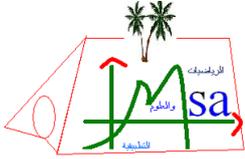
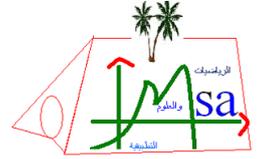


République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieure et de la Recherche Scientifique  
Université de Ghardaia



Faculté des Science et de Technologie  
Département de Mathématique et  
Informatique



&  
Laboratoire de Mathématiques et Sciences  
Appliquées

**Projet de fin d'étude présenté en vue de l'obtention du diplôme de  
Master Informatique**

**Spécialité** : Systèmes Intelligents pour l'Extraction de Connaissances (SIEC)

**THÈME**

---

**Reconnaissance des émotions faciales par  
apprentissage profond**

---

**Présenté par** : Guerradi Noura et Benkhelifa Hafida

Soutenu le : 27/06/2019 devant le jury :

**Président** : D. ZIADI Prof Université de Rouen (France)

**Examineur** : S. Bellaouar MC Université de Ghardaia

**Examineur** : C.A. Kerrache MC Université de Ghardaia

**Encadreur** : S.Oulad-Naoui MC Université de Ghardaia

Année Universitaire : 2018 - 2019

---

**Résumé** — La vision par ordinateur est un champ de recherche fascinant où de nombreux travaux ont été introduits pour des tâches de reconnaissance et de classifications. Les techniques classiques d'apprentissage se heurtent devant les volumes des données à traiter et ont de ce fait, montrés leur limites. L'apprentissage profond s'est imposé récemment comme une alternative. Ce travail utilise l'apprentissage profond et plus précisément les réseaux de neurones convolutionnels pour la classification des émotions faciales en utilisant la base de données FER2013. Nous proposons deux modèles avec une même architecture et la modification consiste à augmenter le nombre d'époques, l'ajout des méthodes de régularisation. Les résultats obtenus montrent que les modifications affectent notre premier modèle proposé et donnent des résultats acceptables d'exactitude égale à 60.52%.

**Mots clés :** Apprentissage profond, Réseaux de neurones convolutionnels, FER2013, Reconnaissance des émotions faciales.

---

---

**Abstract** — Computer vision is a fascinating field of research, and many works have been introduced for recognition and classifications tasks. Classical learning techniques collide in front of the volumes of data to be processed and thus show their limits. Deep learning has recently become an alternative. This work uses deep learning and more precisely convolutional neural networks for classifications of facial emotions using the FER2013 database, where we propose two models with the same architecture and the modification consists of increasing the number of epochs, the addition of regularization methods. The results obtained show that the changes affect our first proposed model gives acceptable results equal to 60.52% of accuracy.

**Keywords :** Deep learning, Convolutional neural networks, FER2013, Facial emotion recognition.

---

---

**ملخص -** تعد رؤية الكمبيوتر مجالًا نشطًا للبحث برزت فيه عدة أعمال على رأسها مهام التعرف والتصنيف . حيث كان سبب ظهورها هو تشابك تقنيات التعلم التقليدية مع تدفق الكم الهائل من البيانات المتوجب أو المراد معالجتها، و بذلك أصبح التعلم العميق بديلاً. يستخدم هذا العمل التعلم العميق و بشكل أدق الشبكة العصبية التلافيفية لتصنيف المشاعر الأساسية للوجه باستخدام قاعدة بيانات FER2013 ، و قد اقترحنا نموذجين لهما نفس البنية مع اختلاف يكمن في اضافة عدد الدورات و طرق التنظيم. أظهرت النتائج التي تم الحصول عليها أنها نتائج مقبولة نسبيًا مقدرة بـ 60.52% .

**الكلمات المفتاحية :** التعلم العميق، الشبكات العصبية التلافيفية، FER2013 ، التعرف على مشاعر الوجه.

---

# *Dédicaces*

Je dédie ce travail

A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études.

A mes chères sœurs pour leurs encouragements permanents, et leur soutien moral.

A mes chers frères, pour leur appui.

A toute ma famille pour leur soutien tout au long de mon parcours universitaire.

A mes meilleurs amis.

# *Dédicaces*

Je dédie ce travail

A mes chers parents qui m'ont soutenu et encouragé durant ces années d'études.

A mes frères, mes sœurs Belkis et Ferdous  
et ceux qui ont partagé avec moi tous les moments d'émotions lors de la réalisation de ce  
travail.

A mes grands parents, qui n'ont jamais cessé de formuler des prières.

A ma famille et mes proches et à ceux qui me donnent de l'amour et de la vivacité.

A tous mes amis qui m'ont toujours encouragé, et à qui je souhaite plus de succès.

# ***Remerciement***

Nous tenons à remercier sincèrement Monsieur ***Slimane Oulad-Naoui*** pour nous avoir proposé ce thème et surtout pour l'attention et pour ses précieux conseils lors de la préparation de ce mémoire. Nous remercions vivement les membres de jury d'avoir accepté l'évaluation de notre travail. Enfin, nous adressons nos gros remerciements à toutes personnes qui nous ont aidés et encouragés durant la réalisation de ce travail.

# Table des matières

Liste des tables	iv
Liste des figures	v
Liste des Abréviations	vi
Introduction générale	1
<b>1 Apprentissage automatique</b>	<b>3</b>
1.1 Introduction . . . . .	4
1.2 Historique . . . . .	4
1.3 Apprentissage classique . . . . .	4
1.3.1 Types d'apprentissage classique . . . . .	5
1.3.2 Domaines d'application de l'apprentissage automatique . . . . .	6
1.3.3 Algorithmes d'apprentissage classique . . . . .	6
1.3.4 Réseaux de neurones (RN) . . . . .	6
1.3.5 Apprentissage d'un réseau de neurones . . . . .	9
1.4 Apprentissage profond (AP) . . . . .	13
1.4.1 Architectures de réseaux de neurones . . . . .	13
1.5 Conclusion . . . . .	17
<b>2 De la reconnaissance faciale à la reconnaissance des émotions faciales</b>	<b>18</b>
2.1 Introduction . . . . .	19
2.2 Systèmes de reconnaissance faciale . . . . .	19
2.2.1 Processus de la reconnaissance faciale . . . . .	19
2.2.2 Limites des systèmes de reconnaissances faciales . . . . .	22

2.3	Analyse des émotions faciales . . . . .	23
2.3.1	Émotions et leurs paramétrages . . . . .	25
2.3.2	Approches REF conventionnelles . . . . .	28
2.3.3	Approches FER basées sur l'apprentissage profond . . . . .	30
2.4	Critères de Performance . . . . .	31
2.4.1	Matrice de confusion . . . . .	31
2.4.2	Mesures d'efficacité . . . . .	32
2.4.3	Représentation graphique des performances . . . . .	32
2.5	Base de données . . . . .	33
2.6	Conclusion . . . . .	34
<b>3</b>	<b>Reconnaissance des émotions faciales par réseaux convolutionnels</b>	<b>35</b>
3.1	Introduction . . . . .	36
3.2	Architecture du réseau . . . . .	36
3.3	Implémentation . . . . .	38
3.3.1	Environnement . . . . .	38
3.3.2	Base de données . . . . .	38
3.3.3	Résultats obtenus et discussion . . . . .	40
3.4	Conclusion . . . . .	49
	<b><i>Conclusion générale</i></b>	<b>50</b>
	<b>Bibliographie</b>	<b>51</b>

# Liste des tables

1.1	Certaines fonctions d'activation . . . . .	8
2.1	Listes des AU liées à certaines émotions. [SC01] . . . . .	27
2.2	Matrice de confusion. . . . .	32
3.1	Nombres des images correspondants au chaque émotions. . . . .	39
3.2	Le résumé des résultats. . . . .	46

# Liste des figures

1.1	Modèle semi-supervisé [Taf]. . . . .	5
1.2	Quelques algorithmes d'apprentissage automatique . . . . .	7
1.3	Schéma d'un neurone biologique(a) /neurone formel(b) [Tho]. . . . .	8
1.4	Perceptron multi couches(trois couche). . . . .	9
1.5	Apprentissage des réseaux de neurones par rétro-propagation de l'erreur . .	10
1.6	Schéma d'un taux d'apprentissage [tau]. . . . .	12
1.7	Différence entre l'apprentissage automatique et profond [img]. . . . .	13
1.8	Convolution [Stu]. . . . .	14
1.9	fonctionnement de Relu [R]. . . . .	15
1.10	Exemples de pooling [R]. . . . .	16
1.11	Aplanissement [SIM18]. . . . .	16
1.12	Architecture de RNN [De]. . . . .	17
2.1	Schéma générale d'un système de reconnaissance des visages . . . . .	19
2.2	Normalisation géométrique. [Hiz09] . . . . .	21
2.3	Normalisation photo-métrique par égalisation d'histogramme. [Bou12] . . .	21
2.4	Exemples de variation de pose. [Aba08] . . . . .	22
2.5	Exemple de port de lunettes. [Khe13] . . . . .	23
2.6	Exemple de changements des expressions faciales. [Khe13] . . . . .	23
2.7	Structure de base des systèmes de reconnaissance des émotions faciales. . .	24
2.8	Les émotions faciales de gauche a droite 1 : dégoût; 2 : peur; 3 : joie; 4 : surprise; 5 : tristesse; 6 : colère. [SC01] . . . . .	25
2.9	Liste de quelques AUs de face inférieure et leurs combinaisons. [Bet12] . .	26
2.10	Liste de certaines UAs de la face supérieure et leurs combinaisons. [Bet12]	26

2.11	Modèle du visage par le standard MPEG-4. [Gha16]	27
2.12	Extraction des caractéristiques. [TKC05]	28
2.13	Courbe de Rappel et Précision(CRP)	33
2.14	Courbe de ROC	33
3.1	Architecture du réseau utilisé.	36
3.2	Exemple des images du FER2013	39
3.3	Configuration du modèle 1	40
3.4	Sur-apprentissage de premier modèle	41
3.5	Configuration du modèle 2	42
3.6	Architecture du modèle 2	42
3.7	Sur-apprentissage de premier modèle	43
3.8	Matrice de Confusion de modèle 02	44
3.9	Nombre des images mal et bien classé par classe	44
3.10	Nombre total des images mal et bien classées	45
3.11	Prédiction correcte de neutre	46
3.12	Prédiction correcte d'heureux	47
3.13	Prédiction correcte de peur	47
3.14	Prédiction correcte de colère	47
3.15	Prédiction correcte de surprise	48
3.16	Prédiction correcte de triste	48
3.17	Prédiction fausse de la surprise	48
3.18	Prédiction fausse de la colère	49

# Liste des Abréviations

AA	Apprentissage automatique
AP	Apprentissage profond
PMC	Perceptron multi couches
RN	Réseaux de neurones
RNA	Réseau de neurone artificiel
RNC	Réseau de neurones convolutifs
RNR	Réseaux de neurones récurrents
REF	Reconnaissance des émotions faciales
FACS	Système de codage des actions faciales
AU	Unités d'action
FAPs	Paramètres d'animation faciale
SVM	Machine à vecteurs de support
ACP	Analyse en composantes principale



# Introduction générale

Au cours de ces dernières années, nous assistons à une véritable révolution scientifique et technologique. Le domaine de la vision par ordinateur en est un exemple typique. Il consiste en la création d'applications intelligentes pour comprendre et extraire des objets de multi types du base de données et des informations des images, vidéos, etc. Cette décennie a connu le développement de l'informatique affective. Il s'agit de l'introduction des émotions à l'ordinateur.

La reconnaissance des émotions faciales est un sujet de recherche actif qui est apparu depuis les années 1872 par les travaux du Darwin. A la fin des années 1970, Suwa et Motoi ont proposé pour la première fois l'analyse automatique des expressions faciales (REF) à partir d'images en traçant le mouvement de 20 points identifiés sur une séquence d'images [Dar72, TKC05]. Sachant que cette activité reste jusqu'à aujourd'hui par le développement des applications de reconnaissance émotionnelle automatique en utilisant diverses méthodes d'apprentissage, une des plus célèbres méthodes est apprentissage profond. Sa puissance réside dans la capacité de traitement d'une large base des données et la qualité de la réponse.

C'est pourquoi nous nous sommes intéressés à étudier et créer un modèle capable de reconnaître les sept émotions faciales (colère, dégoût, peur, heureux, triste, surprise, neutre) à partir des images en utilisant le réseau de neurones convolutionnel. Dans notre projet de fin d'étude, nous expérimentons sur un ordinateur doté de capacités limitées par rapport aux machines disponibles dans les centres de recherche mondiaux; par la suite, nous appliquons ce modèle sur la base d'images Fer2013.

À partir de ces considérations, nous développons trois chapitres pour traiter ce sujet.

- Chapitre 1 : nous divisons ce chapitre en deux parties : dans la première, nous présentons les concepts de base de l'apprentissage automatique et ces méthodes de classification. Puis en deuxième partie nous nous attachons sur l'apprentissage profond notamment les réseaux convolutionnels.
- Chapitre 2 : nous décrivons la reconnaissance des émotions faciales avec une vue sur l'état de l'art à travers les méthodes d'apprentissage classiques et profond.
- Chapitre 3 : nous proposons un modèle pour la reconnaissance des émotions faciales

en utilisant un réseau convolutionnels, montrons la partie expérimentale de notre travail et discutons les résultats obtenus.

N'oublions pas la section importante est la conclusion qui contient une récapitulation et les perspectives d'avenir pour ce travail.

# Chapitre 1

## Apprentissage automatique

## 1.1 Introduction

L'apprentissage est un concept perçu dans notre vie quotidienne comme l'ensemble de mécanismes menant à l'acquisition de savoir-faire. Il se manifeste principalement dès la naissance de l'être humain. Ainsi, un bébé apprend comment répondre à ses besoins en pleurant, à lier sa vision avec ses pieds pour faire des mouvements, puis plus tard comment marcher. Avec le développement technologique, le terme apprentissage n'est pas exclusif à l'homme seulement; mais, les machines aussi apprennent de leur expériences, pour faire face aux besoins de leur concepteurs. Dans ce chapitre, nous présentons dans la première partie les notions fondamentales de l'apprentissage automatique, puis nous nous focalisons dans la deuxième partie, sur l'apprentissage profond.

## 1.2 Historique

L'apprentissage automatique a passé par différentes phases durant son développement dans divers domaines.

Depuis longtemps, l'être humain tentât d'incarner une machine capable de penser. Alan Turing créa en 1950 le «Test de Turing» pour déterminer si un ordinateur dispose d'une véritable intelligence [Tur09], puis en 1952, Arthur Samuel avait une trace dans a écrit le premier programme capable d'apprendre. Ce dernier, pouvait jouer aux dames et s'améliorer avec chaque partie du jeu [Par16]. C'est la raison de l'émergence du terme « apprentissage automatique» en 1959 [Ber18].

Récemment, les réseaux de neurones ont été utilisés fréquemment sachant que les premières modélisations du neurone sont anciennes. C'est en 1943 que Mc Culloch et Pitts ont proposé les premières notions du neurone formel [GBC16].

En 1957 : Frank Rosenblatt développa le modèle du Perceptron. Il construit le premier neuro-ordinateur basé sur ce modèle et l'applique au domaine de la reconnaissance de formes [Tou92].

Proposition en 1986 de l'algorithme de rétro-propagation du gradient pour entraîner un réseau de neurones multi couches [RHW<sup>+</sup>88].

Geoffrey Hinton a forgé le terme « apprentissage profond » en 2006 pour expliquer les nouvelles architectures de réseaux de neurones profonds capables d'apprendre de bien meilleurs modèles [Pac] permettant aux ordinateurs à l'avenir de voir et distinguer les objets et le texte dans les images et les vidéo. Depuis, les travaux sur l'intelligence artificielle sont toujours en cours de développement en particulier l'apprentissage automatique.

## 1.3 Apprentissage classique

Une définition largement acceptée de l'apprentissage automatique (AA) est la suivante: "L'apprentissage automatique est le champ d'étude qui donne aux machines la faculté d'apprendre sans être explicitement programmé." Arthur Samuel, 1959

L'apprentissage automatique (Machine Learning en anglais) est une discipline de l'intelligence artificielle qui signifie que les machines possèdent les capacités de l'extraction

des connaissances à partir de données volumineuses à l'aide des algorithmes pour prendre des décisions au future.

### 1.3.1 Types d'apprentissage classique

En fonction de la qualité et quantité des informations et nos besoins, cela a provoqué la naissance de plusieurs problèmes. Différents types d'apprentissages ont été proposés. Nous présentons ici les plus connues.

#### 1.3.1.1 Apprentissage supervisé

L'apprentissage supervisé a pour but d'établir des règles de comportement à partir d'une base de données contenant des exemples de cas déjà étiquetés. La base de données est en principe un ensemble de couples entrées/sorties  $(X, Y)$ . Le but est d'apprendre à prédire pour toute nouvelle entrée  $X$  la sortie  $Y$  [DGL13]. La prédiction  $Y$  est soit une valeur continue (quantitative) où on parle de régression ou une valeur discrète (qualitative) où c'est le cas de la classification. Un exemple d'un apprentissage supervisé est la prédiction du diabète à travers certains attributs comme l'âge, l'obésité, la génétique, etc.

#### 1.3.1.2 Apprentissage non supervisé

L'apprentissage non supervisé repose sur la construction de groupes (clusters en anglais) des objets similaires à partir d'un ensemble varié d'objets [Bou15], avec la condition que chaque cluster possède le plus de similarité entre ses objets et le plus de dissimilarité aux objets des autres clusters.

Exemple: La segmentation de clients en marketing consiste à diviser l'ensemble de la clientèle en plusieurs groupes homogènes.

#### 1.3.1.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé se situe entre l'apprentissage supervisé et l'apprentissage non supervisé. En plus les données d'entrées de ce type sont constituées d'exemples étiquetés et non étiquetés qui sont représentés dans la figure 1.1. Ce qui peut être très utile quand on a deux types de données, car cela permet d'améliorer la qualité d'apprentissage [CSZ06].

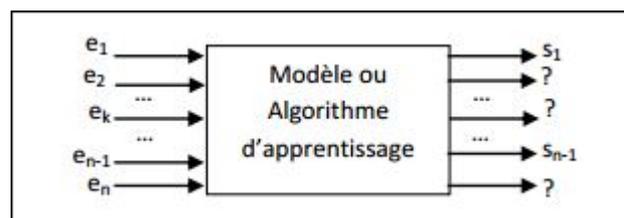


Figure 1.1 – Modèle semi-supervisé [Taf].

#### 1.3.1.4 Apprentissage par renforcement

Les données en entrée dans ce cas sont les mêmes que pour l'apprentissage supervisé, cependant l'apprentissage est guidé par l'environnement sous la forme de récompenses ou de pénalités données en fonction de l'erreur commise lors de l'apprentissage[Taf]. On retrouve ce genre d'apprentissage dans les agent autonomes.

