REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE YAHIA FARES DE MEDEA



Faculté de **T**echnologie

Département du **G**énie **E**lectrique

Mémoire de fin d'études de Master

Filière : télécommunications

Spécialité : système des télécommunications

Reconnaissance des émotions par l'analyse visuelle du visage

Présentés par :

ZOUBIRI Sefouane

HAFSI Mohamed Abderrahmane

Proposés et Dirigés par :

Dr : SOFI Naima

Promotion 2022

Dédicace :

Nous tenons à la fin de ce travail à remercier ALLAH le tout puissant de nous avoir donné la foi et de nous avoir permis d'en arriver là.

Nous dédions ce modeste travail :

- *A nos très chère parentes.*
- *A toute notre famille.*
- *A tous ceux qui nous sont chers.*
- *A tous nos amis.*

Merci à tous

Remerciement

Tout d'abord nous tenons à remercier ALLAH pour nous avoir donné le courage, la force et la volonté pour réussir et de nous avoir éclairci le chemin tout au long de notre vie.

Nous remercions Madame **SOFI Naima** et Monsieur **Abudara salam** pour tout ce qu'ils aient été fait pour nous encadrer dans ce projet de fin d'étude.

ils nous ont dirigé et accompagné de très près, à vrai dire pas à pas, jour par jour, et avec beaucoup de patience Nous tenons aussi à remercier les membres du jury pour leur précieux, temps accordé à l'étude de notre mémoire

Nous remercions tous nos enseignants, collègues et personnels du département de génie électrique pour tout ce qu'ils nous ont apporté pour l'avancement et l'aboutissement de ce travail

Nous remercions toutes les personnes qui nous ont soutenues de près ou de loin dans la réalisation de ce travail.

Résumé

Pour les êtres humains, l'expression du visage est l'un des plus puissants et naturels moyens pour communiquer leurs émotions et leurs intentions.

Un être humain est capable de reconnaitre les expressions du visage sans effort, mais, pour une machine, cette tâche est très difficile.

Aujourd'hui, la reconnaissance d'expression faciale s'avère être l'une des applications les plus pertinentes dans de nombreux domaines à savoir : interaction homme-machine, médecine, sécurité, éducation, . . .

Tout système de reconnaissance automatique des expressions faciales passe par étapes essentielles : la détection du visage, l'extraction des caractéristiques faciales, et la classification des expressions. Un système qui effectue ces opérations, et avec précision, serait crucial pour parvenir à une interaction homme machine comme celle homme aura lieu.

L'efficacité des systèmes de reconnaissance d'expressions faciles est importante pour une bonne interaction homme-machine. Mais la tâche de la reconnaissance de l'expression faciale est liée à plusieurs méthodes qui fournissent des facteurs influençant sur la performance des systèmes FER (Facial recognition System). Ce mémoire présente un travail qui exploite une architecture du réseau CNN «séquentielle», pour la reconnaissance des émotions de base. Notre choix de cette architecture est motivé par son succès dans la classification des images, ce qui a été prouvé par nos résultats obtenus. Nous nous concentrons sur la structure de CNN tout en analysant les convolutions internes dans les réseaux qui influencent directement sur les résultats. Notre système de reconnaissance des émotions de base été entrainé sur la base de données la Fer2013.

Mots clés : réseaux de neurones, CNN, émotion, expressions faciale, séquentielle …

Abstract

For human beings, facial expression is one of the most powerful and natural ways to communicate their emotions and intentions.

A human being is able to recognize facial expressions effortlessly, but for a machine, this task is very difficult.

Today, facial expression recognition is proving to be one of the most relevant applications in many fields such as: human-computer interaction, medicine, security, education, etc…

Any automatic facial expression recognition system goes through essential steps: face detection, facial feature extraction, and expression classification. A system that performs these operations, and with accuracy, would be crucial to achieve a human-machine interaction like the human-human one will take place.

The efficiency of easy expression recognition systems is important for good human-machine interaction. But the task of facial expression recognition is related to several methods that provide influencing factors on the performance of FER (Facial recognition System) systems. This thesis presents a work that exploits a CNN architecture «sequential ", for basic emotion recognition. Our choice of this architecture is motivated by its success in image classification, which has been proven by our results. We focus on the structure of CNN while analysing the internal convolutions in the networks that directly influence the results. Our basic emotion recognition system was trained on the Fer2013 database.

Keywords: neural networks, CNN, emotion, facial expressions, sequential...

Liste des tableaux

TABLEAU	PAGE
Table I.1 Les étapes majeures de Deep Learning	10
Table I.2 La Différence Faite par le Deep Learning	15
Table II.1 Les caractéristique des expressions Faciales	44

Liste des figures

FIGURES	PAGE
Figure I.1 La relation entre l'IA et ML et le Deep Learning	8
Figure I.2 Un processus de Deep Learning	9
Figure I.3 Le procède du ML comparé à celui de Deep Learning	13
Figure I.3 Différents modèles du Deep Learning	19
Figure I.4 Architecture d'un réseau de neurone convolutionnel	21
Figure I.5 Exemple d'une convolution	22
Figure I.6 Exemple de réseau composé de nombreuses couches à Convolution	23
Figure I.7 Diffèrent type de convolution	24
Figure I.8 Application de RELU à partir d'une matrice et d'une image	25
Figure 1.9 Pooling avec un filtre 2*2 et un pas de 2	26
Figure 1.10 à gauche averagepooling et à droite max pooling	26
Figure 1.11 Exemple de mise à plat des images finales en sortie (des filtres +simplifications)	27
Figure 1.12 un exemple d'opération entièrement connecte (FC)	28
Figure II.1 Schéma illustratif des composantes de l'émotion	38
Figure II.2 Canaux de communication émotionnelle et capteurs associé	39
Figure II.3 Exemple des Expressions Faciales	40
Figure II.4 Espace de modélisation A Rousal – Valence	41
Figure II.5 la roue des émotions de Plutchik	42
Figure II.6 Les six expressions faciales	44
Figure II.7 Architecture de système de reconnaissance des expressions Faciales	47
Figure II.8 Détection de visage	47
Figure II.9 AUS et la description de chaque unité d'action faciale	54
Figure II.10 modèle du visage MPEG4	55
Figure II.11 Version 3D du modèle candide	57
Figure III.1 architecture globale de notre système de reconnaissance des émotions	67
Figure III.3 Processus de détection de visage	68
Figure III.4 prétraitement des données	70

Figure III.5 les réseaux neuronaux convolutionnels (CNN)	73
Figure III.6 les réseaux neuronaux convolutionnels (CNN)	74
Figure III.7 Entraînement du modèle	75
Figure III.8 Évaluation du modèle	76
Figure III.9 Test du modèle	77
Figure III.10 Test du modèle	77
Figure III.11 Test du modèle	78
Figure III.12 Test du modèle	78
Figure III.13 Test du modèle	78
Figure III.14 Test du modèle	78
Figure III.15 présente l'état neutre	79
Figure III.16 présente l'état surprise	79
Figure III.17 présente l'état triste	80
Figure III.18 présente l'état heureux	80

Table de matière

Dédicace

Remerciement

Résumé

Abstract

Liste des tableaux

Liste des figures

INTRODUCTION GENERALE : 12

CHAPITRE I : APPRENTISSAGE AUTOMATIQUE 4

1. Introduction : 4

1-1.Quelques mots d'Histoire : 4

1-2. Apprentissage automatique : 4

A- Définition et principe : 4

1.3 Principes : 5

1-4. Types d'apprentissage : 5

1-5-deeplearning et réseaux de neurones : 8

2. L'Apprentissage en Profondeur (Le Deep Learning) : 9

2-1.Historique de Deep Learning : 9

2-2.Définition :.....12
2-3.Pour quoi le Deep Learning :.....12
2-4.Concepts de Deep Learning :.....14
2-5.Machine Learning vs Deep Learning:.....15
2-6.Domaine d'application de Deep Learning :.....16
2-7.Les différentes Architectures du Deep Learning :.....16
A-Les réseaux de neurones convolutifs :.....16
B-Réseau de neurones récurrents :.....17
Modèle génératif :.....18
2-8.Exemples d'Application de Deep Learning *:*.....19
3-Les réseaux de neurones convolutifs :.....19
3-1.Présentation :.....19
3-2.Architecture de réseaux de neurone convolutif :.....20
3-3.La couche de convolution (CONV) :.....21
A-Les différentes convolutions :.....22
B-La convolution classique :.....22
C-La transposed convolution :.....23
D-La séparable convolution :.....23
3-4.La couche RELU :.....23
3-5.Couche de Pooling :.....24
Le Flattening (La mise à plat) :.....26
3-6.La couche entièrement connectée (FC) :.....26
3-7.La couche de perte (LOSS) :.....27
4-Quelques réseaux convolutifs célèbres : [7].....27
5-Contexte émotion- Deep Learning :.....28
Conclusion :.....31
CHAPITRE II : EMOTIONS ET EXPRESSIONS FACIALES.....34
1-Introduction :.....34
2-L'émotion :.....34
2-1.Définition :.....34
2-2.L'émotion en philosophie :.....34
2-3.L'émotion en psychologie :.....35
2-4.Types d'émotion :.....36
A-Émotions primaires :.....36
B-Émotions secondaires :.....36

C-Émotions sociales : 36
2-4-Les composantes de l'émotion : 36
A.La composante comportementale : 37
B.La composante physiologique : 37
2-5.La composante cognitive/subjective : 37
2-6.Canaux de communication émotionnelle : 38
2-7.La reconnaissance des émotions : 39
3-Expression faciale : 39
3-1.Définition : 39
3-2.Représentation de l'Expression Faciale Emotionnelle : 40
A-Le modelé de A Russel-valence : 40
B-La roue des émotions de Plutchik : 41
C-Le modèle discret d'Ekman : 42
3-3.La détection du visage : 47
A-Caractéristiques d'un détecteur de visage : 47
B-Les Méthodes de détection de visage : 48
3-4.Extraction de la caractéristique faciale : 50
A-Les caractéristiques géométriques : 51
B-Les caractéristiques d'apparence : 51
3-5.La classification : 52
4-Système de codification des expressions faciale : 53
4.1.Représentation par le système FACS : 53
4.2.MPEG4 : 54
4.3.Candide : 55
A-Paramètres de forme (Shape Units) : 56
B-Paramètres d'animation (Animation Units) : 56
4.4-Domaines d'applications : 57
5-Méthodes de reconnaissance des émotions de base: 59
5-1.Approche Basée Modèles : 59
5-2.Approche Géométrique : 60
5-3.Approche Basée Règles : 61
5-4.Approches Basées Réseaux de Neurones : 62
6-Contexte de travail : 62
CHAPITRE 3 : CONCEPTION ET IMPLEMENTATION....66
1-Introduction : 66

2-Conception : 66
2-1.Présentation de notre système de reconnaissance des émotions: 66
2-2.Architecture globale de notre système de reconnaissance des émotions : 66
2-3.Conception détaillée de système : 67
A-Détection du visage : 67
B- Extraction de caractéristiques faciales : 68
C- la reconnaissance des émotions faciales à l'aide d'un CNN : 69
Conclusion : 81
Conclusion 83
BIBLIOGRAPHIE : 85

INTRODUCTION

INTRODUCTION GENERALE :

L’informatique affective est le fait que l'ordinateur soit intelligent à comprendre, reconnaître et interpréter l'émotion humaine. Savoir si une machine pourrait un jour ressentir des émotions est une question posée et discutée depuis longtemps, sans qu'une réponse définitive n’ait été apportée. L’expression de l’émotion constitue une dimension importante en communication. Elle permet de définir une relation particulière avec le sentiment d’un individu, de la mettre à jour, de telle sorte qu’une personne peut, de façon fiable, à partir de comportements verbaux ou non verbaux (tels que l’attitude corporelle, les expressions faciales) de déduire l’état affectif actuel d’une autre personne avec laquelle elle interagit.

Le visage étant la partie la plus expressive et communicative d’un être humain, parce qu’elle peut lui montrer différentes expressions émotionnelles exprimant l’émotion intérieure de la personne. Un être humain est capable de détecter les expressions du visage dans une scène sans aucun effort. Mais, le développement d’un système automatique qui effectue cette tâche est assez difficile surtout que la demande d'extraire automatiquement les informations des caractères faciaux augmente toujours. Les expressions faciales provoquent des changements physiologiques sur le visage, tels que le mouvement des sourcils, la position de la bouche ouverts ou fermés, ou encore la manière de regarder fixement les yeux. Selon plusieurs travaux, il existe six expressions

émotionnelles universelles : dégouts, colère, joie, tristesse, surprise et peur. Ces expressions sont identifiables en observant les signes du visage.

Aujourd'hui, la reconnaissance d'expression faciale s'avère être l'une des applications les plus pertinentes dans de nombreux domaines à savoir : interaction homme-machine, médecine, sécurité, éducation, …etc.

Pour réaliser notre travail qui est la reconnaissance des émotions de base nous avons utilisé le Deep Learning, la méthode d`apprentissage qui a montré ses performances ces dernières années et nous avons choisi le CNNs (Réseau de neurones convolutif) comme un modèle de reconnaissance, ce choix est justifié par la simplicité et l'efficacité du modèle. Les réseaux de neurones convolutifs sont devenus l'algorithme principal de la vision par ordinateur.

L'objectif de ce projet de fin d'étude est de concevoir et de réaliser une application permet de reconnaitre les émotions de base exprimées par le visage en utilisant la méthode d'apprentissage profond.

Ce mémoire est composé de trois chapitres, qui sont organisés comme suit :

Le premier chapitre est un état de l'art qui résume le domaine de notre travail : nous commençons par approfondir le concept d'émotion, ensuite nous discutons les expressions faciales et enfin nous présentons les différents systèmes de codifications et de reconnaissances des émotions. Ainsi que quelques travaux de détection des émotions.

Ensuite, le second chapitre, est consacré à la présentation de l'apprentissage profond, où nous donnerons plus de détails sur les réseaux de neurones convolutifs. Enfin nous avons discuté quelques travaux de reconnaissance des émotions basées Deep Learning.

Le dernier chapitre est consacré en premier lieu à la conception de notre approche de reconnaissance des expressions faciales basée sur les CNNs. Ensuite nous présentons les différents outils utilisés dans le développement de notre application, ainsi que les différents résultats obtenus.

Et enfin, nous terminerons ce mémoire par une conclusion générale et quelques perspectives.

CHAPITRE I

APPRENTISSAGE AUTOMATIQUE

CHAPITRE I : APPRENTISSAGE AUTOMATIQUE

1. Introduction :

Au cours des dernières années, l'application de l'intelligence artificielle dans tous les domaines tels que la reconnaissance d'immatriculation, la reconnaissance faciale, la reconnaissance vocale et la reconnaissance des expressions faciales, est évoluée rapidement.

Dans ce chapitre, nous fournissons le contexte théorique nécessaire pour comprendre les méthodes discutées dans le dernier chapitre. Tout d'abord, nous discutons des détails de l'apprentissage automatique, des réseaux de neurones et le deeplearning. Ce chapitre contient la définition et le principe, l'application de chaque élément.

1-1.Quelques mots d'Histoire :

L'apprentissage artificiel est une discipline jeune, à l'instar de l'Informatique et de l'Intelligence artificielle. Il se situe au carrefour d'autres disciplines: philosophie, psychologie, biologie, logique, mathématique [1].les premières études remontent à des travaux de statistique dans les années 1920. C'est après la seconde guerre mondiale que les premières expériences deviennent possibles. Se développent ensuite dans les années 1960 les approches connexionnistes avec des perceptrons, et la reconnaissance des formes. La mise en évidence des limites du perceptron simple arrête toutes les recherches dans ce domaine jusqu'à la renaissance dans les années 1980. Les années 1970 sont dominées par des systèmes mettant l'accent sur les connaissances, les systèmes experts, Les limites de tels systèmes se font sentir dans les années 1980, pendant lesquelles a lieu le retour du connexionnisme avec un nouvel algorithme d'apprentissage [1]. Les mathématiciens commencèrent à s'éloigner du cadre cognitif de l'apprentissage pour envisager le problème sous l'angle de l'optimisation, pendant qu'apparaissaient de nouvelles méthodes comme les arbres de décision ou l'induction de programmes logiques. L'influence de la théorie statistique de l'apprentissage s'est réellement fait sentir dans les années 1990, avec l'ouvrage de Vapnik [2].

1-2. Apprentissage automatique :

A- Définition et principe :

Apprentissage automatique est un sous-ensemble de l'intelligence artificielle (IA) qui est axé sur la création de systèmes qui apprennent ou améliorent les performances en fonction des données qu'ils traitent. L'intelligence artificielle est un terme large qui désigne des systèmes ou des machines imitant l'intelligence humaine. L'apprentissage automatique et l'IA sont souvent abordés ensemble, et les termes sont parfois utilisés de manière interchangeable, mais ils ne veulent pas dire la même chose. Une distinction importante est que, même si l'intégralité d'Apprentissage automatique repose sur l'intelligence artificielle, cette dernière ne se limite pas à l'Apprentissage automatique.

Aujourd'hui, nous utilisons l'apprentissage automatique dans tous les domaines. Lorsque nous interagissons avec les banques, achetons en ligne ou utilisons les médias sociaux, des algorithmes d'Apprentissage automatique entrent en jeu pour optimiser, fluidifier et sécuriser notre expérience. L'Apprentissage automatique et la technologie qui l'entoure se développent rapidement, et nous commençons seulement à entrevoir ses capacités.

1.3 Principes :

Les algorithmes utilisés permettent - dans une certaine mesure - à un système piloté par ordinateurs (un robot éventuellement), ou assisté par ordinateur d'adapter ses analyses, et comportements en réponse en se fondant sur l'analyse de données empiriques provenant d'une base de données ou de capteurs.

La difficulté réside dans le fait que l'ensemble de tous les comportements possibles compte tenu de toutes les entrées possibles devient rapidement trop complexes à décrire dans les langages de programmation disponibles, de sorte qu'on confie en quelque sorte à des programmes le soin d'apprendre de manière à auto-méliorer le système d'analyse ou de réponse (commande adaptative...), ce qui est une des formes que peut prendre l'intelligence artificielle

Ces programmes, selon leur degré de perfectionnement intègrent des capacités en probabilités et statistiques, traitement de données et éventuellement d'analyse de données issues de capteurs, de reconnaissance (reconnaissance vocale, reconnaissance de forme, d'écriture, etc.), de data-mining et d'informatique théorique.

1-4. Types d'apprentissage :

Généralement dans l'apprentissage machine, les tâches à résoudre sont classées en catégories. Ces catégories sont définies selon la façon dont l'apprentissage est reçu ou comment le retour d'informations est donné au système développé.

Il existe plusieurs méthodes d'apprentissage automatique, mais les plus largement utilisées sont l'apprentissage supervisé ou les algorithmes sont basés sur des données d'entrées et de sorties bien spécifiées par l'homme et l'apprentissage non supervisé qui ne fournit pas à l'algorithme des données étiquetées afin de lui pouvoir permettre de définir une structure et de découvrir une logiques dans les données entrées. Auxquelles on ajoute l'apprentissage par renforcement et l'apprentissage profond. Explorons donc ces méthodes plus en détail.

- **L'apprentissage supervisé** : Si les *classes* sont prédéterminées et les *exemples* connus, le système apprend à classer selon un *modèle* de classement ; on parle alors d'apprentissage supervisé (ou d'analyse discriminante). Un expert (ou *oracle*) doit préalablement correctement étiqueter des exemples. L'« *apprenant* » peut alors trouver ou approximer la fonction qui permet d'affecter la bonne *« étiquette »* à ces exemples. Parfois il est préférable d'associer une donnée non pas à une classe unique, mais une probabilité d'appartenance à chacune des classes prédéterminées (on parle alors d'apprentissage supervisé probabiliste).

Ex : L'analyse discriminante linéaire ou les SVM sont des exemples typiques. Autre exemple : en fonction de *points communs* détectés avec les symptômes d'autres patients connus (les *« exemples »*), le système peut catégoriser de nouveaux patients au vu de leurs analyses médicales en risque estimé (probabilité) de développer telle ou telle maladie.

- **L'apprentissage non-supervisé :** (ou classification automatique) : Quand le système ou l'opérateur ne disposent que d'exemples, mais non d'étiquettes, et que le nombre de classe et leur nature n'ont pas été prédéterminés, on parle d'apprentissage non supervisé ou clustering. Aucun expert n'est disponible ni requis. L'algorithme doit découvrir par lui-même la structure plus ou moins *cachée* des données. Le clustering est un algorithme d'apprentissage non supervisé.
Le système doit ici -dans l'espace de description (la somme des données) - cibler les données selon leurs attributs disponibles, pour les classer en groupe *homogènes* d'exemples. La similarité est généralement calculée selon la fonction

de distance entre paires d'exemples. C'est ensuite à l'opérateur d'associer ou déduire du sens pour chaque groupe et pour les patterns d' *apparition* des groupes ou groupes de groupes dans leur *« espace »*. Divers outils mathématiques et logiciels peuvent l'aider. On parle aussi d'analyse des données en régression. Si l'approche est probabiliste (c'est à dire que chaque exemple au lieu d'être classé dans une seule classe est associé aux probabilité d'appartenir à chacune des classes), on parle alors de *« soft clustering »* (par opposition au *« hard clustering »*).
Cette méthode est souvent source de sérendipité.

