



RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
UNIVERSITÉ DE TLEMCEN
FACULTÉ DE TECHNOLOGIE

Département de génie électrique et électronique
Spécialité Automatique
Option: Informatique industrielle

Mémoire de master intitulé

**L'intelligence artificielle pour l'informatique affective:
Reconnaissance de la valence des émotions par
apprentissage profond**

Présenté par:

Mr Ahmed ZIANE et Mlle Siham ARAHMANE

Soutenu le 26/06/2018 devant le jury

<i>PRÉSIDENT :</i>	Mr A. Hadj Abdelkader	MCA, Université de Tlemcen
<i>EXAMINATEUR :</i>	Mr A. Mechernene	MCA, Université de Tlemcen
<i>EXAMINATEUR :</i>	Mr M.H Irid	MCB, Université de Tlemcen
<i>ENCADRANTE :</i>	Mme W. Handouzi	MCB, Université de Tlemcen
<i>CO- ENCADRANT :</i>	Mr A. Rimouche	MCB, Université de Tlemcen

Dédicace

Avec l'aide de ALLAH, tout puissant, j'ai pu achever ce modeste travail que je dédie :

À mes parents qui est grâce à eux que je suis arrivé à ce stade.

Que Dieu les garde et leur accorde longue et prospère vie.

À mes frères Toufik et Mohammed et nièce Marwa Ghizlane.

À toutes les personnes qui m'ont aidé.

ZIANE Ahmed

Dédicace

*À la lumière de mes jours, mon soutien moral, la source de mes efforts et de mes joies,
la flamme de mon cœur, ma vie et mon bonheur, celle qui s'est toujours sacrifié pour
me voir réussir et être heureuse, et que le dieu me la protège à ma mère que j'adore*

*À mon soutien dans la vie et la source de mes efforts et mon courage et qui m'a aidé
beaucoup dans mes études et me donne l'envi de les continuer jusqu'à la fin à mon père*

À tous les membres de ma famille spécialement à ma grande mère, mon frère

Mohammad, mon oncle Abdelnabi et sa femme Zakia et ses filles Amira et Isra

*À tous les personnes et les amis qui m'ont aidé et m'encouragé pendant la période de
mes études universitaires spécialement KADDOUR Zakaria*

*Je dédie ce travail dont le grand plaisir leur revient en premier pour leur conseils,
aides et encouragements*

ARAHMANE Siham

Remerciements

D'abord, nous remercions ALLAH le tout puissant de m'avoir donné le courage, la volonté, patience et la santé pour ce finir ce travail.

Nous tenons beaucoup à remercier Madame Wahida HANDOUZI Maître de conférence à l'Université de Tlemcen qui nous a encadrée avec beaucoup d'efficacité et a été la source de précieux conseils et encouragements, et aussi pour sa patience, et sa disponibilité quand nous avons besoin d'elle.

Nos remerciements s'adressent aussi à Monsieur Ali RIMOUCHE Maître de conférence à l'Université de Tlemcen pour ses conseils et ses aides durant la réalisation de ce mémoire.

Nous remercions l'ensemble des membres du jury qui nous ont fait l'honneur d'accepter de lire et d'évaluer ce mémoire : Monsieur A. HADJ ABDELKADER (Maître de conférence à l'Université de Tlemcen, examinateur et président du jury), Monsieur A. MECHERNENE (Maître de conférence à l'Université de Tlemcen, examinateur) et Monsieur M.H. IRID (Maître de conférence à l'Université de Tlemcen, examinateur).

Nous remercions notre meilleur ami KADDOUR Zakaria pour son aide logistique lors des tests de ce mémoire grâce à la puissance de sa machine.

Résumé

L'informatique affective (affective computing) est une thématique pluridisciplinaire émergente qui regroupe des chercheurs du domaine de l'intelligence artificielle et du traitement du langage naturel jusqu'au domaine des sciences sociales et cognitives. La reconnaissance des émotions et de manière plus large l'informatique affective, a significativement évolué ces dernières années.

Dans ce travail, nous avons proposé une architecture d'un réseau de neurones pour la reconnaissance de la valence des émotions (positives ou négatives). Nous avons comparé plusieurs setups d'entraînement de 80% d'une base de donnée RaFD avec deux fonctions d'activation (Relu et Sigmoid), deux optimiseurs (Adam et SGD) et Softmax dans couche de sortie, nous avons trouvé que le meilleur setup est la fonction Relu dans les couches cachées, optimiseur Adam avec 1000 d'itérations. Le résultat du test de ce setup avec le reste des données donne 95% de précision.

Abstract

Affective computing is an emerging multidisciplinary theme that brings researchers from the field of artificial intelligence and natural language processing to the field of social and cognitive sciences.

The recognition of emotions and more broadly affective computing, has significantly evolved in recent years.

In this work, we proposed an architecture of a network of neurons for the recognition of the valence of emotions (positif or negatif). We tested several training setups of 80% of RaFD dataset with two activation functions (Relu and Sigmoid), two optimizers (Adam and SGD) and Softmax in output layer, we found that the best setup is to use Relu function and Adam optimizer with 1000 epochs.

Sommaire

Introduction générale	1
1 Préliminaires	2
1.1 Introduction	2
1.2 Émotion	2
1.2.1 Type d'émotions	3
1.2.2 La Classification et la Représentation des émotions	5
1.2.3 Comment peut-on-mesurer les émotions?	9
1.3 Apprentissage automatique (Machine learning)	16
1.3.1 Quelques différents types et méthodes de l'apprentissage profond.	17
1.3.2 Les réseaux de neurones une méthode d'apprentissage supervisé :	18
2 Méthodologie	21
2.1 Base de données radboud faces (RaFD) :	21
2.2 Réseaux de neurones :	27
2.2.1 Fonctions d'activation :	30
2.2.2 Fonctions d'optimisation :	33
2.2.3 Réseau de neurones convolutifs (ConvNet) :	41
2.2.4 Les Bibliothèques utilisées :	44
3 Résultats et discussions	48
Conclusion	58
Bibliographie	59

Liste des tableaux

1.1	Les Émotions basiques selon différents auteurs	3
1.2	Les 6 émotions de bases	5
3.1	Résultat après l'entraînement du premier setup.	50
3.2	Résultat après l'entraînement.	52
3.3	Résultat après l'entraînement de troisième setup.	54
3.4	Résultat d'entraînement de quatrième setup.	56
3.5	Résultat de test.	57

Table des figures

1.1	L'approche catégorielle	6
1.2	Représentation des émotions selon deux axes : l'arousal et la valence.	7
1.3	L'expression faciale associée à la joie	8
1.4	L'expression faciale associée à la colère	9
1.5	Quelques composantes des émotions.	9
1.6	Système nerveux autonome (SNA).	11
1.7	Quelques capteurs utilisés dans la littérature.	12
1.8	Echelle d'auto-évaluation (SAM).	13
1.9	Les expressions faciales des six émotions de base selon Ekman [4]	14
1.10	Les différentes étapes d'un système de reconnaissance des expressions faciales [3]	16
1.11	Les différents types de la machine learning [8].	18
1.12	Exemple de réseaux de neurones avec une couche cachée.	19
1.13	Exemples d'un réseaux de neurones profonds	20
2.1	Organigramme représentant les étapes du système de reconnaissance.	21
2.2	Les expressions faciales représentées dans la base RaFD	22
2.3	Emplacement des caméras et des images prises par ces caméras.	22
2.4	Direction de regard pris par caméra de 90.	23
2.5	Détection de visage à l'aide d'openCV.	24
2.6	L'image obtenue après recadrage.	25
2.7	Image RVB avec Tableau multidimensionnel	26
2.8	Image 2.8a à résolution 389×389 , image 2.8b à résolution 48×48	26
2.9	Un nœuds qui reçoit trois entrées.	27
2.10	Le concept d'apprentissage supervisé.	29
2.11	Différentes architectures du réseau de neurones	30
2.12	La fonction Sigmoid.	31
2.13	La fonction ReLU.	33
2.14	Entropie croisée.	40
2.15	Principe de dropout	41
2.16	Architecture d'un réseau convnet avec 3 nœuds de sortie.	42
2.17	Convolution entre image (3×3) et filtre 2×2 par bloc	42
2.18	Procédure de pooling	43
2.19	Architecture du réseau utilisé.	47

3.1	Architecture utilisée pour l'entraînement.	49
3.2	Erreur d'entraînement de Sigmoid avec SGD pour 500 itérations.	49
3.3	Erreur d'entraînement de Sigmoid avec SGD pour 1000 itérations.	50
3.4	Architecture utilisée pour l'entraînement.	51
3.5	Erreur d'entraînement de Sigmoid avec Adam pour 500 itérations.	51
3.6	Erreur d'entraînement de Sigmoid avec Adam pour 1000 itérations.	52
3.7	Architecture utilisée pour l'entraînement.	53
3.8	Erreur d'entraînement de Relu avec SGD pour 500 itérations.	53
3.9	Erreur d'entraînement de Relu avec SGD 1000 itérations.	54
3.10	Architecture utilisée pour l'entraînement.	55
3.11	Erreur d'entraînement de Relu avec Adam pour 500 itérations.	55
3.12	Erreur d'entraînement de Relu avec Adam pour 1000 itérations	56
3.13	Architecture qui nous donne de meilleures performances	58

Introduction générale

L'informatique affective (affective computing) est une thématique pluridisciplinaire émergente qui regroupe des chercheurs du domaine de l'intelligence artificielle et du traitement du langage naturel jusqu'au domaine des sciences sociales et cognitives [1]. La reconnaissance des émotions et de manière plus large l'informatique affective [2], a significativement évolué ces dernières années.

Dans ce rapport le premier chapitre présenté en premier lieu des notions de bases sur les émotions, leurs mesure puis décrit plus particulièrement la reconnaissance automatique par apprentissage profond.

Le second chapitre est tout d'abord consacré à la présentation de la méthodologie suivie pour analyser les expressions faciales, en commençant par le choix et la préparation de la base de données. Ensuite, nous parlerons des réseaux de neurones et leur principe jusqu'à arriver à l'apprentissage profond.

Nous concluons ce chapitre, par la présentation de l'architecture du réseau de neurones qu'on va utiliser. Le troisième chapitre présente les résultats expérimentaux obtenus et les discussions des résultats des tests et comparaisons réalisés durant notre projet de PFE. Enfin, nous concluons ce rapport par une conclusion générale et la proposition de quelques perspectives.

Préliminaires

1.1 Introduction

Dans la complexité continue de la vie sociale les émotions jouent un rôle clé dans la définition et régulation de nos relations avec les autres et plus généralement avec l'environnement qui nous entourent. Nos réactions émotionnelle à d'autres personnes influence la façon dont ses autres personnes réagissent à nous et, dans une certaine mesure, comment les futures rencontres vont se développer. en même temps, notre propre comportement émotionnel est façonné par les pensées et les actes des autres. C'est pourquoi doter la machine des capacités de reconnaissance des émotions par traitement des expressions faciales est un défi scientifique autour duquel se rassemblent différentes communautés (traitement du signal, traitement d'images, intelligence artificielle, robotique, interaction homme-machine, . . . etc)

Dans ce chapitre, nous allons introduire toutes les notions en relation avec notre thématique, comme les émotions et leurs valences, mais aussi une introduction à l'apprentissage profond.

1.2 Émotion

Définition. Les émotions, de façon générale, sont des états motivationnels qui sont constituées d'impulsions, de désirs ou d'aversion ou plus généralement, elles comportent des changements de motivation. Ces émotions poussent l'individu à modifier sa relation avec un objet, un état du monde, un état de soi, ou à maintenir une relation existante malgré des obstacles ou des interférences [3].

De plus l'émotion est un signal que nous envoie notre cerveau pour nous aider à nous adapter à notre environnement. Ce processus est très rapide et tout le monde a le même, mais chaque personne va être plus sensible à une étape qu'à une autre et donc ces mouvements affectifs sont soudains et temporaires, en réaction à un événement, une interaction humaine ou une pensée, accompagnés de changements physiologiques (cœur, respiration, température, sueur, pleurs, rires, tension, relâchement, . . . etc). L'émotion est un terme très large car il prend en considération plusieurs facteurs donc il n'a pas une définition claire, précise et unique sur ce sujet.

Il existe différentes émotions, le groupe des psychologues a proposé une liste d'émotions basiques allant de 2 jusqu'à 18 catégories Ekman et son groupe ont mené différentes études sur les expressions faciales qui ont conduit à conclure qu'il y a 6 émotions de base appelées aussi émotions primaires : Joie, tristesse, surprise, peur, colère et dégoût. Le [Tableau 1.1](#) illustre les émotions primaires selon Izard, Plutchik et Ekman [5].

Auteurs	émotions
Izard (1977)	Joie, surprise, colère, peur, tristesse, mépris, détresse, intérêt, culpabilité, honte,
Plutchik (1980)	Acceptation, colère, anticipation, dégoût, joie, peur, tristesse, surprise
Ekman (1992)	Colère, peur, tristesse, joie, dégoût, surprise

TABLEAU 1.1 – Les Émotions basiques selon différents auteurs

1.2.1 Type d'émotions

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

a. Émotion primaire

Les émotions fondamentales ou primaires sont présentes chez le bébé ainsi que chez le primate, c'est pour cela qu'elles se nomment ainsi. De plus, elles s'identifient de manière équivoque sur le visage, peu importe la culture des personnes qui les manifestent.

Ces émotions sont déclenchées par des événements particuliers ou bien elles se manifestent dans des circonstances précises en provoquant des comportements spécifiques et elles sont à la base de nos réactions qui ne sont pas seulement déterminées par notre jugement rationnel ou notre passé individuel mais aussi par notre passé ancestral. En fait, ces émotions primaires sont comme une matière première, à partir de laquelle on peut fabriquer toutes les autres émotions [3].

b. Émotion secondaire

Les émotions secondaires ont pour base, au départ, un processus de pensée et sont l'aboutissement de l'apprentissage des émotions primaires. Les émotions secondaires sont celles qui sont engendrées à l'évocation de souvenirs et arrivent à maturation à l'âge adulte [3].

c. Émotion sociale

On peut ajouter un autre type des émotions qui sont les émotions sociales. Ces émotions sont inhérentes à la relation aux autres, comme la honte, la culpabilité, la jalousie, la timidité, l'humiliation, . . . etc.

Ces émotions sont constituées à partir des émotions primaires et bien sûr l'éducation et la culture ont un grand rôle sur l'acquisition des émotions sociales. (Tableau 1.2) [3].

Émotion	Déclencheurs et circonstances d'apparition	Comportement
Joie	Désir, Réussite, Bien-être, Accomplissement	Approche
Tristesse	Perte, Deuil	Repli sur soi
Colère	Obstacle, Injustice, Dommage, Atteindre à son intégrité physique ou psychique Limites de la personne, Atteinte au système de valeurs	Attaque
Peur	Menace, Danger, Inconnu	Fuite Sidération Évitement Parfois attaque
Dégoût	Substance ou personne nuisible, Aversion physique ou psychique, Contre quelqu'un, Rejet	
Surprise	Danger immédiat , Inconnu , Imprévu	Retrait Sursaut

TABLEAU 1.2 – Les 6 émotions de bases

1.2.2 La Classification et la Représentation des émotions

On peut classer les émotions selon deux grandes approches, l'approche catégorielle et l'approche dimensionnelle.

a. L'approche catégorielle :

L'approche catégorielle des émotions prend en considération les 6 émotions de bases et essaye de faire la distinction entre ces émotions et en plus de ça elle propose de les classer sous forme de catégories discrètes (Figure 1.1) "selon Le circumplex de Plutchik ". L'intérêt principal de l'approche catégorielle est qu'une fois que les émotions à traiter sont clairement identifiées, il devient simple de les manipuler, aussi bien pour les hommes que pour les machines [5].