### 1.3.2 Domaines d'application de l'apprentissage automatique

L'apprentissage automatique est utilisé dans plusieurs secteurs. Nous citons à titre d'exemple:

- La reconnaissance d'objets (visages, langages naturels, formes, ... )
- Les moteurs de recherches
- La détection de fraudes à la carte de crédit
- La bioinformatique
- L'analyser les émotions révélées par un visage photographié ou filmé
- La génération automatique des images

### 1.3.3 Algorithmes d'apprentissage classique

De nombreux algorithmes ont été proposés selon la classe de l'apprentissage et la tâche considérée. On mentionne certains sont :

- Apprentissage supervisé: régression, k plus proche voisins, arbre de décision, support vecteur machine, réseaux de neurones.etc.
- Apprentissage non supervisé: k moyenne, modèle markov caché. etc.

Afin d'alléger la présentation, nous les avons présenter dans le schéma 1.2. Celui qui voulait plus de détails il suffit de lire les références classiques sur l'apprentissage automatique [Aze18][CM10][Dre08].

### 1.3.4 Réseaux de neurones (RN)

Les méthodes d'apprentissage automatique ont contribué à résoudre plusieurs problèmes dans notre vie quotidienne mais ont échoué dans certains problèmes complexes, ce qui a inspiré certains scientifiques la simulation d'architecture et fonctionnement du cerveau humain dans la machine ce qui a suscité le concept fondamental des réseaux de neurones artificiels. Grâce à leur capacité d'apprentissage, ils sont capable de découvrir des relations complexes entre un grand nombre de variables sans intervention externe.

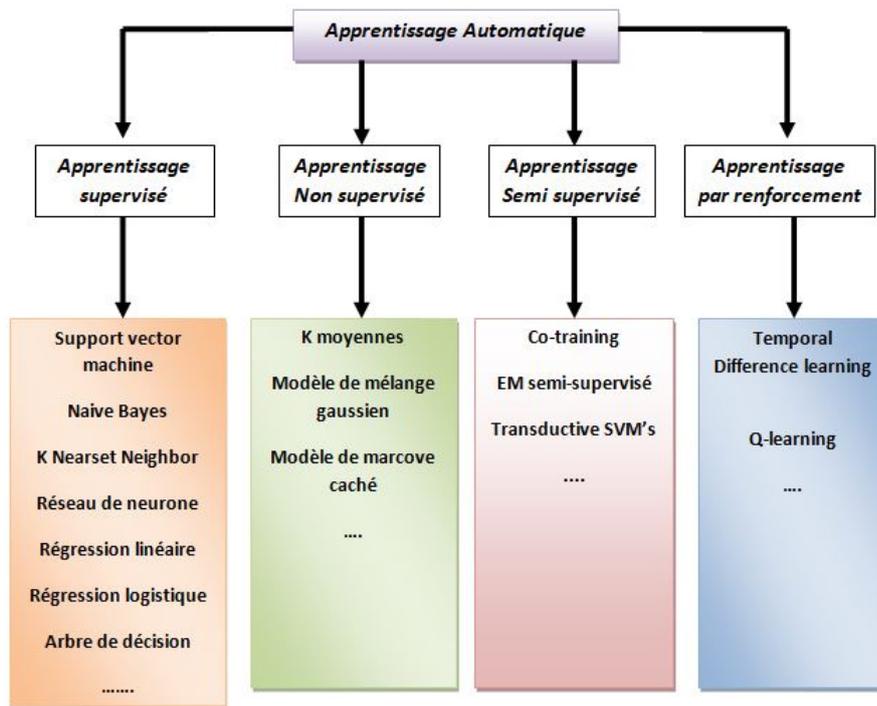


Figure 1.2 – Quelques algorithmes d’apprentissage automatique

Nous présentons dans la section suivantes les éléments de base de construction des réseaux de neurones.

### 1.3.4.1 Neurone formel

Le neurone est une cellule composée d’un corps cellulaire et d’un noyau comme la figure 1.3a présente. Le corps cellulaire se ramifie pour former les dendrites qui acheminent l’information vers les autres neurones par l’axone. La jonction entre deux neurones est appelée une synapse [Tou92].

En fait, cette modélisation nous dirigeons à la conception du neurone formel qui est l’unité élémentaire de traitement de l’information neurobiologique montrée dans la figure 1.3b. Il reçoit un nombre variable d’entrées, chacune étant multipliée par un poids associé à ces entrées qui représentent la force de la connexion. Ces dernières sont ajustées dans une phase d’apprentissage à travers l’algorithme de rétro-propagation par exemple. Ces entrées pondérées sont additionnées puis passées à travers une fonction  $f$  non linéaire appelée fonction d’activation illustrée dans le tableau 1.1 son choix est lié à la nature du problème traité pour produire une sortie par exemple dans un problème de classification binaire ( la fonction d’activation est la fonction sigmoïde ) ou de classification multi-classe ( la fonction d’activation est la fonction softmax ). Cette dernière peut être l’entrée pour un autre neurone si la sortie via  $f$  de la somme pondérée dépasse un seuil  $\theta$  le neurone est activé et transmet sa sortie. Dans le cas contraire, il est inhibé et ne transmet rien [Tou92].

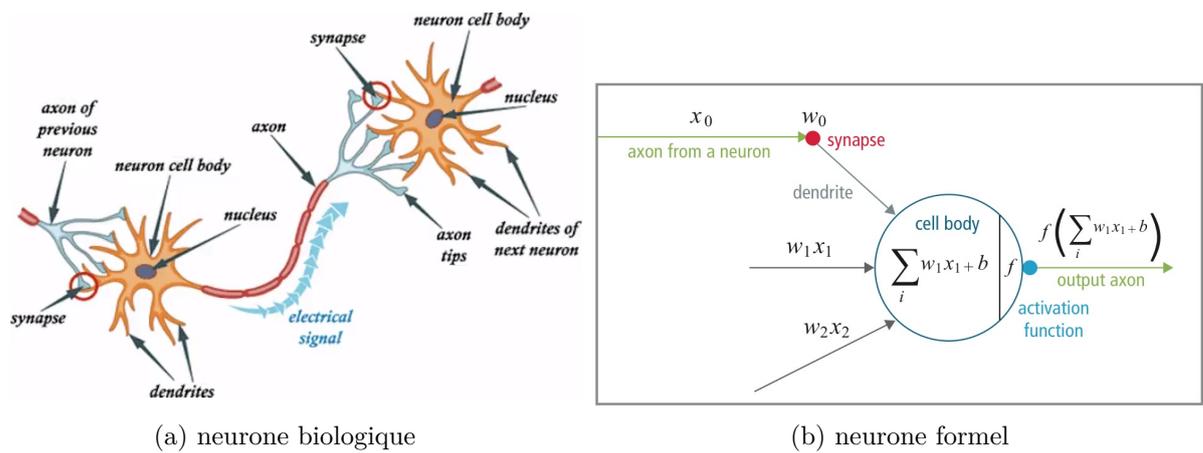


Figure 1.3 – Schéma d'un neurone biologique(a) /neurone formel(b) [Tho].

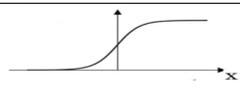
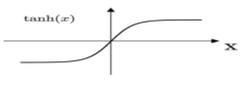
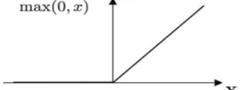
Fonction d'activation	Formule mathématique	Graphe
Sigmoïde	$f(x) = \frac{1}{1 + e^{-x}}$	
Tangente Hyperbolique	$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}$	
Unité Linéaire rectifiée	$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{sinon} \end{cases}$	

Table 1.1 – Certaines fonctions d'activation

### 1.3.4.2 Perceptron multi-couches (PMC)

Le PMC est un réseau de neurone de type feedforward ou non bouclé composé d'un ensemble de neurones formels arrangés par couche de nombres variables des neurones [Tou92]. Chaque neurone d'une couche (i) est connectée à un ensemble de neurones de la couche suivante (i+1), cela permet d'acheminer l'information de la couche entrée vers la couche de sortie, comme illustré par la figure 1.4.

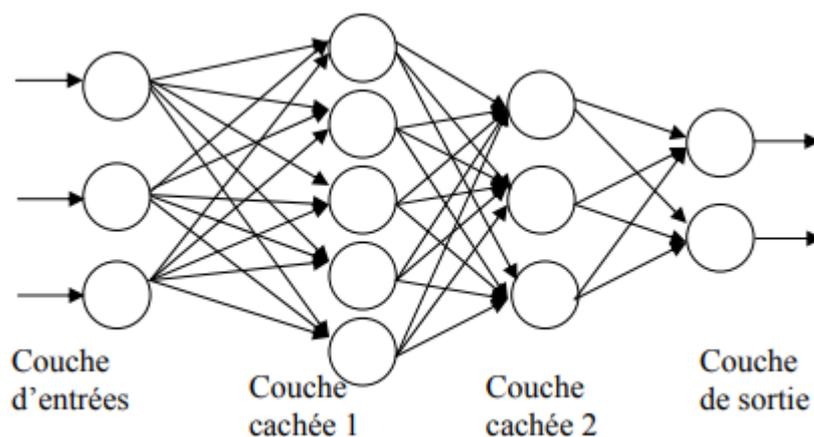


Figure 1.4 – Perceptron multi couches(trois couche).

### 1.3.5 Apprentissage d'un réseau de neurones

L'apprentissage est la phase la plus importante après la construction du réseau. Il réside dans la mise au point et le réglage du réseau pour effectuer la tâche souhaitée tel que la résolution des problème de reconnaissance et de classification. Grâce aux algorithmes d'entraînement qui permettent ensuite d'acquérir d'expérience et la capacité de la généralisation ou prendre des décisions à l'avenir. Il existe deux catégories de l'apprentissage d'un réseau de neurone.

#### 1. *Apprentissage supervisé*

L'apprentissage supervisé implique l'entraînement d'un RN avec les réponses correctes fournies à chaque exemple par un expert, en utilisant une mesure d'erreur bien définie entre la valeur désirée et la valeur obtenue par le réseau pour déterminer la valeur requise à ajuster les poids de chaque connexions afin d'obtenir de bon résultats souvent en minimisant l'erreur commise par le réseau qui est exprimé dans la figure1.5). Un des célèbres algorithmes d'entraînement des RN est l'algorithme de rétro-propagation de l'erreur [BH00].

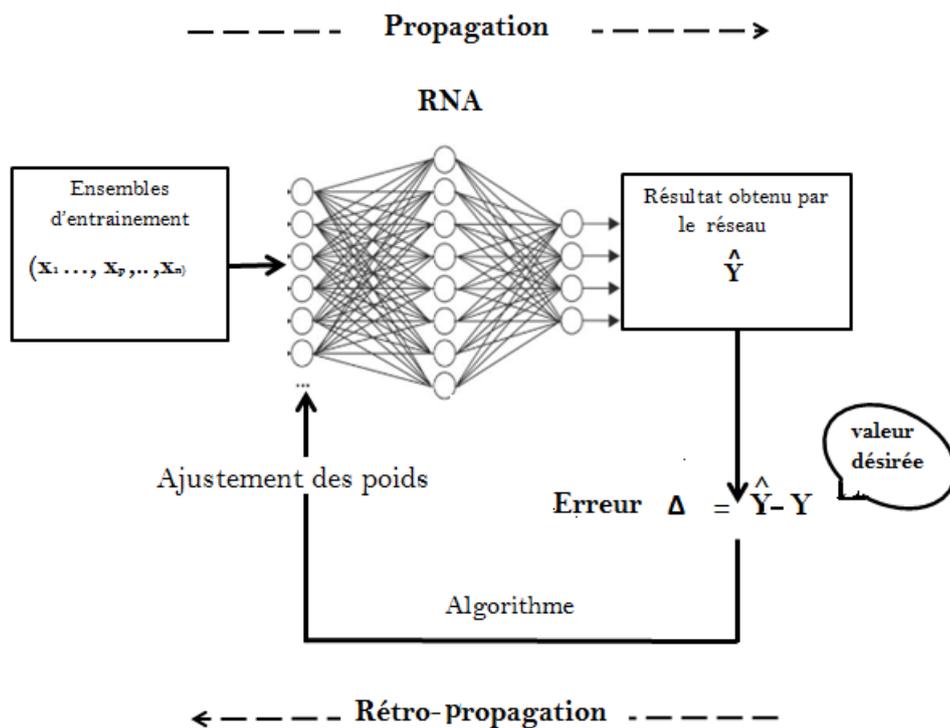


Figure 1.5 – Apprentissage des réseaux de neurones par rétro-propagation de l’erreur

## 2. *Apprentissage non supervisé*

L’apprentissage non supervisé n’exige pas de réponse correcte aux exemples d’entraînement, mais réside dans la capacité des réseaux de neurones à découvrir les caractéristiques et relations distinctives de l’ensemble d’apprentissage [BH00]. Les cartes de kohonen sont les RN utilisés dans ce type d’apprentissage.

### 1.3.5.1 Algorithme de rétro-propagation du gradient

L’algorithme d’apprentissage de rétro-propagation (en anglais backpropagation) existe depuis l’année 1986 par David E. Rumelhart, Geoffrey E. Hinton et Ronald J Williams [RHW<sup>+</sup>88]. Il est l’un des algorithmes les plus célèbres utilisées dans l’entraînement des réseaux feedforward associés au traitement des tâches de classification. Basé sur un processus itératif d’ajustement des poids synaptiques au sens inverse de la propagation des données (rétro-propagation) ceci est effectué au cours de la phase d’apprentissage en calculant l’erreur entre la valeur désirée ( $y_j$ ) et la valeur simulée ( $\hat{y}_j$ ) du réseau afin que la fonction de coût soit minimale.

On peut diviser l’algorithme de rétro-propagation en deux étapes :

#### A. *Étape de propagation*

1. Initialiser aléatoirement les poids du réseau
2. Sélectionner un exemple d’apprentissage  $(x_p, y_p)$  de la base d’apprentissage

3. Calculer par propagation la sortie simulée ( $\hat{y}_j$ ) par la formule suivant

$$\hat{y}_j = f(a_j + b) \quad / \quad a_j = \sum_i \omega_{ij} \cdot x_i \quad / \quad b \text{ est le biais} \quad (1.1)$$

4. Calculer l'erreur  $e_j$  entre la valeur désirée et la valeur simulée de sortie pour cet exemple:

$$e_j = \hat{y}_j - y_j \quad (1.2)$$

### B. Étape de rétro-propagation

5. Rétro-propager l'erreur calculée pour tous les neurones (depuis la sortie jusqu'à l'entrée) :

- Si  $j$  est un neurone de la couche de sortie alors:

$$\delta_j = e_j \cdot f'(a_j) \quad (1.3)$$

- Si  $j$  est un neurone d'une couche cachée alors:

$$\delta_j^c = f'(a_j^c) \cdot \sum_k \delta_j^{c+1} \cdot \omega_{jk}^{c+1} \quad (1.4)$$

Où  $c$  : l'indice de la couche

$\delta$  : appelé gradient local au neurone  $j$

$f'$  : est la dérivée de la fonction  $f$  continue non-linéaire

6. Mettre à jours les poids, où  $\alpha$  est le taux d'apprentissage

$$\Delta \omega_{ij}^c = \alpha \cdot \delta_j^c \cdot y_i^{c-1} \quad (1.5)$$

$$\omega_{ij}^{new} = \omega_{ij}^{old} + \Delta \omega_{ij}^c \quad (1.6)$$

7. Retour à l'étape 2 puis répéter pour chaque exemple  $p$  de l'ensemble d'entraînement jusqu'à ce que l'erreur obtenue franchit un certain seuil définis par le concepteur du réseau.[Tou92][Dec08].

- **Fonction de Coût ( Loss Function )** : fonction de coût est la norme qu'on utilise pour mesurer les performances du modèle correspondant à la nature de problème traité (classification, régression, etc). Cette fonction repose sur une métrique la plus courante est la fonction de l'erreur quadratique associée à les sorties simulées et celles désirées (réelle). Comme la formule mentionné ci-dessous:

$$J = \frac{1}{2} \sum_{j=1}^N (\hat{y}_j - y_j)^2 \quad (1.7)$$

### Facteurs influant sur l'apprentissage

- *Taux d'apprentissage*  $\alpha$  : est un paramètre ( $0 < \alpha \leq 1$ ) très important qui affecte dans la convergence d'entraînement, présente dans la règle d'apprentissage :

- Si sa valeur est très faible, l'apprentissage est lent comme on a mentionné à la figure 1.6a.
- Si sa valeur est appropriée, nous allons obtenir de façon efficace le point requis, comme indiqué à la figure 1.6b.
- Si sa valeur est grande, on risque de dépasser le point plus loin ce qui empêcherait de trouver une bonne solution comme le montre la figure 1.6c.

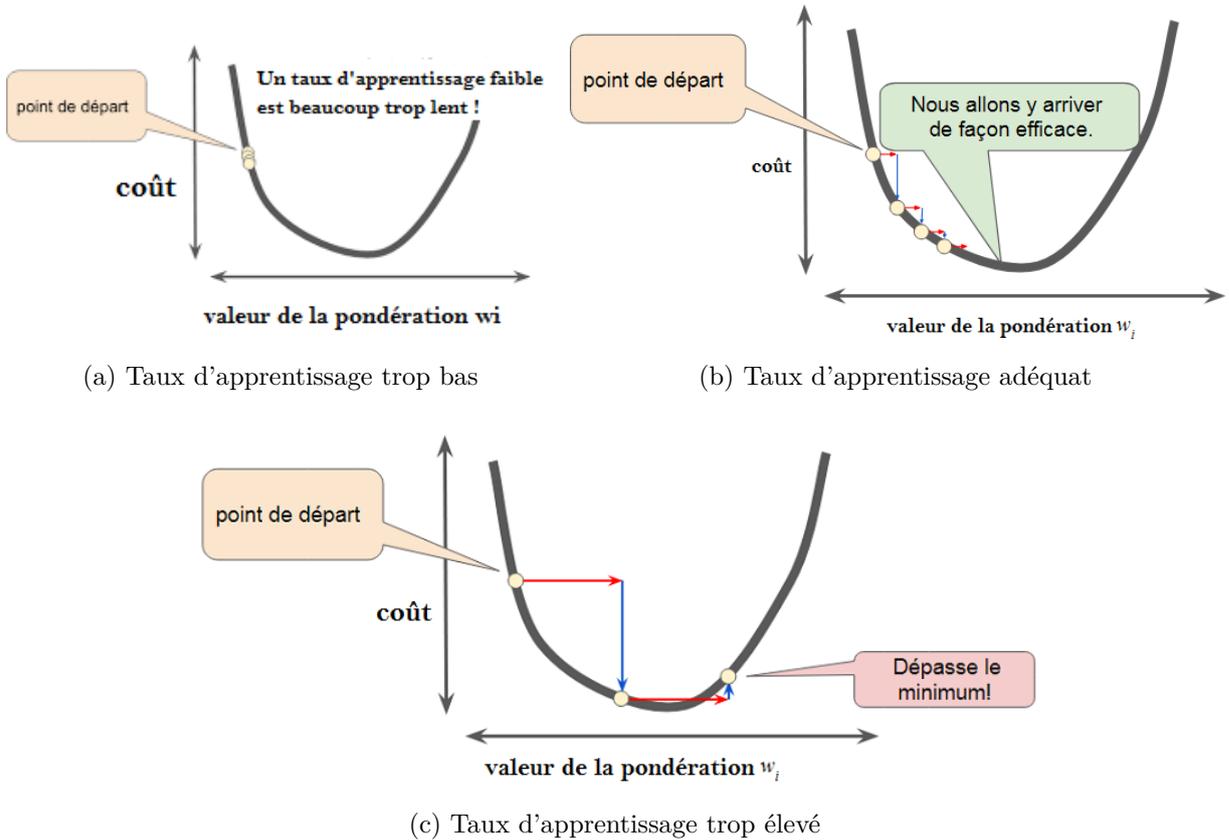


Figure 1.6 – Schéma d'un taux d'apprentissage [tau].

