



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider – BISKRA
Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie
Département d'informatique

N° d'ordre : IVA /M2/2020

Mémoire

Présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : IVA

La Détection de la colère chez le conducteur en utilisant le Deep Learning

Encadré par :
Mme HATTAB DALILA

Présenté par :
DAHAN RAOUNAK LILIA

Septembre 2020

Résumé

La conduite en état de colère est une cause majeure des accidents de la route et expose les conducteurs à des risques routière plus élevé à ceux qui sont à l'état de conduite calme. Par conséquent, l'utilisation des systèmes de reconnaissances des expressions faciales qui aide à mettre les conducteurs attentifs et fait l'alerte dans le cas de détection de colarité peut jouer un rôle très important dans la sécurité des conducteurs et des passagers.

Ce mémoire présente une méthode de détection de colère chez les conducteurs en utilisant une architecture des réseaux de neurones convolutifs approfondis nommé « mini-xception » avec une base de données Fer-2013 pour fournir un classificateur entraîné peut facilement prédire les sept expressions faciales universelles (en colère, Heureux, tristesse, peur, surprise, dégoût, neutre) et estimer la probabilité de ces expressions faciales, notre système déclenche une alerte qui demande le changement du conducteur dans le cas d'estimation de probabilité supérieure ou égale à cinquante pour cent .

Les résultats des tests montrent que le système proposé peut détecter la colère chez un conducteur d'une manière efficace dans peu de temps.

Mots clés : Reconnaissance des expressions Faciales, les expressions faciales universelle, Détection de colère chez le conducteur, réseaux de neurone convolutifs approfondi, classificateur entraîné.

Abstract

Driving in a state of anger is a major cause of traffic accidents and exposes drivers to greater road risks than those in a state of calm driving. Therefore, the use of facial expression recognition systems that help alert drivers and alert when colarity is detected can play a very important role in the safety of drivers and passengers.

This dissertation presents a method of detecting anger in drivers using an architecture of deep convolutional neural networks named 'mini-xception' with fer-2013 database to provide a trained classifier can easily predict the seven universal facial expressions (angry, happy, sadness, fear, surprise, disgust, neutral) and estimate the probability of these facial expressions, our system triggers an alert that requests the change of the driver in the case of a probability estimate greater than or equal to fifty percent .

Test results show that the proposed system can effectively detect anger in a driver in a short time.

Keywords: Recognition of facial expressions, universal facial expressions, Driver anger detection, deep convolutional neural networks, trained classifier.



Remerciements :

Nous tenons à remercier en premier lieu notre grand dieu qui nous a donné la force et le courage pour continuer et nous avoir aidées et éclairer le chemin pour réaliser de ce modeste travail

*J'exprime toute ma gratitude et ma reconnaissance à **Dr HATTAB***

***DALILA** pour avoir bien voulu diriger ce projet, pour m'avoir encouragé et soutenue tout au long de ce travail, et pour sollicitude dont il m'entouré*

Mes remerciements sont aussi adressés à tous les enseignants du Départements d'informatique qui ont contribué à ma formation

J'aimerais également souligner le support de ma sœur, mes frères, mes amies pour leurs encouragements



Dédicace

À MES CHERS PARENTS

Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que vous avez consenti pour mon instruction et mon bien être.

Je vous remercie pour tout le soutien et l'amour que vous me portez depuis mon enfance et j'espère que votre bénédiction m'accompagne toujours.

Que ce modeste travail soit l'exaucement de vos vœux tant formulés, le fruit de vos innombrables sacrifices, bien que je ne vous en acquitterai jamais assez.

Puisse Dieu, le Très Haut, vous accorder santé, bonheur et longue vie et faire en sorte que jamais je ne vous déçoive.

Table des matières

Introduction Générale	1
1 Un tour d’horizon sur l’Expression Faciale	3
1.1 Introduction	3
1.2 Emotion vs Expression Faciale	4
1.2.1 Définitions	4
1.2.2 Les Expression Faciales Emotionnelles EFE	4
1.2.3 Micro et Macro Expression Faciale	4
1.3 L’étude analytique des Expressions Faciales	5
1.3.1 La psychologie des Expressions Faciales	5
1.3.2 Les Expressions Faciales Primaires	5
1.3.3 L’état émotionnelle des six expressions faciales universelles	6
1.3.4 l’étude anatomique Des Expressions Faciales	7
1.4 Différente Représentation des expressions faciales émotionnelles	8
1.4.1 FACS	8
1.4.2 MPEG4	10
1.4.3 CANDIDE	11
1.5 Modules du système de reconnaissance des expressions faciales	13
1.5.1 Détection du visage	14
1.5.1.1 Les techniques de détection du visage	15
1.5.2 Extraction des caractéristiques	15
1.5.2.1 Les caractéristiques géométriques	16
1.5.2.2 Les caractéristiques d’apparence	16
1.5.3 La Reconnaissance des Expressions Faciales	16
1.5.3.1 Les Classes des techniques de REF	16
1.5.3.1.a Les méthodes Globales ou holistique	16
1.5.3.1.b Les Méthodes Locales	17
1.5.3.1.c Les Méthodes Hybrides	17
1.5.3.2 Les Types de Reconnaissance EF	17
1.5.3.2.a La reconnaissance basée sur les images fixe (statique)	18
1.5.3.2.b La reconnaissance basée sur les séquences vidéo (dy-	
namique)	18

1.5.3.3	Les Bases de Données utilisées dans expressions faciales : .	18
1.5.3.3.a	JAFFE	18
1.5.3.3.b	Cohn Kanade	19
1.5.3.3.c	FER2013	20
1.6	Performance et L'état du conducteur dans le cas de « colère »	21
1.6.1	Performance du conducteur	21
1.6.2	L'état Du conducteur « la colère »	21
1.7	Conclusion	22
2	Réseaux de neurones et le Deep Learning	24
2.1	Introduction	24
2.2	L'apprentissage en profondeur (Deep Learning)	25
2.2.1	Définition	25
2.2.2	Histoire de Deep Learning	26
2.2.3	Concepts de Deep Learning	26
2.2.4	Machine Learning vs Deep Learning : où le Deep Learning Fait la difference?	28
2.2.5	Les avantages de Deep Learning	28
2.3	Les Réseaux de neurones Convolutifs (CNN)	28
2.3.1	Les couches constructives de CNN (architecture de CNN)	29
2.3.2	La couche de convolution (CONV)	30
2.3.2.1	Les différentes convolutions	31
2.3.3	La Couche RELU	32
2.3.4	La Couche pooling (POOL)	33
2.3.4.1	Le Flattening (La mise à plat)	33
2.3.5	la Couche entièrement connectée (FC)	34
2.3.6	La Couche de perte (LOSS)	35
2.4	Quelques réseaux convolutifs célèbres	35
2.5	Pourquoi Le CNN?	36
2.6	Conclusion	36
3	Conception d'un système de détection de la colère chez le conducteur en utilisant le Deep Learning	37
3.1	Introduction	37
3.2	Objectifs du système	38
3.3	Conception Générale de Système	38
3.4	Conception détaillée du système	38
3.4.1	La phase d'apprentissage	38
3.4.1.1	la base de données	39
3.4.1.2	L'architecture de modèle « mini-xception »	39
3.4.1.2.a	Motivation et objectif de choix	42

3.4.2	La phase de prédictions (Test)	42
3.4.2.1	la détection du visage	42
3.4.2.2	Le prétraitement	43
3.4.2.3	la classification et reconnaissance des expressions faciales	44
3.5	Conclusion	44
4	Implémentation et résultats Expérimentaux	45
4.1	Introduction	45
4.2	Présentations d'outils	46
4.2.1	Le hardware	46
4.2.2	Le software	46
4.2.2.1	Le langage utilisé	46
4.2.2.2	Les Bibliothèques Utilisés	47
4.3	Implémentation et résultats de notre Système	49
4.3.1	implémentation de la phase d'apprentissage	50
4.3.1.1	Chargement et prétraitement de la base de données	50
4.3.1.1.a	La fonction chargement de donnée (def load_data)	50
4.3.1.1.b	La procédure de prétraitement def preprocess_input	50
4.3.1.2	L'architecture de « mini-xception »	51
4.3.1.3	La fonction d'entraînement	52
4.3.2	Résultat de l'entraînement	53
4.3.3	La phase de prédiction	54
4.4	la performance de notre système	58
4.5	Conclusion	59
	Conclusion générale et perspectives	60

Liste des figures

1.1	Emotions primaires exprimées sur le visage. De gauche à droite : dégoût, peur, joie, surprise, tristesse, colère. [09].	6
1.2	Liste des Action Units relatives aux 6 expressions faciales [15].	9
1.3	Modèle du visage par le standard MPEG-4. (a) Représentation des points caractéristiques FAs. (b) Modèle des paramètres unitaires FAPUs.[13] . .	11
1.4	En haut : CANDIDE-1 avec 79 sommets et 108 surfaces. En bas : CANDIDE-2 avec 160 sommets et 238 surfaces. [17]	12
1.5	CANDIDE-3 avec 113 sommets et 168 surfaces.[17]	13
1.6	Les Modules du système de reconnaissance des expressions faciales	14
1.7	Détection des visages	14
1.8	Les techniques utilisées dans la Reconnaissance des EF.	17
1.9	Exemple de La base JAFFE [13]	19
1.10	Exemple de la Base CK+ [15]	20
1.11	Exemple de la base Fer2013	21
1.12	différent état de colère de gauche à droite (a) : Les yeux sont ouverts, et la bouche également (b) : les sourcils sont rapprochés vers le bas t la bouche est ouverts (c) : la regard direct et agressif, les yeux sont à peine ouverts, les lèvres sont pincées	22
2.1	La relation Entre L'IA et ML et le Deep Learning	25
2.2	Exemple de fonctionnement de Deep Learning [Web 08]	27
2.3	Les réseaux de neurones convolutifs [44]	29
2.4	Simple Exemple d'une Architecture Globale de CNN [45]	30
2.5	exemple d'une opération convolution	31
2.6	Différent type de convolution	32
2.7	application de RELU à une matrice et à une image [49]	32
2.8	Exemple de la phase pooling en utilisant le Max pooling	33
2.9	Exemple de Mise à plat des images finale en sortie (des filtres +simplifications) [49]	34
2.10	un exemple d'opération entièrement connectée (FC)	35
3.1	schéma globale du système	38
3.2	processus de la phase d'apprentissage	39

3.3	l'architecture mini-xception [58]	41
3.4	le processus de la phase de test	42
3.5	le processus de prétraitement	43
3.6	Processus de la classification	44
4.1	le logo python	47
4.2	le logo de tensorflow [web20]	48
4.3	la code source qui charge la base de données « fer-2013 »	50
4.4	la procédure de prétraitement d'entrée de la base de données « fer-2013 »	51
4.5	un aperçu de la code source de l'implémentation de mini-xception	52
4.6	la fonction fit () de l'entraînement de classifieur « mini-xception »	53
4.7	un aperçu sur le résultat général de la phase apprentissage	54
4.8	Traces d'entraînement et précision de la validation	58
4.9	traces d'entraînement de perte	59

Liste des tableaux

1.1	Liste des émotions basiques selon différents théoriciens [03].	5
1.2	l'état émotionnel de six émotions universelles d'Ekman	7
1.3	L'étude anatomique Des Expressions faciales fondamentale du visage [12] .	8
1.4	Les prototypes des 6 émotions de bases avec le Système FACS [16]	10
1.5	les méthodes de Base utilisé dans la détection du Visage	15
1.6	Etiquette d'émotion dans l'ensemble de Donnée Fer 2013[web05]	20
2.1	Les étapes majeures du Deep Learning [42]	26
2.2	la Différence Faite par le Deep Learning[web 10]	28
4.1	L'interface graphique de phase de prédiction	58

Introduction Générale

L'expression faciale est l'un des moyens non verbaux les plus couramment utilisés par les humains pour transmettre les états émotionnels internes et, par conséquent, joue un rôle fondamental dans les interactions interpersonnelles. Bien qu'il existe un large éventail d'expressions faciales possibles, les psychologues ont identifié six expressions fondamentales (la joie, la tristesse, la surprise, la colère la peur et le dégoût) universellement reconnues.

Il est évident qu'un système capable de réaliser une reconnaissance automatique des émotions humaines est une tâche souhaitable pour un ensemble d'applications telles que : L'interaction homme machine, La sécurité, la médecine, l'éducation, etc.

Plusieurs travaux ont été développés dans le domaine de machine Learning, sur la reconnaissance des expressions faciales en utilisant plusieurs méthodes d'extraction de caractéristiques et d'apprentissage automatique, La plupart de ces méthodes ont déployé des fonctionnalités manuelles suivies d'un classifieur, tel que la méthode SVM, Adaboost et la méthode Forêts aléatoires.

Ces travaux ont connu des difficultés à cause de :

- La puissance et le taux de reconnaissance sont faibles dans les grandes bases de données.
- Les résultats de ces méthodes traditionnelles restent limités dans le cadre de traitement de très grande masse de données.
- Les caractéristiques sont définies par des méthodes traditionnelles

Le succès actuel qui résout ces problèmes est dans le Deep Learning ou plus particulièrement les Réseaux de neurone convolutifs CNN parce que :

- CNN apprend à les extraire directement de la base de données d'apprentissage.
- Le réseau CNN est généralement associé à un classifieur qui permet de former le modèle de bout en bout sur l'ensemble de données.
- La disponibilité des bases de données et la puissance de calcul des machines actuelles présentent des valeurs à ajouter pour les CNN.

L'objectif principal de notre projet de fin d'étude est de proposer une architecture permettant de reconnaître certains des expressions faciales universelles et nous nous intéressons plus particulièrement à la « colère » dans le but de minimiser les accidents de la route à cause de cette expression et pour la sécurité de passagers par le lancement d'une alerte qui demande le changement de conducteur par un autre plus calme et plus sage.

Parmi les contributions de cette mémoire, on cite :

- Proposition d'une architecture basée sur l'apprentissage approfondie utilisant le CNN pour la détection des expressions faciales.
- Un système dynamique simple et efficace estimées par la classificateur Mini-xception en utilisant la base de données fer-2013.

Ce mémoire est constitué de quatre chapitres. Les deux premiers sont consacrés aux études théoriques. Le troisième chapitre on présente la conception de notre système et le quatrième est consacré à la description des différents outils utilisés dans le développement de notre application. Une brève description de ces chapitres est faite dans les paragraphes suivants :

Dans le premier chapitre on présente des notions liées aux expressions faciales.

Dans le deuxième chapitre nous le consacrons à la présentation de l'apprentissage profond, où nous détaillons plus les réseaux de neurone convolutifs.

Dans le troisième chapitre on présente la conception de notre système qui consiste à détecter la colère chez le conducteur en utilisant le Deep Learning à base de CNN.

Dans le quatrième chapitre est consacré à la description des différents outils utilisés dans le développement de notre application, ainsi que les différents résultats obtenus. Suivi d'une évaluation de certains paramètres affectant les résultats obtenus. Ce mémoire est clôturé par une conclusion générale et perspectives.

Chapitre 1

Un tour d’horizon sur l’Expression Faciale

Sommaire

1.1	Introduction	3
1.2	Emotion vs Expression Faciale	4
1.3	L’étude analytique des Expressions Faciales	5
1.4	Différente Représentation des expressions faciales émotionnelles	8
1.5	Modules du système de reconnaissance des expressions faciales	13
1.6	Performance et L’état du conducteur dans le cas de « colère »	21
1.7	Conclusion	22

1.1 Introduction

« Ses mots disent une chose, son corps en dit une autre : c’est un menteur. La vérité est écrite sur nos visages. » Popularisée par la série « **Lie To Me** » dans les années 2000, l’utilisation des expressions du visage comme moyen de décoder tout haut ce que nous cachons tout bas Ce n’est que depuis une vingtaine d’années, que l’émotion est explicitement considérée par les linguistes comme un objet de recherche incontournable alors que les philosophes, les anthropologues et, plus massivement, les psychologues ont divisé les expressions faciales en six grandes catégories : **la colère, le dégoût, la peur, la joie, la tristesse et la surprise** ; qui sont aussi appelés émotions primaires Dans ce premier chapitre, nous présentons quelques notions concernant les expressions faciales telles que leurs définitions, leurs différentes théories de codification, ainsi que l’architecture globale des systèmes de reconnaissance faciale. En se basant sur ces dernières,

nous présentons les approches qui aident à la reconnaissance d'expressions faciale, et nous terminons le chapitre par étudie la performance et l'état du conducteur de le cas de colère

1.2 Emotion vs Expression Faciale

1.2.1 Définitions

a. Emotion :

L'émotion est une expérience psychophysiologique complexe et intense (avec un début brutal et une durée relativement brève) de l'état d'esprit d'un individu animal liée à un objet repérable lorsqu'il réagit aux influences biochimiques (internes) et environnementales (externes). Chez les humains, l'émotion inclut fondamentalement « un comportement physiologique, des comportements expressifs et une conscience ». L'émotion est associée à l'humeur, au tempérament, à la personnalité et à la disposition et à la motivation.[web01]

b. Expression Faciale :

Une expression faciale est un ou plusieurs mouvements ou positions des muscles sous la peau du visage. Ces mouvements transmettent l'état émotionnel d'un individu aux observateurs Les expressions faciales sont une forme de communication non verbale. Ils sont un moyen principal de transmettre des informations sociales entre les humains. [web02]

1.2.2 Les Expression Faciales Emotionnelles EFE

Les expressions émotionnelles faciales constituent l'une des plus importantes manières de communiquer les émotions, et ce, via une transmission rapide de l'information. Elles permettent à l'observateur d'inférer les états émotionnels ainsi que les intentions d'autrui et d'anticiper ses gestes, mais aussi de réguler ses propres comportements en conséquence. Ainsi, l'aptitude à reconnaître les émotions influence les capacités d'adaptation à l'environnement et de ce fait, constitue une habileté indispensable au fonctionnement interpersonnel de l'individu [01]

1.2.3 Micro et Macro Expression Faciale

a. Micro Expression Faciale : En général, les micro-expressions sont décrites comme un modèle involontaire du corps humain qui est suffisamment significatif pour être observable, mais qui est trop bref pour transmettre une émotion. Les micro-expressions apparaissant sur le visage sont rapides et sont souvent manquées lors d'une observation

occasionnelle. D'une durée comprise entre 1/25 et 1/3 de seconde (environ 2 à 10 images). Les micro-expressions peuvent être classées en fonction de l'expression est modifiée, en trois types [02] :

- **Expressions simulées** : lorsqu'une micro-expression n'est pas accompagnée par une expression authentique.
- **Expressions neutralisées** : lorsqu'une expression authentique est supprimée et le visage reste neutre.
- **Expressions masquées** : lorsqu'une expression authentique est complètement masquée par une expression falsifiée.

b. Macro Expression Faciale : Les macro expressions durent généralement de 3/4 de seconde à 2 secondes (environ 24 à 60 cadres). Les macros expressions peuvent apparaître sur plusieurs ou une seule région du visage, selon les expressions. Par exemple, les expressions de surprise provoquent généralement des mouvements autour des yeux, du front, des joues et de la bouche, alors que l'expression de la peur génère généralement des mouvements uniquement près des yeux [02]

1.3 L'étude analytique des Expressions Faciales

1.3.1 La psychologie des Expressions Faciales

Le tableau suivant présente une liste des expressions faciales selon les différents théoriciens psychologiques.

