République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université de Mohamed El Bachir El Ibrahimi de Bordj Bou Arreridj
Faculté des Mathématiques et d'Informatique
Département d'informatique

MEMOIRE
Présenté en vue de l'obtention du diplôme
Master en informatique
Spécialité : Ingénierie De l'Informatique Décisionnelle

THEME :
La reconnaissance des émotions par Deep Learning (Étude comparative)

Présenté par :
BOUATTA ABDELGHANI
ABACHE NACEREDDINE

Soutenu publiquement le : 09/09/2023
Devant le jury composé de :
Président : NAILI MAKHLOUF
Examinateur : NOUIOUA MOURAD
Encadreur : MAZA SOFIANE

2022/2023

Dédicace

Je dédie ce travail

A mon père Abdelouahab : Aucune dédicace ne peut exprimer le respect que j'ai toujours eu pour toi. Rien au monde ne vaut les efforts consentis jour et nuit pour mon éducation et mon bien-être. A ma mère Zahia : Tu représentes pour moi le symbole de bonté par excellence, la source de tendresse, et l'exemple de dévouement qui n'a cessé de m'encourager et de prier pour moi.

A mes frères et à mes sœurs.

A mon compagnon Nacereddine et sa famille.

A mes amis... merci beaucoup pour votre aide Merci pour toutes les valeurs de fraternité et pour tous les moments passés ensemble.

Et à toute ma famille et à tous ceux que j'aime.

J'espère que vous serez tous fiers de moi

Dédicace

Je dédie ce travail

A mon père Aucune dédicace ne peut exprimer le respect que j'ai toujours eu pour toi. Rien au monde ne vaut les efforts consentis jour et nuit pour mon éducation et mon bien-être. A ma mère Tu représentes pour moi le symbole de bonté par excellence, la source de tendresse, et l'exemple de dévouement qui n'a cessé de m'encourager et de prier pour moi.

A mes frères et à mes sœurs.

A mon compagnon Abdelghani et sa famille.

A mes amis... merci beaucoup pour votre aide Merci pour toutes les valeurs de fraternité et pour tous les moments passés ensemble.

Et à toute ma famille et à tous ceux que j'aime.

J'espère que vous serez tous fiers de moi

Remerciement

Merci, Allah, de nous avoir donné la capacité d'écrire et de réfléchir, la force d'y croire, la patience d'aller au bout du rêve et le bonheur de lever les mains vers le ciel et de dire "Alhamdoulillah". Tout d'abord, ce travail ne serait pas aussi riche sans l'aide et l'encadrement de Monsieur **MAZA Sofiane**, nous le remercions pour la qualité de son encadrement exceptionnel, pour sa patience sa disponibilité lors de notre préparation à ce mémoire.

Nos remerciements vont également à tous les amis qui nous ont aidés et soutenus de près ou de loin.

Merci à tous ...

Résumé

la reconnaissance des émotions et l'informatique affective dans de nombreux domaines de recherche au cours des dernières décennies ont suscité un grand intérêt. En particulier, les expressions faciales représentent l'un des moyens les plus efficaces pour identifier les éléments caractéristiques du comportement humain et décrire l'état émotionnel.

Dans notre implémentation, nous avons utilisé une technologie de pointe de reconnaissance d'apparence profonde pour analyser FER2013 Dataset , dans le but de compiler une compréhension complète des apparences faciales. Alors, comment pouvons-nous identifier avec précision les émotions en utilisant différentes techniques de reconnaissance des émotions? Grâce à nos expériences, nous introduisons des conceptions innovantes de réseaux cérébraux à convolution profonde. Et faites une comparaison entre différentes architectures CNN telles que RES-NET, VGG, FER-MODEL, SVM-CNN et SEQUENTIEL-MODEL pour apprendre les émotions.

Mots-clés: Deep Learning (DL), Convolutional Neural Network (CNN), Facial Emotion Recognition (FER), Artificial Neural Network (ANN).

Abstract

Emotion recognition and affective computing in many research areas over the past decades have attracted great interest. In particular, facial expressions represent one of the most effective means to identify the characteristic elements of human behavior and to describe the emotional state.

In our implementation, we used state-of-the-art deep appearance recognition technology to analyze FER2013 Dataset, with the goal of compiling a comprehensive understanding of facial appearances. So, how can we accurately identify emotions using different emotion recognition techniques? Through our experiments, we introduce innovative deep convolutional brain network designs. And make a comparison between different CNN architectures like RES-NET, VGG, FER-MODEL, SVM-CNN and SEQUENTIEL-MODEL to learn emotions.