- *Méthode de tirage de l'ensemble d'entraînement*: il existe deux choix pour prendre ces ensembles. L'approche séquentiel (stochastique) met à jour les poids après chaque exemple; contrairement, l'approche par lot (batch) réside dans la présentation de tous les exemples du lot puis la mise à jour des poids.
- *Inertie*: (Momentum) est un terme pour éviter de tomber dans un minimum local. Il est défini par :  $\eta \cdot \Delta\omega_{ij}(n-1)$ ; où 'n' est l'indice de l'itération d'apprentissage, ( $0 < \eta < 1$ ). La formule d'apprentissage devient donc: [Dec08]

$$\Delta\omega_{ij}(n) = \alpha \cdot \delta_j^c \cdot y_i^{c-1} + \eta \cdot \Delta\omega_{ij}(n-1) \quad (1.8)$$

## 1.4 Apprentissage profond (AP)

L'AP est une branche de l'apprentissage automatique qui est apparue dans les dernières années grâce aux deux facteurs principaux suivants: la puissance de calcul des machines et la disponibilité de grandes bases de données. Chacune aidée à développer l'algorithme d'apprentissage des réseaux de neurones artificiels focalisé sur eux l'apprentissage profond pour construire un modèle fiable capable d'apprendre par lui-même. Ainsi la différence fondamentale entre les algorithmes d'AA et les algorithmes d'apprentissage profond réside dans l'étape d'extraction des caractéristiques. Dans les algorithmes d'AA l'extraction des caractéristiques est faite manuellement, c'est une étape difficile et coûteuse et exige un expert alors qu'en apprentissage profond cette étape est exécutée automatiquement par l'algorithme comme indiqué dans la figure 1.7. Cette approche est utilisée dans divers domaines tels que la classification des images, la reconnaissance d'objets, etc.

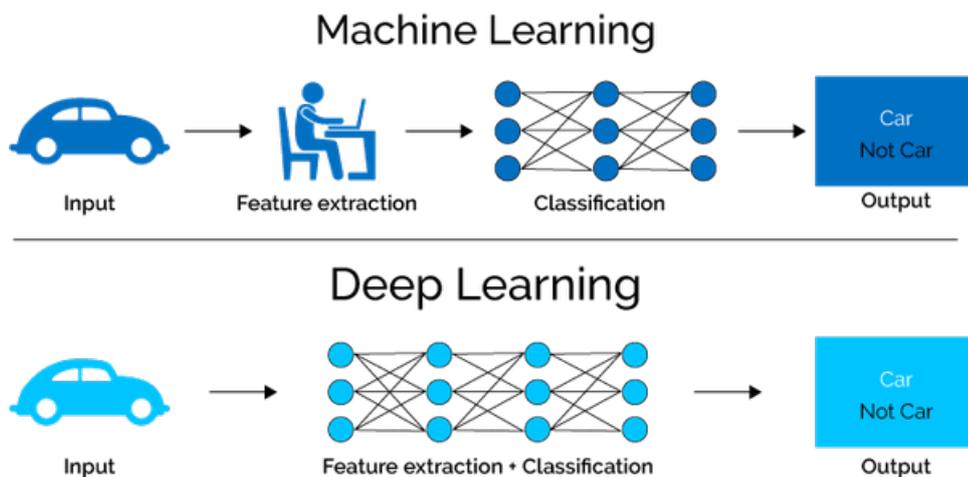


Figure 1.7 – Différence entre l'apprentissage automatique et profond [img].

### 1.4.1 Architectures de réseaux de neurones

Ils existent beaucoup d'architectures distinctes en fonction du champ d'application. Dans la suite, nous présentons les deux architectures les plus courantes.

#### 1.4.1.1 Réseau de neurones convolutionnels (RNC)

Le réseau de neurones convolutifs ou convolutionnels (Convolutional Neural Network ou CNN) existe depuis les années quatre-vingt-dix. Un RNC est un type de réseau feedforward dans lequel la connexion entre les neurones est inspirée de celle du cortex visuel [GBC16]. Il a contribué à résoudre des problèmes de vision par ordinateur tels que classifications des images, la détection des objets, etc. Ce type de réseaux est capable d'extraire des caractéristiques sans intervention d'expert pour construire un modèle fiable d'apprentissage.

Nous détaillons dans la suite, les cinq couches qui jouent un rôle important dans la construction de RNC.

## 1. Couche de convolution

La convolution est une opération mathématique très importante; elle est considérée comme la phase de pré-traitement des données. Elle est notamment plus utilisée dans les images, cette couche permet l'extraction des caractéristiques plus pertinentes en utilisant un nombre de filtres (filtre sobel, gaussian,...) afin d'obtenir qui est appelé la carte de caractéristique (feature Map) dans la figure 1.8 ci-dessous. On peut définir le résultat de convolution  $R$  (une carte caractéristique) de image  $I$  avec un filtre  $K$  par la formule suivante:

$$R(i, j) = \sum_{m=-f}^f \sum_{n=-f}^f I(m, n)K(i - m, j - n) \quad \text{où } \textit{taille} = f(\textit{impaire}) \quad (1.9)$$

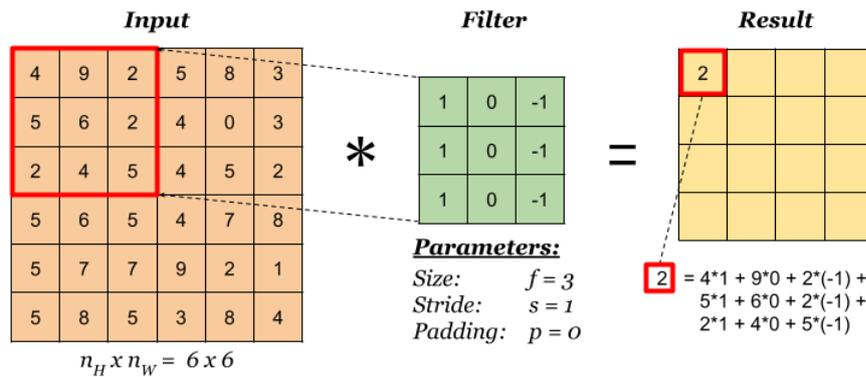


Figure 1.8 – Convolution [Stu].

Il existe trois hyperparamètres permettant de dimensionner le volume de la couche de convolution.

1. *Profondeur*: La profondeur (en anglais deep) correspond au nombre de filtres que nous utilisons pour l'opération de convolution.
2. *Pas*: (en anglais stride) contrôle le chevauchement des champs récepteurs. Plus le pas est petit, plus les champs récepteurs se chevauchent et plus le volume de la sortie sera grand.
3. *Zéro remboursement*: (en anglais zero padding) parfois il est pratique de mettre des zéros à la frontière du matrice de manière à pouvoir appliquer le filtre aux éléments frontière de la matrice d'entrée. Ce dernier permet de contrôler la taille des cartes de caractéristiques [KSH12].

Remarque: On a une image d'entrée de dimension  $W * H * C$  où  $W$  et  $H$  représentent la largeur et la hauteur de l'image et  $C$  le nombre de canaux de couleur ( 1 pour une image en niveaux de gris et 3 pour une image en couleurs), passer par  $D$  différents filtres de taille  $F * F$ . Donc, la taille de l'image après l'opération de convolution est décrit par la formule suivante:

$$W_{out} = \lfloor \frac{W - F + 2P}{S} + 1 \rfloor$$

$$H_{out} = \lfloor \frac{H - F + 2P}{S} + 1 \rfloor$$

Où  $P$  : zéro remboursement

$S$  : pas

Enfin, la taille de la carte de caractéristiques est défini par:  $W_{out} * H_{out} * D$

## 2. Couche de correction(Relu)

C'est l'abréviation d'unité de rectification linéaire de la forme mathématique suivante :

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{sinon} \end{cases}$$

Une des fonctions d'activation développées par Nair Hinton [Vie18] opérant différemment car sa transformation n'écrase pas la valeur en entrée, en plus de supprimer toutes les informations négatives de la couche précédente comme illustré dans la figure 1.9. Elle permet un entraînement plus rapide. Elle est appliquée très souvent en sortie d'une couche de convolutions plus fiable et plus efficace. En fait, les recherches ont révélé que ReLU réalise de meilleures performances que les fonctions sigmoïde et tangent hyperbolic [Vie18].

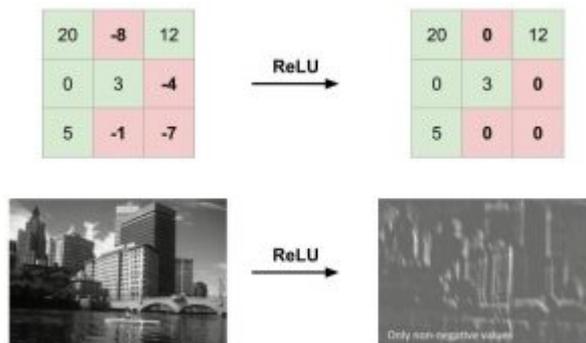


Figure 1.9 – fonctionnement de Relu [R].

## 3. Couche d'agrégation spatiale (pooling)

le pooling est une étape indispensable dans la construction de réseaux de convolution. L'agrégation spatiale aide à réduire la taille spatiale des cartes de caractéristiques. Ainsi, la nouvelle taille est calculée par le même formule de convolution la différence dans d'absence de hyperparamètre zéro remboursement :  $W_{out} * H_{out}$

$$W_{out} = \lfloor \frac{W - F}{S} + 1 \rfloor$$

$$H_{out} = \lfloor \frac{H - F}{S} + 1 \rfloor$$

Ce qui en résulte une diminution des paramètres du réseau et donc le coût de calcul, tout en gardant les principales caractéristiques de l'image. Il est judicieux la plupart

du temps d'insérer une couche de pooling entre deux couches de convolution successives d'une architecture CNN pour contrôler le sur-apprentissage (l'overfitting). Un autre intérêt des opération d'agrégation est la résistance aux transformation. Il existe plusieurs fonctions de pooling comme pooling-moyenne qui prend la valeur moyenne de toutes les valeurs recouvertes par la zone, mais la fonction la plus courante est le max-pooling qui consiste à prendre la valeur maximale voir la figure 1.10.

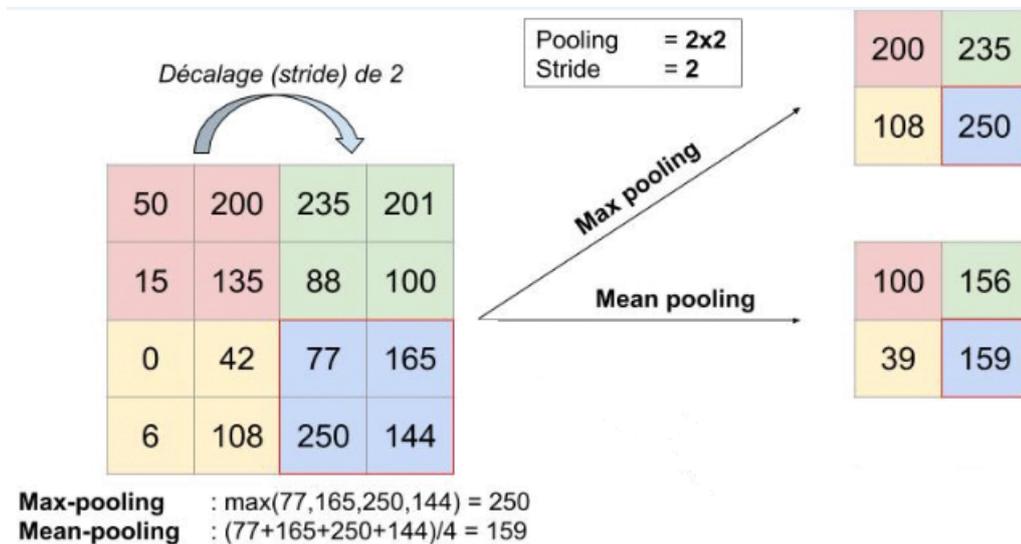


Figure 1.10 – Exemples de pooling [R].

#### 4. **Aplanissement**

Cette opération (flattening en anglais) consiste à mettre à plat toutes les données d'une couche de convolution dans un seul vecteur comme on a présenté à la figure 1.11 afin par exemple d'insérer ces données dans le réseau de neurones artificiel (considéré comme couche d'entrée de RNA), on le passer à une autre couche souvent complètement connectée.

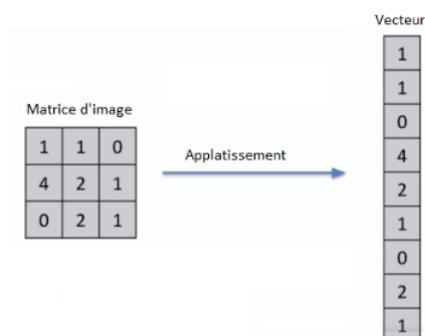


Figure 1.11 – Aplanissement [SIM18].

## 5. *Couche entièrement connectée*

Ce terme «entièrement connecté» (en anglais full connected layer) signifie que chaque neurone de la couche précédente est connecté à chaque neurone de la couche suivante pour garantir que l'information a bien circulé dans le réseau. Sachant qu'au niveau de cette couche se trouve le raisonnement de haut niveau du réseau de neurone qui contribue à construire un modèle de classification.

### 1.4.1.2 Réseau de neurones récurrents

Le réseau de neurone récurrent (RNN) est un type de réseau feedback ou bouclé qui prend en compte les données passées en vue de leur utilisation ultérieure dans leur fonctions de décision. Ils sont pour cela particulièrement adaptés à plusieurs tâches de traitement automatique des langues (TAL), notamment celles qui consistent à prédire une information séquentielle telle que traitement de séquence d'ADN plus les vidéos, les séries temporelle, etc.[DT16].

Toutefois qu'il existe d'autres structures de réseaux spécialisés dans certain tâches on peut citer par exemple le réseau long short-term memory, les réseau à port(GRU).

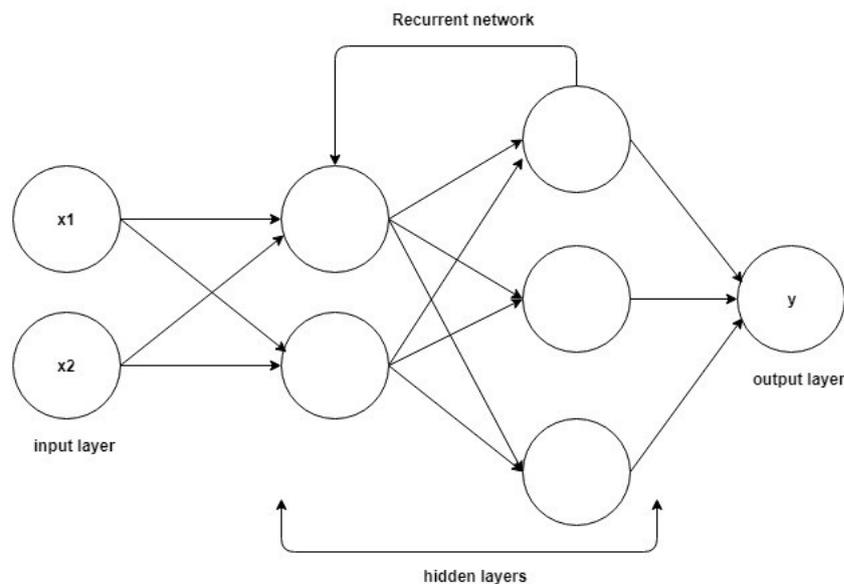


Figure 1.12 – Architecture de RNN [De].

## 1.5 Conclusion

Dans ce chapitre, nous avons survolé le sujet de l'apprentissage automatique avec une explication simple de certains concepts tels que les classes d'apprentissage (supervisé, non supervisé, ...) et les méthodes de classification par réseaux de neurones. Puis nous avons traité l'apprentissage profond et ses types en particulier les architectures convolutives. Dans le prochain chapitre, nous allons traiter le principe et les facteurs affectant la reconnaissance des émotions faciales.

## **Chapitre 2**

### **De la reconnaissance faciale à la reconnaissance des émotions faciales**

## 2.1 Introduction

Parmi les principaux canaux de communication entre personnes on trouve les émotions faciales parce qu'elles permettent d'améliorer les conversations humaines. Avec le développement de notre monde, l'automatisation est devenue une partie importante de nos vies. Où l'homme cherche à utiliser l'ordinateur dans tous les domaines, l'un de ces domaines est l'identification des personnes et la reconnaissance des émotions à partir du visage. Ces processus de reconnaissances sont faciles pour l'homme; est-il également facile sur la machine comme est sur les humains? de cela le concept de reconnaissance faciale automatique est apparu, qui accompagné des travaux importants contribue à résoudre les problèmes d'identification et de vérification des individus et plus que ça, on a passé à la reconnaissance d'émotions internes qui sont visibles sur la visage des personnes. Dans ce chapitre nous allons faire un tour sur les méthodes de reconnaissance faciale et on passons au système d'analyse des émotions faciale avec une vue sur l'état de l'art à travers les méthodes d'apprentissage classiques et l'apprentissage profond.

## 2.2 Systèmes de reconnaissance faciale

La reconnaissance faciale est un système qui repose sur la manipulation de l'image ou la vidéo pour extraire automatiquement les traits du visage, afin de les conserver en tant que marqueurs numériques dans une base de données où chaque personne dispose d'une signature numérique unique. Ces systèmes fonctionnent sur deux modes soit l'identification ou l'authentification des personnes sans aucune autre information ou aide.

### 2.2.1 Processus de la reconnaissance faciale

Les systèmes de reconnaissance faciale passe par plusieurs étapes de traitement. Les sections suivantes présentent l'essentiel de ces traitements qui sont montrés dans la figure 2.1.