Auteurs	Emotions
Izard (1977)	Joie, surprise, colère, peur, tristesse, mépris, détresse, intérêt, culpabilité, honte, amour
Plutchik (1980)	Acceptation, colère, anticipation, dégoût, joie, peur, tristesse, surprise
Kemper (1981)	Peur, colère, dépression, satisfaction
Ekman (1992)	Colère, peur, tristesse, joie, dégoût, surprise

TABLE 1.1 – Liste des émotions basiques selon différentes théoriciens [03].

1.3.2 Les Expressions Faciales Primaires

Les recherches actuelles en sciences cognitives montrent que les émotions jouent un rôle crucial dans l'acquisition et la maîtrise de compétences aussi bien sociales [04] que

cognitives et académiques [05]. Déjà en 1872, Charles Darwin expliquait que la capacité à reconnaître et à exprimer des expressions émotionnelles représentait un avantage adaptatif [06]. En effet, cette capacité permet au récepteur d'être informé rapidement sur son environnement, sur l'état de sa relation avec autrui, sur les besoins de son interlocuteur et d'inférer la potentielle réaction de l'émetteur et, ainsi, de sélectionner la tendance à l'action la plus pertinente [07]. Les émotions primaires (ou dites «de base ») peuvent être observées chez la plupart des vertébrés. Leur nombre varie selon les études, cependant on en retient généralement six : **la joie, la surprise, la peur, la colère, le dégoût et la tristesse** (une septième émotion, naturel, est parfois ajoutée). Selon Paul Ekman, ces six émotions sont primaires de par leurs propriétés universelles [08]. C'est à dire qu'une expression faciale universelle est une émotion connue de la même façon aux différentes cultures. Chaque émotion de six émotions primaires est associée à une expression faciale particulière exprimées sur le visage. (Voir figure 1.1)

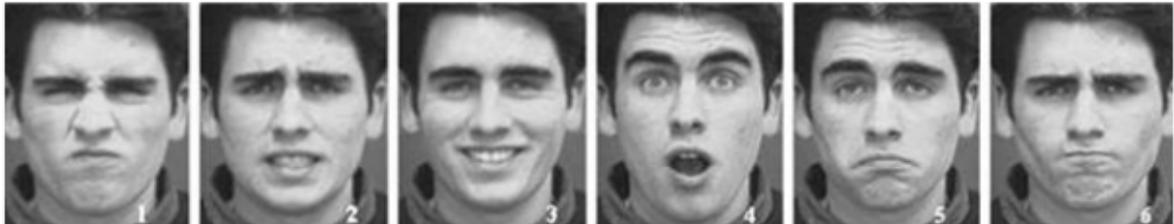


FIGURE 1.1 – Emotions primaires exprimées sur le visage. De gauche à droite : dégoût, peur, joie, surprise, tristesse, colère. [09].

1.3.3 L'état émotionnelle des six expressions faciales universelles

Les six émotions d'Ekman sont devenues universelles car elles sont les émotions les plus fréquemment étudiées et utilisées. Dans ce qui suit un tableau descriptif qui exprime l'état émotionnel de ces émotions

Emotions	Etat émotionnelle
Joie	C'est l'état d'une personne dans une condition de satisfaction intense [10]. Elle est due par rapport au désir, à la réussite, au bien-être, et l'accomplissement [11], mais aussi l'approche.
Tristesse	C'est l'état d'une personne qui souffre moralement suite à une insatisfaction et des soucis [10]. Elle est souvent due soit à une perte, ou un deuil, ou un obstacle [11], la personne se replie souvent sur soi.
Colère	C'est l'état d'une personne dans une réaction violente et agressive lors d'une contrariété [10]. Elle est souvent due soit à une injustice, ou un dommage, atteinte au système de valeurs [11]. La plupart des personnes en ce moment attaquent.
Dégout	C'est l'état d'une personne qui a une répugnance pour certains aliments ou à un manque d'appétit [10]. Elle est souvent due soit à un rejet, ou contre quelqu'un, ou à une aversion [11]. La plupart des personnes en ce moment préfèrent se retirer.
Peur	C'est l'état d'une personne menacée par un danger réel ou imaginaire [11]. Elle est souvent due soit à une menace, ou à un danger, ou à des inconnus [11]. La plupart des personnes en ce moment préfèrent prendre la fuite
Surprise	C'est l'état d'une personne étonnée par quelque chose d'inattendu [10]. Elle est souvent due soit à un danger immédiat, ou à un imprévu, ou à des inconnus [10]. La plupart des personnes en ce moment préfèrent prendre la fuite ou elles sursautent.

TABLE 1.2 – l'état émotionnel de six émotions universelles d'Ekman

1.3.4 l'étude anatomique Des Expressions Faciales

L'étude des expressions faciales ne peut se faire sans l'étude anatomique de visage car la production de ces derniers est la conséquence de déformation au niveau de traits permanents du visage. Le tableau suivant représente les changements des traits faciaux ou différentes expressions faciales

Expressions	Distance entre paupières	Distance entre œil et sourcil	Distance entre les coins de la bouche	Distance entre lèvre supérieure et lèvre inférieure	Distance entre les coins de l'œil et de la bouche
Joie 	Accroît ou décroît	Accroît ne change pas ou décroît	Accroît	Ne change pas ou accroît	Décroît
Surprise 	Accroît	Accroît	Ne change pas ou décroît	Accroît	Ne change pas ou accroît
Dégout 	Décroît	Décroît	Accroît ne change pas ou décroît	Accroît	Accroît ne change pas ou décroît
Colère 	Décroît ou accroît	Décroît	Ne change pas ou décroît	Accroît ne change pas ou décroît	Ne change pas ou accroît
Tristesse 	Décroît	Accroît	Ne change pas ou accroît	Ne change pas ou accroît	Ne change pas ou décroît
Peur 	Ne change pas ou accroît	Ne change pas ou accroît	Accroît ne change pas ou décroît	Ne change pas ou accroît	Accroît ne change pas ou décroît

TABLE 1.3 – L'étude anatomique Des Expressions faciales fondamentale du visage [12]

1.4 Différente Représentation des expressions faciales émotionnelles

Les représentations faciales relient les mouvements des muscles faciaux aux émotions exprimées. Elles représentent une sorte de dictionnaire utile à la reconnaissance des émotions. Dans ce qui suit nous définissons trois représentations des systèmes de codification utilisée dans les expressions faciales émotionnelles [13].

1.4.1 FACS

Le Facial Action Coding System (FACS) est une méthode permettant de décrire les mouvements faciaux observables. Cette classification est le fruit des recherches des psychologues : Paul Ekman, Wallace, Hager et Friesen dans les années 70. [web03]

Le système de codage des actions faciales FACS est probablement l'étude la plus connue sur l'activité faciale. C'est un système qui a été développé pour faciliter la mesure ob-

jective de l'activité faciale pour des investigations comportementales de la science sur le visage [14]

FACS est conçu pour les observateurs humains pour détecter les changements subtils indépendants de l'aspect facial provoqué par des contractions des muscles faciaux [14]

Les changements au niveau de l'expression faciale sont décrits avec les FACS en termes de 46 unités d'action différentes. Qui sont anatomiquement liées à la contraction d'un muscle spécifique du visage ou d'un ensemble de muscles [14]

La figure 1.5 illustre quelques images des AUs et la description de chaque unité d'action faciale

					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
					
Nose Wrinkler	Upper Lip Raiser	Lip Corner Puller	Lip Corner Depressor	Lower Lip Depressor	Chin Raiser
					
Lip Stretcher	Lip Tightener	Lip Pressor	Lips part	Jaw Drop	Mouth Stretch

FIGURE 1.2 – Liste des Action Units relatives aux 6 expressions faciales [15].

A partir d'une série de code FACS il est possible de reconstituer l'expression faciale codée, et donc d'y attribuer un sens. Par exemple, les expressions faciales émotionnelles ont fait l'objet de nombreuses études en lien avec le codage FACS. [web03]

Aujourd'hui les EFE sont considérées comme universelles et dépassent les considérations ethniques, culturelles et sexuelles. Pour chaque expression faciale il existe donc des prototypes de ces expressions faciales émotionnelles[web03]

Le tableau suivant présente la reconstruction des expressions faciales universelles à partir des unités d'actions (AUs) [web03]

Les émotions	Les unité d'actions (AUs)
La colère	AU4+AU5+AU7+AU23
La joie	AU6+AU12
La tristesse	AU1+AU4+AU15
La surprise	AU1+AU2+AU5+AU26
La peur	AU1+AU2+AU4+AU5+AU20+AU26
Le dégoût	AU9+AU15+AU16

TABLE 1.4 – Les prototypes des 6 émotions de bases avec le Système FACS [16]

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[15].
- 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 entraînant 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 [14]

1.4.2 MPEG4

Un deuxième codage très utilisé pour la description des expressions faciales est défini par les paramètres du standard de compression MPEG-4. Utilisé principalement pour l'animation et la synthèse du visage [13]. MPEG-4 définit FAPs (Facial Animation Parameters) qui est un ensemble de paramètres [13], permettent de définir des mouvements élémentaires du visage de manière transposable. En effet, il est difficile de définir les mouvements élémentaires des muscles de manière absolue, mais on peut considérer l'intensité de leur déplacement relative à certaines distances pertinentes comme constant 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 points sont effectuées pour former des unités de mesure qui servent à la description des mouvements musculaires (Facial Animation Paramètres - équivalents des Actions Unitaires d'Ekman). [15] Les Facial Animation Parameter Units (**FAPU**) permettent de définir des mouvements élémentaires du visage de manière transposable. 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 [12] Le modèle ainsi défini doit

être applicable sur des visages de différentes tailles. Pour ce faire, les paramètres d'animation faciale FAPs sont définis en fonction des unités de paramètres d'animation faciale (**FAPU** : Face Animation Parameter Units). Ces dernières sont les rapports des distances entre les caractéristiques faciales clés d'un visage à l'état neutre (représentation (b) figure 1.3)[13]

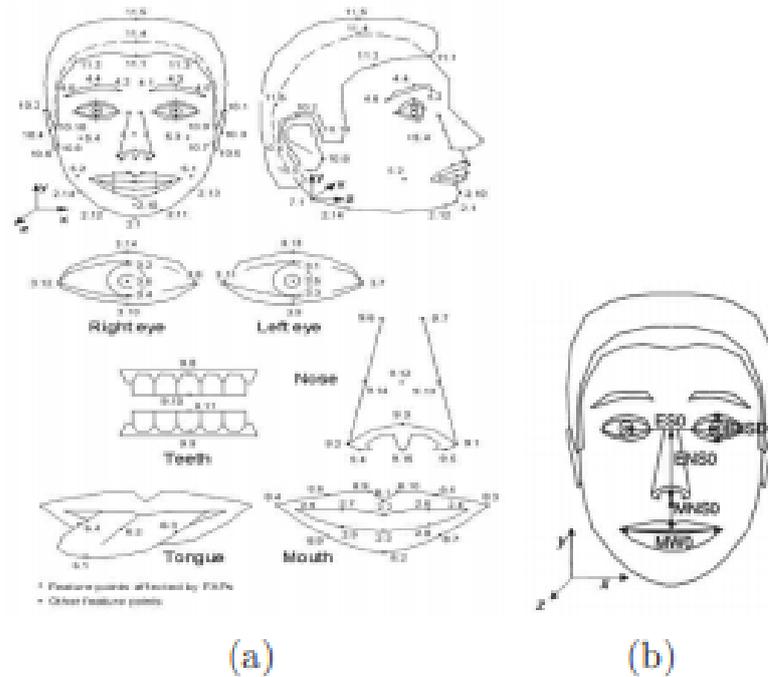


FIGURE 1.3 – Modèle du visage par le standard MPEG-4.
 (a) Représentation des points caractéristiques FAS.
 (b) Modèle des paramètres unitaires FAPUs.[13]

1.4.3 CANDIDE

CANDIDE est un masque facial paramétré spécialement développé pour le codage basé sur un modèle de visages humains. Son faible nombre de polygones (environ 100) permet une reconstruction rapide avec une faible puissance de calcul CANDIDE est contrôlé par des Unités d'action (UA). Les globales correspondent à des rotations autour de trois axes. L'action locale Les unités contrôlent les imitations du visage afin que différentes expressions puissent être obtenues [17].

Le CANDIDE original, décrit dans le rapport par Rydfalk [18], contenait 75 sommets et 100 Triangles. Cette version est rarement utilisée, La version la plus répandue est le modèle CANDIDE standard qui est légèrement modifié le modèle avec 79 sommets, 108 surfaces et 11, Unités d'action (voir la figure 1.4 en haut).

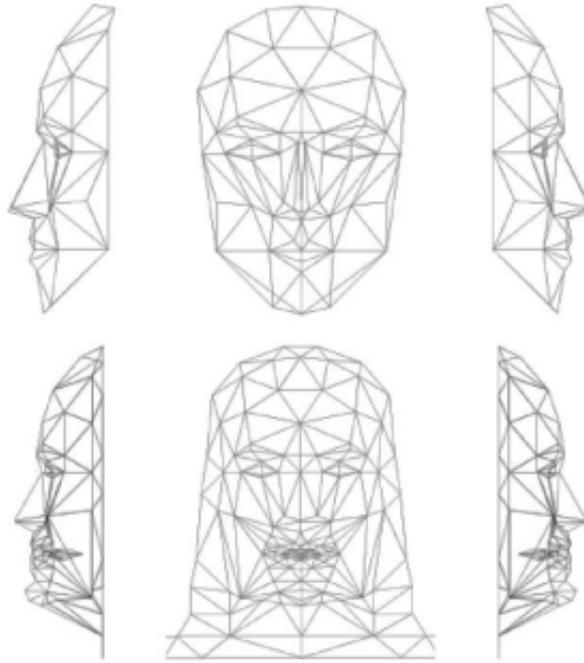


FIGURE 1.4 –
 En haut : CANDIDE-1 avec 79 sommets et 108 surfaces.
 En bas : CANDIDE-2 avec 160 sommets et 238 surfaces. [17]

Plus tard, Bill Welsh de British Telecom [19] a créé une autre version avec 160 sommets et 238 triangles couvrant toute la tête frontale et les épaules. Cette version, connue sous le nom de CANDIDE-2 est également incluse dans le progiciel CANDIDE, mais est livrée avec seulement six Unités d'action.(Figure 1.4 En bas)

Au cours des années 90, plusieurs modèles de visage avancés ont été créés par différents groupes de recherche et entreprises, Malgré cela, le modèle CANDIDE est encore largement utilisé, car sa simplicité en fait un bon outil pour les tâches d'analyse d'images et d'animation de faible complexité. Le Candide-3 est la dernière version avec des changements simple présenté dans la figure 1.5 suivante [17]

Les modifications dans la version 3 par rapport au version-1 (candide -1) sont [17] :

- Trois sommets (92 - 94) au milieu du nez
- Des sommets ont été ajoutés aux yeux, faisant les moins bruts (sommets 95-110).
- Les unités d'action définies dans CANDIDE-1 sont étendu pour inclure les nouveaux sommets.
- L'animation peut être réalisée par MPEG-4, FAP ainsi que par unités d'action.



FIGURE 1.5 – CANDIDE-3 avec 113 sommets et 168 surfaces.[17]

1.5 Modules du système de reconnaissance des expressions faciales

Un système qui effectue une reconnaissance automatique des expressions faciales est généralement composé de trois modules principaux. [20]

- **Le premier module** : consiste à détecter et enregistrer le visage dans un image ou une séquence d'image d'entrée
- **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
- **Les Modules intermédiaires** : peuvent être utilisés entre ces modules principaux pour améliorer les résultats de détection, d'extraction de caractéristiques ou de classification.

La figure suivante illustre l'architecture suivie dans les systèmes de reconnaissance d'expression faciale

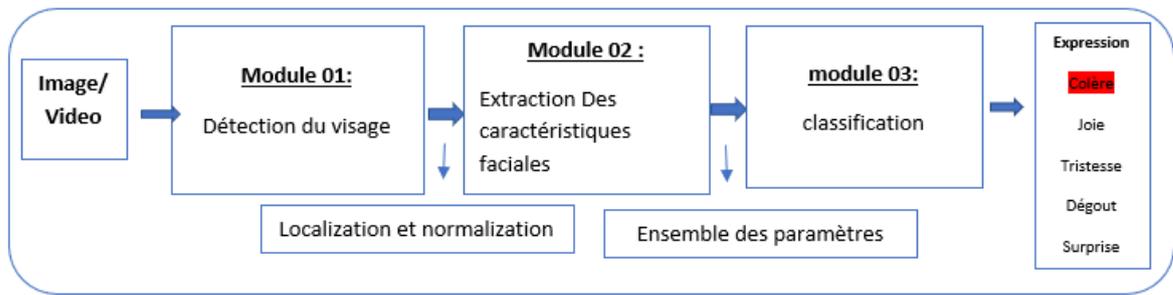


FIGURE 1.6 – Les Modules du système de reconnaissance des expressions faciales

1.5.1 Détection du visage

La première étape d'un système de reconnaissance automatique des EF 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 [20], Cette tâche préliminaire nécessaire et obligatoire à la plupart des techniques d'analyse du visage. Les techniques utilisées sont généralement issues du domaine de la reconnaissance des formes, En effet, le problème peut être vu comme la détection de caractéristiques communes à l'ensemble des visages humains : il s'agit de comparer une image à un modèle générique de visage et d'indiquer s'il y a ou non ressemblance. [21]

La sortie d'un détecteur de visage indique le nombre de visages présents dans l'image. De plus, la plupart des détecteurs de visage actuels sont aussi des localisateurs de visages : ils renvoient une localisation des visages détectés (la rectangle englobante). [16]

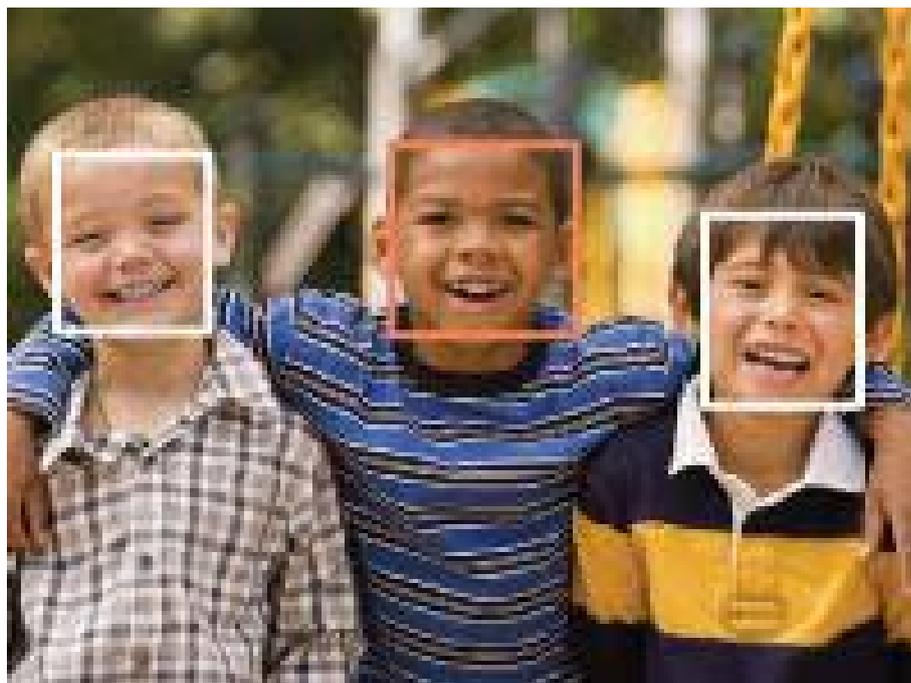


FIGURE 1.7 – Détection des visages

1.5.1.1 Les techniques de détection du visage

Il existe une multitude de techniques consacrées à la détection du visage. Mais pour la plupart il est d'intérêt que ces techniques se basent sur des éléments du visage qui sont le moins susceptibles aux changements. Les méthodes de base les plus utilisées sont dans le tableau suivant [22] :

Les méthodes	Définitions
Traitement automatique du visage	C'est une technologie rudimentaire, elle caractérise les visages par des distances et des proportions entre des points particuliers comme les deux yeux, le nez, les coins de la bouche. Aussi éprouvé que les autres technologies, le traitement automatique du visage est le plus efficace dans des situations de capture d'image avec peu d'éclairage [23].
Eigenface	Les « Eigenfaces » furent le premier type de caractérisation utilisé avec succès dans des traitements faciaux tels que la détection et la reconnaissance du visage [24] Elle utilise une représentation des éléments caractéristiques d'une image de visage à partir d'images modèles en niveau de gris. Des variantes d'Eigenface sont fréquemment utilisées comme base pour d'autres méthodes de reconnaissance [25].
L'analyse des points particuliers	Elle est la technique d'identification faciale la plus utilisée. Cette dernière se rapproche de Eigenface, mais elle est capable de s'adapter à des changements d'aspect facial (sourire, froncement des sourcils, ...). Les ingénieurs numériques l'utilisent souvent [23].
LDA (Linear discriminant analysis) fisher	Elle fait partie des techniques d'analyse discriminante prédictive. Il s'agit d'expliquer et de prédire l'appartenance d'un individu à une classe (groupe) prédéfinie à partir de ses caractéristiques mesurées à l'aide de variables prédictives [26].
Méthode (LBP)	Les LBP (Modèles binaires locaux) ont également été utilisés pour la détection du visage ou ce dernier est subdivisé en sous-régions carrées de taille égale sur lesquelles sont calculées les caractéristiques LBP. Les vecteurs obtenus sont ensuite concaténés pour obtenir le vecteur de caractéristiques final.[27]
Filtre Haar	La détection du visage dans ce filtre est réalisée par un filtre multi-échelles de Haar. Les propriétés d'un visage sont décrites dans un fichier XML. Elles ne sont pas choisies au hasard et reposent sur un échantillon de quelques centaines d'images tests [23]. Nous allons voir plus des détails sur cette méthode.

