République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abderrahmane Mira Bejaia
Faculté Des Sciences Exactes
Département d'Informatique

Mémoire de fin d'études

Pour l'obtention du diplôme de Master en Intelligence Artificielle

---

Thème
Reconnaissance des Expressions Faciales avec le Deep Learning

---

Réalisé par :
M. Bouguettaya Rabah

Encadré par :
M. Khammari Mohammed

Soutenu le 14 septembre 2023, Devant le jury composé de :

Dr. Mouloud ATMANI : M.C.A - Président
Mme. Dalila KESSIRA : M.A.A - Examinatrice

Promotion : 2022/2023

Remerciements

Je tiens à exprimer ma profonde gratitude envers le Bon Dieu qui m'a donné la force nécessaire pour réaliser ce modeste travail ainsi qu'à toutes les personnes ayant contribué de près ou de loin à la réalisation de ce mémoire.

Tout d'abord, je tiens à remercier notre encadrant Monsieur KHAMMARI Mohammed de nous avoir orienté et aidé à mener à bien ce projet. Je lui exprime toute ma reconnaissance pour son soutien constant, ses conseils précieux et son expertise. Sa disponibilité et sa passion pour la recherche m'a inspiré et encouragé à repousser mes limites.

Nous sommes également reconnaissants envers les membres du jury, Monsieur. Mouloud ATMANI, Madame. Dalila KESSIRA, pour leur temps, leur attention et leurs commentaires constructifs qui ont grandement amélioré ce travail.

Je remercie ma très chère maman, amis et familles respectives pour leur soutien indéfectible tout au long de mes études. Leurs encouragements, leurs conseils et leur amour ont été une source de motivation et de réconfort dans lequel je puise mon énergie et ma force pour ne jamais abandonner face aux obstacles.

Enfin, je suis reconnaissant envers l'université A. Mira de Béjaïa pour avoir fourni les enseignements et les ressources nécessaires à la réalisation de cette recherche.

Mes sincères remerciements vont également à toutes les personnes que nous n'avons pas citées ici, mais qui ont joué un rôle important dans ce projet.

Table des matières

Table des figures

Liste des tableaux

Liste des abréviations

CNN	*Convolutional Neural Network*
DNN	*Deep Neural Network*
IA	*Intelligence Artificielle*
ML	*Machine Learning*
QA	*Question Answering*
RNN	*Réseaux de neurones récurrents*
ResNet	*Residual Neutral Network*
FER	*Facial Expression Recognition*
HOG	*Histogram of oriented gradients*
TAN	*Tree augmented bayésien naïf*
HMM	*Hidden Markov model*
CK	*Cohn-Kanade*
TFD	*TensorFlow Datasets*
CONV	*Convolution*
VGG	*Visual Geometry Group*
V3	*Version 3*
LBP	*local binary pattern*
JAFFE	*Japanese Female Facial Expression*
SVM	*support vector machines*
SFEW	*Static Facial Expression in the Wild*
MLP	*Multilayer perceptron*
ILSVRC	*ImageNet Large Scale Visual Recognition Challenge*
ReLU	*A rectified linear unit*
BSD	*Berkeley Source Distribution*
CPU	*Central Processing Unit*
GPU	*Graphics processing unit*
TPU	*Thermoplastic polyurethane*
RAM	*Random-access memory*
VRAM	*Video random-access memory*
FACS	*Facial Action Coding System*

Introduction générale

Les expressions faciales jouent un rôle irremplaçable dans la communication non verbale. Elles communiquent l'émotion et signalent les intentions, la vigilance, la douleur et les traits de personnalité. Les émotions peuvent être exprimées à la fois verbalement et non verbalement. Il existe de nombreux canaux tels que la voix, le visage et les gestes corporels à travers lesquels l'information non verbale est transmise aux observateurs. En outre, Mehrabian et Ferris [63] ont indiqué que l'expression faciale du locuteur contribue à 55% à l'effet du message parlé, alors que la partie verbale et la partie vocale qu'avec 7% et 38% respectivement. Les expressions faciales peuvent non seulement changer le flux de la conversation, mais aussi fournir aux auditeurs un moyen de communiquer une grande quantité d'informations au locuteur sans même prononcer un seul mot. Lorsque l'expression faciale ne coïncide pas avec les mots parlés, alors l'information véhiculée par le visage prend plus de poids dans le décodage des informations.

L'analyse automatique de l'expression du visage est un problème qui affecte d'importantes applications dans de nombreux domaines tels que l'interaction homme-machine. En fait, bien que les nouvelles technologies soient présentes dans notre vie quotidienne, ils ne fournissent pas une interface adéquate qui les rend plus abordables pour les utilisateurs. Par conséquent, l'informatique affective, en améliorant l'interaction homme-ordinateur, permet aux ordinateurs d'être plus adaptés à l'homme et non pas l'inverse.

L'intérêt de la recherche est de permettre au système informatique de reconnaître les expressions et d'utiliser les informations émotives intégrées dans les interfaces homme-machine.

Le succès actuel des réseaux de neurones convolutifs (CNN) dans la classification d'images s'est étendu au problème de la reconnaissance de l'expression faciale. Le deep Learning et plus particulièrement CNN sont apparus spécialement pour résoudre les problèmes rencontrés de machine Learning. L'un des ingrédients les plus importants pour le succès de ces méthodes est la disponibilité de grandes quantités de données d'entraînement.

Le Convolutional Neural Networks est l'une des structures réseau les plus représentatives de la technologie d'apprentissage en profondeur et a connu un grand succès dans le domaine du traitement et de la reconnaissance d'images.

L'objectif de ce mémoire consiste à proposer une approche de reconnaissance des expressions faciales en se basant sur la méthode des réseaux de neurones convolutifs. Ce dernier sera constitué en quatre chapitres et organisé comme suit :

– Dans le premier chapitre, nous présentons l'expression faciale et le processus d'analyse automatique de l'émotion.
– Ensuite, le second chapitre, nous le consacrons à la présentation de l'apprentissage profond, où nous donnerons plus de détails sur les réseaux de neurones convolutifs.
– La conception de notre approche de reconnaissance des expressions faciales basée sur les CNN est présentée dans le troisième chapitre.
– Le dernier chapitre est consacré à la description des différents outils utilisés dans le développement de notre application, ainsi que les différents résultats obtenus.
– Et enfin, nous terminerons ce mémoire par une conclusion générale et quelques perspectives.

Chapitre 1

EXPRESSIONS FACIALES

1.1 Introduction

Les expressions faciales sont une forme de communication humaine riche et complexe. Elles transcendent les barrières linguistiques et permettent de transmettre des émotions, des intentions et des réactions instantanément.

La capacité à comprendre ces expressions est essentielle dans de nombreuses interactions humaines, qu'elles soient sociales, professionnelles ou cliniques. Imaginez un monde où les machines pourraient décoder les émotions humaines, offrant une compréhension des besoins, des désirs et des sentiments.

Cette vision devient de plus en plus tangible grâce à l'intersection entre la vision par ordinateur et le deep learning, un domaine qui est en constante évolution. Ainsi, offre la capacité d'extraire des informations subtiles et complexes à partir de données brutes. Dans le contexte de la reconnaissance des expressions faciales, il peut analyser des milliers de caractéristiques du visage en un instant, permettant ainsi une compréhension plus approfondie des émotions humaines.

Ce mémoire se penche sur la convergence de la reconnaissance des expressions faciales et du deep learning. Notre objectif principal est d'explorer les avancées récentes dans le domaine, en examinant les méthodes, les avantages et les défis associés à l'utilisation du deep learning pour la compréhension des expressions faciales. Nous nous efforçons de répondre à des questions cruciales telles que : comment les réseaux neuronaux profonds peuvent-ils améliorer la précision de la reconnaissance des expressions faciales ? Quels sont les défis et les limitations actuels de cette approche ? En quoi cette technologie a-t-elle le potentiel de révolutionner divers domaines, de la psychologie à la technologie d'assistance ?

Notre travail est structuré en plusieurs chapitres, chacun explorant une facette différente de la reconnaissance des expressions faciales avec le deep learning. Nous commencerons par examiner les fondements théoriques de ce domaine, avant de plonger dans la méthodologie utilisée pour notre recherche, les résultats obtenus et les discussions sur les implications de ces résultats. Enfin, nous conclurons en discutant des perspectives futures pour cette technologie prometteuse.

1.2 Problématique

L'interaction homme-machine à longtemps se limiter ses recherches au développement de techniques fondées sur l'usage du triplet écran-clavier-souris. Aujourd'hui, elle se dirige vers de nouveaux paradigmes : l'utilisateur doit pouvoir évoluer sans obstacles dans son milieu naturel ; les doigts, la main, le visage ou les objets familiers sont envisagés comme autant de dispositifs d'entrée/sortie, la frontière entre les mondes électronique et physique tend à devenir floue.

Ces nouvelles formes d'interaction ont besoin généralement de capturer du comportement observable d'un utilisateur et de son environnement. Elles se basent pour cela sur des techniques de vision par ordinateur. Les générations futures d'environnement Homme-Machine deviendront multimodales en intégrant de nouvelles informations, tire son origine de la prise en compte de la parole et/ou des expressions faciales, pour faire passer l'utilisation des machines en une manière directe et naturelle.

L'état émotionnel de l'être humain affecte d'une manière directe son comportement et son rendement dans leurs tâches quotidiennes. Pour cela, la détection de son expression faciale qui va préciser son émotion devient une tâche indispensable pour préciser son émotion avant d'effectuer son travail, tel que la conduite, la robotique sociale et le traitement médical.

Ainsi, plusieurs questions se posent et ouvrent sur les problématiques suivantes Comment modéliser les émotions en tenant compte de leur complexité ? Comment effectuer l'échange émotionnel lors d'une interaction homme-machine ou machine-machine ? Comment modéliser l'aspect psychologique et émotionnel humain en informatique en se basant sur les différentes théories, théorèmes ?

1.3 Définition de l'expression faciale

Tout d'abord, il est primordial de connaitre la différence entre la reconnaissance d'expressions faciales et la reconnaissance d'émotions. Les émotions sont la combinaison de plusieurs facteurs observés comme : un ton élevé ou étouffé, les gestes des mains ou mouvement de la tête, une posture droite ou relâchée, un regard évasif ou appuyé en plus des expressions du visage (froncement du nez, sourcils abaissés, bouche ouverte).
La figure 1.1 montre une représentation globale de comment sont générées les émotions et est-ce que ceci peut avoir un impact sur l'expression faciale et vice vers ça.

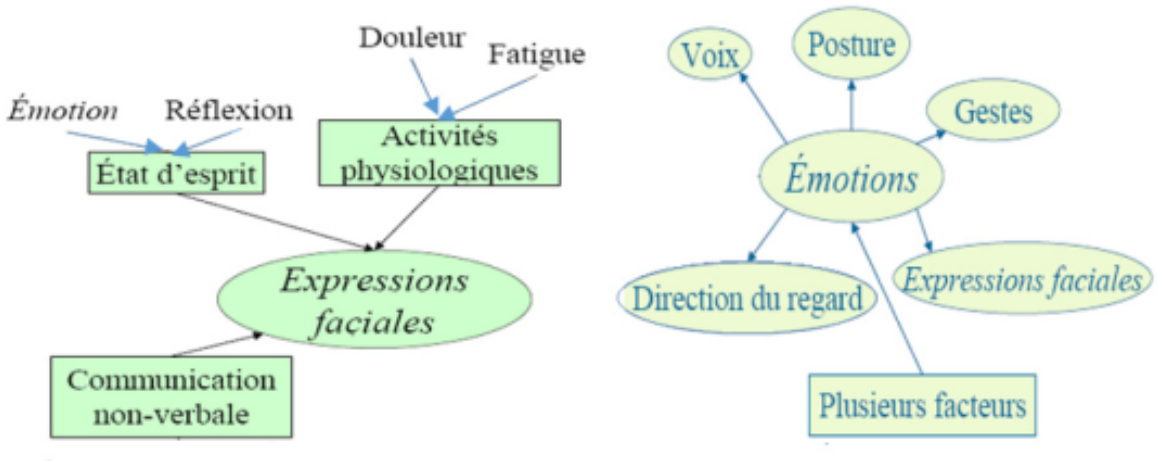

FIGURE 1.1 – Générateurs de l'expression faciale et de l'émotion [35].

En revanche, les émotions ne sont pas la seule origine des expressions faciales. En effet, celles-ci peuvent provenir de l'état d'esprit (réflexion, ennui), de l'activité physiologique (la douleur ou la fatigue) et de la communication non verbale (émotion simulée, clignotement de l'œil, froncement des sourcils). Néanmoins, sept émotions de base correspondent chacune à une expression faciale unique, et ce, quelles que soient l'ethnicité et la culture de la personne observée, ces émotions sont : la colère, le dégoût, l'étonnement, la joie, le mépris, la peur et la tristesse.

La reconnaissance des expressions faciales consiste à classer les déformations des structures faciales et les mouvements faciaux uniquement à partir des informations visuelles. La reconnaissance des émotions, quant à elle, nécessite une interprétation qui exige une information contextuelle plus complète.

FIGURE 1.2 – joie, colère, surprise [2].

1.4 Différent type d'expression faciale

Le tableau 1-1 : ci-dessous présente une classification des expressions faciales selon certains chercheurs à base d'émotion et d'inclusion.

Théoriciens	**Émotions de base**	**Base d'inclusion**
Plutchik	Acceptation, colère, anticipation, dégoût, joie, peur, tristesse, surprise	Relation aux processus biologiques adaptatifs
Arnold	Colère, aversion, courage, abattement, désir, désespoir, peur, haine, espoir, amour, tristesse	Relation aux tendances d'action
Frijda	Désir, bonheur, intérêt, surprise, émerveillement, chagrin	Désir, bonheur, intérêt, surprise, émerveillement, chagrin
McDougall	Colère, dégoût, exaltation, peur, soumission, émotion tendre, émerveillement	Relation à l'instinct
Ekman, Friesen et Ellsworth	Colère, dégoût, peur, joie, tristesse, surprise	Expressions faciales universelles

TABLE 1.2 – Tableau sur les différentes classifications d'expressions faciales [21].

1.5 Domaine d'application

L'analyse de l'impact de toute situation, contenu, produit ou service, censé susciter des réponses faciales volontaires ou involontaires, est d'un intérêt majeur, car cette analyse permet de détecter l'état cognitif et affectif de la personne impactée et par conséquent comprendre et même prédire ses intentions, ses actions et ses réactions. De ce fait, l'analyse des expressions faciales trouve de nombreuses applications dans des domaines divers tels que :

1.5.1 Psychologie et Recherche Comportementale

- L'analyse des expressions faciales est utilisée pour étudier les émotions humaines, la réaction à des stimuli et les troubles émotionnels.

- Elle est également employée pour examiner les réponses émotionnelles dans des études de psychologie expérimentale et clinique.

1.5.2 Santé Mentale et Bien-être

- Les applications mobiles et les dispositifs portables utilisent la reconnaissance des expressions pour surveiller le bien-être émotionnel et offrir des conseils ou des interventions en cas de détresse émotionnelle.

- Les thérapeutes peuvent utiliser ces technologies pour évaluer l'état émotionnel de leurs patients et adapter leurs interventions en conséquence.

1.5.3 Éducation et Apprentissage

- Les systèmes éducatifs utilisent la détection des émotions pour évaluer l'engagement des élèves et personnaliser l'enseignement en temps réel.

- Les tuteurs virtuels peuvent détecter la frustration ou la confusion chez les étudiants et leur fournir un soutien adapté.

1.5.4 Interaction Homme-Machine

- Les interfaces homme-machine basé sur la reconnaissance des expressions peuvent rendre les interactions avec les ordinateurs et les robots plus naturelles et intuitives.

- Cela inclut les jeux vidéo qui s'adaptent au comportement du joueur en fonction de ses émotions.

1.5.5 Sécurité et Surveillance

- La surveillance vidéo utilise la reconnaissance des expressions pour détecter des comportements suspects ou des signaux d'alarme, comme la colère ou la peur, dans les environnements publics ou privés.

- Les contrôles de sécurité des aéroports peuvent bénéficier de la détection d'émotions pour identifier des comportements anormaux.

1.5.6 Technologie d'Assistance

- Les technologies d'assistance pour les personnes atteintes de troubles du spectre autistique ou de déficiences émotionnelles utilisent la reconnaissance des expressions pour aider ces individus à interagir avec les autres de manière plus efficace.

- Les technologies d'assistance pour les personnes atteintes de troubles du spectre autistique ou de déficiences émotionnelles utilisent la reconnaissance des expressions pour aider ces individus à interagir avec les autres de manière plus efficace.

1.5.7 Publicité et Marketing

- Les entreprises utilisent la reconnaissance des expressions pour évaluer les réactions des consommateurs à des publicités, des produits ou des expériences de magasinage.

- Cela permet d'ajuster les campagnes marketing en fonction des émotions suscitées.

1.5.8 Santé et Médecine

- Dans le domaine médical, la détection des expressions faciales peut être utilisée pour évaluer la douleur des patients, détecter des signes de détresse dans les soins aux patients âgés, ou surveiller les émotions des patients atteints de troubles de l'humeur.

1.5.9 Jeux et Divertissement

- L'industrie du jeu vidéo utilise la reconnaissance des expressions pour créer des expériences de jeu plus immersives, où les personnages réagissent aux émotions du joueur.

- Les parcs d'attractions peuvent également intégrer cette technologie pour personnaliser les expériences des visiteurs.

1.5.10 Ressources Humaines et Entretiens d'Embauche

- Les entreprises peuvent utiliser la reconnaissance des expressions pour évaluer la réaction des candidats lors d'entretiens d'embauche.

- Elle peut également être utilisée pour évaluer l'engagement des employés ou détecter des signes de stress dans le milieu de travail.

1.6 Processus d'analyse automatique de l'émotion

Depuis le milieu des années 70, différentes approches ont été proposées pour l'analyse des expressions faciales, des images faciales statiques ou des séquences d'images, elles s'accordent tous sur trois étapes de base qui sont la détection du visage, l'extraction des expressions faciales et la classification de ces dernières. Un système qui effectue une reconnaissance automatique des expressions faciales est généralement composé de trois modules principaux.

La première étape consiste à détecter et enregistrer la région du visage dans les images. Par la suite, l'extraction des informations nécessaires qui décrivent au mieux l'expression. À la fin, en se basant sur ces informations, l'image sera affectée à une catégorie d'expressions à l'aide d'un classifier.

La figure 1-3 schématise les différentes étapes d'un système de reconnaissance des expressions faciales. D'autres filtres ou modules de prétraitement de données peuvent être utilisés entre ces modules principaux pour améliorer les résultats de détection, d'extraction de caractéristiques ou de classification.

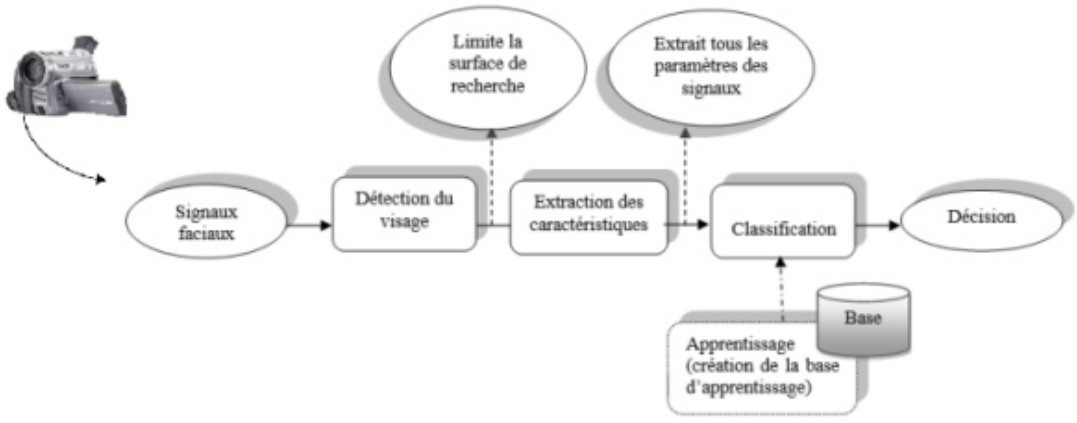

FIGURE 1.3 – Architecture d'un système de reconnaissance des expressions faciales [40].