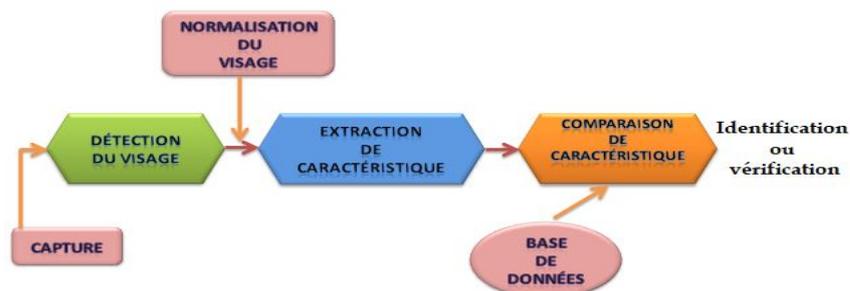


Figure 2.1 – Schéma générale d'un système de reconnaissance des visages

#### 2.2.1.1 Capture

C'est la phase préliminaire qui consiste à capturer l'image naturelle et la transformer vers une matrice dont la valeur de chaque élément représente une intensité discrète de

la lumière en cas de photo noire et blanc ou couleur cette matrice est appelée image digitale. En effet, l'image de bonne qualité donne de meilleurs résultats dans la phase de la reconnaissance. Donc il faut capturer l'information pertinente sans bruit. On trouve plusieurs types de capteurs d'image tels que: les captures classique 2D, les captures 3D, Les captures en proche infrarouge. Chaque type de capteur présence des avantages et des inconvénients. Dans la reconnaissance des visages, on peut utiliser les capteurs 3D par exemple pour s'affranchir des problèmes de pose; mais leur prix excessif ne permet pas une utilisation à grande échelle. Les captures en proche infrarouge sont utilisées pour éliminer les problèmes d'illumination [Hiz09].

### 2.2.1.2 Détection du visage

La détection de visage dans une image est la deuxième phase de traitement avant la reconnaissance. Le traitement repose sur la détection de la région du visage qui contient les yeux, la bouche et le nez. Il existe des méthodes de détection du visage on peut le catégorisées en 4 méthodes.

- **Méthodes basées sur les connaissances**

Ces méthodes sont basées sur la connaissance des principaux caractéristique du visage et les relations qui existent entre eux. Par exemple, un visage apparaît dans une image avec deux yeux dans des positions relatives entre un nez, et une bouche [CTY<sup>+</sup>03].

- **Méthodes basées sur les caractéristiques invariantes**

Ces approches se basent sur des caractéristiques structurelles (traits faciaux, texture, couleur de la peau) qui existent même quand la pose, le point de vue, ou les conditions d'illumination varient [Abd10] .

- **Méthodes basées sur la mise en correspondance de modèles**

Ces méthodes consiste à l'utilisation des modèles prédéfinis des visages ou une partie de visage (bouche, œil, nez). Une comparaison s'effectue entre chaque modèle de l'ensemble existant et l'image entrée pour identifier la présence de visage ou non dans l'image [VWMVB10].

- **Méthodes basées sur l'apprentissage**

Ces méthodes se basent sur le même principe des modèles prédéfinis des visages mais les modèles sont ici des modèles appris à partir d'un ensemble d'images d'apprentissage qui doivent permettre de caractériser la variabilité de l'apparence d'un visage. Ces méthodes présentent l'avantage de s'exécuter très rapidement mais demandent un long temps d'entraînement [CTY<sup>+</sup>03].

Ces méthodes de détection du visage sont liées aux qualités de l'image de visage extraite, donc on peut effectuer des améliorations sur l'image avant de passer aux étapes suivantes, voici quelque amélioration :

- **Normalisation géométrique** : la normalisation géométrique consiste à extraire la zone du visage de l'image originale, ensuite une rotation du visage est effectuée afin d'aligner l'axe des yeux avec l'axe horizontal. Enfin, une réduction proportionnelle

à la distance entre les centres des deux yeux est appliquée. On obtient alors une image de visage dont la distance entre les centres des yeux est fixe. Les dimensions de l'image du visage sont calculées à partir de la distance à obtenir entre les centres des deux yeux [Bou12]. La figure 2.2 illustre la méthode de la normalisation géométrique.

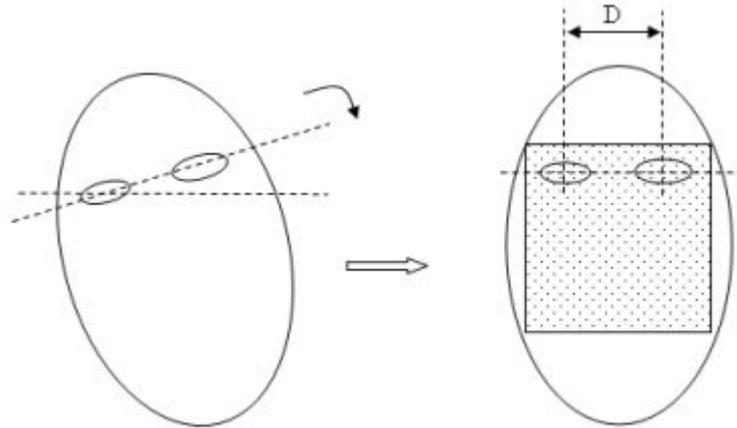


Figure 2.2 – Normalisation géométrique. [Hiz09]

- **Normalisation photo-métrique** : ce pré-traitement est nécessaire pour régler l'éclairage dans une image et minimiser l'influence de l'illumination. Cela on peut être effectué soit par des méthodes simples tell que l'égalisation d'histogrammes comme elle présente dans la figure 2.3 et correction de gamme, ou par des méthodes complexes telles que le lissage anisotropique [Hiz09].

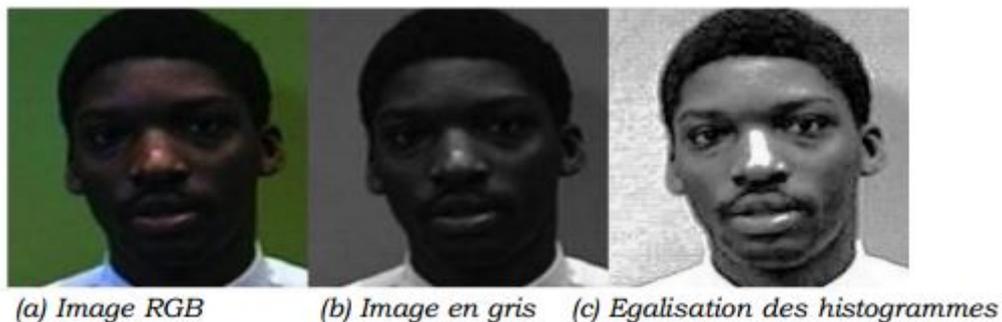


Figure 2.3 – Normalisation photo-métrique par égalisation d'histogramme. [Bou12]

### 2.2.1.3 Extraction de caractéristiques

Après la détection de la zone de visage, le système va extraire Les caractéristiques du visage à l'aide de deux pratiques différentes. La première pratique repose sur l'extraction des régions entières du visage, elle est souvent implémentée avec une approche globale de reconnaissance de visage. La deuxième pratique extrait des points particuliers des différentes régions caractéristiques du visage, tels que les coins des yeux, de la bouche et du nez. Elle est présentée comme une approche locale de reconnaissance [Aba08].

#### 2.2.1.4 Comparaison des caractéristiques

Selon les caractéristiques extraites à l'étape précédent, les algorithmes de comparaison sont différents. On trouve dans la littérature plusieurs approches telles que le calcul de distance, calcul de similarité. D'autres méthodes se basent sur la classification des caractéristiques par un classifieur (machine à vecteurs de support, classifieur bayésien, etc.) [Hiz09].

### 2.2.2 Limites des systèmes de reconnaissances faciales

Les systèmes de reconnaissance faciale se heurtent à de nombreuses difficultés qui entravent le processus de la reconnaissance parce que les visages ne sont pas des objets fixes. De plus, les environnements d'acquisition des images sont très variables. Alors, nous rappelons ci-dessous les difficultés que les systèmes doivent pouvoir gérer.

#### 2.2.2.1 Changement d'illumination

Les changements des degrés d'éclairage affectent négativement le processus de reconnaissance faciale. En effet, la clarté des détails du visage est fortement liée au changement du degré d'éclairage, donc l'illumination c'est lui qui affectant l'apparence des détails du visage.

#### 2.2.2.2 Vrais jumeaux

Le terme jumeaux désigne que les individus partagent les mêmes caractéristiques de visage (les visages sont plus identiques). Donc, la vérification automatique de visage est diminuée à cause des jumeaux, ne pourra jamais détecter les différences très subtiles qui existent entre les jumeaux.

#### 2.2.2.3 Variation de pose

Le taux de reconnaissance des visages diminue lorsque des différences entre les visages des images apparaissent. La variation de pose lors de la capture d'image est considérée comme un problème majeur pour les systèmes de reconnaissance faciale. Où la rotation ne permet pas de détecter le visage à partir de la photo quand l'angle de pose, supérieure à  $30^\circ$  [Aba08]. La figure 2.4 présente le cas de variation de pose du visage lors de la capture d'image.



Figure 2.4 – Exemples de variation de pose. [Aba08]

#### 2.2.2.4 Présence ou absence des composants structurels

Le visage peut fournir plusieurs informations sur une personne et la présence des composants structurels telle que la barbe, la moustache, ou bien les lunettes peut modifier énormément les caractéristiques faciales telles que la forme, la couleur ou la taille du visage. De plus, ces composants peuvent cacher les caractéristiques faciales de base comme la figure 2.5 illustre un exemple de présence des lunettes, causant ainsi une défaillance du système de reconnaissance.



Figure 2.5 – Exemple de port de lunettes. [Khe13]

#### 2.2.2.5 Émotions faciales

Étant donné que les visages sont des objets instables, donc il existe un autre facteur qui influe sur l'apparence du visage est les émotions. Les informations faciales causées par les émotions faciales se retrouvent principalement dans la partie inférieure du visage et les informations faciales sur la partie supérieure du visage restent presque constantes [Aba08]. Toutefois, l'émotion faciale résulte inévitablement une diminution du taux de reconnaissance, la figure 2.6 présent des facteurs (l'émotion et un composant structurel) influent sur la phase de reconnaissance faciale. Identifier le visage avec l'émotion qu'il porte est un problème difficile.



Figure 2.6 – Exemple de changements des expressions faciales. [Khe13]

La section suivante dresse un panorama des travaux sur l'analyse automatique des émotions faciales.

### 2.3 Analyse des émotions faciales

Les émotions jouent un rôle essentiel dans notre vie puisque elles permettent d'améliorer la conversation humaine. L'analyse des émotions faciales est un sujet de recherche actif pour les spécialistes du comportement depuis 1872 par les travaux de Charles Darwin

[Dar72]. Et ceux du psychologue Paul Ekman et ses collègues depuis les années 1970, où cette équipe définit en 1978 l'ensemble des émotions de base et établit qu'elles sont universelles (se présentent dans tous les cultures de la même façon) [MVJP17].

La reconnaissance des émotions faciales(REF) à partir d'images est un sujet de recherche intéressant où Suwa et Motoi ont présenté une première tentative d'analyse automatique des expressions faciales en traçant le mouvement de 20 points identifiés sur une séquence d'images en 1978 [TKC05].

Bien que la reconnaissance des émotions faciales (REF) est un sujet de recherche depuis de nombreuses années dans le domaine de la vision par ordinateur les systèmes de reconnaissance des émotions trouvés sont utilisés dans divers domaines tels que la communication, les sciences du comportement, les jeux vidéo, l'animation, la psychiatrie, etc. Ce domaine présente encore de nombreux défis liés à la complexité des émotions, les changements dans la pose du visage, les conditions d'éclairage et les variations entre les individus en termes d'attributs tels que l'âge, le sexe [JK18].

Dans cette partie, nous présentons les travaux de recherches sur la REF automatisés que nous divisons en deux groupes selon que les caractéristiques sont créées manuellement ou générées via un réseau de neurones profonds.

La figure 2.7 présente la structure de base des systèmes de reconnaissance des émotions faciales qui partagent les mêmes premières étapes des systèmes de reconnaissance faciale. Le système repose sur la détection et la localisation de la zone du visage pour passer à l'étape de l'extraction de caractéristiques du visage selon deux approches : les approches basées sur les caractéristiques géométriques ou les approches basées sur les caractéristiques d'apparence ou une approche hybride. Enfin, l'étape de classification qui utilise les méthodes d'apprentissage traditionnelles (tel que SVM, RNA) ou d'apprentissage profond tel que les réseaux convolutionnels .

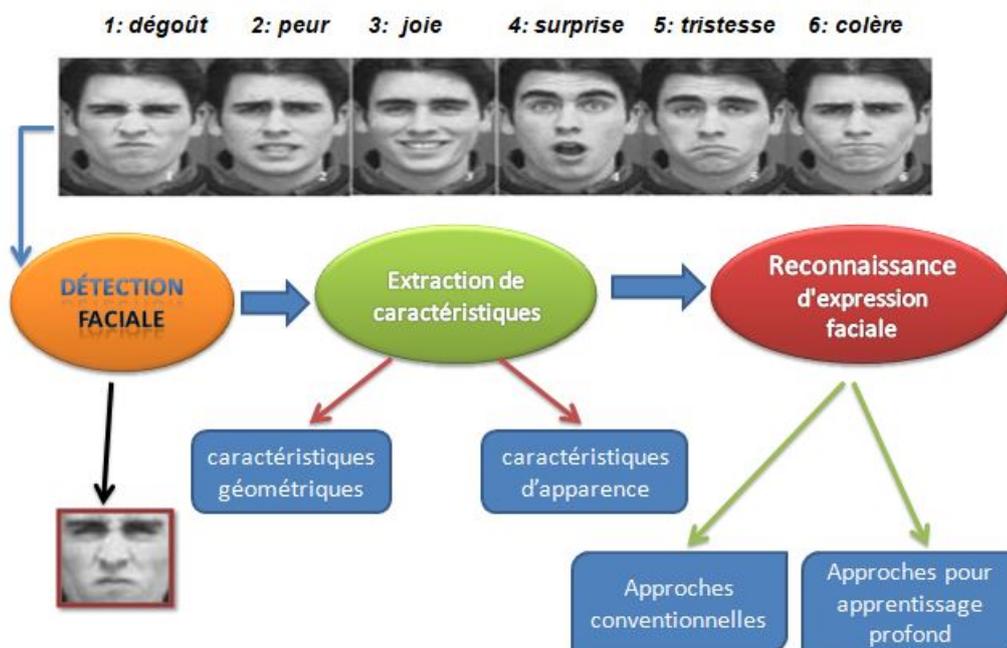


Figure 2.7 – Structure de base des systèmes de reconnaissance des émotions faciales.

### 2.3.1 Émotions et leurs paramétrages

Les émotions sont divisées en différentes catégories selon plusieurs chercheurs, dans cette section nous présentons quelque groupement des émotions.

**Les émotions de base** : sont les six émotions humaines de base universelles présente par Ekman et Friesen en 1978 il s'agit de : la joie, la surprise, la colère, le tristesse, la peur, le dégoût [MVJP17]. Celles-ci sont montrées dans la figure 2.8.



Figure 2.8 – Les émotions faciales de gauche a droite 1 : dégoût; 2 : peur; 3 : joie; 4 : surprise; 5 : tristesse; 6 : colère. [SC01]

**Les émotion secondaires** : c'est la deuxième catégorisation des émotions sont décrites par le psychologue Robert Plutchik comme résultant de combinaisons de 2 émotions primaires voisines dans le cercle des émotions, sont: amour, soumission, crainte, désappointement, remords, mépris, agressivité, optimisme [Plu84].

**Les micro expressions** : sont les expressions que les gens expriment lorsqu'ils tentent de dissimuler leurs véritables émotions, contrairement aux expressions spontanées régulières, les micro expressions ne durent que 1/25 ème de seconde. [Bet12]. Elles sont de ce fait difficile a détecter.

Les divers émotions faciales peuvent être paramétrées ou causées en fonction des actions musculaires. Donc pour la détection des émotions il faut tout d'abord détecter l'ensemble des ces paramètres. Jusqu'à présent, il existe deux systèmes importants dans la création de ces ensembles de paramètres.

- Le système de codage des actions faciales (FACs) est le système le plus utilisé qui code les différents mouvements faciaux possible présenté par Ekman et Friesen en 1978 [Ekm97]. FACs contient une liste de 46 mouvements faciaux appelés «unités d'action» présentent dans les figures 2.9 et 2.10. Chaque unité d'action (AU) correspond à un mouvement des muscles faciaux et possède un code numérique. Le FACs définit 32 actions atomiques plus les 14 descripteurs d'action tels que la direction du regard et la pose de la tête.

Par exemple, la joie est décrite par l'unité AU6 (Cheek Raiser) et AU12 (Lip Corner Puller). Dans le tableau 2.1, nous présentons les combinaisons des unités d'actions qui codent les émotions.

- Les paramètres d'animation faciale (FAPs) c'est le deuxième codage utilisé pour la détection des émotions, ce codage est défini par le standard MPEG-4 en 1998 où il utilisé principalement pour l'animation des déformation du visage [Bet12]. Les valeurs des paramètres d'animation du codage FAPs sont 84 points définis sur un visage neutre présentent dans la figure 2.11a. Pour la généralisation du modèle

NEUTRAL	AU 9	AU 10	AU 12	AU 20
				
Lips relaxed and closed.	The infraorbital triangle and center of the upper lip are pulled upwards. Nasal root wrinkling is present.	The infraorbital triangle is pushed upwards. Upper lip is raised. Causes angular bend in shape of upper lip. Nasal root wrinkle is absent.	Lip corners are pulled obliquely.	The lips and the lower portion of the nasolabial furrow are pulled back laterally. The mouth is elongated.
AU15	AU 17	AU 25	AU 26	AU 27
				
The corners of the lips are pulled down.	The chin boss is pushed upwards.	Lips are relaxed and parted.	Lips are relaxed and parted; mandible is lowered.	Mouth stretched open and the mandible pulled downwards.
AU 23+24	AU 9+17	AU9+25	AU9+17+23+24	AU10+17
				
Lips tightened, narrowed, and pressed together.				
AU 10+25	AU 10+15+17	AU 12+25	AU12+26	AU 15+17
				
AU 17+23+24	AU 20+25			
				

Figure 2.9 – Liste de quelques AUs de face inférieure et leurs combinaisons. [Bet12]

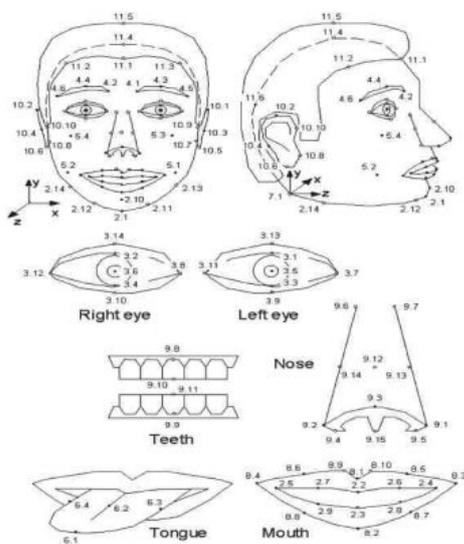
NEUTRAL	AU 1	AU 2	AU 4	AU 5
				
Eyes, brow, and cheek are relaxed.	Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together	Upper eyelids are raised.
AU 6	AU 7	AU 1+2	AU 1+4	AU 4+5
				
Cheeks are raised.	Lower eyelids are raised.	Inner and outer portions of the brows are raised.	Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.
AU 1+2+4	AU 1+2+5	AU 1+6	AU 6+7	AU 1+2+5+6+7
				
Brows are pulled together and upward.	Brows and upper eyelids are raised.	Inner portion of brows and cheeks are raised.	Lower eyelids cheeks are raised.	Brows, eyelids, and cheeks are raised.