TABLE 1.5 – les méthodes de Base utilisé dans la détection du Visage

1.5.2 Extraction des caractéristiques

Après la détection et le suivi des visages, l'étape suivante consiste à extraire les traits représentatifs et distinctifs du visage. L'extraction des caractéristiques efficaces de l'image du visage détecté est cruciale pour une reconnaissance réussie des expressions faciales. L'extraction optimale des caractéristiques doit être effectuée de manière à minimiser les variations d'expression intra classe tout en maximisant les variations inter classes. Les fonctionnalités sont généralement divisées en deux Méthode : [28]

1.5.2.1 Les caractéristiques géométriques

Représentent la forme et l'emplacement des composants faciaux (les yeux, les sourcils, le nez, et la bouche) ou des points caractéristiques du visage prédéfinis, qui sont extraits pour former un vecteur de caractéristiques pour représenter la géométrie du visage

1.5.2.2 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.

1.5.3 La Reconnaissance des Expressions Faciales

Cette étape est considérée comme la dernière étape dans un système automatique de reconnaissance des expressions faciales en fonction des caractéristiques extraites.

1.5.3.1 Les Classes des techniques de REF

Ils existent différentes approches qui ont été développées dans le domaine de reconnaissance des expressions faciales à savoir, les approches globales, les approches locales, et enfin les approches hybrides [29],[30]

1.5.3.1.a Les Méthodes Globales ou Holistique

qui traitent le visage dans son ensemble comme une seule entité. Elles considèrent donc la région entière du visage et ne font pas la différence entre les différents composants du visage, Cette représentation basée sur l'apparence globale à l'avantage de conserver implicitement toute l'information de texture et de forme utile pour différencier les visages. Cependant il est très sensible à la variation de l'éclairage et de la pose, la majorité des méthodes appartenant à cette catégorie sont des méthodes statiques de réduction de dimension de l'espace de description. Les techniques les plus utilisés dans ces méthodes sont : **PCA**, **k PCA**, **EBGM**, **LLE**.

1.5.3.1.b Les Méthodes Locales

qui représentent un visage par un ensemble de régions sur les quelles des statiques sont calculées afin de créer une description du visage. Bien que ces méthodes soient simples et que leur coût de stockage soit très réduit, l'utilisation de quelques paramètres géométriques n'est pas suffisante pour représenter les variations des visages. Les techniques les plus connu des méthodes locales sont : AAM, LBP, SVM.

1.5.3.1.c Les Méthodes Hybrides

les méthodes Hybrides combinent des propriétés globale et locale, cette méthode est donc comme le système de perception humain, utilisent à la fois des caractéristique locale et globale de la région du visage afin d'identifier [29], Cela permet d'obtenir plus de précision et combiner les avantages des deux méthodes avec les techniques les pertinente : **LFA, EBGM, LG** . Grâce aux combinaisons de ces deux méthodes (locale et globale), les résultats seront les plus efficaces en termes de robustesse.

La figure suivante résume la classification des algorithmes pour la reconnaissance des expressions faciales [31]

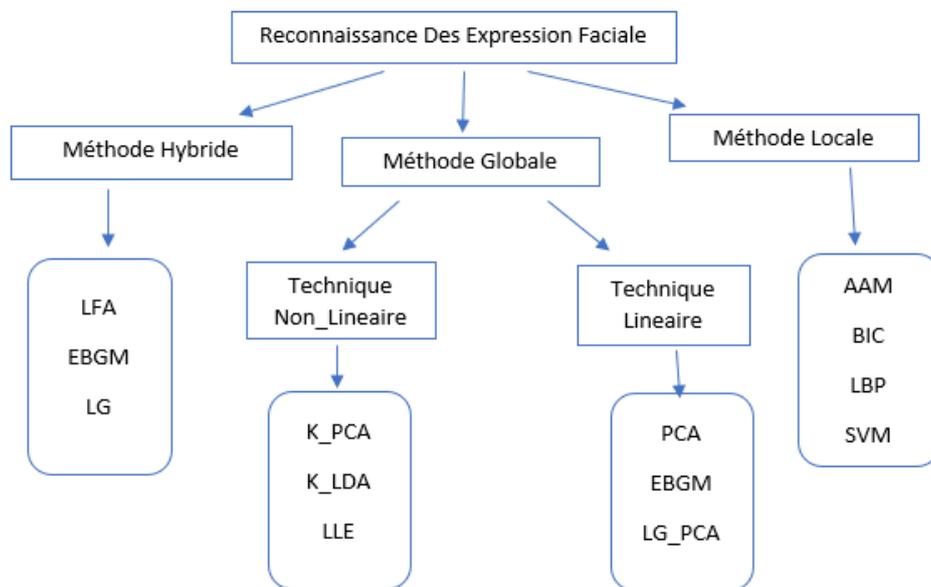


FIGURE 1.8 – Les techniques utilisées dans la Reconnaissance des EF.

1.5.3.2 Les Types de Reconnaissance EF

Nombreuses méthodes de reconnaissance des EF ont été proposées durant ces 30 dernières années, suivant deux grands axes [20] :

1.5.3.2.a La reconnaissance basée sur les images fixe (statique)

La Reconnaissance Statique est la Reconnaissance qui est basée sur les image fixe, cette dernière n'utilise pas d'information temporelle pour les images d'entrée. Elle utilise que les informations de l'image d'entrée courante. L'image d'entrée peut être une image statique ou une image d'une séquence traitée indépendamment des autres images de la séquence. Plusieurs méthodes peuvent être trouvées dans la littérature pour la reconnaissance d'expressions faciales telles que les NNs[32], SVM [33,34], BN [35]

1.5.3.2.b La reconnaissance basée sur les séquences vidéo (dynamique)

Dans ces dernières années et très récemment les approches dynamiques de classification des Expression Faciales (EF) sont basée sur le deep Learning qui sont devenues de plus en plus dominante et une large propagation. Par conséquent, plusieurs travaux basés sur CNN et Deep Belief Networks (DBN) ont été proposés pour la classification dynamique des expressions faciales

1.5.3.3 Les Bases de Données utilisées dans expressions faciales :

Plusieurs bases de données contenant des informations qui permettent l'évaluation des systèmes de reconnaissance des Expression Faciales sont disponibles sur le marché [39]. De façon à obtenir l'évaluation la plus objective possible, les bases de données doivent être générales et indépendantes des restrictions ou des hypothèses liées à un domaine précis. Ainsi, les performances des méthodes peuvent être référencées et comparées plus facilement.[13]

1.5.3.3.a JAFFE

La base de données JAFFE (Japanese Female Facial Expression Database) comporte 219 images statiques. Ces images concernent dix femmes japonaises qui simulent les six émotions universelles Tristesse , joie , la colère , le dégoût , surprise , peur et l'expression neutre (voir le Figure 1.8) [13]

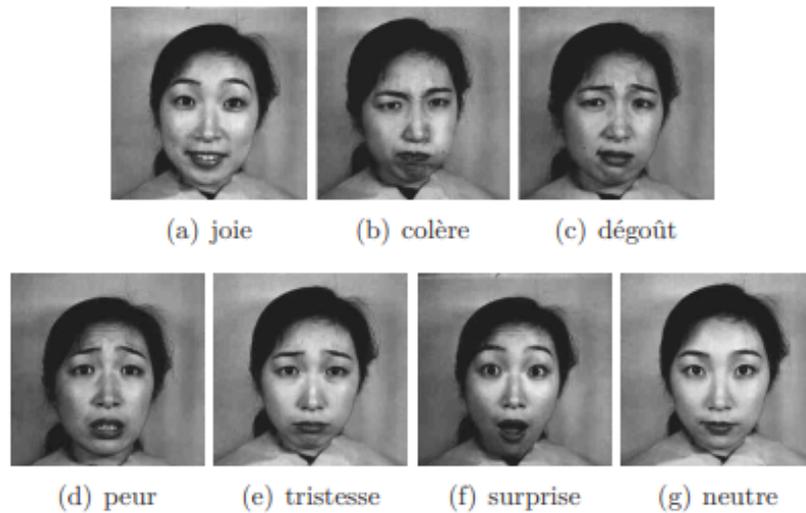


FIGURE 1.9 – Exemple de La base JAFFE [13]

1.5.3.3.b Cohn Kanade

En 2000, la base de données Cohn-Kanade (CK) a été publiée dans le but de promouvoir la recherche sur détecter les expressions faciales individuelles. Depuis, le CK la base de données est devenue l'un des bancs d'essai les plus utilisés pour le développement et l'évaluation d'algorithmes [36]

La première version de cette base de données d'expressions simulées par des acteurs, cette dernière contient 97 sujets, âgés de 18 à 30 ans. Ces sujets sont constitués de 65% de femmes, 15% d'afro-américains et 3% d'asiatiques et de latinos. Les séquences d'images présentées commencent par l'expression neutre et se terminent par le maximum de l'expression [13]

Ck+ est considérée comme la version améliorée dans [36] comporte deux types d'expressions, posées et spontanées. Les expressions posées présentent une continuation de la première version, le nombre de vidéos augmentant de 22% et le nombre de sujets de 27%. Les séquences d'images sont organisées de la même manière que les séquences de la première version [13]

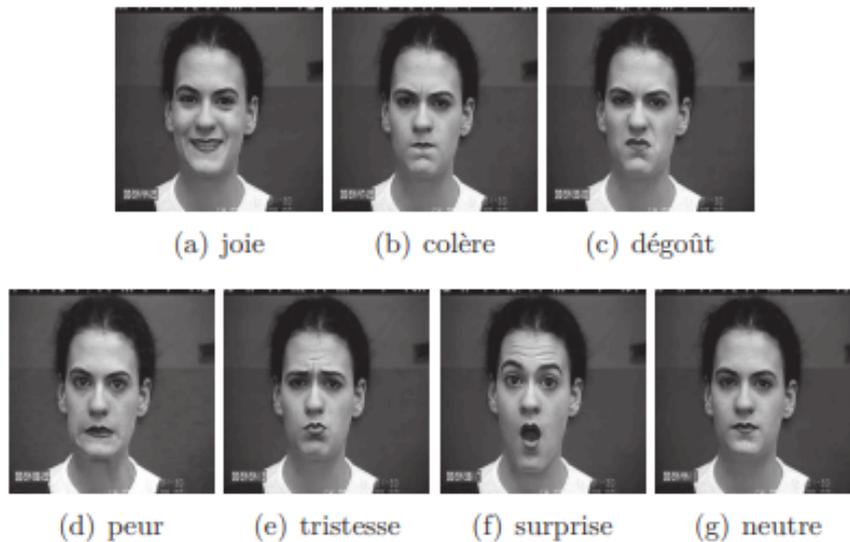


FIGURE 1.10 – Exemple de la Base CK+ [15]

1.5.3.3.c FER2013

Fer-2013 est un ensemble de données open-source qui est d'abord, créé par Pierre-Luc Carrier et Aaron Courville, puis partagé publiquement pour un concours Kaggle, peu avant ICML 2013.

La base de données a été créée à l'aide de réseaux de neurones dans les domaines de la reconnaissance visuelle d'objets, de l'estimation de la pose humaine, de la vérification du visage et bien d'autres. Même dans le domaine du FER, les résultats ont jusqu'à présent été prometteurs, et la plupart des gagnants du défi de reconnaissance de l'expression faciale ont utilisé des réseaux de neurones profonds [40].

Cet ensemble de données se compose de 35,887 niveaux de gris, images de visage de taille 48×48 avec diverses émotions tell que (la joie, la colère, peur, triste, surprise, dégoût, neutre)

Label	Nombre images	Emotions
0	4593	En colère
1	547	Dégoût
2	5121	Peur
3	8989	La joie
4	6077	Tristesse
5	4002	Surprise
6	6198	Neutre

TABLE 1.6 – Etiquette d'émotion dans l'ensemble de Donnée Fer 2013[web05]

Les 3 589 images restantes représentent les images de test [web05]

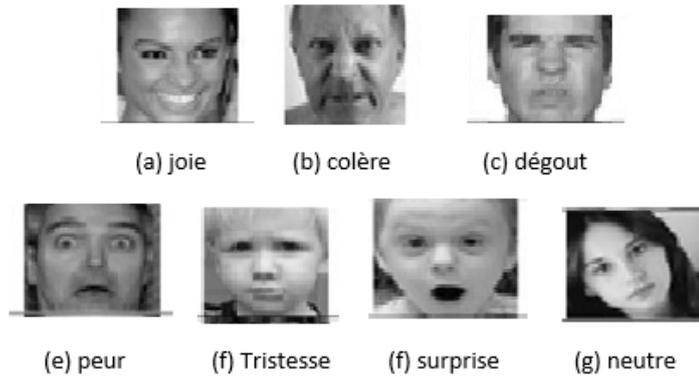


FIGURE 1.11 – Exemple de la base Fer2013

1.6 Performance et L'état du conducteur dans le cas de « colère »

1.6.1 Performance du conducteur

Les mesure objective de la manière dont le conducteur contrôle le véhicule est un élément clé de ses performances. Ces mesures constituent le moyen le moins invasif de détecter l'état du conducteur car il n'y pas d'interaction direct avec celui-ci, d'une part, ces Mesures résultent directement de l'intervention du conducteur dans le contrôle du véhicule, telle que : la direction, l'accélérateur et les freins.

Les Véhicules sont progressivement équipés de systèmes de détection des métriques du conducteur, ces mesures sont donc particulièrement appropriées. Ces mesures du conducteur se composent de la position de la voie, de l'avance et de l'angle du volant [37]

Les performances peuvent être également influencées par des plusieurs facteurs tels que :

- L'Expérience du conducteur
- Les distractions et les conditions de conduite
- Et plus nécessairement les comportements **cognitifs agressif**

1.6.2 L'état Du conducteur « la colère »

La colère est l'un des émotions négatives qui peuvent grandement la première cause des accidents dans une session de conduite. La capacité du conducteur à conduire peut-être déterminée par la façon dont il se comporte au volant.

Les comportements révélateurs de la colère ou d'autres situations de conduite dangereuse, elles se manifestent par plusieurs formes telle que : regard vague, direct et agressif, lèvres pincées, les sourcils sera rapprochés vers le bas, fermeture des yeux, etc. [38] (voir le figure

1.8)

Une des manifestations particulièrement remarquables de la colère en conduite provient de son expression par les conducteurs. Quatre moyens d'expression de la colère ont été identifiés en conduisant qui sont :

- L'expression verbale agressive qui peut se manifester par des insultes ou des cris
- L'expression personnelle physique agressive qui regroupe les possibles agressions physiques envers les autres conducteurs
- L'expression de la colère par l'utilisation du véhicule, ce qui conduit à des appels de phares, des obstructions de voies et à des comportements délibérément dangereux pour les autres usagers [39]

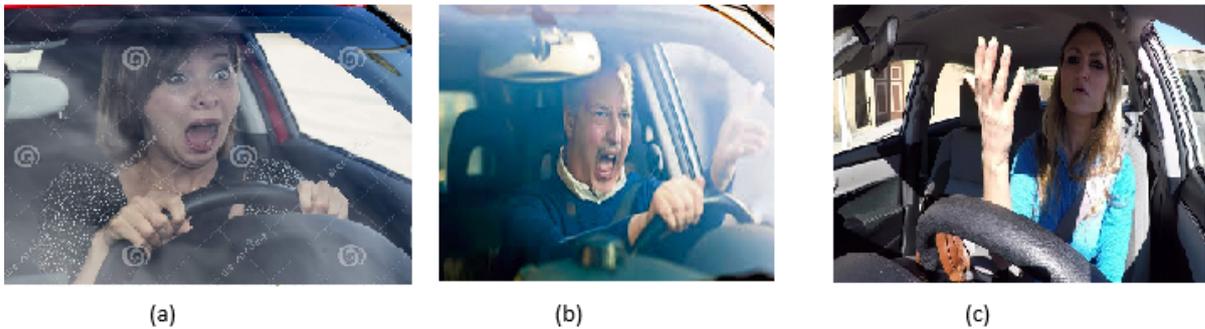


FIGURE 1.12 – différents états de colère de gauche à droite

- (a) : Les yeux sont ouverts, et la bouche également
- (b) : les sourcils sont rapprochés vers le bas et la bouche est ouverte
- (c) : le regard direct et agressif, les yeux sont à peine ouverts, les lèvres sont pincées

1.7 Conclusion

Dans ce chapitre, nous avons présenté une revue de littérature des expressions faciales telle que : les définitions, les études psychologiques et anatomiques et les différentes représentations des expressions faciales et également les modules suivis dans la reconnaissance de ces derniers et finalement nous avons mis au point la performance et l'état de conducteur dans ces différentes dimensions.

Dans le chapitre suivant, nous allons voir l'une des stratégies les plus suivies dans les systèmes de reconnaissance automatique des expressions faciales qui est le Deep Learning. Nous allons par conséquent répondre aux questions suivantes :

- Qu'est-ce que le Deep Learning ?
- Quels sont les mécanismes de l'implémentation de DL ?
- Pourquoi le DL est important et où fait la différence ?