Ex : Un épidémiologiste pourrait par exemple dans un ensemble assez large de victimes de cancers du foie tenter de faire émerger des hypothèse explicatives, l'ordinateur pourrait différentier différents groupes, qu'on pourrait ensuite associer par exemple à leur provenance géographique, génétique, à l'alcoolisme ou à l'exposition à un métal lourd ou à une toxine telle que l'aflatoxine.

. Apprentissage non supervisé :

Dans l'apprentissage non supervisé, la machine reçoit des données non étiquetées. On lui demande de découvrir les schémas qui sous-tendent les données, p. ex. une structure en grappes, une variété en basses dimensions, ou un arbre et un graphique de faible densité.

. Apprentissage par renforcement :

L'apprentissage par renforcement permet d'analyser et d'optimiser le comportement d'un agent en fonction du retour d'informations de l'environnement. Les machines essaient différentes situations pour déterminer les actions les plus avantageuses, plutôt que de simplement recevoir des instructions sur les actions à entreprendre. Ce qui distingue l'apprentissage par renforcement des autres techniques, ce sont l'apprentissage par essais et erreurs et la récompense différée. L'apprentissage par renforcement est un modèle d'apprentissage comportemental. L'algorithme reçoit des informations grâce à l'analyse des données, de sorte que l'utilisateur est orienté vers le meilleur résultat. L'apprentissage par renforcement diffère des autres types d'apprentissage supervisé, car le système n'est pas entraîné à partir d'un ensemble de données : il apprend par essais et erreurs. Par conséquent, une série de décisions a pour

effet de « renforcer » le processus, car celui-ci convient le mieux pour résoudre le problème.

1-5-DeepLearning et réseaux de neurones :

L'IA est basée sur une démarche d'apprentissage afin de reproduire une partie de l'intelligence humaine à travers une application, un système ou un processus. La reconnaissance faciale, la perception visuelle et autres sont des exemples de systèmes d'intelligence artificielle.

La machine Learning (ML) est un sous-domaine de l'IA qui utilise les réseaux neuronaux artificiels (ANN) pour imiter la façon dont les êtres humains prennent des décisions. La machine Learning permet aux ordinateurs de développer des modèles d'apprentissage par eux-mêmes, sans aucune programmation, à partir de gros ensembles de données.

La couche immédiatement inférieure est occupée par le Deep Learning (DL), est l'une des nombreuses approches du machine Learning qui a connu un grand succès dans ces cinq dernières années, pour illustrer la relation entre ces termes, nous pouvons utiliser des cercles concentriques :

• Intelligence artificielle IA : le cercle le plus large est l'idée qui a émergé en premier dans ce domaine.

• Apprentissage automatique (machine Learning) : au milieu, il a prospéré plus tard après l'IA.

• Apprentissage approfondie (Deep Learning) : le plus petit cercle est une expansion de l'IA actuellement.

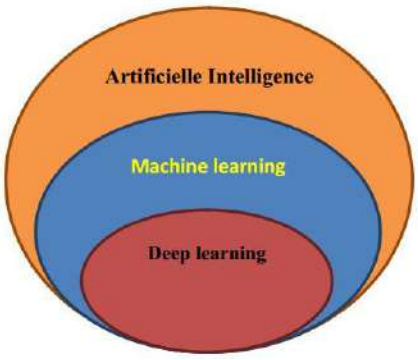


Figure I.1 La relation entre l'IA et ML et le Deep Learning

Le terme "Deep Learning" ou Apprentissage profond, a été introduit pour la première fois au ML par Dechter en1986, et aux réseaux neuronaux artificiels par Aizenberg et AL en 2000.

Le Deep Learning est un nouveau domaine de recherche du ML, qui a été introduit dans le but de rapprocher le ML de son objectif principal : l'intelligence artificielle. Il concerne les algorithmes inspirés par la structure et le fonctionnement du cerveau. Ils peuvent apprendre plusieurs niveaux de représentation dans le but de modéliser des relations complexes entre les données.



Figure I.2 Un processus de Deep Learning

Le Deep Learning est basé sur l'idée des réseaux de neurones artificiels et il est taillé pour gérer de larges quantités de données en ajoutant des couches au réseau. Un modèle de Deep Learning a la capacité d'extraire des caractéristiques

A partir des données brutes grâce aux multiples couches de traitement composé de multiples transformations linéaires et non linéaires et apprendre sur ces caractéristiques petit à petit à travers chaque couche avec une intervention humaine minime.

2. L'Apprentissage en Profondeur (Le Deep Learning) :

2-1.Historique de Deep Learning :

Table I.1 Les étapes majeures de Deep Learning

Années	Contributeur	Contribution
300 AC	Aristotle	Introduction de

		l'associationnisme, début de l'histoire des humains qui essayent de comprendre le cerveau.
1873	Alexander Bain	Introduction du Neural Groupings comme les premiers modèles de réseaux de neurones.
1943	McCulloch and Pitts	Introduction du McCulloch-Pitts (MCP) modèle considérer comme ''ancêtrel des réseaux de neurones artificiels.
1949	Donald Hebb	considérer comme le père des réseaux de neurones, il introduit la règle d'apprentissage de Hebb qui servira de fondation pour les
Années	Contributeur	Contribution
300 AC	Aristotle	Introduction de l'associationnisme, début de l'histoire des humains qui essayent de comprendre le cerveau.
1873	Alexander Bain	Introduction du Neural Groupings comme les premiers modèles de réseaux de neurones.
1943	McCulloch and Pitts	Introduction du

		McCulloch-Pitts (MCP) modèle considérer comme ''ancêtrel des réseaux de neurones artificiels.
1949	Donald Hebb	considérer comme le père des réseaux de neurones, il introduit la règle d'apprentissage de Hebb qui servira de fondation pour les
		bidirectionnels.
	Hochreiter and Schmidhuber	introduction de LSTM, qui ont résolu le problème du vanishing gradient dans Les Réseaux de neurones récurrents.
	Geoffrey Hinton	Introduction des Deepbelief Network.
	Salakhutdinov and Hinton	introduction des deep Boltzmann Machines.
	Alex Krizhevsky	Introduction d'AlexNet qui remporta le challenge imageNet.

2-2.Définition :

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

Dans l'apprentissage automatique, un programme analyse un ensemble de données afin de tirer des règles qui permettront de tirer des conclusions sur des nouvelles données. L'apprentissage profond est basé sur ce qui a été appelé, par analogie, des « Réseaux de neurones artificiels », composés de milliers d'unités (les « neurones ») qui effectuent chacune de petites opérations simples. Les résultats d'une première couche de « neurones » servent d'entrée aux calculs d'une deuxième couche et ainsi de suite.

Par exemple, pour la reconnaissance visuelle, des premières couches d'unités identifient des lignes, des courbes, des angles… des couches supérieures identifient des formes, des combinaisons de formes, des objets, des contextes…

Les progrès de l'apprentissage profond ont été possibles notamment grâce à l'augmentation de la puissance des ordinateurs et au développement de grandes bases de données (« Big Data »). [10]

2-3.Pour quoi le Deep Learning :

Les algorithmes de ML fonctionnent bien pour une grande variété de problèmes, cependant ils ont échoué à résoudre quelques problèmes majeurs de l'IA tels que la reconnaissance vocale et la reconnaissance d'objets.

L'IA émotionnelle regroupe un ensemble de cas d'utilisation de la machine Learning pour lesquelles les techniques d'apprentissage profond sont particulièrement bien adaptées.

Le principe fondamental de ML est de chercher à extraire des caractéristiques « features » qui aident à la classification et pour les extraire il y a deux possibilités :

• Soit à l'aide d'une expertise métier, celle d'un physionomiste par exemple, pour construire manuellement et directement ces variables prédictives à partir d'un visage. L'inconvénient de ce processus d'extraction manuel est qu'il comporte une part

d'arbitraire et qu'il n'exploite pas nécessairement l'intégralité de l'information disponible dans une photo ou dans une vidéo.

Soit en utilisant un réseau de neurones profond capable d'extraire les caractéristiques par collection de toute information utile disponible dans les données brutes et pour bénéficier de cet atout il faudra cependant renoncer à interpréter les prédictions d'un tel modèle et disposer d'un grand nombre d'exemples étiquetés. Comme dans beaucoup de problèmes d'IA c'est là que réside l'une des principales difficultés : construire des bases de données assez riches pour entraîner les algorithmes de Deep Learning à reconnaître des émotions. [7]

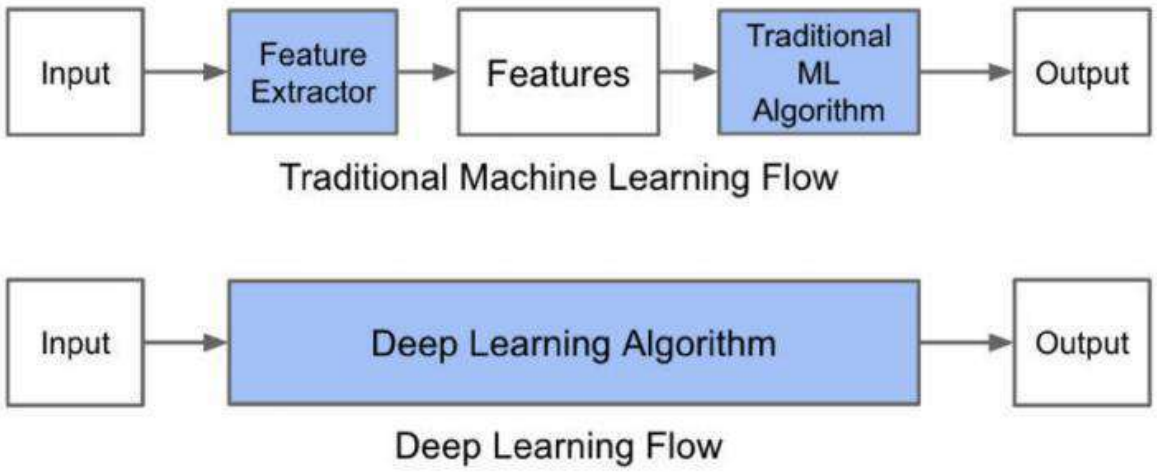


Figure I.3 Le procède du ML comparé à celui de Deep Learning

Les différents algorithmes du Deep Learning ne sont apparus qu'à l'échec de l'apprentissage automatique tentant de résoudre une grande variété de problèmes de l'intelligence artificielle (l'IA) afin :

• D'améliorer le développement des algorithmes traditionnels dans les 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.

Donc l'apprentissage en profondeur utilise des réseaux de neurones pour apprendre des représentations utiles de caractéristiques directement à partir de données. [4]

2-4.Concepts de Deep Learning :

Les concepts de Deep Learning constituent une classe d'algorithmes d'apprentissage automatique dont les caractéristiques sont les suivantes :

• Ils utilisent différentes couches d'unité de traitement non linéaire pour l'extraction et la transformation des caractéristiques. Chaque couche prend en entrée la sortie de la précédente. Les algorithmes peuvent être supervisés ou non supervisés et leurs applications comprennent la reconnaissance de modèles ou la classification statistique.

• Ils fonctionnent avec un apprentissage à plusieurs niveaux de détails ou de représentations des données. À travers les différentes couches on passe de paramètres de bas niveau à des paramètres de plus haut niveau.

• Ces différents niveaux correspondent à différents niveaux d'abstraction des données.

• Ce nouveau champ d'étude a pour objectif d'avancer vers les capacités d'intelligence artificielle. Ses architectures permettent aujourd'hui de donner du sens à des données sous forme d'image, de son ou de texte.

Le système de Deep Learning est un système basé sur les réseaux de neurones artificiels, qui est constitué d'un ensemble de couches cachées. Le mot profond (apprentissage profond) vient de ce nombre des couches cachées le point différencie entre un perceptron classique, et un système de Deep Learning est que pour le premier les entrées des réseaux sont les caractéristiques de l'image, mais pour le deuxième ce sont les pixels bruts de l'image d'entrée. En effet dans un système de Deep Learning chaque couche est considérée comme un niveau d'abstraction de l'image.

La couche 1 est capable d'extraire des caractéristiques de niveau d'abstraction inférieure à celui de la couche 2, tandis que la couche 3 a une qualité plus importante. À partir de ces caractéristiques le système est capable de reconnaitre les objets avec un taux d'erreur petit. [4]

2-5.Machine Learning vs Deep Learning:

Table I.2 La Différence Faite par le Deep Learning

Machine Learning	Deep Learning
L'apprentissage automatique utilise les algorithmes pour analyser les données, tirer des leçons de ces données et prendre des décisions éclairées en fonction de ce qu'il apprit.	L'apprentissage profond structure les algorithmes en couche pour créer un « Réseau neuronal artificiel » capable d'apprendre et de prendre des décisions intelligentes par lui-même.
Peut s'entrainer avec moins de données d'entrainement.	Nécessite de grands ensembles de données pour l'entrainement.
La sortie est sous forme numérique. Pour les applications de classification est de notification.	La sortie peut être sous n'importe quelle forme, y compris des éléments de forme libre tels que du texte et du son libre.
Réglage limité pour le réglage des hyper paramètres.	Peut être réglé de différentes manières.
S'entraine sur CPU.	S'entraine sur GPU pour une formation appropriée.

2-6 .Les avantages de Deep Learning :

Ils existent plusieurs avantages qui distinguent le Deep Learning, on cite :

• La robustesse pour comprendre et utiliser de nouvelles données.

• Il gère tout à un niveau d'abstraction beaucoup plus élevé que les réseaux de neurones standard.

• Il obtient ses résultats plus rapidement. Il apprend au fil du temps plutôt qu'en un éclair.

• Il peut gérer de grandes quantités de données pour de petits réseaux avec un coût d'apprentissage bien moindre.

2-7.Domaine d'application de Deep Learning :

Ces techniques se développent dans le domaine de l'informatique appliquées aux NTIC (reconnaissance visuelle - par exemple d'un panneau de signalisation par un robot ou une voiture autonome - et vocale notamment) à la robotique, à reconnaissance ou comparaison de formes, la sécurité, la santé, etc…, la pédagogie assistée par l'informatique, et plus généralement à l'intelligence artificielle.

L'apprentissage profond peut par exemple permettre à un ordinateur déformable d'analyser par exemple les émotions révélées par un visage photographié ou filmé, ou analyser les mouvements et position des doigts d'une main, ce qui peut être utile pour traduire le langage des signes, améliorer le positionnement automatique d'une caméra, etc… Elles sont utilisées pour certaines formes d'aide au diagnostic médical (ex : reconnaissance automatique d'un cancer en imagerie médicale), ou de prospective ou de prédiction (ex : prédiction des propriétés d'un sol filmé par un robot). [10]

2-8.Les différentes Architectures du Deep Learning :

Bien qu'il existe un grand nombre de variantes d'architectures profondes. Il n'est pas toujours possible de comparer les performances de toutes les architectures, car elles ne sont pas toutes évaluées sur les mêmes ensembles de données. Le Deep Learning est un domaine à croissance rapide, et de nouvelles architectures, variantes ou algorithmes apparaissent toutes les semaines**.**

A-Les réseaux de neurones convolutifs :

Convolutional Neural Network (CNN) (réseaux de neurones convolutifs) sont un type de réseau de neurones spécialisés en traitement de données ayant une topologie semblable à une grille. Qui se sont avérés très efficaces dans des domaines tels que la reconnaissance et la classification d'images et vidéos. CNN a réussi à identifier les visages, les objets, panneaux de circulation et auto-conduite des voitures. Récemment, les CNN ont été efficaces dans plusieurs tâches de traitement du langage naturel (telles que la classification des phrases).

Dans le ML, un réseau convolutif est un type de réseau de neurones feed-forward, il a été inspiré par des processus biologiques. Il existe cinq (5) principales opérations illustrées dans le CNN à savoir :

– La couche convolution.

– La couche RectifiedLinear Unit.

– La couche pooling.

– La couche entièrement connectée.

– La couche de perte.

B-Réseau de neurones récurrents :

Les humains ne commencent pas leurs pensées à zéro à chaque seconde. Lorsqu'ont un livre on comprend chaque mot en fonction de la compréhension des mots précédents. On n'oublie pas tout et recommençons à réfléchir à nouveau. Nos pensées ont une persistance.

RNN (les réseaux de neurones récurrents) traitent ce problème. Ce sont des réseaux avec des boucles, permettant aux informations de persister.

L'idée derrière les RNN est d'utiliser des informations séquentielles. Dans un réseau neuronal traditionnel, nous supposons que toutes les entrées (et les sorties) sont indépendantes les unes des autres. Mais pour de nombreuses tâches, c'est une très mauvaise idée. Si on veut prédire le prochain mot dans une phrase, il faut connaitre les mots qui sont venus avant. Les RNN sont appelés récurrents car ils exécutent la même tâche pour chaque « élément d'une séquence, la sortie étant dépendante des calculs précédents. Une autre façon de penser les RNN est qu'ils ont une mémoire qui capture l'information sur ce qui a été calculé jusqu'ici. En théorie, les RNN peuvent utiliser des informations dans des séquences arbitrairement longues, mais dans la pratique, on les limite à regarder seulement quelques étapes en arrière.

Les RNN ont connu un grand succès dans de nombreuses taches de traitement du langage naturel. Les plus grands exploits des RNN ont été accomplis par l'architecture LSTM, car ils sont bien meilleurs pour capturer des dépendances à long terme. [10]

• Modélisation du langage et génération de texte.

• Traduction automatique.

• La reconnaissance vocale description des images.

Modèle génératif :

Si les modèles discriminatifs comme (CNN, RNN) sont utilisés pour prédire les données du label et de l'entrée, tant dis 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. Voici quelques exemples de modèle génératif : [4]

• Boltzmann Machines.

• Restricted Boltzmann Machines.

• DeepBelief.

• GenerativeAdversarial Networks.

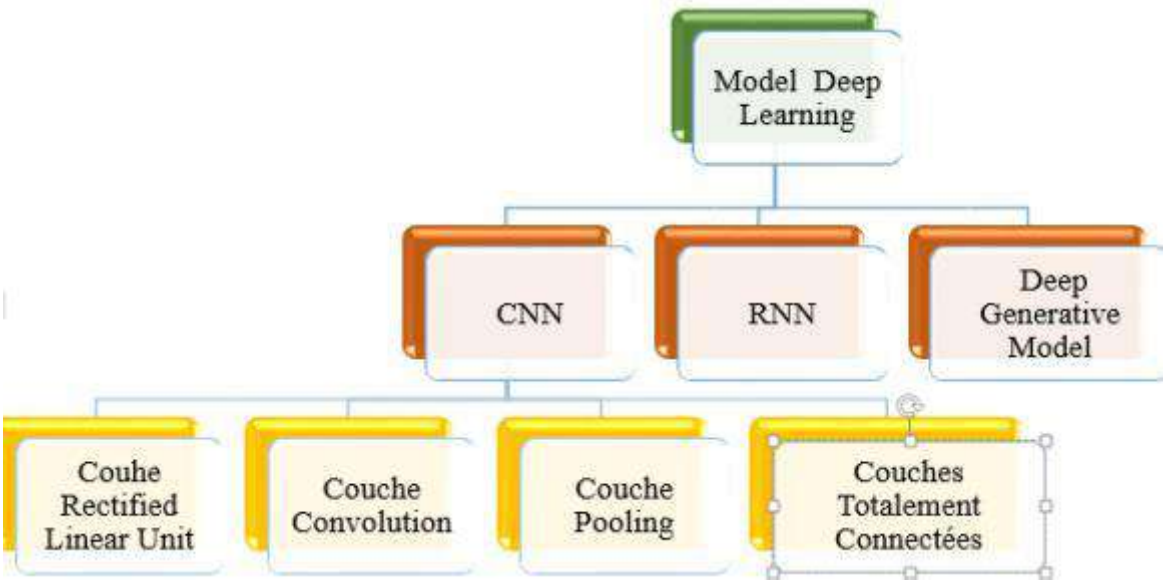


Figure I.3 Différents modèles du Deep Learning

2-9.Exemples d'Application de Deep Learning *:*

Les applications exploitantes le Deep Learning sont utilisées dans divers secteurs, de la conduite automatisée aux dispositifs médicaux. Grâce au Deep Learning nous pouvons maintenant : [4]

– Faire une colorisation des images en noir et blanc.