Figure 2.10 – Liste de certaines UAs de la face supérieure et leurs combinaisons. [Bet12]

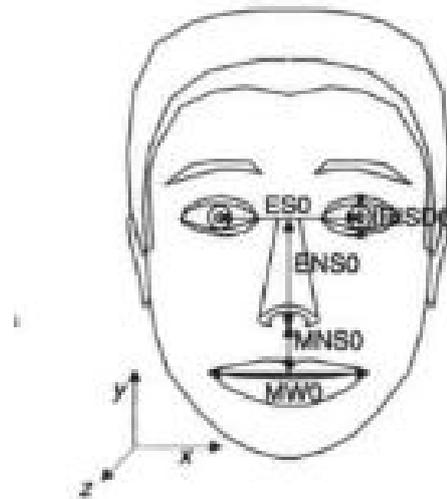
Table 2.1 – Listes des AU liées à certaines émotions. [SC01]

	UAs
FACs	face supérieure: 1, 2, 4-7, 43, 45, 46 face inférieure: 9-18, 20, 22-28; autre :21, 31, 38, 39
Émotions	joie : 6, 12, 25 surprise : 1, 2, 5, 26, 27 colère : 4, 5, 7, 10, 17, 22-26 tristesse : 1, 4, 6, 11, 15, 17 peur : 1, 2, 4, 5, 20, 25, 26, 27 dégoût : 9, 10, 16, 17, 25, 26

sur différents visages, les FAPs sont définis en fonction des unités de paramètres d’animation faciale (FAPUs : Face Animation Parameter Units) [Gha16]. Ces paramètres sont les distances entre les caractéristiques faciales clés d’un visage à l’état neutre qui sont montrés à la figure 2.11b.



(a) Les points caractéristiques FAs



(b) Modèle des paramètres unitaires FA-PU<sub>s</sub>

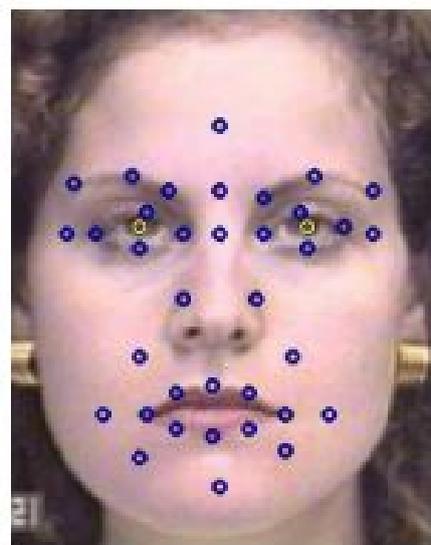
Figure 2.11 – Modèle du visage par le standard MPEG-4. [Gha16]

### 2.3.2 Approches REF conventionnelles

Pour les systèmes REF automatiques conventionnelles. Ils passent par trois étapes principales. Premièrement, une image de visage est détectée à partir d'une image d'entrée. Deuxièmement, l'extraction des caractéristiques soit l'utilisation des caractéristiques géométriques présentées dans la figure 2.12a ou d'apparence présentées dans la figure 2.12b ou un hybride des caractéristiques géométriques et d'apparence sont détectés à partir de la région de visage. Troisièmement, les classifieurs des émotions tels qu'une machine à vecteurs de support (SVM), une analyse en composantes principales(ACP), etc. sont utilisés pour produire les résultats de la reconnaissance à l'aide des caractéristiques extraites [Ko18].



(a) Caractéristiques géométriques



(b) Caractéristiques d'apparence

Figure 2.12 – Extraction des caractéristiques. [TKC05]

#### 2.3.2.1 Machine à vecteurs de support (SVM)

Ghimire et Lee ont utilisé les caractéristiques géométriques basées sur la position et l'angle de 52 points de visage sont extraites des séquences d'images d'expression faciale. Ils utilisaient Deux méthodes, la première repose sur l'utilisation d'algorithme AdaBoost multi-classe avec une précision de reconnaissance de 95.17%, la deuxième utilisent le classifieur SVM avec une précision de 97.35% les fonctionnalités renforcées est obtenue avec la base de données d'expressions faciales Extended Cohn-Kanade (CK +) [GL13].

S.L.Happy et Aurobinda Routray ont proposé une méthodologie qui utilise les caractéristiques d'apparence des points sélectionnés du visage et les points sélectionnés qui possède le plus de variance entre deux images et alors les caractéristiques sont extraites. La classification est fait par SVM avec une précision de 94.09% sur les bases de données d'expression faciale CK + et JAFFE montrent l'efficacité du système proposé pour la classification entre les six émotions de base [BTA<sup>+</sup>15].

La méthode du Guoying Zhao et al. qui utilise des séquences vidéos proche infrarouge (NIR) et des descripteurs de caractéristiques LBP-TOP (Local binary patterns from three orthogonal planes). Cette étude utilise des composants basés sur les caractéristiques du visage qui combine les informations géométriques(NIR) et d'apparence(LBP-

TOP) du visage sur une nouvelle base de données d'expression faciale Oulu-CASIA NIR VIS à l'aide du classifieur SVM qui a donné de bons résultats contre les variations d'éclairage [GZ11].

### **2.3.2.2 Analyse en composantes principales (ACP)**

S.L.Happy et al. ont utilisé l'histogramme LBP (Local Binary Patterns) de différentes tailles de blocs d'une image de visage en tant que vecteurs de caractéristiques et une classification des diverses expressions faciales à l'aide de l'analyse en composantes principales avec une précision de plus 97% mais la méthodologie est limitée à la classification de l'image frontale uniquement. Cependant, la rotation du visage ou les occlusions dégradent les performances du système [Ko18].

Yang J, Zhang DD, Frangi AF et Yang JY ont proposé une nouvelle technique qui repose sur l'analyse deux dimensionnelle en composantes principales (2DPCA) qui est basé sur des matrices d'images 2D et développée pour la représentation d'images pour éliminer la transformation du matrice d'images en un vecteur avant l'extraction de caractéristiques. L'expérience a été réalisée sur trois bases de données d'images de faces: ORL, AR et Yale. Le taux de reconnaissance dans tous les essais était plus élevé avec 2DPCA avec un pourcentage de précision 96.1% et avec PCA donne 94.7% [YZFY04].

### **2.3.2.3 Modèle de Markov caché (MMC)**

Montse Pardàs et Antonio Bonafonte ont présentés un système d'analyse vidéo pour la reconnaissance des émotions faciale en utilisant l'algorithme Active Contour pour l'extraction des caractéristiques du visage sur la base de données Cohn-Kanade. Le classifieur utilisée est le Modèle de Markov caché pour détecter l'émotion faciale de la séquence de vidéo avec une efficacité globale de 84% sur les 6 émotions de base [Bet12].

Petar S. Aleksic et Aggelos Katsaggelos ont présentés un système automatique de reconnaissance des émotions faciales. Le système utilise des paramètres d'animation faciale (FAP) comme caractéristiques de classification, par exemple les FAP décrivent le mouvement des contours et des sourcils. L'expérimentation fait sur la base de données Cohn-Kanade, le système montre que l'amélioration des performances est possible en utilisant multistream-MMC (MS-HMM) où le système donne une précision de reconnaissance des émotions quand il utilise le MMC seulement : 88,73% et quand on utilise MS-MMC qui basée sur les poids des FAPS où la lèvre externe ayant plus poids que les sourcils, et donne une précision de 93.66% [AK06].

### **2.3.2.4 Réseau de neurones artificiels (ANN)**

Le Hoang Thai, Nguyen Do Thai Nguyen et Tran Son Hai ont travaillé sur la reconnaissance des six émotions faciales de base sur la base de données JAFFE en utilisant la méthode canny pour la recherche des régions locales (sourcil gauche - droit, yeux et bouche), un réseau de neurones artificiels pour la classification des caractéristiques extraite par la méthode ACP. Le résultat expérimental de la méthode Canny-PCA-ANN donne une résultat d'accuracy estimés de 85.7% [TNH11].

Isidoros Perikos, Epaminondas Ziakopoulos et Ioannis Hatzilygeroudis ont présentés un système de détection automatique des émotions. Ce système détecte tout d'abord le visage, mesure les déformations faciales distinctives de l'homme, telles que les yeux, les sourcils et la bouche, et en extrait les caractéristiques principales. Un réseau de neurones multicouches est utilisé pour la classification. Le système a été évalué sur des images de visages humains tirées de la base de données JAFFE et Les résultats de ce système montrent une bonne performance. Plus précisément, parmi l'ensemble de test qui contient 73 images, sélectionnez correctement l'état émotionnel sur 56 images donc, la précision générale du système est de 76.7% ce qui indique de bonnes performances [PZH14].

Mais toutes ces méthodes traditionnelle et d'autres peuvent fonctionner rapidement et efficacement seulement avec une petite quantité de données. Donc, en termes de performances, restent inférieures que les méthodes d'apprentissage profond.

### 2.3.3 Approches FER basées sur l'apprentissage profond

A l'opposé des approches traditionnelles, l'approche basée sur l'apprentissage profond est devenue une approche générale de l'apprentissage par machine, donnant des meilleurs résultats dans de nombreuses études sur la vision par ordinateur tel que la REF automatiques avec la disponibilité de données volumineuses et le principal avantage des algorithmes d'apprentissage profond est de supprimer les techniques de pré-traitement d'image en permettant un apprentissage de début à la fin directement à partir d'images d'entrées, ils existent des méthodes basées sur la détection des UAs afin de classifier l'émotion en utilisant des réseaux profonds. Les architectures les plus utilisées dans cette approche sont les réseaux neurones convolutifs.

#### 2.3.3.1 Réseau neurones convolutionnel (CNN)

Heechul Jung, Sihaeng Lee Junho, Yim Sunjeong et Park Junmo Kim ont utilisé deux types différents de RNC. Le premier extrait les caractéristiques d'aspect temporel des séquences d'images (deep temporal appearance network (DTAN)), tandis que le second extrait les caractéristiques de géométrie temporelle à partir de points de repère temporels du visage (deep temporal geometry network (DTGN)). Ces deux modèles sont combinés à l'aide d'une nouvelle méthode d'intégration afin d'améliorer les performances de la reconnaissance de l'expression faciale sur les bases de données CK + et Oulu-CASIA. La méthode DTGAN donne des résultats plus précis que les méthodes traditionnelles [JLY<sup>+</sup>15].

Young-Hyen Byeon et Keun-Chang Kwak sont intéressés à la reconnaissance des émotions faciales basée sur la vidéo. Ils ont proposé un réseau 3D-CNN (3D-Convolutional Neural Networks) qui utilise des groupes de 5 images consécutives en entrée avec une taille de 64x44 sur la base de données d'expression faciale basée sur la vidéo CNU contenant 10 personnes, l'expérimentation révèle que la méthode présentée montre une bonne performance avec une précision de 95% [BK14].

Peter Burkert, Felix Trier, Muhammad Zeshan Afzal, Andreas Dengel et Marcus Liwicki ont proposé une architecture CNN pour la REF nommée FeatEx qui se compose de plusieurs couches convolutives de différentes tailles, ainsi que de Max Pooling avec une activation ReLU. L'évaluation des performances a été réalisée sur les bases Extended

Cohn-Kanade(CKP) et MMI. Ce système a montré une performance de détection des émotions une grande performance sur CKP 99.6% et 98. 63% pour MMI [BTA<sup>+</sup>15].

### 2.3.3.2 Approche hybride entre RNC et réseaux de neurones récurrents

Les approches basées sur RNC pour la reconnaissance des émotions faciales ne peuvent pas refléter les variations temporelles des composants faciaux de cela une approche hybride a été développée entre RNC pour les caractéristiques spatiales et LSTM (un type spécial de RNR) pour les caractéristiques temporelles des images consécutives.

Behzad Hasani et Mohammad H. Mahoor ont présenté des réseaux de convolution 3D qui se composent de couches 3D Inception-ResNet suivies d'une unité LSTM qui extraient l'ensemble les relations spatiales au sein des images faciales, ainsi que les relations temporelles entre différentes images de la vidéo. Les points de repère du visage sont également utilisés comme entrées dans le réseau, ce qui met l'accent sur l'importance des composants du visage à la génération l'émotion faciale. Ils ont évalué la méthode proposée sur quatre bases de données bien connues : CK +, MMI, FERA et DISFA [HMM17].

Jain Deepak Kumar, Zhang Zhang, Huang Kaiqi, Kaiqi Huang ont proposé une méthode d'apprentissage en profondeur basée sur des modèles optimaux à plusieurs angles (MAOP-DL) pour résoudre le problème des changements soudains d'éclairage. Initialement, cette approche soustrait l'arrière-plan et isole l'avant-plan des images, puis extrait les motifs de texture et les principales caractéristiques des points du visage. Les caractéristiques pertinents sont ensuite extraites de manière sélective et un LSTM-CNN est utilisé pour prédire l'étiquette requise pour l'émotion faciale. Les principales conclusions du travail proposé sont : une analyse d'image claire, une gestion efficace des variations de pose, éclairage et l'alignement du visage. Le MAOP-DL proposé valide son efficacité dans deux bases de données standard CK + et MMI [JZH17].

Mengyi Liu, Shaoxin Li, Shiguang Shan et Xilin Chen ont proposé une architecture profonde nommée AUDN (réseau inspiré par l'unité d'action) inspirés de la théorie psychologique selon laquelle les émotions peuvent être décomposées en plusieurs unités d'action faciales. L'évaluation des performances a été réalisée sur sept catégories d'émotions, y compris le neutre, avec une précision de 93,7% pour la base de données CK + et de 75,85% pour la base de données MMI [JK18].

## 2.4 Critères de Performance

l'évaluation des performances d'un système de classification des émotions faciales possède un grande importance car ces mesures de performances peuvent être utilisées pour l'apprentissage avec l'optimisation dans les hyperparamètres du classifieur. Dans cette section nous allons présentons les indicateurs de performance en classification.

### 2.4.1 Matrice de confusion

Pour mesurer les performances d'un classifieur, il existe quatre types de classes on peut le détecté sont:

- Vrai positif VP: élément de la classe C est correctement prédit.
- Vrai négatif VN: élément de la classe  $\bar{C}$  correctement prédit.
- Faux positif FP: élément de la classe  $\bar{C}$  mal prédit.
- Faux négatif FN: élément de la classe C mal prédit.

Ces informations peuvent être visualisés sous forme de tableau nommé matrice de confusion. [Taf] Le tableau 2.2 présente les classes prédites du classifieur par rapport aux classes réelles dans le cas d'un classifieur binaire .

Table 2.2 – Matrice de confusion.

	C	$\bar{C}$
Prédit C	Vrai positif (VP)	Faux positif (FP)
Prédit $\bar{C}$	Faux négatif (FN)	Vrai négatif (VN)

### 2.4.2 Mesures d'efficacité

La matrice de confusion donne la possibilité de calculer plusieurs mesures résumons leur contenue dans les métriques ci\_ après :

- **La précision** : c'est la proportion des émotions bien classés pour une classe donnée. [Taf] La précision est donnée par l'équation :

$$Précision = \frac{VP}{VP + FP}$$

- **Le rappel** : est le nombre de reconnaissances correctes de l'émotion C sur le nombre réel d'images avec émotion C. [Ko18] On le calcule par l'équation :

$$Rappel = \frac{VP}{VP + FN}$$

- **L'exactitude(accuracy)** : c'est le taux de bonne classification ou le rapport entre les résultats réels (les vrais positifs et les vrais négatifs) sur le nombre total de cas examinés [Ko18]. Donnée par l'équation :

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN}$$

### 2.4.3 Représentation graphique des performances

Il est possible de se présenter ou visualiser les performances du classifieur avec une représentation graphique par courbes. Deux courbes sont les plus utilisées dans la littérature de recherches sont CRP et ROC.

- Courbe Rappel-Précision(CRP): utilise comme abscisse les valeurs de rappel obtenues par l'algorithme de classification, alors que l'axe des ordonnées est réservé aux valeurs de performance obtenues en précision. (les valeurs sont des taux compris entre 0 à 1). [Taf] LA figure 2.13 présente un exemple de courbe de précision et rappel.
- Courbe ROC (Receiver Operator Characteristic): c'est une représentation du taux de vrais positifs en fonction du taux de faux positifs [Taf]. LA figure 2.14 présente un courbe de ROC.

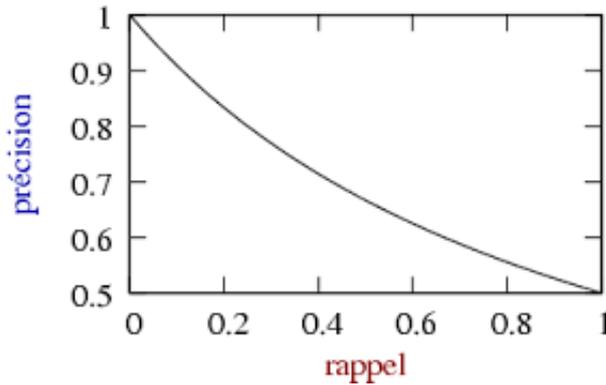


Figure 2.13 – Courbe de Rappel et Précision(CRP)

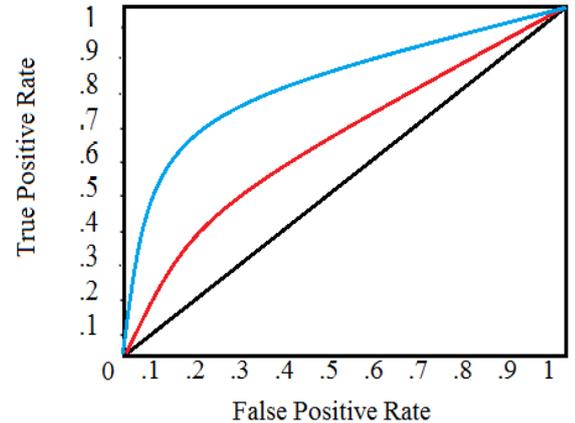


Figure 2.14 – Courbe de ROC

## 2.5 Base de données

L'étape de l'apprentissage ou l'évaluation des modèles d'apprentissage automatique consiste à utiliser une ou plusieurs base de données. Dans les méthodes de la reconnaissance des émotions faciales les bases de données utilisées doivent être générales non liées à une seule personne ou d'une région. Nous présentons ici les bases de données les plus utilisées dans la littérature.