Nous allons également parler sur les réseaux de neurone convolutifs (CNN) qu'est l'une des méthodes de l'apprentissage automatique qui repose sur l'apprentissage profonde (Deep Learning).

Chapitre 2

Réseaux de neurones et le Deep Learning

Sommaire

2.1	Introduction	24
2.2	L'apprentissage en profondeur (Deep Learning)	25
2.3	Les Réseaux de neurones Convolutifs (CNN)	28
2.4	Quelques réseaux convolutifs célèbres	35
2.5	Pourquoi Le CNN ?	36
2.6	Conclusion	36

2.1 Introduction

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 autre sont des exemples de systèmes d'intelligence artificielle.

Le 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. Le 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 connue un grand succès dans ces cinq dernières années, Pour illustrer la relation entre ces termes, nous pouvons utiliser des cercles concentriques :

- Intelligence artificielle IA (**intelligence artificielle**) : le cercle 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

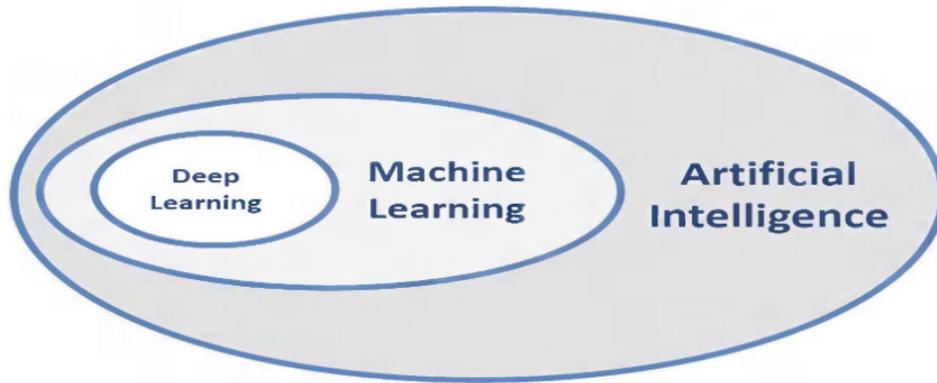


FIGURE 2.1 – La relation Entre L'IA et ML et le Deep Learning

Dans ce chapitre, nous allons concentrer beaucoup plus sur Le « **Deep Learning** », l'élément le plus intéressant dans les entreprises du monde numérique et qui booste l'Intelligence Artificielle, cet outil d'apprentissage basé sur des « réseaux de neurones artificiels », est utilisé par Google Now, pour la reconnaissance faciale, reconnaissances des Expressions faciale ou reconnaissance vocale.

2.2 L'apprentissage en profondeur (Deep Learning)

2.2.1 Définition

L'apprentissage en profondeur (Deep Learning : DL) est une branche de l'apprentissage automatique qui enseigne aux ordinateurs à faire ce qui vient naturellement aux humains : apprendre de l'expérience. Les algorithmes d'apprentissage automatique utilisent des méthodes de calcul pour "apprendre" des informations directement à partir de données sans s'appuyer sur une équation prédéterminée comme modèle. DL est une classe de techniques d'apprentissage automatique qui exploitent de nombreuses couches de traitement d'informations non linéaires pour l'extraction et la transformation de fonctions supervisées ou non supervisées, et pour l'analyse et la classification de modèles.

Les caractéristiques et les concepts de niveau supérieur sont donc définis en termes de niveaux inférieurs, et une telle hiérarchie de caractéristiques est appelée architecture pro-

fonde. La plupart de ces modèles sont basés sur un apprenti sage non supervisé des représentations [41].

2.2.2 Histoire de Deep Learning

Année	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 considéré comme L'ancêtre des réseaux de neurones artificielles
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 réseaux de neurones modernes
1958	Frank Rosenblatt	introduction du premier perceptron
1974	Paul Werbos	introduction de la retro propagation
1980	Teuvo Kohonen	introduction des cartes auto organisatrices
1980	Kunihiko Fukushima	introduction du Neocognitron, qui a inspiré les réseaux de neurones convolutif
1982	John Hopfield	introduction des réseaux de Hopfield
1985	Hilton and Sejnowski	introduction des machines de Boltzmann
1986	Paul Smolensky	introduction de Harmonium, qui sera connu plus tard comme machines de Boltzmann restreintes.
1986	Michael I. Jordan	définition et introduction des réseaux de neurones récurrent
1990	Yann LeCun	introduction de LeNet et montra la capacités des réseaux de neurones profond
1997	Schuster and Paliwal	introduction des réseaux de neurones récurrent bidirectionnelles
1997	Hochreiter and Schmidhuber	introduction de LSTM, qui ont résolu le problème du vanishing gradient dans les réseaux de neurones récurrent
2006	Geoffrey Hinton	introduction des Deep belief Network
2009	Salakhutdinov and Hinton	introduction des Deep Boltzmann Machines
2012	Alex Krizhevsky	introduction de AlexNet qui remporta le challenge ImageNet

TABLE 2.1 – Les étapes majeures du Deep Learning [42]

2.2.3 Concepts de Deep Learning

Les concepts de Deep Learning constituent une classe d'algorithmes d'apprentissage automatique dont les caractéristiques sont les suivantes [43] :

- Elles 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
- Elles fonctionnent avec un apprentissage à plusieurs niveaux de détails ou de représentations des données. A 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 davantage 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 (comme il est illustré dans la figure 9)

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.

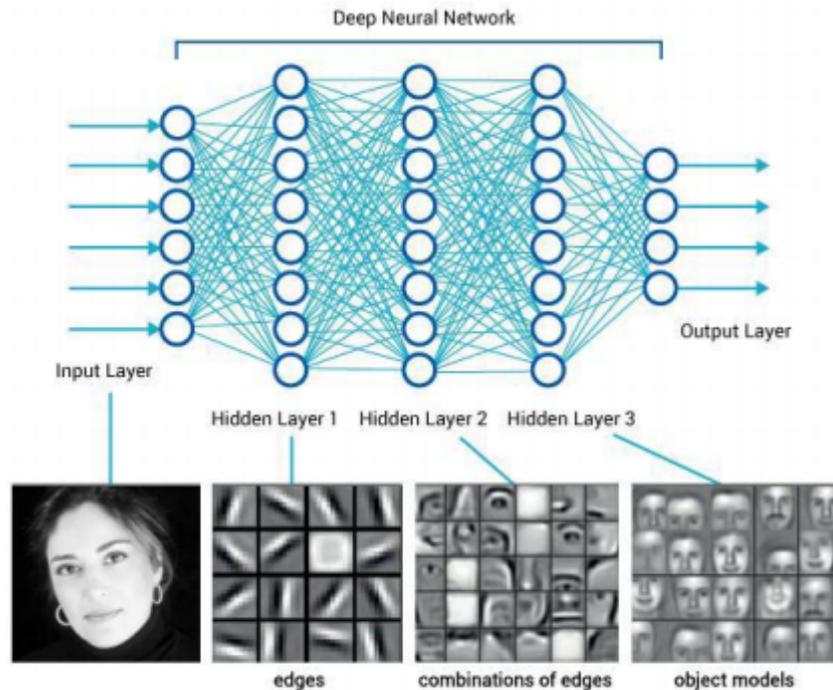


FIGURE 2.2 – Exemple de fonctionnement de Deep Learning [Web 08]

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 reconnaître les objets avec un taux d'erreur petit.

2.2.4 Machine Learning vs Deep Learning : où le Deep Learning Fait la difference ?

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

TABLE 2.2 – la Différence Faite par le Deep Learning[web 10]

2.2.5 Les avantages de Deep Learning

Permis les avantages qui distinguent le Deep Learning, on cite [web 11] :

- 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.

2.3 Les Réseaux de neurones Convolutifs (CNN)

Les réseaux de neurones convolutifs sont à ce jour les modèles les plus performants pour des images. Désignés par l'acronyme CNN, (de l'anglais Convolutional Neural Network), ils comportent deux parties bien distinctes. En entrée, une image est fournie sous la forme d'une matrice de pixels. Elle a deux dimensions pour une image aux niveaux de gris.

La couleur est représentée par une troisième dimension, de profondeur 3 pour représenter les couleurs fondamentales [Rouge, Vert, Bleu].

La première partie d'un CNN est la partie convolutive à proprement parler. Elle fonctionne comme un extracteur de caractéristiques des images. Une image est passée à travers d'une succession de filtres, ou noyaux de convolution, créant de nouvelles images appelées cartes de convolutions. Certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum local. Enfin, les cartes de convolutions sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN.

Ce code CNN en sortie de la partie convolutive est ensuite branché en entrée d'une deuxième partie, constituée de couches entièrement connectées. Le rôle de cette partie est de combiner les caractéristiques du code CNN pour classer l'image.

La sortie est une dernière couche comportant un neurone par catégorie. Les valeurs numériques obtenues sont généralement normalisées entre 0 et 1, de somme 1, pour produire une distribution de probabilité sur les catégories. [44]

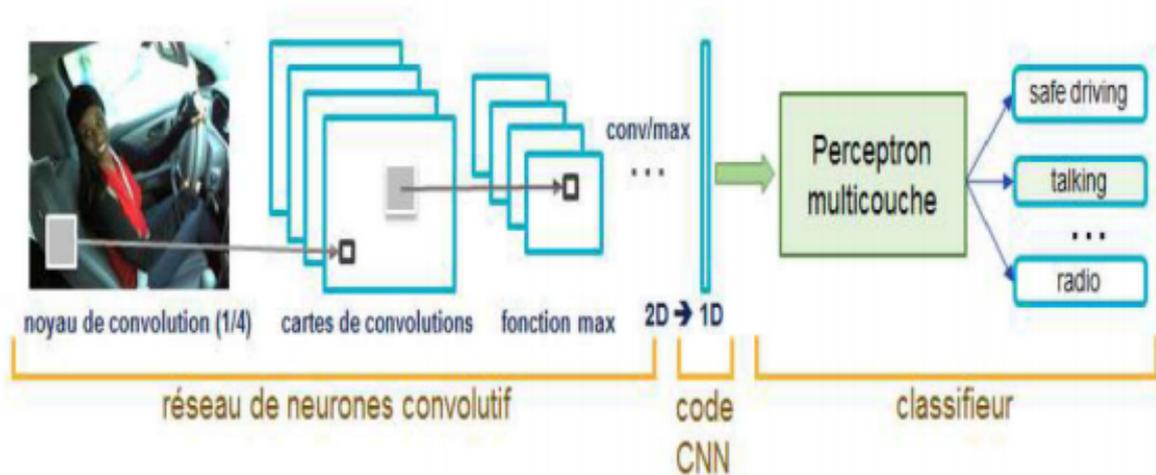


FIGURE 2.3 – Les réseaux de neurones convolutifs [44]

2.3.1 Les couches constructives de CNN (architecture de CNN)

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 [45]

- 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) [44]

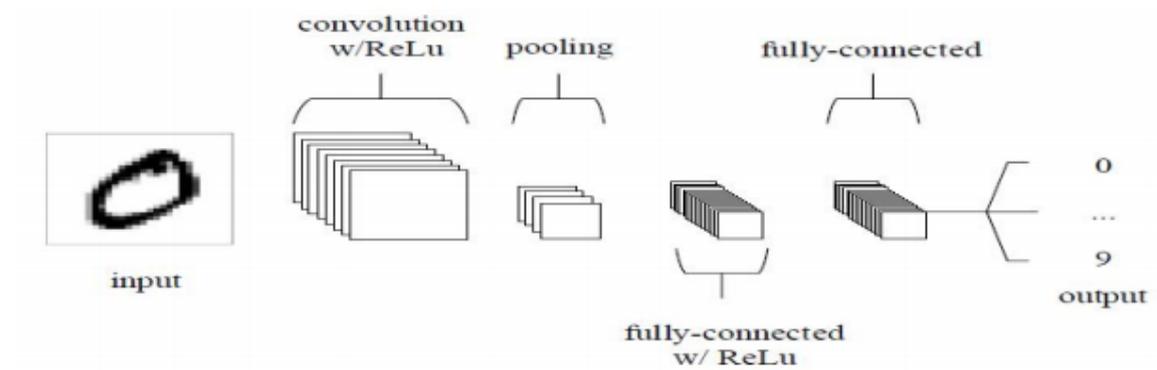


FIGURE 2.4 – Simple Exemple d’une Architecture Globale de CNN [45]

2.3.2 La couche de convolution (CONV)

La couche de convolution est le bloc de construction de base d’un CNN, Il existe trois éléments principaux qui entrent dans l’opération de convolution :

- Image d’entrée (input)
- Détecteur de caractéristiques (Filtre / Kernel)
- Carte des caractéristiques(output)

La couche de convolution reçoit donc en entrée plusieurs images, et calcule la convolution de chacune d’entre elles avec chaque filtre. Les filtres correspondent exactement aux features que l’on souhaite retrouver dans les images, On obtient pour chaque paire (image, filtre) une carte d’activation, ou **feature map**, qui nous indique où se situent les features dans l’image : plus la valeur est élevée, plus l’endroit correspondant dans l’image ressemble à la feature.

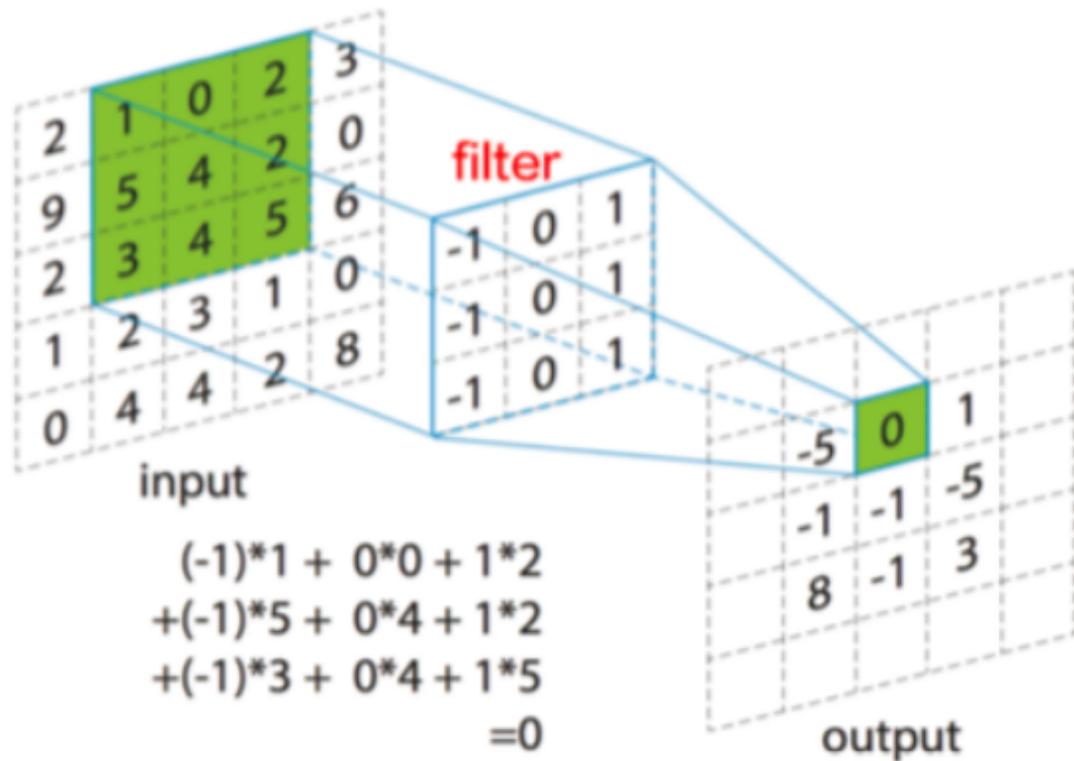


FIGURE 2.5 – exemple d’une opération convolution

Ainsi, lors de l’utilisation d’un CNN, les trois hyperparamètres importants que nous devons choisir sont : la profondeur, le pas et la marge [46]

- **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.

2.3.2.1 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. [web12]

- **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.

- **La dilated convolution** : 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.
- **La transposed convolution** : qui construit la sortie comme si on inversait une convolution sur l'image
- **La séparable convolution** : qui est une convolution décomposable en convolutions plus simples

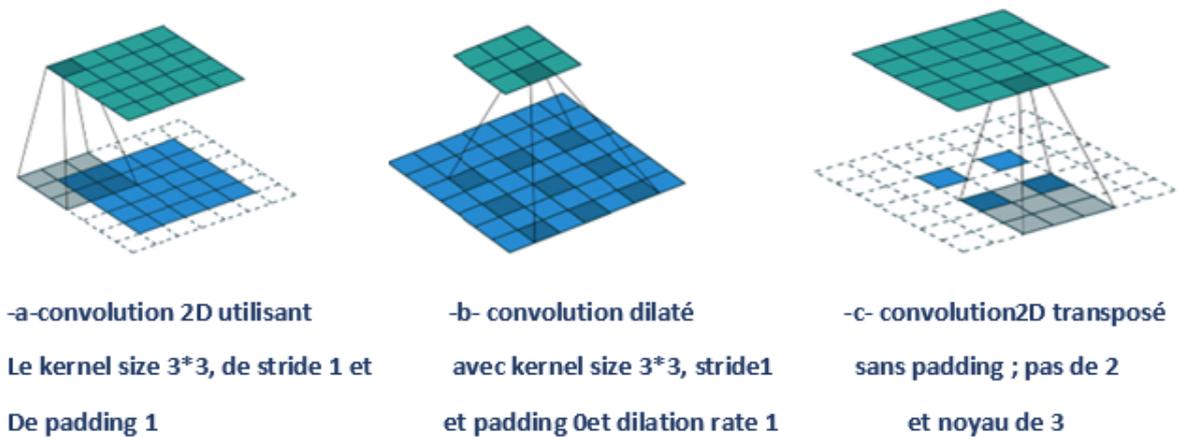


FIGURE 2.6 – Différent type de convolution

2.3.3 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 [45]. La fonction Relu $\mathbf{F}(\mathbf{x})=\max(0, \mathbf{x})$ Cette fonction force les neurones à retourner des valeurs positives. [47] [48]

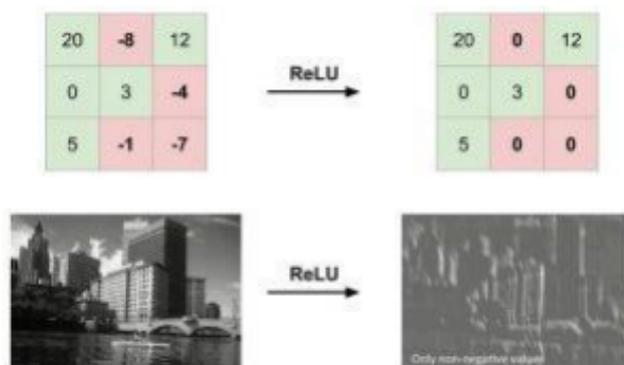


FIGURE 2.7 – application de RELU à une matrice et à une image [49]

2.3.4 La Couche pooling (POOL)

Une architecture atypique d'un réseau convolutif se compose de trois types de couches différentes. D'abord une couche convolutive pour générer un ensemble d'activations linéaires ensuite, on les fait passer à travers une couche d'activation non linéaire telle que Rectified Linear Unit (ReLU), enfin on utilise la fonction pooling [46].

Il existe plusieurs types de pooling [web 12] :

- **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 mean pooling** : (ou average pooling), 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 sum pooling** : c'est la moyenne sans avoir divisé par le nombre de valeurs (on ne calcule que leur somme).