– Ajouter des sons à des films silencieux.

– Faire de la traduction automatique.

– Faire de classification des objets en photographies.

– Générer d'écriture automatique.

– Génération de légende d'image.

– Jeu automatique.

3-Les réseaux de neurones convolutifs :

Durant cette phase nous baserons uniquement sur le CNN qui est l'approche utilisée dans notre système de reconnaissance des émotions.

3-1.Présentation :

Les réseaux de neurones convolutifs sont à ce jour les modèles les plus performants dans des domaines tels que la reconnaissance et la classification des images.

CNN ont réussi à identifier les visages, les objets, panneaux de circulation et auto-conduite des voitures, durant cette phase de formation, nous nous baserons uniquement sur le CNN. Désignés par l'acronyme CNN, (de l'anglais Convolutional Neural Network) [5], Le nom « réseau de neurones convolutif » indique que le réseau emploie une opération mathématique appelée convolution. La convolution est une opération linéaire spéciale. Les réseaux convolutifs sont simplement des réseaux de neurones qui utilisent la convolution à la place de la multiplication matricielle dans au moins une de leurs couches. Ils comportent deux parties principales. S'il y on a en entrée, une image qu'elle doit être sous la forme d'une matrice de pixels, de 2 dimensions pour une image

en niveaux de gris et en 3 dimensions si elle est en couleur, pour représenter les couleurs fondamentales [Rouge, Vert, Bleu]. Cette image passe par la première partie d'un CNN qui est la partie convolutive.

Elle fonctionne comme un extracteur de caractéristiques des images. L'image passe à travers une succession de filtres, ou noyaux de convolution, pour la transformé en nouvelles images appelées cartes de convolutions « featuremaps ». Certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum local. Au final, les cartes de convolutions sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN. Le résultat en sortie de la partie convolutive est branché en entrée d'une deuxième partie, constituée de couches entièrement connectées (perceptron multicouche) qui consiste à combiner les caractéristiques de tous le réseau pour classer l'image et à la sortie qui est une couche comportant un neurone par classe, on obtient des valeurs numériques généralement normalisées entre 0 et 1, pour présenter la distribution de probabilité sur les classes. [7]

Les CNN permettent de traiter directement de grandes quantités de données que sont les images pour plusieurs raisons:

• Les images sont corrélées spatialement (les valeurs des pixels adjacents sont généralement très proches) et les couches de convolutions permettent de créer des liens entre ces données corrélées spatialement.

• Les couches de convolutions sont généralement suivies de couches de sous échantillonnage (pooling) qui diminuent grandement la taille des données. [4]

3-2.Architecture de réseaux de neurone convolutif :

Les CNN sont composés de trois types de couches : des couches convolutives, des couches de regroupement et des couches entièrement connectées. Lorsque ces couches sont empilées, une architecture CNN a été formée. Ces couches sont traitées indépendamment :

• La couche de convolution (CONV), qui traite les données d'un champ récepteur.

• La couche de pooling (POOL), qui permet de compresser l'information en réduisant la taille de l'image intermédiaire (souvent par sous-échantillonnage).

• La couche de correction (Relu), souvent appelée par abus 'Relu' en référence à la fonction d'activation (Unité de rectification linéaire).

• La couche "entièrement connectée" (FC), qui est une couche de type perceptron.

• La couche de perte (LOSS). [5]

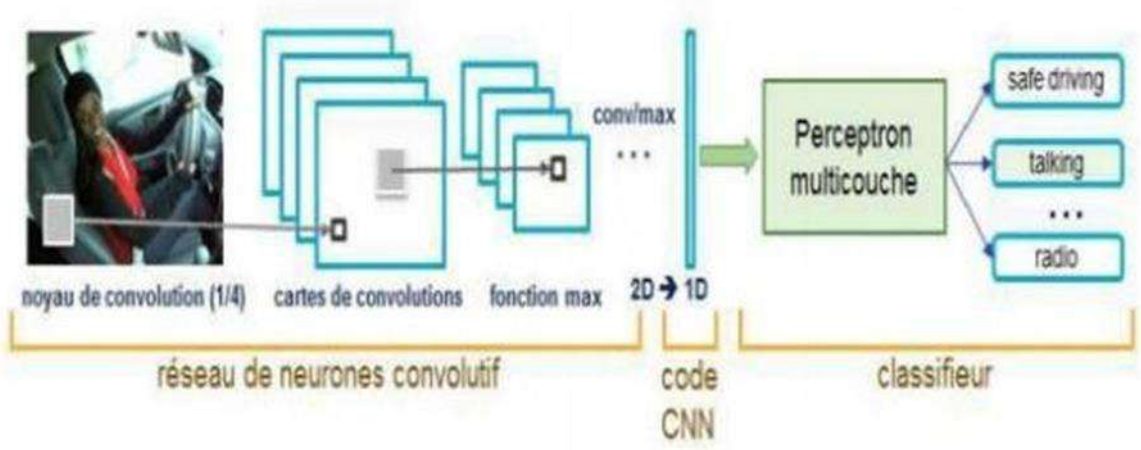


Figure I.4 Architecture d'un réseau de neurone convolutionnel

3-3.La couche de convolution (CONV) :

La convolution est le cœur du réseau de neurones convolutif. A l'origine, une convolution est un outil mathématique (on parle de produit de convolution) très utilisé en retouche d'image, car il permet d'en faire ressortir l'extraction des caractéristiques à partir des images d'entrées, afin d'appliquer un bon filtre. En fait, une convolution prend simplement en entrée une image et un filtre (qui est une autre image), effectue un calcul, puis renvoie une nouvelle image (généralement plus petite). [4]

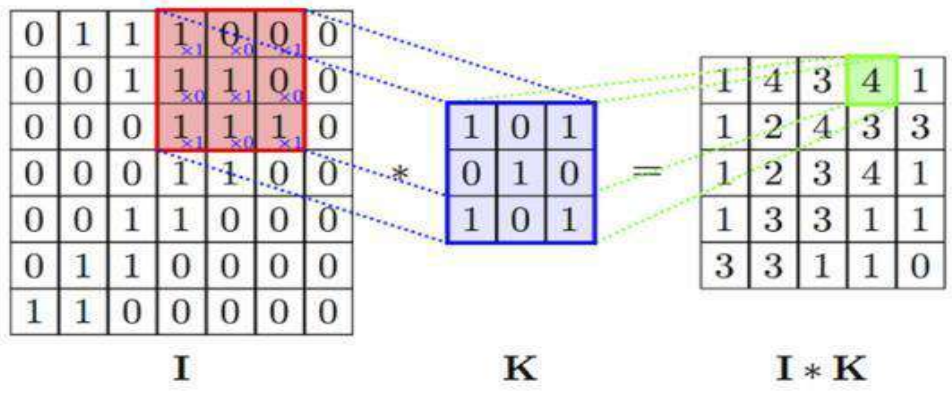


Figure I.5 Exemple d'une convolution

Trois hyper paramètres permettent de dimensionner le volume de la couche de convolution (aussi appelé volume de sortie) : la profondeur, le pas et la marge. [6]

• **Profondeur de la couche** : nombre de noyaux de convolution (ou nombre de neurones associés à un même champ récepteur).

• **Le pas** : contrôle le chevauchement des champs récepteurs. Plus le pas est petit, plus les champs récepteurs se chevauchent et plus le volume de sortie sera grand.

• **La marge (à 0) ou zéro padding** : parfois, il est commode de mettre des zéros à la frontière du volume d'entrée. La taille de ce 'zéro-padding' est le troisième hyper paramètre. Cette marge permet de contrôler la dimension spatiale du volume de sortie. En particulier, il est parfois souhaitable de conserver la même surface que celle du volume d'entrée.

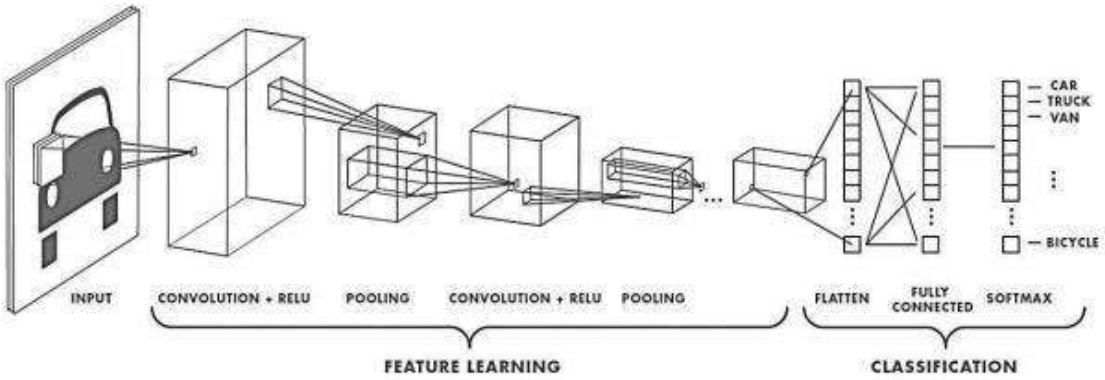


Figure I.6 Exemple de réseau composé de nombreuses couches à convolution

Des filtres sont appliqués à chaque image utilisée pour l'apprentissage à différentes résolutions, et la sortie de chaque image convoluée est utilisée comme entrée de la couche suivante.

A-Les différentes convolutions :

Il existe plusieurs types de convolutions, même si en général on utilise celle de base, il peut s'avérer utile de connaître les outils à notre disposition. [5]

B-La convolution classique :

qui représente le décalage du noyau entre chaque calcul, et le padding qui est la manière dont on peut « dépasser » de l'image pour appliquer la convolution.

C- convolution dilaté : identique à la convolution à ceci près que le kernel est éclaté (on prend, par exemple, un pixel sur deux pour calculer la convolution). Il y a un paramètre supplémentaire : la dilation rate, qui est le nombre de pixels à ignorer.

D- convolution transposé :

qui construit la sortie comme si on inversait une convolution sur l'image.

E- convolution séparé :

qui est une convolution décomposable en convolutions plus simples.

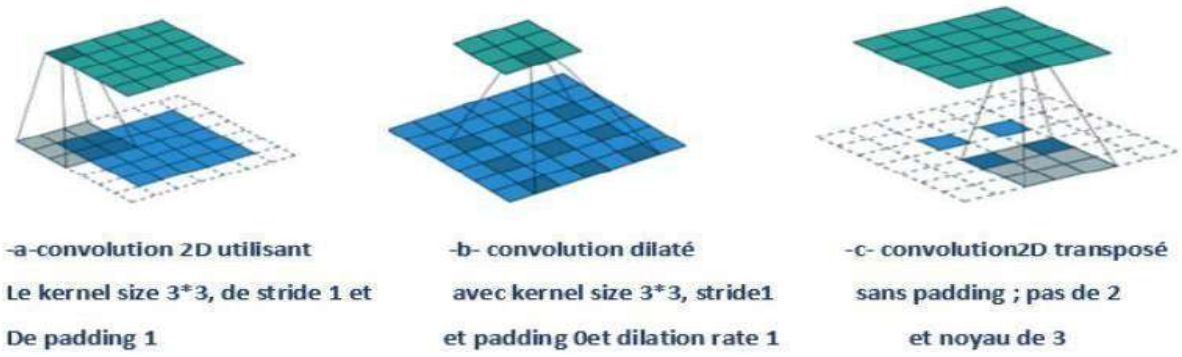


Figure I.7 Diffèrent type de convolution

3-4.La couche RELU :

C'est une couche pour 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 fonction Relu F(x)=max (0, x), cette fonction force les neurones à retourner des valeurs positives [5]. La fonction d'activation est souvent une fonction non-linéaire. Leur but est de permettre aux réseaux de neurones d'apprendre des fonctions plus complexes qu'une simple régression linéaire car le fait de multiplier les poids d'une couche cachée est juste une transformation linéaire. [7]

RectifiedLinear Unit (ReLU) : Elle est utilisée après chaque opération de convolution, ou toutes les valeurs de pixels négatifs sont mises à zéro. Le but de ReLU est d'introduire la non-linéarité dans notre CNN, puisque la plupart des données du monde réel, puisque la plupart de caractéristiques appliquées à l'une des cartes d'entrée donne une carte de sortie qui est également appelée carte de caractéristiques rectifiées. [4]

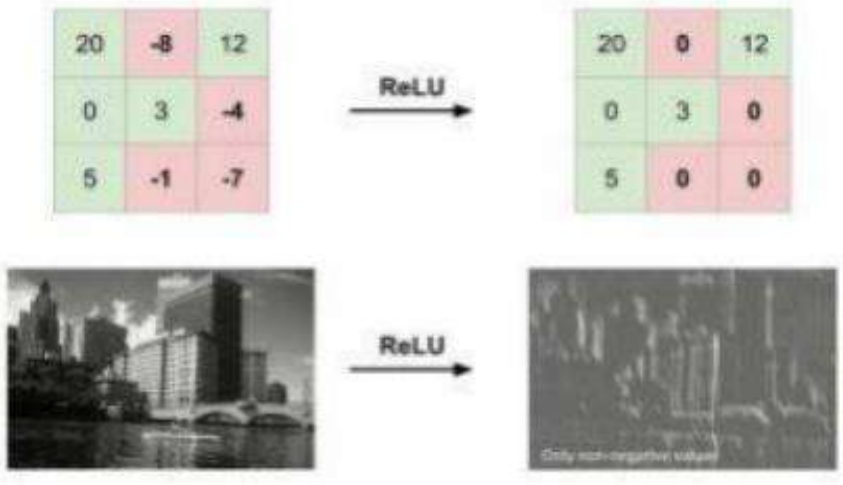


Figure I.8 Application de RELU à partir d'une matrice et d'une image

3-5.Couche de Pooling :

Ce type de couche est souvent placé entre deux couches de convolution : elle reçoit en entrée plusieurs featuremaps, et applique à chacune d'entre elles l'opération de pooling. L'opération de pooling (ousub-sampling) consiste à réduire la taille des images, tout en préservant leurs caractéristiques importantes.

Pour cela, on découpe l'image en cellules régulières, puis on garde au sein de chaque cellule la valeur maximale. En pratique, on utilise souvent des cellules carrées de petite taille pour ne pas perdre trop d'informations. Les choix les plus communs sont des cellules adjacentes de taille 2 x 2 pixels qui ne se chevauchent pas, ou des cellules de taille 3 x 3 pixels, distantes les unes des autres d'un pas de 2 pixels (qui se chevauchent donc). On obtient en sortie le même nombre de featuremaps qu'en entrée, mais celles-ci sont bien plus petites. La couche de pooling permet de réduire le nombre de paramètres et de calculs dans le réseau. On améliore ainsi l'efficacité du réseau et on évite le sur-apprentissage. [4]

Il existe plusieurs types de pooling : Max, Moyenne, Somme, etc.

Max pooling : qui revient à prendre la valeur maximale de la sélection. C'est le type le plus utilisé, car il est rapide à calculer (immédiat), et permet de simplifier efficacement l'image.

• **Le meanpooling :** (ou averagepooling), soit la moyenne des pixels de la sélection

: on calcule la somme de toutes les valeurs et on divise par le nombre de valeurs. On obtient ainsi une valeur intermédiaire pour représenter ce lot de pixels.

• **Le sumpooling :** c'est la moyenne sans avoir divisé par le nombre de valeurs (on ne calcule que leur somme).

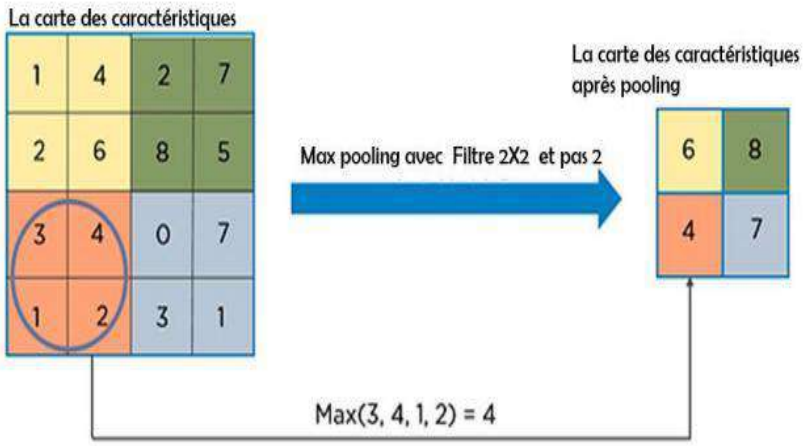


Figure 1.9 Pooling avec un filtre 2*2 et un pas de 2

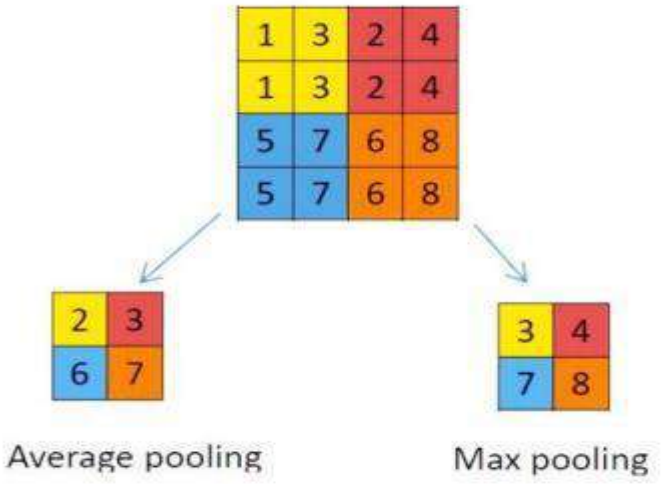


Figure 1.10 à gauche averagepooling et à droite max pooling

Le Flattening (La mise à plat) :

Dernière étape de la partie "extraction des informations", le flattening consiste simplement à mettre bout à bout toutes les images (matrices) que nous avons (résultat de pooling) pour en faire un (long) vecteur. Les pixels (en réalité ce ne sont plus des images ou des pixels, mais des matrices de nombres, donc les pixels sont ces nombres) sont récupérés ligne par ligne et ajoutés au vecteur final. [5]

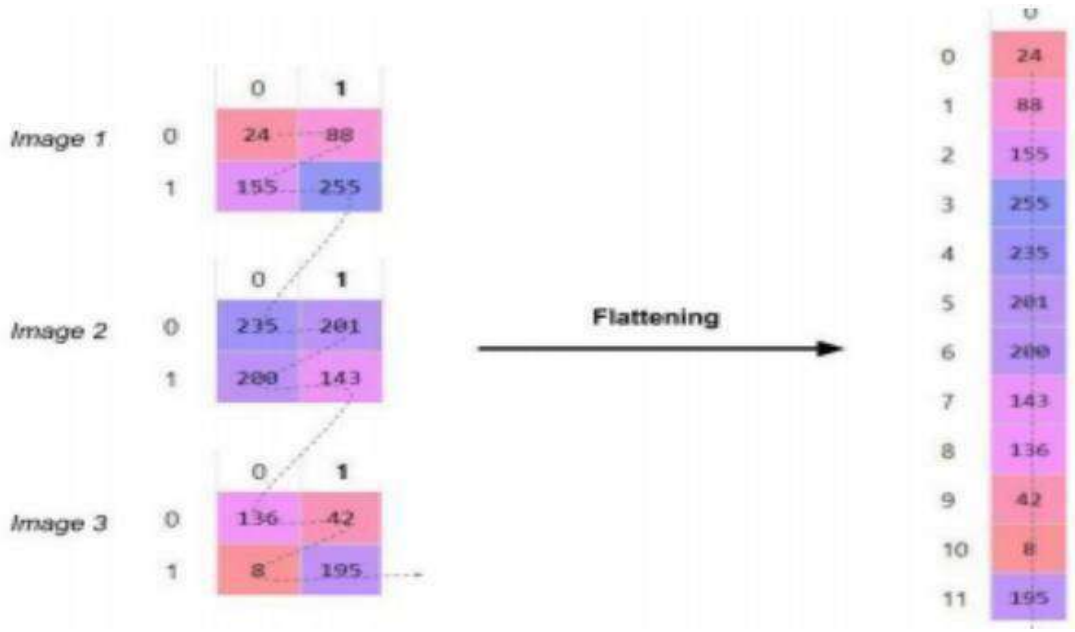


Figure 1.11 Exemple de mise à plat des images finales en sortie (des filtres +simplifications)

3-6.La couche entièrement connectée (FC) :