- **Cohn-Kanade**

La base de données Cohn-Kanade (CK+) est la base la plus utilisée en FER. Cette base de données contient 593 séquences d'images de 123 sujets et des étiquettes d'émotions basées sur les FACS. Toutes les images de séquence incluent les repères de la face, le code FACS et les étiquettes d'émotion. Les émotions sont classées en huit classes : (0 = neutre, 1 = colère, 2 = mépris, 3 = dégoût, 4 = peur, 5 = heureux, 6 = tristesse, 7 = surprise). Les participants sont des adultes âgés de 18 à 50 ans et sont aussi de 81% Euro-américains, 13% afro-américains et 6% autres groupes. Les images de la base ont des résolutions en pixels de (640 × 480) et (640 × 490) [LCK<sup>+</sup>10].

- **MMI**

La base de données MMI a été conçue en 2002 par Maja Pantic, Michel Valstar et Ioannis Patras comme une ressource pour la construction et l'évaluation d'algorithmes de reconnaissance de l'émotions faciale [Gro]. Elle contient 213 séquences

d'images et plus de 2900 vidéos de 20 participants de différentes origines (américains, asiatiques, européens). La taille d'origine de chaque image faciale est de (720 × 576) [Gha16].

- **JAFFE**

C'est une base de données des expressions faciales féminines japonaises contient 213 images de sept émotions faciales (six émotions faciales de base et une neutre) posées par dix femmes différentes. La base de données a été assemblée par Michael Lyons, Miyuki Kamachi et Jiro Gyoba. La taille d'origine de chaque image faciale est de (256 x 256) [Ko18].

- **FER2013**

La base de données FER2013 a été préparé par Pierre-Luc Carrier et Aaron Courville dans le cadre d'un projet de recherche [23]. Elle Consiste en images de visage en niveaux de gris de taille (48 x 48) pixels. Elle contient 35888 images, elles sont étiquetés en 7 classes : (0 = colère, 1 = dégoût, 2 = peur, 3 = heureux, 4 = triste, 5 = surprise, 6 = neutre).

## 2.6 Conclusion

Dans ce deuxième chapitre, nous présentons d'une manière générale l'architecture des systèmes de reconnaissance des visages et nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance de visages notamment les émotions faciales qui sont le principal point dans cette étude. Nous avons aussi donné une vue sur l'état de l'art des travaux sur la reconnaissance des émotions faciales. Les détails sur les différentes étapes de notre travail sur la reconnaissance des émotions faciale par apprentissage profond et les méthodes utilisées présentons au chapitre suivant.

## **Chapitre 3**

# **Reconnaissance des émotions faciales par réseaux convolutionnels**

## 3.1 Introduction

Dans ce chapitre, nous parlons sur l'architecture retenue dans notre travail qui basée sur le réseaux de neurones convolutionnels. Quelle base de données on a exploitée ? et quelles sont les couches du réseau de neurones utilisée ? quelles sont les entrées et les sorties de réseau et quel langage de programmation nous avons choisi ?. Toutes les questions qui concernant le réseau CNN utilisé dans notre travail seront débattues en détail dans ce chapitre.

## 3.2 Architecture du réseau

Les réseaux des neurones convolutionnels sont devenu un sujet de recherche très actif dans le domaine de la vision par ordinateur. Leur usage s'est élargit à divers systèmes tels que la classification des images, la reconnaissance des objets et plus particulièrement dans la reconnaissance des émotions faciales qui est le sujet de notre fin d'étude.

Nous avons que l'architecture d'un système varie en fonction du champ d'application. A cette fin, nous avons proposé une architecture d'un réseau convolutionnel adapté à notre problème qui est de classifier les sept émotions à partir d'une image.

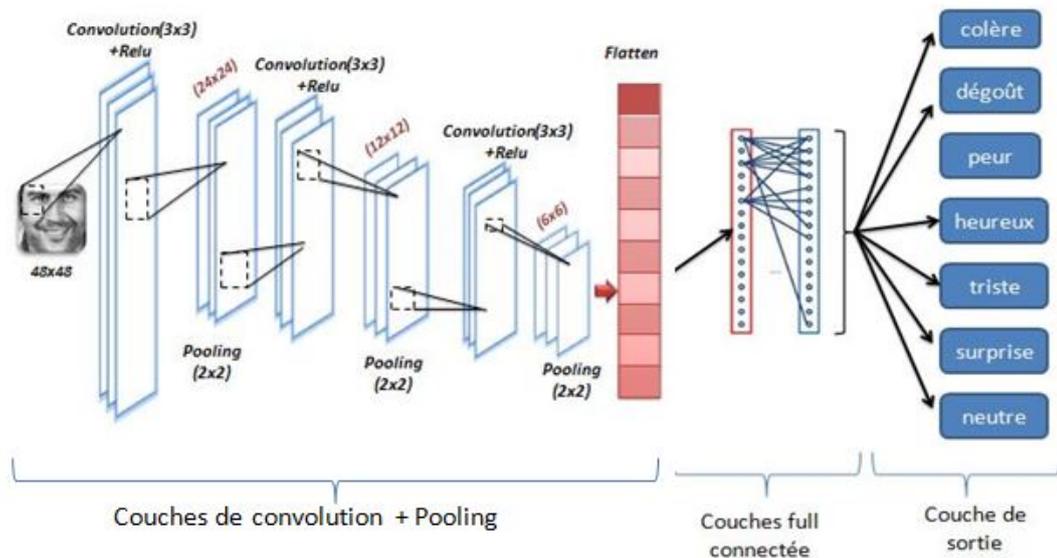


Figure 3.1 – Architecture du réseau utilisé.

Voici, dans ce qui suite, une description des couches formant le réseau proposé.

1. *Couche d'entrée:* est le point de départ de notre système. Elle présente les données de traitement sous la forme des images de taille (48x48) en niveau de gris. Le choix ici est dicté par la base de données retenue et des considérations matérielles.
2. *Couche de convolution:* une couche prend comme entrée une image de taille 48 x 48 et passe par différents filtres pour aider à l'extraction des caractéristiques pertinentes.

pour cela nous avons utilisés cette couche 3 fois avec des filtres de taille (3 x 3), un pas et zéro rembourrage de 1. Chacune de nos couches de convolution est suivie d'une fonction d'activation linéaire rectifier (relu). Une couche de convolution est introduit par la fonction Conv2D du modèle séquentiel d'un réseau convolutif dans notre environnement:

Conv2D(filters, kernel size, activation, padding, stride).

- Filtres: c'est le nombre de filtres de convolution.
  - Kernel size: un entier ou tuple / liste de 2 entiers, spécifiant la taille (hauteur et la largeur) du filtre de convolution 2D.
  - Activation: fonction d'activation à utiliser. Divers choix sont disponibles (relu, elu,..)
3. *Couche d'agrégation spatiale* : c'est l'étape qui, permet de réduire la taille spatiale des cartes caractéristiques locale. Nous avons utilisés le pooling de taille (2 x 2) et un pas de 2 de type max après chaque couche de convolution qui est définit par:

MaxPooling(pool size, stride).

- Pool size = la taille de fenêtre de max pooling.
  - Stride = facteur par lequel réduire. Par exemple 2 divisera à la moitié son entrée
4. *Couche entièrement connectée* : c'est là dernière couche qui contient le raisonnement de haut niveaux, où les données sont, sous la forme d'un vecteur de dimension ( n x 1). Dans cette couche, nous avons utilisés deux couches entièrement connectées avec une fonction d'activation relu, où cette couche définit par:

Dense(units, activation=None).

- Units: entier positif, c'est la dimension de la couche.
- Activation: fonction d'activation à utiliser.

Et une couche de sortie de classification qui contient la fonction softmax qui permet de calculer la distribution de probabilité des sept classes où la formule utilisée est définit par:

Dense(num classe, activation)

- Num classe: le nombre de classe de classification 7 dans notre cas.
- Activation: fonction d'activation à utiliser. Dans cette couche nous avons utilisé la fonction softmax.

## 3.3 Implémentation

Dans cette section, nous présentons l'environnement qui a suivi notre implémentation ainsi que les résultats obtenus.

### 3.3.1 Environnement

Pour l'entraînement de notre réseau, nous avons utilisé un ordinateur ayant les caractéristiques suivantes :

- Marque de PC: Samsung
- Système d'exploitation: windows 7
- Processeur: Intel(R) Core(TM) i3-2350M CPU 2.30GHz
- Carte graphique: Intel(R) HD Graphics 3000
- RAM: 4Go

Pour le développement nous avons retenu le langage Python pour sa simplicité et popularité dans ce domaine. Voici le détail de notre environnement de développant:

- Langage de programmation: Python 3.6
- Environnement de développement python: Pycharm-2019.1.2
- Bibliothèques: pour faire l'entraînement et le teste de base de données nous avons utilisé les bibliothèques suivantes :
  1. TensorFlow : est une bibliothèque pour les calculs numériques à haute performance permet de créer et de déployer facilement des applications d'apprentissage automatique [MMJ18], publiée par Google (<https://www.tensorflow.org/>).
  2. Keras : est une bibliothèque de haut niveau destinée à l'apprentissage profond, écrite en Python. Elle a été développée pour permettre une expérimentation rapide et une recherche de qualité (<https://keras.io/>) [MMJ18].
  3. Numpy : la bibliothèque NumPy (<http://www.numpy.org/>) permet d'effectuer des calculs numériques avec Python. Elle introduit une gestion facilitée des vecteurs et matrices.
  4. Matplotlib : est une bibliothèque riche en fonctionnalités permettant de tracer des graphiques en 2D de haute qualité à l'aide de python (<https://matplotlib.org/>).

### 3.3.2 Base de données

La base de données utilisée dans notre expérimentation est FER2013. Elle été préparée par Pierre-Luc Carrier et Aaron Courville [FER]. Elles est disponible sur la plate forme des compétitions en science des données kaggle.

La base d'image de FER2013 se compose de 35888 images de visages en niveaux de gris comme indiqué dans la figure 3.2. Chaque image à une taille de 48x48, ces images sont étiquetés en 7 classes présenter dans le tableau 3.1 sous dessus. Dans cette base, on trouve 28 709 images pour l'apprentissage et 3 589 images pour le test. Nous choisissons cette base parce qu'elle contient des images en niveaux de gris éliminé les images en couleurs qui contient des données n'est pas nécessaires pour notre entraînement. Et une taille des images a une résolution (48 x 48) qui garantit que l'émotion est détectable [TKC05]. FER2013 nous épargnons de l'étape indispensable de pré-traitement des images souvent très coûteuse.

Émotions	Nombres des images
Colère = 0	4593
Dégoût = 1	547
Peur = 2	5121
Heureux = 3	8989
Triste = 4	6077
Surprise = 5	4002
Neutre = 6	6198

Table 3.1 – Nombres des images correspondants au chaque émotions.



Figure 3.2 – Exemple des images du FER2013

### 3.3.3 Résultats obtenus et discussion

Afin de montrer les résultats obtenus pour notre modèle, on illustre dans ce qui suit les résultats en termes d'exactitude (accuracy) et d'erreur ainsi que la matrice de confusion. Dans notre étude nous avons utilisés deux modèles.

#### A. Modèle 1

Ce modèle à été entraîné pour 11 époques que nous illustrons dans la figure 3.3.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 48, 48, 32)	320
conv2d_2 (Conv2D)	(None, 48, 48, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 24, 24, 32)	0
conv2d_3 (Conv2D)	(None, 24, 24, 64)	18496
conv2d_4 (Conv2D)	(None, 24, 24, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(None, 12, 12, 64)	0
conv2d_5 (Conv2D)	(None, 12, 12, 128)	73856
max_pooling2d_3 (MaxPooling2D)	(None, 6, 6, 128)	0
flatten_1 (Flatten)	(None, 4608)	0
dense_1 (Dense)	(None, 2304)	10619136
dense_2 (Dense)	(None, 2304)	5310720
dense_3 (Dense)	(None, 7)	16135
=====		
Total params: 16,084,839		
Trainable params: 16,084,839		
Non-trainable params: 0		

Figure 3.3 – Configuration du modèle 1

Après l'entraînement de ce modèle nous constatons les remarques suivantes à partir des courbes d'accuracy et loss manifeste dans la figure 3.4a et la figure 3.4b qui représentent le développement des résultats d'accuracy et de loss par rapport les époques que: les accuracy du modèle 1 sur l'ensemble d'entraînement et test sont de 0.99 et 0.55, respectivement. Nous remarquons que les résultats obtenues sur l'ensemble de test en terme d'accuracy et de loss sont mauvaises par rapport aux résultats d'ensemble d'apprentissage. Il est clair que nous somme devons un cas de sur-apprentissage ( Overfitting ). Le sur-apprentissage survient quand le modèle apprend l'ensemble d'entraînement très bien et ne parvient pas à généraliser sur

l'ensemble de test. Ce qui est visible par les écarts entre les deux courbes de la figure 3.4.

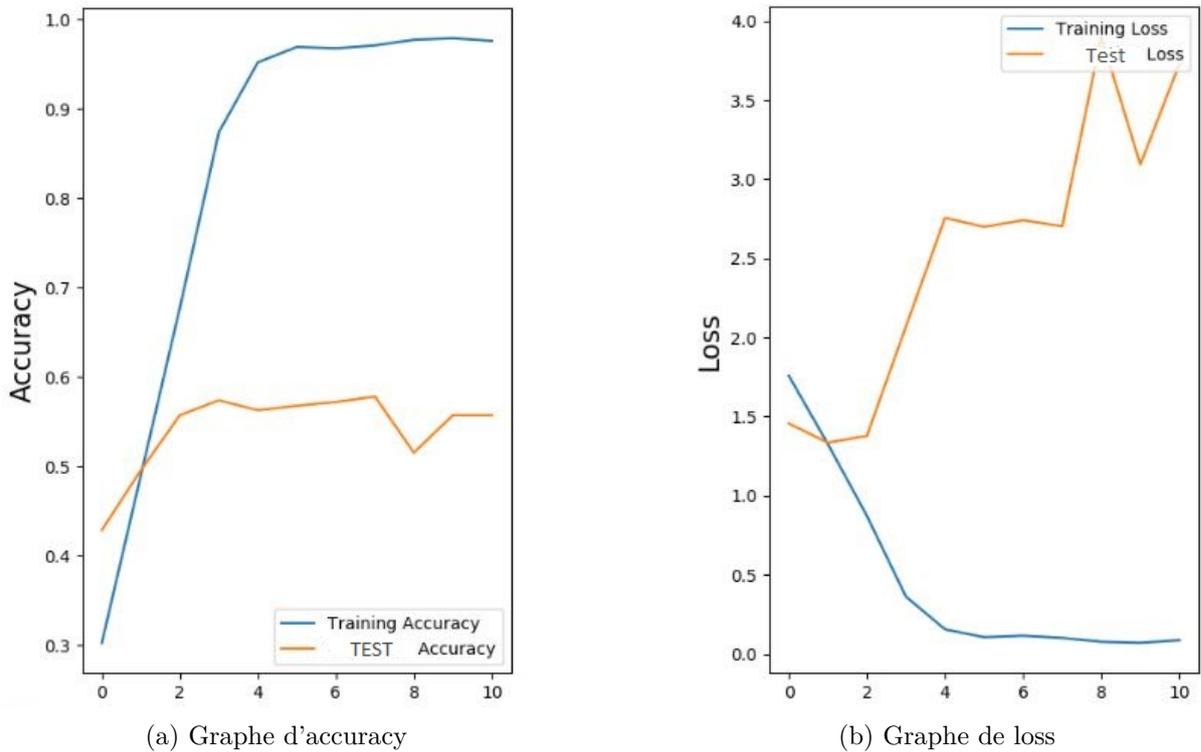


Figure 3.4 – Sur-apprentissage de premier modèle

## B. Modèle 2

Dans ce modèle, nous traitons le problème de sur-apprentissage obtenu en utilisant les méthodes connues dans le domaine de l'apprentissage automatique. Dans notre expérimentation, nous avons retenus : la régularisation L2 et le dropout.

- Dropout: qui consiste à désactiver aléatoirement des noeuds à chaque époque pour cela en l'ajoute après chaque couche entièrement connectée dans le modèle précédent le dropout qui égale à 0,5.
- régularisation l2: permet de mettre à jours la fonction de coût en ajoutant un terme de régularisation, ce terme on ajouter au dernier couche de convolution et les couches entièrement connectées avec 20 époques comme apparaît dans la figure 3.5et 3.6.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 48, 48, 32)	320
conv2d_2 (Conv2D)	(None, 48, 48, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 24, 24, 32)	0
conv2d_3 (Conv2D)	(None, 24, 24, 64)	18496
conv2d_4 (Conv2D)	(None, 24, 24, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(None, 12, 12, 64)	0
conv2d_5 (Conv2D)	(None, 12, 12, 128)	73856
max_pooling2d_3 (MaxPooling2D)	(None, 6, 6, 128)	0
flatten_1 (Flatten)	(None, 4608)	0
dense_1 (Dense)	(None, 2304)	10619136
dense_2 (Dense)	(None, 2304)	5310720
dense_3 (Dense)	(None, 7)	16135
Total params: 16,084,839		
Trainable params: 16,084,839		
Non-trainable params: 0		

Figure 3.5 – Configuration du modèle 2

```

model = Sequential()

model.add(Conv2D(32, kernel_size=(3, 3), activation='relu', padding="same", strides=(1, 1), input_shape=(48, 48, 1)))
model.add(Conv2D(32, kernel_size=(3, 3), padding="same", strides=(1, 1), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2), strides=(2, 2)))

model.add(Conv2D(64, kernel_size=(3, 3), activation='relu', padding="same", strides=(1, 1)))
model.add(Conv2D(64, kernel_size=(3, 3), padding="same", activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2), strides=(2, 2)))

model.add(Conv2D(128, kernel_size=(3, 3), padding="same", strides=(1, 1), kernel_regularizer=regularizers.l2(0.0001), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2), strides=(2, 2)))

model.add(Flatten())
model.add(Dropout(0.5))

model.add(Dense(2304, activation='relu', kernel_regularizer=regularizers.l2(0.0001)))
model.add(Dropout(0.5))

model.add(Dense(2304, activation='relu', kernel_regularizer=regularizers.l2(0.0001)))
model.add(Dropout(0.5))

model.add(Dense(num_classes, activation='softmax'))

```

Figure 3.6 – Architecture du modèle 2

D'après les figures 3.7a et 3.7b, nous trouvons que l'accuracy et loss du modèle 2 sur l'ensemble d'entraînement et test est de 0.6117, 1.22 et 0.6052, 1.25 respective-

ment. Cela indique que le problème de surapprentissage sont relativement améliorée puisque les courbes d'entraînement de test ne sont pas divergentes.