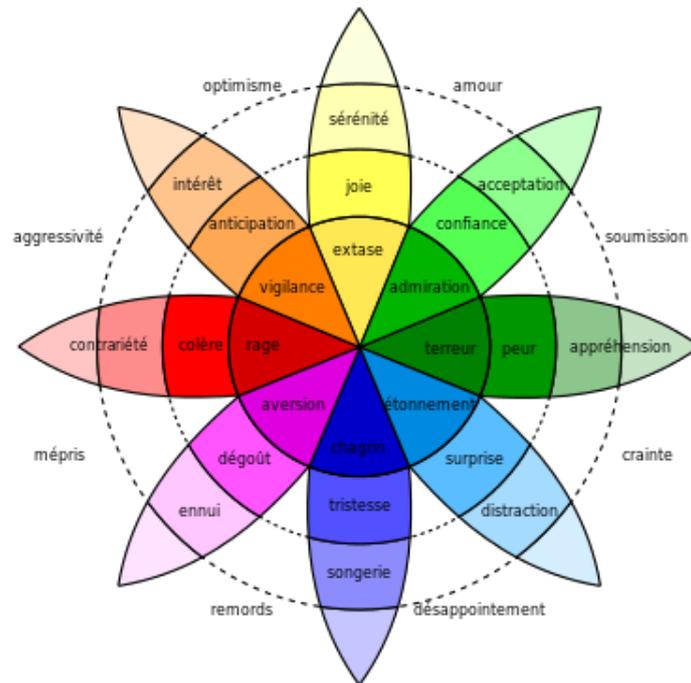


FIGURE 1.1 – L'approche catégorielle

b. L'approche dimensionnelle :

Cette approche s'est développée en parallèle avec l'approche catégorielle dans le but de compenser aux défauts de cette dernière. L'approche dimensionnelle propose la représentation des états émotionnels sur des axes abstraits : la valence et l'arousal, la dimension de valence est la plus utilisée. Elle consiste en la description des états émotionnels sur un axe positif/négatif. Ainsi l'activation représente le niveau d'excitation corporelle qui s'exprime par des réactions physiologiques, comme l'accélération du cœur, la transpiration....[5]. Pour bien définir les différentes émotions à l'aide de l'approche dimensionnelle on doit prendre en compte trois facteurs : la valence, le degré d'activation physiologique (ou l'arousal) et la dominance (contrôle). En général, deux dimensions principales sont mises en avant. D'une part, la valence émotionnelle,

c'est-à-dire le caractère positif ou négatif de l'expérience émotionnelle et, d'autre part, la dimension de l'intensité ou le degré d'activation de l'expérience émotionnelle (l'arousal). Donc une émotion positive est une émotion qui a une valence positive et une émotion négative c'est l'émotion qui a une valence négative. On plus le degré de l'émotion change selon l'arousal [3] Si on prend par exemple le contentement a une valence positive et un niveau d'arousal moyen par contre la colère à une valence négative et un très haut niveau d'arousal. Et donc à l'aide d'un schéma (Figure 1.2) on peut définir la qualité d'une émotion.

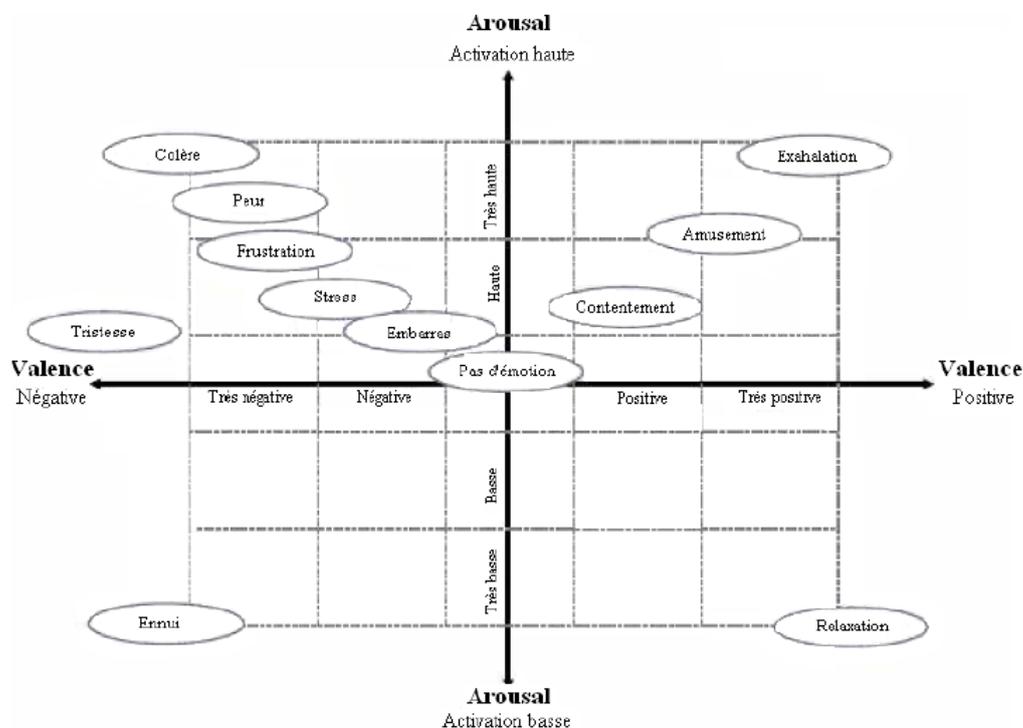


FIGURE 1.2 – Représentation des émotions selon deux axes : l'arousal et la valence.

c. Exemple d'émotion positive :

On a pris comme exemple la joie : c'est une émotion de satisfaction plus ou moins durable, et la vraie joie est profonde, elle vient du cœur, du corps tout entier. On sait si la personne qui est devant nous est heureuse à l'aide de son visage donc à l'aide de

son expression faciale. On voit plusieurs changements sur le visage et les indices les plus importants de la joie sont :

les yeux sont légèrement plissés c'est -à-dire que la paupière inférieure couvre en partie l'œil, et le mouvement du triangle infra orbital c'est-à-dire les joues sont remontées, formant des plis en dessous de l'œil, ont plus la partie basse de la joue est levée latéralement, et on voit aussi que le zygomatique travaille également lors de cette expression c'est-à-dire le coin des lèvres est de ce fait remonté obliquement et le menton aussi est étiré de manière oblique, et on peut ajouter que la personne peut monter ses dents si elle le désire. On peut comprendre mieux ces expressions à l'aide de la (Figure 1.3).



FIGURE 1.3 – L'expression faciale associée à la joie

d. Exemple d'émotion négative :

On a pris comme exemple la colère : la colère est considérée comme une émotion secondaire à une blessure physique ou psychique, un manque, une frustration. On connaît si la personne qui est devant nous est en colère à l'aide de l'expression faciale. On voit que la partie intérieure des sourcils est légèrement abaissée et l'œil est demi-ouvert et la partie en dessous de l'œil est ridée. En plus le nez est retroussé, des rides

apparaissent ainsi au niveau des joues et les narines sont étirées et élargies, le menton remonte un peu et on voit aussi que la bouche est fermée. On peut remarquer que la lèvre supérieure remonte également. (Figure 1.4).

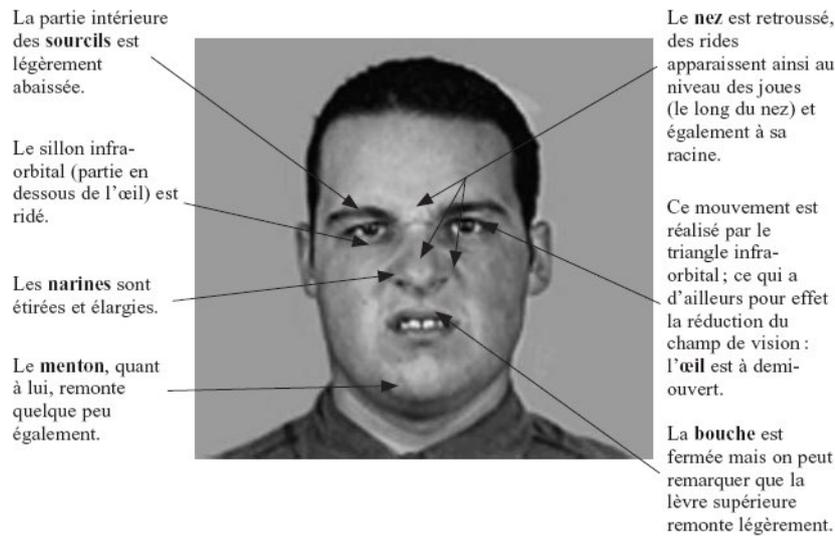


FIGURE 1.4 – L'expression faciale associée à la colère

1.2.3 Comment peut-on mesurer les émotions ?

On peut obtenir l'état émotionnel des humains à partir de plusieurs indices qui sont classés selon trois ensembles (Figure 1.5)

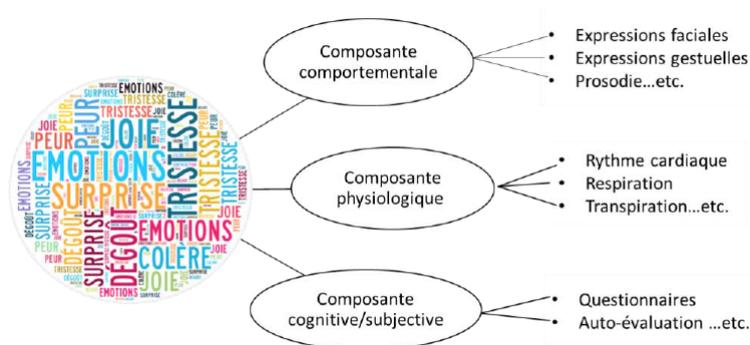


FIGURE 1.5 – Quelques composantes des émotions.

a. La composante physiologique :

On peut traiter des expériences émotionnelles par la composante physiologique à l'aide de plusieurs facteurs comme le rythme cardiaque, la respiration et la transpiration. Cette méthode permet au cerveau de préparer l'action car le système nerveux autonome (SNA) est responsable de la détection de ce qui se passe dans le corps et la régulation des réponses involontaires, y compris celles du cœur et les muscles qui contrôlent notamment la constriction des vaisseaux de sang, les voies respiratoires et le tractus gastro-intestinal [5].

Le SNA contient deux parties : le parasympathique et le sympathique qu'on définit comme suit :

La première partie qui est le parasympathique est responsable du ralentissement de la fréquence cardiaque et la détente des muscles lisses et la deuxième partie qui est le sympathique du SNA est responsable de l'inverse, à savoir augmenter le rythme cardiaque et la constriction des vaisseaux sanguins qui provoquent entre autres effets une augmentation de la pression artérielle. La réponse sympathique est plus lente et plus durable que la réponse parasympathique et qui est associée à la réaction se battre ou fuir. Le rôle de cette dernière est de mettre l'organisme en état d'alerte et de le préparer à l'activité. (Figure 1.6)

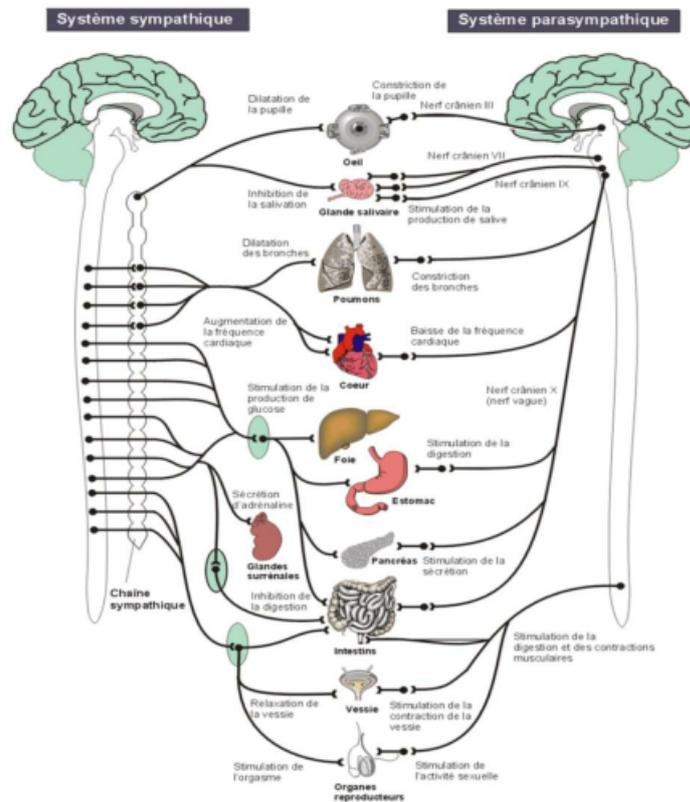


FIGURE 1.6 – Système nerveux autonome (SNA).

Différents capteurs permettant l'évaluation de ces changements et permettent de distinguer les différentes émotions. On va citer quelques'uns dans la suite.

- Température cutanée (SKT)
- Conductance de la peau (SKC)
- Volume respiratoire (VR)
- La pression sanguine volumique (Blood volume pulse, BVP)
- Le signal électromyographie (EMG)
- Le signal électrocardiographie (ECG)

On peut obtenir la fréquence ou la période cardiaque à partir des trois derniers signaux (BVP, EMG, ECG), il s'agit d'une importante corrélation physiologique avec l'émotion

vu que la fréquence cardiaque (FC) est contrôlée à la fois par la réponse parasympathique (FC décroissant) et la réponse sympathique (FC croissant) et donc toutes ces mesures physiologiques et bien d'autres contribuent à la reconnaissance de l'état émotionnel des individus. Elles sont utilisées dans la reconnaissance de l'émotion soit seule ou avec différentes combinaisons [5].



FIGURE 1.7 – Quelques capteurs utilisés dans la littérature.

Dans notre travail on ne s'est pas intéressé à cette composante car on n'a pas les moyennes (les capteurs et le vrai espace de travail) pour faire cette expérience et aussi parce que cette dernière peut ne pas être fiable car : peut-être les capteurs et les matériaux utilisés ne fonctionnent pas bien ou bien les candidates peuvent avoir peur juste en voyant les capteurs.

b. La composante cognitive :

Cette méthode se base sur le ressenti de la personne et sa perception de la situation. Il existe des échelles permettant de mesurer le ressenti émotionnel, notamment le SAM (Self Assesment Manikin) utilisé pour l'évaluation des états émotionnels phasiques et le VAS (Visual Analog Scale) qui permet d'évaluer les états émotionnels toniques. Des questionnaires contenant plusieurs items sont utilisés pour reconnaître l'état émotionnel.