Après les couches de convolution et pooling, le raisonnement de haut niveau dans le réseau neuronal se fait via des couches totalement connectées. Dans les réseaux de neurones convolutifs, chaque couche agit comme un filtre de détection pour la présence de caractéristiques spécifiques ou de motifs présents dans les données d'origine. Les premières couches d'un réseau convolutif détectent des caractéristiques qui peuvent être reconnues et interprétées facilement. Les couches ultérieures détectent de plus en plus des caractéristiques plus abstraites .La dernière couche du réseau convolutif est capable de faire une classification ultra-spécifique en combinant toutes les caractéristiques spécifiques détectées par les couches précédentes dans les données d'entrée .Les couches totalement connectées font les mêmes tâches que celles des ANN (réseaux de neurone artificielle) standard et tenteront de produire des notes de classe à partir des

activations, pour les utiliser pour la classification. Il est également suggéré d'utiliser ReLu entre ces couches pour améliorer les performances. 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. [5]

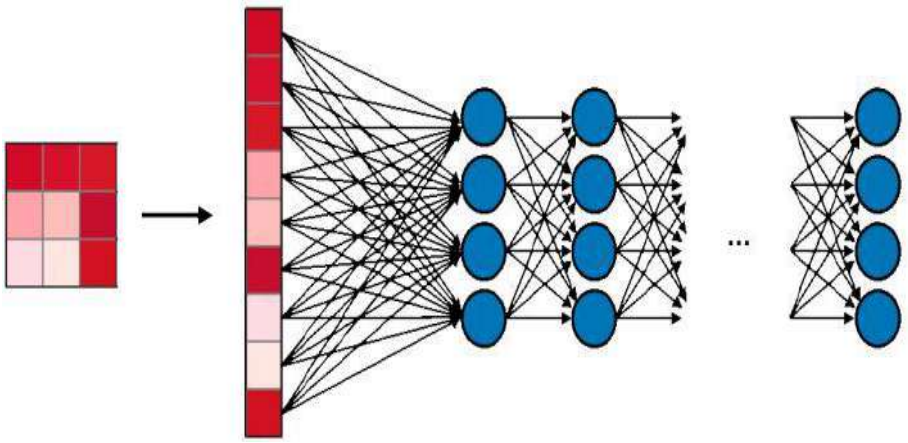


Figure 1.12 un exemple d'opération entièrement connecte (FC)

3-7.La couche de perte (LOSS) :

La couche de perte spécifie comment l'entrainement du réseau pénalise l'écart entre le signal prévu et réel. Elle est normalement la dernière couche dans le réseau. Diverses fonctions adaptées à différentes tâches peuvent y être utilisées. La perte « Soft max » est utilisée pour prédire une seule classe parmi K classes mutuellement exclusives. La perte par entropie croisée sigmoïde est utilisée pour prédire K valeurs de probabilité indépendante dans [0,1]. La perte euclidienne est utilisée pour régresser vers des valeurs réelles. [6]

4-Quelques réseaux convolutifs célèbres : [7]

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

• **AlexNet :** Le premier travail qui a popularisé les réseaux convolutifs dans la vision par ordinateur était AlexNet, développé par Alex Krizhevsky, Ilya Sutskever et Geoff Hinton. Ce CNN été soumis au défi de la base ImageNet en 2012 et a nettement surpassé ses concurrents. Le réseau avait une architecture très similaire à LeNet, mais était plus profond, plus grand et comportait des couches convolutives empilées les unes

sur les autres (auparavant, il était commun de ne disposer que d'une seule couche convolutifs toujours immédiatement suivie d'une couche de pooling).**ZFnet :** C'était une amélioration d'AlexNet en ajustant les hyper-paramètres de l'architecture, en particulier en élargissant la taille des couches convolutifs et en réduisant la taille du noyau sur la première couche.

• **GoogLeNet :** C'est un modèle de Google. Sa principale contribution a été le développement d'un module inception qui a considérablement réduit le nombre de paramètres dans le réseau (4M, par rapport à AlexNet avec 60M). En outre, ce module utilise le global averagepooling ce qui élimine une grande quantité de paramètres. Il existe également plusieurs versions de GoogLeNet, parmi elles, Inception-v4 et Xception.

• **ResNet :**Residual network développé par Kaiming He et al. Été le vainqueur de **ILSVRC 2015**. Il présente des sauts de connexion et une forte utilisation de la batch normalisation. Il utilise aussi le global AVG pooling au lieu du PMC à la fin.

• **VGG Net :** Il s'agit d'une structure du Visual Geometry Group d'Oxford réalisée par Andrea Vedaldi et Andrew Zisserman (en 2017).

• **MobiLeNet :** est une classe de réseaux neuronaux convolutionnels profonds légers qui sont beaucoup plus petits en taille et plus rapides en performances que de nombreux autres modèles populaires. Il utilise des convolutions séparables dans le sens de la profondeur, ce qui signifie essentiellement qu'il effectue une seule convolution sur chaque canal de couleur plutôt que de combiner les trois et de l'aplatir. [28]

• **MobiLeNetV2 :** MobileNetV2 est très similaire au MobileNet d'origine, sauf qu'il utilise des blocs résiduels inversés avec des fonctionnalités de goulot d'étranglement. Il a un nombre de paramètres considérablement inférieur à celui du MobileNet d'origine. Les MobileNet prennent en charge toute taille d'entrée supérieure à 32 x 32, avec des tailles d'image plus grandes offrant de meilleures performances**.** [29]

5-Contexte émotion- Deep Learning :

Il existe plusieurs travaux qui ont exploité le deep Learning pour la reconnaissance des émotions citons :

Mehendale et Ninad proposent une technique appelée la reconnaissance des émotions faciales à l'aide de réseaux neuronaux convolutifs (FERC). La FERC est basée sur un réseau de neurones convolutifs (CNN) en deux parties : La première partie supprime l'arrière-plan de l'image, et la seconde partie se concentre sur l'extraction du vecteur de caractéristiques faciales.

Dans le modèle FERC, le vecteur d'expression (EV) est utilisé pour trouver les cinq types différents d'expression faciale régulière. Supervision ont été obtenues à partir de la base de données stockée de 10.000 images (154 personnes). Il a été possible de mettre correctement en évidence l'émotion avec une précision de 96%, en utilisant un EV de longueur 24 valeurs. Le CNN à deux niveaux fonctionne en série, et la dernière couche du perceptron ajuste les poids et les valeurs de l'exposant

à chaque itération. Le FERC diffère des stratégies généralement suivies avec le CNN à un seul niveau, ce qui améliore la précision.

En outre, une nouvelle procédure de suppression de l'arrière-plan, appliquée avant la génération de l'EV, évite de traiter les multiples problèmes qui peuvent survenir (par exemple la distance de la caméra). FERC a été testé de manière extensive avec plus de 750 000 images en utilisant l'expression étendue de Cohn-Kanade, les visages Caltech, CMU et NIST. Ils sont attendus à ce que la détection des émotions de FERC soit utile dans de nombreuses applications telles que l'apprentissage prédictif, les détecteurs de mensonges, etc. [11]

Ainsi que dans un autre travail de Sreenidhi, Surabi, Dhanya et Sahithi l'apprentissage profond est proposé pour analyser les émotions académiques. Cette méthode combine le réseau de neurones convolutifs (CNN) et le réseau de mémoire à long terme (LSTM) pour l'apprentissage profond. L'apprentissage profond afin de reconnaître et d'analyser le modèle continu d'émotions académiques faciales des étudiants et ainsi reconnaître les émotions scolaires. Grâce à cette méthode, le système d'apprentissage en ligne peut comprendre rapidement et précisément la progression de l'apprentissage des étudiants, et leur proposer des matériels pédagogiques approprié pour améliorer leurs résultats scolaires et leur motivation. Les résultats expérimentaux du site ont montré que la précision de reconnaissance du modèle CNN plus LSTM était de 72,47 et 72,47 respectivement.

LSTM étaient respectivement de 72,47 et 84,33 %. La combinaison de deux réseaux neuronaux a amélioré la précision d'environ 12 % par rapport à celle du modèle CNN seul. [12]

Wenbinzhou, justincheng, xingyu lei, Bedrich Benes et Nicoletta Adamo ont basés sur un réseau neuronal profond convolutif qui classifie les émotions des personnes capturées par une caméra web. Pour leur résultat de classification, ils adoptent le modèle de Russel de l'affect central dans lequel toute émotion particulière peut être placée dans l'un des quatre quadrants suivants : agréable-actif, agréable-inactif, désagréable-actif, et désagréable-inactif. Ils ont recueilli des données à partir de divers ensembles de données qui ont été normalisées et utilisées pour former le modèle d'apprentissage profond.

Ils utilisent les couches entièrement connectées du réseau VGGS qui a été entraîné sur des expressions faciales humaines étiquetées manuellement. Ils ont testé leur application en divisant les données en deux parties 80:20 et en ré-entrainant le modèle. La précision globale du test pour toutes les émotions détectées était de 66%. Ils ont une application fonctionnelle qui est capable de signaler l'état émotionnel de l'utilisateur à environ cinq images par seconde sur un ordinateur portable standard équipé d'une webcam. Le détecteur d'état émotionnel sera intégré dans un système d'agent pédagogique affectif où il servira de retour à un agent éducatif intelligent animer. Servira de retour d'information à un tuteur éducatif intelligent et animé**.** [13]

Neha, Nivya, Pooja HC Sekar et KeerthanaKumar visent à construire un système dont L'objectif est d'analyser l'émotion d'un utilisateur en utilisant une vidéo en temps réel. Ceci est réalisé en utilisant des réseaux neuronaux convolutifs [CNN]. Si l'émotion est analysée comme étant dépression, elle doit être traitée à un stade précoce. Au fur et à mesure que les symptômes s'aggravent, les capacités mentales d'un individu sont devenues incontrôlable, ce qui conduit à un trouble. Si l'émotion est analysée comme étant une dépression, une fenêtre pop-up de chatbot apparaît à l'écran et l'utilisateur peut partager ses sentiments avec le chatbot. Sentiments avec le chatbot réalisé à l'aide de la bibliothèque Tkinter.Cela permet d'améliorer l'humeur de l'utilisateur, d'analyser le niveau de dépression et d'aider l'utilisateur à s'en sortir. Dépression et d'aider l'utilisateur à sortir de cette humeur. Si l'émotion de l'utilisateur est triste, une évaluation continue

est effectuée évaluation continue est effectuée afin de classer la tristesse et la dépression. [14]

Spiers, Daniel LIatas visent à construire un système qui capture des images faciales du monde réel via la caméra frontale sur un ordinateur portable. Le système est capable de traiter / reconnaitre les données capturées image et prédire un résultat en temps réel. Dans ce système, ils exploiter la puissance de la technique d'apprentissage en profondeur pour apprendre un soin du visage modèle de reconnaissance des émotions (FER) basé sur un ensemble des images faciales. Enfin, des expériences sont menées pour évaluer leur modèle utilisant une base de données publique largement utilisée**.** [15]

M.KalpanaChowdary, Tu N.Nguen et D.JudeHemanth proposent des approches d'apprentissage par transfert

Conclusion :

Dans ce chapitre nous avons présenté les notions importantes qui sont en relation avec l'apprentissage profond (définition, architectures…etc.). Ensuite nous avons focalisé notre attention sur les réseaux de neurones convolutifs CNN qui sont exploités pour la réalisation de notre projet. Nous avons présenté après quelques travaux de reconnaissance des émotions basées sur le DeepLearning. Le prochain chapitre est consacré à la présentation du prinicipe des emotions et l'expression faciale..

CHAPITRE II

Emotions et expressions faciales

CHAPITRE II : EMOTIONS ET EXPRESSIONS FACIALES

1-Introduction :

L'expression faciale est l'une des plus puissants moyens de la communication émotionnelle inter humain. Pour automatiser une telle communication nous sommes face à la réalisation d'une interaction homme-ordinateur, pour ce faire l'ordinateur doit être capable d'interagir naturellement avec l'utilisateur, de la même manière que l'interaction homme-homme a lieu.

Dans ce chapitre nous présentons les principales théories émotionnelles, avec un accent particulier sur les expressions faciales. Ensuite, nous décrivons les étapes de traitement au cours desquelles un système d'analyse d'expression faciale peut être décomposé tout en discutant les diverses techniques utilisées à chaque étape, après nous citons les différents systèmes de codifications des émotions. Enfin nous présentons les différentes méthodes de reconnaissance des émotions de base.

2-L'émotion :

2-1.Définition :

Qu'est-ce qu'une émotion ? Cette question n'est pas seulement une interrogation scientifique ; mais c'est aussi le titre d'un célèbre article sur l'émotion écrit en 1884 : « What is an émotion ? ».

Il existe plus d'une centaine de définitions différentes du mot « émotion » et le seul point sur lequel tout le monde s'accorde c'est que le concept est difficile à définir. Une première difficulté vient probablement du fait qu'il s'agit d'un mot du langage populaire. De plus, la définition varie en fonction du point de vue théorique adopté par les différents auteurs. [1]

2-2.L'émotion en philosophie :

En philosophie l'émotion peut se définir principalement comme une manifestation de la vie affective, généralement accompagnée d'un état de conscience agréable ou pénible.

L'émotion est un trouble de durée variable, une rupture d'équilibre. Le trouble est parfois violent, et entraîne une augmentation des mouvements (colère, enthousiasme), ou au contraire, un arrêt des mouvements (peur ou "coup de foudre" en amour).

L'émotion agit donc tantôt à la façon d'un excitant, tantôt à la façon d'un stupéfiant. Les répercussions sur le corps peuvent aller jusqu'à la syncope, mais se limite le plus souvent à des manifestations physiques minimes (rougissement, pâleur, …). Aussi le terme émotion a été définit par autres philosophe comme :

Kant : "L'émotion est le sentiment d'un plaisir ou d'un déplaisir actuel qui ne laisse pas le sujet parvenir à la réflexion. Dans l'émotion, l'esprit surpris par l'impression perd l'empire sur lui-même" (Anthropologie du point de vue pragmatique).

Sartre : "Une émotion est une transformation du monde".

Alain : "L'émotion est un régime de mouvement qui s'établit dans le cœur sans la permission de la volonté, et qui change soudainement la couleur des pensées". [1]

2-3.L'émotion en psychologie :

En psychologie, une émotion est une réaction à un stimulus affectif, environnemental ou psychologique. L'émotion ressentie par rapport à une situation est propre à chaque individu, à son passé et son histoire de vie, à ses capacités intellectuelles, et à son état psychologique. Les émotions impliquent des répercussions physiques du ressentiment psychologique initial : la tristesse peut provoquer les larmes, la peur peut déclencher un cri, ou une perte urinaire parfois, et la joie peut générer un grand sourire, voire même des larmes.

Une même situation implique des émotions différentes suivant l'individu concerné, le contexte et l'implication. L'émotion est à différencier de la sensation qui fait appel à une perception sensorielle, et du sentiment qui est un affectif ressenti, mais en l'absence de manifestation physique.

Toutefois, malgré ces difficultés, une définition semble s'être imposée. En 2013, David et AL. Considèrent que l'émotion est « un processus rapide, focalisé sur un événement et constitué de deux étapes : un mécanisme de déclenchement fondé sur la pertinence de l'événement (par exemple, l'événement est-il pertinent pour mes buts ?), et une réponse émotionnelle à plusieurs composantes (les tendances à l'action, les

réactions du système nerveux autonome contrôlant par exemple le rythme cardiaque, les expressions et les sentiments) ». [1]

2-4.Types d'émotion :

Les émotions sont des séquences courantes et vives qui interviennent tout au long de nos journées et en donnent le ton. Il existe trois types d'émotions : émotions primaires ou dites de base, émotions secondaires et émotions sociales. [17]

A-Émotions primaires :

Les émotions primaires sont déclenchées par des événements particuliers ou bien elles se manifestent dans des circonstances précises en provoquant des comportements spécifiques **(Table 1.1)** elles sont à la base de nos réactions qui ne sont pas seulement déterminées par notre jugement rationnel ou notre passé individuel mais aussi par notre passé ancestral. En fait, ces émotions primaires sont comme une matière première, à partir de laquelle on peut fabriquer toutes les autres émotions.

B-Émotions secondaires :

Les émotions secondaires ont pour base, au départ, un processus de pensée et sont l'aboutissement de l'apprentissage des émotions primaires. Les émotions secondaires sont celles qui sont engendrées à l'évocation de souvenirs et arrivent à maturation à l'âge adulte.

C-Émotions sociales :

On parle aussi des émotions sociales qui sont inhérentes à la relation aux autres, comme la culpabilité, la honte, la jalousie, la timidité, l'humiliation, etc. Toutes ces émotions sont apprises et sont constituées à partir des émotions primaires.

L'éducation et la culture sont fortement impliquées dans l'acquisition des émotions sociales.

2-5-Les composantes de l'émotion :

Une émotion est brève et est toujours déclenchée par un événement spécifique, mais, d'après cette définition, c'est aussi un phénomène dynamique qui présente de multiples composantes.

Kleinginna et AL, ont analysé et classifié 92 définitions, en montrant qu'il y a peu de points communs entre elles et que la plupart sont trop vagues. Ils ont suggéré une définition opératoire que nous utilisons ici : Les émotions sont le résultat de l'interaction de facteurs subjectifs et objectifs, réalisés par des systèmes neuronaux ou endocriniens, qui peuvent :

– Induire des expériences telles que des sentiments d'éveil, de plaisir ou de déplaisir.

– Générer des processus cognitifs tels que des réorientations pertinentes sur le plan perceptif, des évaluations, des étiquetages.

– Activer des ajustements physiologiques globaux.

– Induire des comportements qui sont, le plus souvent, expressifs, dirigés vers un but et adaptatifs.

Dans cette définition, on peut noter la présence de trois composantes fondamentales de l'émotion. [1]

A.La composante comportementale :

Renvoie, quant à elle, à toutes les manifestations comportementales (fuite vs approche) et expressives d'une émotion, qui sont donc dirigées vers l'extérieur. On peut ainsi citer les postures ou la tonalité de la voix. Toutefois, la composante comportementale la plus étudiée est sans doute l'expression faciale, considérée comme le canal majeur de communication émotionnelle tout simplement parce que le visage est une source considérable de « contact » à autrui.

B.La composante physiologique :

Réfère à l'ensemble des manifestations physiologiques concomitantes à un événement émotionnel. On distingue souvent celles liées au système endocrinien, celles du système nerveux autonome, et celles de l'activité cérébrale.

2-6.La composante cognitive/subjective :

Correspond aux changements d'état mental liés à l'émotion. En d'autres termes, cette composante désigne la composante subjective de l'expérience émotionnelle. Pour

illustrer l'interaction entre ces composantes, reprenons le célèbre exemple du psychologue William James :



Figure II.1 Schéma illustratif des composantes de l'émotion

2-7.Canaux de communication émotionnelle :

Depuis les travaux de Picard en 1997, de nombreux systèmes ont été développés pour la reconnaissance d'émotions. Ils sont basés sur divers canaux de communication (visage, voix, geste, réactions physiologiques et nerveuses). Les systèmes de reconnaissance ont donc au fil des années, considéré les différents canaux de communication émotionnelle, en commençant par le visage et la voix, pour ensuite s'intéresser à la gestuelle et enfin aux réactions physiologiques, comme le montre la figure suivante (**Figure 1.2)** : [3]



Figure II.2 Canaux de communication émotionnelle et capteurs associé

2-8.La reconnaissance des émotions :

Il y a Trois domaines principaux où les systèmes de reconnaissance d'émotions se sont développés : [3]

- La reconnaissance des émotions à partir de la parole (langage parlé).
- La reconnaissance des émotions à partir de signaux du système nerveux (EEG).
- La reconnaissance des émotions à partir d'images fixes ou de vidéos

(analyse d'expressions faciales).

Nous nous intéressons à la reconnaissance d'émotions par l'analyse des expressions faciales.

3-Expression faciale :

3-1.Définition :

L'expression faciale est une mimique faciale chargée de sens, c'est un changement dans le visage, perceptible visuellement, dû à l'activation ou mouvements

de l'un ou de plusieurs muscles composant le visage. Ces mouvements transportent l'état émotionnel de l'individu aux observateurs. Les émotions sont au centre de l'explication des expressions faciales.[2]



Figure II.3 Exemple des Expressions Faciales

3-2.Représentation de l'Expression Faciale Emotionnelle :

A-Le modelé de A Russel-valence :

Est une représentation non discrète en liant les expressions l'une à l'autre. Ce modèle été proposé par James A Russel. Il a proposé un système à deux axes bipolaires, à savoir l'arousal (ou excitation) et la valence. La valence va de la tristesse au bonheur, tandis que l'excitation va de l'ennui ou de la somnolence à l'excitation frénétique. Selon Russell, les émotions se situent sur un cercle dans cet

espace bidimensionnel, et sont caractérisées par des catégories floues regroupées sur les axes A Rousal et Valence. La **(Figure I.3)** présente cet espace de représentation des expressions.