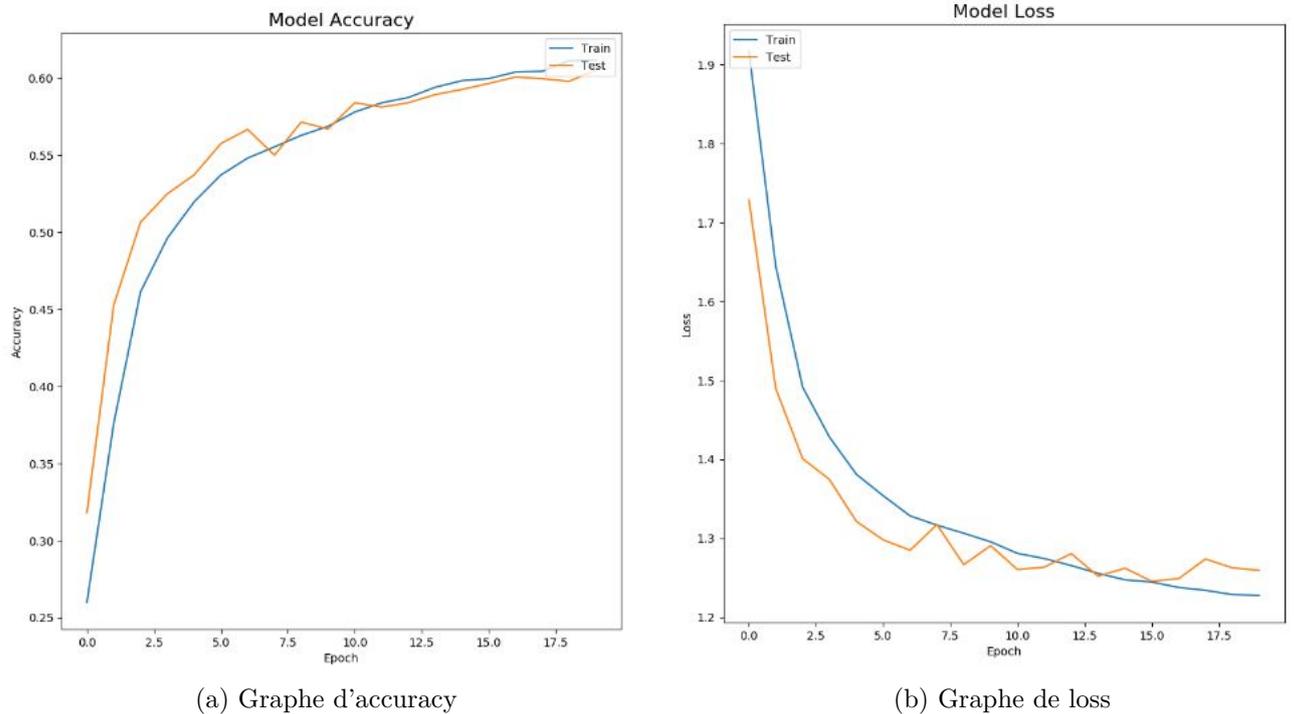


Figure 3.7 – Sur-apprentissage de premier modèle

La figure 3.8 représente la matrice de confusion qui permet d'évaluer la performance de notre modèle. La matrice illustre de près la classification des images dans chaque classe émotionnelle. A titre d'exemple, le modèle a bien classé les images heureux et surprises et il a mal classé les images de dégoût et peur.

La figure 3.9 illustre la portion des images bien et mal classées de chaque classe. Nous remarquons que la classe heureux possède le meilleur score avec 698 images classées comme heureux contre 197 images avec fausse prédiction par apport la totalité des images heureux. La classe dégoût quant à elle possède la mauvaise classification où 15 images seulement ont été classées comme dégoût vraiment. De la figure 3.10 nous concluons que la totalité des images est bien classées avec une proportion de 60.52% qui est équivalent à 2172 images alors que le taux de la mal-classification représente par 39.48%. Notons que ces résultats sont proches des performance atteintes par les travaux de l'état de l'art de la base FER2013 [LHHK13].

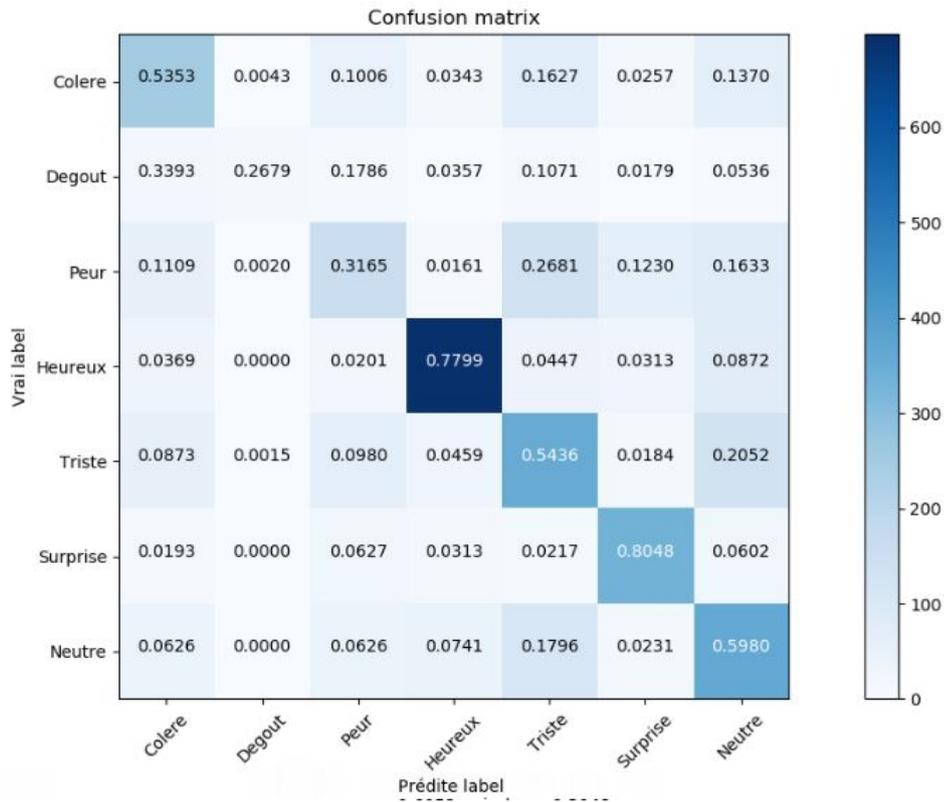


Figure 3.8 – Matrice de Confusion de modèle 02

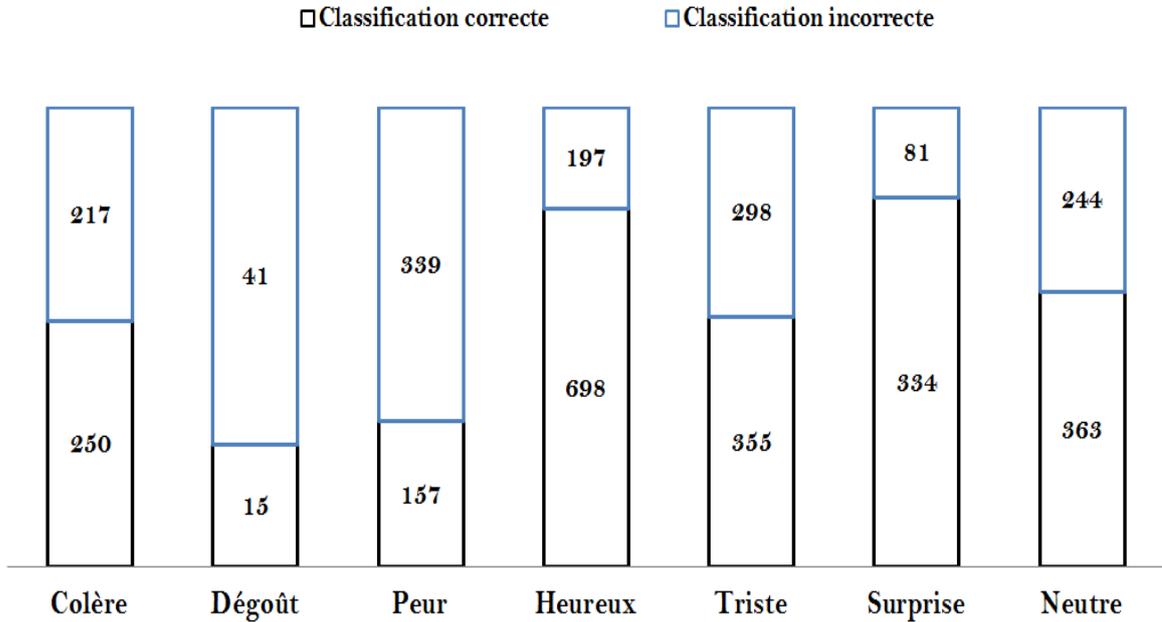


Figure 3.9 – Nombre des images mal et bien classé par classe

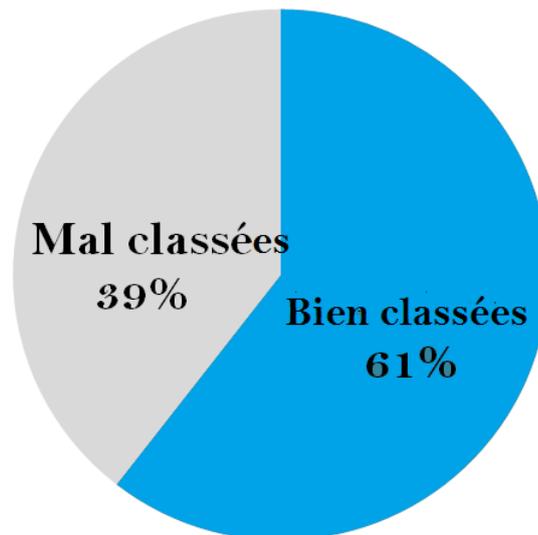


Figure 3.10 – Nombre total des images mal et bien classées

### C. Tableau des résultats

Le tableau 3.2 ci-dessous résume les architectures et les différents résultats obtenus par les deux modèles proposés. Le tableau montre que le deuxième modèle donne de bons résultats par rapport le modèle 1 où nous pouvons voir que l’accuracy d’entraînement et du test du modèle 2 sont acceptable face au modèle 1 qui tombe dans le cas de sur apprentissage et ce résultat acceptable revient à l’ajout de la régularisation et dropout au modèle 2.

Le tableau 3.2 aussi montre que le temps d’exécution est très coûteux, ceci revient à la taille importante de la base de données et aux caractérisation de notre machine utilisée qui contient un CPU et les réseaux profonds nécessite l’utilisation des GPUs ou autres environnements de haut calculs.

**Remarque** : nous avons en abrégé les mots suivants afin d’assurer leur clarté dans le tableau :

Couche de convolution : CONV

Couche d’agrégation spatiale : Pooling

Couche entièrement connectée : EC

Temps d’exécution: temps

Table 3.2 – Le résumé des résultats.

Modèle	Architecture				Époque	Accuracy		Erreur	Temps (s)
	CONV	Pooling	EC	Dropout		Apprentissage	Test		
1	3	3	3	non	11	98%	55.72%	44.28%	31765 (8h)
2	3	3	3	oui	20	61.17%	60.52%	39.48%	53842 (14h)

Notre deuxième réseau de neurones convolutionnel donne un résultat égale à 60.52% alors il est acceptable par apport a l'état de l'art de la base de données FER2013 [LHHK13]. Donc, aussi la base de données influe positivement/négativement au résultat obtenus. Pour des bons résultats en terme de performance et temps d'exécution, il faut utilise les GPUs et une base de données de grande taille que la base FER2013.

#### D. Résultats d'application

Dans cette section, nous présentons l'ensemble des figure de 3.11 à 3.18 qui montrent les résultats de prédictions de notre réseau sur des exemples varies.