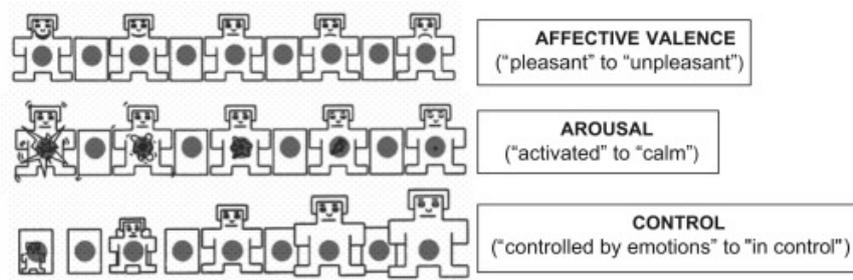


FIGURE 1.8 – Echelle d'auto-évaluation (SAM).

Cette méthode est subjective, les sujets peuvent mal évaluer leurs émotions on n'a pas les moyennes pour savoir s'ils disent la vérité ou pas.

c. La composante comportementale

On peut révéler le type des émotions à l'aide d'un ensemble de traits comportementaux tels que les expressions faciales, l'intonation de la voix, la posture... etc, qui génèrent un langage détectable par les autres individus et facilitent notre vie quotidienne avec les autres. Cette méthode est la plus célèbre et la plus facile à étudier car elle a un grand intérêt pour les chercheurs [5].

d. L'expression faciale :

Une émotion implique généralement une expression faciale correspondante mais l'inverse n'est pas vrai : il est possible de mimer une expression représentant une émotion sans pour autant ressentir cette émotion. Alors que les expressions dépendent des individus et des cultures, on distingue généralement un nombre limité d'émotions universellement reconnues.

Les visages véhiculent des informations riches qui constituent deux catégories, la partie des indices de l'identité individuelle et la partie des expressions de communication (verbale et non verbale), d'intentions et d'émotions entre individus, via, en particulier, la

direction du regard et les expressions faciales. L'être humain est aussi particulièrement doué pour reconnaître les émotions associées à des expressions faciales et donc il peut communiquer avec d'autres personnes présentes d'une façon beaucoup plus rapide qu'avec le langage. Ceci lui permet entre autres de donner aux personnes de son entourage des feedback sur leurs actions, pour qu'elles puissent savoir de quelle façon celles-ci sont perçues, et ainsi de modifier si nécessaire leur projet d'origine [3, 4]. (Figure 1.9).



FIGURE 1.9 – Les expressions faciales des six émotions de base selon Ekman [4]

L'essentiel de l'information d'une expression faciale est contenue dans la déformation des traits permanents principaux du visage, caractérisée par un changement, perceptible visuellement et donc cette déformation est due à l'activation volontaire ou non de l'un ou de plusieurs des 44 muscles composant le visage. Il émet en permanence des signes dont le décodage, pas seulement renseigne sur l'état émotionnel de cette personne, mais aussi éclaire sur ce qui est dit. Aujourd'hui, l'analyse assistée par ordinateur du visage et de ses expressions est un domaine émergent et les applications sont nombreuses comme l'informatique affective [6].

Cette méthode a un grand succès dans le monde car elle est facile à étudier mais elle a aussi des inconvénients car dans les bases des données on fait l'étude sur les acteurs

qui peuvent mentir dans c'est juste des expressions faciales sans émotions et sans sentiment.

e. La reconnaissance automatique des émotions :

Le visage étant la partie la plus expressive et communicative d'un être humain, il représente un centre d'intérêt majeur dans les recherches actuelles concernant l'amélioration de l'interaction Homme-Machine (IHM) pour l'établissement d'un dialogue entre les deux entités.

La reconnaissance des expressions faciales s'appuie sur l'analyse des changements des traits permanents du visage. Les mouvements de la tête et des traits mobiles du visage (clignements, bâillement, suivie du regard. . .) sont aussi des informations intéressantes à exploiter pour la reconnaissance des expressions faciales.

Pour analyser les expressions faciales à l'aide des systèmes automatiques, il faut passer généralement par trois étapes principales :

La première étape consiste à détecter le visage, qui permet de limiter la zone de recherche et ensuite, l'extraction des informations nécessaires qui décrivent au mieux l'expression et à la fin, en se basant sur ces informations, l'image sera affectée à une catégorie d'expressions à l'aide d'un classifieur.(Figure 1.10) [3].

1. Détection de visage à étudier.
2. Extraction de composants importantes (déformations).
3. Classification de l'expression faciale pour avoir l'expression faciale.

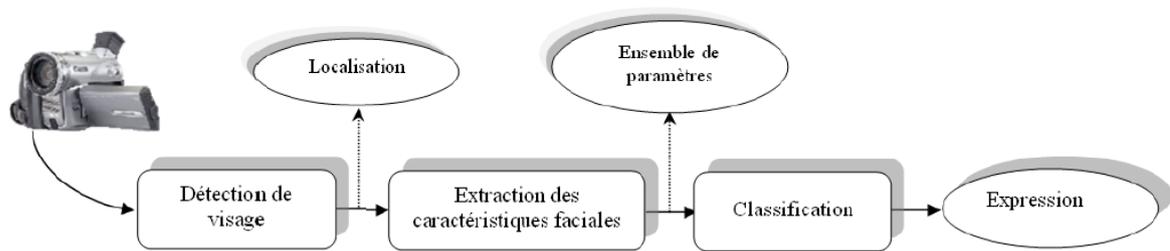


FIGURE 1.10 – Les différentes étapes d’un système de reconnaissance des expressions faciales [3]

1.3 Apprentissage automatique (Machine learning)

La Machine Learning est une méthode d’intelligence artificielle permettant aux ordinateurs d’apprendre sans avoir été programmés explicitement à cet effet et pour apprendre et se développer, les ordinateurs ont besoin de données à analyser et sur lesquelles s’entraîner et il s’agit d’une science moderne permettant de découvrir des patterns et d’effectuer des prédictions à partir de données en se basant sur des statistiques, sur du forage de données, sur la reconnaissance de patterns et sur les analyses prédictives.

Cette technique a été créée pour résoudre les problèmes pour lesquels les modèles analytiques sont difficilement disponibles. L’idée principale de Machine Learning est de réaliser un modèle en utilisant les données d’apprentissage lorsque les équations et les lois ne sont pas prometteuses.

La Machine Learning est très efficace dans les situations où les insights doivent être découverts à partir de larges ensembles de données diverses et changeantes, c’est-à-dire : le Big Data. Pour l’analyse de telles données, la Machine Learning se révèle nettement plus efficace que les méthodes traditionnelles en termes de précision et de vitesse.[13]

1.3.1 Quelques différents types et méthodes de l'apprentissage profond.

De nombreuses techniques d'apprentissage automatique ont été développés pour résoudre des problèmes dans divers domaines. Ces techniques d'apprentissage automatique peuvent être classées en trois types en fonction de la méthode d'apprentissage (Figure 1.11).

a. L'apprentissage supervisé :

L'apprentissage supervisé (supervised learning en anglais) est une tâche d'apprentissage automatique consistant à apprendre une fonction de prédiction à partir d'exemples labellisés. On distingue les problèmes de régression des problèmes de classification. Ainsi, on considère que les problèmes de prédiction d'une variable quantitative sont des problèmes de régression tandis que les problèmes de prédiction d'une variable qualitative sont des problèmes de classification [7, 8] (Figure 1.11).

b. L'apprentissage non supervisé :

L'apprentissage non supervisé consiste à apprendre sans superviseur. Il s'agit d'extraire des classes ou groupes d'individus présentant des caractéristiques communes. La qualité d'une méthode de classification est mesurée par sa capacité à découvrir certains ou tous les motifs cachés.

On distingue l'apprentissage supervisé et non supervisé. Dans le premier apprentissage, il s'agit d'apprendre à classer un nouvel individu parmi un ensemble de classes prédéfinies : on connaît les classes a priori. Tandis que l'apprentissage non supervisé, le nombre et la définition des classes n'étant pas données à priori [7, 8] (Figure 1.11).

c. L'apprentissage par renforcement :

Ce type d'apprentissage doit apprendre, à partir d'expériences et ce qu'il convient de faire en différentes situations, de façon à optimiser une récompense quantitative au cours du temps.

Un paradigme classique pour présenter les problèmes d'apprentissage par renforcement consiste à considérer un agent autonome, plongé au sein d'un environnement, et qui doit prendre des décisions en fonction de son état courant. En retour, l'environnement procure à l'agent une récompense, qui peut être positive ou négative.

L'agent cherche, au travers d'expériences itérées, un comportement décisionnel (appelé stratégie ou politique, et qui est une fonction associant à l'état courant l'action à exécuter) optimal, en ce sens qu'il maximise la somme des récompenses au cours du temps (Figure 1.11).

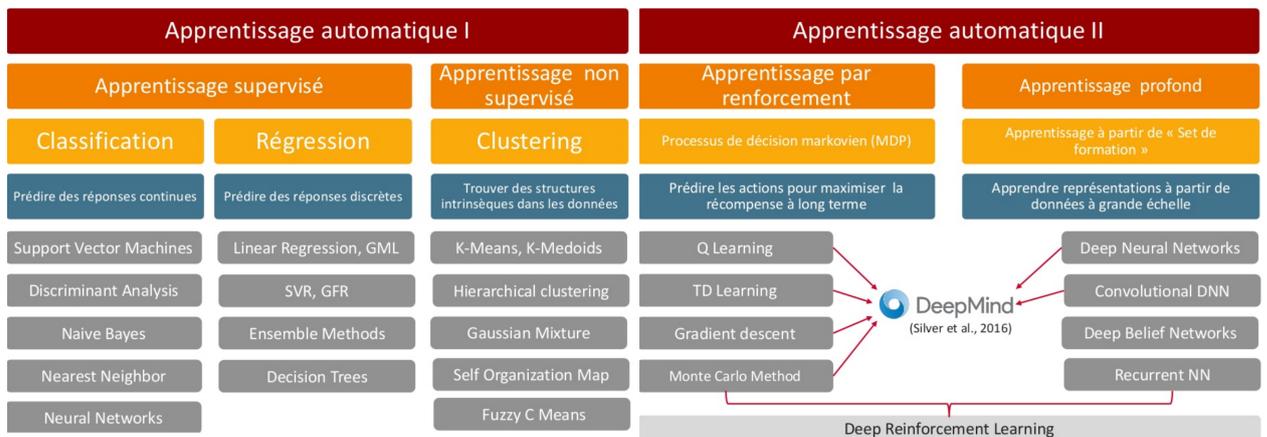


FIGURE 1.11 – Les différents types de la machine learning [8].

1.3.2 Les réseaux de neurones une méthode d'apprentissage supervisé :

Définition.

Un réseau de neurones artificiels est une technique de calcul qui se base sur le fonctionnement des neurones. Les neurones biologiques (ou neurones formels) opèrent mécaniquement à partir de règles précises. Si les entrées dépassent un certain seuil de référence, les neurones émettent un signal. Les réseaux de neurones artificiels sont par exemple utilisés dans l'informatique.

Les poids entre les unités sont les principaux moyens de stockage de l'information à long terme dans les réseaux de neurones. La mise à jour des poids est la principale façon dont le réseau de neurones apprend de nouvelles informations. Le réseau de neurones artificiels capable d'apprentissage, capables de mémoriser l'information, et capable de traiter des informations et donc un réseau neuronal s'inspire du fonctionnement des neurones biologiques et prend corps dans un ordinateur sous forme d'un algorithme. Le réseau neuronal peut se modifier lui-même en fonction des résultats de ses actions, ce qui permet l'apprentissage et la résolution de problèmes sans algorithme, donc sans programmation classique.

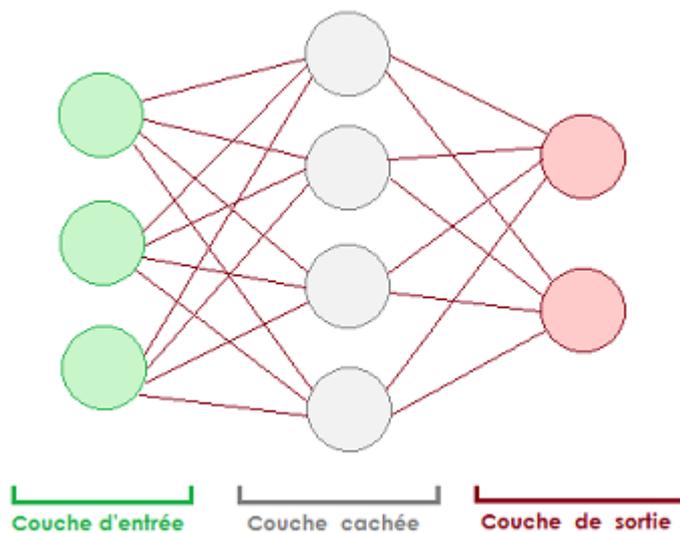


FIGURE 1.12 – Exemple de réseaux de neurones avec une couche cachée.

C'est quoi le deep learning ?

Deep Learning est une technique d'apprentissage automatique qui utilise réseau neuronal profond et bien sur que le réseau neuronal profond est le multicouche réseau de neurones contenant deux couches masquées ou plus. L'apprentissage profond se situe à la place du produit final de l'apprentissage automatique, et la règle d'apprentissage devient l'algorithme qui génère le modèle (le réseau neuronal profond) à partir des données d'apprentissage. En général, la raison pour laquelle le réseau neuronal multicouche a pris 30 ans pour résoudre les problèmes du réseau neuronal monocouche était l'absence de la règle d'apprentissage, qui a finalement été résolue par l'algorithme de rétropropagation(voir 2.2.2 b.). En revanche, la raison pour laquelle des autres 20 années se sont écoulées jusqu'à l'introduction de Deep Learning basé sur les réseaux neurones profonds était la mauvaise performance. L'entraînement de rétropropagation avec les couches cachées supplémentaires a souvent entraîné des performances plus médiocres. Deep Learning a fourni une solution à ce problème [7].

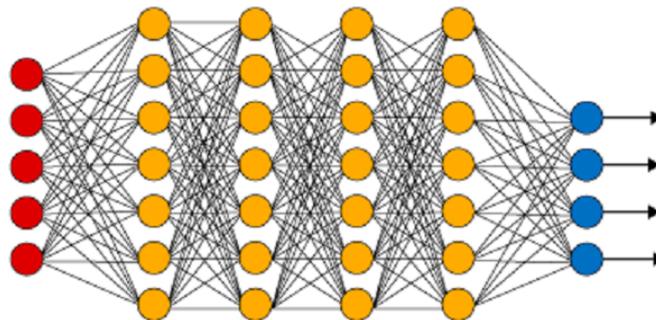


FIGURE 1.13 – Exemples d'un réseaux de neurones profonds

Méthodologie

Introduction

Dans ce chapitre on va parler sur la méthode de travail, comment on a préparé la base de données ? et quelles sont les bibliothèques et la forme de l'architecture du réseau de neurones (couches cachées) utilisé ? quelles sont les entrées et les sorties de réseau ? Toutes ces questions seront répondues en détail.



FIGURE 2.1 – Organigramme représentant les étapes du système de reconnaissance.