Bien qu'un espace continu puisse représenter toutes les expressions possibles, cela n'est pas garanti par la théorie de Russell et n'a d'ailleurs pas été démontré. Il n'est pas clair comment une expression doit être projetée sur l'espace ou, vice versa, comment définir les régions dans l'espace de A Rousal/ Valence qui correspondent à une certaine expression. Ce modèle est assez subjectif, ce qui est problématique lorsque l'on cherche à reconnaître de manière automatique des expressions. Nous sommes donc arrivés à la conclusion que ce système n'est pas adapté pour la vision par ordinateur. [8]

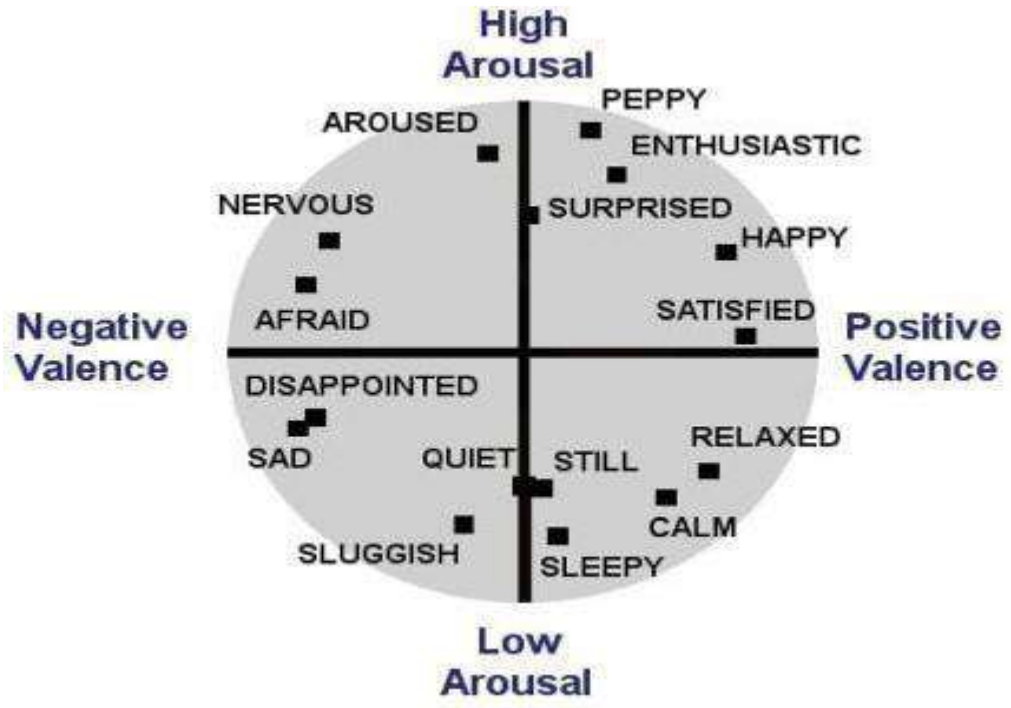


Figure II.4 Espace de modélisation A Rousal - Valence

B-La roue des émotions de Plutchik :

Cette représentation significative des émotions a été proposée par Plutchik.

Dans son modèle, les émotions humaines sont modélisées selon huit émotions

primaires. Les émotions sont organisées en paires d'opposés : la joie contre la tristesse, la confiance contre le dégoût, la peur contre la colère et l'anticipation contre la surprise. La **(Figure 1.5)** représente la roue des émotions de Plutchik qui peut être considérée comme un modèle hybride entre le modèle d'Ekman et le modèle A Rousal/Valence. Ainsi, les émotions les plus complexes sont définies comme une combinaison d'émotions basiques. Par exemple, l'amour est considéré comme la combinaison de la joie et de la confiance. [8

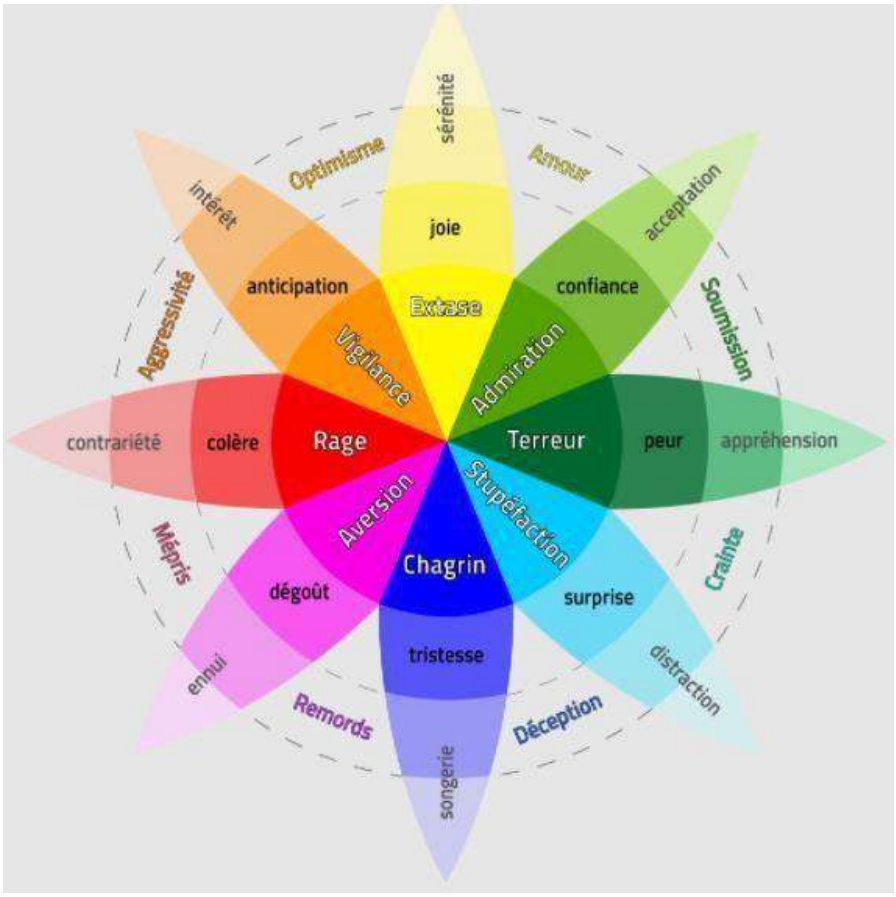


Figure II.5 la roue des émotions de Plutchik

C-Le modèle discret d'Ekman :

Les travaux fondamentaux dans le domaine de l'analyse des expressions du visage sont dus principalement à Charles Darwin, Guillaume Duchenne de Boulogne au XIXème siècle et plus récemment Paul Ekman.

Charles Darwin, est le premier à traiter l'universalité des expressions du visage et à proposer une théorie évolutionniste sur la formation des expressions.

L'argument principal est que les expressions des enfants et des nouveaux nés existent aussi chez les adultes. D'après lui, l'expression des émotions est un processus nécessaire à la survie. Ainsi, les expressions non-verbales sont aussi importantes que les interactions verbales dans le processus de communication humain.

Au XIXème siècle, Guillaume Duchenne de Boulogne est le premier à localiser individuellement les différents muscles faciaux par activation électrique. Il est un des premiers à livrer à la communauté scientifique un ensemble de photographies montrant l'activation des différents muscles faciaux.

Paul Ekman, psychologue, s'intéresse à partir du milieu des années 1960 aux expressions et émotions humaines. Il établit qu'il existe un nombre limité d'expressions reconnues par tous, indépendamment de la culture. Il met donc en évidence.

L'universalité de certaines émotions innées qui correspondent aux six émotions de base :[2]

- **La colère :** La colère est une réaction émotionnelle forte et peut également être une émotion dangereuse car elle pourrait provoquer la violence. La source de la colère a de nombreuses raisons (par exemple nous pourrions ressentir de la colère contre un obstacle sur la voie du succès, quand quelqu'un veut nous faire du mal, les menaces verbales, les réclamations provoquent aussi l'émotion de la colère). La colère a un impact important sur tout le corps. L'augmentation de la pression sanguine, le visage rouge et la tension dans les muscles.

- **La tristesse :** La tristesse apparaît quand une personne souffre. L'origine de la tristesse est typiquement une perte de quelque chose. Cette émotion calme et non impulsive, est souvent accompagnée de larmes.

- **La joie** : La joie est une émotion positive souvent associée à un sourire sur le visage. L'émotion de joie apparaît lors de la réalisation des objectifs.

- **La surprise :** La surprise est une émotion soudaine. Cela vient sans réfléchir et ne dure que peu de temps. Le début est une situation inattendue ou faussementattendue. Par conséquent, une émotion surprenante peut être positive ou négative. La surprise ne peut être anticipée. Si la situation s'accompagne d'un temps de réflexion, la réaction ultérieure ne sera pas une surprise. La surprise entraîne souvent d'autres émotions, en l'occurrence, la joie ou la tristesse.

- **Le dégoût :** Le dégoût est une émotion négative généralement évoquée par l'odorat, le goût ou la vision. Contrairement aux autres émotions, les objets évoquant le

dégoût ne sont pas universels, mais culturels ou personnels, par exemple aliments. La réponse physiologique extrême est le vomissement.

- **La peur :** La peur est induite par des situations dangereuses ou stressantes. On peut ressentir la peur des événements futurs, par exemple peur de la misère. Pour éprouver la peur, le corps de la personne se prépare à une évasion ou à une défense contre toute attaque possible.



Figure II.6 Les six expressions faciales

Lors de la production d'une expression faciale, un ensemble de distorsions faciales apparaît dans les traits permanents du visage ainsi qu'au niveau du cœur et de la respiration et autre niveau, Le tableau suivant illustre ces distorsions.

Table II.1 Les caractéristique des expressions Faciales

Expression faciale	Les déformations
Joie 	Les coins des lèvres tirés vers le haut. La bouche peut être ouverte et les dents peuvent être visibles. Les joues peuvent être soulevées. Les rides sous la paupière inférieure ainsi que les rides aux coins des yeux peuvent apparaître. La respiration est ample. Tout le corps qui est suractivé. Le rythme cardiaque s'accélère.

Dégout 	La lèvre supérieure est levée, plus le dégoût est grand plus la dentition supérieure est visible. Les rides apparaissent sur le nez. Un ralentissement de la fréquence cardiaque. Des réactions de fermeture des sens. Des sensations au niveau gastro-intestinal et au niveau de la gorge, avec les nausées ou l'envie de vomir.
Surprise 	La remontée des parties interne et externe des sourcils. Des rides horizontales pouvant apparaitre sur le front. L'ouverture de la bouche et des yeux.
Colère 	Le visage est contracté : froncement des sourcils, mâchoires serrées, narines pincées. Il peut y avoir rougeur ou pâleur du visage. La personne a des gestes de nervosité. Sa voix est forte. La respiration est courte. Accélération du rythme cardiaque. Augmentation de la pression sanguine.
Peur	Les yeux sont ouverts et les pupilles deviennent larges. Les sourcils sont soulevés et tirés vers l'intérieur. Les paupières supérieures sont soulevées, ce qui a pour effet d'afficher un sanpaku supérieur (c-à-d, un blanc des yeux visible). Les muscles de la bouche se tendent et abaissent souvent la lèvre inférieure. La personne semble fixer quelque chose comme si elle

	ne pouvait s'en détacher. La personne peut ressentir des palpitations, des tremblements, fourmillements dans les membres. - Ralentissement du rythme cardiaque. Respiration rapide. Pression sanguine élevée.
Tristesse 	Remontée de la partie interne des sourcils. - Abaissement et rapprochement des sourcils. - Abaissement des coins externes des lèvres, qui signe la tristesse. - Les sourcils ont une position oblique. - Le regard est absent, les yeux sont brillants, humides, parfois surviennent des pleurs. - Des rides apparaissent sur le front. - Il peut y avoir un sentiment d'oppression. - Ralentissement du rythme cardiaque.

3-3-Architecture de système de reconnaissance des expressions faciale (FER) :

Un système qui effectue une reconnaissance automatique des expressions faciales est généralement composé de trois modules principaux, comme illustre dans la **(Figure 1.7)** le premier module consiste à détecter et enregistrer la région du visage dans les images ou les séquences d'images d'entrée. Il peut s'agir d'un détecteur de visage dans chaque image. Le deuxième module consiste à extraire et représenter les changements faciaux causés par les expressions faciales. Le dernier module détermine une similarité entre l'ensemble des caractéristiques extraites et un ensemble de caractéristiques de référence. D'autres filtres ou modules de prétraitement de données peuvent être utilisés entre ces modules principaux pour améliorer les résultats de détection, d'extraction de caractéristiques ou de classification. [7]

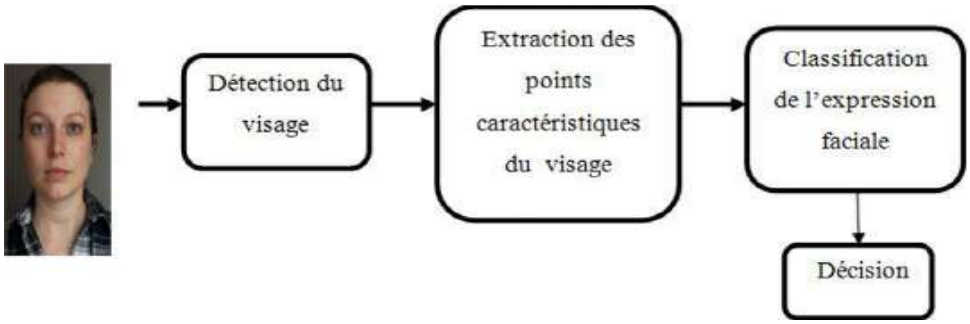


Figure II.7 Architecture de système de reconnaissance des expressions Faciales

3-3.1 .La détection du visage :

La première étape d'un système de reconnaissance automatique des expressions faciales consiste à localiser la région du visage, c'est l'étape de détection du visage. Le rôle principal de cette étape est de déterminer la présence ou pas de visage humain dans une image. [5]



Figure II.8 Détection de visage

A-Caractéristiques d'un détecteur de visage :

Un détecteur de visage idéal doit posséder les caractéristiques suivantes :

• Robustesse : il doit être capable de gérer les variations d'apparence selon la pose, la taille, l'éclairage, l'occlusion, les milieux complexes, les expressions faciales et la résolution.

• Rapidité : il doit être assez rapide pour effectuer le traitement en temps réel.

• Simplicité : il doit être simple. Par exemple, le temps de formation doit être court, le nombre de paramètres doit être petit, et les échantillons d'apprentissage doivent pouvoir être recueillis à moindre coût. [9]

B-Les Méthodes de détection de visage :

Plusieurs méthodes sont développées pour la détection de visage qui sont divisées en quatre catégories : [4]

- Méthodes basées sur des caractéristiques invariantes.
- Méthodes basées sur la mise en correspondance.
- Méthodes basées sur l'apparence.
- Méthodes basées sur les connaissances acquises.

➢ **Méthodes basées sur les caractéristiques invariantes :**

Ce type de méthodes a pour objectif de trouver les caractéristiques structurelles même si le visage est dans différentes positions, conditions lumineuses ou changement d'angle de vue. Le problème avec cette méthode est que la qualité des images peut être sévèrement diminuée à cause de l'illumination, le bruit ou l'occlusion ce qui empêche l'algorithme de fonctionner correctement. Cependant, il existe plusieurs propriétés ou caractéristiques invariantes du visage dont les principales sont les suivantes : [1]

Caractéristiques du visage :

Cette méthode utilise les plans d'arrêtes appelés « Canny detecto » et des heuristiques pour supprimer tous les groupes d'arrêtes sauf celles qui représentent les contours du visage. Une ellipse est déduite comme frontière entre l'arrière-plan et le visage. Celle-ci est décrite comme étant formée des points de discontinuité dans la fonction de luminance (intensité) de l'image. Le principe de base consiste à reconnaitre des objets dans une image à partir de modèle des contours connu au préalable. Pour réaliser cette tâche, deux méthodes peuvent être appliquées : la transformée de Hough et la distance de Hausdorff.

Texture :

La texture de l'être humain est distinctive et peut être utilisée pour séparer les visages par rapport à d'autres objets. Augusteijin et Skufca en 1993 ont développé une méthode de détection de visages sur une image en se basant uniquement sur la texture. Le calcul de cette dernière se fait en utilisant les caractéristiques de second ordre sur des sous- images. Dans cette méthode, trois types de caractéristiques sont prises en considération : la peau, les cheveux et le reste des composants de visage.

Couleur de la peau :

Pour les visages, la couleur de la peau de l'être humain est une caractéristique spécifique, c'est pourquoi elle a été utilisée pour la détection des visages. En effet la couleur de la peau est différente selon les personnes et leur origine (Africain,

Européen, Asiatique...). Dans ce contexte plusieurs études ont démontré que la plus grande différence s'étend largement entre intensité plutôt que la chrominance lumineuse.

Caractéristiques multiples :

Il existe plusieurs méthodes qui combinent les différentes caractéristiques faciales pour la détection des visages. La majorité entre elles adoptent des propriétés globales : couleur de la peau, la forme, la taille du visage, pour trouver les candidats puis les vérifier localement en se basant sur les caractéristiques détaillées comme les yeux, les sourcilles, le nez et les cheveux. On trouve dans le contexte de reconnaissance des émotions, beaucoup de travaux utilisant ce type de méthode. Citons par exemple le travail de Singh et AL en 2003.

- **Méthodes basées sur la mise en correspondance :**

Des modèles caractéristiques d'un visage entier ou de sous-partie de visage (bouche, Œil, nez) sont créés. La localisation se fait par corrélation de ces modèles avec les candidats. On trouve deux techniques : [4]

Modèle prédéfini :

Les sous-gabarits souvent utilisés pour la détection frontale sont les yeux, le nez, la bouche et les contours du visage. Chaque sous-gabarit est défini par une segmentation des lignes. On commence par une comparaison entre les sous-images et les contours de gabarit pour détecter les visages candidats. Puis on met les sous gabarits comme les yeux et le nez sur les positions candidates. On peut conclure que la détection s'effectue en deux étapes : la première consiste à la détection des régions candidates, la deuxième est l'examen des détails pour déterminer les caractéristiques du visage. Cette méthode est utilisée par Sakai et AL. [1]

Modèle déformable :

Cette approche est utilisée dans le but de modéliser les caractéristiques faciales

qui s'adaptent élastiquement par rapport au modèle du visage présent. Dans cette méthode, les caractéristiques faciales sont décrites par des gabarits paramétrés. Une fonction est définie pour relier les contours, les sommets et les angles dans l'image d'entrée, pour faire correspondre les paramètres sur les gabarits. La meilleure adaptation du modèle élastique est de trouver la fonction énergétique en minimisant les paramètres. Cette méthode a été utilisée par Kanade et AL. [1]

- **Méthode basée sur les apparences :**

Ces méthodes utilisent le même principe que celui présenté au point précédent mais se basent sur des modèles appris à partir d'un ensemble d'essai. Ces méthodes présentent l'avantage de s'exécuter très rapidement mais demandent un long temps d'entrainement. Les méthodes appartenant à cette catégorie ont montré de bons résultats par rapport aux trois autres types de méthodes. On peut citer parmi celles-ci, la méthode basée sur les réseaux de neurones de Rowley et AL, la méthode de Schneiderman et Kanade basée sur un classifieur de Bayes naif ainsi que le fameux algorithme de Viola et Jones fonctionnant en temps réel. [4]

3-3 .2 .Extraction de la caractéristique faciale :

Les points caractéristiques du visage sont principalement situés autour des composants faciaux tels que les yeux, la bouche, les sourcils, le nez et le menton. La détection des points caractéristiques du visage commence habituellement à partir d'une boite englobant rectangulaire renvoyée par un détecteur de visage qui localise ce

dernier. L'extraction de caractéristiques géométriques telles que les contours des composants faciaux, les distances faciales, etc. fournit les emplacements ou les caractéristiques d'apparence peuvent être calculées. De ce fait, les méthodes d'extraction des caractéristiques pour l'analyse d'expression peuvent être séparés en deux types d'approches : les méthodes basées sur les caractéristiques géométriques et les méthodes basées sur l'apparence :

A-Les caractéristiques géométriques :

Représentent la forme et l'emplacement des composants du visage (y compris la bouche, les yeux, les sourcils et le nez). Les composants faciaux ou les traits

faciaux sont extraits pour former un vecteur de caractéristiques représentant la géométrie du visage. Cette approche contient des différents modèles pour obtenir la reconnaissance des expressions faciales comme :

- Modèle de forme active (ASM).
- Modèles d'apparence active (AAM). [4]

B-Les caractéristiques d'apparence :

Représentent les changements d'apparence (texture de la peau) du visage, tels que les rides et les sillons. Ces caractéristiques d'apparence peuvent être extraites sur tout le visage ou sur des régions spécifiques du visage. Selon les différentes méthodes d'extraction des caractéristiques, les effets de la rotation de la tête dans le plan et les différentes échelles de prise de vue du visage peuvent être éliminés par une normalisation de ce dernier avant l'extraction des caractéristiques ou par une représentation des caractéristiques avant l'étape de reconnaissance d'expression.