1.6.1 Détection de visage

La détection de visage consiste à déterminer la présence ou l'absence de visages dans une image et en cas de présence à déterminer sa localisation. C'est une tâche préliminaire nécessaire et fondamentale à la plupart des techniques d'analyse du visage. Les techniques utilisées sont généralement issues du domaine de la reconnaissance des formes. En effet, le problème peut être vu comme la détection de caractéristiques communes à l'ensemble des visages humains : il s'agit de comparer une image à un modèle générique de visage et d'indiquer s'il y a ou non ressemblance.

La sortie d'un détecteur de visage indique le nombre de visages présents dans l'image. De plus, la plupart des détecteurs de visage actuels sont aussi des localisateurs de visages : ils renvoient une localisation des visages détectés (une boîte englobante par exemple). Les principales difficultés sont la robustesse aux différentes identités, poses du visage, expressions faciales et aux variations d'illumination.

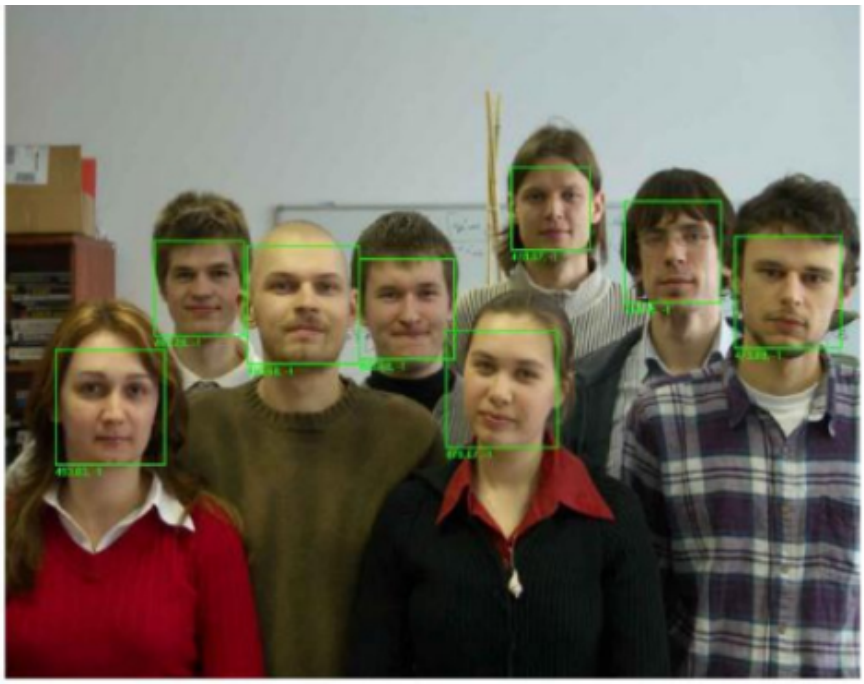

FIGURE 1.4 – Détection de visage [17].

1.6.2 Problèmes rencontrés lors de la détection de visage

La reconnaissance automatique des émotions faciales est un domaine de recherche en constante évolution, mais il rencontre encore plusieurs problèmes et défis importants. détection de visages, la reconnaissance de visages, la reconnaissance d'expressions faciales et la reconnaissance de genre. L'orientation du visage, Les déformations de l'apparence du visage causées par différentes expressions telles que le profil de la tête, l'éclairage.

1.6.2.1 Profil de la tête

Le taux de détection du visage baisse quand des variations de pose sont présentes dans les images. La variation de pose est considérée comme un problème majeur pour les systèmes de détection faciale.

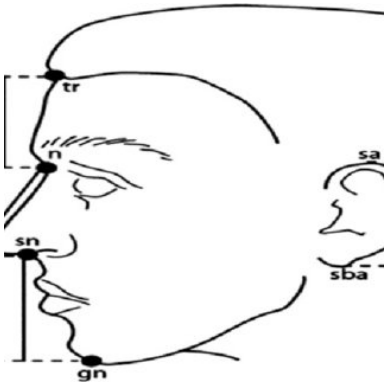

FIGURE 1.5 – montre un exemple de visage de profile [19].

1.6.2.2 Eclairage

Le problème de l'éclairage est un vieux problème dans la vision de la machine. L'intensité et la direction de l'éclairage pendant la prise de vue affectent l'apparence du visage. Ces changements dans l'éclairage peuvent révéler des ombres qui mettent en évidence ou cachent certaines caractéristiques du visage. Par exemple, un visage vu sous une lumière bleue est totalement différent d'un visage vu sous une lumière rouge. [18]

FIGURE 1.6 – montre un exemple de changement d'éclairage [47].

1.6.3 Occlusion

Un visage peut être partiellement masqué par des objets ou par le port d'accessoires tels que les lunettes, le chapeau, l'écharpée, livre et autres accessoires. Cela affecte l'extraction et la reconnaissance des caractéristiques d'un visage. [16]

FIGURE 1.7 – montre des exemples d'occlusion [74].

1.6.4 Extraction des caractéristiques

Une fois que le visage est détecté dans une image, la prochaine étape est l'extraction de caractéristiques du visage montré, qu'on appelle aussi les points caractéristiques ou ("Landmarks").Ces points permettent d'encadrer les régions telles que les yeux, la bouche, le nez, les sourcils, etc. La détection des points caractéristiques du visage commence habituellement à partir d'une boite englobante rectangulaire renvoyée par un détecteur de visage qui localise ce dernier. L'extraction de caractéristiques géométriques telles que les contours des composants faciaux, les distances faciales, etc. fournit les emplacements ou les caractéristiques d'apparence peuvent être calculées. En raison de la grande variabilité dans les types de visages, il est très difficile pour la machine d'extraire les traits faciaux. De ce fait, les méthodes d'extraction des caractéristiques pour l'analyse d'expression peuvent être séparées en deux types d'approches : les méthodes basées sur les caractéristiques géométriques et les méthodes basées sur l'apparence. [68]

1.6.4.1 Les caractéristiques géométriques

Représentent la forme et l'emplacement des composants du visage (y compris la bouche, les yeux, les sourcils et le nez). Les composants faciaux ou les traits faciaux sont extraits pour former un vecteur de caractéristiques représentant la géométrie du visage.

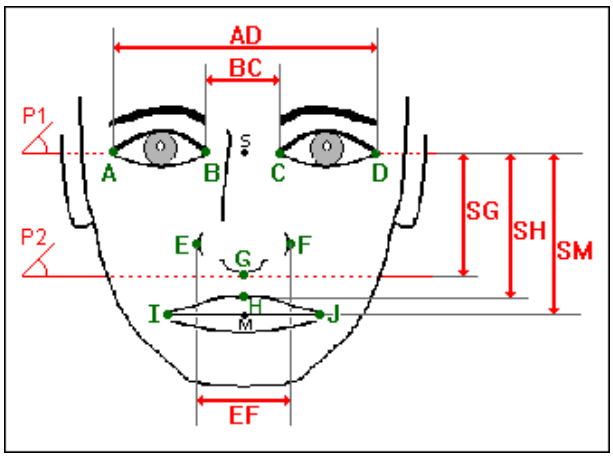

FIGURE 1.8 – Modèle géométrique du visage [52].

1.6.4.2 Les caractéristiques d'apparence

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

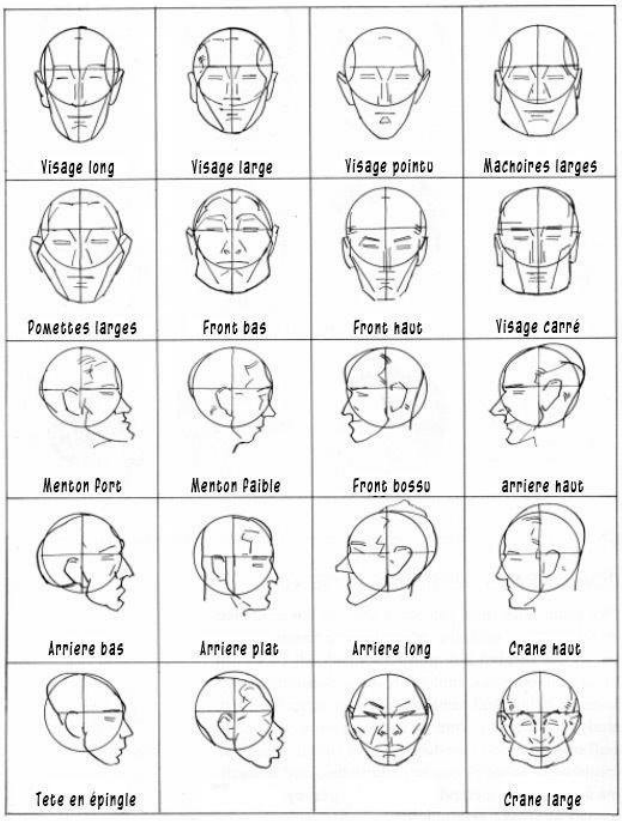

FIGURE 1.9 – differents-types-de-visage [4].

1.6.5 Problèmes rencontrés à la reconnaissance automatique des émotions faciales

Voici quelques-uns des problèmes courants auxquels les systèmes de reconnaissance automatique des émotions faciales peuvent être confrontés [64] [54].

1.6.5.1 Variabilité des Expressions

Les expressions faciales peuvent varier considérablement d'une personne à l'autre, ainsi qu'au sein d'une même personne dans des contextes différents. Cette variabilité peut rendre la tâche de reconnaissance difficile.

1.6.5.2 Influence de l'Âge et du Sexe

L'âge et le sexe peuvent influencer la manière dont les émotions sont exprimées sur le visage. Les modèles doivent prendre en compte ces facteurs pour améliorer la précision.

1.6.5.3 Étiquetage des Données

L'étiquetage des données d'apprentissage pour l'entraînement des modèles de reconnaissance d'émotions peut être sujet à des erreurs, ce qui peut avoir un impact sur la qualité des modèles.

1.6.5.4 Sur-apprentissage

Les modèles de deep learning peuvent être sujets au sur-apprentissage, ce qui signifie qu'ils sont trop adaptés aux données d'entraînement spécifiques et ne généralisent pas bien sur de nouvelles données.

1.6.5.5 Interprétation Subjective

L'interprétation des émotions à partir des expressions faciales peut être subjective, et différentes personnes peuvent attribuer des émotions différentes à la même expression.

1.6.5.6 Ambiguïté Émotionnelle

Parfois, il peut y avoir de l'ambiguïté dans les expressions faciales, où une expression peut être interprétée de différentes manières en fonction du contexte.

1.7 Etat de l'art

Chao Qi et all [77] : Dans cet article, une nouvelle approche de reconnaissance d'expressions est présentée, basée sur la cognition et les modèles binaires mappés, ainsi que des motifs binaires cartographiés.
Dans un premier temps, l'approche est basée sur l'opérateur LBP pour extraire les contours du visage. Ensuite, l'établissement d'un modèle pseudo-3-D est utilisé pour segmenter la zone du visage en six sous-régions d'expression faciale. Dans ce contexte, les sous-régions et les images globales d'expression faciale utilisent la méthode LBP pour l'extraction des caractéristiques, puis utilisent deux classifications, à savoir la machine à vecteur de support et la méthode softmax, avec deux types de modèles de classification des émotions : le modèle d'émotion de base et le modèle d'émotion circumplex. Enfin, ils ont réalisé une expérience comparative sur l'expansion de l'ensemble de données d'expression faciale Cohn-Kanade (CK +) et les ensembles de données de test recueillies auprès de dix volontaires. Les résultats expérimentaux montrent que la méthode peut éliminer efficacement les facteurs de confusion dans l'image. Et le résultat de l'utilisation du modèle émotionnel circumplex est manifestement meilleur que celui du modèle émotionnel traditionnel. En se référant à des études pertinentes sur la cognition humaine, ils ont vérifié que les yeux et la bouche expriment d'avantage d'émotions.

Deepak Ghimire et all [36] : Les expressions faciales véhiculent des indices non verbaux qui jouent un rôle important dans les relations interpersonnelles et sont largement utilisées dans l'interprétation des émotions, les sciences cognitives et les interactions sociales. Dans cet article, ils ont analysé différentes manières de représenter les caractéristiques géométriques et présentent un système de reconnaissance d'expressions faciales (RIF) entièrement automatique utilisant des caractéristiques géométriques saillantes. Dans l'approche de la RFE basée sur les caractéristiques géométriques, la première étape importante consiste à initialiser et à suivre un ensemble dense de points faciaux au fur et à mesure que l'expression évolue dans le temps sur des images consécutives. Dans le système proposé, les points du visage sont initialisés à l'aide de l'algorithme EBGM (elastic bunch graph matching) et le suivi est effectué à l'aide du suiveur Kanade-Lucas-Tomaci (KLT). ils ont extrait les caractéristiques géométriques des points, des lignes et des triangles composés des résultats du suivi des points du visage. Les caractéristiques de ligne et de triangle les plus discriminantes sont extraites à l'aide de l'algorithme AdaBoost multiclasses sélectif avec l'aide de la classification ELM (machine d'apprentissage extrême). Enfin, les caractéristiques géométriques pour la RFE sont extraites des lignes et des triangles boostés composés de points du visage. La précision de la reconnaissance à l'aide des caractéristiques du point, de la ligne et du triangle est analysée indépendamment. Les performances du système FER proposé sont évaluées sur trois ensembles de données différents : les ensembles de données d'expression faciale CK+, MMI et MUG.

Michael Lyons et all [58] Une méthode d'extraction d'informations sur les expressions faciales à partir d'images est présentée. Les images d'expressions faciales sont codées à l'aide d'un ensemble de filtres de Gabor multiorientation multirésolution qui sont ordonnés topographiquement et alignés approximativement avec le visage. L'espace de similarité dérivé de cette représentation est comparé à un espace dérivé des évaluations sémantiques des images par des observateurs humains. Les résultats montrent qu'il est possible de construire un classificateur d'expressions faciales avec le codage de Gabor des images faciales comme étape d'entrée. La

représentation de Gabor présente un degré significatif de plausibilité psychologique, une caractéristique de conception qui peut être importante pour les interfaces homme-machine.

Michel F Valstar et all [86] La distinction automatique entre les expressions posées et spontanées est un problème non résolu. Des études antérieures des sciences cognitives ont indiqué que la séparation automatique des expressions posées et spontanées est possible en utilisant la modalité du visage. Cependant, on sait peu de choses sur les informations provenant des mouvements de la tête et des épaules. Dans ce travail, nous proposons (i) de distinguer les sourires posés des sourires spontanés en fusionnant les modalités de la tête, du visage et des épaules, (ii) d'étudier quelles modalités portent des informations importantes et comment les modalités sont liées les unes aux autres, et (iii) dans quelle mesure la dynamique temporelle de ces signaux contribue à résoudre le problème. Un tracker de tête cylindrique est utilisé pour suivre le mouvement de la tête et deux techniques de filtrage de particules pour suivre le mouvement du visage et des épaules. La classification est effectuée par des méthodes à noyau combinées à des techniques d'apprentissage d'ensemble. ils ont étudié deux aspects de la fusion multimodale : le niveau d'abstraction (c'est-à-dire la fusion précoce, moyenne et tardive) et la règle de fusion utilisée (c'est-à-dire les critères de somme, de produit et de poids). Les résultats expérimentaux de 100 vidéos montrant des sourires posés et de 102 vidéos montrant des sourires spontanés sont présentés. Les meilleurs résultats ont été obtenus avec la fusion tardive de toutes les modalités, 94,0 % des vidéos ayant été classées correctement.

Uroš Mlakar et all [66] Cet article propose un système efficace de sélection des caractéristiques appliqué à un système de reconnaissance des expressions faciales (FER). Ce système, capable de reconnaître sept émotions prototypiques, y compris l'expression neutre, est basé sur un descripteur d'histogramme de gradient orienté (HOG) et des vecteurs de caractéristiques différentielles. La sélection des caractéristiques des émotions a été effectuée à l'aide d'un algorithme d'évolution différentielle multiobjectif modifié de manière appropriée. Le nombre de caractéristiques utilisées a été minimisé, tandis que la précision de reconnaissance des émotions des classificateurs de la machine à vecteurs de support a été maximisée simultanément. Des stratégies de sélection des "caractéristiques spécifiques aux émotions" et des "caractéristiques les plus discriminantes pour toutes les émotions" ont été développées, cette dernière stratégie s'étant avérée plus efficace à l'aide du test statistique de Friedman. Ce système FER indépendant de la personne avec la sélection de caractéristiques proposée a été validé sur trois bases de données d'évaluation couramment utilisées, où le taux moyen de reconnaissance des émotions était de 98,37 % sur la base de données Cohn Kanade, de 92,75 % sur la base de données JAFFE et de 84,07 % sur la base de données MMI, tandis que le nombre de caractéristiques utilisées a diminué jusqu'à 89 % par rapport à la longueur du vecteur de caractéristiques de différence d'origine.

Ira Cohen et all [24] Les expressions faciales sont la façon la plus expressive dont les humains manifestent leurs émotions. Dans ce travail, ils ont fait état de plusieurs avancées dans la construction d'un système de classification des expressions faciales à partir d'une entrée vidéo continue. Introduit et testé différents classificateurs de réseaux bayésiens pour la classification des expressions à partir de vidéos, en se concentrant sur les changements dans les hypothèses de distribution et les structures de dépendance des caractéristiques. En particulier, ils ont utilisé des classificateurs Naive-Bayes et changeons la distribution de Gaussien à Cauchy, et des

classificateurs Naive Bayes gaussiens à arbre renforcé (TAN) pour apprendre les dépendances entre les différentes caractéristiques de mouvement du visage. Ils ont introduit également une reconnaissance d'expression faciale à partir d'une entrée vidéo en direct en utilisant des indices temporels. Exploiter les méthodes existantes et proposons une nouvelle architecture de modèles de Markov cachés (HMM) pour segmenter et reconnaître automatiquement les expressions faciales humaines à partir de séquences vidéo. L'architecture effectue à la fois la segmentation et la reconnaissance des expressions faciales automatiquement à l'aide d'une architecture à plusieurs niveaux composés d'une couche de HMM et d'une couche de modèle de Markov. Explorer la reconnaissance des expressions en fonction de la personne et indépendamment de la personne et comparons les différentes méthodes.

Franck Davoine et all [28] Dans cet article, nous nous intéressons à l'extraction automatique des traits de visages (yeux, sourcils, nez, bouche, menton) ainsi qu'à la reconnaissance des six expressions faciales définies par Ekman. Exploitons pour cela des versions modifiées du modèle actif d'apparence initialement proposé par Cootes et al. qui permet de représenter à la fois la forme et la texture d'un visage. L'extraction des traits faciaux est faite à l'aide d'un modèle actif d'apparence hiérarchique, calculé à partir des réponses de visages à des bancs de filtres de Gabor. Deux modèles d'expressions faciales sont ensuite proposés, calculés à patir du modèle d'apparence standard (non hiérarchique), pour reconnaître puis supprimer ou modifier l'expression d'un visage inconnu.

Maja Pantic et all [76] La reconnaissance automatique des gestes faciaux (c'est-à-dire l'activité des muscles du visage) devient rapidement un domaine d'intérêt intense dans le domaine de la recherche en vision artificielle. Dans cet article, il a été présenté un système automatisé qu'ils ont mis au point pour reconnaître les gestes faciaux dans des images statiques de visages en couleur, en vue frontale et/ou de profil. Une approche multidétecteur de la localisation des caractéristiques du visage est utilisée pour échantillonner spatialement le contour du profil et les contours des composants du visage tels que les yeux et la bouche. À partir des contours extraits des caractéristiques faciales, ils ont extait dix points de repère du contour du profil et 19 points de repère des contours des composantes faciales. Sur la base de ces points, 32 actions musculaires faciales individuelles (UA) survenant seules ou en combinaison sont reconnues à l'aide d'un raisonnement basé sur des règles. À chaque UA notée, l'algorithme utilisé associe un facteur indiquant la certitude avec laquelle l'UA pertinente a été notée. Un taux de reconnaissance de 86 % est atteint.