2.1 Base de données radboud faces (RaFD) :

Pour notre expérience, nous avons utilisé la base de données radboud faces (RaFD) [10]. Cette base de données créée par l'équipe de recherche à l'université radboud aux Pays-Bas, RaFD est une ensemble d'images de 67 modèles affichant 8 expressions émotionnelles (Figure 2.2) avec 3 directions de vue (droite, gauche et frontal) et 5 caméras à angle différent (Figure 2.3).

Modèle de RaFD :

- 20 hommes caucasiens.
- 19 femmes caucasiennes.

2 Méthodologie

- 4 garçons caucasiens.
- 6 filles caucasiennes.
- 18 hommes marocains.

Les expressions faciales de RaFD représenté sur la figure suivante :

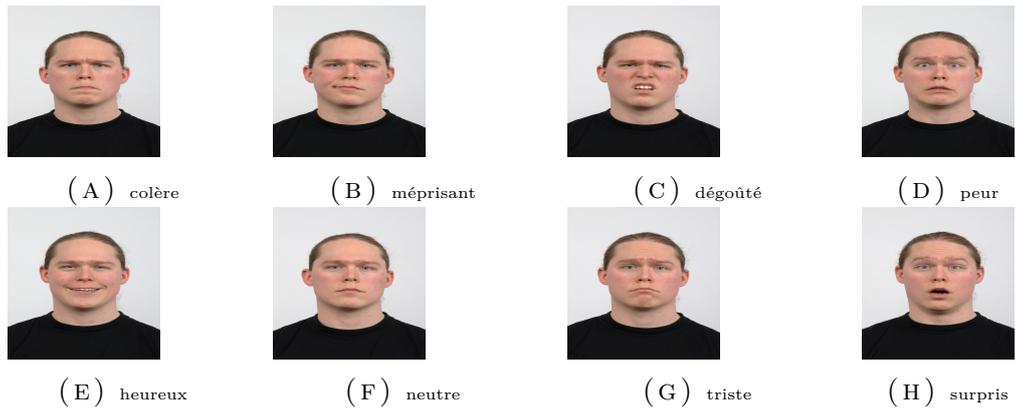


FIGURE 2.2 – Les expressions faciales représentées dans la base RaFD

Les différents angles des caméras :

Les caméras sont situées à différents angles, la figure suivante nous montre leurs emplacements et des exemples des photos prises par ces dernières.



FIGURE 2.3 – Emplacement des caméras et des images prises par ces caméras.

Direction de regard :

Chaque modèle a été montrée une expression avec trois directions du regard différent sans changer l'orientation de la tête.



(A) droite



(B) gauche



(C) frontal

FIGURE 2.4 – Direction de regard pris par caméra de 90.

Résolution des images sont 681×1024 pixels, et les noms des images de RaFD sont nommé de manière suivante :

- Le nom la base de données.
- Référence de l'angle de caméra
- Numéro de modèle.
- Ethnicité du modèle.
- Sexe du modèle.

- émotion.
- Direction de regard.

a. Prétraitement de la base de données :

Avant d'utiliser la base de données dans notre expérience, nous changeons la base pour s'adapter à notre réseau de neurones, pour cela nous suivrons ces étapes :

1. Sélection des émotions :

nous choisissons les images de direction du regard frontal de la caméra d'angle de 90° avec 3 expressions (neutre, colère, heureux).

2. Détection de visage :

Dans cette étape nous intéressons au le visage à l'aide de openCV¹.

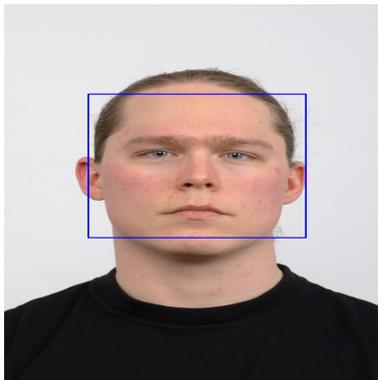


FIGURE 2.5 – Détection de visage à l'aide d'openCV.

Après avoir la reconnaissance faciale, nous avons recadré les images et nous changeons la couleur des images en gris.

1. Open Computer Vision (OpenCV) <http://opencv.org> : est une bibliothèque pour traitement d'images en temps réel, utilisé dans vision par ordinateur et l'apprentissage automatique.



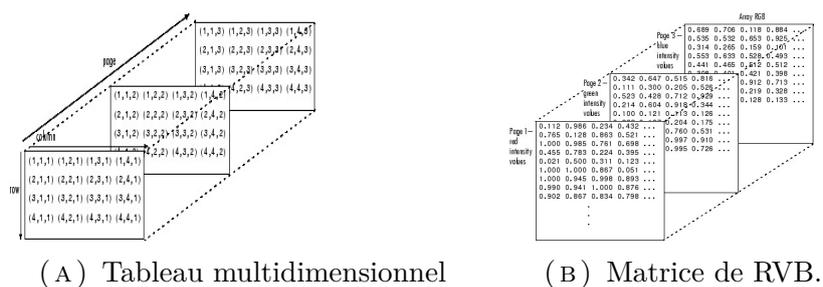
FIGURE 2.6 – L'image obtenue après recadrage.

Pourquoi l'image en gris ?

L' image colorée contient beaucoup de donnée qui ne peut pas être nécessaires pour notre entraînement, et l'image colorée est composée de 3 couches (R : Rouge, V : Vert et B : Blue) comme il est indiqué la (Figure 2.7c), c'est un tableau tridimensionnel² (Figure 2.7a), l'image grise est de seulement 2 dimensions, et les valeurs vont de 0 à 255 (entiers non signés de 8 bits)[11].

2. tableau tridimensionnel utilise les indices supplémentaires pour indexation :

- la ligne (the row)
- la colonne (the column)
- page pour représenter la dimension de tableau [14].



(C) Les couches d'image RVB.

FIGURE 2.7 – Image RVB avec Tableau multidimensionnel

3. Dimension des images :

Les images obtenues à partir de l'étape 2 leur résolution des images est 389×389 , donc on a redimensionné à 48×48 pixels puis on a enregistré les images sur le nom "subject_numéro de modèle_émotion".



FIGURE 2.8 – Image 2.8a à résolution 389×389 , image 2.8b à résolution 48×48

Cette résolution (48×48) nous garantit que l'émotion est claire et visible, et le choix de la résolution inférieure en raison de la taille du vecteur de pixel utilisé pour l'entraînement dans le réseau de neurones.

Après la base de données est prête. Maintenant, dans la section suivante nous allons parler du réseau de neurones que va utiliser.

2.2 Réseaux de neurones :

Dans les dernières années le réseau de neurones de neurones est devenu un sujet de recherche scientifique phare, le réseau de neurones est largement utilisé comme un modèle pour machine learning, dans cette partie on va parler sur l'architecture, et le réseau qui est utilisé dans traitement des images.

Pour mieux comprendra le mécanisme du réseau de neurones, considérons un nœud qui reçoit trois entrées (Figure 2.9).

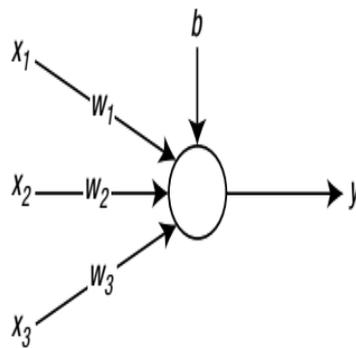


FIGURE 2.9 – Un nœuds qui reçoit trois entrées.

Le cercle indique le nœud et la flèche indique le signal.

x_1, x_2, x_3 : les entrées de signal.

w_1, w_2, w_3 : sont les poids pour les signaux correspondants.

b : biais.

y : sortie de nœud.

Avant le signal d'entrée atteint le nœud nous avons multiplié par le poids.

$$v = w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3 + b \quad (2.2.1)$$

L'équation de somme pondérée (2.2.1) sous forme matricielle devient :

$$v = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + b \quad (2.2.2)$$

La sortie du nœud y doit être activée par la fonction d'activation (voir [section 2.2.1](#)) notée par $\varphi(\cdot)$.

$$y = \varphi(v) \quad (2.2.3)$$

Apprentissage supervisé d'un réseau neuronal, ce qui signifie que le réseau compare la sortie du réseau avec la sortie désirée c'est-à-dire calculer l'erreur et ajuster les poids pour réduire l'erreur. La [Figure 2.10](#) résume le concept d'apprentissage supervisé, pour les calculs d'erreur, nous utilisons des fonctions d'optimisation (plus de détails voir [section 2.2.2](#)).

La procédure d'apprentissage supervisé suit l'algorithme suivant :

1. Initialiser les poids.
2. Prenez la "entrée" des données d'entraînement, et entrez-la dans le réseau de neurones, obtenir la sortie du réseau neuronal et calculer l'erreur à partir de la sortie désirée.

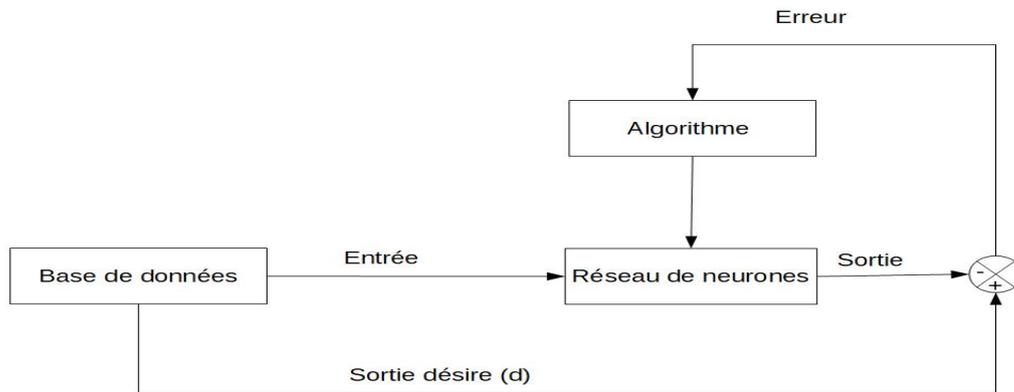


FIGURE 2.10 – Le concept d'apprentissage supervisé.

$$e = d - y \quad (2.2.4)$$

où : d : entrée, et y : sortie désirée.

3. Ajustez les poids pour réduire l'erreur (chaque fonction a des méthodes d'ajuster les poids (2.2.2)) .
4. Répétez les étapes 2-3 pour toutes les données d'entraînement.

Les réseaux de neurones se composent de de 3 couches :

— Couche d'entrée :

C'est une première couche dans le réseau de neurones, où les nœuds de la couche d'entrée transmettent les signaux d'entrée aux nœuds de la couche suivants, par conséquent ils ne calculent pas la somme pondérée.

— Couches cachées :

Ce sont des couches non visibles où leurs entrées sont les sorties de la couche d'entrée et leurs sorties sont l'entrée de la couche de sortie, les nombres de couches

cachées définissent l'architecture de réseau. La figure (2.11) représente différentes architectures.

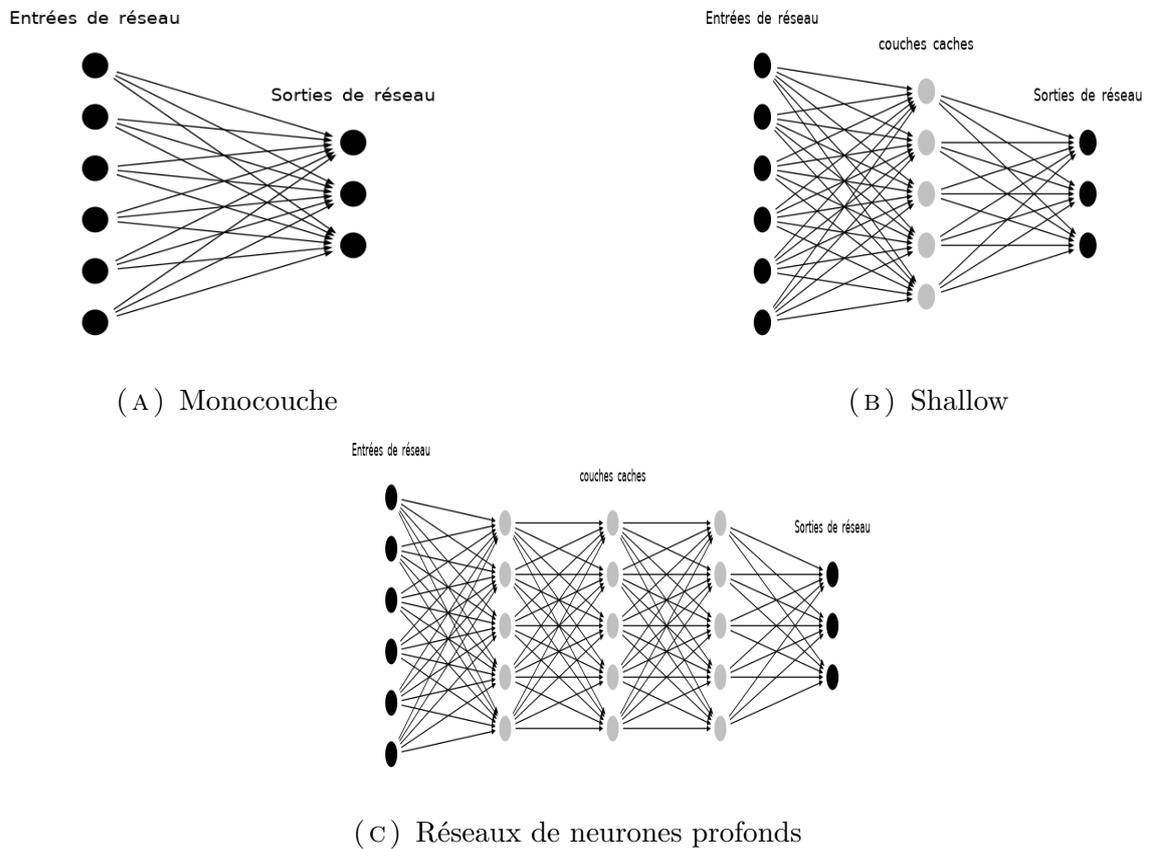


FIGURE 2.11 – Différentes architectures du réseau de neurones

— Couche de sortie :

C'est la dernière couche de réseau, la sortie de ces nœuds devient le résultat final du réseau de neurones.

2.2.1 Fonctions d'activation :

Les fonctions d'activation ou de transfert sont des fonctions mathématiques qui établissent des bornes pour la sortie des neurones, les réseaux de neurones peuvent

utiliser plusieurs fonctions d'activation différentes. il existe beaucoup de fonction d'activation nous nous sommes intéressé que par les fonctions les plus utilisées.

a. Sigmoid

Sigmoid est définie :

$$\varphi(x) = \frac{1}{1 + e^x} \quad (2.2.5)$$

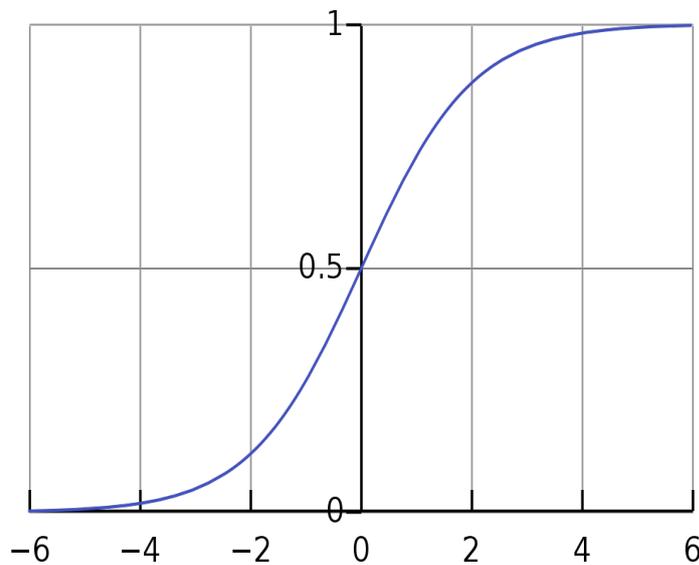


FIGURE 2.12 – La fonction Sigmoid.