Diverses techniques ont déjà été proposées pour l'extraction des caractéristiques faciales. Elles peuvent être classées en 3 grandes familles : analyse bas niveau, analyse intermédiaire et analyse haut niveau. [5][1]

- A**nalyse bas niveau :**

Les techniques de bas niveau supposent que les pixels de l'objet à segmenter possèdent des caractéristiques homogènes et différentes du fond, la segmentation peut être effectuée par l'identification et la séparation des classes objets et fond. Pour cela, différentes solutions ont été proposées. Certaines utilisent un simple seuillage d'une

grandeur colorimétrique, alors que d'autres mettent en œuvre des techniques de classification plus évoluées.

Kapmann et Zhang ont exploité leurs travaux antérieurs qui sont basés sur la corrélation entre l'image originale et des modèles de coins d'œil et de la bouche, pour l'extraction des coins de l'œil afin de les représenter avec un modèle déformable. [1]

- **Analyse intermédiaire :**

A un niveau d'analyse intermédiaire, on cherche à détecter des caractéristiques indépendantes des conditions lumineuses et de l'orientation des visages. Les contours actifs (ou snakes) introduits par Kass et Witkin à la fin des années 80 sont des courbes qui peuvent se déformer progressivement de manière à s'approcher au plus près du contour d'un objet. Cette déformation est guidée par la minimisation d'une fonction d'énergie. [1]

- **Analyse haut niveau :**

Les méthodes précédentes d'extraction des caractéristiques faciales sont des processus à forme libre. Elles n'intègrent aucune connaissance a priori sur les formes admissibles. Contrairement aux méthodes de haut niveau qui sont basées sur des modèles de caractéristiques des formes à segmenter, obtenus de manière heuristique ou statique. Ces modèles génériques sont déformés de manière à être adaptés aux contours de l'objet.

Plusieurs travaux ont utilisé les modèles actifs d'apparence (Active Appearence Models AAM) qui permettent de construire un modèle statique de l'objet à segmenter incluant à la fois la forme et les niveaux de gris. [1]

3-3.3.La classification :

La dernière étape d'un système automatique d'analyse d'expression faciale est la reconnaissance de l'expression faciale en fonction des caractéristiques extraites. Certains systèmes classent directement les expressions. De nombreux classifieurs ont été appliqués à la reconnaissance d'expression tels que : [4]

- Réseaux de neurone (Neural Networks, NN).
- Machines à vecteurs de support (Support Vector Machine, SVM).

- Analyse discriminante linéaire (Linear discriminant analysis, LDA).
- K-plus proche voisin (K Nearest Neighbor, KNN).
- Régression logistique multinomiale (Multinomial Regression Logistic, MRL).
- Modèles de markov cachés (Hidden Markov Model, HMM).
- Réseaux bayésiens (Bayesian Network, BN).

4-Système de codification des expressions faciale :

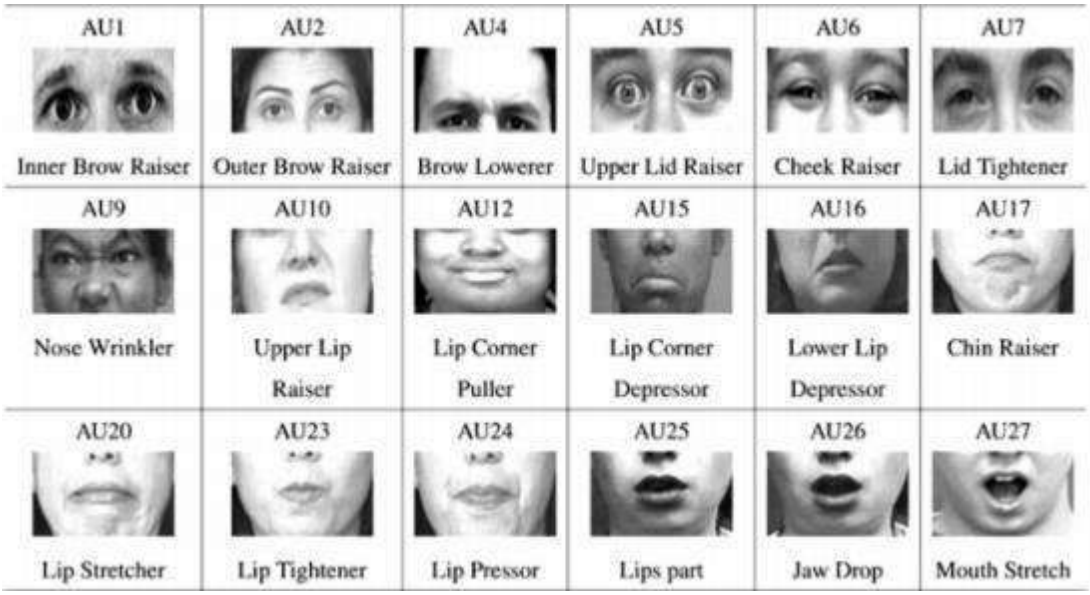
Les représentations faciales relient les mouvements des muscles faciaux aux émotions exprimées. Elles représentent une sorte de dictionnaire utile à la reconnaissance des émotions.

4.1.Représentation par le système FACS :

On se basant sur l'universalité, Ekman et Lzard ont développé des méthodes de mesure des comportements du visage. En particulier, ils ont créé le système FACS (Facial Action Coding System), largement utilisé et reconnu. Il utilise une quarantaine de caractéristiques anatomiques indépendantes, et définit une taxonomie de toutes les expressions faciales.

Le système FACS est probablement le standard le plus populaire utilisé pour classifier systématiquement les expressions physiques des émotions du visage, et est utilisé par des psychologues mais aussi par des infographistes. Ce système définit 46 Unités d'action, qui sont autant de contraction ou de relaxation d'un ou plusieurs muscles, et dont l'association définit une expression faciale. La philosophie de base du système consistait à former des experts pour la reconnaissance et l'interprétation des unités d'action, mais désormais, le système est aussi utilisé pour automatiser la reconnaissance des unités d'action et donc des expressions, ainsi que pour la simulation graphique de visage.

La **(Figure II.9)** illustre quelques images des AUS et la description de chaque unité d'action faciale.



Bien que FACS soit un système de description bénéficiant d'un grand succès, il souffre cependant de quelques inconvénients : [15]

- Complexité : Il faut 100 heures d'apprentissage pour maîtriser les principaux concepts.

- Difficulté de manipulation par une machine : FACS a d'abord été créé pour des psychologues, certaines mesures restent floues et difficilement évaluables par une machine.

- Unité d'action peuvent se produire dans plus de 7000 combinaisons complexes différentes entrainant des bombements et divers mouvements d'entrée et de sortie des images planes de caractéristiques faciales permanents qui sont difficiles à détecter dans les images faciales à deux dimensions.

4.2.MPEG4 :

La norme de codage vidéo MPEG-4 dispose d'un modèle du visage humain développé par le groupe d'intérêt Face and Body AdHoc Group. C'est un modèle 3D

articulé. Ce modèle est construit sur un ensemble d'attributs faciaux, appelés Facial Feature Points (FFP). Des mesures sur ces FFP sont effectuées pour former des

unités de mesure (Facial Animation Parameter Units) qui servent à la description des mouvements musculaires (Facial Animation Parameters - équivalents des Actions Unitaires d'Ekman).

Les Facial Animation Parameter Units (FAPU) permettent de définir des mouvements élémentaires du visage ayant un aspect naturel. En effet, il est difficile de définir les mouvements élémentaires des muscles de manière absolue : le déplacement absolu des muscles d'une personne à l'autre change, mais leurs déplacements relatifs à certaines mesures pertinentes sont constants. C'est ce qui permet d'animer des visages de manière réaliste et peut permettre de donner des expressions humaines à des personnages non-humains. Comme exemples de FAPU, on peut citer la largeur de la bouche, la distance de séparation entre la bouche et le nez, la distance de séparation entre les yeux et le nez, etc. Par exemple, l'étirement du coin de la lèvre gauche (Facial Animation Parameter 6 stretch corner lip) est défini comme le déplacement vers la droite du coin de la lèvre gauche d'une distance égale à la longueur de la bouche. Les FAPUs sont donc des mesures qui permettent de décrire des mouvements élémentaires et donc des animations. [2]

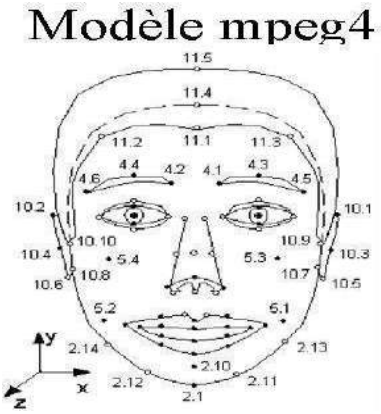


Figure II.10 modèle du visage MPEG4

4.3.Candide :

Est un modèle du visage. Il est composé d'un modèle en fil de fer représentant un visage générique et d'un ensemble de paramètres :

A-Paramètres de forme (Shape Units) :

Ces paramètres permettent d'adapter le modèle générique à un individu particulier. Ils représentent les différences inter-individus et sont au nombre de 12 :

- Hauteur de la tête.
- Position verticale des sourcils.
- Position verticale des yeux.
- Largeur des yeux.
- Hauteur des yeux.
- Distance de séparation des yeux.
- Profondeur des joues.
- Profondeur du nez.
- Position verticale du nez.
- Degré de courbure du nez (s'il pointe vers le haut ou non).
- Position verticale de la bouche.
- Largeur de la bouche.

B-Paramètres d'animation (Animation Units) :

Ces paramètres représentent les différences intra-individus. Les différentes actions faciales. Ils sont composés d'un sous-ensemble de FACS et d'un sous-ensemble des FAPs de MPEG-4. Les FAPs sont définis par rapport à leur FAPU correspondant. Ces paramètres, qu'ils soient d'animation ou de forme, sont représentés sous forme d'une liste de points du modèle de fil de fer à mettre à jour.

Candide permet de voir clairement la différence entre les AUs de FACS et les FAPs de MPEG-4 : les Aus de FACS sont exprimées de manière absolue, à la différence des FAPs qui sont exprimés par rapport à des mesures du visage (les FAPUs). [2]

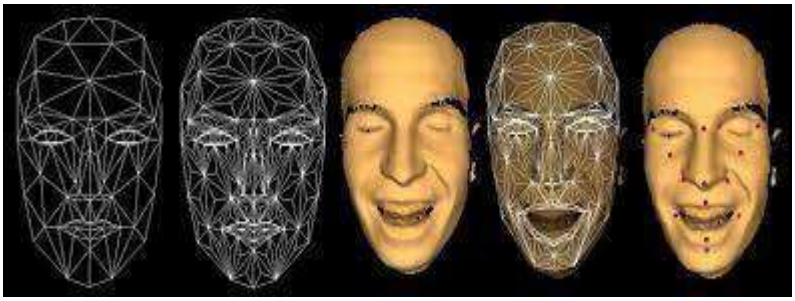


Figure II.11 Version 3D du modèle candide

4.4-Domaines d'applications :

De nombreuses études ont été menées sur l'analyse automatique de l'expression faciale en raison de son importance pratique en robotique sociable, traitement médical, surveillance de la fatigue du conducteur et de nombreuses autres interactions homme-machine systèmes. Ainsi que dans le domaine de la vision par ordinateur et de l'apprentissage automatique. Divers systèmes de reconnaissance de l'expression faciale (FER) ont été explorés pour coder les informations d'expression à partir de représentations faciales. [7]

Par exemple, la reconnaissance précise des émotions faciales est importante en psychiatrie pour comprendre les interactions avec le patient, au-delà des mots.

Pour les nouveau-nés, l'évaluation de la douleur par l'analyse de l'expression du visage (système néonatal de codage facial) donne aussi d'importantes informations d'ordre médical.

Les découvertes psychologiques des mimiques faciales sont également utilisées par les polices à travers le monde : Scotland Yard (Grande Bretagne), les services canadiens, le FBI et la CIA (Etats-Unis), etc. Aussi, des centaines d'enquêteurs ont été formés à démêler le vrai du faux :« Plus la personne parle, plus il est facile de dire si elle ment, parce qu'une bonne partie de son cerveau étant occupée à faire des phrases,

elle laisse échapper toutes sortes de choses » soutient Ekman. Ou à identifier un danger imminent :

« Les visages reflètent des expressions qui sont un signal d'avertissement d'intention dangereuse » précise-t-il encore.

La justice se sert également des mêmes méthodes. Clignement ou mouvement des yeux, froncement du front ou des sourcils, dilatation des pupilles, rougissement ou blêmissement facial, œil gauche ou droit plus grand : tous ces signes perçus et interprétés avec justesse livrent à l'avocat initié une foule d'informations précieuses sur la crédibilité de son interlocuteur devant la Cour ou en négociation. « Toutes ces observations peuvent orienter la négociation ou la plaidoirie de l'avocat » fait valoir Mme Gagnon, juriste canadien. Certains experts révèlent qu'en cas de contradiction entre les mots et les attitudes, ils choisissent de se fier principalement aux comportements qui représentent plus de la moitié du langage non verbal. Cependant, certains critiquent des risques de dérives.

Autre domaine d'application : l'emploi. L'expression faciale a une influence directe sur l'impression que la personne donne d'elle-même à ses interlocuteurs dans le cadre d'une entrevue. Lors d'un entretien d'embauche, le recruteur peut percevoir dans le langage non verbal du candidat des indications sur son caractère, son émotivité, sa franchise. Si ses yeux se promènent un peu partout dans la salle, si son regard est vide ou bien encore si elle fixe le plancher, elle démontre soit du désintérêt, soit un manque de confiance en elle. Le candidat devient alors moins intéressant pour l'employeur potentiel. Pour réussir un entretien d'embauche, il faut regarder son interlocuteur avec intérêt, les yeux clairement ouverts et sans peur de regarder l'autre en face mais sans le fixer constamment. Un contact oculaire est une bonne façon de s'attirer la sympathie du recruteur. Le sourire pas seulement avec la bouche mais avec tout votre visage aussi s'avère contagieux. Il traduit l'aisance, ce que recherche l'employeur. Voir le sourire de quelqu'un se transposer dans ses yeux est une excellente façon de démontrer qu'on est amical et ouvert à la conversation.

Il en est de même dans le monde des affaires, des négociations. Les techniques de langage non verbal permettent de décrypter les intentons réelles ou cachées de la partie adverse. Des indications indispensables aux implications cruciales.

Pour terminer, le monde de l'art, verbal ou visuel, repose aussi en partie sur la reconnaissance des expressions faciales. Le mime, sans parole, est capable de

transmettre des émotions par de simples attitudes. Tout comme les personnages de bandes dessinées et dessins animés, également croqués avec des traits immédiatement identifiables, sans avoir besoin du texte ou du son. Leurs faciès caricaturaux forcent leur état d'esprit. Enfin, les comédiens utilisent également les expressions faciales pour raconter les émotions intérieures du rôle qu'ils interprètent.

Ainsi, quel que soit l'événement de la vie, nous sommes suspendus malgré nous ou intentionnellement à la traduction faciale de nos émotions les plus intimes et profonde. [30]

5-Méthodes de reconnaissance des émotions de base:

Elles peuvent être divisées globalement en quatre approches principales qui sont : l'approche basée modèles, l'approche basée règles, l'approche basée géométrie et l'approche basée réseaux de neurones.

5-1.Approche Basée Modèles :

Cette approche consiste à voir le visage comme un tout. L'analyse consiste alors à mesurer la ressemblance du visage observé à un modèle (connu ou appris). Les méthodes de cette approche font généralement appel à des méthodes de mise en correspondance de modèles. L'intérêt de ces méthodes est qu'elles peuvent être appliquées de manière plus locale. Ainsi la majorité de ces méthodes qui sont utilisées pour le visage sont aussi utile pour n'importe quel objet et plus particulièrement pour les composantes du visage. [9]

Avantages :

Les avantages principaux de ces méthodes sont :

- Qu'elles peuvent donc s'adapter à beaucoup de problèmes.
- Ces méthodes sont généralement moins sensibles au bruit que les méthodes classiques, puisqu'il existe un modèle sous-jacent de comparaison.

• En plus elles sont robustes par rapport à l'occultation pour des visages de la base d'apprentissage et même sur des visages non connus (hors de la base

d'apprentissage).

Inconvénients :

• Lorsqu'on veut appliquer ces méthodes sur des différentes poses ou illuminations nous sommes besoins d'une base de données d'apprentissage alors la construction de cette base devient difficile.

• De plus, la taille de la matrice d'apprentissage est fixée donc les images doivent être redimensionnées alors la construction d'un modèle adapté à ce problème est très longue.

• Ces méthodes ne nous donnent pas des mesures mais seulement des paramètres, par exemple il est difficile d'implémenter une tache qui indique quel est le degré, en pixel, de la bouche, tant qu'il est facile d'implémenter une tache qui décide si la bouche est ouverte ou fermée.

• Ces méthodes sont donc bien adaptées à l'extraction d'informations sur des composantes dont certains états sont difficiles à déterminer (le gonflement de la joue par exemple).

5-2.Approche Géométrique :

Depuis peu de temps, les informaticiens utilisent l'analyse géométrique pour représenter l'information faciale car elle donne des mesures précises qui ne donnent pas les approches basée modèle. Dans cette approche, les mouvements sont détectés à l'aide des calcule des distances entres les points caractéristiques entre deux images l'une à l'état neutre et l'autre expressive.

Lien et AL proposent une méthode hybride basée sur : premièrement suivi des points faciaux (points autour des contours des yeux, des sourcils, du nez et de la bouche manuellement localisés sur la première image de la séquence vidéo), deuxièmement, le flux optique et troisièmement détection des rides afin d'extraire des informations sur l'expression.

Tian et AL utilisent deux réseaux de neurones séparés afin de reconnaitre 6 unités d'actions de la partie supérieure du visage et 10 de la partie inférieure, leur méthode est basée sur les traits permanents du visage ainsi que les traits

transitoires. Les traits faciaux sont initialisés manuellement sur la première image et suivis dans le reste des images de la séquence vidéo. La reconnaissance des expressions faciales est réalisée par la combinaison des unités d'actions des deux parties supérieure et inférieure du visage.

Cohen et AL développent un système basé sur un algorithme de suivi des traits faciaux afin d'extraire les mouvements locaux des composantes faciales. Ces mouvements forment les entrées d'un réseau bayésien utilisé pour reconnaitre les six expressions faciales. La représentation basée composantes requiert une détection efficace et fiable des composantes faciales pour faire face à la variation de l'illumination, du mouvement principal significatif et de la rotation. [9]

Inconvénients :

- Le problème majeur des méthodes de géométrique est la détection non exacte des points caractéristiques du visage sur lesquels des mesures géométriques seront calculées à cause des variations de l'illumination, de l'occlusion et de beaucoup d'autres facteurs.

5-3.Approche Basée Règles :

Une seule parmi les méthodes rencontrées dans la littérature est basée sur des règles. Elle a été proposée par Pantic et Rothkrant. La méthode réalise une codification automatique en unités d'action. Un multi détecteur automatique de composantes du visage est appliqué. Depuis les contours des composantes, des modèles de composantes sont extraits. La différence est ensuite calculée entre le modèle détecté et le modèle correspondant du visage à l'état neutre. Basé sur les connaissances acquises de FACS, les règles de production classifient le modèle détecté dans la classe AUs appropriée. La méthode proposée a été testé sur un ensemble de 49 vues. Les taux de reconnaissance étaient de 92% pour les unités d'actions de la partie supérieure du visage et 86% pour celles du bas du visage. La classification finale dans une des six catégories correspondantes aux six expressions universelles est réalisée à l'aide de la description

linguistique donnée par Eckman. La performance globale du système a été testée sur l'ensemble de 265

vues représentant six expressions faciales affichées par 8 sujets. Le taux de reconnaissance réalisé était de 91%. [9]

Inconvénients :

- Le problème majeur de ces méthodes est qu'aucune méthode n'arrive à reconnaitre les 46 unités d'actions définies dans FACS.