Ahmed Maalej et all [59] Dans cet article, nous abordons le problème de la reconnaissance des expressions faciales en 3D. il a été proposé une analyse géométrique locale des surfaces faciales couplée à des techniques d'apprentissage automatique pour la classification des expressions. Un calcul de la longueur du chemin géodésique entre des taches correspondantes, utilisant un cadre riemannien, dans un espace de forme fournit une information quantitative sur leurs similitudes. Ces mesures sont ensuite utilisées comme entrées dans plusieurs méthodes de classification. Les résultats expérimentaux démontrent l'efficacité de l'approche proposée. En utilisant des classificateurs multiboosting et des machines à vecteurs de support (SVM), nous avons obtenu des taux de reconnaissance moyens de 98,81 % et 97,75 %, respectivement, pour la reconnaissance des six expressions faciales prototypiques sur la base de données BU-3DFE. Une étude comparative utilisant le même cadre expérimental montre que l'approche proposée

est plus performante que les travaux antérieurs.

Dennis Núñez Fernández [31] Une expression faciale est un ou plusieurs mouvements ou positions des muscles sous la peau du visage. Ces mouvements traduisent l'état émotionnel d'un individu aux observateurs. Plusieurs travaux ont été développés dans le domaine de l'apprentissage automatique sur la reconnaissance des expressions faciales en utilisant plusieurs algorithmes pour l'extraction des caractéristiques (statistiques ou structurelles) et des classifieurs. Ces travaux ont prouvé leurs puissances en termes du taux de reconnaissance sur les petites bases de données, tout de même, ces résultats restent limités dans le cadre de traitement de très grande masse de données. Avec l'apparition du concept de deep Learning (apprentissage en profondeur) et les bases de données volumineuses, un nouvel axe de recherche est développé. Ce projet consiste à proposer une approche de reconnaissance des expressions faciales basée sur le deep Learning et plus particulièrement les réseaux de neurones convolutifs. Un test expérimental a été fait sur la base de données FER 2013, le taux de reconnaissance obtenu 61,24 % et les résultats sont prometteurs.

Hui Ding et all [32] Les ensembles de données relativement petits disponibles pour la recherche sur la reconnaissance d'expressions rendent très difficile l'entraînement des réseaux profonds pour la reconnaissance d'expressions. Bien que le réglage fin puisse partiellement atténuer le problème, les performances restent inférieures aux niveaux acceptables, car les caractéristiques profondes contiennent probablement des informations redondantes provenant du domaine pré-entraîné. Dans cet article, ils ont présenté FaceNet2ExpNet, une nouvelle idée pour former un réseau de reconnaissance d'expressions basé sur des images statiques. ils ont tout d'abord proposé une nouvelle fonction de distribution pour modéliser les neurones de haut niveau du réseau d'expression. Sur cette base, un algorithme de formation en deux étapes est soigneusement conçu. Dans la phase de pré-entraînement, ils ont entraîné les couches convolutives du réseau d'expression, régularisées par le réseau de visage ; dans la phase d'affinage, ajouté des couches entièrement connectées aux couches convolutives pré-entraînées et entraîné l'ensemble du réseau conjointement. La visualisation montre que le modèle formé avec notre méthode capture une sémantique d'expression de haut niveau améliorée. Des évaluations sur quatre bases de données publiques d'expressions, CK+, Oulu-CASIA, TFD, et SFEW démontrent que notre méthode obtient de meilleurs résultats que l'état de l'art.

1.8 Conclusion

Dans ce chapitre, nous avons défini la reconnaissance de l'expression, ensuite, nous avons cité les différents problèmes rencontrés ainsi que les divers domaines d'application de cette dernière dans le monde réel. Enfin, nous avons présenté quelques travaux connexes sur la reconnaissance de l'expression à partir des images faciales. Dans le chapitre suivant, on s'intéressera de plus près au domaine de l'apprentissage profond et des réseaux de neurones, plus particulièrement au réseau de neurones convolutif.

Chapitre 2

APPRENTISSAGE PROFOND ET RÉSEAUX NEURONAUX

2.1 Introduction

Le Deep Learning représente une évolution significative du domaine du Machine Learning, ayant pour objectif premier de rapprocher ce dernier de son aspiration fondamentale : l'intelligence artificielle. Cette approche se fonde sur des algorithmes inspirés par la structure et le fonctionnement du cerveau humain, leur permettant d'apprendre et de représenter plusieurs niveaux de complexité dans le but de modéliser des relations intrinsèquement complexes entre les données.

Dans ce chapitre, nous aborderons diverses notions essentielles liées aux réseaux de neurones, notamment leurs définitions et les différentes topologies qui les caractérisent. Nous plongerons ensuite plus en profondeur dans le domaine du Deep Learning, explorant les différentes architectures existantes qui en découlent. Notre attention sera principalement portée sur un type spécifique de réseau de neurones, à savoir les Réseaux de Neurones à Convolution (CNN, Convolutional Neural Networks), en raison de leur importance cruciale dans le traitement des données visuelles et la reconnaissance d'objets.

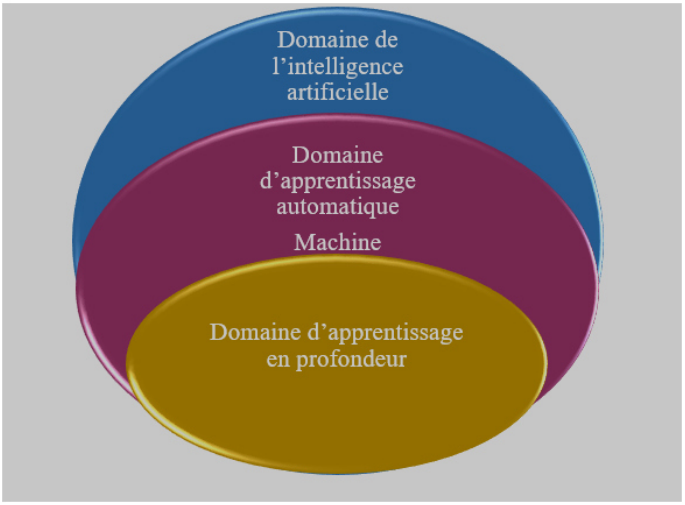

FIGURE 2.1 – La relation entre l'intelligence artificielle, le ML et le deep Learning [5].

2.2 Réseaux De Neurones

Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique de réseaux est évidemment un réseau.

2.2.1 Définition

Un réseau de neurones artificiels est un système dont la conception est à l'origine inspirée du fonctionnement des neurones biologiques, et qui par la suite s'est rapproché des méthodes statistiques. Les réseaux de neurones artificiels sont des réseaux fortement connectés par des processeurs élémentaires fonctionnant en parallèle.

Chaque processeur élémentaire (neurone artificiel) calcule une sortie unique sur la base des informations qu'ils reçoivent. Les points essentiels qu'on peut retenir sur les réseaux de neurones sont les suivants [67] :

– Ce sont des systèmes composés de neurones répartis en plusieurs couches connectées entre elles.
– Ces systèmes peuvent résoudre de divers problèmes statistiques en général, et spécialement des problèmes de classification, en calculant à partir de l'entrée du réseau le score (ou la probabilité) de chaque classe. La classe attribuée à l'objet, c'est celle de la probabilité la plus élevée.
– L'entrée de chaque couche, les données sont traitées et transformée en calculant une combinaison linéaire, puis en appliquant une fonction non-linéaire, appelée fonction d'activation. Les coefficients de la combinaison linéaire définissent les paramètres (ou poids) de la couche.
– La dernière couche calcule les probabilités finales à partir d'une fonction d'activation (classification binaire) ou la fonction softmax (classification multiclasses).
– On associe aussi une fonction de perte loss-function à la couche finale pour calculer l'erreur de classification. Il s'agit en général de l'entropie croisée (accurancy).
– On calcule les poids des couches par rétro-propagation du gradient : Calcule les paramètres qui minimisent la fonction de perte régularisée progressivement en partant de la dernière couche à la première couche, L'optimisation se fait avec une descente du gradient stochastique.

2.2.2 Topologie

Chaque réseau de neurones est connecté entre eux de diverses manières. Dans la figure suivante. Nous pouvons distinguer deux familles de réseaux de neurones : non bouclés ou statiques (a) et (b) et bouclés (dynamiques) (c) et (d).

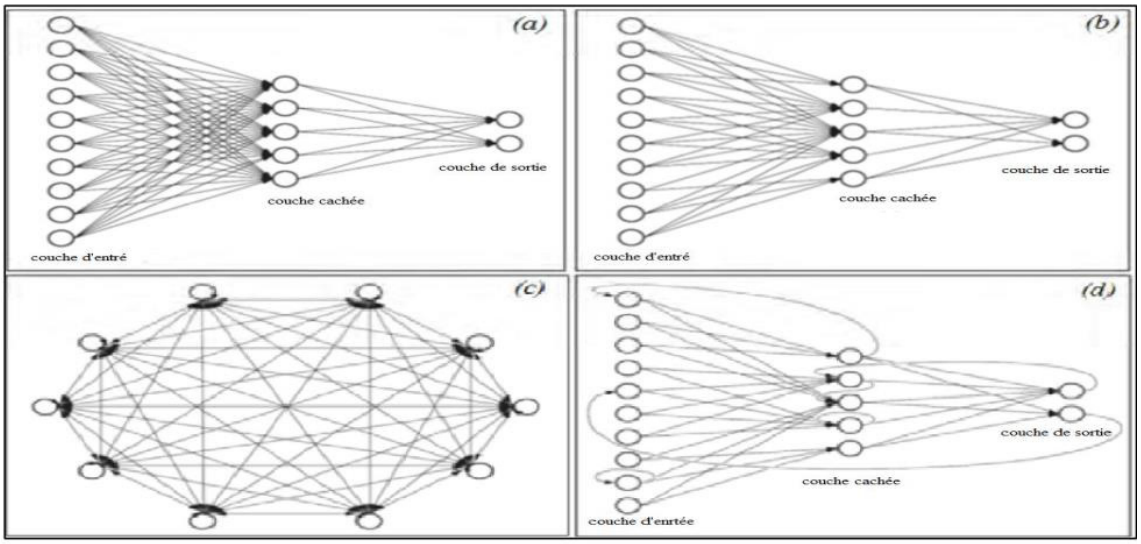

FIGURE 2.2 – Topologie des Réseaux de neurones artificiels [22].

2.2.3 Les différentes Architectures du Deep Learning

Bien qu'il existe un grand nombre de variantes d'architectures profondes[figure 2.5]. Il n'est pas toujours possible de comparer les performances de toutes les architectures, car elles ne sont pas toutes évaluées sur les mêmes ensembles de données. Le Deep Learning est un domaine à croissance rapide, et de nouvelles architectures, variantes ou algorithmes apparaissent toutes les semaines.

2.2.3.1 Les Réseaux de Neurones Convolutifs

Convolutional Neural Network (CNN) sont un type de réseau de neurones spécialisés pour le traitement de données ayant une topologie semblable à une grille. Qui se sont avérés très efficaces dans des domaines tels que la reconnaissance et la classification d'images et vidéos. CNN a réussi à identifier les visages, les objets, panneaux de circulation et auto-conduite des voitures [67]. Récemment, les CNN ont été efficaces dans plusieurs tâches de traitement du langage naturel (telles que la classification des phrases) [20] [44] [25]. Dans le ML, un réseau convolutif est un type de réseau de neurones feed-forward, il a été inspiré par des processus biologiques. [61]

Il existe quatre (4) principales opérations illustrées dans le CNN et présenté dans la figure 2.3 :

– La couche convolution
– La couche Rectified Linear Unit
– La couche Pooling
– La couche entièrement connectée

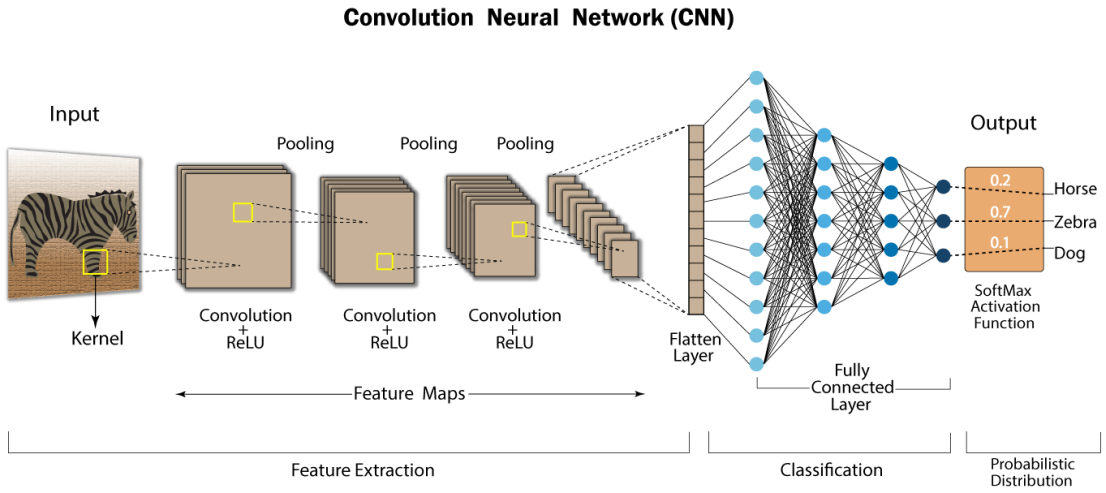

FIGURE 2.3 – Convolutionnal Neural Network [27].

2.2.3.2 Réseau de Neurones Récurrents

L'idée derrière les RNN est d'utiliser des informations séquentielles. Dans un réseau neuronal traditionnel, nous supposons que toutes les entrées (et les sorties) sont indépendantes les unes des autres. Mais pour de nombreuses tâches, c'est une très mauvaise idée. Si on veut prédire le prochain mot dans une phrase, il faut connaître les mots qui sont venus avant. Les RNN sont appelés récurrents, car ils exécutent la même tâche pour chaque élément d'une séquence, la sortie étant dépendante des calculs précédents. Une autre façon de penser les RNN est qu'ils ont une « mémoire » qui capture l'information sur ce qui a été calculé jusqu'ici, [figure 2.4]. En théorie, les RNN peuvent utiliser des informations dans des séquences arbitrairement longues, mais dans la pratique, on les limite à regarder seulement quelques étapes en arrière [62]. Il est utilisé pour :

– La modélisation du langage et génération de texte
– La traduction automatique
– La reconnaissance vocale
– Et la description des images

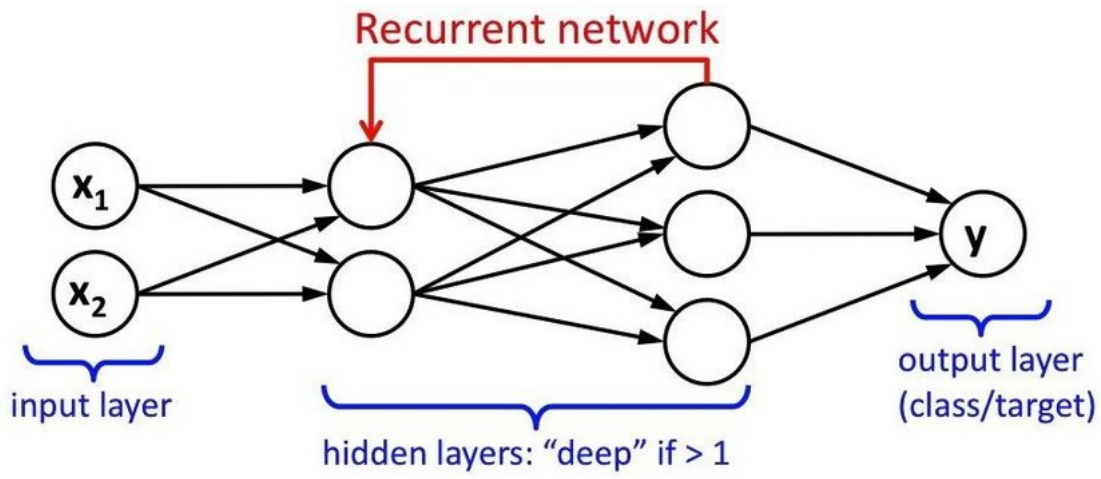

FIGURE 2.4 – Recurrent-neural-networkRNN-or-Long-Short-Term-MemoryLSTM [65].

2.2.3.3 Modèle Génératif

Si les modèles discriminatifs comme (CNN, RNN) sont utilisés pour prédire les données du label et de l'entrée, tant disque le modèle génératif décrit comment générer les données, il apprend et fait des prédictions en utilisant la loi de Bayes. [71] Cependant, les modèles génératifs sont capables de bien plus que la simple classification, comme par exemple générer de nouvelles observations.
Voici quelques exemples de modèle génératif :
– Boltzmann Machines [12]
– Restricted Boltzmann Machines [79]
– Deep Belief Networks [48]
– Deep Boltzmann Machines [82]
– Generative Adversarial Networks
– Generative Stochastic Networks [38]
– Adversarial auto encoders [29]

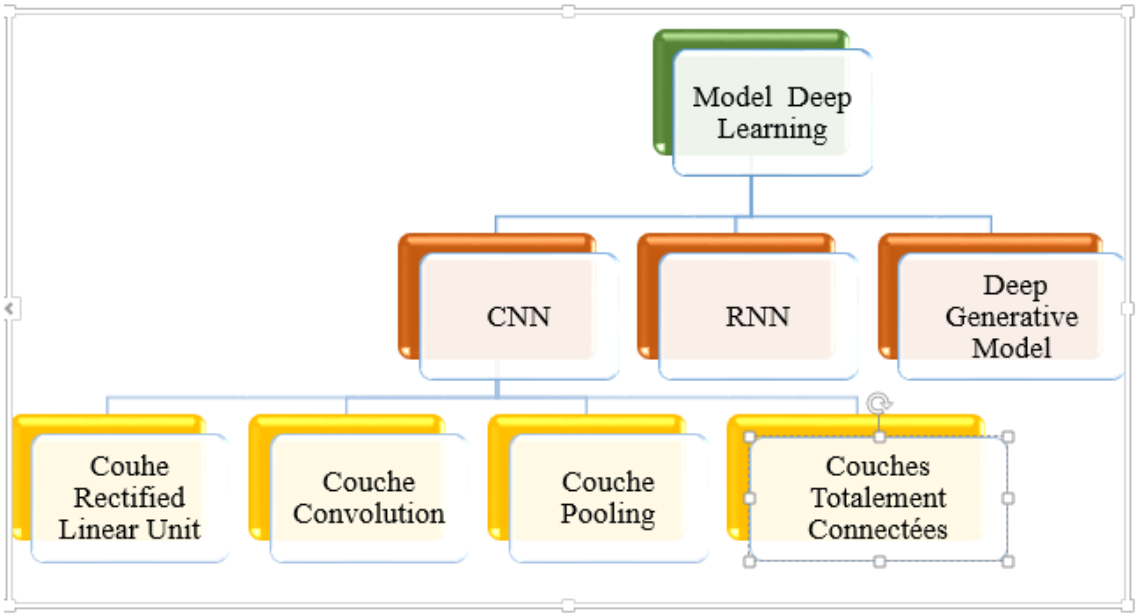

FIGURE 2.5 – Différents modèles du Deep Learning [23].

2.3 L'Apprentissage En Profondeur (Le Deep Learning)

Le terme "Deep Learning" ou Apprentissage profond, a été introduit pour la première fois au ML par Dechter (1986), et aux réseaux neuronaux artificiels par Aizenberg et al (2000) [13].

2.3.1 Définition

L'apprentissage en profondeur est un ensemble d'algorithmes d'apprentissage automatique qui tentent d'apprendre à plusieurs niveaux, correspondant à différents niveaux d'abstraction. Il a la capacité d'extraire des caractéristiques à partir des données brutes grâce aux multiples couches de traitement composé de multiples transformations linéaires et non linéaires et apprendre sur ces caractéristiques petites à petit à travers chaque couche avec une intervention humaine minime comme sur la figure 2.7 [89].

FIGURE 2.6 – Un processus de Deep Learning : les images sont transmises à un réseau, qui apprend automatiquement les caractéristiques et classe les objets [90].