La sortie de sigmoid est très lisse et donne de bons dérivés continus, ce qui fonctionne bien pendant l'entraînement d'un réseau de neurones, la fonction sigmoid utilisée pour produire une probabilité pour une classification binaire.

b. Softmax :

Softmax est définie :

$$\varphi(v_i) = \frac{e^{v_i}}{\sum_{k=1}^M e^{v_k}} \quad (2.2.6)$$

où :

- $v_i = w_i \times x$: somme des poids du i-ème nœud de sortie.
- M : nombre du nœud de sortie.

La fonction softmax utilisée dans couche de sortie pour multi-classification en conjonction avec la fonction d'entropie croisée (2.2.2 i.), et représente la probabilité pour chaque classe de multi-classification.

c. Unité Linéaire Rectifié (ReLU) :

La fonction Unité Linéaire Rectifié (en anglais : Rectified Linear Unit) renvoie 0 pour les entrées négatives et conserve la même valeur pour les entrées positives, elle est définie :

$$\varphi(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad (2.2.7)$$

$$\varphi(x) = \max(0, x)$$

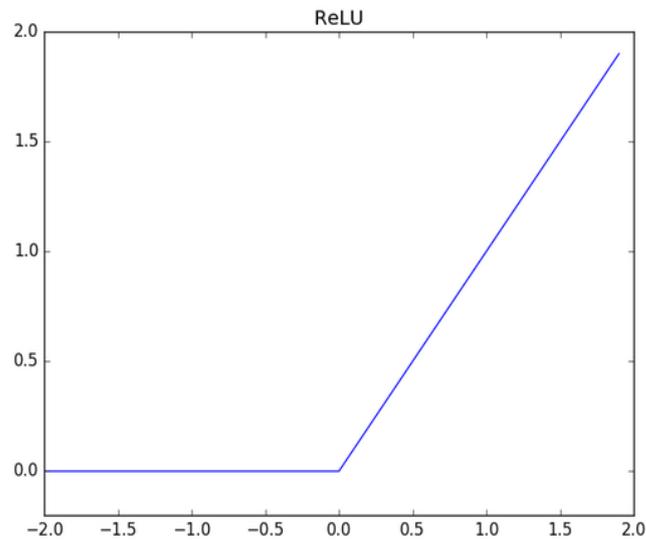


FIGURE 2.13 – La fonction ReLU.

Relu est plus utilisée dans les couches cachées, car elle est linéaire et non bornée.

2.2.2 Fonctions d'optimisation :

Les fonctions d'optimisations sont des fonctions mathématiques utilisées pour calculer l'erreur et mettre à jour les poids pour que l'erreur devienne minimale.

a. Règle delta :

Cette règle est utilisée dans les réseaux monocouches [Figure 2.11a](#), l'algorithme de règle delta :

1. Initialiser le poids.
2. Calculer l'erreur par la formule suivante :

$$e_i = d_i - y_i$$

où : d_i : est la sortie correcte i , et y_i : sortie désirée de nœud i .

3. Calculer les mises à jour de poids selon les critères suivants :

$$\Delta w_{ij} = \alpha e_i x_j$$

où :

α : taux d'apprentissage ($0 < \alpha \leq 1$) déterminer le changement de poids par le temps.

- si la valeur est trop élevée, la sortie ne converge pas.
- si la valeur est trop faible, le calcul atteint la solution trop lente.

x_j : la sortie d'entrée de nœud j .

w_{ij} : le poids entre la sortie de nœud i et l'entrée de nœud j .

4. Ajustez les poids :

$$w_{ij} \leftarrow w_{ij} + \Delta w_{ij}$$

5. Effectuez les 2 à 4 pour toutes les données d'entraînement.
6. Répétez les étapes 2 à 5 jusqu'à ce que l'erreur atteigne un niveau de tolérance acceptable.

b. Back-propagation :

Algorithme de back-propagation calcule l'erreur et met à jour les poids à partir de la couche de sortie du réseau vers les couches cachées comme l'algorithme suivant :

1. Initialiser le poids.
2. Calculer le premier delta δ .

$$e = d - y$$

$$\delta = \varphi'(v)e$$

3. Propager vers l'arrière et calculer les deltas des prochains nœuds.

$$e^{(k)} = w^T \delta$$

$$\delta^{(k)} = \varphi'(v^{(k)})e^{(k)}$$

k : numéro de couche cachée.

4. Répéter l'étape 3 jusqu'à atteindre toutes les couches cachées.

5. Ajuster les poids suivant :

$$\Delta w_{ij} = \alpha \delta_i x_j$$

$$w_{ij} \leftarrow w_{ij} + \Delta w_{ij}$$

6. Répétez les étapes 2 à 5 pour toutes les données d'entraînement.

7. Répétez les étapes 2 à 6 jusqu'à ce que le réseau de neurones soit correctement entraînés.

c. Momentum :

Momentum notée m est un terme ajouté à la règle delta (a.) pour ajuster le poids :

$$\Delta w = \alpha \delta x$$

$$m = \Delta w + \beta m^-$$

$$w = w + m$$

$$m^- = m$$

où :

m^- : momentum précédent.

β : est une constante positive $\beta < 1$. Les étapes suivantes montrent comment momentum change avec le temps :

$$m(0) = 0$$

$$m(1) = \Delta w(1) + \beta m(0) = \Delta w(1)$$

$$m(2) = \Delta w(2) + \beta m(1) = \Delta w(2) + \beta \Delta w(1)$$

$$m(3) = \Delta w(3) + \beta m(2) = \Delta w(3) + \beta \{\Delta w(2) + \beta \Delta w(1)\}$$

⋮

d. Adam

Estimateur de moment adaptatif (Adam) est une technique d'optimisation qui a un taux d'apprentissage adaptatif pour chaque paramètre ou poids, Adam garde les carrés

des gradients passés.

$$m_{ij} = \beta_1 m_{ij} + (1 - \beta_1) \frac{\partial v}{\partial w_{ij}}$$
$$g_{ij} = \beta_2 g_{ij} + (1 - \beta_2) \left(\frac{\partial v}{\partial w_{ij}} \right)^2$$

m_{ij} s : taux de décroissance.

g_{ij} : gardien pour chaque poids w_{ij} . β_1 et β_2 sont des constants < 1 .

La moyenne normalisée des gradients \hat{m}_{ij} et la moyenne des gradients carrés \hat{v}_{ij} sont calculées comme suit :

$$\hat{m} = \frac{m}{1 - \beta_1}$$
$$\hat{v} = \frac{v}{1 - \beta_2}$$

La règle de mise à jour finale pour chaque poids w_{ij} est la suivante :

$$w_{ij} = w_{ij} - \frac{\eta}{\sqrt{\hat{v} + \epsilon}} \hat{m}$$

avec : η taux d'apprentissage. [11]

e. Algorithme du gradient stochastique (SGD) :

Algorithme du gradient stochastique (en anglais : the Stochastic Gradient Descent) calcule l'erreur pour chaque donnée d'entraînement et ajuste les poids immédiatement exemple si nous avons N points de données d'entraînement, SGD ajuste les poids N fois, la performance du réseau de neurones devient tordue pendant entraînement.

Le nom stochastique implique le comportement aléatoire du processus de l'entraînement.

Le SGD calcule les mises à jour de poids comme :

$$\Delta w_{ij} = \alpha \varphi'(v_i) e_i x_j$$

où :

α : taux d'apprentissage.

$\varphi'(v_i)$: dérivé de fonction d'activation où $\varphi'(v_i) = \varphi(v_i)(1 - \varphi(v_i))$.

x_j : la sortie d'entrée de nœud j .

w_{ij} : le poids entre la sortie de nœud i et l'entrée de nœud j .

f. Batch :

Dans la méthode batch, chaque mise à jour de poids est calculée pour toutes les erreurs des données d'apprentissage, et la moyenne des mises à jour de poids est utilisée pour ajuster les poids, cette méthode utilise toutes les données d'apprentissage et les mises à jour une seule fois.

Le batch calcule la mise à jour du poids comme suit :

$$\Delta w_{ij} = \frac{1}{N} \sum_{k=1}^N \Delta w_{ij}(k)$$

où :

$\Delta w_{ij}(k)$: mise à jour de poids pour k données d'entraînements.

N : nombre donnée d'entraînement.

le batch prend beaucoup de temps dans l'entraînement.

g. Mini Batch :

La méthode basée sur deux méthodes SGD et Batch, le mini batch sélectionne une partie de l'ensemble de données d'apprentissage et les utilise pour l'apprentissage par batch, utilisation de la vitesse de SGD il calcule les mises à jour de poids des données

sélectionnées et entraîne le réseau de neurones avec la moyenne des poids mises à jour.

h. Fonction de coût :

La fonction de coût est concept mathématique associé à la théorie de l'optimisation, lié à l'apprentissage supervisé du réseau de neurones, plus l'erreur du réseau neurones est grande plus la valeur de fonction de coût est élevée.

$$J = \frac{1}{2} \sum_{i=1}^M (d_i - y_i)^2 \quad (2.2.8)$$

Fonction de coût (2.2.8) est le carré de la différence entre la sortie du réseau y et la sortie correcte d , l'erreur devient nulle si $y = d$.

i. Entropie croisée :

L'équation (2.2.9) s'appelle l'entropie croisée.

$$E = -d_i \ln(y_i) - (1 - d_i) \ln(1 - y_i) \quad (2.2.9)$$

L'équation est contractée pour une expression plus simple :

$$E = \begin{cases} -\ln(y) & d = 1 \\ -\ln(1 - y) & d = 0 \end{cases} \quad (2.2.10)$$

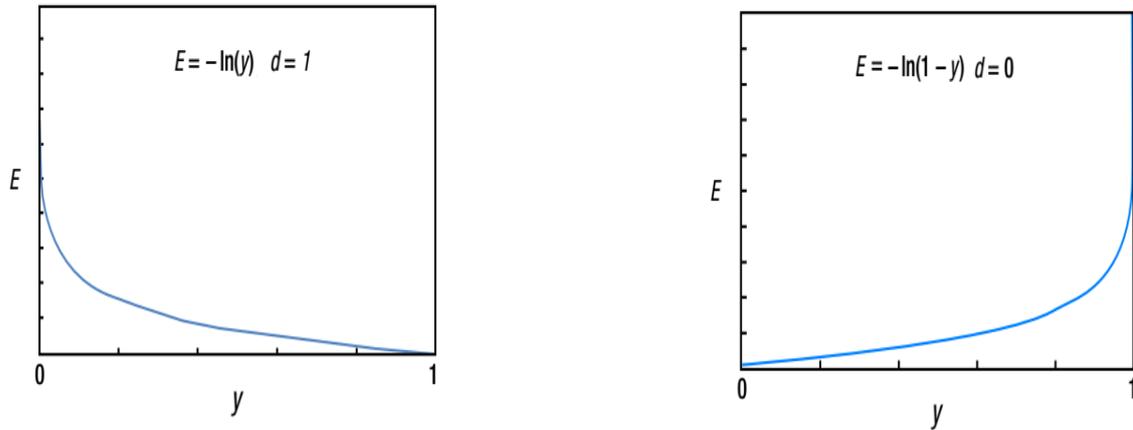
(A) Entropie croisée pour $d = 1$.(B) Entropie croisée pour $d = 0$.

FIGURE 2.14 – Entropie croisée.

La sortie $y = 1$ (Figure 2.14a), c'est-à-dire que l'erreur $d - y = 0$, la valeur de la fonction de coût est égale à 0, par contre lorsque la sortie $y \rightarrow 0$, c'est-à-dire que l'erreur augmente, la valeur de la fonction de coût augmente, donc cette fonction de coût est proportionnelle à l'erreur.

La Figure 2.14b la sortie $y = 0$ donc l'erreur $e = 0$, lorsque $y \rightarrow 1$ l'erreur augmente, dans ce cas la fonction coût est proportionnelle à l'erreur. On conclut que la fonction coût est très sensible à l'erreur.

j. Dropout :

Dropout consiste à sélectionner de façon aléatoire les nœuds pendant l'entraînement c'est-à-dire certains nœuds sont sélectionnés au hasard et leurs sorties sont mises à zéro pour désactiver les nœuds, la figure suivante explique comment cela fonctionne :

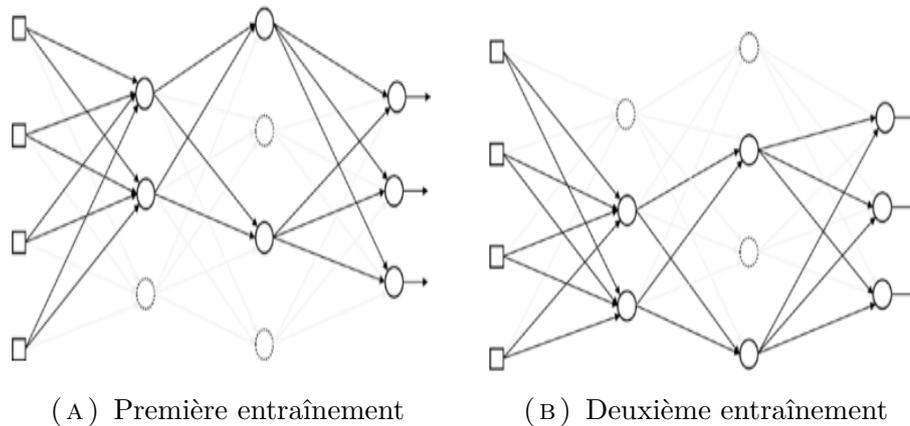


FIGURE 2.15 – Principe de dropout

2.2.3 Réseau de neurones convolutifs (ConvNet) :

Réseau de neurones convolutifs largement utilisé pour la reconnaissance d'image, cette technique importante pour l'amélioration des couches profondes pour le traitement de l'information (images).

La convolution de deux signaux nous donne un autre signal contenant une meilleure représentation caractéristique. Convolution à temps continue.

$$s(t) = (f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau$$

La [Figure 2.16](#) montre une architecture d'un réseau Convnet.

a. Couches convolutionnelles

L'objectif de cette couche est de détecter les caractéristiques telles que les bords, les taches de couleur et d'autres éléments visuels, la convolution d'image avec le filtre créé des images appelées cartes de caractéristiques de sortie, plus de filtres dans les couches de convolution plus des caractéristiques ont détecté.