Keywords: Deep Learning (DL), Convolutional Neural Network (CNN), Facial Emotion Recognition (FER), Artificial Neural Network (ANN)

ملخص

بات التعرف على المشاعر والحوسبة العاطفية في العديد من مجالات البحث على مدار العقود الماضية اهتمامًا كبيرًا. على وجه الخصوص ، تمثل تعابير الوجه واحدة من أكثر الوسائل فعالية لتحديد العناصر المميزة للسلوك البشري ووصف الحالة العاطفية.

في تطبيقنا ، استخدمنا أحدث تقنيات التعرف على المظهر العميق لتحليل مجموعة بيانات FER2013 ، بهدف تجميع فهم شامل لمظاهر الوجه. إذًا، كيف يمكننا تحديد المشاعر بدقة باستخدام تقنيات التعرف على المشاعر المختلفة ؟ من خلال تجاربنا ، نقدم تصميمات مبتكرة لشبكات الدماغ التلافيفية العميقة. وإجراء مقارنة بين بنيات لشبكة CNN المختلفة مثل RES-NET , VGG , FER-MODEL , SVM-CNN و SEQUENTIEL-MODEL لمعرفة العواطف.

الكلمات المفتاحية: التعلم العميق (DL)، الشبكة العصبية التلافيفية (CNN)، التعرف على مشاعر الوجه (FER)، الشبكة العصبية الاصطناعية (ANN).

Table des matières

Liste des abréviations

ASM :	Modèle de Forme Actif.
ANN :	Réseau de neurones artificiels DL Apprentissage en profondeur
CK+ :	Cohn-Kanade étendu
CNN :	Réseau de neurones convolutifs
CONV :	Convolutif
FACS :	système de codage des actions faciales
FC :	Entièrement connecté
FER :	Reconnaissance des expressions faciales
FERC :	reconnaissance des émotions faciales par réseaux de neurones convolutifs
IA :	Intelligence Artificielle
KNN :	K-plus proche voisin
K-NN :	K-plus proche voisin.
LSTM :	Mémoire longue à court terme
NN :	Réseaux de neurones
PISCINE :	Mise en commun
RELU :	Unité Linéaire Rectifiée
RNN :	Réseaux de neurones récurrents
SVM :	Machines à vecteurs de support.

Liste des figures

Liste des tableaux

INTRODUCTION GENERALE

Introduction générale

En raison de la rapidité croissante des mises à jour en science et technologie, divers aspects de notre vie quotidienne, tels que les machines, ont subi des changements rapides. Nous utilisons des machines dans tout, de l'éducation, de l'e-learning, des soins de santé et de la technologie d'assistance, offrant ainsi la possibilité à des machines intelligentes de nous propulser à accomplir une variété de tâches, même très complexes.

Les machines ont plusieurs moyens de détecter leur environnement à travers des caméraset des capteurs, et alors que le contact humain augmente, les interactions doivent devenir plus fluides et plus naturelles. La reconnaissance des émotions faciales est l'une des utilisations qui pourrait aider les machines à mieux remplir cette fonction.

Les émotions jouent un rôle important non seulement dans nos relations avec les autres, mais également dans notre façon d'utiliser les ordinateurs. L'informatique affective est un domaine qui se concentre sur les émotions des utilisateurs lorsqu'ils interagissent avec des ordinateurs et des applications. Étant donné que l'état émotionnel d'une personne affecte l'attention, la résolution de tâches et la prise de décisions, la vision de l'informatique affectiveest de permettre aux systèmes de reconnaître les émotions humaines et de les manipuler pour augmenter la productivité et l'efficacité du travail informatique. Le problème complexe de la détection automatique des émotions humaines est devenu un domaine de recherche impliquant un nombre croissant de scientifiques qui se concentrent sur divers domaines tels que l'intelligence artificielle, la vision par ordinateur et la psychologie. Sa popularité découle d'une large gamme d'applications potentielles.

La reconnaissance des expressions humaines lors des interactions est complexe. Dans le domaine de l'informatique affective, elle capture et modélise la signification des expressions faciales. Récemment, la reconnaissance des expressions faciales est devenue un sujet très actif dans la communauté de la vision par ordinateur. Dans la communauté de la vision par ordinateur, le terme "reconnaissance des expressions faciales" fait généralement référence àla classification des caractéristiques faciales en l'une des sept émotions de base ou universelles suivantes : bonheur, tristesse, peur, dégoût, surprise, colère et neutralité.