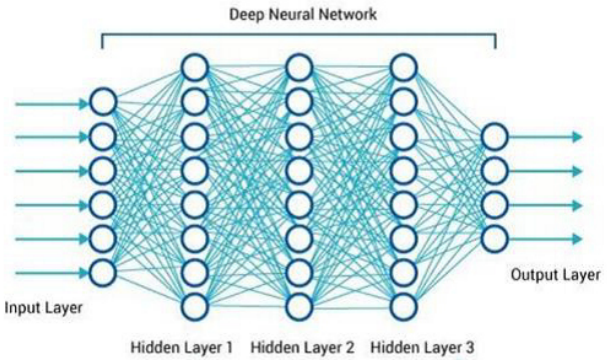

FIGURE 2.7 – Schéma illustratif de DL avec plusieurs couches [31].

2.3.2 Pour quoi le choix du Deep Learning ?

Les algorithmes de ML fonctionnent bien pour une grande variété de problèmes. Cependant, ils ont échoué à résoudre quelques problèmes majeurs de l'IA telle que la reconnaissance vocale et la reconnaissance d'objets [67]. Tout d'abord, les différents algorithmes du deep Learning ne sont apparus qu'à l'échec de l'apprentissage automatique tentant de résoudre une grande variété de problèmes de l'intelligence artificielle (l'IA) :

– Afin d'améliorer le développement des algorithmes traditionnels dans de telles tâches de l'IA.
– De développer une grande quantité de données telle que les big data.
– De s'adapter à n'importe quel type de problème.
– D'extraire les caractéristiques de façon automatique. [Figure 2.8]

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

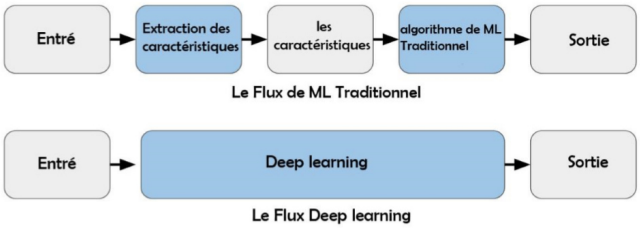

FIGURE 2.8 – Comparaison entre la machine Learning et le Deep Learning [20].

2.3.3 Réseaux de Neurones Convolutifs CNN

Durant cette phase nous présentant uniquement sur le CNN, qui sera le réseau de neurone utilsé par notre système automatique pour l'extraction de nos caractéristiques.

2.3.3.1 Présentation

Les réseaux de neurones convolutifs CNN (Convolutional Neural Network) sont les structures les plus performantes dans des domaines tels que la reconnaissance et la classification d'images. CNN a réussi à identifier les visages, les objets, panneaux de circulation et auto-conduite des voitures, durant cette phase de formation, nous nous baserons uniquement sur le CNN. Le nom « réseau de neurones convolutif » indique que le réseau emploie une opération mathématique appelée convolution. La convolution est une opération linéaire spéciale. Les réseaux convolutifs sont simplement des réseaux de neurones qui utilisent la convolution à la place de la multiplication matricielle dans moins une de leurs couches.
Ils comportent deux parties principales. S'il y a on a en entrée, une image qu'elle doit être sous la forme d'une matrice de pixels de 2 dimensions pour une image en niveaux de gris et en 3 dimensions si elle est en couleur, pour représenter les couleurs fondamentales [Rouge, Vert, Bleu].

Cette image passe par la première partie d'un CNN qui est la partie convolutive et elle fonctionne comme un extracteur de caractéristiques des images. L'image passe à travers une succession de filtres ou noyaux de convolution, pour la transformée en nouvelles images appelées cartes de convolutions «feature maps ». Certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum local. Au final, les cartes de convolutions sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN.

Le résultat en sortie de la partie convolutive est branché en entrée d'une deuxième partie, constituée de couches entièrement connectées (perceptron multicouche) qui consiste à combiner les caractéristiques de tout le réseau pour classer l'image et à la sortie qui est une couche comportant un neurone par classe, on obtient des valeurs numériques généralement normalisées entre 0 et 1, pour présenter la distribution de probabilité sur les classes. [56]

2.3.3.2 Architecture des Réseaux de Neurone Convolutifs

Comme nous l'avons mentionnée précédemment, les réseaux de neurones convolutifs sont basés sur le perceptron multicouche(MLP). L'architecture CNN est formée par un empilement de couches de traitement indépendantes comme c'est montrée dans la figure 2.9 :

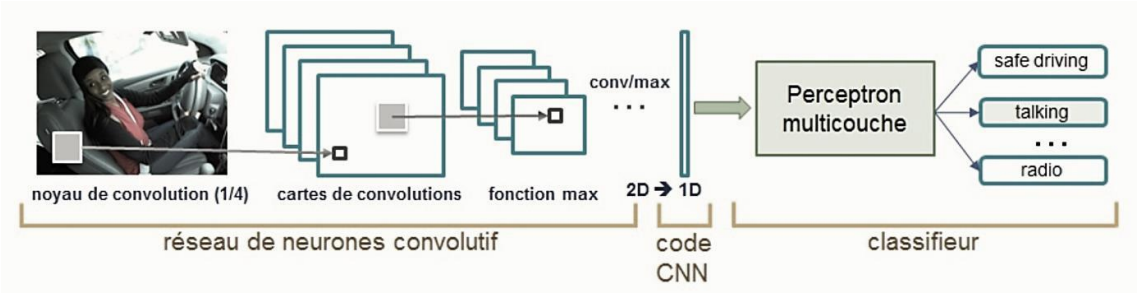

FIGURE 2.9 – Architecture standard d'un réseau de neurone convolutionnel [41].

2.3.3.3 Les différentes couches

1. La Couche De Convolution (CONV) Trois hyper paramètres permettent de dimensionner le volume de la couche de convolution (aussi appelé volume de sortie) : La profondeur, le pas et la marge.
   – Profondeur de la couche : nombre de noyaux de convolution (ou nombre de neurones associés à un même champ récepteur).
   – Le pas : contrôle le chevauchement des champs récepteurs. Plus le pas est petit, plus les champs récepteurs se chevauchent et plus le volume de sortie sera grand.
   – La marge (à 0)' ou 'zero padding ' : parfois, il est commode de mettre des zéros à la frontière du volume d'entrée. La taille de ce 'zero-padding' est le troisième hyper-paramètre. Cette marge permet de contrôler la dimension spatiale du volume de sortie.
   En particulier, il est parfois souhaitable de conserver la même surface que celle du volume d'entrée [49].

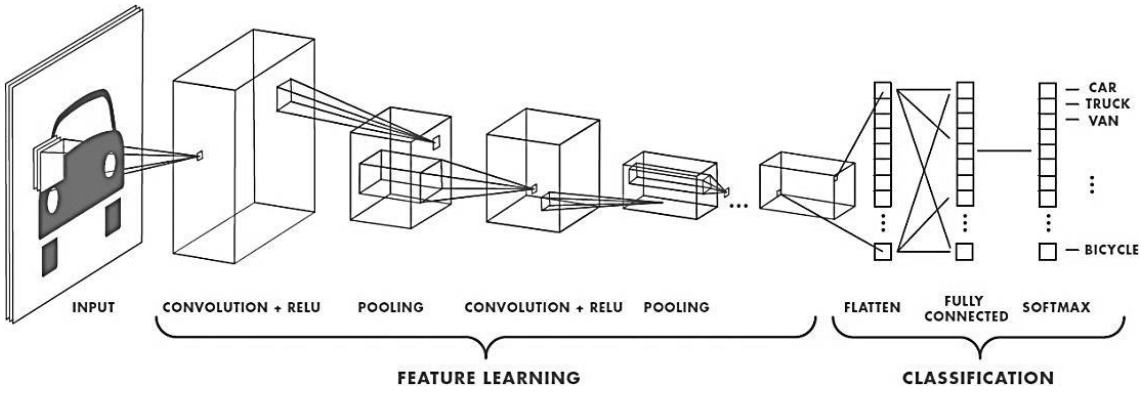

FIGURE 2.10 – Exemple de réseau composé de nombreuses couches à convolution. Des filtres sont appliqués à chaque image utilisée pour l'apprentissage à différentes résolutions, et la sortie de chaque image convoluée est utilisée comme entrée de la couche suivante [14].

Dans la terminologie du réseau convolutif, le premier argument de la convolution est souvent appelé l'entrée (input) et le second argument comme noyau (kernel). La sortie est parfois appelée la carte des caractéristiques (feature map).

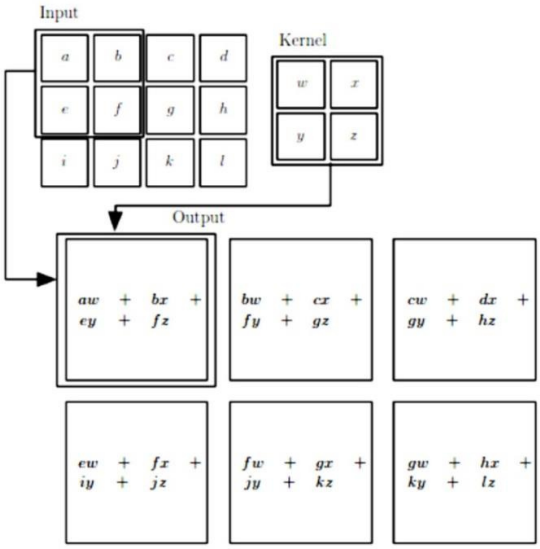

FIGURE 2.11 – Exemple d'une convolution 2D [37].

2. La Couche De Pooling Le pooling est une forme de sous-échantillonnage de l'image, il permet de réduire progressivement la taille des représentations afin de réduire la quantité de paramètres et de calcul dans le réseau ainsi que l'invariance aux petites translations, il est donc fréquent d'insérer périodiquement une couche de pooling entre deux couches convolutives successives d'une architecture CNN pour contrôler le sur-apprentissage.
   L'opération de pooling créait aussi une forme d'invariance par translation. La couche de pooling fonctionne indépendamment sur chaque tranche de profondeur de l'entrée et la redimensionne uniquement au niveau de la surface. La forme la plus courante est une couche de mise en commun avec des filtres de taille 2x2 (largeur/hauteur) et comme valeur de sortie la valeur maximale en entrée. On parle dans ce cas de « Max-Pool 2x2 » [49].
   Le pooling permet de gros gains en puissance de calcul. Cependant, en raison de la réduction agressive de la taille de la représentation et donc de la perte d'information associée, la tendance actuelle est d'utiliser de petits filtres (type 2x2) comme c'est démontré dans la figure 2.12. Il est aussi possible d'éviter la couche de pooling mais cela implique un risque sur-apprentissage plus important.
   Il existe plusieurs types de pooling :
   – Le « max pooling », qui revient à prendre la valeur maximale de la sélection. C'est le type le plus utilisé, car il est rapide à calculer (immédiat), et permet de simplifier efficacement l'image.
   – Le « mean pooling » (ou average pooling), soit la moyenne des pixels de la sélection : On calcule la somme de toutes les valeurs et on divise par le nombre de valeurs. On obtient ainsi une valeur intermédiaire pour représenter ce lot de pixels comme le montre la figure 2.13.
   – Le « sum pooling », c'est la moyenne sans avoir divisé par le nombre de valeurs (on ne

calcule que leur somme).

– La séparable convolution, qui est une convolution décomposable en convolutions plus simples.

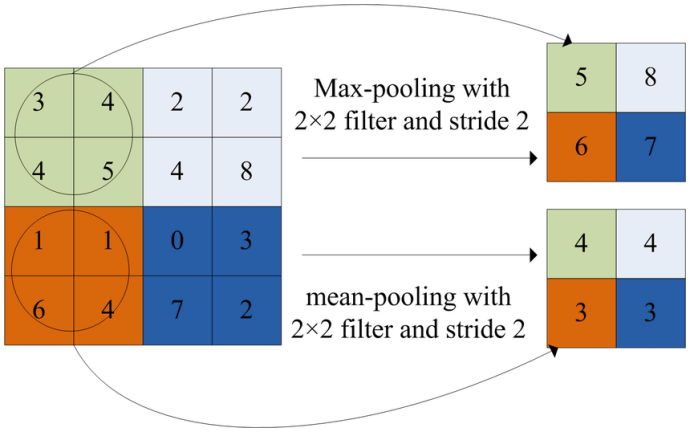

FIGURE 2.12 – Pooling avec un filtre 2x2 et un pas de 2 [60].

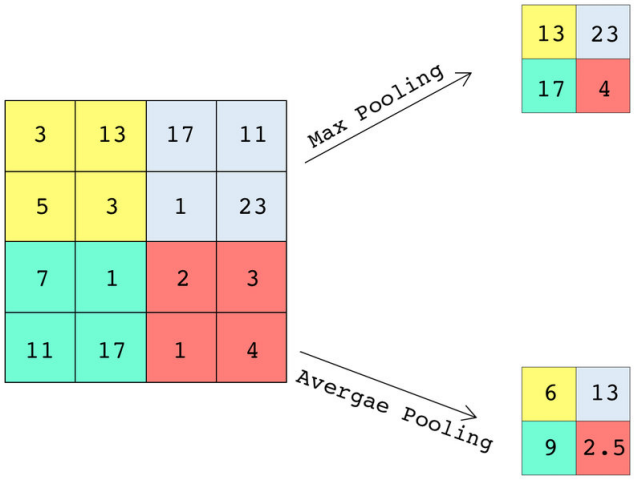

FIGURE 2.13 – (Exemple d'opérations de pooling maximum et de pooling moyen. Dans cet exemple, une image 4x4 est sous-échantillonnée en 2x2 en prenant la valeur maximale ou la valeur moyenne de chaque sous-région [15].

3. Couches Entièrement Connectées Après l'extraction des caractéristiques des entrées, on attache à la fin du réseau un perceptron ou bien un MLP (multi layer perceptron). Le perceptron prend comme entrée les caractéristiques extraites et produit un vecteur de N dimensions ou N est le nombre de classes ou chaque élément est la probabilité d'appartenance à une classe.
   Chaque probabilité est calculée à l'aide de la fonction softmax dans le cas où les classes sont exclusivement mutuelles. Le terme « entièrement connecté » implique que chaque neurone dans la couche précédente est connecté à chaque neurone sur la couche suivante comme le montre la figure 2.14 [49].

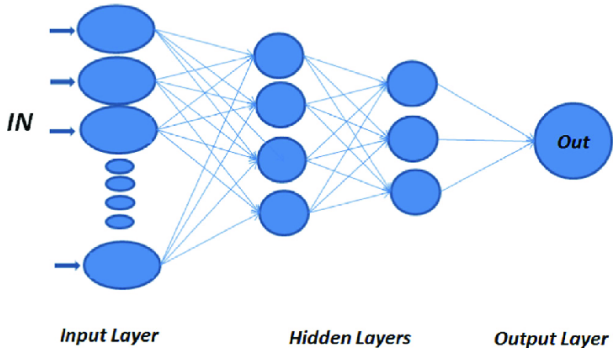

FIGURE 2.14 – Schéma représentatif des connexions entre les neurones [33].

4. Les Fonctions d'Activation La fonction d'activation est une fonction mathématique appliquée à un signal en sortie d'un neurone artificiel. Le terme de "fonction d'activation" vient de l'équivalent biologique "potentiel d'activation", seuil de stimulation qui, une fois atteint entraîne une réponse du neurone. La fonction d'activation est souvent une fonction non-linéaire.
   Leur but est de permettre aux réseaux de neurones d'apprendre des fonctions plus complexes qu'une simple régression linéaire, car le fait de multiplier les poids d'une couche cachée est juste une transformation linéaire.
   La figure 2.15 présente un exemple de fonctions d'activations comme Le ReLu (Rectified Linear Units) : Elle est utilisée après chaque opération de convolution, ou toutes les valeurs de pixels négatifs sont mises à zéro.

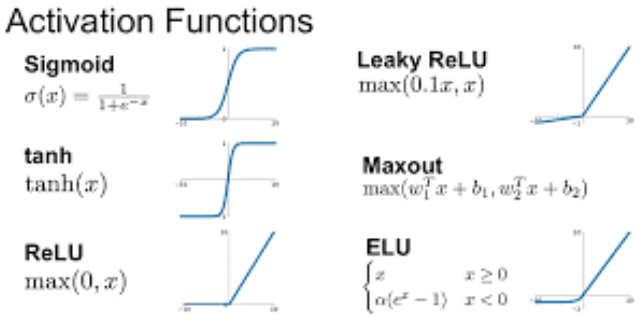

FIGURE 2.15 – Quelques fonctions d'activation [8].

2.4 Quelques réseaux convolutifs célèbres

Nous allons citer les réseaux convolutifs les plus célèbres :

2.4.1 LeNet

Les premières applications réussies des réseaux convolutifs ont été développées par Yann LeCundans les années 1990. Parmi ceux-ci, le plus connu est l'architecture LeNet utilisée pour lire les codes postaux, les chiffres, etc. La figure 2.16 montre l'architecture de LeNet [53].

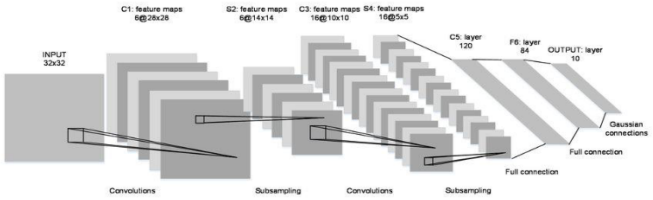

FIGURE 2.16 – La figure montre une architecture de LeNet [10].

2.4.2 AlexNet

Le premier travail qui a popularisé les réseaux convolutifs dans la vision par ordinateur était AlexNet, développé par Alex Krizhevsky, Ilya Sutskever et Geoff Hinton en 2012. Ce CNN était soumis au défi de la base ImageNet et a nettement surpassé ses concurrents. Le réseau avait une architecture très similaire à LeNet, mais était plus profond, plus grand et comportait des couches convolutives empilées les unes sur les autres (auparavant, il était commun de ne disposer que d'une seule couche convolutive toujours immédiatement suivie d'une couche de pooling).La figure 2.17 montre l'architecture de AlexNet [49].

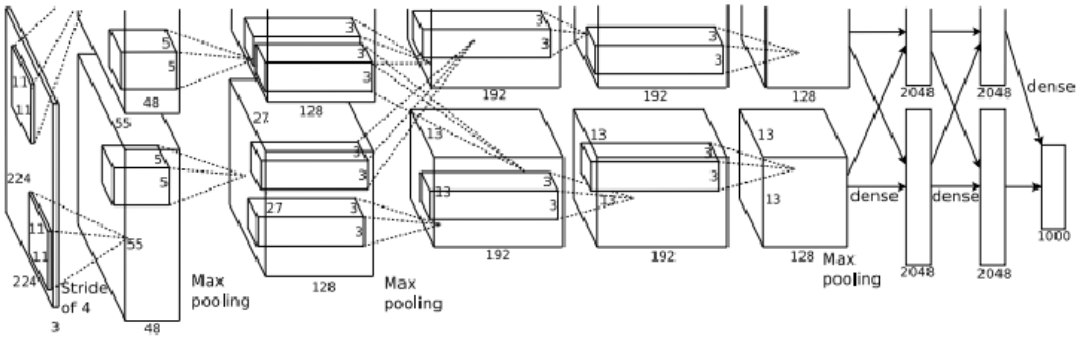

FIGURE 2.17 – La figure montre une architecture de AlexNet [6].

2.4.3 Overfeat

C'est un classificateur d'image basé sur un réseau convolutionnel et un extracteur de fonctionnalités. Il a été formé sur le jeu de données Image Net et a participé au concours Image Net 2013. La figure 2.18 montre l'architecture de Overfeat [80].

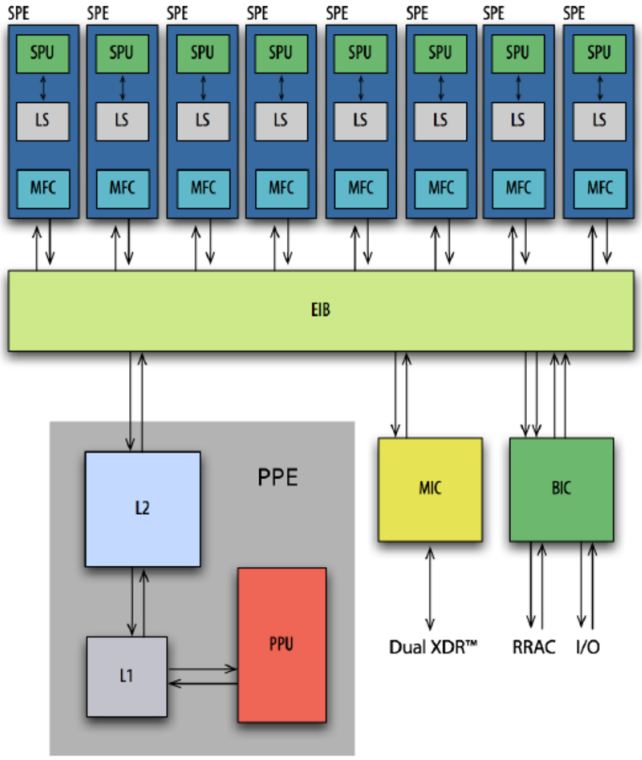

FIGURE 2.18 – La figure montre une architecture de Overfeat [7].