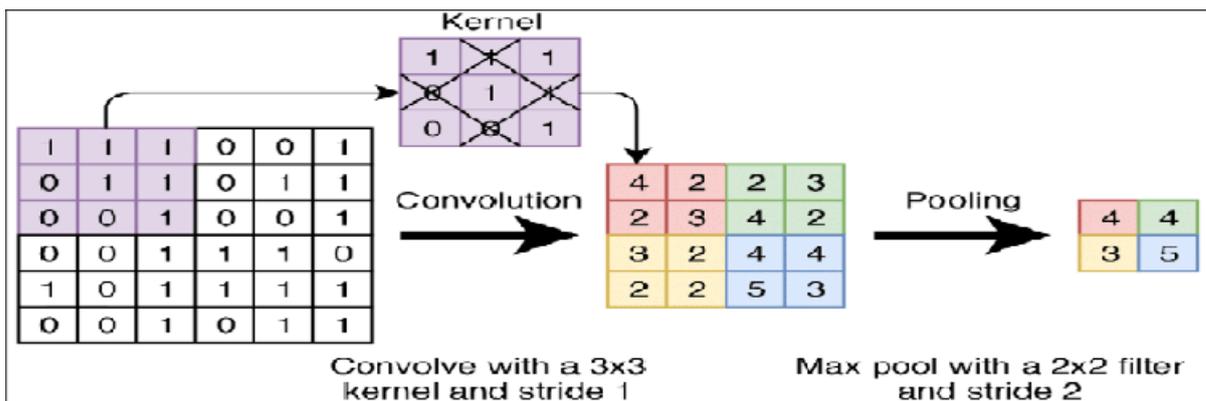


FIGURE 2.8 – Exemple de la phase pooling en utilisant le Max pooling

2.3.4.1 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 [49].

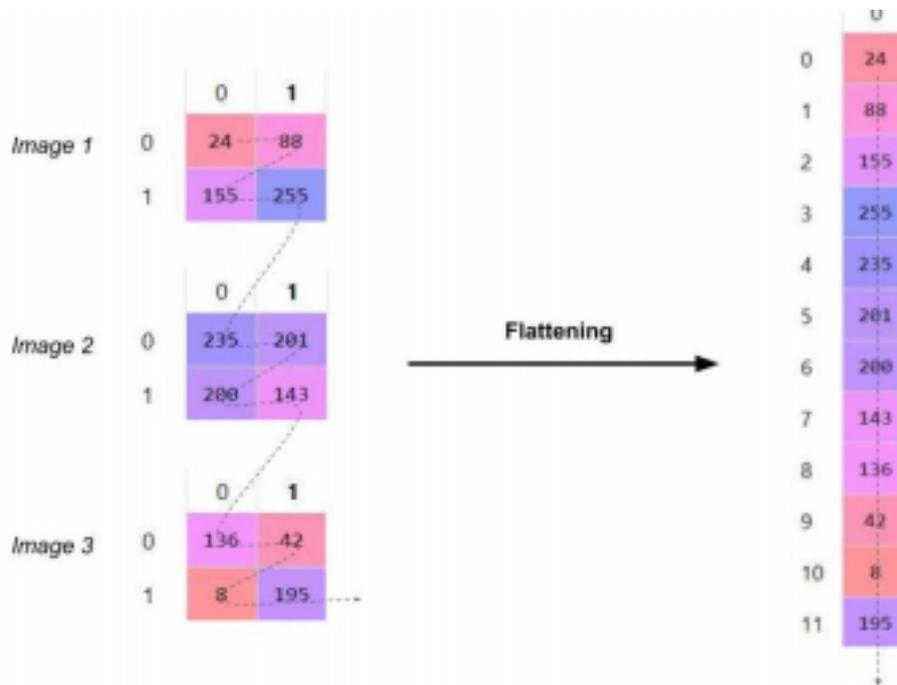


FIGURE 2.9 – Exemple de Mise à plat des images finale en sortie (des filtres +simplifications) [49]

2.3.5 la Couche entièrement connectée (FC)

Enfin, 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 [45]. 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 [45].

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. [47] [48] [web 13]

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 [50]

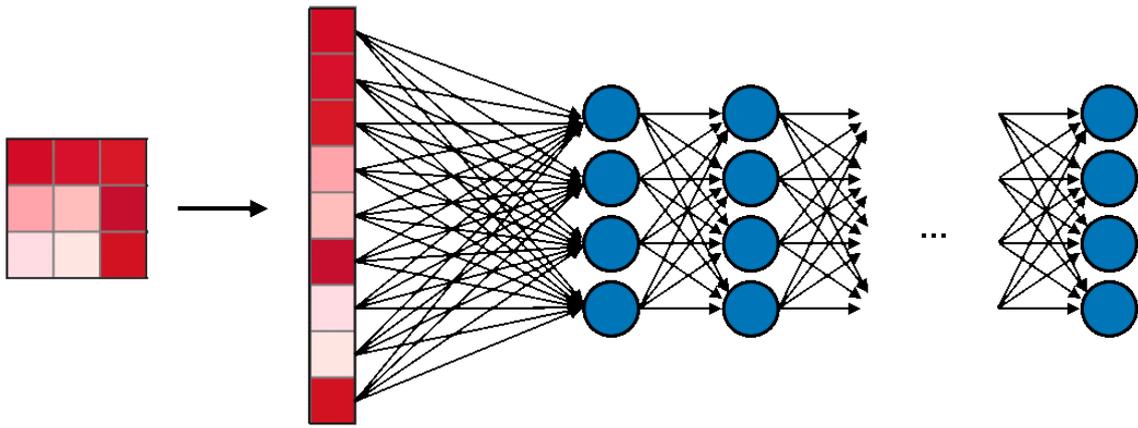


FIGURE 2.10 – un exemple d'opération entièrement connectée (FC)

2.3.6 La Couche de perte (LOSS)

La couche de perte spécifie comment l'entraînement du réseau pénalise l'écart entre le signal prévu et réel. Elle est normalement la dernière couche dans le réseau. Diverses fonctions 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. [44]

2.4 Quelques réseaux convolutifs célèbres

- **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. [51] [42]
- **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 [52]. AlexNet a été soumis au défi ImageNet ILSVRC [53], 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). [42]
- **Inception V3** : Ce type d'architecture, introduit en 2014 par Szegedy et al. Dans [55] utilise des blocs avec des filtres de différentes tailles qui sont ensuite concaténés pour extraire des caractéristiques à différentes échelles [web14]
- **Xception** : Cette architecture a été proposée par François Chollet (le créateur

de Keras) et la seule chose qu'il apporte à Inception est qu'il effectue de manière optimale les circonvolutions pour qu'elles prennent moins de temps. Ceci est réalisé en séparant les convolutions 2D en 2 convolutions 1D [54] [web14]

- **VGG 16** : Cette architecture, qui a été l'une des premières à apparaître, a été introduite par Simonyan et Zisserman en 2014. Il s'agit d'une architecture simple, utilisant uniquement des blocs composés d'un nombre incrémentiel de couches convolutives avec des filtres de taille 3x3. En outre, pour réduire la taille des cartes d'activation obtenues, des blocs de pooling max sont intercalés entre les blocs convolutifs, réduisant de moitié la taille de ces cartes d'activation. Enfin, un bloc de classification est utilisé, composé de deux couches denses de 4096 neurones chacune, et la dernière couche, qui est la couche de sortie, de 1000 neurones. [web14]

2.5 Pourquoi Le CNN ?

Un Avantage majeur des réseaux convolutifs est l'utilisation d'un poids unique associé aux signaux entrant dans tous les neurones d'un même noyau de convolution. Cette méthode réduit l'empreinte mémoire, améliore les performances et permet une invariance du traitement par translation. C'est le principal avantage du CNN par rapport au MLP, qui lui considère chaque neurone indépendant et donc affecte un poids différent à chaque signal entrant. Lorsque le volume d'entrée varie dans le temps (vidéo ou son), il devient intéressant de rajouter un paramètre de temporisation (delay) dans le paramétrage des neurones. On parlera dans ce cas de réseau neuronal à retard temporel (TDNN).

Comparés à d'autres algorithmes de classification de l'image, les réseaux de neurones convolutifs utilisent relativement peu de pré-traitement. Cela signifie que le réseau est responsable de faire évoluer tout seul ses propres filtres (apprentissage sans supervision), ce qui n'est pas le cas d'autres algorithmes plus traditionnels. L'absence de paramétrage initial et d'intervention humaine est un atout majeur des CNN [46]

2.6 Conclusion

Dans ce Chapitre, nous avons présenté les concepts de base concernant le Deep Learning et Le réseaux de neurone convolutifs (CNN) car l'utilisation de ce genre d'algorithmes est en pleine expansion, et cela dû aux résultats qu'il fournit. Nous allons faire de même et utiliser ce concept dans ce travail pour concevoir un système de détection de la colère du conducteur. Nous donnons beaucoup de détails de conception de ce système dans le chapitre suivant.

Chapitre 3

Conception d'un système de détection de la colère chez le conducteur en utilisant le Deep Learning

Sommaire

3.1	Introduction	37
3.2	Objectifs du système	38
3.3	Conception Générale de Système	38
3.4	Conception détaillée du système	38
3.5	Conclusion	44

3.1 Introduction

Dans les chapitres précédents, nous avons présenté la phase théorique concernant la reconnaissance des expressions faciales et le Deep Learning avec un outil de CNN (convolutional neural network)

Dans ce chapitre, nous allons mettre en pratique ce concept et proposer un système de détections de la colère et les autres expressions faciales universelles tel que : la joie, le dégoût, la peur, la surprise, la tristesse et l'expression de neutre. C'est pour cela nous avons organisé ce chapitre comme suit :

Dans La première partie, nous allons présenter la méthodologie de travail suivie et ces étapes nécessaires pour obtenir les résultats souhaités dans un modèle descriptif puis nous allons décrire en détails chaque étape.

3.2 Objectifs du système

Notre travail se dirige dans le cadre des systèmes de reconnaissance d'expressions faciales où nous concentrons sur la détection de la colère chez le conducteur.

Comme paradigme d'apprentissage, on va utiliser l'approche profonde ou le Deep Learning qui a prouvé son efficacité dans ce genre de problème. Pour le montage du système, on va utiliser les réseaux de neurones convolutionnels CNN.

3.3 Conception Générale de Système

Dans ce travail, nous avons divisé notre application en deux parties :

- La première partie est la phase d'entraînement utilisé les modèles CNN profond avec une base de données et préparer un classificateur pour la deuxième partie.
- La deuxième partie est la phase complétive consiste à prédire les expressions faciales en utilisant un classificateur entraîné dans la première partie. Le schéma suivant donne une idée sur le fonctionnement de notre système.

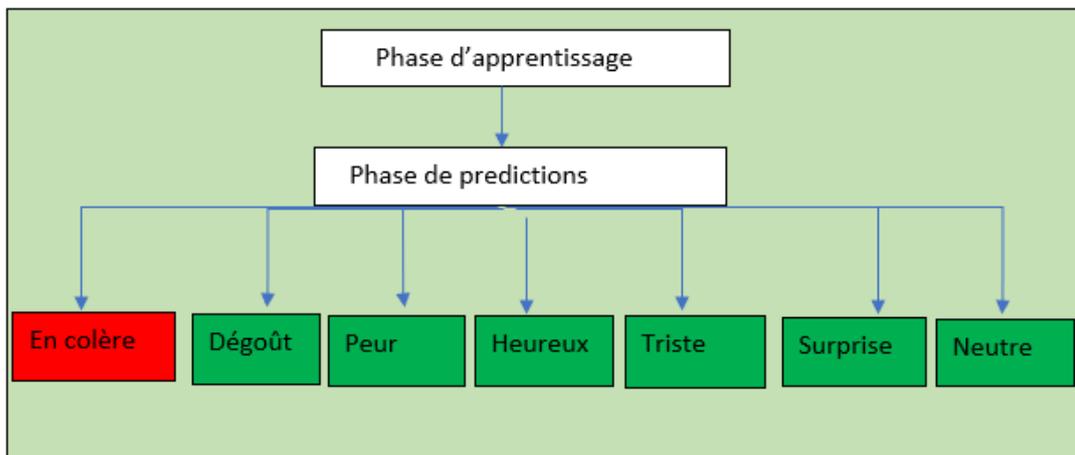


FIGURE 3.1 – schéma globale du système

3.4 Conception détaillée du système

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

3.4.1 La phase d'apprentissage

Le but de la phase d'apprentissage est de créer un classificateur qui peut classer les expressions faciales, cette phase consiste à utiliser un modèle CNN qualifié avec une base

de données, le schéma suivant illustre le processus de la phase d'apprentissage.

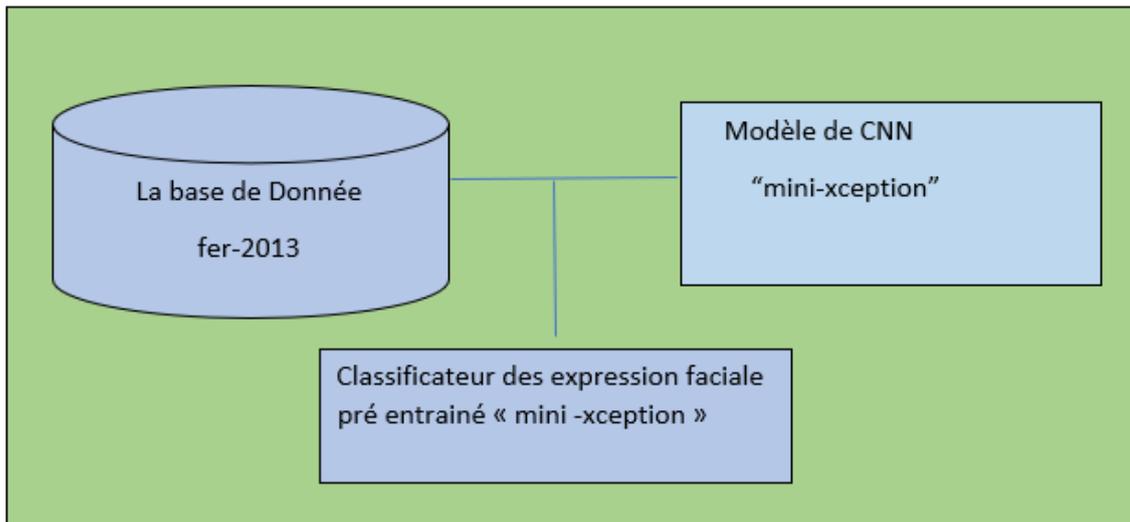


FIGURE 3.2 – processus de la phase d'apprentissage

3.4.1.1 la base de données

Nous avons utilisé l'ensemble de données « fer-2013 » fourni par le site Web de Kaggle, qui se compose d'environ 35887 images bien structurées en niveaux de gris de 48×48 pixels de visages.

Les images sont traitées de manière que les faces sont presque centrées et chaque face occupe à peu près la même quantité d'espace dans chaque image. Chaque image doit être classée dans l'une des sept classes qui expriment différentes émotions faciales. Ces émotions faciales ont été classées comme suit : 0 = En colère, 1 = Dégoût, 2 = Peur, 3 = Heureux, 4 = Triste, 5 = Surprise et 6 = Neutre.

Les images données sont divisées en trois ensembles différents qui sont des ensembles d'apprentissage, de validation et de test. Il y a environ 29 000 images d'entraînement, 4 000 images de validation et 4 000 images à tester.[61]

Les auteurs ont conclu après quelques expériences préliminaires que la précision de la prédiction des expressions faciales pour cet ensemble de données est d'environ 65% [56]

3.4.1.2 L'architecture de modèle « mini-xception »

Mini-xception est une architecture proposée par Octavio Arriaga, Matias Valdenegro-Toro et Paul Plöger et inspiré par l'architecture de modèle très connu appelé Xception, ce dernier est une pile linéaire de couches de convolution séparables en profondeur avec des connexions résiduelles, fonctionnant dans des modules répétés, Ce Framework a 36 couches convolutives, la ligne de base pour l'extraction d'entités, organisées en 14 modules avec des connexions résiduelles autour d'elles sauf pour le premier et le dernier. [57]

L'architecture dérivée mini-Xception avec beaucoup moins de paramètres et moins de profondeur a été développée dans le but d'obtenir le meilleur rapport précision sur nombre de paramètres.

La réduction du nombre de paramètres présente deux avantages principaux :

- Réduire la puissance de traitement nécessaire pour faire fonctionner le réseau et moins de paramètres donnent un meilleur
- Généralisation dans le cadre d'un rasoir Occam « principe de parcimonie » ou encore « principe d'économie ».

Cette architecture combine l'utilisation de modules résiduels (VGG16) et de convolutions séparables en profondeur (Xception). De plus, il élimine les dernières couches entièrement connectées afin de réduire le nombre de paramètres, grâce à l'utilisation d'une couche Global Average Pooling.[58]

Mini-Xception, comprend quatre modules de convolution séparables en profondeur résiduels. Chaque convolution est suivie d'une opération de normalisation par lots et d'une fonction d'activation ReLU. Enfin, une opération de pooling moyenne globale suivie d'une fonction d'activation softmax est implémentée pour sortir une prédiction. La figure 3.3 représente l'architecture de mini-xception.

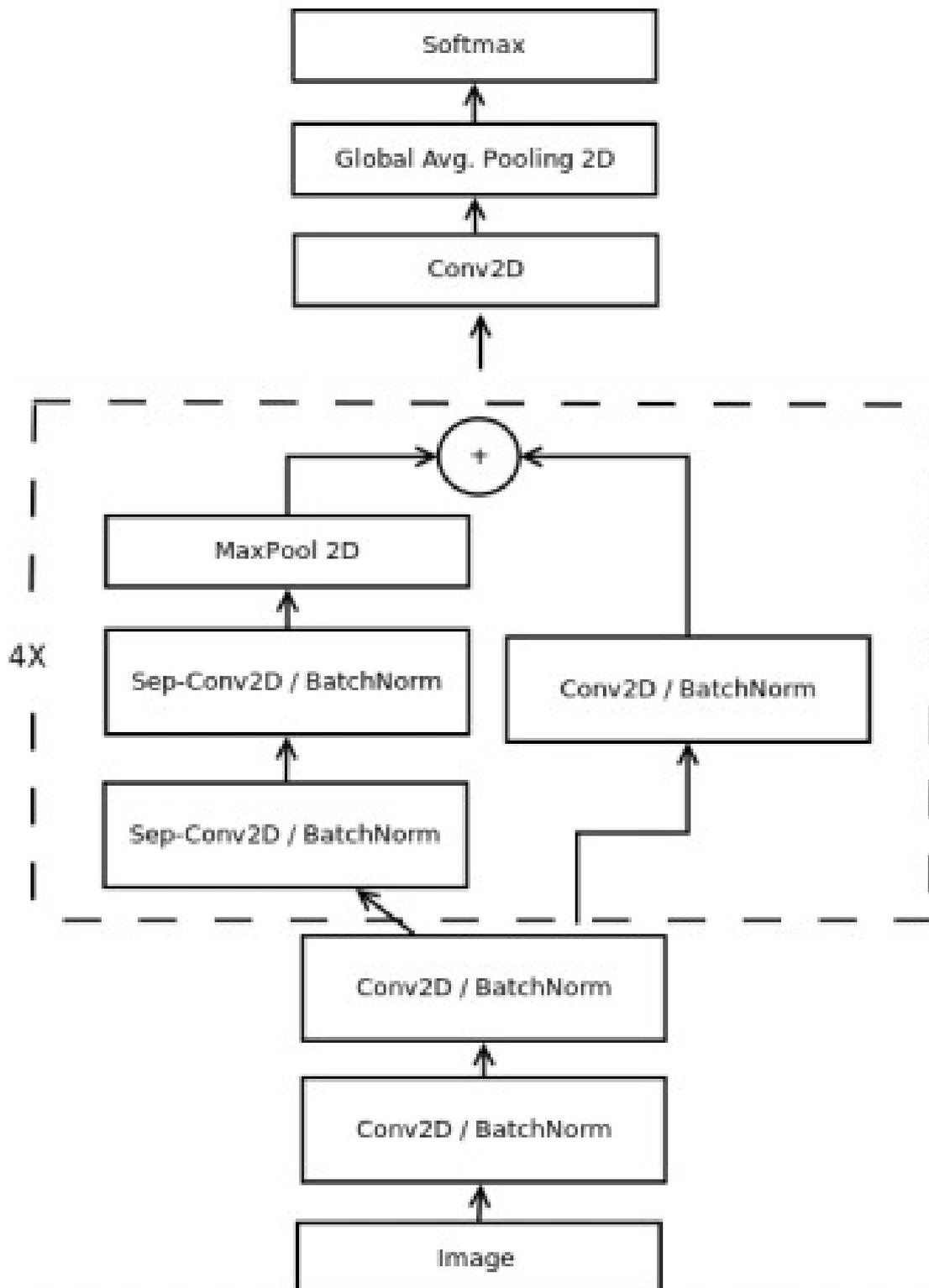


FIGURE 3.3 – l'architecture mini-xception [58]

et en temps réel des objets dans une image, cette méthode est l'une des méthodes la plus connue et la plus utilisée [59].