L'opération de convolution commence dans le coin supérieur gauche de la sous-matrice

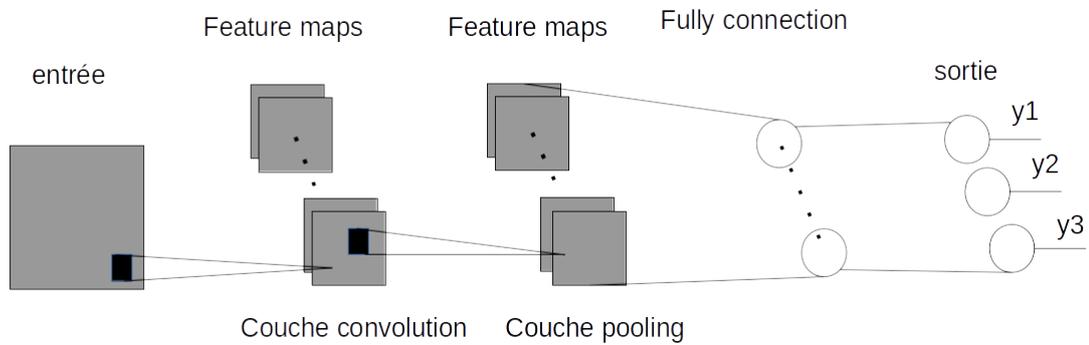


FIGURE 2.16 – Architecture d'un réseau convnet avec 3 nœuds de sortie.

qui est la même taille que le filtre de convolution, pour comprendre la procédure comment la convolution de l'image avec filtre nous considérons l'exemple suivant :

$$\begin{bmatrix} 1 & 3 & 5 \\ 6 & 7 & 8 \\ 2 & 4 & 9 \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 11 \\ 10 & 16 \end{bmatrix}$$

FIGURE 2.17 – Convolution entre image (3×3) et filtre 2×2 par bloc

La procédure de calcul :

$$(1 \times 1) + (3 \times 0) + (6 \times 0) + (7 \times 1) = 8$$

$$(3 \times 1) + (5 \times 0) + (7 \times 0) + (8 \times 1) = 11$$

$$(6 \times 1) + (7 \times 0) + (2 \times 0) + (4 \times 1) = 10$$

$$(7 \times 1) + (8 \times 0) + (4 \times 0) + (9 \times 1) = 16$$

b. Couches pooling

La couche pooling réduit la taille de l'image, car elle combine les pixels voisins d'une certaine zone de l'image en une seule valeur représentative, ces pixels voisins sont généralement sélectionnés à partir de la matrice carrée.

La valeur représentative est généralement définie comme la moyenne ou le maximum des pixels sélectionnés.

Maximum pooling : la valeur maximale de pixel est considérée comme une valeur représentative.

Moyenne pooling : la valeur moyenne des pixels est considérée comme une valeur représentative.

Procédure de calcul pooling :

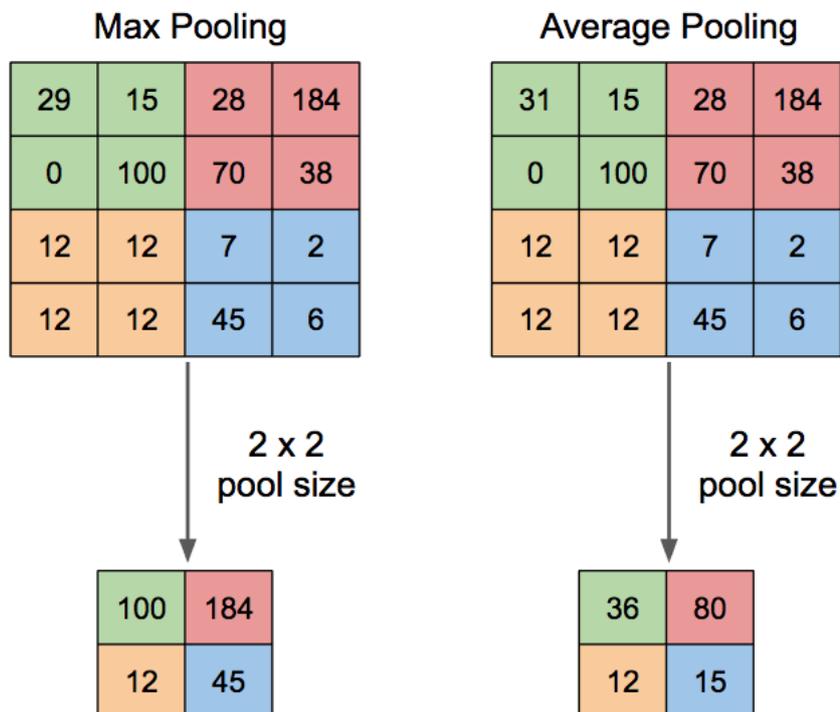


FIGURE 2.18 – Procédure de pooling
[15]

c. Couches fully connected

Les neurones de cette couche reçoivent les poids des couches précédentes, mais ne partagent pas le poids entre eux.

Chaque ce neurone sera connecté à tous les neurones de la couche précédente pour classification.

2.2.4 Les Bibliothèques utilisées :

Pour faire l'entraînement et le tester de base de données nous avons utilisé les bibliothèques suivantes :

a. Diverses interfaces pour le système d'exploitation(OS)

Diverses interfaces pour le système d'exploitation : cette bibliothèque permet d'utiliser les fonctionnalités dépendantes du système d'exploitation, par exemple ouvrir un fichier, définit le contenu d'un fichier sous forme de liste, ...etc. [16]

b. Numpy

Numpy : est un package fondamental pour les calculs scientifiques sous forme de tableaux [17].

c. Matplotlib

Matplotlib : est une bibliothèque graphique permet de tracer les graphes en 2D [18].

d. PIL

PIL(Python imaging library) : ce package contient des fonctions permet de gérer les images, par exemple ouvrir une image, obtenir des valeurs pixels, ...etc [19].

e. TensorFlow

TensorFlow : est une bibliothèque pour les calculs numériques à haute performance, utilisée beaucoup dans apprentissage automatique, publiée par Google [20].

f. Theano

Theano : est une bibliothèque permet à nous de définir, optimiser et évaluer des expressions mathématiques impliquant les tableaux [21, 22].

g. Scipy

Scipy : est une bibliothèque pour l'utilisation mathématique, sciences et ingénierie [23].

h. Scikit-learn

Scikit-learn : est une bibliothèque pour apprentissage automatique contient des outils simples et efficaces pour l'exploration et analyse des données, il est programmé pour les besoins de Scipy, Numpy et Matplotlib [24].

i. Keras

Keras : est une bibliothèque haute niveau programmée pour apprentissage profond, elle a été développée dans le but d'accélérer l'expérimentation (utilisation systématique des expériences) [25].

Pourquoi nous utilisons Keras ?

Voilà quelques arguments :

- Keras est une bibliothèque open source et écrite en python.
- Simple et facile à utiliser parce qu'il offre une fonctionnalité d'aide nous pour faire l'entraînement de base de données .
- Supporte multi-GPU et le calcul parallèle.
- Enregistrement de modèle d'entraînement pour faire le test.
- Keras est soutenu par les sociétés qui utilisent le deep learning comme Google, Microsoft, Nvidia, ...etc

Conclusion

D'après ce qu'on a vu dans ce chapitre :

Pour notre but de travail est la reconnaissance de la valence positive et négative en utilisent un réseau convnet profond de 8 couches cachées (3 couches de convolution, 3 couches pooling et 2 couches fully connected) de vecteur d'entrée de taille $[2304 \times 1]$ et vecteur de sortie $[3 \times 1]$. La [Figure 2.19](#) montre l'architecture utilisée dans l'entraînement. Pour l'entraînement de notre réseau, nous avons utilisé un ordinateur de caractéristiques suivantes :

- **Marque de pc** : Dell Inspiron 7567.
- **Processeur** : Intel i7-7700HQ.
- **Carte graphique** : Nvidia GeForce GTX 1050Ti 4GB GDDR5.
- **RAM** : 16GB (2x8GB) DDR4 .

Dans le chapitre suivant on va discuter les résultats d'entraînement et de test avec différentes fonctions d'activation et d'optimisation.

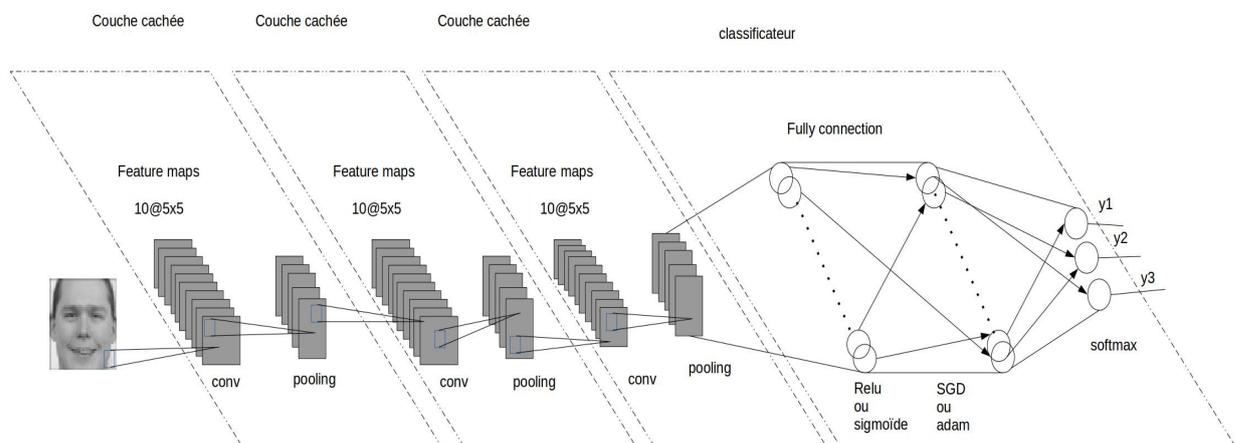


FIGURE 2.19 – Architecture du réseau utilisé.

Résultats et discussions

Dans ce mémoire, nous avons proposé un réseau de neurones pour la reconnaissance de la valence des émotions. Dans ce chapitre nous allons présenter les résultats de test de plusieurs mixtures entre les fonctions d'activations des couches cachées (sigmoïd et Relu) pour 500 et 1000 d'itérations. Ainsi qu'avec deux optimisateurs Adam et SGD (voir 2.2.2 d. et e.). Le programme utilisé dans notre expérience de base de donnée jaffe¹ se trouve sur plate-forme github².

Pendant l'entraînement nous considérons que biais est nulle, 50% de dropout et softmax comme une fonction d'activation de couche de sortie.

On a pris pour l'entraînement et le test 80% de la base (159 images) et 20% (42 images) respectivement.

1. Premier setup :

L'architecture de réseau utilisée pour l'entraînement indiqué dans la figure 3.1, on a utilisé la Sigmoid comme une fonction d'activation dans les couches cachées avec l'optimiseur SGD.

1. jaffe : est une base de données contient 7 expressions faciale des femmes japonaises

2. <https://github.com/PrudhviRaj12/Facial-Emotion-Detection-Using-Convolutional-Neural-Networks-and-Representational-Autoencoder-Units>

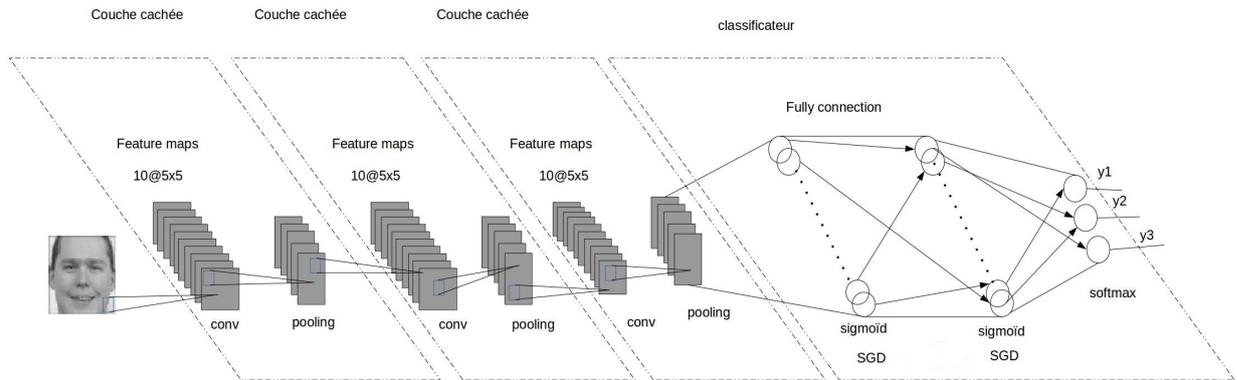


FIGURE 3.1 – Architecture utilisée pour l'entraînement.

a) Entraînement de réseau pour 500 itérations :

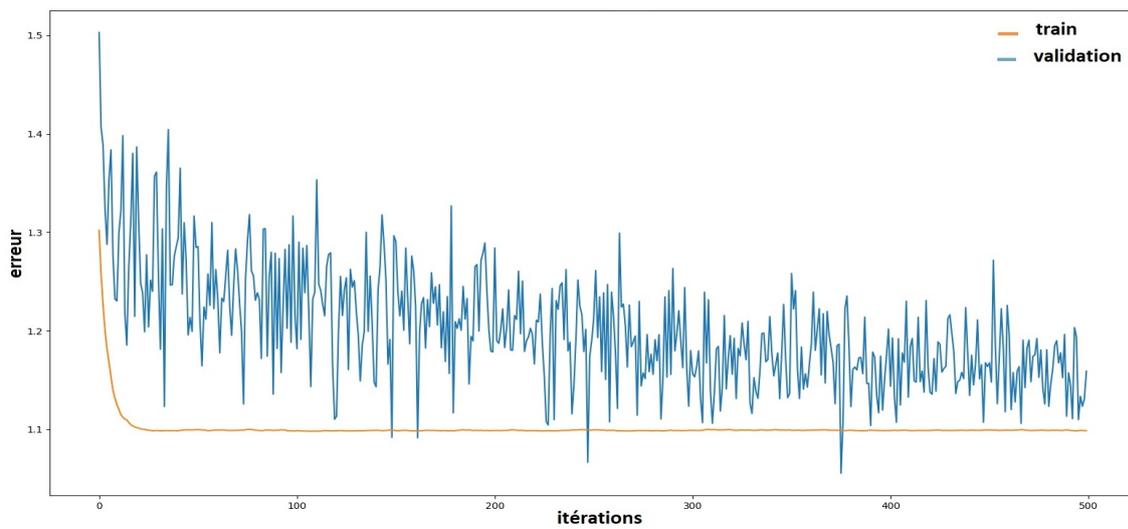


FIGURE 3.2 – Erreur d'entraînement de Sigmoid avec SGD pour 500 itérations.