2.4.4 Inception V3

Ce type d'architecture, introduit en 2016 par Szegedy et al, utilise des blocs avec des filtres de différentes tailles qui sont ensuite concaténés pour extraire des caractéristiques à différentes échelles.La figure 2.19 montre l'architecture de Inception V3 [84].

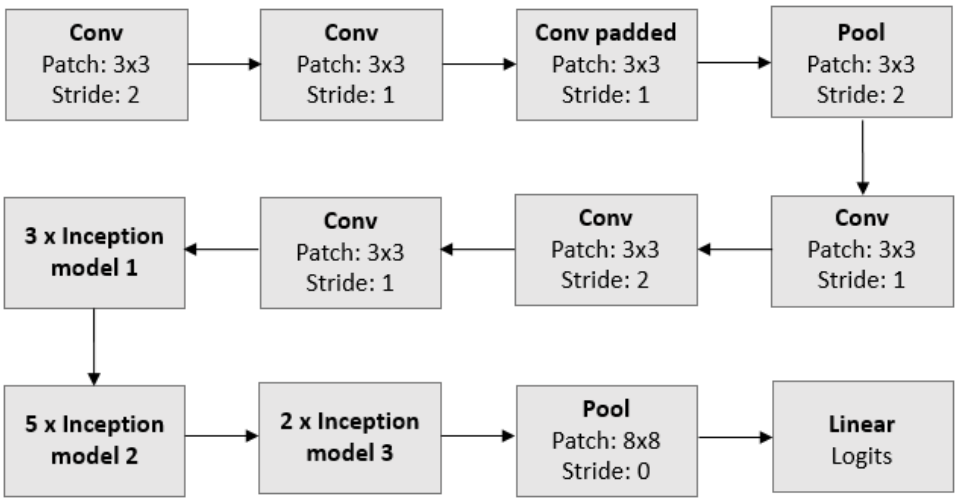

FIGURE 2.19 – La figure montre une architecture de Inception V3 [72].

2.4.5 ResNet : Residual Neural Network

ResNet est une architecture de réseau profond avec 152 couches. Ceci a été développé par Microsoft en 2016. Il a remporté l'ILSVRC 2016 avec un taux d'erreur de 3,6 %, ce qui est considéré comme meilleur que la précision au niveau humain. Les couches résiduelles dans ResNet calculent les changements dans l'entrée. Ceci est ensuite ajouté à l'entrée pour produire la sortie. ResNet-50 est l'un des premiers à adopter la normalisation des batchs. La figure 2.20 montre l'architecture de ResNet [42].

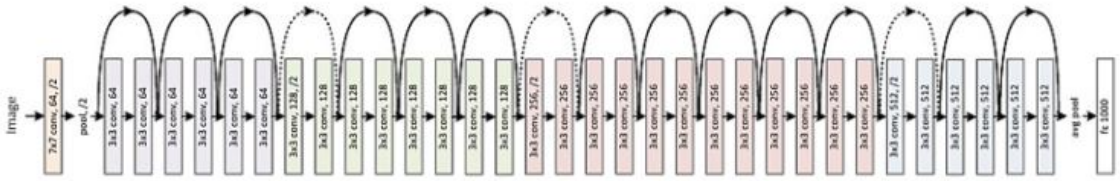

FIGURE 2.20 – La figure montre une architecture de ResNet [9].

2.4.6 VGG-16

Ren Simonyan et Andrew Zisserman d'Oxford Vision Geometry Group (VGG) ont proposé le VGG-16 qui a 13 couches de convolution et trois couches entièrement connectés. VGGNet utilise des filtres de taille 3x3 comparant de 11x11 de AlexNet et 7x7 de ZFNet. Les auteurs donnent l'intuition derrière cela qu'avoir deux filtres de taille 3x3 consécutifs donne un champ récepteur efficace de 5x5, et trois séries de filtres de taille 3x3 donnent un champ. Récepteur de filtres 7x7, mais en utilisant cela, nous pouvons utiliser un nombre beaucoup moins élevé d'hyper-paramètres. Ils ont aussi proposé une variante plus profonde VGG-19 [81]. La figure 2.21 montre l'architecture de VGG-16.

FIGURE 2.21 – La figure montre une architecture de VGG-16 [11].

2.5 Classification

La classification automatique des images consiste à attribuer automatiquement une classe à une image à l'aide d'un système de classification. On retrouve ainsi la classification d'objets, de scènes, de textures, la reconnaissance de visages, d'empreintes digitale et de caractères. La classification des images consiste à repartir systématiquement des images selon des classes établies au préalable, classer une image lui fait correspondre une classe, marquant ainsi son émotion avec d'autres images. En général, reconnaitre une image est une tâche aisée pour un humain au fil de son existence, il a acquis des connaissances qui lui permettent de s'adapter aux variations qui résultent différentes conditions d'acquisition. Il lui est par exemple relativement simple de reconnaitre un objet dans plusieurs orientations, partiellement caché par un autre de près ou de loin et selon diverses illuminations. L'objectif de la classification d'images est d'élaborer un système capable d'affecter une classe automatiquement à une image. Ainsi, ce système permet d'effectuer une tâche d'expertise qui peut s'avérer couteuse à acquérir pour un être humain en raison notamment de contraintes physiques comme la concentration, la fatigue ou le temps nécessité par un volume important de données images important.

Les applications de la classification automatique d'images sont nombreuses et vont de - L'analyse de documents à la médecine en passant par le domaine militaire. - Ainsi, on retrouve des applications dans le domaine médical comme la reconnaissance de cellules et de tumeurs, la reconnaissance d'écriture manuscrite pour les chèques, les codes postaux. - Dans le domaine urbain comme la reconnaissance de panneaux de signalisation, la reconnaissance de piétons, la détection de véhicules, la reconnaissance de bâtiments pour aider à la localisation. - Dans le domaine de la biométrie comme la reconnaissance de visage, d'empreintes, d'iris. Le point commun a toutes ces applications est qu'elles nécessitent la mise en place d'une chaine de traitement à partir des images disponibles composée de plusieurs étapes afin de fournir en sortie une décision. [30] La figure 2.22 montre un exemple de classification de l'émotion à partir d'un CNN.

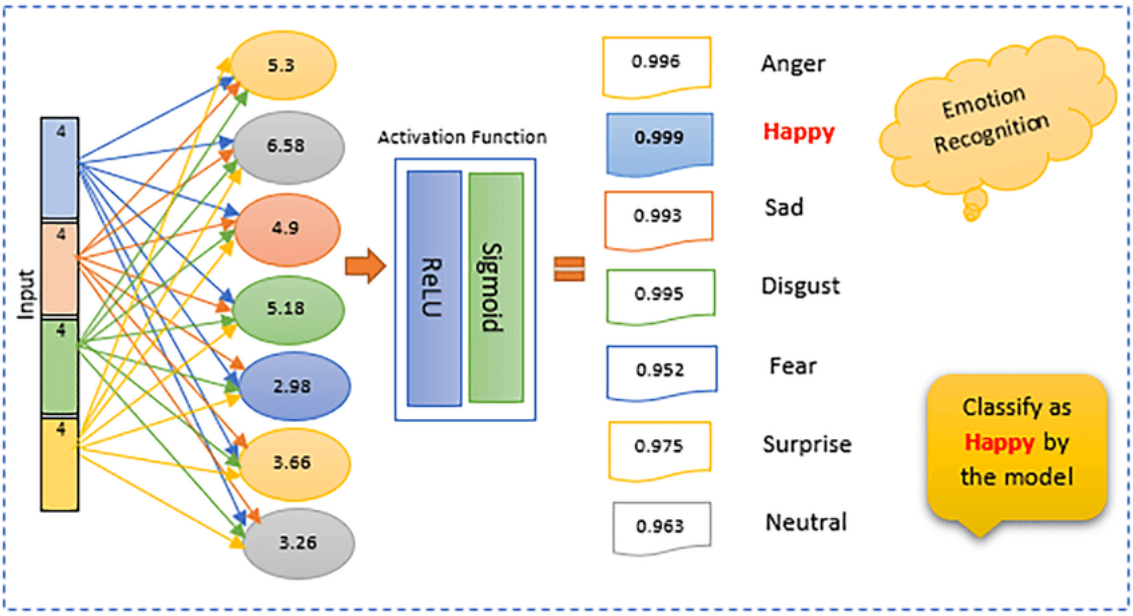

FIGURE 2.22 – Exemple de classification de l'émotion à partir d'un CNN [46].

2.6 Conclusion

Dans ce chapitre, nous avons vu ce que les réseaux de neurones et leurs différents types, ensuite, on a essayé d'expliquer les réseaux de neurones convolutifs CNN et leur structure, et ses différentes couches. Le CNN à quatre principales opérations : convolution, la fonction non-linéarité (ReLU), Pooling et couche entièrement connectée.

Première opération après la détection de visage est la convolution pour l'extraction de caractéristiques de l'image d'entrée. La deuxième opération est la fonction non-linéarité (ReLU) pour remplacer toutes les valeurs de pixels négatives par zéro. Troisième opération est la Pooling pour réduire progressivement la taille de la carte de caractéristiques rectifiée. Enfin une couche entièrement connectée pour la classification.

À la fin, nous avons présenté quelques exemples d'architectures, dont parmi eux, ceux que nous allons utiliser dans notre modèle VGG16 pour l'extraction et un deuxième CNN pour la classification, ceci se fera dans le prochain chapitre.

Chapitre 3

CONCEPTION

3.1 Introduction

La reconnaissance des expressions faciales est un problème important, qui trouve des applications dans différents domaines. Plusieurs méthodes traditionnelles ont été utilisées dans la reconnaissance d'expression telle que les SVM, Adaboost et l'apprentissage profond (et principalement les CNN) permet de supprimer ou réduire fortement la dépendance des modèles physiques. Dans ce chapitre, nous présentons notre conception en suivant les éléments suivants : Tout d'abord, on commence par présenter comment le système est censé fonctionner. Il s'agit du modèle conceptuel présentant les grandes fonctionnalités du système. Ensuite, nous passons à une description plus détaillée, nous décrivons en détail la conception du modèle proposé en donnant les détails de chaque module de la conception, ceci remplaçant les méthodes traditionnelles utilisées par un deux réseaux de neurones. Le premier fera l'extraction des caractéristiques et le deuxième pour la classification ; enfin, nous définissons par la suite les paramètres et les détails techniques relatifs à l'analyse des expressions ainsi que l'architecture utilisée. La figure 3.1 montre de base de notre système automatique de reconnaissance de l'émotion.

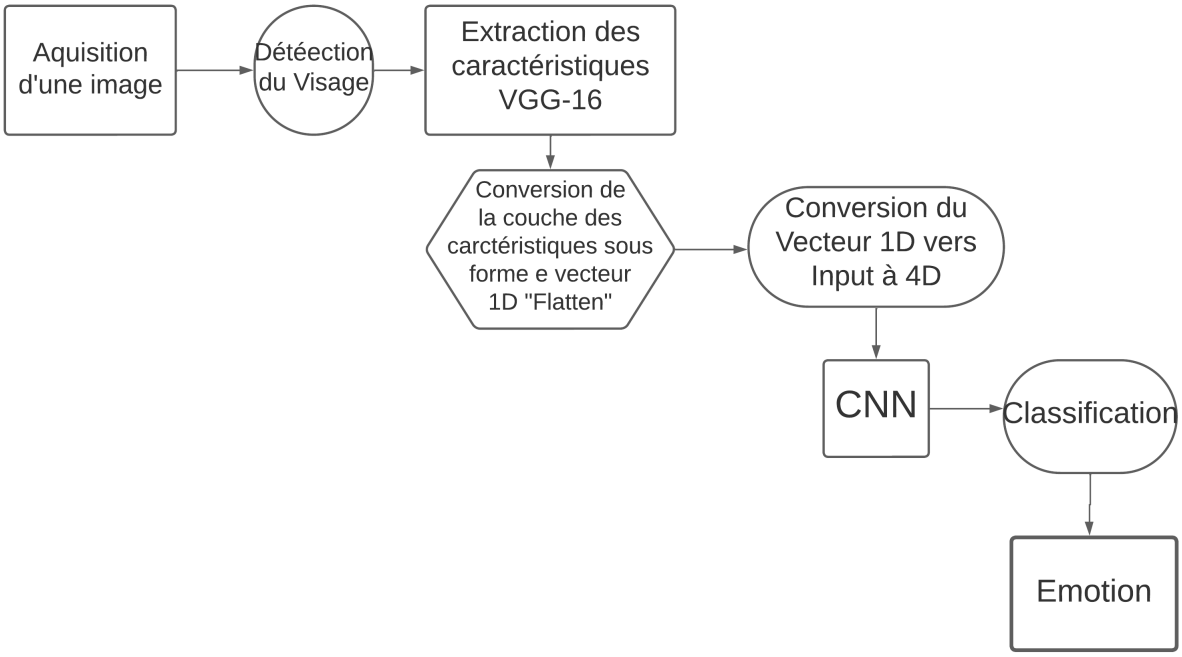

FIGURE 3.1 – Schéma récapitulatif de notre système

3.2 Détection du visage

La méthode que nous allons utiliser pour la détection de visage est la méthode de Viola et Jones.

La méthode de Viola et Jones : les filtres de Haar

1. **Explication** Elle a la particularité d'utiliser des caractéristiques très simples, mais très nombreuses. Une première innovation de la méthode est l'introduction des images intégrales, qui permettent le calcul rapide de ces caractéristiques. Une deuxième innovation importante est l'élection de ces caractéristiques par boosting, en interprétant les caractéristiques comme des classifiés. Enfin, la méthode propose une architecture pour combiner les classifieurs boostés en un processus en cascade, ce qui apporte un net gain en temps de détection.

2. **Caractéristiques** Une caractéristique est une représentation synthétique et informative, calculée à partir des valeurs des pixels. Les caractéristiques utilisées ici sont les caractéristiques pseudo Haar. Elles sont calculées par la différence des sommes de pixels de deux ou plusieurs zones rectangulaires adjacente. La figure 3.2 montre un exemple de caractéristiques pseudo-Haar [87]. Explication : Voici deux zones rectangulaires adjacentes, la première en blanc, la deuxième en noire. Les caractéristiques seraient calculées en soustrayant la somme des pixels noirs à la somme des pixels blancs. Les caractéristiques sont

calculées à toutes les positions et à toutes les échelles dans une fenêtre de détection de petite taille, typiquement de 24x24 pixels ou de 20x15 pixels. Un très grand nombre de caractéristiques par fenêtre est ainsi généré [87].

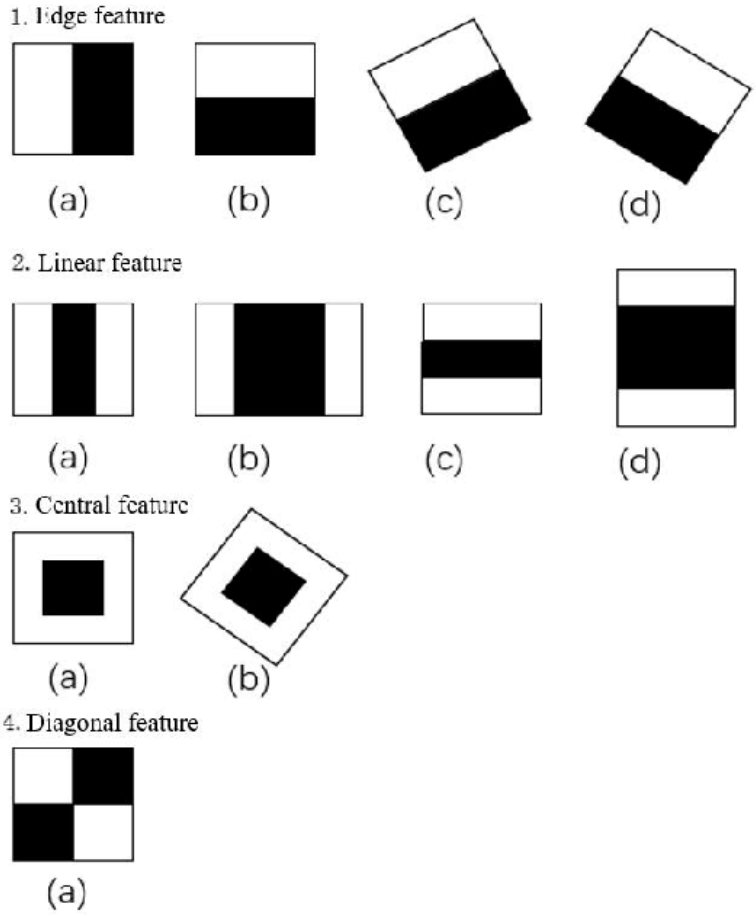

FIGURE 3.2 – Exemple de caractéristiques pseudo-Haar [57].

3. **Sélection de caractéristiques par boosting** Le boosting est un principe qui consiste à construire un classifier ≪ fort ≫ à partir d'une combinaison pondérée de classifieur ≪ faibles ≫, c'est-à-dire donnant en moyenne une meilleure réponse qu'un tirage aléatoire. La valeur seuil de la caractéristique doit être trouvée pour l'apprentissage du classifieur faible qui va permettre de mieux séparer les exemples positifs des négatifs. Dans ce cas, le classifieur se réduit alors à un couple (caractéristique, seuil) [87].

   Adaboost : La méthode d'AdaBoost (Adaptive Boosting) est un algorithme de boosting qui permet de combiner plusieurs hypothèses pour créer une autre hypothèse plus performante. Pour la détection du visage, l'algorithme AdaBoost a été proposé avec une architecture de cascade sur laquelle on peut appliquer la détection du visage Pour les caractéristiques de l'algorithme, un classifieur faible est établi par une caractéristique rectangulaire, qui est l'équivalent d'une caractéristique faible, Cet algorithme a été proposé par Viola et Jones [87], son résultat est le plus abouti connu à ce jour sur la détection de visages. Il est rapide quand il s'agit d'appliquer un nombre important de caractéristiques de rectangle à une petite région d'une fenêtre candidate, le classifieur fort est formé par quelques-unes qui sont choisies et combinées.

4. **Cascade de classifieurs** Une cascade de classifieurs est un arbre de décisions où chaque étape est entrainée pour détecter un maximum d'objets intéressants tout en excluant une certaine fraction des objets non intéressants.La figure 3.3 montre une illustration de l'architecture en cascade. Les fenêtres sont traitées séquentiellement par les classifieurs, qui prennent : Une décision d'acceptation ; la fenêtre contient l'objet et l'exemple est alors passé au classifieur suivant. Une décision de rejet ; la fenêtre ne contient pas l'objet et dans ce cas l'exemple est définitivement écarté.

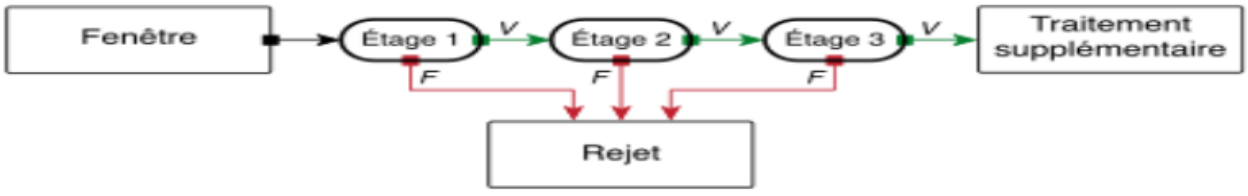

FIGURE 3.3 – Illustration de l'architecture de la cascade : les fenêtres sont traitées séquentiellement par les classifieurs, et rejetées immédiatement si la réponse est négative (F) [3]

.