Pour la détection des visages, des cascades de Haar ont été utilisées dans les méthodes de Viola et Jones qui sont connues sous le nom de détecteurs Viola Jones, sont des classificateurs qui détectent un objet dans une image ou une vidéo.

Les caractéristiques de Haar détectent trois zones sombres sur le visage, par exemple les sourcils. L'ordinateur est entraîné pour détecter deux régions sombres sur le visage et leur emplacement à l'aide d'un calcul rapide des pixels.

Les cascades de Haar ont réussi à supprimer les données d'arrière-plan non requises de l'image et à détecter la région du visage à partir d'une image. [60]

3.4.2.2 Le prétraitement

Les stratégies de prétraitement utilisées sont la conversion de l'image en niveaux de gris, la normalisation et le redimensionnement de l'image [60].

- Normalisation : La normalisation d'une image est effectuée pour supprimer les variations d'éclairage et obtenir une image de visage améliorée.
- Conversion en niveau de gris : la conversion en niveau de gris est le processus de conversion d'une entrée d'image colorée en une image dont la valeur de pixel dépend de l'intensité de la lumière sur l'image. cette conversation est effectuée car les images colorées sont difficiles à processus par un algorithme.
- Redimensionnement : L'image est redimensionnée pour supprimer les parties inutiles de l'image. Cela réduit la mémoire requise et augmente la vitesse de calcul.

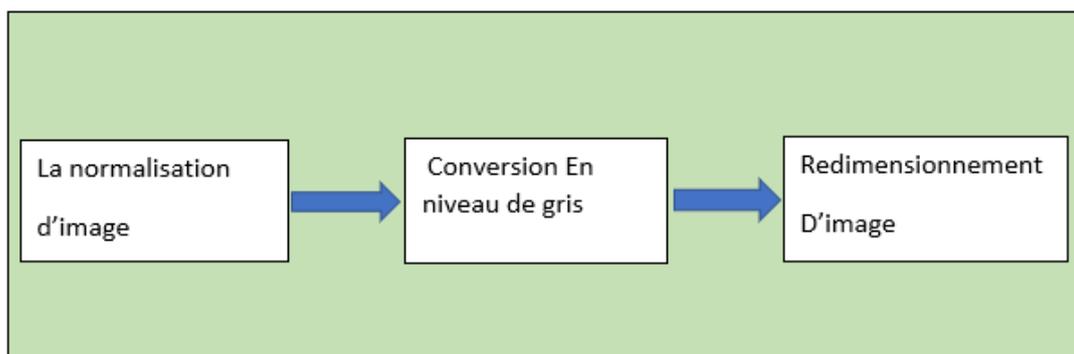


FIGURE 3.5 – le processus de prétraitement

3.4.2.3 la classification et reconnaissance des expressions faciales

Dans cette étape, le système classe l'image dans l'une des sept expressions universelles - le bonheur, Tristesse, colère, surprise, dégoût, peur et neutralité comme indiqué dans l'ensemble de données FER2013, l'entraînement a été fait à l'aide d'un modèle CNN afin de créer un classificateur prêt de prédire les expressions faciales après détecter le visage et faire des prétraitements sur ses images. La classificateur « mini-xception » a appris ses connaissances par la méthode appelée **Transfert Learning** (apprentissage par transfert).[60]

Le concept d'apprentissage par transfert peut être utilisé pour détecter l'émotion dans des images capturées en temps réel.

Ce classificateur se compose de poids et de valeurs pré-entraînés, qui peuvent être utilisés pour implémenter un nouveau problème de détection des expressions faciales, Comme le modèle généré contient déjà des poids, FER devient plus rapide pour les images en temps réel.

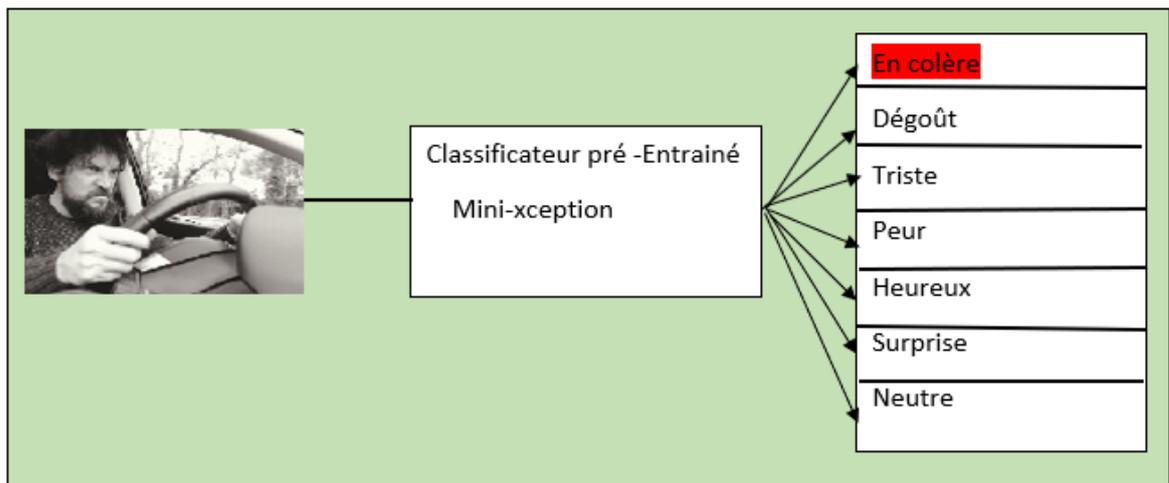


FIGURE 3.6 – Processus de la classification

3.5 Conclusion

Dans ce chapitre, nous avons introduit la conception suivie dans notre programme ainsi que l'architecture de modèle adoptée pour la classification des expressions faciales afin de détecter la colère du conducteur, Les détails de l'implémentation et les résultats obtenus seront présentés dans le chapitre suivant.

Chapitre 4

Implémentation et résultats

Expérimentaux

Sommaire

4.1	Introduction	45
4.2	Présentations d'outils	46
4.3	Implémentation et résultats de notre Système	49
4.4	la performance de notre système	58
4.5	Conclusion	59

4.1 Introduction

Dans ce chapitre, nous allons décrire la mise en œuvre des différentes étapes de détections de la colère et les autres expressions faciales universelles, c'est pour cela, nous allons commencer au premier temps la présentation des outils utilisés dans notre programme (software et hardware) ensuite nous allons voir notre implémentation en expliquant le principe des étapes suivis pour détecter la colère en temps réel.

La dernière partie concerne les résultats obtenus et quelques discussions sur les performances de notre programme.

4.2 Présentations d'outils

4.2.1 Le hardware

La configuration du matériel utilisé dans notre implémentation est comme ce qui suit :

Processeur (CPU)	Intel(R) Core (TM) i5-5200U CPU @ 2.20GHz 2.19 GHz
Mémoire (RAM)	8 Go
Carte graphique	NVIDIA GEFORCE 920 M
Type de système	Windows 10, 64bits
Ecran	15 "

4.2.2 Le software

4.2.2.1 Le langage utilisé

Python :

Python est un langage de script de haut niveau, structuré et open source. Il est multiparadigme et multi-usage. Développé à l'origine par Guido van Rossum en 1989, il est, comme la plupart des applications et outils open source, maintenu par une équipe de développeurs un peu partout dans le monde. Conçu pour être orienté objet, il n'en dispose pas moins d'outils permettant de se livrer à la programmation fonctionnelle ou impérative ; c'est d'ailleurs une des raisons qui lui vaut son appellation de « langage agile », parmi les autres raisons, citons la rapidité de développement (qualité propre aux langages interprétés), la grande quantité de modules fournis dans la distribution de base ainsi que le nombre d'interfaces disponibles avec des bibliothèques écrites en C, C++ ou Fortran. Il est également apprécié pour la clarté de sa syntaxe, ce qui l'oppose au langage Perl.

Ce langage de programmation présente de nombreuses caractéristiques intéressantes tel que :

- Python est portable, non seulement sur les différentes variantes d'Unix, mais aussi sur les OS propriétaires : MacOS, BeOS, NeXTStep, MS-DOS et les différentes variantes de Windows. Un nouveau compilateur, baptisé JPython, est écrit en Java et génère du byte code Java.
- Python est gratuit, mais on peut l'utiliser sans restriction dans des projets commerciaux.
- Python convient aussi bien à des scripts d'une dizaine de lignes qu'à des projets complexes de plusieurs dizaines de milliers de lignes.
- Python gère ses ressources (mémoire, descripteurs de fichiers...) sans intervention du programmeur, par un mécanisme de comptage de références (proche, mais différent, d'un ramasse-miettes).

- Python intègre, comme Java ou les versions récentes de C++, un système d'exceptions, qui permettent de simplifier considérablement la gestion des erreurs. [web 17]



FIGURE 4.1 – le logo python

4.2.2.2 Les Bibliothèques Utilisés

OpenCV :

OpenCV (Open Source Computer Vision Library) est une bibliothèque de logiciels open source de vision par ordinateur et d'apprentissage automatique. OpenCV a été conçu pour fournir une infrastructure commune pour les applications de vision par ordinateur et pour accélérer l'utilisation de la perception de la machine dans les produits commerciaux. Étant un produit sous licence BSD, OpenCV permet aux entreprises d'utiliser et de modifier facilement le code. [web18]

Theano :

Theano est une bibliothèque Python qui vous permet de définir, d'optimiser et d'évaluer efficacement des expressions mathématiques impliquant des tableaux multidimensionnels.

Caractéristiques de Theano :

- Utilisation transparente d'un GPU : Effectuez des calculs gourmands en données beaucoup plus rapidement que sur un processeur.
- Génération de code C dynamique : Évaluez les expressions plus rapidement.
- Optimisations de vitesse et de stabilité : Obtenez la bonne réponse pour le journal $(1 + x)$ même lorsque x est vraiment minuscule.
- Différenciation symbolique efficace - Theano fait vos dérivées pour les fonctions avec une ou plusieurs entrées. [web 19]

TensorFlow :

TensorFlow est un Framework de programmation pour le calcul numérique qui a été rendu

Open Source par Google en Novembre 2015. Depuis son release, TensorFlow n'a cessé de gagner en popularité, pour devenir très rapidement l'un des Framework les plus utilisés pour le Deep Learning et donc les réseaux de neurones. Son nom est notamment inspiré du fait que les opérations courantes sur des réseaux de neurones sont principalement faites via des tables de données multi-dimensionnelles, appelées Tenseurs (Tensor). Un Tensor à deux dimensions est l'équivalent d'une matrice. Aujourd'hui, les principaux produits de Google sont basés sur TensorFlow : Gmail, Google Photos, Reconnaissance de voix [44].



FIGURE 4.2 – le logo de tensorflow [web20]

Keras :

Keras est une API de réseaux de neurones de haut niveau, écrite en Python et capable de fonctionner sur TensorFlow ou Theano. Il a été développé en mettant l'accent sur l'expérimentation rapide. Être capable d'aller de l'idée à un résultat avec le moins de délai possible est la clé pour faire de bonnes recherches. Il a été développé dans le cadre de l'effort de recherche du projet ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System), et son principal auteur et mainteneur est François Chollet, un ingénieur Google. En 2017, l'équipe TensorFlow de Google a décidé de soutenir Keras dans la bibliothèque principale de TensorFlow. Chollet a expliqué que Keras a été conçue comme une interface plutôt que comme un cadre d'apprentissage end to end. Il présente un ensemble d'abstractions de niveau supérieur et plus intuitif qui facilitent la configuration des réseaux neuronaux indépendamment de la bibliothèque informatique de backend. Microsoft travaille également à ajouter un backend CNTK à Keras aussi [44].

Scikit-learn :

Scikit-learn est une bibliothèque libre Python dédiée à l'apprentissage automatique. Elle est développée par de nombreux contributeurs notamment dans le monde académique par des instituts français d'enseignement supérieur et de recherche comme Inria et Télécom ParisTech. Elle comprend notamment des fonctions pour estimer des forêts aléatoires, des régressions logistiques, des algorithmes de classification, et les machines à vecteurs de support. Elle est conçue pour s'harmoniser avec des autres bibliothèques libre Python, notamment NumPy et SciPy [44]

Numpy :

Numpy est la bibliothèque de base pour le calcul scientifique en Python. Il fournit un objet de tableau multidimensionnel hautes performances et des outils pour travailler avec

ces tableaux, Cette bibliothèque est hautement optimisée pour les opérations numériques avec une syntaxe proche de celle MATLAB. [web 21]

Le principe fondamental de NumPy est l'apport de tableaux multi-dimensionnels. Aussi, NumPy peut être vu comme la base du calcul matriciel dans Python, et a été conçu dans l'optique de résoudre des problèmes mathématiques et scientifiques. Le module NumPy nous fournit un ensemble de milliers de fonctions mathématiques très utiles [web 22]

Pandas :

Pandas est une bibliothèque écrite pour le langage de programmation Python permettant la manipulation et l'analyse des données. Elle propose en particulier des structures de données et des opérations de manipulation de tableaux numériques et de séries temporelles. Pandas est un logiciel libre sous licence BSD2 [web 23]

Les fonctionnalités de pandas sont [web23] :

- L'objet DataFrame pour manipuler des données aisément et efficacement avec des index pouvant être des chaînes de caractères ;
- Des outils pour lire et écrire des données structurées en mémoire depuis et vers différents formats : fichiers CSV, fichiers textuels, fichier du tableur Microsoft Excel, base de données SQL ou le format rapide et permettant de gérer de gros volume de données nommé **HDF5** ;
- Fusion et jointure de large volume de données
- Analyse de séries temporelles

Matplotlib :

Matplotlib est une bibliothèque du langage de programmation Python destinée à tracer et visualiser des données sous formes de graphiques Elle peut être combinée avec les bibliothèques python de calcul scientifique Numpy et SciPy. Matplotlib est distribuée librement et gratuitement sous une licence de style BSD. Sa version stable actuelle (la 2.0.1 en 2017) est compatible avec la version 3 de Python. [web24]

Plusieurs points rendent cette bibliothèque intéressante :

- Export possible en de nombreux formats matriciels (PNG, JPEG...) et vectoriels (PDF, SVG...)
- Documentation en ligne en quantité, nombreux exemples disponibles sur internet
- Forte communauté très active
- Bibliothèque haut niveau : idéale pour le calcul interactif

4.3 Implémentation et résultats de notre Système

Notre application est constituée de deux phases principales

- La première de phase est d'apprentissage
- La deuxième phase est de prédiction

Dans ce qui suit, nous voyons l'implémentations et les résultats de chacun d'eux

4.3.1 implémentation de la phase d'apprentissage

Le but de cette phase est de faire l'entraînement de notre modèle choisi «mini_xception» en utilisant la base de données fer-2013, les étapes d'implémentation sont les suivants

4.3.1.1 Chargement et prétraitement de la base de données

Dans la base « fer-2013 » installée à partir de kaggle.com, il existe des images sous forme de chaînes de valeurs de niveaux de gris, chaque image est donnée sous forme de chaîne images de taille (48 × 48) stockée en tant que vecteur de ligne sous le format (.csv).

Le code ci-dessous charge la base de données et faire les traitements des images et les préparé pour les servir au modèle CNN choisi par la suite.

