine à la fois Intéressant 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 faciale.

Partant de ce principe, nous avons conçu et implémenté un système de détection et de reconnaissance faciale dans une séquence d'images, en utilisant la technique de réseau de neurone convolutionnels.

Durant notre étude, nous nous sommes attachés à :

- Donner une vue générale sur la reconnaissance faciale dont nous avons présenté ses concepts de base.
- Faire une étude sur l'apprentissage profond et son exploitation dans reconnaissance faciale
- Concevoir et implémenter un système permettant de faire la détection et la reconnaissance faciale dans une séquence d'image basé sur la technique CNN. La détection du visage dans chaque frame de la vidéo nous a permis de faire le suivi de la personne reconnue. Nous avons pu ajouter aussi un module de détection des expressions faciales.

Le système que nous avons proposé est un noyau d'un système de détection et de reconnaissance faciale, il peut être amélioré par l'ajout de plusieurs composantes telles que :

- Le suivi d'une personne suspecte dans les grandes surfaces en faisant des captures des visages des visiteurs dès leurs entrées.
- La découverte de l'identité d'un criminel, en intégrant une base de données contenant les criminels connus dans une ville.

Références

- [1] <http://thesis.univ-biskra.dz/944/4/Chap%201%20Syst%C3%A8me%20RV%20sept%2012.pdf> 19-06-2022
- [2] <https://www.lemagit.fr/definition/Biometrie> 19-6-2022
- [3] <https://www.vie-publique.fr/questions-reponses/271221-sept-questions-sur-la-reconnaissance-faciale> 19-06-2022
- [4] <https://www.journaldunet.fr/business/dictionnaire-du-marketing/1208194-reconnaissance-faciale-definition-et-traduction/> 19-06-2022
- [5] Mémoire de Master Azzoune , I., & Kheldoun, N. (2020). Une approche IA pour la reconnaissance des. Bouira.
- [6] Mémoire de master Bouzit , D. (2019). Reconnaissance de visage basée sur une approche. Guelma.
- [7] Mémoire de master CHETTOUH , H. (2020). Montage d'un Système de Reconnaissance des Expressions Faciales avec le Deep Learning. Biskra.
- [8] Thèse de doctorat HAMOUDI, L. (2011). Application de techniques d'apprentissage pour. LILLE.
- [9] Mémoire de Master WASSIM, M. (2021). Contrôle d'accès des véhicules par la reconnaissance automatique de leur. Biskra.
- [10] Mémoire de Master AKCHA, I., & AMMARI, A. (2020). Développement d'un système de reconnaissance faciale. blida.
- [11] <https://www.thalesgroup.com/fr/europe/france/dis/gouvernement/inspiration/biometrie> 19.06.2022
- [12] <https://medium.com/@appstud/lab-appstud-reconnaissance-de-visages-grace-au-machine-learning-85a612b368db> 19.06.2022
- [13] <https://developer.ibm.com/articles/cc-machine-learning-deep-learning-architectures/> 19-06-2022 c
- [14] <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1445304-python-definition-et-utilisation-de-ce-langage-informatique/>
- [15] T.Sakai et al, « Computer analysis and classification of human faces », in Proc, First USA Japan computer Conference, p .27, 1972.
- [16] G. Yang and T. S. Huang. "Human Face Detection in Complex Background," Pattern Recognition, vol. 27, no. 1, pp. 53-63, 1994

- [17] L. Xiaohua et al « Face detection using simplified gabor features and hierarchical regions in a cascade of classifiers ». *Pattern Recognit Lett* 30:717–728, 2009
- [18] Y. Taigman et al, « Deepface: Closing the gap to human-level performance in face verification. In *CVPR*, pages 1701–1708, 2014
- [19] Y. Sun et al, « Deep learning face representation by joint identification-verification. In *NIPS*, pages 1988– 1996, 2014.
- [20] Y. Sun et al, « Deepid3: Face recognition with very deep neural networks. arXiv:1502.00873, 2015
- [21] F. Schroff et al, « Facenet: A unified embedding for face recognition and clustering. In *CVPR*, pages 815– 823, 2015.
- [22] J. Liu et al, « Targeting ultimate accuracy: Face recognition via deep embedding. arXiv:1506.07310, 2015
- [23] O. Parkhi et al. « Deep face recognition. In *BMVC*, volume 1, page 6, 2015
- [24] X. Wu et al, « A light cnn for deep face representation with noisy labels arXiv:1511.02683, 2015.
- [25] X. Wu et al, « A light cnn for deep face representation with noisy labels arXiv:1511.02683, 2015 *ECCV*, pages 499–515. Springer, 2016
- [26] W. Liu et al. « Large-margin softmax loss for convolutional neural networks. In *ICML*, pages 507–516, 2016.
- [27] X. Zhang et al, « Range loss for deep face recognition with long-tail. arXiv:1611.08976, 2016.
- [28] R. Ranjan, et al, « L2-constrained softmax loss for discriminative face verification. arXiv:1703.09507, 2017
- [29] F. Wang et al, « Normface: l_2 hypersphere embedding for face verification. arXiv:1704.06369, 2017
- [30] Y. Liu et al, « Rethinking feature discrimination and polymerization for large-scale recognition, arXiv:1710.00870, 2017
- [31] M. Hasnat et al. « von mises-fisher mixture model-based deep learning: Application to face verification. arXiv:1706.04264, 2017
- [32] J. Deng, Y. Zhou, and S. Zafeiriou. Marginal loss for deep face recognition. In *CVPR Workshops*, volume 4, 2017
- [33] W. Li et al, « Sphreface: Deep hypersphere embedding for face recognition. In *CVPR*, volume 1, 2017

- [34] X. Qi et al, « Face recognition via centralized coordinate learning. arXiv:1801.05678, 2018
- [35] F. Wang et al, « Additive margin softmax for face verification. arXiv:1801.05599, 2018.
- [36] H. Wang, et al, « Large margin cosine loss for deep face recognition, 2018
- [37] J. Deng, et al, « Additive angular margin loss for deep face recognition. arXiv:1801.07698, 2018
- [38] Y. Zheng, et al, « Convex feature normalization for face recognition. In CVPR, June 2018.
- [39] HADJER, C. (2020). Montage d'un Système de Reconnaissance des Expressions Faciales avec le Deep Learning. Biskra.