5-4.Approches Basées Réseaux de Neurones :

Un réseau de neurones peut être vu comme une fonction ayant un certain nombre d'entrées et un certain nombre de sorties. Le principe de l'apprentissage est de donner en entrée du réseau un certain nombre d'exemples et de fixer la sortie à la valeur désirée. Une méthode d'apprentissage permet alors au neurone de s'adapter au mieux pour qu'il affiche la même sortie quand on lui donnera des donnés proches des données d'apprentissage. [2]

Avantages :

- Les réseaux de neurones sont généralement utilisés pour leur faible sensibilité au bruit (robustesse au bruit) et leur capacité d'apprentissage.

Inconvénients :

- La plupart de ces méthodes sont testées uniquement sur des images utilisées au cours de l'apprentissage, c'est pourquoi on ne peut prévoir le comportement de ces méthodes dans le cas de sujets inconnus. En outre, la plupart de ces méthodes exigent une intervention manuelle.

6-Contexte de travail :

Plusieurs travaux s'intéressent à la reconnaissance des émotions vue l'importance déjà citée de ce domaine. Nous présentons dans ce qui suit quelques-uns :

Qi-rong Mao, Xin-yu Pan, Yong-zhao Zhan et Xiang-jun Shen proposent une

approche de reconnaissance des émotions en temps réel basée sur les caractéristiques d'expression faciale 2D et 3D capturées par les capteurs Kinect. Pour capturer la déformation du maillage 3D pendant l'expression faciale, ils combinent les caractéristiques des unités d'animation (AUs) et des positions des points de caractéristique (FPPs) suivies par Kinect. Un algorithme de fusion basé sur des profils émotionnels améliorés (PEI) et une confiance maximale sont proposé pour reconnaître les émotions avec ces caractéristiques d'expression faciale en temps réel. Des expériences sur un ensemble de données d'émotions et une vidéo en temps réel montrent les performances supérieures de leur méthode. [18]

Matthew Ratliff et Eric Patterson discutent d'un cadre pour la classification des états émotionnels, basé sur des images fixes du visage. La technique qu'ils présentent implique la création d'un modèle d'apparence active (AAM) formé sur des images de visage à partir d'une base de données accessible au public pour représenter la clé de variation de forme et de texture à la reconnaissance d'expression. Les paramètres de l'AAM sont utilisés comme caractéristiques pour un schéma de classification capable d'identifier avec succès les visages liés aux six émotions universelles. Les résultats de leur étude démontrent l'efficacité des AAM pour capturer l'importante structure faciale pour l'identification de l'expression et aident également à suggérer un cadre pour le développement futur. [19]

Miss Sneha et Dr Desai proposent une technique de reconnaissance des émotions faciales, cette technique utilise un LBP modifié, qui effectuent une comparaison horizontale et verticale des pixels de voisinage, pour générer une faciale initiale discriminante représentation. Ensuite, un PSO modifié est proposé pour effectuer l'optimisation des fonctionnalités. Le système de reconnaissance des émotions faciales se compose de trois étapes : 1) Extraction de caractéristiques ;

2) Optimisation des fonctionnalités ; 3) Reconnaissance des émotions. Premièrement, ils utilisent des motifs binaires locaux modifiés (LBP), c'est- à-dire des voisinages horizontaux et verticaux. Comparaison LBP, pour extraire la représentation faciale initiale. Ensuite, l'algorithme PSO proposé est utilisé pour identifier les caractéristiques discriminantes et significatives pour différencier les expressions faciales

distinctes. SVM (Machine à vecteur de soutien) et Les classificateurs SVM multiples sont utilisés pour reconnaître six expressions faciales.

: 1) Bnheur ; 2) Tristesse ; 3) Colère ; 4) la peur ; 5) Surprise ; 6) Le dégoût. [20]

6-Conclusion :

Dans ce chapitre nous avons expliqué dans un premier temps la définition des émotions. Ensuite, nous avons discuté les théories et les représentations les plus connues dans la reconnaissance des expressions faciales à savoir : les approches de détection de visage dans des images ainsi que l'extraction des caractéristiques et les techniques de codifications. Et enfin, les méthodes de reconnaissance des émotions de base.

En se basant sur l'état de l'art présenté dans ce chapitre nous avons choisi le domaine des réseaux de neurones pour la reconnaissance des émotions de base et particulièrement les réseaux de neurones convolutifs.

Dans le chapitre suivant nous présenterons l'apprentissage profond en citant les différents types de réseaux de neurones et on expliquant en détailles notre approche de reconnaissance des émotions basée CNN.

CHAPITRE III

CONCEPTION ET IMPLEMENTATION

CHAPITRE 3 : CONCEPTION ET IMPLEMENTATION

1-Introduction :

La reconnaissance des expressions faciales est un problème important en vision artificielle, et plusieurs méthodes ont été utilisées pour le résoudre, telles que : le deep learning (principalement CNN).

Dans ce chapitre, nous décrivons en détail la conception de notre méthode proposée pour développer un système de base de reconnaissance des émotions. Par conséquent, nous décrirons les outils (logiciels et matériels) utilisés dans la mise en œuvre.

La dernière partie traite des résultats obtenus et de quelques captures d'écran de notre application.

2-Conception :

2-1.Présentation de notre système de reconnaissance des émotions:

Nous proposons la conception d'un système de reconnaissance des émotions de base en temps réel basé sur CNN. Le système consiste à détecter des visages humains à travers des caméras, pour apprendre les expressions faciales avec un taux de précision qui correspond à sept expressions courantes, à savoir la joie, le dégoût, la peur, la colère, la tristesse, la surprise et les états neutres.

2-2.Architecture globale de notre système de reconnaissance des émotions :

Toutes les solutions de reconnaissance des émotions sont construites selon la même architecture globale, divisée en trois modules principaux qui fonctionnent indépendamment :

- ✓ Détection de visage.
- ✓ Extraction de fonctionnalités.
- ✓ Classement.

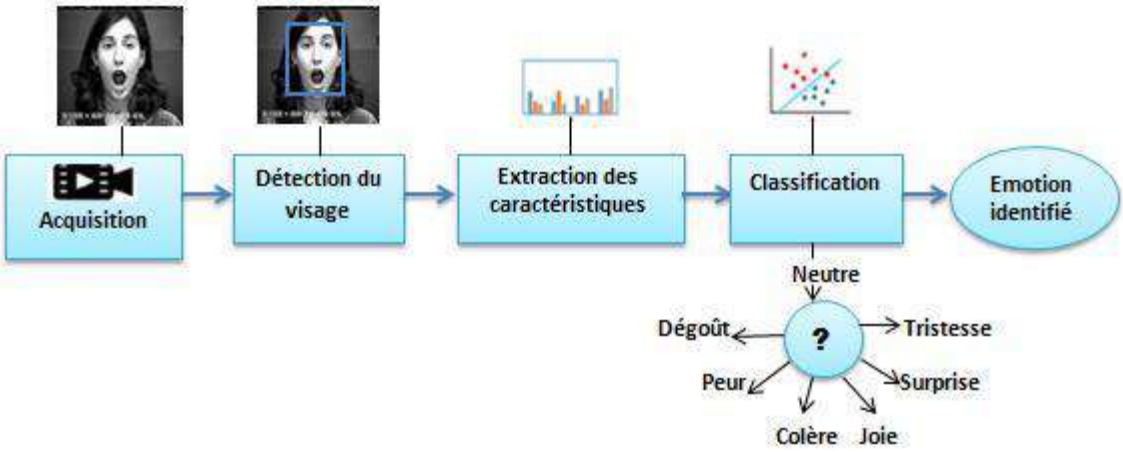


Figure III.1 architecture globale de notre système de reconnaissance des émotions

2-3.Conception détaillée de système :

Dans ce qui suit, nous allons détailler chacune des étapes décrites ci-dessus.

A-Détection du visage :

Le problème de détection de visage consiste à identifier la présence d'un visage dans une image et à déterminer l'emplacement et l'échelle du visage.

L'efficacité des systèmes de reconnaissance des émotions dépend essentiellement de la méthode utilisée pour localiser le visage dans l'image. Dans notre méthode, nous utilisons l'algorithme Viola-Jones pour détecter diverses parties du visage humain telles que la bouche, les yeux, le nez, les narines, les sourcils, les lèvres et oreilles.

Principe de la méthode de Viola et Jones :

La méthode de Viola et Jones consiste à balayer une image à l'aide d'une fenêtre de détection de taille initiale 24px par 24px (dans l'algorithme original) et de déterminer si un visage y est présent. Lorsque l'image a été parcourue entièrement, la taille de la fenêtre est augmentée et le balayage recommence, jusqu'à ce que la fenêtre fasse la taille de l'image. L'augmentation de la taille de la fenêtre se fait par un facteur

multiplicatif de 1.25. Le balayage, quant à lui, consiste simplement à décaler la fenêtre d'un pixel.

Ce décalage peut être changé afin d'accélérer le processus, mais un décalage d'un pixel assure une précision maximale. Cette méthode est une approche basée sur l'apparence, qui consiste à parcourir l'ensemble de l'image en calculant un certain nombre de caractéristiques dans des zones rectangulaires qui se chevauchent. Elle a la particularité d'utiliser des caractéristiques très simples, mais très nombreuses.

La bibliothèque OpenCV présente une implémentation de cet algorithme sous le nom « détecteur en cascades de Haar ». Le point fort de cette méthode est la rapidité de détection ce qui la rend capable de s'exécuter en temps réel et de répondre aux exigences du traitement vidéo.

Toutefois, elle présente quelques limites telles que la difficulté de détection simultanée du plusieurs vues de même objet, la durée nécessaire à la phase d'apprentissage des cascades est relativement assez grande et le nombre d'échantillons d'apprentissage est important.

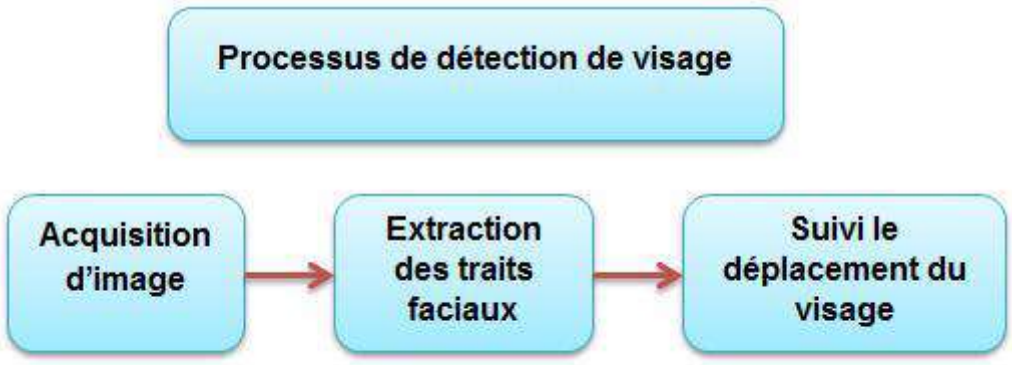


Figure III.3 Processus de détection de visage

B- Extraction de caractéristiques faciales :

Une fois le visage est détecté dans l'image entrée, le système lance le processus d'extraction des caractéristiques qui va convertir les données des pixels

à des représentations et configurations plus réduites et optimales, pour que la représentation extraite soit utilisée dans le processus de la classification, cette

étape réduit les dimensions de l'image en entrée en gardant les données les plus utiles pour la classification.

L'étape d'extraction représente le cœur du système de reconnaissance, on extrait de l'image les informations qui seront sauvegardées en mémoire pour être utilisées plus tard dans la phase de décision (classification).

C- la reconnaissance des émotions faciales à l'aide d'un CNN :

Pour créer notre modele de reconnaissance des émotions , nous avons suivre les etapes suivantes :

Étape 1 : Préparation du Data set :

- nous avons utilisé un dataset disponible en ligne « **fer2013 »** Les données consistent en des images de visages en niveaux de gris de 48*48 pixels. Ce dataset est divisé en deux ensembles de donné un pour l'entrainement et l'autre pour le teste.
- La tâche consiste à classer chaque visage dans l'une des sept catégories (0 = colère, 1 = dégoût, 2 = peur, 3 = joie, 4 = tristesse, 5 = surprise, 6 = neutralité) en fonction de l'émotion exprimée dans l'expression faciale.
- L'ensemble de données contient deux colonnes, "émotion" et "pixels".
- La colonne "emotion" contient un code numérique allant de 0 à 6, inclus, pour l'émotion présente dans l'image.
- La colonne "pixels" contient une chaîne de caractères entre guillemets pour chaque image. Le contenu de cette chaîne est constitué de valeurs de pixels séparées par des espaces et classées par ordre majeur de rangée.

Étape 2 : prétraitement des données :

Le prétraitement des données est l'une des étapes importantes du pipeline d'apprentissage automatique.

Nous avons chargez le fichier de données. Puis nous avons appliquez des techniques de prétraitement d'images telles que le redimensionnement, le remodelage, la conversion en niveaux de gris et la normalisation.

Ensuite, nous avons fait une augmentation des données puisque les réseaux de neuronnes sont avides de données. La meilleure pratique consiste à utiliser un grand ensemble de données, car cela réduit l'overfit et augmente les chances de généralisation.

Nous utilisons donc l'augmentation des données en utilisant ImageDataGenerator() dans keras.

```
[3]: train_data_dir='train'
     validation_data_dir='test'

[19]: train_datagen = ImageDataGenerator(
          rescale=1./255,
          rotation_range=30,
          shear_range=0.3,
          zoom_range=0.3,
          width_shift_range=0.4,
          height_shift_range=0.4,
          horizontal_flip=True,
          fill_mode='nearest')
      validation_datagen = ImageDataGenerator(rescale=1./255)

[20]: train_generator = train_datagen.flow_from_directory(
       train_data_dir,
       color_mode='grayscale',
       target_size=(img_rows,img_cols),
       batch_size=batch_size,
       class_mode='categorical',
       shuffle=True)
```

Figure III.4 : prétraitement des données

Étape 3 : Construire le modèle en utilisant les réseaux neuronaux convolutionnels (CNN) :

Un réseau neuronal convolutif (ConvNet/CNN) est un algorithme d'apprentissage profond qui peut prendre une image d'entrée, attribuer de l'importance (poids et biais apprenables) à divers aspects/objets de l'image, et être capable de les différencier les uns des autres. Le prétraitement requis dans un ConvNet est beaucoup moins important que dans d'autres algorithmes de classification. Alors que dans les méthodes primitives, les filtres sont conçus à la main, avec un entraînement suffisant, les ConvNet peuvent apprendre ces filtres/caractéristiques.

Le CNN est composé principalement de deux types de couches.

1.Les couches cachées / Partie d'extraction de caractéristiques :

- Convolutions

- mise en commun

2. La partie classificateur :

Ici, nous avons utilisé Keras avec TensorFlow comme back-end pour construire des réseaux neuronaux. Pour aller plus loin dans la généralisation, nous avons les couches suivantes à ajouter dans le Keras

1. Couche de convolution
2. Couche de mise en commun
3. Normalisation par lot
4. Couche d'activation
5. Couche d'exclusion
6. Couche d'aplatissement
7. Couche dense

Convolution : une multiplication matricielle avec filtre -> détecteur de caractéristiques.

Normalisation par lot :

Elle est utilisée pour stabiliser, voire accélérer, le processus d'apprentissage en normalisant les entrées des couches.

Elle est réalisée en appliquant une transformation qui maintient l'activation moyenne proche de 0 et l'écart type de l'activation proche de 1.

Voici les points intuitifs et les idées liées à la normalisation par lot

1. normalisation => le processus tend à suivre une courbe en forme de cloche connue sous le nom de distribution normale.
2. rétropropagation => mise à jour couche par couche en arrière de la sortie vers l'entrée en utilisant une estimation de l'erreur qui suppose que les poids dans la couche avant la couche actuelle sont fixes.
3. le gradient indique comment mettre à jour chaque paramètre en supposant que les autres couches ne changent pas.
4. toutes les couches changent pendant une mise à jour → cette procédure de mise à jour conduit à poursuivre éternellement une cible mouvante.

5. normalisation des lots => technique permettant de coordonner la mise à jour de plusieurs couches du modèle => reparamétrisation du réseau.
6. il s'agit de normaliser la moyenne et la variance de chaque unité en dist normale
7. il s'agit de normaliser les entrées des couches pour chaque mini-batch

Activation(ELu) : couche d'activation(couche non linéaire)

1. une convention est d'appliquer après la conv pour introduire de la non-linéarité dans un système qui a couche calculé des opérations linéaires dans le Conv
2. L'unité linéaire exponentielle (ELU) est une fonction d'activation des réseaux de neurones. Contrairement aux ReLU, les ELU ont des valeurs négatives qui leur permettent de rapprocher les activations unitaires moyennes de zéro, comme la normalisation par lots, mais avec une complexité de calcul inférieure. elle est largement utilisée par rapport aux fonctions non linéaires (sigmoïde, tanh) pour son apprentissage rapide sans précision.
3. Sans ces fonctions non linéaires (fonctions d'activation), le réseau serait un grand classificateur linéaire qui pourrait être simplifié en multipliant les matrices de poids (en tenant compte de bais). Il ne ferait rien d'intéressant, comme la classification d'images, etc.

Pooling : implique les types suivants - max, avg ,global max, globalavg

1. Ordre : conv > activation > pooling
2. Pour réduire les dimensions de la carte des caractéristiques en réduisant les paramètres à apprendre et la quantité de calculs.
3. il résume davantage la carte des caractéristiques au lieu de positionner précisément les caractéristiques générées par la couche conv. Cela rend le modèle plus robuste aux variations de la position des caractéristiques dans l'image.
4. lorsque le réseau veut détecter des caractéristiques de plus haut niveau à partir de blocs de construction de bas niveau (détection de coins à partir de bords), nous n'avons pas besoin d'être rigides quant à la position exacte. nous avons besoin d'une invariance de translation au niveau des caractéristiques. nous insérons donc le pooling
5. Cela permet de surmonter le problème de la sensibilité à l'emplacement des caractéristiques.
6. invariance locale de la traduction

Dropout : est une technique de régularisation

1. les neurones sont éliminés de façon aléatoire pendant la formation
2. Cet effet rend le réseau moins sensible aux poids spécifiques des neurones.
3. meilleure généralisation - moins d'overfit

```
model = Sequential()

model.add(Conv2D(32,(3,3),padding='same',kernel_initializer='he_normal',input_shape=(img_rows,img_cols,1)))
model.add(Activation('elu'))
model.add(BatchNormalization())
model.add(Conv2D(32,(3,3),padding='same',kernel_initializer='he_normal',input_shape=(img_rows,img_cols,1)))
model.add(Activation('elu'))
model.add(BatchNormalization())
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.2))

model.add(Conv2D(64,(3,3),padding='same',kernel_initializer='he_normal'))
model.add(Activation('elu'))
model.add(BatchNormalization())
model.add(Conv2D(64,(3,3),padding='same',kernel_initializer='he_normal'))
model.add(Activation('elu'))
model.add(BatchNormalization())
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.2))

model.add(Conv2D(256,(3,3),padding='same',kernel_initializer='he_normal'))
model.add(Activation('elu'))
model.add(BatchNormalization())
model.add(Conv2D(256,(3,3),padding='same',kernel_initializer='he_normal'))
model.add(Activation('elu'))
model.add(BatchNormalization())
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.2))

model.add(Flatten())
model.add(Dense(64,kernel_initializer='he_normal'))
```

Figure III.5 les réseaux neuronaux convolutionnels (CNN)

model.summary() : donne les informations sur l'architecture et la configuration de notre réseau neuronal.