Le code se compose de deux fonctions :

4.3.1.1.a La fonction chargement de donnée (def load_data)

La fonction de chargement de donnée « fer-2013.csv » servent à :

- Lire le fichier de l'extension csv
- Convertir les séquences de pixels de chaque ligne en images de dimension 48*48
- Faire retourner les images qui contient des visages et des étiquettes d'émotions

```
def load_fer2013():
    data = pd.read_csv(dataset_path)
    pixels = data['pixels'].tolist()
    width, height = 48, 48
    faces = []
    for pixel_sequence in pixels:
        face = [int(pixel) for pixel in pixel_sequence.split(' ')]
        face = np.asarray(face).reshape(width, height)
        face = cv2.resize(face.astype('uint8'), image_size)
        faces.append(face.astype('float32'))
    faces = np.asarray(faces)
    faces = np.expand_dims(faces, -1)
    emotions = pd.get_dummies(data['emotion']).as_matrix()
    return faces, emotions
```

FIGURE 4.3 – la code source qui charge la base de données « fer-2013 »

4.3.1.1.b La procédure de prétraitement def preprocess_input

Cette procédure d'entrée est une procédure standard pour l'apprentissage, dans ce dernier :

- Les images sont redimensionnées à [0,1]
- Puis en le divisant par 255
- Puis soustraction par 0.5
- Et finalement en fait la multiplication par 2 modifiant le va jusqu'à [-1,1], [-1 ,1]

```
def preprocess_input(x, v2=True):
    x = x.astype('float32')
    x = x / 255.0
    if v2:
        x = x - 0.5
        x = x * 2.0
    return x
```

FIGURE 4.4 – la procédure de prétraitement d'entrée de la base de données « fer-2013 »

4.3.1.2 L'architecture de « mini-xception »

Dans cette partie de la phase nous allons implémenter la srtucture de notre modèle adopté dans un classe nommé `cnm.py` en utilisant la bibliothèque Keras, Le code de notre architecture implémenté est le suivant :

```

def mini_XCEPTION(input_shape, num_classes, l2_regularization=0.01):
    regularization = l2(l2_regularization)

    # base
    img_input = Input(input_shape)
    x = Conv2D(8, (3, 3), strides=(1, 1), kernel_regularizer=regularization,
              use_bias=False)(img_input)
    x = BatchNormalization()(x)
    x = Activation('relu')(x)
    x = Conv2D(8, (3, 3), strides=(1, 1), kernel_regularizer=regularization,
              use_bias=False)(x)
    x = BatchNormalization()(x)
    x = Activation('relu')(x)

    # module 1
    residual = Conv2D(16, (1, 1), strides=(2, 2),
                     padding='same', use_bias=False)(x)
    residual = BatchNormalization()(residual)

    x = SeparableConv2D(16, (3, 3), padding='same',
                       kernel_regularizer=regularization,
                       use_bias=False)(x)
    x = BatchNormalization()(x)
    x = Activation('relu')(x)
    x = SeparableConv2D(16, (3, 3), padding='same',
                       kernel_regularizer=regularization,
                       use_bias=False)(x)
    x = BatchNormalization()(x)
    x = Activation('relu')(x)
    x = MaxPooling2D((3, 3), strides=(2, 2), padding='same')(x)
    x = layers.add([x, residual])

    # module 2
    residual = Conv2D(32, (1, 1), strides=(2, 2),
                     padding='same', use_bias=False)(x)
    residual = BatchNormalization()(residual)

    x = SeparableConv2D(32, (3, 3), padding='same',
                       kernel_regularizer=regularization,
                       use_bias=False)(x)
    x = BatchNormalization()(x)
    x = Activation('relu')(x)
    x = SeparableConv2D(32, (3, 3), padding='same',
                       kernel_regularizer=regularization,
                       use_bias=False)(x)
    x = BatchNormalization()(x)
    x = Activation('relu')(x)
    x = MaxPooling2D((3, 3), strides=(2, 2), padding='same')(x)
    x = layers.add([x, residual])
    
```

FIGURE 4.5 – un aperçu de la code source de l’implémentation de mini-xception

4.3.1.3 La fonction d’entraînement

Après le chargement de donnée et faire implémenter l’architecture CNN, l’étape suivante consiste à faire l’apprentissage en utilisant la fonction « fit () » avec les paramètres

suivants :

- Données d'entraînement (xtrain)
- Données target (ytrain)
- Données de validation (x_test, y_test)
- Nombre d'époque

Le code est dans le suivant (voir figure 4.5)

```
model.fit_generator(data_generator.flow(x_train, y_train,  
                                     batch_size),  
                  steps_per_epoch=len(x_train) / batch_size,  
                  epochs=num_epochs, verbose=1,  
                  callbacks=callbacks,  
                  validation_data=(x_test,y_test))
```

FIGURE 4.6 – la fonction fit () de l'entraînement de classifieur « mini-xception »

4.3.2 Résultat de l'entraînement

Le nombre d'époques est le nombre de fois où le modèle parcourt les données. Plus nous avons d'époques, plus le modèle s'améliorera, jusqu'à un certain point. Puis, le modèle cessera de s'améliorer à chaque époque.

```

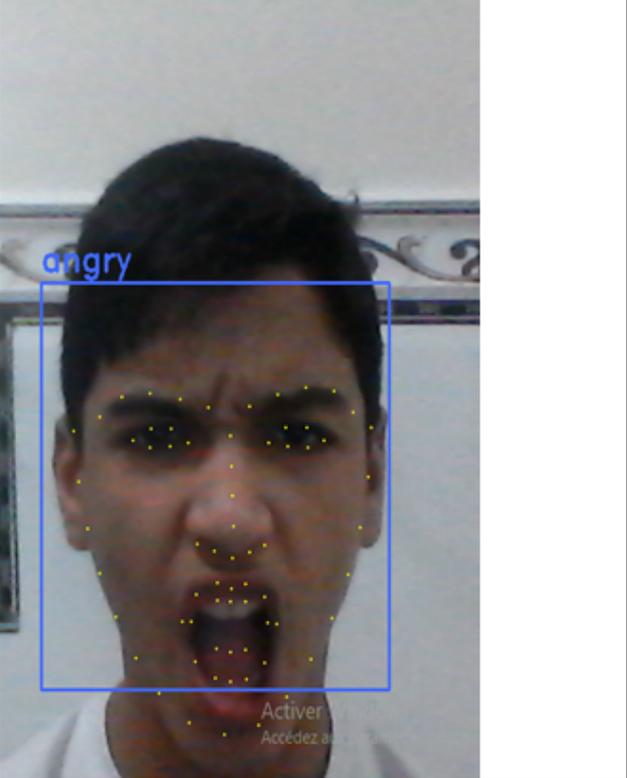
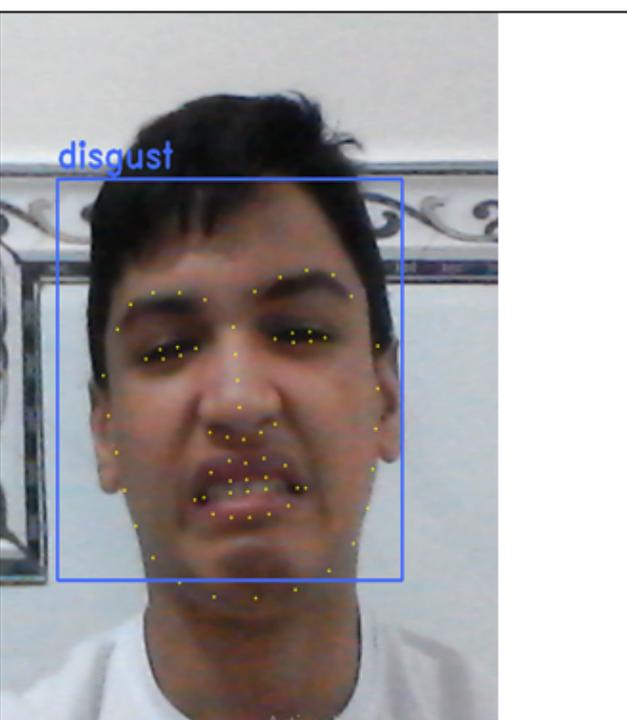
=====>.....] - ETA: 2:35 - loss: 0.9807 - acc: 0.6343
737/897 [=====>.....] - ETA: 2:34 - loss: 0.9807 - acc: 0.6341
738/897 [=====>.....] - ETA: 2:33 - loss: 0.9808 - acc: 0.6341
739/897 [=====>.....] - ETA: 2:32 - loss: 0.9806 - acc: 0.6343
740/897 [=====>.....] - ETA: 2:31 - loss: 0.9813 - acc: 0.6340
741/897 [=====>.....] - ETA: 2:30 - loss: 0.9812 - acc: 0.6340
742/897 [=====>.....] - ETA: 2:29 - loss: 0.9812 - acc: 0.6340
743/897 [=====>.....] - ETA: 2:28 - loss: 0.9811 - acc: 0.6341
744/897 [=====>.....] - ETA: 2:27 - loss: 0.9811 - acc: 0.6342
745/897 [=====>.....] - ETA: 2:26 - loss: 0.9813 - acc: 0.6343
746/897 [=====>.....] - ETA: 2:25 - loss: 0.9816 - acc: 0.6341
747/897 [=====>.....] - ETA: 2:24 - loss: 0.9816 - acc: 0.6341
748/897 [=====>.....] - ETA: 2:23 - loss: 0.9816 - acc: 0.6341
749/897 [=====>.....] - ETA: 2:22 - loss: 0.9816 - acc: 0.6341
750/897 [=====>.....] - ETA: 2:21 - loss: 0.9812 - acc: 0.6343
751/897 [=====>.....] - ETA: 2:20 - loss: 0.9813 - acc: 0.6343
752/897 [=====>.....] - ETA: 2:19 - loss: 0.9814 - acc: 0.6344
753/897 [=====>.....] - ETA: 2:18 - loss: 0.9811 - acc: 0.6344
754/897 [=====>.....] - ETA: 2:18 - loss: 0.9811 - acc: 0.6344
755/897 [=====>.....] - ETA: 2:17 - loss: 0.9810 - acc: 0.6345
756/897 [=====>.....] - ETA: 2:16 - loss: 0.9808 - acc: 0.6345
757/897 [=====>.....] - ETA: 2:15 - loss: 0.9811 - acc: 0.6345
758/897 [=====>.....] - ETA: 2:14 - loss: 0.9809 - acc: 0.6344

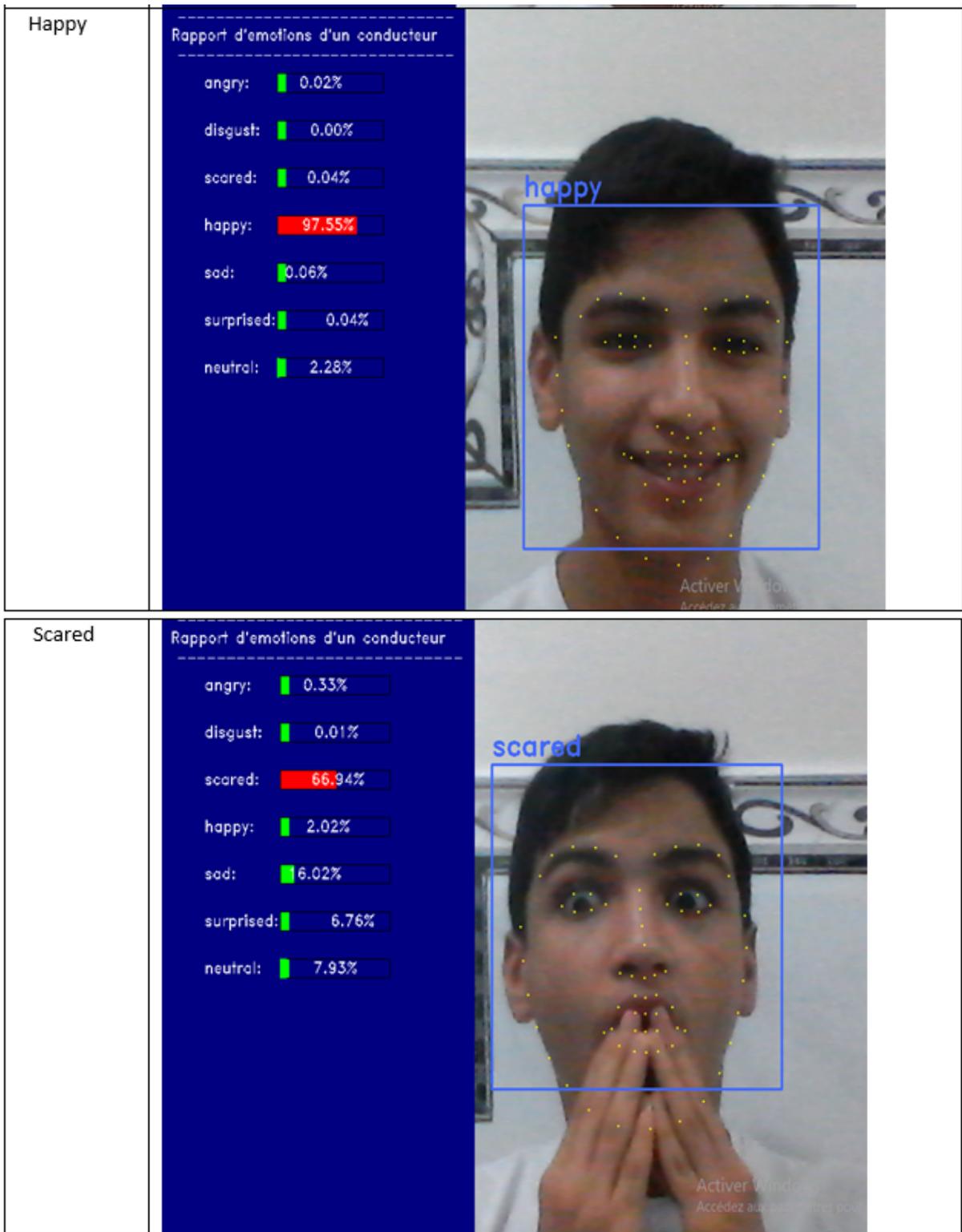
```

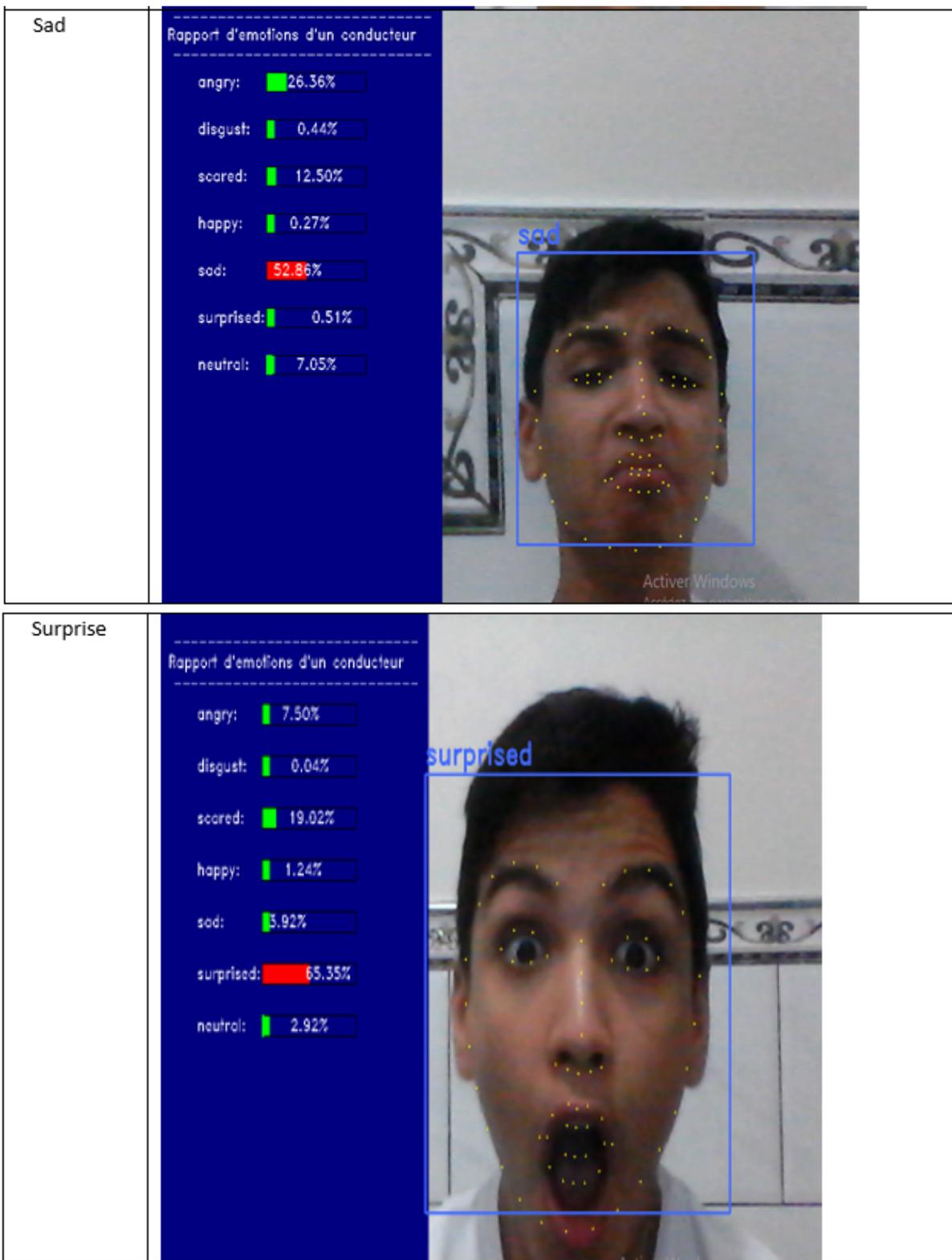
FIGURE 4.7 – un aperçu sur le résultat général de la phase apprentissage

4.3.3 La phase de prédiction

La phase de prédictions est la phase ou on peut évaluer notre classifieur afin de classer les six expressions faciales pour détecter la colère. le tableau suivant représente les résultats obtenus d'évaluation de cette classificateur (voir table 4.1)

L'expressions faciales	Les résultats obtenus	
Angry	<p>Rapport d'émotions d'un conducteur</p> <ul style="list-style-type: none"> angry: 70.98% disgust: 13.89% scared: 7.25% happy: 0.42% sad: 4.11% surprised: 0.39% neutral: 2.95% 	
Disgust	<p>Rapport d'émotions d'un conducteur</p> <ul style="list-style-type: none"> angry: 25.02% disgust: 65.84% scared: 1.37% happy: 0.69% sad: 2.97% surprised: 0.09% neutral: 4.02% 	





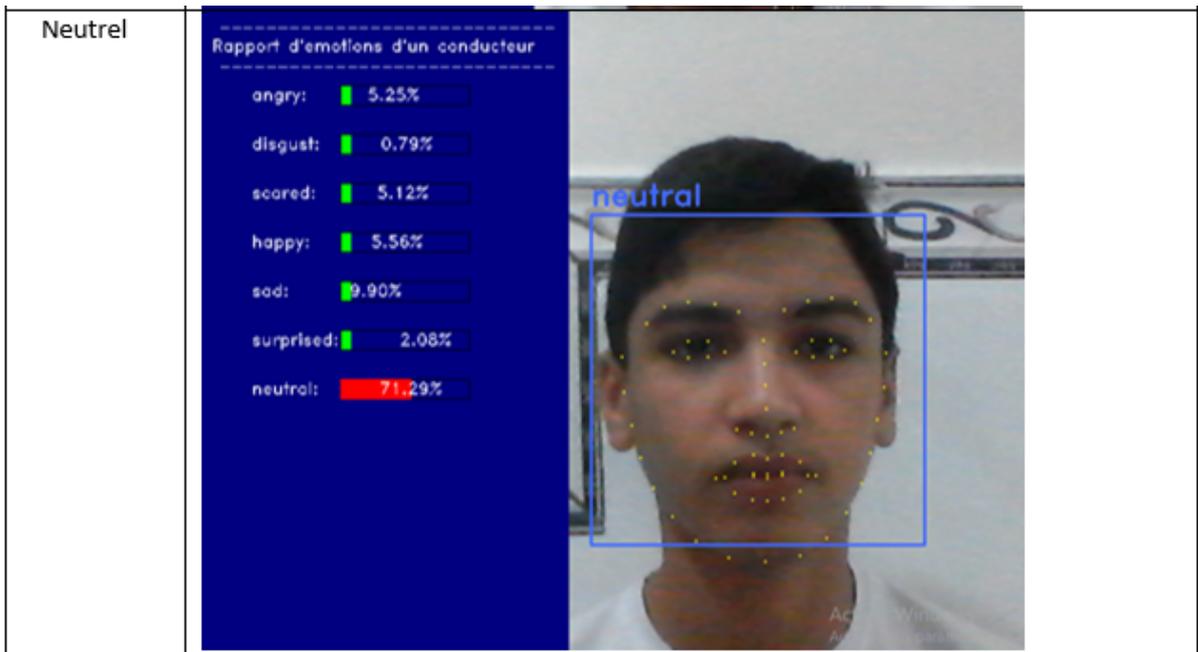


TABLE 4.1 – L’interface graphique de phase de prédiction

Et parce que nous avons fait notre étude sur l’expression faciales «la colère», nous avons ajouté un message par son (vocale) à notre application dans le cas de détection la colarité du conducteur, ce message d’alerte demande le changement si le taux de la colère est Supérieur ou égal à cinquante (50)

4.4 la performance de notre système

Notre modèle était capable de classer correctement une séquence d’image consécutives à partir de vidéos, il détectait les expressions avec une confiance de 65% dans la plupart de notre test. et Inférieur à 10% au niveau de confiance de la perte. Pour visualiser la fonction de perte et La précision, nous avons exécuté plus de 50 époques, ce qui a donné la figure suivante :