3.3 Extractions des caractéristiques

Une fois le visage détecté dans l'image, le système lance le processus d'extraction des caractéristiques qui va convertir les données (pixels) à des représentations et configuration plus réduite et optimal pour que la représentation extraite soit utilisée dans le processus de la classification , cette étape réduit les dimensions de l'image en entrée en gardant les données les plus utiles

pour la classification.

L'étape d'extraction représente le cœur du système de reconnaissance, on extrait de l'image les informations qui seront sauvegardées en mémoire pour être utilisées plus tard dans la phase de décision (classification).L'extraction de caractéristiques du visage sera faite en utilisant le modèle VGG16 qui est une technique couramment utilisée en vision par ordinateur et en traitement d'images.

Le VGG16 est un réseau de neurones convolutifs pré-entraîné qui a été initialement développé pour la classification d'images. Cependant, il peut également être utilisé pour extraire des caractéristiques utiles d'une image, y compris les caractéristiques du visage. Voici comment cela fonctionne en théorie :

3.3.1 Prétraitement de l'image

L'image du visage est d'abord prétraitée pour la rendre plus facile à analyser par le CNN. Cela implique généralement la normalisation de l'image, la suppression du bruit et la conversion de l'image en un format compatible avec le CNN. Les étapes de prétraitement courantes comprennent :

3.3.1.1 Normalisation de l'image

La normalisation de l'image consiste à ajuster les valeurs des pixels de l'image afin qu'elles aient une distribution uniforme. Cela permet au CNN d'apprendre plus efficacement.

3.3.1.2 Suppression du bruit

Le bruit peut interférer avec l'extraction des caractéristiques. Les techniques de suppression du bruit courantes comprennent la suppression du bruit par moyenne et la filtre suppression du bruit médian.

1. **Filtre moyenneur** Les filtres moyenneurs, comme leur nom l'indique, calculent la moyenne, éventuellement pondérée, des pixels situés dans le voisinage de chaque pixel.

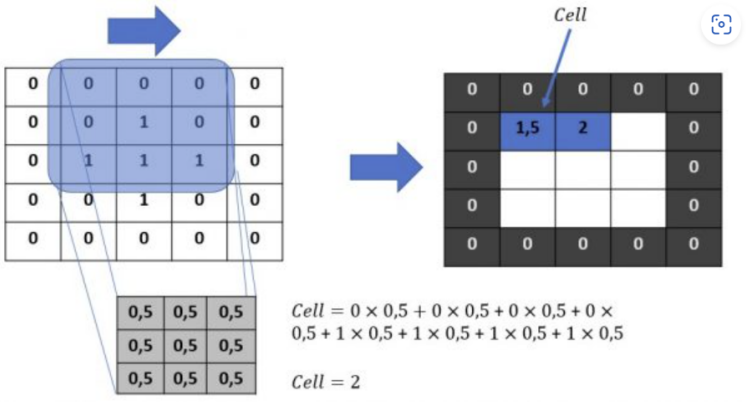

FIGURE 3.4 – filtre moyenneur [85].

FIGURE 3.5 – Exemple : filtre moyenneur non appliqué [43]

FIGURE 3.6 – Exemple : Application du filtre moyenneur [43]

2. **Filtre gaussien**

$$\frac{1}{16} \times \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

FIGURE 3.7 – filtre gaussien [45]

FIGURE 3.8 – Exemple : filtre gaussien non appliqué [43]

FIGURE 3.9 – Exemple : Application du filtre gaussien [43]

3.3.2 Extraction des caractéristiques avec VGG-16

Le CNN utilise une série de couches de convolution et de couches de pooling pour extraire les caractéristiques de l'image. Les couches de convolution appliquent un filtre à l'image pour identifier les caractéristiques locales. Les couches de pooling réduisent la taille de la représentation de l'image tout en préservant les caractéristiques les plus importantes.

3.3.2.1 Architecture de VGG-16

Le VGG16 est composé de 16 couches de convolution et de sous-échantillonnage (pooling), suivies de quelques couches entièrement connectées. Les couches de convolution sont responsables de la détection des caractéristiques visuelles de bas niveau telles que les bords, les textures, etc. Tandis que les couches entièrement connectées sont utilisées pour la classification.

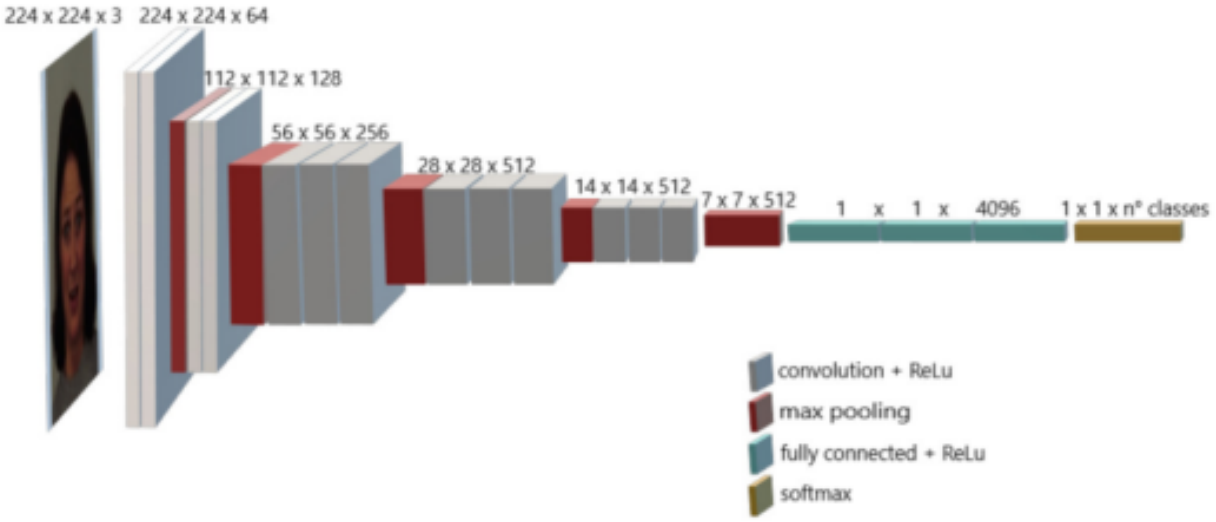

FIGURE 3.10 – Architecture de VGG-16 [51].

3.3.2.2 Extraction des caractéristiques avec VGG16

Les couches de convolution du VGG16 sont organisées en 5 blocs de convolution. Chaque bloc de convolution comprend deux ou trois couches de convolution, suivies d'une couche de pooling. Le dernier bloc de convolution est suivi d'une couche dense qui produit la représentation finale de l'image.

La représentation finale de l'image est un vecteur de caractéristiques. La longueur du vecteur de caractéristiques est de 128. Ce vecteur est utilisé pour identifier le visage dans une base de données ou pour comparer deux visages etc.

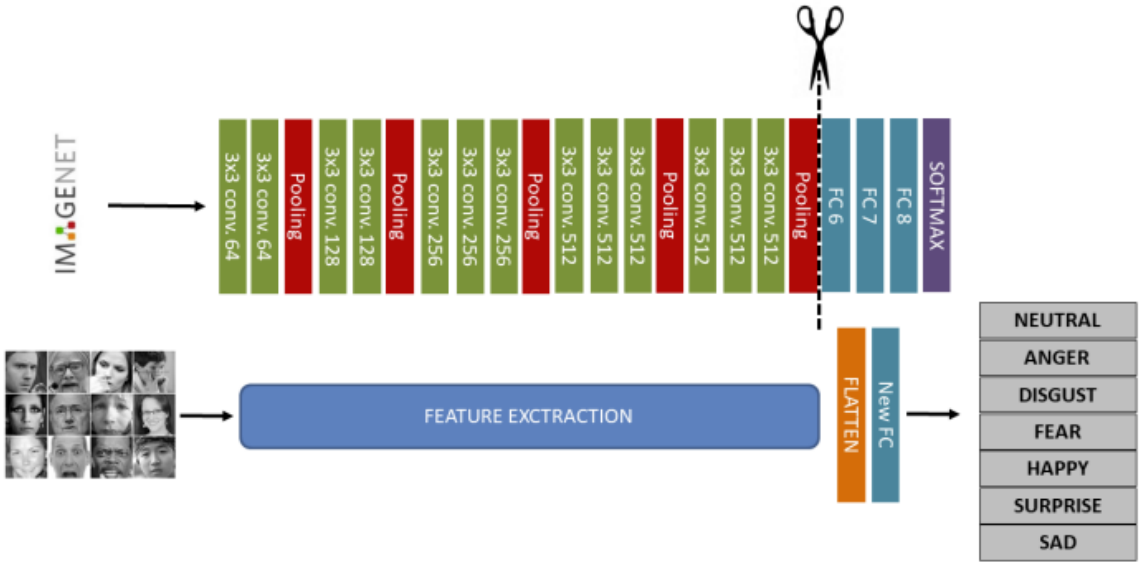

FIGURE 3.11 – Extraction des caractéristiques avec VGG16 [55].

3.4 Classification

Une fois que les caractéristiques des images ont été extraites avec VGG-16 [Figure 4.10], elles peuvent être utilisées pour prédire l'émotion du visage. La couche fully-connected du CNN est utilisée pour prédire la classe de l'image. Pour cela, notre choix, c'est porté vers ResNet-50 [42] car il est consideré comme meilleur que la précision au niveau humain.

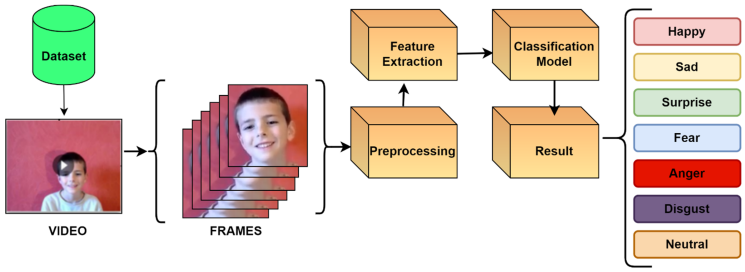

FIGURE 3.12 – Classification de l'émotion avec ResNet-50 [78].

3.5 Conclusion

Dans ce chapitre, nous avons détaillé la structure globale de notre modèle ainsi son fonctionnement, en passant par la détection de visage avec la méthode Viola et Jones, puis prétraitement de données et l'extraction des caractéristiques et pour finir le choix de ResNet-50 pour la classification des émotions à partir des caractéristiques extraite en utilisant VGG-16.

Chapitre 4

IMPLEMENTATION

4.1 Introduction

L'objectif de ce chapitre est de présenter les étapes de l'implémentation de l'approche proposée dans le cadre d'un système de reconnaissance des expressions faciales et les différentes étapes de réalisation. Nous nous sommes intéressés à l'utilisation de réseau neuronal convolutif profond, et ce, avec l'utilisation de VGG16 pour l'extraction des caractéristiques et de ResNet-50 pour la classification des émotions.
Nous commençons tout d'abord par la présentation des ressources, du langage et de l'environnement de développement que nous avons utilisés. Puis les étapes de la réalisation du modèle.Nous poursuivons ce chapitre par la présentation des différents résultats expérimentaux obtenus, quelques captures d'écrans de notre application et une petite discussion sont données à la fin de ce chapitre.

Réference : Dell Inspiron 13 5310
Processeur : Intel Evo I5-11300H
Ecran : 13.3" FHD+
RAM : 8 GO
ROM : 256GO

4.1.1 Environnement de développement

Les outils d'apprentissage profond permettent aux sciences des données (Datas Scientists) de créer des programmes capables d'amener un ordinateur ou une machine à apprendre comme le cerveau humain et à traiter des données et des modèles avant d'exécuter des décisions.
La présentation suivante détaille certains d'outils les plus couramment utilisés et les plus importants pour le développement de notre approche basée sur le CNN.

1. **Python**
   Python est un langage de programmation interprété, orienté objet, de haut niveau et à sémantique dynamique. La syntaxe de Python est simple et facile à apprendre, privilégie la lisibilité et réduit donc le coût de la maintenance des programmes. Python prend en charge les modules et les packages, ce qui encourage la modularité des programmes et la

réutilisation du code Python est un langage de programmation de haut niveau, polyvalent et très populaire. [1]

FIGURE 4.1 – Python-logo [1]

2. **OpenCv**
   OpenCV (Open Source Computer Vision Library) est une bibliothèque logicielle open source de vision par ordinateur et d'apprentissage automatique. OpenCV a été construit pour fournir une infrastructure commune pour les applications de vision par ordinateur et pour accélérer l'utilisation de la perception artificielle dans les produits commerciaux. Étant un produit sous licence BSD, OpenCV permet aux entreprises d'utiliser et de modifier facilement le code. La bibliothèque compte plus de 2500 algorithmes optimisés, ce qui inclut un ensemble complet d'algorithmes de vision par ordinateur et d'apprentissage automatique classiques et de pointe. Ces algorithmes sont spécialisés dans le traitement d'images en temps réel, d'où ils peuvent être utilisés pour détecter et reconnaître des visages, identifier des objets, classer des actions humaines dans des vidéos, suivre les mouvements de la caméra, suivre des objets en mouvement, extraire des modèles 3D d'objets, etc. [50]

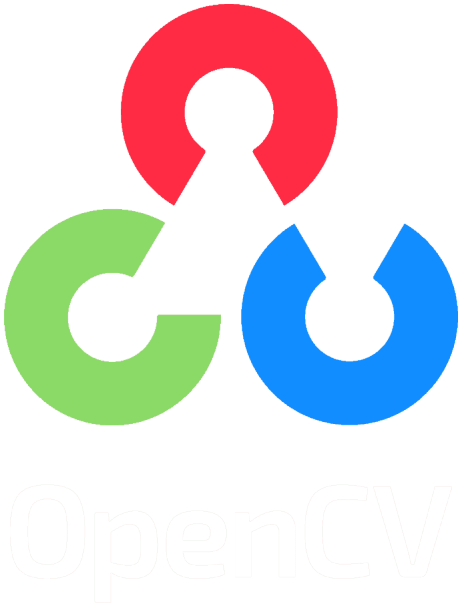

FIGURE 4.2 – OpenCV-logo [50]

3. **NumPy**
   NumPy est le package de base pour le calcul scientifique en Python. Il s'agit d'une bibliothèque Python qui fournit un objet tableau multidimensionnel, divers objets dérivés

(tels que des tableaux et des tableaux cachés) et une variété de procédures pour des opérations rapides sur les tableaux, y compris la forme, la logique, le contrôle, l'organisation, les changements, I/O, transformées de Fourier discrètes, algèbre variable linéaire de base, opérations statistiques de base, simulation stochastique et bien plus encore. [70]

FIGURE 4.3 – Numpy-logo [70]

4. **Matplotlib**
   Matplotlib est une bibliothèque complète permettant de créer des visualisations statiques, animées et interactives en Python. Matplotlib rend les choses faciles faciles et les choses difficiles possibles. Elle permet de créer des graphiques de qualité professionnelle, ainsi que des figures interactives pouvant être zoomées, panoramiques et actualisées tout en personnalisant le style visuel et la mise en page. Elle permet aussi de s'exporter vers de nombreux formats de fichiers. [75]
5. **TensorFlow**
   TensorFlow est une plateforme open-source pour la création d'applications d'apprentissage automatique. Il s'agit d'une bibliothèque de mathématiques symboliques qui utilise le flux de données et la programmation différentiable pour effectuer diverses tâches axées sur la formation et l'inférence de réseaux neuronaux profonds. Elle permet aux développeurs de créer des applications d'apprentissage automatique en utilisant divers outils, bibliothèques et ressources communautaires. [69]

FIGURE 4.4 – tensorflow-logo [69]

6. **Google Colaboratory**
   Google Colab est un produit de Google, comme son nom l'indique. Il s'agit essentiellement d'un environnement de bloc-notes gratuit qui fonctionne entièrement dans le nuage informatique. Il dispose de fonctionnalités qui aident à modifier des documents de la même manière que travaille Google Docs. Colab prend en charge de nombreuses bibliothèques d'apprentissage automatique populaires et de haut niveau qui peuvent être facilement chargées dans votre notebook. Google Colab nous offre trois types de runtime pour nos

ordinateurs portables : CPUs, GPUs, et TPUs, Colab nous offre un total de 12 heures d'exécution continue. Après cela, toute la machine virtuelle est effacée et nous devons repartir de zéro à cause de limite de l'utilisation des ressources google Colab. Nous pouvons exécuter plusieurs instances CPU, GPU et TPU simultanément sur Google collab, mais les ressources sont partagées entre ces instances. Pour notre apprentissage, nous avons utilisé les ressources de Colab avec RAM 13 GO et 110 GO de VRAM. [39]

FIGURE 4.5 – Google colab-logo [39]

4.2 Base de Données

1. **CK+**
   C'est la base de données la plus largement utilisée pour évaluer les systèmes FER. Elle contient 593 séquences vidéo de 123 sujets. Les séquences varient en durée de 10 à 60 images et montrent un passage d39 ;une expression faciale neutre à l'expression maximale. Les séquences varient en durée de 10 à 60 images et montrent un passage d'une expression faciale neutre à l'expression maximale. Parmi ces vidéos, 327 séquences de 118 sujets sont étiquetées avec sept étiquettes d'expression de base (colère, mépris, dégoût, peur, bonheur, tristesse et surprise) basées sur le Facial Action Coding System (FACS) [26].

FIGURE 4.6 – BDD CK+ [43]

4.3 Implémentation de notre modèle

4.3.1 Prétraitement

Le paramètre de rescale dans chacun de ces objets est défini sûr $\frac{1}{255}$. Cela signifie que les images seront redimensionnées pour avoir une plage de valeurs de pixels de [0, 1]. Il s'agit d'une pratique courante dans le traitement d'images, car elle rend les images plus compatibles avec les algorithmes d'apprentissage automatique.

En plus du paramètre rescale, chacun des objets ImageDataGenerator peut être configuré avec divers autres paramètres.

Ces paramètres peuvent être utilisés pour augmenter artificiellement les données d'entraînement, ce qui peut contribuer à améliorer les performances du modèle d'apprentissage automatique.

L'objet train_datagen est configuré pour appliquer les augmentations suivantes aux images de formation :

Zoom : les images peuvent être zoomées ou dézoomées de manière aléatoire jusqu'à 15 %.
Décalages horizontaux et verticaux : les images peuvent être décalées de manière aléatoire jusqu'à 20 % horizontalement et verticalement.
Cisaillement : les images peuvent être cisaillées de manière aléatoire jusqu'à 15 %.

Les objets test_datagen et validation_datagen ne sont pas configurés pour appliquer des augmentations aux images.

La méthode flow_from_directory de chaque générateur de données est utilisée pour charger les images du répertoire spécifié et créer un lot d'images sur lesquelles le modèle doit s'entraîner. Le paramètre target_size spécifie la taille des images après leur prétraitement. Le paramètre batch_size spécifie le nombre d'images dans chaque lot.
Le paramètre shuffle spécifie si les images doivent être mélangées avant chaque époque. Le paramètre class_mode spécifie le type d'étiquettes utilisées. Dans ce cas, les étiquettes sont catégorielles, ce qui signifie que chaque image peut appartenir à l'une d'un ensemble de catégories.