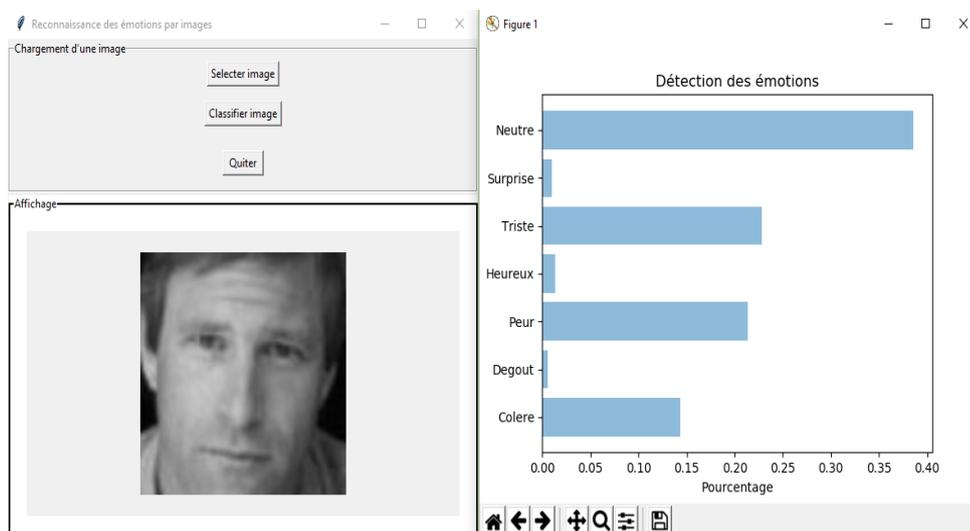


Figure 3.11 – Prédiction correcte de neutre

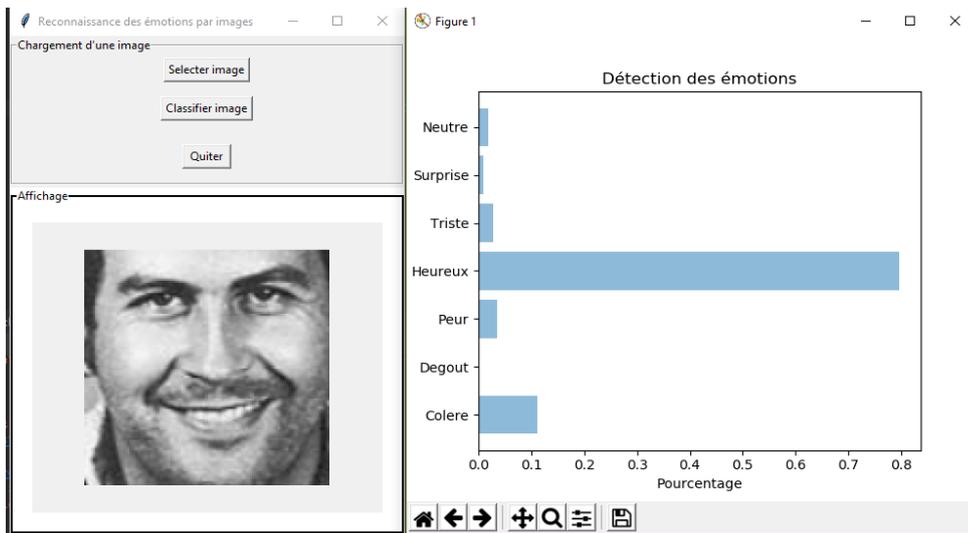


Figure 3.12 – Prédiction correcte d'heureux

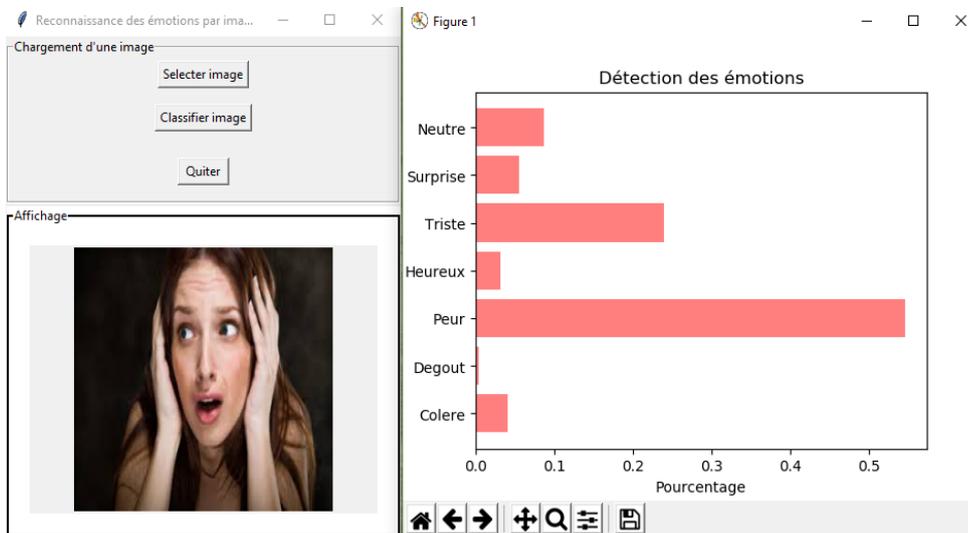


Figure 3.13 – Prédiction correcte de peur

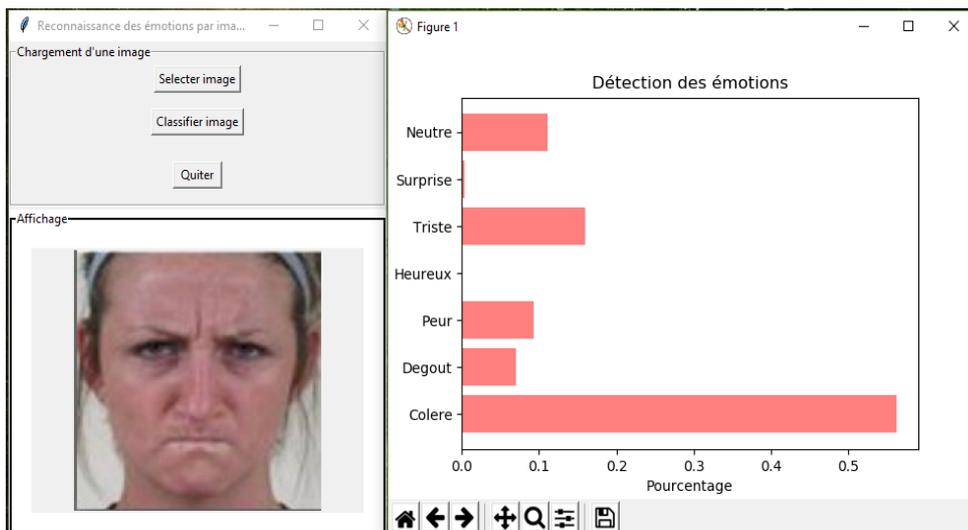


Figure 3.14 – Prédiction correcte de colère

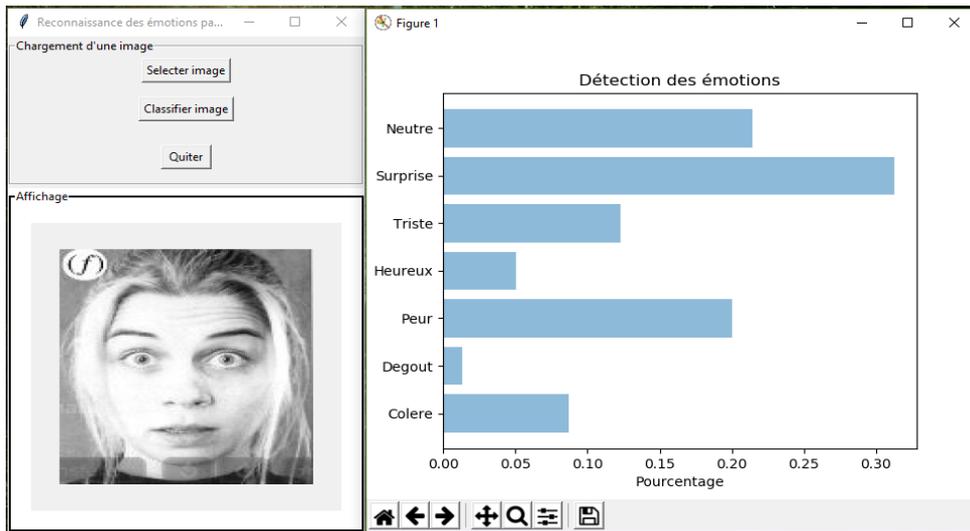


Figure 3.15 – Prédiction correcte de surprise

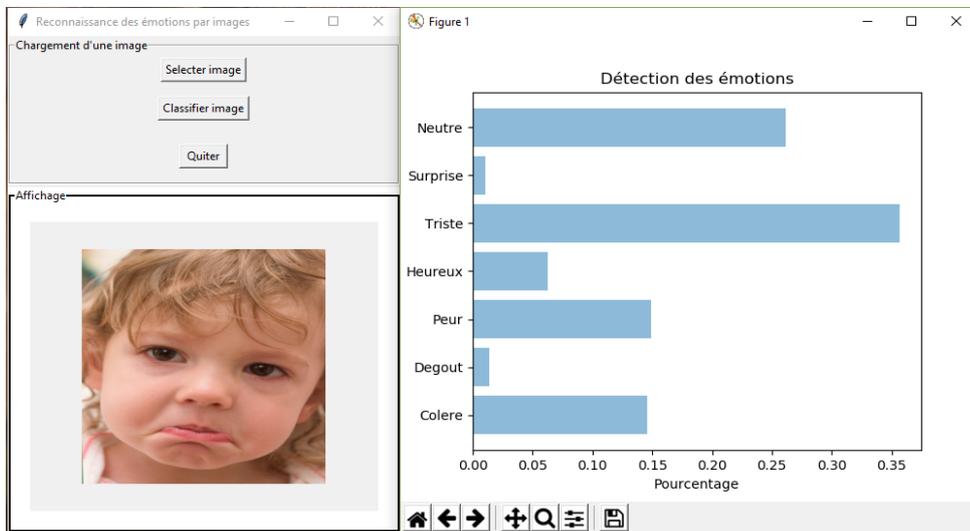


Figure 3.16 – Prédiction correcte de triste

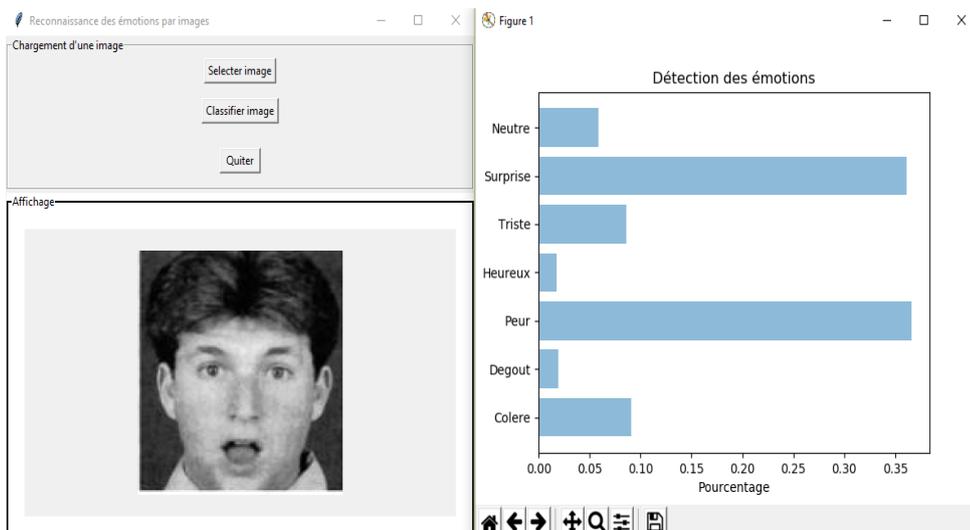


Figure 3.17 – Prédiction fautive de la surprise

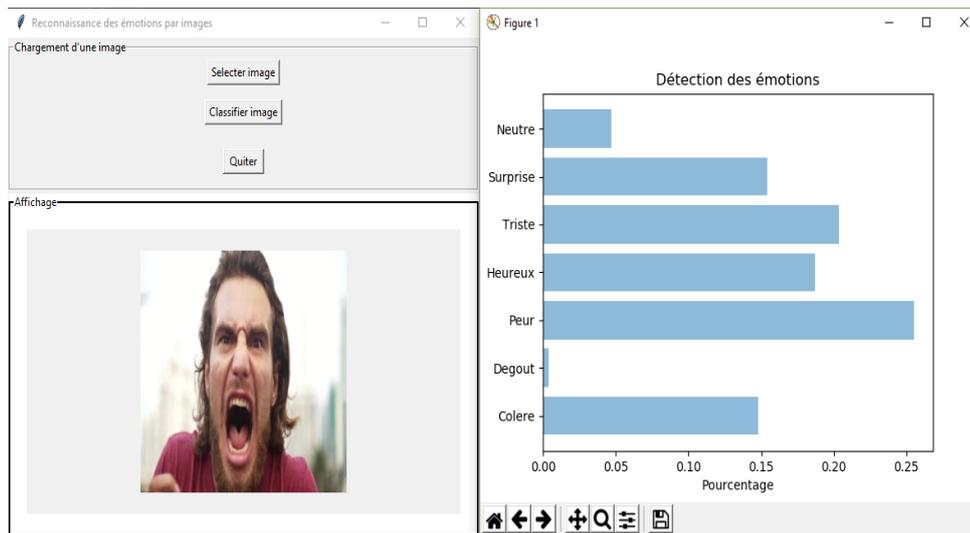


Figure 3.18 – Prédiction fautive de la colère

### 3.4 Conclusion

Dans ce troisième chapitre, nous avons présenté les architectures des réseaux de neurones convolutionnels utilisées et les différents résultats obtenus sur la base FER2013. Pour ces architectures, pour notre application de reconnaissance des émotions faciales. Cependant, nous avons vu d'après les résultats que les moyens de la régularisation des réseaux RNC et le nombre d'époque permettent d'améliorer les architectures des réseaux, et qu'il est entraîné difficile de trouver des paramètres adéquates à ce type de systèmes. Nous examinons, nous sommes satisfaits des résultats obtenus.

# Conclusion générale

Dans ce projet, nous avons abordé les notions de base de l'apprentissage automatique avec ses méthodes, et spécialement l'apprentissage profond. Nous avons montré en particulier de rôle des réseaux convolutionnels dans le traitement des images grâce aux leurs couches constitutifs notamment: les couches de convolution, de pooling, et les couches entièrement connectées. Le but est la conception et l'implémentation d'une application de reconnaissance des émotions.

Notre travail consiste à mettre au point un modèle capable de classifier ou reconnaître une émotion (colère, dégoût, peur, heureux, triste, surprise, neutre) à partir d'une image en utilisant l'apprentissage profond notamment les réseaux de neurones convolutionnels qui se base sur l'opération d'extraction des cartes caractéristiques et la construction automatique de ce modèle du base de données FER2013, Il nous a fallu résoudre le problème de surapprentissage rencontré. Sachant que les possibilités de notre unité de traitement ( CPU ) sont simples par rapport aux unité de traitements modernes (GPU) où l'apprentissage profond exige des ressources énormes (mémoire, processus,..). Aussi le réglage des nombreux paramètres impliqué dans ce type de projet est un problème fastidieux.

Ce travail peut être étendu sur plusieurs axes. Voici un ensemble des points des perspectives retenues : porter ce travail sur des environnements haut de calcul: GPU, Cloud, Clusters. Et explorer les avantages des facteurs d'architectures de paramètres, hyperparamètres permettant d'améliorer la qualité du réseau. Aussi étudier l'impact du l'apprentissage par transfert de réseau semblables. Si nous avons plus de temps, nous aurions amélioré et développé cette application pour la reconnaissance des émotions faciales à partir d'images quelconques : images en couleurs, poses dans la nature, etc. Également la reconnaissance des émotions faciales à partir de séquences vidéos, ainsi que la détection des émotions d'un groupe de personnes. Enfin, étant donné que les images faciales sont symétriques dans la majorité des cas, il est légitime d'expérimenter notre réseau convolutionnel sur la moitié des images seulement. Cela permet en principe de réduire le volume de données traitées, et par conséquent le temps nécessaire au traitement, nous conjecturons que cela ne dégrade pas la qualité de la réponse.

# Bibliographie

- [Aba08] Souhila Guerfi Ababsa. Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2d/3d. *Evry-Val d'Essonne*, 2008.
- [Abd10] Faiza Abdat. Reconnaissance automatique des émotions par données multimodales: expressions faciales et signaux physiologiques. *Université de Metz, France*, 2010.
- [AK06] Petar S Aleksic and Aggelos K Katsaggelos. Automatic facial expression recognition using facial animation parameters and multistream hmms. *IEEE Transactions on Information Forensics and Security*, 1(1):3–11, 2006.
- [Aze18] C.A. Azencott. *Introduction au Machine Learning*. Dunod, 2018.
- [Ber18] Jérôme Berthier. Que (nous) font les big data? *Revue internationale et stratégique*, (2):89–99, 2018.
- [Bet12] Vinay Bettadapura. Face expression recognition and analysis: the state of the art. *arXiv preprint arXiv:1203.6722*, 2012.
- [BH00] Imad A Basheer and Maha Hajmeer. Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, 43(1):3–31, 2000.
- [BK14] Young-Hyen Byeon and Keun-Chang Kwak. Facial expression recognition using 3d convolutional neural network. *International journal of advanced computer science and applications*, 5(12), 2014.
- [Bou12] Sofiane Boudjellal. *Détection et identification de personne par méthode biométrique*. PhD thesis, Université Mouloud Mammeri, 2012.
- [Bou15] Mohamed Bouguessa. *Forage de données. Notes de cours*. Université du Québec à Montréal. 2015.
- [BTA<sup>+</sup>15] Peter Burkert, Felix Trier, Muhammad Zeshan Afzal, Andreas Dengel, and Marcus Liwicki. Dexpression: Deep convolutional neural network for expression recognition. *arXiv preprint arXiv:1509.05371*, 2015.
- [CM10] A. Cornuéjols and L. Miclet. *Apprentissage artificiel*. Algorithmes (Paris). Eyrolles, 2010.
- [CSZ06] O Chapelle, B Schölkopf, and A Zien. Semi-supervised learning, ser. adaptive computation and machine learning. 2006.

- [CTY<sup>+</sup>03] Cheng-Chin Chiang, Wen-Kai Tai, Mau-Tsuen Yang, Yi-Ting Huang, and Chi-Jaung Huang. A novel method for detecting lips, eyes and faces in real time. *Real-time imaging*, 9(4):277–287, 2003.
- [Dar72] Charles Darwin. 1965. the expression of the emotions in man and animals. *London, UK: John Marry*, 1872.
- [De] Debarko De. Rnn or recurrent neural network for noobs. <<https://hackernoon.com/rnn-or-recurrent-neural-network-for-noobs-a9afbb00e860>>.
- [Dec08] Benoit Decou. *Neurocomputing - Réseaux de Neurones*. mars-mai 2008.
- [DGL13] Luc Devroye, László Györfi, and Gábor Lugosi. *A probabilistic theory of pattern recognition*, volume 31. Springer Science & Business Media, 2013.
- [Dre08] G. Dreyfus. *Apprentissage statistique*. Algorithmes (Paris). Eyrolles, 2008.
- [DT16] Marco Dinarelli and Isabelle Tellier. Étude de réseaux de neurones récurrents pour l’étiquetage de séquences. In *Traitement Automatique des Langues Naturelles (TALN)*, 2016.
- [Ekm97] Rosenberg Ekman. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.
- [FER] FER2013. Challenges in representation learning: Facial expression recognition challenge. <<https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data>>.
- [GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [Gha16] Sonia Gharsalli. *Reconnaissance des émotions par traitement d’images*. PhD thesis, Université d’Orléans, 2016.
- [GL13] Deepak Ghimire and Joonwhoan Lee. Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. *Sensors*, 13(6):7714–7734, 2013.
- [Gra16] Daniel Graupe. *Deep learning neural networks: Design and case studies*. World Scientific Publishing Company, 2016.
- [Gro] Intelligent Behaviour Understanding Group. Intelligent behaviour understanding group (ibug), department of computing, imperial college london.
- [GZ11] Matti Taini Stan Z. Li Matti Pietikäinen Guoying Zhao, Xiaohua Huang. Facial expression recognition from near-infrared videos. *Image and Vision Computing*, 29(9):607–619, 2011.
- [Hiz09] Walid Hizem. *Capteur intelligent pour la reconnaissance de visage*. PhD thesis, Evry, Institut national des télécommunications, 2009.

- [HMM17] Behzad H. Hasani; Mohammad Mahoor. Facial expression recognition using enhanced deep 3d convolutional neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.
- [img] Machine learning intro. <<https://www.slideshare.net/adunnantakaew/machine-learning-intro-109381827>>.
- [JK18] Mira Jeong and Byoung Chul Ko. Driver’s facial expression recognition in real-time for safe driving. *Sensors*, 18(12):4270, 2018.
- [JLY<sup>+</sup>15] Heechul Jung, Sihaeng Lee, Junho Yim, Sunjeong Park, and Junmo Kim. Joint fine-tuning in deep neural networks for facial expression recognition. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.
- [JZH17] Deepak Kumar Jain, Zhang Zhang, and Kaiqi Huang. Multi angle optimal pattern-based deep learning for automatic facial expression recognition. *Pattern Recognition Letters*, 2017.
- [Khe13] Bouchra Khefif. *Mise au point d’une application de reconnaissance faciale*. PhD thesis, Université Abou Bakr Belkaid – Tlemcen, 2013.
- [Ko18] Byoung Ko. A brief review of facial emotion recognition based on visual information. *sensors*, 18(2):401, 2018.
- [KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [LCK<sup>+</sup>10] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 94–101. IEEE, 2010.
- [LHHK13] Minhoo Lee, Akira Hirose, Zeng-Guang Hou, and Rhee Man Kil. Neural information processing: 20th international conference. In *ICONIP*, pages 3–7, 2013.
- [MMJ18] Navin Kumar Manaswi, Navin Kumar Manaswi, and Suresh John. *Deep Learning with Applications Using Python*. Springer, 2018.
- [MVJP17] Brais Martinez, Michel F Valstar, Bihan Jiang, and Maja Pantic. Automatic analysis of facial actions: A survey. *IEEE transactions on affective computing*, 2017.
- [Pac] Víctor González Pacheco. Une brève histoire de l’apprentissage automatique. <<http://www.synergicpartners.com/en/espanol-una-breve-historia-del-machine-learning/>>.
- [Par16] Zaven Paré. La virtuosité artificielle des robots. *Sémiotique et vécu musical: Du sens à l’expérience, de l’expérience au sens*, page 171, 2016.

- [Plu84] Robert Plutchik. Emotions: A general psychoevolutionary theory. *Approaches to emotion*, 1984:197–219, 1984.
- [PZH14] Isidoros Perikos, Epaminondas Ziakopoulos, and Ioannis Hatzilygeroudis. Recognizing emotions from facial expressions using neural network. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 236–245. Springer, 2014.
- [R] Lambert R. Focus: Le réseau de neurones convolutifs. <<http://penseeartificielle.fr/focus-reseau-neurones-convolutifs/>>.
- [RHW<sup>+</sup>88] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.
- [SC01] Karen L Schmidt and Jeffrey F Cohn. Human facial expressions as adaptations: Evolutionary questions in facial expression research. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, 116(S33):3–24, 2001.
- [SIM18] Florent SIMON. Deep learning, le réseau à convolution. <<https://www.supinfo.com/articles/single/8037-deep-learning-reseau-convolution>>, 05/10/2018.
- [Stu] Convolutional neural networks (cnn) introduction. <<https://indoml.com/2018/03/07/student-notes-convolutional-neural-networks-cnn-introduction/#more-6446>>.
- [SWL13] Peijia Shen, Shangfei Wang, and Zhilei Liu. Facial expression recognition from infrared thermal videos. In *Intelligent Autonomous Systems 12*, pages 323–333. Springer, 2013.
- [Taf] Mokhtar Taffar. *Initiation à l'apprentissage automatique. Support de Cours. Université de Jijel*.
- [tau] Réduction de la perte : le taux d'apprentissage. <<https://developers.google.com/machine-learning/crash-course/reducing-loss/learning-rate?hl=fr>>.
- [Tho] Nicolas Thome. Apprentissage, réseaux de neurones, département informatique national des arts et métiers (cnam).
- [TKC05] Ying-Li Tian, Takeo Kanade, and Jeffrey F Cohn. Facial expression analysis. In *Handbook of face recognition*, pages 247–275. Springer, 2005.
- [TNH11] Le Hoang Thai, Nguyen Do Thai Nguyen, and Tran Son Hai. A facial expression classification system integrating canny, principal component analysis and artificial neural network. *arXiv preprint arXiv:1111.4052*, 2011.
- [Tou92] Claude Touzet. *les réseaux de neurones artificiels, introduction au connexionnisme*. EC2, 1992.
- [Tur09] Alan M Turing. Computing machinery and intelligence. In *Parsing the Turing Test*, pages 23–65. Springer, 2009.
- [Vie18] João Santos Crispiniano Vieira. Eaglenet-a deep multi-task learning cnn for emotion, age and gender recognition. 2018.

- [VWMVB10] Mathieu Van Wambeke, Benoit Macq, and Christian Van Brussel. Reconnaissance et suivi de visages et implémentation en robotique temps-réel. *Université Catholique de Louvain Ecole Polytechnique de Louvain. Mémoire de fin d'études*, 2010.
- [YZFY04] Jian Yang, David D Zhang, Alejandro F Frangi, and Jing-yu Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE transactions on pattern analysis and machine intelligence*, 2004.
- [ZB09] Wenlong Zheng and Suchendra M Bhandarkar. Face detection and tracking using a boosted adaptive particle filter. *Journal of Visual Communication and Image Representation*, 20(1):9–27, 2009.