Analyse de la courbe :

Nous remarquons que l'erreur de validation n'est pas nulle et reste stable en valeur 1.1 après 27 itérations, au contraire de l'erreur faite dans l'entraînement n'est pas stable et varie fortement.

b) Entraînement de réseau pour 1000 itérations :

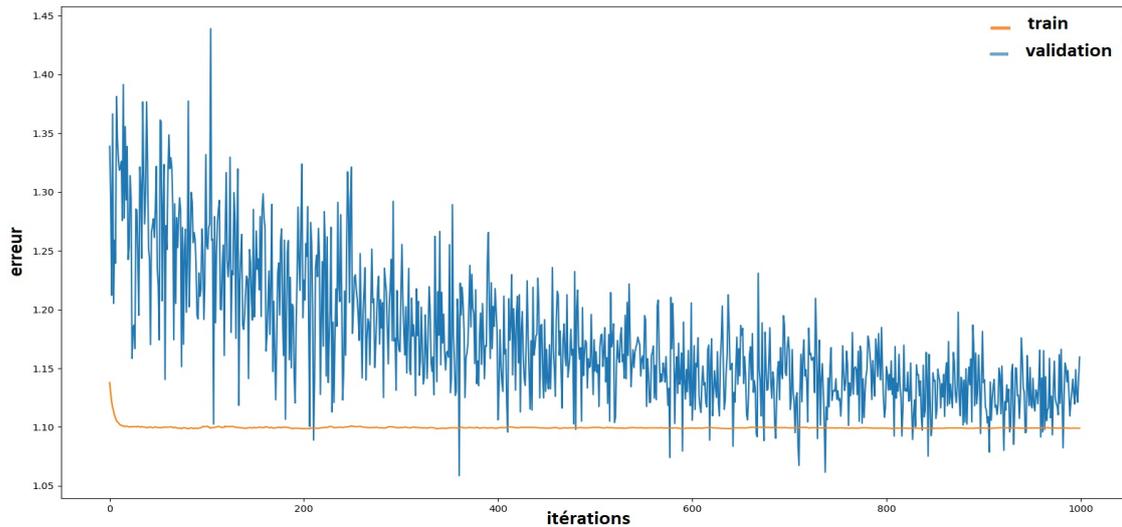


FIGURE 3.3 – Erreur d'entraînement de Sigmoid avec SGD pour 1000 itérations.

Analyse de la courbe :

Nous remarquons que l'erreur de validation ne tend pas vers zéro et reste stable à la valeur 1.1 après 27 itérations, au contraire de l'erreur faite dans l'entraînement n'est pas stable et varie fortement aussi.

On résume les résultats du premier setup dans un tableau.

Itérations	500	1000
Précision	33%	33%
Erreur de validation	1.1	1.1
Erreur d'entraînement	1.16	1.16

TABLEAU 3.1 – Résultat après l'entraînement du premier setup.

2. Deuxième setup :

L'architecture du réseau utilisé pour l'entraînement indiqué dans la figure 3.4. On a utilisé la Sigmoid comme une fonction d'activation dans les couches cachées avec l'optimiseur Adam.

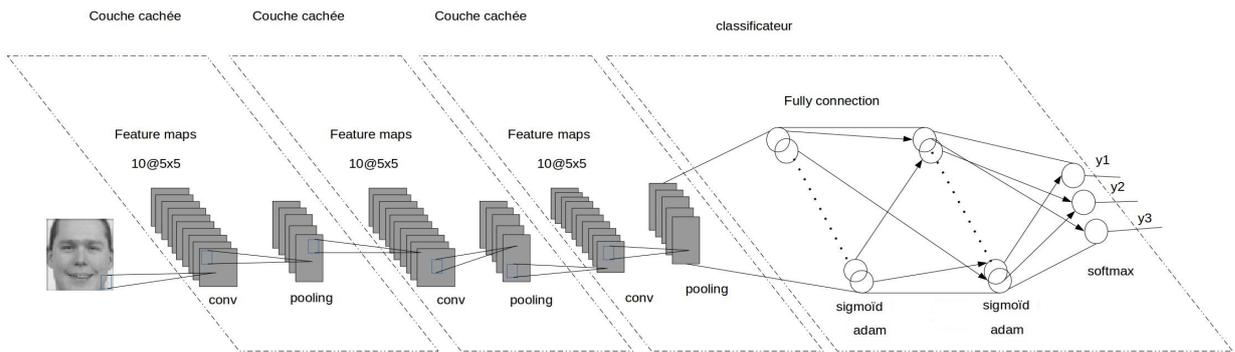


FIGURE 3.4 – Architecture utilisée pour l'entraînement.

a) Entraînement de réseau 500 itérations :

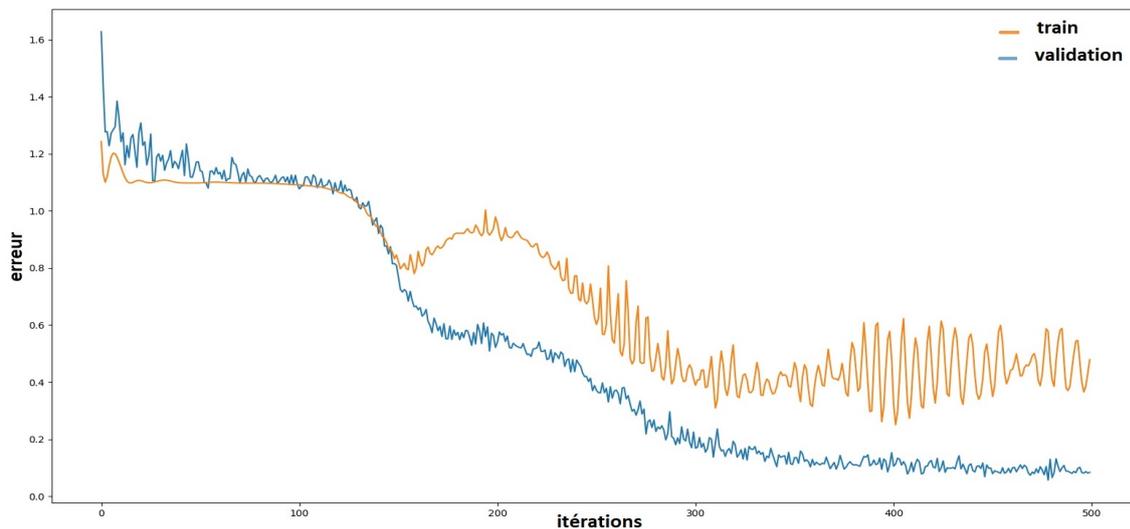


FIGURE 3.5 – Erreur d'entraînement de Sigmoid avec Adam pour 500 itérations.

Analyse de la courbe :

Nous remarquons que les erreurs de validation et d'entraînement ne sont pas nulles ils restent instables.

b) Entraînement de réseau pour 1000 itérations :

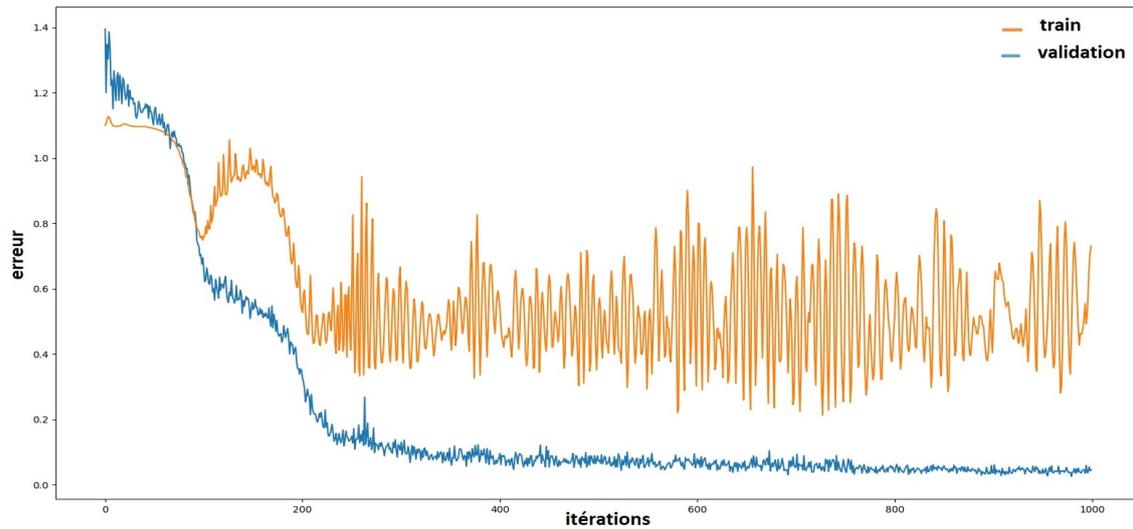


FIGURE 3.6 – Erreur d'entraînement de Sigmoid avec Adam pour 1000 itérations.

Analyse de la courbe :

Nous remarquons que les erreurs de validation et entraînement ne sont pas nulles et restent instables comme les résultats pour 500 itérations.

On résume les résultats du deuxième setup sur un tableau.

Itérations	500	1000
Précision	96%	97%
Erreur de validation	0.5	0.72
Erreur d'entraînement	0.15	1.18

TABLEAU 3.2 – Résultat après l'entraînement.

3. Troisième setup :

L'architecture de réseau utilisée pour l'entraînement indiqué dans la figure 3.7, on a utilisée Relu comme une fonction d'activation dans les couches cachées avec optimiseur SGD.

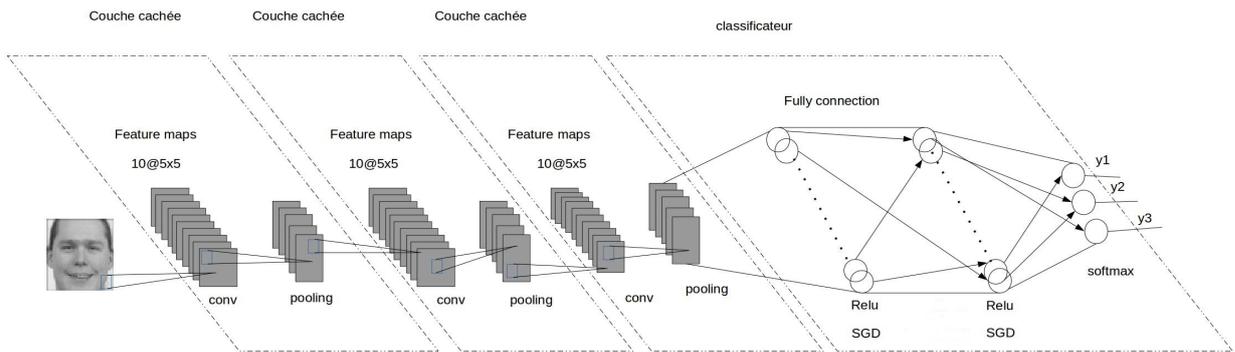


FIGURE 3.7 – Architecture utilisée pour l'entraînement.

a) Entraînement de réseau pour 500 itérations :

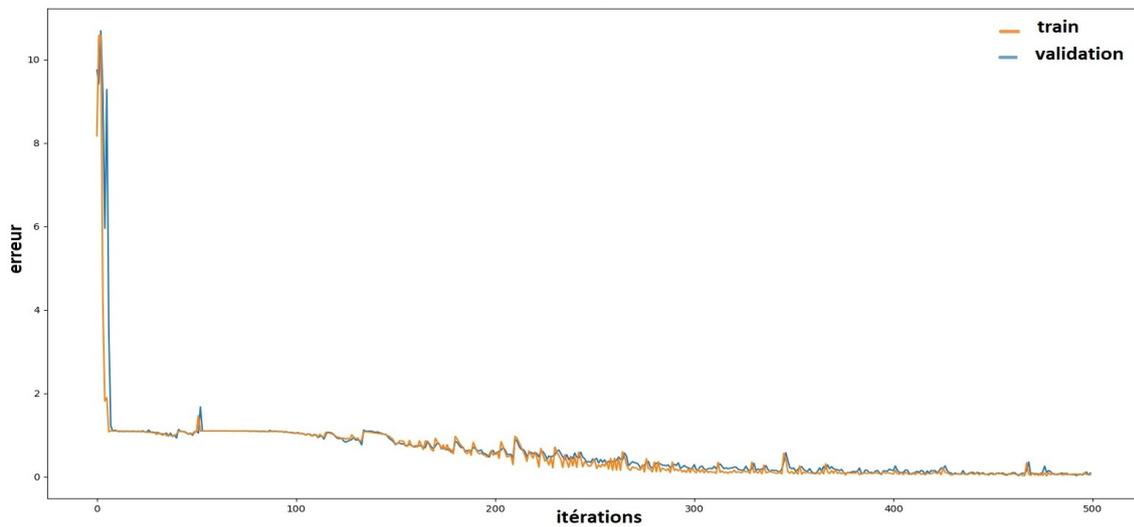


FIGURE 3.8 – Erreur d'entraînement de Relu avec SGD pour 500 itérations.

Analyse de la courbe :

Nous remarquons que les erreurs devient nulle après 470 itérations.

b) Entraînement de réseau pour 1000 itérations :

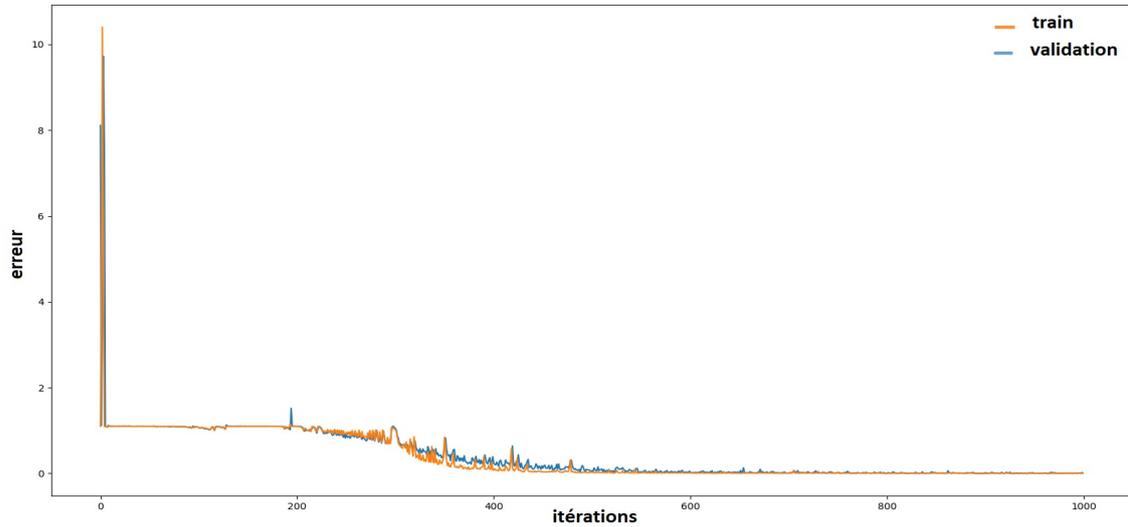


FIGURE 3.9 – Erreur d'entraînement de Relu avec SGD 1000 itérations.

Analyse de la courbe :

Nous remarquons que les erreurs devient et restons nulle après 470 itérations.

On résume les résultats du troisième setup sur le tableau suivant.

Itérations	500	1000
Précision	99%	100%
Erreur de validation	0	0.0
Erreur d'entraînement	0	0

TABLEAU 3.3 – Résultat après l'entraînement de troisième setup.

4. Quatrième setup :

L'architecture de réseau utilisée pour l'entraînement indiqué dans la figure 3.13, on a utilisée Relu comme une fonction d'activation dans les couches cachées avec optimiseur adam.