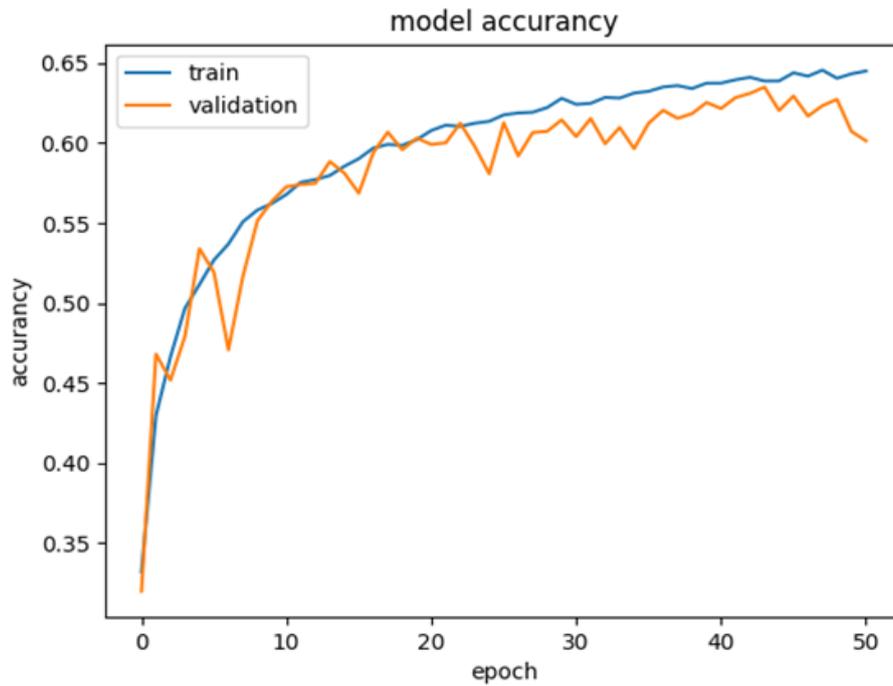


FIGURE 4.8 – Traces d’entrainement et précision de la validation

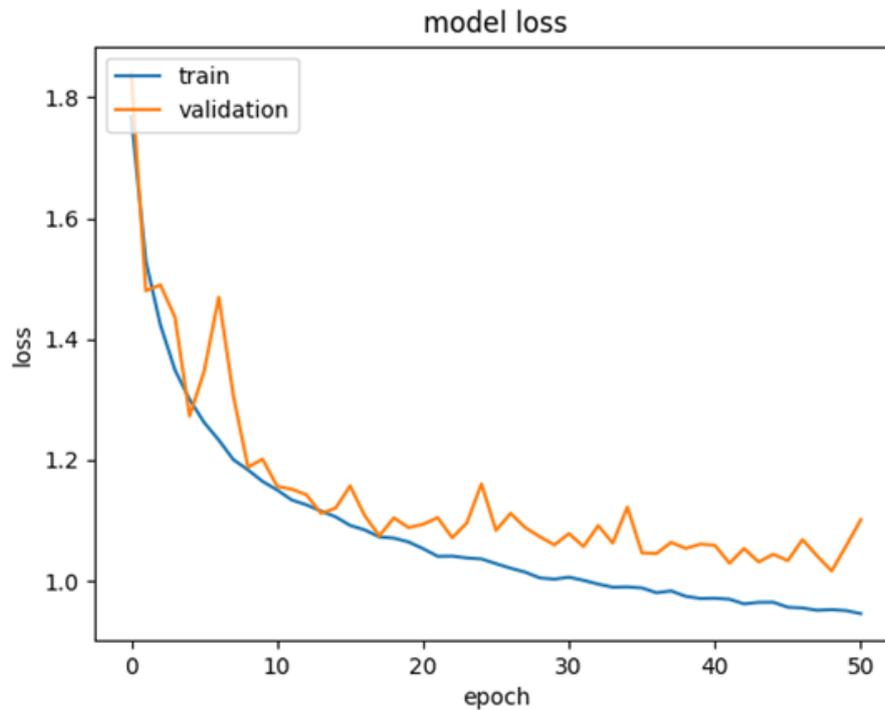


FIGURE 4.9 – traces d’entrainement de perte

4.5 Conclusion

Dans de ce chapitre, nous avons défini les différentes bibliothèques utilisées dans notre application, puis nous avons introduite les détails de notre implémentation ainsi que nous avons vu les résultats obtenir et présenté les performances de notre système de détection de la colère .

Conclusion générale et perspectives

Nous avons présenté à travers ce mémoire de fin d'étude une étude théorique ce que concerne les expressions faciales et l'apprentissage approfondie en se basant sur les réseaux de neurone convolutifs (CNN).

Nous avons par la suite utilisé un nouvel modèle de CNN approfondi nommé Mini-xception avec une base de données fer-2013 ce qui nous fournit un très bon résultat d'entraînement et un classificateur à la capacité de classifier les expressions faciales avec un gain de performance arrive à 65% de précision et 10% de perte.

L'utilisation d'un classificateur pré entraîné dans la phase de test montre la flexibilité, fiabilité et la robustesse de notre application, par conséquent, il a besoin d'utiliser du matériel très puissant pour se servir et répondre dans un temps très court.

Comme perspectives de recherches futures, nous envisageons de :

- Tester sur notre modèle d'autre base de données autre que celle de Fer2013+, Cohn-Kanade+
- Travailler avec les modèles déjà entraînés tels que le VGG16, Inception et faire une comparaison des Résultats obtenus.
- Appliquer notre modèle sur des images 3D acquises par des caméras de profondeur.
- Implémenter sur une autre plateforme pour qu'elle soit par exemple utilisée avec des systèmes embarqués.

Bibliographie

- [01] : Marie-Hélène Cigna, Jean-Pierre Guay, Patrice Renaud, La reconnaissance émotionnelle faciale : validation préliminaire de stimuli virtuels dynamiques et comparaison avec les Pictures of Facial Affect (POFA), Volume 48, Numéro 2, Automne 2015, p. 237–263, october2015
- [02] : Matthew Adam Shreve, Automatic Macro- and Micro-Facial Expression Spotting and Applications, University of South Florida, January 2013
- [03] : Franck Berthelon. Modélisation et détection des émotions à partir de données expressives et contextuelles. Interface homme-machine. Thèse de doctorat, Université Nice Sophia Antipolis, Français, , 2013.
- [04] : BARTHASSAT J., GENTAZ E. « Le développement de l'empathie au cours de l'enfance : le point de vue psychologique », Méd. Enf, 33 : 356-9, 2013
- [05] : THEUREL A., GENTAZ E., Entraîner les compétences émotionnelles à l'école, Approche neuropsychologique des apprentissages chez l'enfant (ANAE),27 (139) : 545-55; 2015
- [06] : DARWIN C., The Expression of the Emotions in Man and Animals, impression anastaltique Culture et Civilisation, 1872
- [07] : DEZECACHE G., MERCIER H., SCOTT-PHILLIPS T.C., An evolutionary approach to emotional communication » J. Pragmatics, 2013; 59, Part B : 221-33.
- [08] : EKMAN P., SORENSON E.R., FRIESEN W.V., Pan-cultural elements in facial displays of emotion, Science; 164 (3875) : 86-8n, 1969
- [09] : Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition : History, Trends, and Affect-related Applications Ciprian A. Corneanu, Marc Oliu, Jeffrey F. Cohn, and Sergio Escalera 2016.
- [10] : Padgett, C., Cottrell, G. W. Representing face images for emotion classification. In Advances in neural information processing systems (pp. 894-900), 1997.
- [11] : Lanitis, A., Taylor, C. J., Cootes, T. F. (1997). Automatic interpretation and coding of face images using flexible models. IEEE Transactions on Pattern Analysis and machine intelligence, 19(7), 743-756.

- [12] : KHADOUDJA GHANEM, Reconnaissance des Expressions Faciales à Base d'Informations Vidéo ; Estimation de l'Intensité des Expressions Faciales, thèse de doctorat, Université Mentouri de Constantine, octobre 2010
- [13] : Sonia GHARSALLI, Reconnaissance des émotions par traitement d'images, thèse de doctorat, UNIVERSITÉ D'ORLÉANS, juillet 2016
- [14] : Youcef Aichour. Détection automatique des expressions faciales à partir de séquences d'images pour l'évaluation de l'Etat des facultés d'une personne. Mémoire de master. Université du Québec à trois Rivières.2007
- [15] : Daniel LlatasSpiers, Facial emotion detection using deep Learning, Institution de l'information et technologie, UPPSALA UNIVERSITE, juin 2016
- [16] : OUAKKAF Hassiba, BELGASMI, RECONNAISSANCE AUTOMATIQUE DES EXPRESSIONS FACIALES PAR SUPPORT VECTOR MACHINE, mémoire de master, Université Larbi Ben M'hidi Oum El Bouaghi, juin2017
- [17] : J. Ahlberg, CANDIDE-3 – an updated parameterized face, Report No. LiTH-ISYR-2326, Dept. of Electrical Engineering, Linköping University, Sweden, 2001
- [18] : M. Rydfalk, CANDIDE, a parameterized face, Report No. LiTH-ISY-I-866, Dept. of Electrical Engineering, Linköping University, Sweden, 1987.
- [19] : B. Welsh, Model-Based Coding of Images, PhD dissertation, British Telecom Research Lab, Jan. 1991.
- [20] : Khadija Lekdioui. Reconnaissance d'états émotionnels par analyse visuelle du visage et apprentissage machine. Synthèse d'image et réalité virtuelle. Université Bourgogne Franche-Comté ; Université Ibn Tofail. Faculté des sciences de Kénitra, 2018.
- [21] : Catherine Soladié, Représentation invariante des expressions faciales applications en analyse multimodales des émotions, Ecole Doctorale MATISSE, thésis décembre 2013.
- [22] : Boukerrouche Youssef, Zerriouh Ahmed, Mise au point d'une application de détection et reconnaissance faciale, mémoire de master, UNIVERSITÉ ABOU BEKR BELKAID DE TLEMCEN,2018
- [23] : P. Jonathon Phillips and Hyeon Joon Moon and Syed A. Rizvi and Patrick J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms", IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 22, number 10, pages 1090-1104, October 2000
- [24] : M. Turk, A. Pentland, Eigenfaces for recognition. J. of Cognitive Neuroscience 3, pages 72– 86,1991.

-
- [25] : M. A. Turk and A. P. Pentland, “Face recognition using eigenfaces,” in Proc. IEEE Conf. Computer Vision Pattern Recognition, pages 586–591, Jun. 1991
 - [26] : S. Der Phillips J., P. J. Rauss. Feret recognition algorithm development and test results. Army Research Laboratory technical report, ARL-TR-995, 1996.
 - [27] : Livre Eigenfaces for Recognition Matthew Turk- Alex Pentland Vision and Modeling Group The Media Laboratory Massachusetts Institute of Technology Journal of cognitive neuroscience pages 71-86, 1991
 - [28] : Mina Navraan, Tarbiat Modares, Nasrollah Moghadam Charkari, Automatic Facial Emotion Recognition Method Based on Eye Region Changes, Faculty of Electrical and Computer Engineering, University, Tehran, 2016
 - [29] : Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A. : Face recognition : A literature survey. *Acm Computing Surveys* 35(4), pages : 399–458, 2003
 - [30] : Amel Aissaoui, Reconnaissance Bimodale de visage par fusion de caractéristiques visuelle et profondeur, thèse de doctorat, université des sciences et Technologie de Lille, juin 2014
 - [31] : Serge KOMANDA BASEMA, Identification des personnes par reconnaissance de visage pour la sécurité d’une institution bancaire, Institut supérieur pédagogique de Bukavu- RDC, 2010
 - [32] : Tian, Y.-I., Kanade, T., and Cohn, J. F. Recognizing action units for facial expression analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 23(2) : pages 97–115, 2001.
 - [33] : Ghimire, D., Jeong, S., Lee, J., and Park, S. H. Facial expression recognition based on local region-specific features and support vector machines. *Multimedia Tools and Applications*, 76(6) : pages 7803–7821, 2017
 - [34] : Carcagnì, P., Coco, M., Leo, M., and Distantè, C. Facial expression recognition and histograms of oriented gradients : a comprehensive study. *Springer Plus*, 4(1) : 1, 2015
 - [35] : Cohen, I., Sebe, N., Gozman, F., Cirelo, M. C., and Huang, T. S. (2003b). Learning Bayesian network classifiers for facial expression recognition both labeled and unlabeled data. In *Computer Vision and Pattern Recognition, 2003. Proceedings. IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2003
 - [36] : Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, The Extended Cohn-Kanade Dataset (CK+) : A complete dataset for action unit and emotion-specified expression Robotics Institute, Carnegie Mellon Univer-

sity, Pittsburgh, PA, 152131 Department of Psychology, University of Pittsburgh, Pittsburgh, PA, 2010

- [37] : SHABNEM ABTAHI, driver drowsiness monitoring based on yawning detection, these de doctorat, université Ottawa, Canada,2012
- [38] : G. Yang and T.S. Huang, Human Face Detection in a Complex Background, Pattern Recognition, Vol. 27, No 1, pp. 53–63, 1994
- [39] : Franck Techer. Impact de la colère sur le traitement de l'information et la performance de conduite sur simulateur. Psychologie et comportements. Université de Nantes, 2016. Français
- [40] : Ali Mollahosseini, Behzad Hasani, Michelle J. Salvador, Hojjat Abdollahi, David Chan, Mohammad H. Mahoor ; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 58-65, 2016
- [41] : ALOUACHE Rayhana, CHIA Radhia, ÉVALUATION DE LA PERFORMANCE D'UN CAPTEUR LOGICIEL EN UTILISANT L'APPRENTISSAGE EN PROFONDEUR, Mémoire de Master, UNIVERSITÉ MOHAMED BOUDIAF - M'SILA,2019
- [42] : Y. Moualek Djaloul, "Deep Learning pour la classification des images", Mémoire de master, Université de Tlemcen, Algérie, 2017.
- [43] : Daha Boubaker Medileh Mounir, Un miroir intelligent interactif basé sur Raspberry Pi, Mémoire de Master, UNIVERSITÉ ECHAHID HAMMA LAKHDAR EL OUED,2018
- [44] : Mr. Mokri Mohammed Zakaria., « Classification des images avec les réseaux de neurones convolutionnels », mémoire de master, Université Abou Bakr Belkaid Tlemcen,2017
- [45] : OULMI Mehdi - KALOUNE Salim, Classification d'objets avec le Deep Learning, mémoire de master, Université Akli Mohand Oulhadj de Bouira,2018
- [46] : Abdelaziz HABBA et Omar ISHAK, La classification des images satellitaires par l'apprentissage profonde (Deep Learning), Mémoire de master, Université Ahmed Draia – Adrar ,2019
- [47] : O'SHEA Keiron, NASH Ryan. An introduction to convolutional neural networks. 2015
- [48] : WANG Peng, XU Jiaming, XU Bo, LIU Chenglin, ZHANG Heng, WANG Fangyuan HAO Hongwei. Semantic clustering and convolutional neural networks for short text categorization. 2015

- [49] : BELABEND Meriem, Reconnaissance d'iris par les réseaux de neurones convolutionnels, Mémoire de master, université BADJI MOKHTAR,2019
- [50] : Van den Oord, A., Dieleman, S., Schrauwen, B. (2013). Deep content-based music recommendation. In Advances in neural information processing systems (pp. 2643-2651).
- [51] : Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
- [52] : A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, in Advances in neural information processing systems, pp. 1097–1105, 2012
- [53] : O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision (IJCV), vol. 115, no. 3, pp. 211–252, 2015
- [54] : Francois Chollet, Xception : Deep Learning with Depthwise Separable Convolutions, April 2017
- [55] : C. Szegedy et al., Inception : «Going Deeper with Convolutions», september2014
- [56] : Casian Miron, George porusniuc, Radu timofte, convolutional neural network architecture for facial expression , conference paper , November 2019
- [57] : Arriaga, O., Valdenegro-Toro, M., Plöger, P,Real-time Convolutional Neural Networks for Emotion and Gender Classification.2017
- [58] : Chollet, F. (2017). Xception : Deep learning with depthwise separable convolutions. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition
- [59] : Khefif Bouchra, Mise au point d'une application de reconnaissance faciale, Université Abou Bakr Belkaid – Tlemcen, Mémoire master, 2013
- [60] : Isha Talegaonkar, Kalyani Joshi, Shreya Valunj, Rucha Kohok, Anagha Kulkarni, Real Time Facial Expression Recognition using Deep Learning ,International Conference on Communication and Information Processing,2019
- [61] : Azar Fazel,Shima Alizadeh,Convolutional Neural Networks for Facial Expression Recognition,conference paper ,Apr 2017
- [web01] : <https://fr.wikipedia.org/wiki/%C3%89motion> consulté le : 20 novembre 2019

- [web02] : <https://www.definitions.net/definition/facial+expression> consulté le : 20 novembre 2019
- [web03] : <https://www.la-communication-non-verbale.com/2013/03/facial-action-coding-system-6734/> consulté le : 15 décembre 2019
- [web04] : <http://visagetechologies.com/mpeg-4-face-and-body-animation/> consulté le : 15 mars 2020
- [web05] : <https://mc.ai/deep-learning-lab-episode-3-fer2013/> consulté le 15 mars 2020
- [web06] : <https://www.cairn.info/revue-bulletin-de-psychologie-2008-2-page-115.htm> consulté le 18 mars 2020
- [web 07] : <https://www.talend.com/fr/resources/ai-vs-machine-learning-vs-deep-learning/> consulté le : 18mars2020
- [web 08] : <https://medium.com/diaryofawannapreneur/deep-learning-for-computer-vision-for-the-average-person-861661d8aa61> consulté le : 18 mars 2020
- [web09] : <https://www.mygreatlearning.com/blog/is-deep-learning-better-than-machine-learning/> consulté le : 02 avril 2020
- [web10] : <https://www.mygreatlearning.com/blog/is-deep-learning-better-than-machine-learning/> consulté le 02 avril 2020
- [web11] : <https://mc.ai/32-advantages-and-disadvantages-of-deep-learning/> consulté le 16 avril 2020
- [web 12] : https://penseeartificielle.fr/focus-reseau-neurones-convolutifs/4_Le_flattening_ou_ consulté le 20 avril 2020
- [web13] : <http://adventuresinmachinelearning.com/convolutional-neural-networks-tutorial-tensorflow> consulté le : 20 avril 2020
- [web14] : <https://towardsdatascience.com/convolutional-neural-networks-most-common-architectures-6a2b5d22479d> consulté le : 20 avril 2020
- [web17] : https://fr.wikibooks.org/wiki/Programmation_Python/Introduction consulté le : 31 aout 2020
- [web18] : <https://opencv.org/about/> consulté le : 02 septembre 2020
- [Web19] : <http://deeplearning.net/software/theano/> consulté le : 02 septembre 2020
- [web 20] : <https://www.tensorflow.org/> consulté le : 02 septembre 2020
- [web 21] : <https://cs231n.github.io/python-numpy-tutorial/numpy> consulté le : 03 septembre 2020

- [web22] : <https://code.tutsplus.com/fr/tutorials/introducing-numpy-cms-28703> consulté le : 03 septembre 2020
- [web23] : <https://fr.wikipedia.org/wiki/Pandas> consulté le : 03 septembre 2020
- [web 24] : <https://matplotlib.org/> consulté le : 04 septembre 2020