Au cours du développement de ce travail, des algorithmes d'apprentissage automatique et des techniques d'apprentissage en profondeur ont été utilisés sur des images montrant les émotions faciales suivantes : bonheur, tristesse, colère, surprise, dégoût, neutralité et peur pour extraire des caractéristiques sémantiques à partir d'images faciales.

L'objectif de ce projet final est implémenté différentes architectures CNN telles que RES-NET, VGG, FER-MODEL, SVM-CNN et SEQUENTIEL-MODEL avec étude comparative. Cette thèse est composée de trois chapitres, organisés comme suit :

Le premier chapitre plonge dans l'océan du savoir, offrant un aperçu exhaustif de notre domaine d'étude. Il débute par une exploration en profondeur du concept d'émotion, puis plonge dans l'univers des expressions faciales, pour enfin dévoiler une panoplie de systèmes de codage et de reconnaissance des émotions, tout en exposant quelques travaux novateurs dans le domaine de la détection d'émotion.

Le second chapitre s'élance dans le domaine de l'apprentissage profond, où les rouages des réseaux de neurones convolutifs sont minutieusement disséqués. Il clôture son exposé en abordant diverses initiatives liées à la reconnaissance des émotions, toutes ancrées dans le terreau de l'apprentissage profond.

Quant au dernier chapitre, il amorce son périple en détaillant la mise en œuvre de différentes variantes de réseaux de neurones convolutifs. Par la suite, il lève le voile sur les outils employés pour l'élaboration de notre application, dévoilant les résultats obtenus et orchestrant une comparaison éclairante entre ces diverses architectures de CNN dédiées à l'exploration des émotions.

Enfin, cette thèse clôture son voyage intellectuel par une conclusion générale, offrant un horizon panoramique sur les perspectives à venir.

CHAPITRE I

RECONNAISSANCE DES EXPRISSIONS FACIALES

Chapitre I : Reconnaissance des expressions faciales

I.1. Introduction

Ce chapitre présent le domaine d'application de la reconnaissance des expressions faciales, puisnous décrivons les différentes émotions de base existantes et leurs propriétés. Enfin, nous avons évoqué le processus de reconnaissance des expressions faciales.

I.2. Historique

La technologie de reconnaissance faciale a une longue histoire, dont les racines remontent auxannées 1960. Voici un bref aperçu de son histoire :

- Dans les années 1960, le psychologue Paul Ekman a commencé sa recherche pionnière sur les expressions faciales, qui allait devenir le fondement de la reconnaissance moderne des expressions faciales. Ekman a identifié six émotions de base qui s'expriment universellement à travers les expressions faciales : le bonheur, la tristesse, la colère, la peur, la surprise et le dégoût. Il a également développé le système de codage d'action faciale (FACS), un système complet pour mesurer objectivement les expressions faciales [1].
- 1960 - Woody Bledsoe, Helen Chan Wolf et Charles Bisson développent le premier système dereconnaissance faciale au Stanford Research Institute (SRI). Le système a utilisé une tablette RAND pour tracer et analyser 21 traits du visage[2].
- Années 1970 - Takeo Kanade développe un algorithme de reconnaissance des visages sur les photographies. Cet algorithme utilisait une technique appelée eigenfaces, qui consistait à analyser les principales composantes des visages [3].
- Dans les années 1980 et 1990, des informaticiens ont commencé à développer des algorithmesde vision par ordinateur pour reconnaître les expressions faciales. L'un des premiers algorithmes était la méthode de suivi des points caractéristiques (FPT), qui détectait les repères faciaux et suivait leur mouvement pour reconnaître les expressions[4].
- Années 1990 - Des chercheurs du Massachusetts Institute of Technology (MIT) développent un système de reconnaissance faciale basé sur un réseau de neurones appelé Eigenface. Ce système était capable d'identifier les visages avec une grande précision[5].
- Dans les années 2000, des algorithmes d'apprentissage automatique tels que les

machines à vecteurs de support (SVM) et les réseaux de neurones artificiels (ANN) ont été appliqués à la reconnaissance des expressions faciales, permettant une reconnaissance plus précise et plus robuste. Avec l'essor de l'apprentissage en profondeur dans les années 2010, les réseaux de neurones convolutifs (CNN) sont devenus la méthode de pointe pour la reconnaissance des expressions faciales [6].