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_6 (Conv2D)           (None, 48, 48, 32)        320

 activation_9 (Activation)   (None, 48, 48, 32)        0

 batch_normalization_8 (Batc  (None, 48, 48, 32)       128
 hNormalization)

 conv2d_7 (Conv2D)           (None, 48, 48, 32)        9248

 activation_10 (Activation)  (None, 48, 48, 32)        0

 batch_normalization_9 (Batc  (None, 48, 48, 32)       128
 hNormalization)

 max_pooling2d_3 (MaxPooling  (None, 24, 24, 32)       0
 2D)

 dropout_5 (Dropout)         (None, 24, 24, 32)        0

 conv2d_8 (Conv2D)           (None, 24, 24, 64)        18496

 activation_11 (Activation)  (None, 24, 24, 64)        0

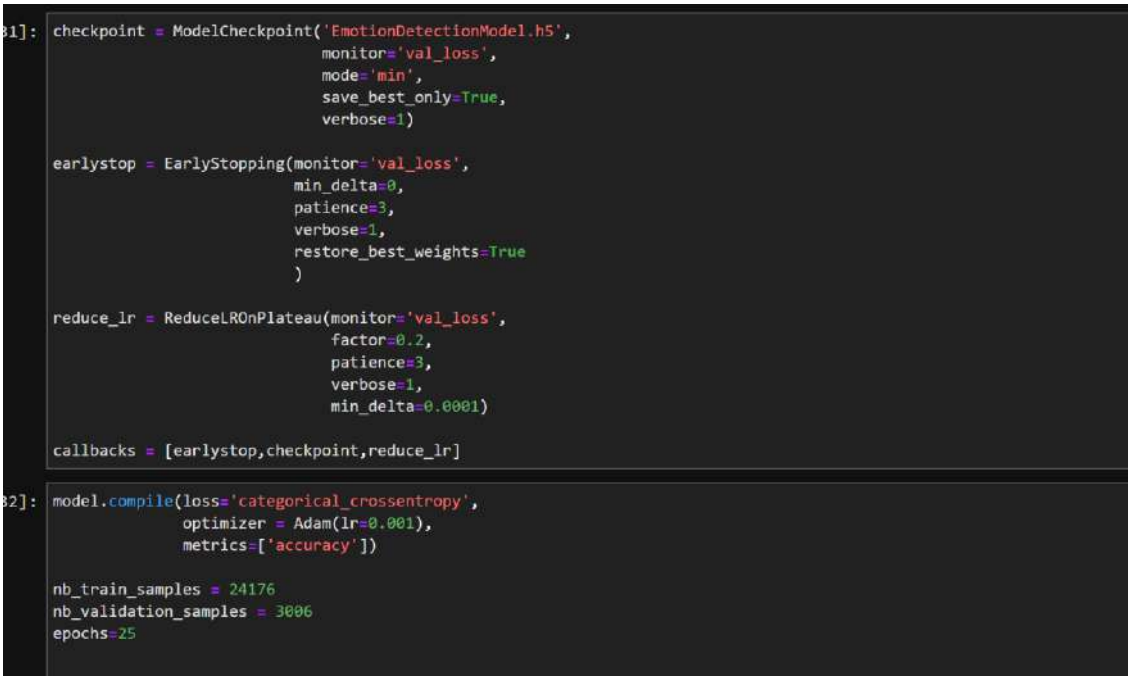
 batch_normalization_10 (Bat  (None, 24, 24, 64)       256
 chNormalization)

 conv2d_9 (Conv2D)           (None, 24, 24, 64)        36928
```

Figure III.6 les réseaux neuronaux convolutionnels (CNN)

Étape 5 : Entraînement du modèle :

Il est maintenant temps ‘entrainer notre modèle. L'entraînement n'est rien d'autre qu'une boucle d'apprentissage. Nous avons défini ci-deeous des hyperparamètres tels que le nombre d'époques, la taille du lot, le taux d'apprentissage, etc.



```
31]: checkpoint = ModelCheckpoint('EmotionDetectionModel.h5',
                                  monitor='val_loss',
                                  mode='min',
                                  save_best_only=True,
                                  verbose=1)

     earlystop = EarlyStopping(monitor='val_loss',
                               min_delta=0,
                               patience=3,
                               verbose=1,
                               restore_best_weights=True
                               )

     reduce_lr = ReduceLROnPlateau(monitor='val_loss',
                                   factor=0.2,
                                   patience=3,
                                   verbose=1,
                                   min_delta=0.0001)

     callbacks = [earlystop,checkpoint,reduce_lr]

32]: model.compile(loss='categorical_crossentropy',
                   optimizer = Adam(lr=0.001),
                   metrics=['accuracy'])

     nb_train_samples = 24176
     nb_validation_samples = 3006
     epochs=25
```

Figure III.7 Entraînement du modèle

EarlyStopping() :

- est utilisé pour arrêter la formation lorsqu'une métrique surveillée a cessé de s'améliorer.

ReduceLROnPlateau()

- est utilisée pour réduire le taux d'apprentissage lorsque la métrique a cessé de s'améliorer.
- Les paramètres sont les suivants

1. moniteur, patience, min_delta
2. factor = 0.1 ==> taux d'apprentissage réduit à 10% (lr*0.1)
3. verbose ==> 1 : calme, 1 : msgs de mise à jour

ModelCheckpoint() :

- pour sauvegarder le modèle keras ou les poids du modèle à une certaine fréquence
- Voici les paramètres utilisés

1. filepath
2. monitor → val_acc ou val_loss

3. save_best_only = True

Étape 6 : Évaluation du modèle :

Nous avons évaluer le taux d'apprentissage de notre modele en regardant sa précision.

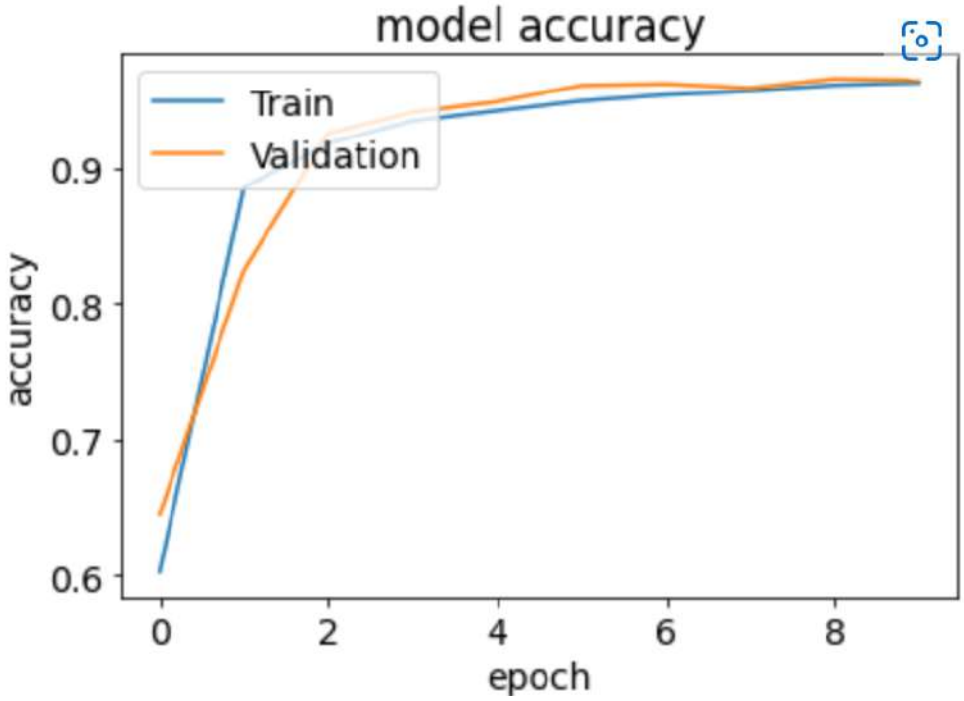


Figure III.8 : Évaluation du modèle

Étape 7 : Enregistrement du modèle :

Puisque nous avons été satisfaits par sa précision, nous avons sauvegardé notre modele. Cela implique à la fois les poids sous un fichier « EmotionDetectionModel2.h5 »pour l'appelé dans l'étape de test.

Nous n'avons pas besoin de sauvegarder l'architecture mais à chaque fois que nous chargeons le modèle, nous devons à nouveau définir l'architecture du modèle avant de l'utiliser

Étape 8 : Test du modèle :

Dans cette étape nous avons fait des prédictions. Testons notre modèle avec quelques images.

Avant de tester, nous avons du charger notre modèle. Et nous devons charger également le classificateur que nous avons utilisé pour détecter le visage d'une personne devant la caméra haarcascade_frontalface_default.

```
face_classifier=cv2.CascadeClassifier('haarcascade_frontalface_default.xml')
classifier = load_model('EmotionDetectionModel2.h5')
```

Figure III.9 Test du modèle

Ensuite nous avons défini une variable class_labels pour stocker le nom des classes ou les types d'émotions que nous allons prédire ainsi qu'une variable cap pour stocker la valeur renvoyée par la méthode cv2.VideoCapture. Ici, la valeur 0 dans VideoCapture est utilisée pour indiquer à la méthode d'utiliser la webcam principale d'un ordinateur portable.

```
class_labels=['Angry','Happy','Neutral','Sad','Surprise']
cap=cv2.VideoCapture(0)
```

Figure III.10 Test du modèle

Nous avons créé des boîtes autour des visages détectés par le classificateur dans le flux de la caméra.

Nous avons ensuite ajouté quelques éléments relatifs à OpenCV ici pour prétraiter les images.

```
while True:
    ret,frame=cap.read()
    labels=[]
    gray=cv2.cvtColor(frame,cv2.COLOR_BGR2GRAY)
    faces=face_classifier.detectMultiScale(gray,1.3,5)
```

Figure III.11 Test du modèle

- Cette fonction s'exécute indéfiniment jusqu'à ce que quelqu'un appuie sur la touche "q" du clavier.
- La première ligne après lit une image à partir du flux de caméra en cours d'exécution et la stocke dans la variable frame.

- La fonction detecMultiScale utilise le classificateur en cascade haar et détecte les visages dans l'image et renvoie une liste contenant les coordonnées du rectangle autour du visage et la stocke dans les visages variables.

Une image à partir de flux vidéo peut être composée de nombreux visages de nombreuses personnes, alors nous avons ajouté quelques lignes de code pour parcourir les éléments de la liste pour obtenir les différentes valeurs de rectangle pour différents visages.

```
for (x,y,w,h) in faces:
    cv2.rectangle(frame,(x,y),(x+w,y+h),(255,0,0),2)
    rgray=gray[y:y+h,x:x+w]
    rgray=cv2.resize(rgray,(48,48),interpolation=cv2.INTER_AREA)
```

Figure III.12 Test du modèle

La variable rgray stocke uniquement le visage de l'image stockée qui a le rectangle autour de lui dans l'itération actuelle, sous la forme d'un tableau de valeurs stockées dans la variable gray , puis nous avons du redimensionner le tableau à une taille (48,48) avec une interpolation INTER_AREA. Le redimensionnement d'une image a besoin d'un moyen pour calculer les valeurs de pixel pour la nouvelle image à partir de l'originale. Les cinq méthodes d'interpolation fournies avec OpenCV sont INTER_NEAREST, INTER_LINEAR, INTER_AREA, INTER_CUBIC et INTER_LANCZOS4. Nous avons utilisé la méthode INTER_AREA qui peut calculer les pixels en utilisant la relation pixel-aire.

Ensuite, le tableau rgray est converti en flottant et les valeurs de pixel qu'il contient sont divisées par 255,0 pour obtenir des valeurs comprises entre 0 et 1 afin qu'il soit plus facile pour le modèle de prédire les résultats et qu'il soit stocké dans la variable R. La valeur dans R est convertie en tableau pour utiliser la prédiction à l'aide du modèle.

```
    R=rgray.astype('float')/255.0
    R=img_to_array(R)
    R=np.expand_dims(R,axis=0)

    preds=classifier.predict(R)[0]
    label=class_labels[preds.argmax()]
    label_position=(x,y)
    cv2.putText(frame,label,label_position,cv2.FONT_HERSHEY_SIMPLEX,2,(0,255,0),3)
```

Figure III.13 Test du modèle

Finalement, nous avons utilisé la méthode cv2.putText pour afficher l'émotion détectée à l'écran à côté du cadre et la méthode cv2.imshow affiche le résultat.

Puis nous avons ajouter une commande pour fermer la fenetre et liberer le cadre de la caméra

```
cap.release()
cv2.destroyAllWindows()

1/1 [==============================] - 1s 575ms/step
```

Figure III.14 Test du modèle

Quelques résultats de notre système sont montré ci-dessous .

- Pour l'etat neutre :

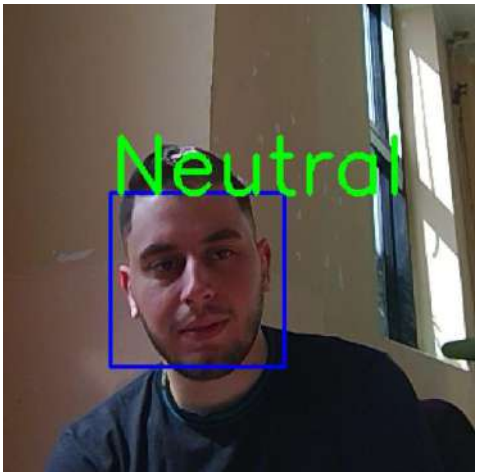


Figure III.15 présente l'état neutre

- Pour l'etat surpris :

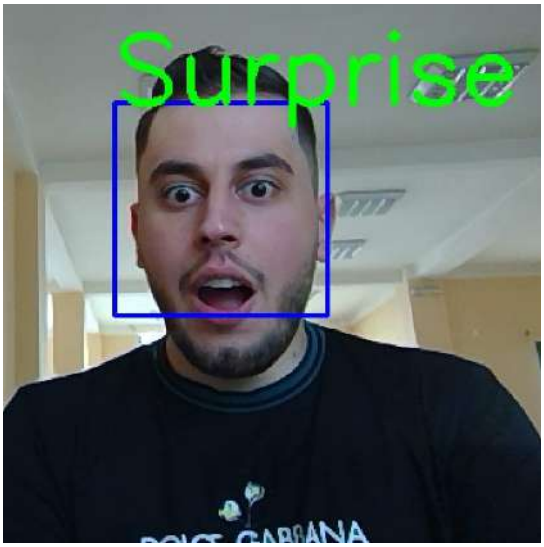


Figure III.16 présente l'état surprise

- Pour l'etat triste :

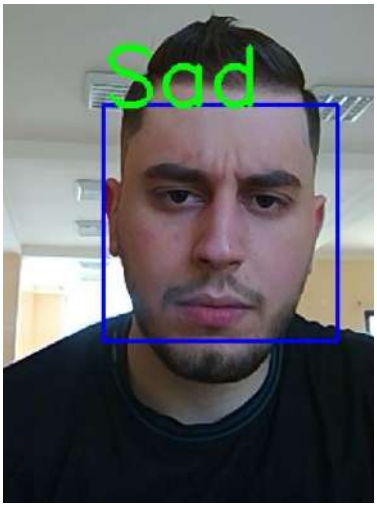


Figure III.17 présente l'état triste

- Pour l'etat heureux :



Figure III.18 présente l'état heureux

Conclusion :

Dans ce chapitre, nous avons présenté la conception de notre système de reconnaissance des émotions de base. En utilisant le modèle séquentielle, afin de révéler les performances d'un système FER par apprentissage profond. Ensuite on a projeté la lumière sur l'environnement de travail, côté matériel côté et logiciel pour réaliser notre système. Nous avons vu les résultats obtenus et nous avons présenté les performances de notre système.

CONCLUSION

Conclusion

La capacité de l'ordinateur à reconnaitre les émotions de la personne constitue un nouveau défi pour la recherche scientifique moderne. D'autant plus que la communication entre humaine et les appareils électroniques a augmenté considérablement, les chercheurs cherchent à développer des programmes intelligents capables de comprendre les expressions faciales en peu de temps.

Ces dernières années, le Deep Learning et plus particulièrement les réseaux de neurones convolutionnels (CNN) ont apparu spécialement pour résoudre les problèmes rencontrés de la machine Learning. Le CNN est l'une des structures réseaux les plus représentatives de la technologie d'apprentissage en profondeur qui a connu un grand succès dans le domaine du traitement et de la reconnaissance d'images.

Nous avons présenté à travers ce mémoire une étude théorique qui concerne les émotions, les expressions faciales et l'apprentissage approfondi en se basant sur les réseaux de neurone convolutifs (CNN).

Dans ce travail, nous avons développé un système de reconnaissance des expressions faciale basé principalement sur le modèle Séquentielle. Le système a été entrainé sur la base de données Fer2013.

Notre travail, comme tous les travaux de recherche présente quelques perceptives, qui sont considéré comme des futurs travaux :

- Tester notre modèle sur d'autres bases des données plus volumineuses.
- Ajouter des données supplémentaires d'entrainement à la base de données fer2013.
- Développer notre modèle en utilisant des architecture plus développées en essayant de les optimiser.
- Inclure plus de classes des émotions secondaires pour reconnaitre les micros expressions.

BIBLIOGRAPHIE

BIBLIOGRAPHIE :

[1] Belhouchette, K. (2018). MODELISATION DE L'ETAT AFFECTIF DANS LES SEQUENCES VIDEO. Algérie/Oum El Bouaghi.

[2] GHANEM, K. (2009-2010). Reconnaissance des Expressions Faciales à Base d'Informations Vidéo,Estimation de l'Intensité des Expressions faciale. Algérie/Constantine.

[3] Hamdi, H. (2012). Plate-forme multimodale pour la reconnaissance d'émotions via l'analyse de signaux physiologiques. French.

[4] Nadia, A. I. (2019-2020). Une approche IA pour la reconnaissance des expressions faciales. Algérie/Bouira.

[5] LILIA, D. R. (2020). La Détection de la colère chez le conducteur en utilisant le Deep Learning. Algérie/BISKRA.

[6] Brahim, B. M. (2016/2017). L`apprentissage profond (Deep Learning) pour la classification et la recherche d'images par le contenu. Algérie/OUARGLA.

[7] Foued, N. (2019). Reconnaissance d'expression faciale à partir d'un visage réel. Algérie/Guelma.

[8] Crenn, A. (2019). Reconnaissance d'expressions corporelles dans des mouvements de personnes en vue de la synthèse de style. Lyon.

[9] BELGASMI, O. H. (2016-2017). RECONNAISSANCE AUTOMATIQUE DES EXPRESSIONS FACIALES PAR SUPPORT VECTOR MACHINE. Algérie/Oum El Bouaghi.

[10] Zakaria, M. M. (03/07/2017). Classification des images avec les réseaux de neurones convolutionnels. Algérie/Tlemcen.

[11] Mehendale, N. (2019, july 16). Facial emotion recognition using convolutional neural networks (FERC. springer nature jurnal, 8.

[12] R, S. M. (July-2020). Human Emotion Recognition System using Deep Learning Techniques. International Journal of Engineering Research & Technology (IJERT), 5.

[13] Wenbin Zhou, J. C. (AG 2020). Deep Learning-Based Emotion Recognition from Real-Time Videos. USA.

[14] Neha S1, N. P. (Aug 2020). Emotion Recognition and Depression Detection using Deep Learning. Journal of Engineering and Technology (IRJET), 6

[15] Spiers, D. L. (Juni 2016). Facial Emotion Detection Using Deep Learning.

[16]. M. Kalpana Chowdary, T. N. (07 January 2021). Deep learning-based facial emotion recognition for human–computer interaction applications. Récupéré sur https://doi.org/10.1007/s00521-021-06012-8

[17] Khalf, F. (29 Mar 2018). Reconnaissance automatique des émotions par données multimodales : expressions faciales et des signaux physiologiques.

[18] Xin-yu Pan, Y.-z. Z.-j. (2015, April). Using Kinect for real-time emotion recognition via facial expressions .

[19] Patterson, M. S. (February 2008). EMOTION RECOGNITION USING FACIAL EXPRESSIONS WITH ACTIVE APPEARANCE MODELS . USA.

[20] Mane1, M. S. (May-June 2017). Intelligent Facial Emotion Recognition using modified-PSO. Kolhapur, India.

[21] udemy. (s.d.). Récupéré sur La Formation Complète PyCharm: https://www.udemy.com/course/la-formation-complete-pycharm/

[22] jetbrains. (s.d.). Récupéré sur L'IDE Python pour développeurs professionnels: https://www.jetbrains.com/fr-fr/pycharm/

[23] Wikipédia.(s.d.). Récupérésur opencv:

https://fr.wikipedia.org/wiki/OpenCV#Traitement_d'images

[24] TensorFlow : tout savoir sur la bibliothèque Machine Learning open source. (s.d.). Récupéré sur LEBIGDATA: https://www.lebigdata.fr/tensorflow-definition-tout-savoir

[25] PHOTOSHOP. (s.d.). Récupéré sur Fonctionnalités de Python: https://fr.photo-

555.com/7018858-python-features#menu-1

[26] (s.d.). Récupéré sur keras: https://keras.io/

[27] Programmation d'interface graphique Python avec Tkinter. (s.d.). Récupéré sur RealPython: https://realpython.com/python-gui-tkinter/

[28] MobileNet et MobileNetV2. (s.d.). Récupéré sur keras: https://keras.io/api/applications/mobilenet/

[29] transférer l'apprentissage à l'aide de Mobilenet et Keras. (s.d.). Récupéré sur Towards Data Science: https://towardsdatascience.com/transfer-learning-using-mobilenet-and-keras

[30] Expression faciale : le métalangage du visage qui trahit nos émotions. (s.d.). Récupéré sur Noémie Grynberg: http://noemiegrynberg.e-monsite.com/

[31] Mark Sandler and Andrew Howard, G. R. (2018, april Tuesday 3). MobileNetV2: The Next Generation of On-Device Computer Vision Networks. Récupéré sur Google Blig: https://ai.googleblog.com/2018/04/mobilenetv2-next-generation-of-on.html

[32] kaggle. (s.d.). Récupéré sur : https://www.kaggle.com/