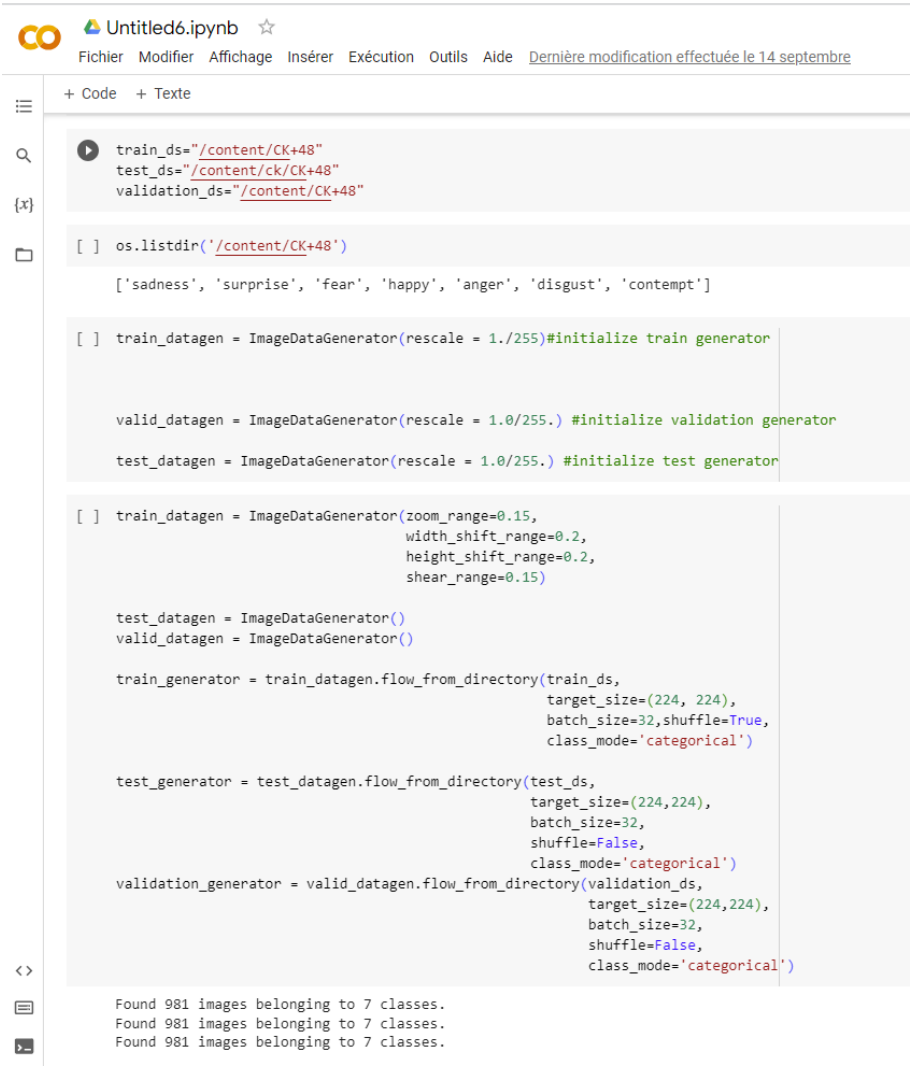

FIGURE 4.7 – Prétraitement de donnée

4.3.2 Importation du VGG-16

Chargement de la classe VGG16 du module keras.applications charge le modèle VGG16, qui est un réseau neuronal convolutif pré-entraîné sur l'ensemble de données ImageNet.

FIGURE 4.8 – transfert learning de VGG-16

4.3.3 Trie de données

Nous devons tout d'abord trier notre jeu de donnée afin de créer un tableau NumPy contenant toutes les images prétraitées dans images-array et un tableau NumPy contenant les étiquettes correspondantes dans labels-array. Ces données peuvent ensuite être utilisées pour former notre modèle d'apprentissage automatique pour la classification des émotions.

```python
import os
import cv2  # Utilisez OpenCV pour charger les images
import numpy as np

# Chemin vers le répertoire racine contenant les dossiers d'émotions
data_root = "/content/CK+48"

# Liste des émotions correspondant aux noms des sous-répertoires
emotions = os.listdir(data_root)

# Initialisez des listes pour stocker les images et les étiquettes
all_images = []
all_labels = []

# Parcourez chaque dossier d'émotion
for emotion_label, emotion_folder in enumerate(emotions):
    emotion_path = os.path.join(data_root, emotion_folder)

    # Parcourez chaque image dans le dossier
    for image_filename in os.listdir(emotion_path):
        image_path = os.path.join(emotion_path, image_filename)

        # Chargez l'image avec OpenCV
        image = cv2.imread(image_path)

        # Prétraitez l'image (redimensionnez et normalisez)
        image = cv2.resize(image, (224, 224))
        image = image / 255.0  # Normalisation

        # Ajoutez l'image et l'étiquette à la liste
        all_images.append(image)
        all_labels.append(emotion_label)

# Convertissez les listes en tableaux NumPy
images_array = np.array(all_images)
labels_array = np.array(all_labels)
```

FIGURE 4.9 – Trie de notre jeu de donnée

4.3.4 Flattenisation du tableau des caractéristiques extraites

La "flattenisation" des caractéristiques, également appelée "aplatissement" en français, est une étape courante dans le traitement d'images ou de données pour les introduire dans une couche entièrement connectée d'un réseau de neurones ou d'un modèle d'apprentissage automatique. L'objectif est de convertir des caractéristiques multidimensionnelles en un vecteur unidimensionnel (aplatissement) pour qu'elles puissent être utilisées comme entrée dans une couche dense.

La méthode reshape() de NumPy est utilisée pour remodeler le tableau de caractéristiques en un vecteur unidimensionnel. L'utilisation de -1 dans la deuxième dimension permet à NumPy de calculer automatiquement la taille de cette dimension en fonction des autres dimensions, ce qui est utile lorsque vous avez des caractéristiques de forme variable.

Après avoir effectué cette opération d'aplatissement, on peut utiliser le vecteur résultant comme entrée pour une couche dense (entièrement connectée) de votre réseau de neurones ou pour toute autre opération de traitement que vous souhaitez effectuer sur ces caractéristiques aplaties.

```
[ ] from keras.applications import VGG16
    import numpy as np

    # Entraînement de l'extracteur de caractéristiques VGG16
    vgg16 = VGG16(weights="imagenet", include_top=False)
    features = vgg16.predict(images_array)

    Downloading data from https://storage.googleapis.com/tensorflow/ke
    58889256/58889256 [==============================] - 0s 0us/step
    31/31 [==============================] - 617s 20s/step

[ ] print(features.shape)

    (981, 7, 7, 512)

[ ] flattened_features = features.reshape(features.shape[0], -1)
    print(flattened_features.shape)

    (981, 25088)
```

FIGURE 4.10 – flattenisation des caractéristiques

4.3.5 Chargement de ResNet50

Chargement de la classe ResNet50 du module keras.applications charge le modèle ResNet50, qui est un réseau neuronal convolutif pré-entraîné sur l'ensemble de données ImageNet.

```
[ ] from keras.applications import ResNet50

    img_width=224
    img_height=224
    # Chargement du modèle ResNet50 pré-entraîné
    resnet = ResNet50(weights='imagenet', include_top=False, input_shape=(img_width, img_height,3))
    resnet.summary()
```

FIGURE 4.11 – Chargement de ResNet50

4.3.6 Développement du modèle ResNet50 pour le préparer à l'entrainement

Après que notre modèle ResNet50 pré-entrainé soit prêt, nous devons convertir la couche de sortie du VGG-16 contenant les caractéristiques de fait qu'elle soit compatible à être une entrée pour notre CNN.

```
[ ] from keras.applications import ResNet50
    from keras.layers import Flatten, Dense
    from keras.models import Model

    # Charger le modèle ResNet-50 pré-entraîné sans les couches fully connected (top)
    resnet_model = ResNet50(weights='imagenet', include_top=False,input_shape=(224,224,3))

    # Créer une nouvelle couche fully connected (top) pour la classification
    x = Flatten()(resnet_model.output)
    x = Dense(128, activation='relu')(x)
    predictions = Dense(7, activation='softmax')(x)

    # Créer le modèle complet en utilisant la sortie des couches fully connected
    model = Model(inputs=resnet_model.input, outputs=predictions)

    # Compilez le modèle
    model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])

    # Ensuite,entraîner le modèle en utilisant 'flattened_features' comme entrée
    # (labels_array) comme étiquettes de formation.
```

FIGURE 4.12 – Préparation de ResNet50 pour l'entrainement

Le code de la figure 5.13 entraînera le modèle personnalisé pendant 10 époques. La méthode fit() de la classe Model prend trois arguments : le générateur de données d'entraînement, le générateur de données de validation et le nombre d'époques.

Le générateur de données de formation sera utilisé pour entraîner le modèle, et le générateur de données de validation sera utilisé pour évaluer les performances du modèle sur des données invisibles. Le nombre d'époques spécifie le nombre de fois où le modèle sera entraîné sur les données d'entraînement.

Le résultat de la méthode fit() sera un objet History, qui contient des informations sur la perte et la précision de la formation et de la validation pour chaque époque.

textLe code de la figure créera un Pandas DataFrame de la perte et de la précision de la formation et de la validation pour chaque époque.

L'objet history.history : contient un dictionnaire de métriques, et la variable de résultats sera un DataFrame avec les colonnes suivantes :

époque : le numéro de l'époque.

perte : la perte d'entraînement pour l'époque.

val_loss : la perte de validation pour l'époque.

précision : la précision de l'entraînement pour l'époque.

val_accuracy : la précision de validation pour l'époque.

```
history =model.fit(train_generator, validation_data=validation_generator, epochs=10)

Epoch 1/10
31/31 [==============================] - 1675s 54s/step - loss: 2.3784 - accuracy: 0.2538 - val_loss: 1.6086 - val_accuracy: 0.3282
Epoch 2/10
31/31 [==============================] - 1662s 54s/step - loss: 1.7157 - accuracy: 0.3496 - val_loss: 1.4300 - val_accuracy: 0.4985
Epoch 3/10
31/31 [==============================] - 1665s 55s/step - loss: 1.5083 - accuracy: 0.4261 - val_loss: 1.2087 - val_accuracy: 0.5576
Epoch 4/10
31/31 [==============================] - 1659s 54s/step - loss: 1.3567 - accuracy: 0.4913 - val_loss: 0.9235 - val_accuracy: 0.6901
Epoch 5/10
31/31 [==============================] - 1660s 54s/step - loss: 1.0179 - accuracy: 0.6239 - val_loss: 0.5739 - val_accuracy: 0.8359
Epoch 6/10
31/31 [==============================] - 1672s 54s/step - loss: 0.8344 - accuracy: 0.6911 - val_loss: 0.4409 - val_accuracy: 0.8379
Epoch 7/10
31/31 [==============================] - 1678s 55s/step - loss: 0.6257 - accuracy: 0.7696 - val_loss: 0.4723 - val_accuracy: 0.8063
Epoch 8/10
31/31 [==============================] - 1663s 54s/step - loss: 0.6310 - accuracy: 0.7554 - val_loss: 0.4962 - val_accuracy: 0.8186
Epoch 9/10
31/31 [==============================] - 1670s 54s/step - loss: 0.5344 - accuracy: 0.8135 - val_loss: 0.2904 - val_accuracy: 0.9032
Epoch 10/10
31/31 [==============================] - 1658s 54s/step - loss: 0.4003 - accuracy: 0.8491 - val_loss: 0.2355 - val_accuracy: 0.9215
```

FIGURE 4.13 – Entrainement du modèle

4.4 Evaluation de notre modèle

```
[ ]  y_test_1 = test_generator.classes
     y_pred_1 =model.predict(test_generator)
     y_pred_1 = np.argmax(y_pred_1,axis=1)

     31/31 [==============================] - 371s 12s/step

[ ]  results =model.evaluate(test_generator)

     31/31 [==============================] - 374s 12s/step - loss: 0.2355 - accuracy: 0.9215

[ ]  from sklearn.metrics import classification_report

     print(classification_report(y_test_1, y_pred_1))
```

	precision	recall	f1-score	support
0	0.72	0.99	0.83	135
1	1.00	0.72	0.84	54
2	1.00	0.95	0.98	177
3	0.77	0.99	0.87	75
4	1.00	0.92	0.96	207
5	0.96	0.63	0.76	84
6	0.99	0.99	0.99	249
accuracy			0.92	981
macro avg	0.92	0.88	0.89	981
weighted avg	0.94	0.92	0.92	981

FIGURE 4.14 – Evaluation des résultats de notre modèle

test_generator est un générateur de données pour alimenter notre modèle de manière itérative avec des données de test, telles que y_test_1 = test_generator.classes
.classes : est un attribut de ce générateur qui permet d'extraire les étiquettes de classe réelles associées aux échantillons de test.
Ainsi, y_test_1 sera un tableau contenant les étiquettes de classe réelles pour les échantillons de test.

y_pred_1 = model.predict(test_generator) : est utilisé par le modèle (model) pour faire des prédictions sur les données de test (test_generator).
model.predict() : prend les échantillons de test en entrée et retourne les prédictions du modèle sous forme de probabilités pour chaque classe. En d'autres termes, pour chaque échantillon de test, on obtient un tableau de probabilités, où chaque valeur représente la probabilité que l'échantillon appartienne à une classe particulière.

y_pred_1 = np.argmax(y_pred_1, axis=1) : Après avoir obtenu les probabilités pour chaque classe à partir de l'étape précédente (y_pred_1), on utilise np.argmax() pour extraire l'indice de la classe avec la probabilité la plus élevée pour chaque échantillon.
axis=1 : signifie que la recherche de l'indice maximum est effectuée le long de l'axe des colonnes, c'est-à-dire pour chaque échantillon individuel, ce qui nous donne l'indice de la classe prédite.

En résumé, ces trois étapes vous permettent d'obtenir deux tableaux :

- y_test_1 : Les étiquettes de classe réelles pour les échantillons de test. - y_pred_1 : Les classes prédites par notre modèle pour ces mêmes échantillons, obtenues en sélectionnant la classe avec la probabilité la plus élevée parmi les probabilités prédites. Ces prédictions peuvent

ensuite être utilisées pour évaluer la performance de votre modèle. Précision (Precision) : Il s'agit de la capacité du modèle à prédire correctement les exemples positifs parmi toutes les prédictions positives. Une précision élevée signifie que le modèle fait peu de fausses prédictions positives.

Rappel (Recall) : Le rappel mesure la capacité du modèle à identifier correctement tous les exemples positifs. Un rappel élevé signifie que le modèle parvient à capturer la plupart des exemples positifs.

F1-score : Le F1-score est une mesure qui combine à la fois la précision et le rappel en une seule métrique. Il est particulièrement utile lorsque les classes sont déséquilibrées.

Support : Le support est le nombre d'exemples réels de chaque classe dans l'ensemble de test.

Le rapport de classification nous donne ces métriques pour chaque classe présente dans nos données. Il fournit également une moyenne pondérée de ces métriques pour évaluer la performance globale du modèle.

4.5 Comparaison avec l'état de l'art

Nous avons remarqué en comparant les résultats de notre système ainsi que ceux testés sur la même base de données CK+, que notre apprentissage profond permet d'avoir des performances asser convaincante en combinant pour cela deux réseaux de neurones différents, et ces performances sont à qui vaut celles qui utilisent le machine learning tel que SVM, etc. Ce qui nous permet de déduire que les CNN permettent de travailler et d'avoir de bons résultats, et ce, malgré un nombre important de données, comme nous le montre le tableau [4.2].

Auteurs	**Techniques**	**Expressions faciales**	**Bases de données**	**Types**	**Sujets**	**Performance**
Chao Qi et all [77]	LPB, SVM	Six	CK+	Images	50	97%
Deepak Ghimire et all [36]	EBGM, ELM	Six	CK+	Images	40	96%
Michael Lyons et all [58]	PCA, LDA	Six	CK+	193 images	9	92% 75%
Uroš Mlakar et all [66]	HOG, SVM	Six	CK+	Images	Pas précisé	90%
Hui Ding et all [32]	CNN	Six	CK+	Images	123 sujets	91%
Notre Modèle	VGG16, ResNet50	Sept	CK+	Images	981 sujets	93%

TABLE 4.2 – Tableau comparaison avec l'état de l'art en terme l'accuracy

4.6 Conclusion

Dans ce dernier chapitre, nous avons présenté un modèle pour la classification des expressions faciales en se basant sur les techniques développées dans le domaine du Deep Learning et en proposant notre architecture composé de deux réseaux de neurones convolutif(CNN).
Le premier qui est VGG-16 permettra d'extraire les caractéristiques faciales ; Le second qui est ResNet50 fera la classification des émotions et ceci en utilisant la couche flatten de VGG-16. Nous avons montré que notre modèle est globalement efficace, mais qui peut s'améliorer encore plus, notamment en augmentant le nombre d'époques et en l'entraînant sur une base de données plus large.

Conclusion générale et perspectives

L'approche globale de l'analyse automatique des expressions faciales comprend généralement trois étapes. Étant donné une image d'entrée ou une séquence d'images, la première étape consiste à localiser le visage, détecter un ensemble de points faciaux ou région du visage. Une fois le visage détecté, l'étape suivante concerne l'extraction des caractéristiques du visage. L'étape finale prend comme entrée le vecteur de caractéristiques extrait précédemment pour effectuer la tâche de classification en utilisant une technique d'apprentissage automatique.

À travers ce mémoire, nous avons fait une généralité sur la reconnaissance des expressions faciales et l'utilisation deux réseaux de neurones convolutifs qui exploite les convolutions séparables en profondeur. En parcourant les différents chapitres, nous avons décrit et clarifié la définition de la reconnaissance des expressions faciales, l'architecture du système, les objectifs atteints, l'exposition des résultats obtenus et leur discussion lors de différents tests réalisés.

Notre système utilise un classifieur d'expressions faciales créé à l'aide de Deep learning avec une architecture CNN "VGG-16" pour l'extraction des caractéristiques, puis utiliser ses caractéristiques convertie en un vecteur Flatten à une dimension par une deuxième architecture CNN "ResNet-50", cette dernière fera de la classification des émotions extraite en sortie l'émotion attendue ou on peut dire détecté.

Ce travail nous a permis d'acquérir de nouvelles connaissances et d'approfondir d'autres, que ce soit dans l'aspect théorique : les expressions faciales, techniques et algorithmes utilisés, Deep Learning ou l'aspect pratique : Python, les diverses librairies dédiées au traitement d'images et la vision artificielle, Google Colab...

Parmi les perspectives ouvertes de ce projet :

- Inclure plus de classes, donc étendre le CK+ pour reconnaitre les micro-expressions, ce qui permettra plus de précision pour indiquer l'état émotionnel.
- Validation des résultats sur d'autres bases de données visage.
- L'utilisation d'autres méthodes de reconnaissance et de détection, et la combinaison avec d'autres pour la conception d'un système hybride plus performant.
- Fiabiliser le système en diminuant la sensibilité aux contraintes d'éclairage, de pose et d'occultation par l'utilisation de nouvelles techniques de normalisation.
- Accroitre le taux de précision en entrainant le modèle sur des bases, des données plus volumineuses et en augmentant le nombre d'époques.
- Étendre notre système pour l'acquisition des images 3D et inclure plus de classes d'expressions.