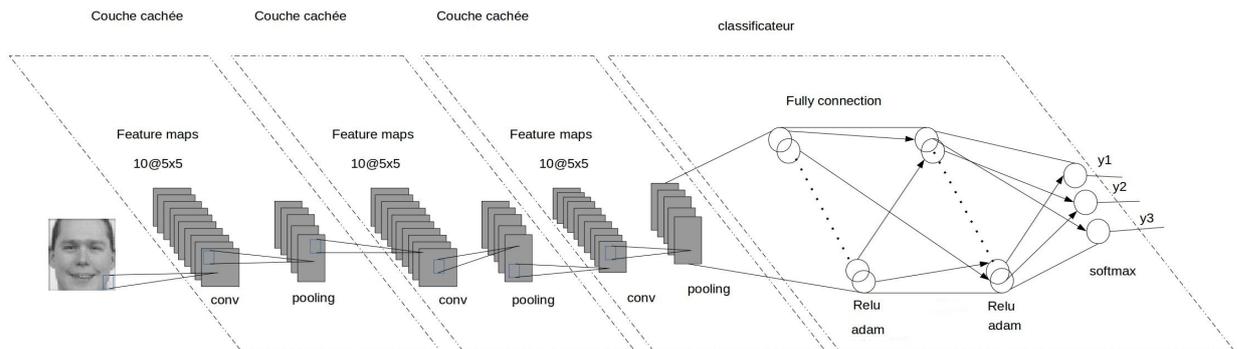


FIGURE 3.10 – Architecture utilisée pour l'entraînement.

a) Entraînement de réseau pour 500 itérations :

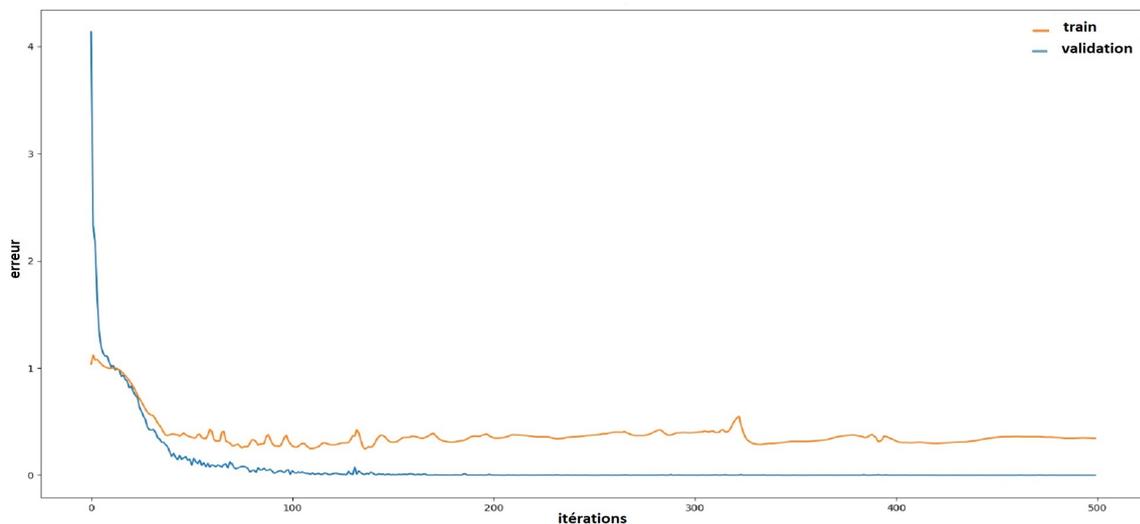


FIGURE 3.11 – Erreur d'entraînement de Relu avec Adam pour 500 itérations.

Analyse de la courbe :

Nous remarquons que l'erreur de validation ne pas nulle à la fin d'entraînement leur valeur exacte est 0.36 on contraire que l'erreur faite en entraînement devient nulle après 170 itérations.

b) Entraînement de réseau pour 1000 itérations :

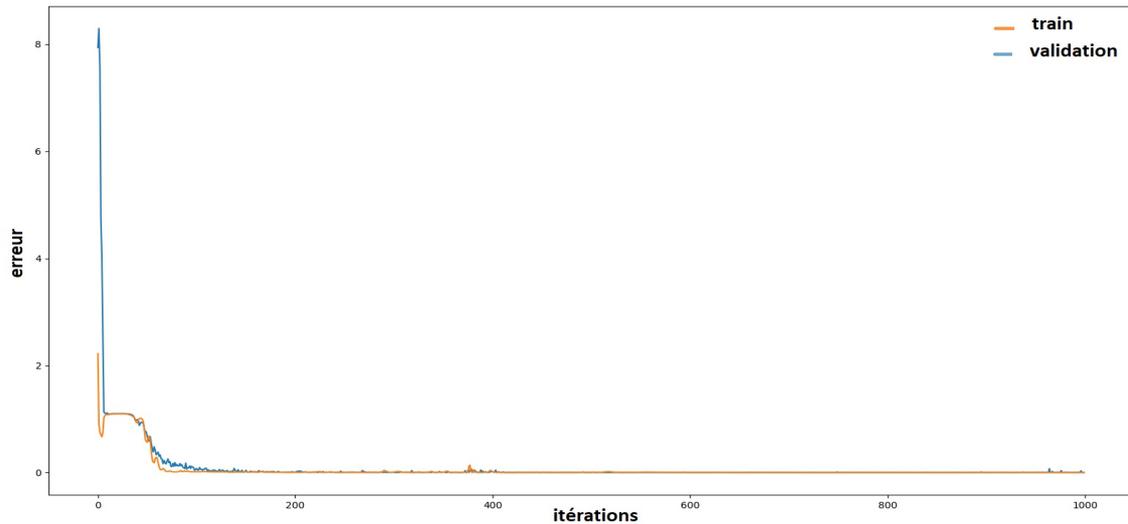


FIGURE 3.12 – Erreur d'entraînement de Relu avec Adam pour 1000 itérations

Analyse de la courbe :

Nous remarquons que l'erreur de validation devient nulle après 90 itérations avec une petite changement dans intervalle $[364, 400[$, après 400 itérations l'erreur nulle jusqu'à la fin. L'erreur d'entraînement devient nulle après 150 itérations.

On résume les résultats de quatrième setup sur le tableau suivant.

Itérations	500	1000
Précision	99%	100%
Erreur de validation	0.36	0
Erreur d'entraînement	0	0

TABLEAU 3.4 – Résultat d'entraînement de quatrième setup.

D'après la comparaison entre les résultats des différents setups, on a remarqué que la Relu donne de meilleurs (4b, Figure 3.12) résultats après 90 itérations.

Teste (validation) :

D'après les résultats obtenue depuis le modèle de la meilleure résultats (la figure Figure 3.12) nous avons testé le reste de la base de données avec ce modèle, le tableau suivant montre les résultats du test

Nombre de prédictions correctes	40
Nombre total d'images	42
Précision	95%

TABLEAU 3.5 – Résultat de test.

Sur 42 images d'expressions faciales le modèle arrive a reconnaître 40 expressions faciales avec la précision de 95%. Matrice de confusion :

$$\begin{bmatrix} 12 & 0 & 2 \\ 0 & 14 & 0 \\ 0 & 0 & 14 \end{bmatrix}$$

La première ligne de matrice de confusion, le modèle a une confusion de reconnaître entre la valence de l'émotion négative (colère) et la valence de l'émotion neutre.

La deuxième et la troisième ligne de la matrice de confusion, le modèle a reconnu l'autre valence de l'émotion .

Discussion :

Par la comparaison des résultats d'entraînement et de test, on conclut que la fonction Relu est le meilleur choix pour une utilisation en couche cachée que la Sigmoid, avec un l'optimiseur Adam nous obtenons meilleure performance.

Conclusion

Dans ce mémoire, nous avons proposé un setup pour une architecture d'un réseau de neurones pour la reconnaissance de la valence des émotions. Nous avons comparé plusieurs setups et nous avons trouvé que le meilleur setup est la fonction Relu dans les couches cachées, l'optimiseur Adam avec 1000 itérations.

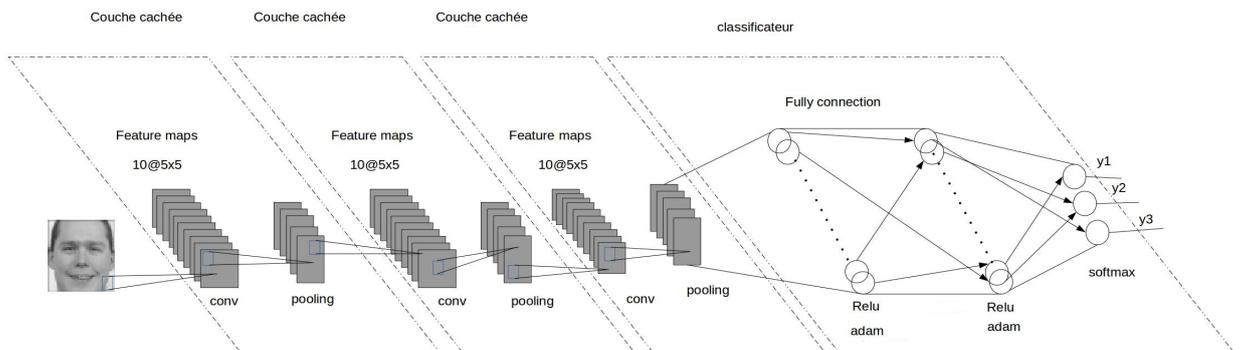


FIGURE 3.13 – Architecture qui nous donne de meilleures performances

Les résultats sont de 95% avec une erreur qui s'annule avant les 200 itérations. Notre réseau a beaucoup de pistes d'application comme l'informatique affective dans laquelle on essaye de donner à la machine la faculté de reconnaître l'émotion de l'utilisateur et de la simuler. Dans un travail futur on va essayer de proposer notre travail dans une conférence internationale en ciblant la reconnaissance des émotions de bases selon Ekman par le traitement des expressions faciales.

Bibliographie

- [1] S. Poria, E. Cambria, R. Bajpai, et A. Hussain, " A review of affective computing : From unimodal analysis to multimodal fusion", Inf. Fusion, vol. 37, p. 98-125, 2017.
- [2] R. W. Picard, " Affective Computing for HCI. ", in HCI (1), 1999, p. 829-833.
- [3] F. Abdat, "Reconnaissance automatique des émotions par données multimodales : expressions faciales et signaux physiologiques," Université de Metz, France, 2010.
- [4] P. Ekman, "An argument for basic emotions," Cogn. Emot., vol. 6, no. 3, pp. 200, 169, 1992.
- [5] W. Handouzi, Traitement d'information mono-source pour la validation objective d'un modèle d'anxiété : application au signal de pression sanguine volumique. Université de Lorraine, 2014.
- [6] K. Gouizi, " Analyse et Traitement des Signaux Physiologiques en vue De la Reconnaissance des Emotions. ", Thesis, 2009.
- [7] P. Kim, MATLAB Deep Learning : With Machine Learning, Neural Networks and Artificial Intelligence. Apress, 2017.
- [8] Felipe Sanchez Garzon, " AI Apprentissage Automatique, Machine Learnig ", 16 :14 :11 UTC.
- [9] P. Borne, M. Benrejeb, et J. Haggège, Les réseaux de neurones : présentation et applications. Editions OPHRYS, 2007.
- [10] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, et A. van Knippenberg, " Presentation and validation of the Radboud Faces Database ", Cognition & Emotion, vol. 24, no 8, p. 1377-1388, déc. 2010.
- [11] S.Pattanayak, Pro Deep Learning with TensorFlow : A Mathematical Approach to Advanced Artificial Intelligence in Python. Apress, 2017.
- [12] " Réseau de neurones artificiels ", Wikipédia. 24-avr-2018.
- [13] <https://www.lebigdata.fr/machine-learning-et-big-data>, accédé le : 03-07-2018 19 :29 :18
- [14] <https://www.mathworks.com/help/matlab/math/multidimensional-arrays.html>, accédé le : 23-06-2018 09 :53 :06
- [15] <https://pythonmachinelearning.pro/introduction-to-convolutional-neural-networks-for-vision-tasks/>, accédé le : 03-07-2018 19 :40 :40

- [16] <https://docs.python.org/fr/3/library/os.html>, accédé le : 27-05-2018 19 :59 :29
- [17] <http://www.numpy.org/>, accédé le : 27-05-2018 19 :50 :29
- [18] <https://matplotlib.org/index.html>, accédé le : 04-06-2018 01 :29 :55
- [19] <https://pillow.readthedocs.io/en/5.1.x/>, accédé le : 27-05-2018 21 :00 :34
- [20] <https://www.tensorflow.org/>, accédé le : 04-06-2018 01 :15 :55
- [21] <http://www.deeplearning.net/software/theano/>, accédé le : 26-05-2018 19 :27 :01
- [22] <https://github.com/Theano/>, accédé le : 26-05-2018 19 :28 :01
- [23] <https://www.scipy.org/>, accédé le : 23-06-2018 10 :59 :33
- [24] <http://scikit-learn.org/stable/>, accédé le : 26-06-2018 19 :38 :01
- [25] <https://keras.io/>, accédé le : 04-06-2018 01 :20 :55

Résumé

L'informatique affective (affective computing) est une thématique pluridisciplinaire émergente qui regroupe des chercheurs du domaine de l'intelligence artificielle et du traitement du langage naturel jusqu'au domaine des sciences sociales et cognitives.

La reconnaissance des émotions et de manière plus large l'informatique affective, a significativement évolué ces dernières années.

Dans ce travail, nous avons proposé une architecture d'un réseau de neurones pour la reconnaissance de la valence des émotions (positives ou négatives). Nous avons comparé plusieurs setups d'entraînement de 80% d'une base de donnée RaFD avec deux fonctions d'activation (Relu et Sigmoid), deux optimiseurs (Adam et SGD) et Softmax dans couche de sortie, nous avons trouvé que le meilleur setup est la fonction Relu dans les couches cachées, optimiseur Adam avec 1000 d'itérations. Le résultat du test de ce setup avec le reste des données donne 95% de précision

Abstract

Affective computing is an emerging multidisciplinary term that brings researchers from the field of artificial intelligence and natural language processing to the field of social and cognitive sciences.

The recognition of emotions and more broadly affective computing, has significantly evolved in recent years.

In this work, we proposed an architecture of a network of neurons for the recognition of the valence of emotions (positif or negatif). We tested Serval training setups of 80% of RaFD dataset with two activation functions (Relu and Sigmoid), two optimizers (Adam and SGD) and Softmax in the output layer, we found that the best setup is to use Relu function and Adam optimizer with 1000 epochs.

المخلص

الحوسبة العاطفية هو موضوع ناشى متعدد التخصصات يجمع باحثين من مجال الذكاء الاصطناعي و معالجة التصرفات الطبيعية في مجال العلوم الاجتماعية و المعرفية. لقد اصبح التعرف على المشاعر و الحوسبة العاطفية أكثر فعالية على نطاق واسع و تطور بشكل كبير في السنوات الأخيرة. في هذا العمل اقترحنا بنية لشبكة العصبية لتعرف على تعبير الوجه لقاعدة البيانات و قمنا بتدريب الشبكة العصبية.

DISCIPLINE : AUTOMATIQUE