Bibliographie

[1] https://www.python.org/about/

[2] *ARTNATOMY*. https://www.sundayseeds.com/habits-tension-aging. Version : 2005

[3] *Identification des personnes par reconnaissance de visage pour la sécurité d'une institution bancaire.* https://www.memoireonline.com/01/14/8585/m_Identification-des-personnes-par-reconnaissance-de-visage-pour-la-securite-d-une-in html. Version : 2010

[4] *Topologies du visage*. 2012. – Disponible sur : https://www.apprendre-a-dessiner.org/comment-dessiner-et-deformer-un-visage, Consulté le : 08/09/2023

[5] *Montage d'un Système de Reconnaissance des Expressions Faciales avec le Deep Learning*. 2020. – Disponible sur : http://archives.univ-biskra.dz/bitstream/123456789/15765/1/chettouh_hadjer.pdf, Consulté le : 08/09/2023

[6] *Architecture of AlexNet (Krizhevsky et al., 2012)*. ResearchGate. https://www.researchgate.net/figure/Architecture-of-AlexNet-Krizhevsky-et-al-2012_fig2_337486420. Version : Year. – Accessed on September 9, 2023

[7] *Architecture of the CELL processor*. ResearchGate. https://www.researchgate.net/figure/Architecture-of-the-CELL-processor_fig1_30045856. Version : Year. – Accessed on September 9, 2023

[8] *Fig. 3 : The basic activation functions of the neural networks.* ResearchGate. https://www.researchgate.net/figure/Fig-3-The-basic-activation-functions-of-the-neural-networksNeural-Networks_fig3_350567223. Version : Year. – Accessed on September 9, 2023

[9] *A Hybrid GA-PSO Method for Evolving Architecture and Short Connections of Deep Convolutional Neural Networks*. ResearchGate. https://www.researchgate.net/figure/ResNet-architecture-image-taken-from-11_fig1_331671014. Version : Year. – Accessed on September 9, 2023

[10] *The LeNet-5 Architecture : a convolutional neural network.* ResearchGate. https://www.researchgate.net/figure/The-LeNet-5-Architecture-a-convolutional-neural-network_fig4_321586653. Version : Year. – Accessed on September 9, 2023

[11] *Two Person Interaction Recognition Based on Effective Hybrid Learning.* ResearchGate. https://www.researchgate.net/figure/portrays-the-VGG16-model-for-ImageNet-40-It-has-13-convolutional-layers-and-three_fig2_331562880. Version : Year. – Accessed on September 9, 2023

[12] Ackley, David H. ; Hinton, Geoffrey E. ; Sejnowski, Terrence J. : A learning algorithm for Boltzmann machines. In : *Cognitive science* 9 (1985), Nr. 1, S. 147–169

[13] Aizenberg, Igor ; Aizenberg, Naum N. ; Vandewalle, Joos P. : *Multi-valued and universal binary neurons : Theory, learning and applications*. Springer Science & Business Media, 2000

[14] Alamsyah, Andry ; Saputra, Muhammad Apriandito A. ; Masrury, Riefvan A. : Object detection using convolutional neural network to identify popular fashion product. In : *Journal of Physics : Conference Series* Bd. 1192 IOP Publishing, 2019, S. 012040

[15] Aljaafari, Nura : *Ichthyoplankton classification tool using Generative Adversarial Networks and transfer learning*, Diss., 2018

[16] Angulu, Raphael ; Tapamo, Jules R. ; Adewumi, Aderemi O. : Age Estimation via Face Images : A Survey. In : *EURASIP Journal on Image and Video Processing* 2018 (2018), Nr. 1, S. 1–35. http://dx.doi.org/10.1186/s13640-018-0278-6. – DOI 10.1186/s13640–018–0278–6

[17] assistant, F.U.S.I.A. : *Board of Directors of F.U.S.I.A. 1998 l-r : Ivan Oljelund II, Agnes Källström,Sebastian Berggren, as released by image creator F.U.S.I.A.* 2010. – Disponible sur : https://commons.wikimedia.org/wiki/File:Face_detection_example_openCV.jpg, Consulté le : 07/09/2023

[18] B. F. Lisa, M. Stacy M. M. Aleix et D. P. S. A. Ralph R. A. Ralph : Emotional Expressions Reconsidered :Challenges to Inferring Emotion From Human Facial Movements. In : *Psychol Sci Public Interest* (2019), S. 1–68

[19] Batista, Wilson ; Navarro, Marcus ; Maia, Ana : Development of a phantom and a methodology for evaluation of depth kerma and kerma index for dental cone beam computed tomography. In : *Radiation protection dosimetry* 157 (2013), 07. http://dx.doi.org/10.1093/rpd/nct174. – DOI 10.1093/rpd/nct174

[20] Bengio, Yoshua ; Courville, Aaron ; Vincent, Pascal : Representation learning : A review and new perspectives. In : *IEEE transactions on pattern analysis and machine intelligence* 35 (2013), Nr. 8, S. 1798–1828

[21] Camras, Linda ; Plutchik, Robert ; Kellerman, Henry : Emotion : Theory, Research, and Experience. Vol. 1. Theories of Emotion, 1981, S. 370

[22] CHETTOUH, HADJER : Montage d'un Système de Reconnaissance des Expressions Faciales avec le Deep Learning. https://theses-algerie.com/1746526159641876/memoire-de-master/universite-mohamed-khider---biskra/montage-d-un-syst%C3%A8me-de-reconnaissance-des-expressions-faciales-avec-le-deep-learning

[23] CHETTOUH, HADJER : Montage d'un Système de Reconnaissance des Expressions Faciales avec le Deep Learning.

[24] Cohen, Ira ; Sebe, Nicu ; Garg, Ashutosh ; Chen, Lawrence S. ; Huang, Thomas S. : Facial expression recognition from video sequences : temporal and static modeling. In : *Computer Vision and image understanding* 91 (2003), Nr. 1-2, S. 160–187

[25] Collobert, Ronan ; Weston, Jason : A unified architecture for natural language processing : Deep neural networks with multitask learning. In : *Proceedings of the 25th international conference on Machine learning*, 2008, S. 160–167

[26] Culjak, Ivan ; Abram, David ; Pribanic, Tomislav ; Dzapo, Hrvoje ; Cifrek, Mario : A brief introduction to OpenCV. In : *2012 proceedings of the 35th international convention MIPRO* IEEE, 2012, S. 1725–1730

[27] dalita dabasish : *Basics of CNN in Deep Learning.* 2022. – Disponible sur : https://www.analyticsvidhya.com/blog/2022/03/basics-of-cnn-in-deep-learning/, Consulté le : 08/09/2023

[28] Davoine, Franck ; Abboud, Bouchra : Van Mô Dang,«. In : *Analyse de visages et d'expressions faciales par modèle actif d'apparence», Université de Technologie de Compiègne, France* (2004)

[29] Denton, Emily L. ; Chintala, Soumith ; Fergus, Rob u. a. : Deep generative image models using a[U+FFFC] laplacian pyramid of adversarial networks. In : *Advances in neural information processing systems* 28 (2015)

[30] Desir, Chesner : *Classification automatique d'images, application à l'imagerie du poumon profond*, Université de Rouen, Diss., 2013

[31] Dian, Diallo Nene A. : La Reconnaissance Des Expressions Faciales, 2019

[32] Ding, Hui ; Zhou, Shaohua K. ; Chellappa, Rama : Facenet2expnet : Regularizing a deep face recognition net for expression recognition. In : *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* IEEE, 2017, S. 118–126

[33] Enab, Khaled ; Ertekin, Turgay : Artificial neural network based design for dual lateral well applications. In : *Journal of Petroleum Science and Engineering* 123 (2014), S. 84–95

[34] Fernández, Dennis N. : Multi-subject Continuous Emotional States Monitoring by Using Convolutional Neural Networks. In : *2019 International Conference on Control of Dynamical and Aerospace Systems (XPOTRON)* IEEE, 2019, S. 1–4

[35] GHANEM, KHADOUDJA : *Reconnaissance des Expressions Faciales à Base d'Informations Vidéo ; Estimation de l'Intensité des ExpressionsFaciales.* 2010. – Disponible sur : https://bu.umc.edu.dz/theses/informatique/GHA5753.pdf., Consulté le : 07/09/2023

[36] Ghimire, Deepak ; Lee, Joonwhoan ; Li, Ze-Nian ; Jeong, Sunghwan : Recognition of facial expressions based on salient geometric features and support vector machines. In : *Multimedia Tools and Applications* 76 (2017), S. 7921–7946

[37] Goodfellow, Ian ; Bengio, Yoshua ; Courville, Aaron : *Deep learning.* MIT press, 2016

[38] Goodfellow, Ian ; Pouget-Abadie, Jean ; Mirza, Mehdi ; Xu, Bing ; Warde-Farley, David ; Ozair, Sherjil ; Courville, Aaron ; Bengio, Yoshua : Generative adversarial nets. In : *Advances in neural information processing systems* 27 (2014)

[39] https://research.google.com/colaboratory/faq.html

[40] HADJER, CHETTOUH : *Montage d'un Système de Reconnaissance des Expressions Faciales avec le Deep Learning.* 2020. – Disponible sur : http://archives.univ-biskra.dz/bitstream/123456789/15765/1/chettouh_hadjer.pdf, Consulté le : 07/09/2023

[41] Hakim, Belhadjer ; Brahim, Sarouer : *Classification des images avec les réseaux de neurones convolutionnels*, Université Mouloud Mammeri, Diss., 2018

[42] He, Kaiming ; Zhang, Xiangyu ; Ren, Shaoqing ; Sun, Jian : Deep residual learning for image recognition. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, S. 770–778

[43] Hsu, Lun-Kai ; Tseng, Wen-Sheng ; Kang, Li-Wei ; Wang, Yu-Chiang F. : Seeing through the expression : Bridging the gap between expression and emotion recognition. In : *2013 IEEE International Conference on Multimedia and Expo (ICME)* IEEE, 2013, S. 1–6

[44] Hubara, Itay ; Courbariaux, Matthieu ; Soudry, Daniel ; El-Yaniv, Ran ; Bengio, Yoshua : Quantized neural networks : Training neural networks with low precision weights and activations. In : *The Journal of Machine Learning Research* 18 (2017), Nr. 1, S. 6869–6898

[45] https://openclassrooms.com/fr/courses/5060661-initiez-vous-aux-traitements-de-base-d
5217251-analysez-le-filtrage-spatial-et-la-convolution-par-masque

[46] Khattak, Asad ; Asghar, Muhammad U. ; Batool, Ulfat ; Asghar, Muhammad Z. ; Ullah, Hayat ; Al-Rakhami, Mabrook ; Gumaei, Abdu : Automatic detection of citrus fruit and leaves diseases using deep neural network model. In : *IEEE access* 9 (2021), S. 112942–112954

[47] Koey, Mili : *Conseils pour peindre les ombres et les lumières.* 2019. – Disponible sur : https://www.clipstudio.net/comment-dessiner/archives/165619, Consulté le : 07/09/2023

[48] Krizhevsky, Alex ; Sutskever, Ilya ; Hinton, Geoffrey E. : Imagenet classification with deep convolutional neural networks. In : *Advances in neural information processing systems* 25 (2012)

[49] Krizhevsky, Alex ; Sutskever, Ilya ; Hinton, Geoffrey E. : ImageNet Classification with Deep Convolutional Neural Networks. 25 (2012), S. 1–9

[50] In : Kumar, Suresh ; Raju, Viswanadha ; Maheswari, Uma : *1 OpenCV libraries for computer vision*. 2023. – ISBN 9783110756722, S. 1–22

[51] Kumov, Vyacheslav ; Samorodov, Andrey : Recognition of Genetic Diseases Based on Combined Feature Extraction From 2D Face Images, 2020, 1-7

[52] Laskri, Mohamed T. ; Chefrour, Djallel : Who_Is : système d'identification des visages humains. In : *Revue Africaine de Recherche en Informatique et Mathématiques Appliquées* 1 (2002)

[53] LeCun, Yann ; Bottou, Léon ; Bengio, Yoshua ; Haffner, Patrick : Gradient-based learning applied to document recognition. In : *Proceedings of the IEEE* 86 (1998), Nr. 11, S. 2278–2324

[54] Lekdioui, Khadija : *Reconnaissance d'états émotionnels par analyse visuelle du visage et apprentissage machine*, Université Bourgogne Franche-Comté ; Université Ibn Tofail. Faculté des ..., Diss., 2018

[55] Leone, Alessandro ; Caroppo, Andrea ; Manni, Andrea ; Siciliano, P. : Vision-Based Road Rage Detection Framework in Automotive Safety Applications, 2021

[56] Li, Haoxiang ; Lin, Zhe ; Shen, Xiaohui ; Brandt, Jonathan ; Hua, Gang : A convolutional neural network cascade for face detection. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, S. 5325–5334

[57] Lu, Wen-Yao ; Ming, YANG : Face detection based on viola-jones algorithm applying composite features. In : *2019 International Conference on Robots & Intelligent System (ICRIS)* IEEE, 2019, S. 82–85

[58] Lyons, Michael ; Akamatsu, Shigeru ; Kamachi, Miyuki ; Gyoba, Jiro : Coding facial expressions with gabor wavelets. In : *Proceedings Third IEEE international conference on automatic face and gesture recognition* IEEE, 1998, S. 200–205

[59] Maalej, Ahmed ; Amor, Boulbaba B. ; Daoudi, Mohamed ; Srivastava, Anuj ; Berretti, Stefano : Shape analysis of local facial patches for 3D facial expression recognition. In : *Pattern Recognition* 44 (2011), Nr. 8, S. 1581–1589

[60] Mao, Keming ; Lu, Duo ; E, Dazhi ; Tan, Zhenhua : A case study on attribute recognition of heated metal mark image using deep convolutional neural networks. In : *Sensors* 18 (2018), Nr. 6, S. 1871

[61] Matsugu, Masakazu ; Mori, Katsuhiko ; Mitari, Yusuke ; Kaneda, Yuji : Subject independent facial expression recognition with robust face detection using a convolutional neural network. In : *Neural Networks* 16 (2003), Nr. 5-6, S. 555–559

[62] Medsker, Larry R. ; Jain, LC : Recurrent neural networks. In : *Design and Applications* 5 (2001), Nr. 64-67, S. 2

[63] Mehrabian, Albert ; Ferris, Susan : Inference of Attitudes from Non-Verbal Communication in Two Channels. In : *Journal of consulting psychology* 31 (1967), 07, S. 248–52. `http://dx.doi.org/10.1037/h0024648`. – DOI 10.1037/h0024648

[64] Migneault, Francis C. ; Granger, Éric : *Évaluation de méthodes de reconnaissance de visages pour l'identification d'individus à partir d'une image de référence*. 2016

[65] Mishra, Vidushi ; Agarwal, Smt M. ; Puri, Neha : Comprehensive and comparative analysis of neural network. In : *International Journal of Computer Application* 2 (2018), Nr. 8, S. 126–137

[66] Mlakar, Uroš ; Fister, Iztok ; Brest, Janez ; Potočnik, Božidar : Multi-objective differential evolution for feature selection in facial expression recognition systems. In : *Expert Systems with Applications* 89 (2017), S. 129–137

[67] NACER, Foued : Reconnaissance d'expression faciale à partir d'un visage réel. (2019)

[68] Navraan, Mina ; Charkari, Nasrollah M. ; Mansoorizadeh, Muharram : Automatic Facial Emotion Recognition Method Based on Eye Region Changes. In : *Journal of Information Systems and Telecommunication (JIST)* 4 (2016), Nr. 4, S. 221–231

[69] In : Nelli, Fabio : *Deep Learning with TensorFlow*. 2023. – ISBN 978–1–4842–9531–1, S. 289–321

[70] In : Nelli, Fabio : *The NumPy Library*. 2023. – ISBN 978–1–4842–9531–1, S. 45–72

[71] Ng, Andrew ; Jordan, Michael : On discriminative vs. generative classifiers : A comparison of logistic regression and naive bayes. In : *Advances in neural information processing systems* 14 (2001)

[72] Nguyen, Long D. ; Lin, Dongyun ; Lin, Zhiping ; Cao, Jiuwen : Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation. In : *2018 IEEE international symposium on circuits and systems (ISCAS)* IEEE, 2018, S. 1–5

[73] Oord, Aaron Van d. ; Dieleman, Sander ; Schrauwen, Benjamin : Deep content-based music recommendation. In : *Advances in neural information processing systems* 26 (2013)

[74] Ouamane, Abdelmalik : *Reconnaissance Biométrique par Fusion Multimodale du Visage 2D et 3D*, Université Mohamed Khider-Biskra, Diss., 2015

[75] In : Pajankar, Ashwin : *Matplotlib*. 2017. – ISBN 978–1–4842–2877–7, S. 159–165

[76] Pantic, Maja ; Rothkrantz, Leon J. : Facial action recognition for facial expression analysis from static face images. In : *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 34 (2004), Nr. 3, S. 1449–1461

[77] Qi, Chao ; Li, Min ; Wang, Qiushi ; Zhang, Huiquan ; Xing, Jinling ; Gao, Zhifan ; Zhang, Huailing : Facial expressions recognition based on cognition and mapped binary patterns. In : *IEEE Access* 6 (2018), S. 18795–18803

[78] Rathod, Manish ; Dalvi, Chirag ; Kaur, Kulveen ; Patil, Shruti ; Gite, Shilpa ; Kamat, Pooja ; Kotecha, Ketan ; Abraham, Ajith ; Gabralla, Lubna A. : Kids' Emotion Recognition Using Various Deep-Learning Models with Explainable AI. In : *Sensors* 22 (2022), Nr. 20, S. 8066

[79] Salakhutdinov, Ruslan ; Mnih, Andriy ; Hinton, Geoffrey : Restricted Boltzmann machines for collaborative filtering. In : *Proceedings of the 24th international conference on Machine learning*, 2007, S. 791–798

[80] Sermanet, Pierre ; Eigen, David ; Zhang, Xiang ; Mathieu, Michaël ; Fergus, Rob ; LeCun, Yann : Overfeat : Integrated recognition, localization and detection using convolutional networks. In : *arXiv preprint arXiv :1312.6229* (2013)

[81] Simonyan, Karen ; Zisserman, Andrew : Very deep convolutional networks for large-scale image recognition. In : *arXiv preprint arXiv :1409.1556* (2014)

[82] Srivastava, Nitish ; Salakhutdinov, Ruslan R. ; Hinton, Geoffrey E. : Modeling documents with deep boltzmann machines. In : *arXiv preprint arXiv :1309.6865* (2013)

[83] Szegedy, Christian ; Toshev, Alexander ; Erhan, Dumitru : Deep neural networks for object detection. In : *Advances in neural information processing systems* 26 (2013)

[84] Szegedy, Christian ; Vanhoucke, Vincent ; Ioffe, Sergey ; Shlens, Jon ; Wojna, Zbigniew : Rethinking the inception architecture for computer vision. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, S. 2818–2826

[85] `https://datacorner.fr/image-processing-6/`

[86] Valstar, Michel F.; Gunes, Hatice; Pantic, Maja : How to distinguish posed from spontaneous smiles using geometric features. In : *Proceedings of the 9th international conference on Multimodal interfaces*, 2007, S. 38–45

[87] Viola, Paul; Jones, Michael u. a. : Robust real-time object detection. In : *International journal of computer vision* 4 (2001), Nr. 34-47, S. 4

[88] `https://www.dell.com/support/home/fr-fr/product-support/product/inspiron-13-5310-laptop/drivers`

[89] Youcef, Moualek D. : Deep Learning pour la classification des images. In : *Université Abou Bakr Belkaid Tlemcen Année universitaire* 2017 (2016)

[90] Zhang, Bin : *Étiquetage des images et autres tâches quotidiennes dans l'équipe de projet d'intelligence artificielle*. `https://travaux.master.utc.fr/formations-master/ingenierie-de-la-sante/ids114/`. Version : 2021. – Mémoire de Stage, réf n° IDS114

Résumé

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

La reconnaissance des émotions est l'un des domaines scientifiques les plus complexes. Ces dernières années, de plus en plus d'applications tentent de l'automatiser. Ces applications innovantes concernent plusieurs domaines comme l'aide aux enfants autistes, les jeux vidéo, l'interaction homme-machine.

Cependant, le succès du Deep Learning, a poussé les chercheurs à exploiter les différents types d'architectures de cette technique, pour obtenir de meilleures performances.

Nous proposons dans ce travail un système capable de détecter et d'identifier l'utilisateur à travers ses expressions faciales afin de reconnaître son état émotionnel. Le système utilise un classifieur d'expressions faciales basé sur l'apprentissage profond (Deep learning) et qui applique un algorithme de réseaux de neurones convolutifs (VGG-16) pour l'éxtraction des caractéristiques, puis applique un deuxièeme algorithme de réseaux de neurones convolutifs (ResNet-50) pour la classification de l'émotion.

Les expériences ont été menées afin de vérifier la faisabilité du système proposé. Son objectif est la validation de la détection des visages et la reconnaissance de l'utilisateur et de ses émotions à travers ses expressions faciales avec la base de données CK+.

Mots clés : aprentissage profond ; deep learning ; Réseaun de neurones convolutif : VGG-16, ResNet-50.

ABSTRACT

Facial expression is one of the most common non-verbal means used by humans to convey internal emotional states and therefore plays a fundamental role in interpersonal interactions. Although there is a wide range of possible facial expressions, psychologists have identified six basic expressions (joy, sadness, surprise, anger, fear and disgust) that are universally recognised.

Emotion recognition is one of the most complex areas of science. In recent years, more and more applications have attempted to automate it. These innovative applications cover a wide range of fields, including helping autistic children, video games and human-computer interaction.

However, the success of Deep Learning has prompted researchers to exploit the different types of architecture for this technique, in order to achieve better performance.

In this work, we propose a system capable of detecting and identifying users through their facial expressions in order to recognise their emotional state. The system uses a facial expression classifier based on deep learning which applies a convolutional neural network algorithm (VGG-16) for feature extraction and then applies a second convolutional neural network algorithm (ResNet-50) for emotion classification.

The experiments were carried out to verify the feasibility of the proposed system. Its objective is to validate face detection and recognition of the user and his emotions through his facial expressions with the CK+ database.

Keywords : convolutional neural network : VGG-16, ResNet-50, Data Base : CK+.