- Années 2010 - La technologie de reconnaissance faciale se généralise, des entreprises comme Facebook et Google l'utilisant pour marquer des photos et à d'autres fins. Cependant, des préoccupations concernant la confidentialité et l'exactitude se posent également [7].
- En 2019, la Haute Cour du Royaume-Uni a statué que l'utilisation de la technologie de reconnaissance faciale automatique pour rechercher des personnes dans les foules était légale [W1].

Aujourd'hui, la reconnaissance des expressions faciales est utilisée dans diverses applications, notamment l'interaction homme-machine, l'analyse des émotions et la psychologie clinique. La technologie a évolué au point de pouvoir reconnaître non seulement les six émotions de base, mais aussi des expressions plus subtiles et même des micro-expressions qui ne durent qu'une fraction de seconde.

I.3. Définitions

I.3.1. Reconnaissance des émotions faciales

La reconnaissance des émotions faciales (FER) est une technologie qui permet aux machines d'identifier, d'interpréter et de répondre aux émotions humaines en analysant les expressions faciales. Cela implique l'utilisation d'algorithmes informatiques et de techniques de traitement d'images pour détecter et reconnaître les émotions affichées sur les visages humains[8].

Les systèmes FER fonctionnent généralement en capturant une image ou une vidéo du visage d'une personne, puis en utilisant des algorithmes d'apprentissage automatique pour analyser les traits du visage et déterminer l'état émotionnel de la personne. Les algorithmes sont formés sur de grands ensembles de données d'expressions faciales pour reconnaître les modèles et les corrélations entre les traits du visage et des émotions spécifiques.

Le système peut alors classer l'état émotionnel de la personne en fonction d'une ou plusieurs des six émotions de base : bonheur, tristesse, colère, peur, surprise et dégoût. Certains systèmes peuvent également détecter des émotions plus complexes telles que le mépris, la gêne ou la honte.

La technologie FER a diverses applications, notamment le marketing, l'interaction homme-machine, la sécurité et les soins de santé. Par exemple, le FER peut être utilisé en marketing pour évaluer les réactions émotionnelles des consommateurs à la publicité ou à la conception de produits. Dans le domaine de la santé, le FER peut être utilisé pour surveiller les patients souffrant de troubles mentaux et évaluer leurs états émotionnels.

Cependant, la technologie FER soulève des préoccupations éthiques liées à la confidentialité, aux préjugés et à l'utilisation abusive des données personnelles. Par conséquent, il est essentiel d'élaborer et de mettre en œuvre des directives et des réglementations éthiques pour le développement et le déploiement des systèmes FER [9].

Figure 1: *quelques exemples d'expressions faciales [W2].*

Voici décrit les expressions faciales de base qui sont communément reconnues dans toutes les cultures :

- **Bonheur :** Les coins de la bouche sont relevés et les joues peuvent être relevées, créant des pattes d'oie autour des yeux.
- **Tristesse :** Les coins de la bouche sont tournés vers le bas, les sourcils sont abaissés et les yeux peuvent apparaître tombants.
- **Colère :** Les sourcils sont abaissés et rapprochés, les yeux sont rétrécis et la bouche peut être fermée ou tournée vers le bas.
- **Peur :** Les sourcils sont relevés et rapprochés, les yeux sont élargis et la bouche peut être légèrement ouverte.
- **Surprise :** Les sourcils sont relevés et les yeux sont élargis, parfois avec la bouche ouverte.
- **Dégoût :** Le nez peut être plissé, la lèvre supérieure peut être relevée et les coins de la bouche peuvent être tournés vers le bas [10].

I.3.2. Domaine d'application de la reconnaissance des expressionsfaciales

La reconnaissance des émotions, qui est la capacité d'identifier et d'interpréter les émotions humaines, a de nombreuses applications pratiques dans divers aspects de notre vie quotidienne. L'un des domaines les plus prometteurs et passionnants dans ce domaine est la reconnaissance des expressions faciales. Cela implique l'utilisation d'algorithmes d'apprentissage automatique pour analyser les traits et les mouvements du visage afin de déduire l'état émotionnel d'une personne [11].

La reconnaissance des expressions faciales est de plus en plus explorée dans plusieurs domaines, notamment les sciences du comportement, la médecine, la sécurité et l'interface homme-machine. En sciences du comportement, la reconnaissance des expressions faciales peut être utilisée pour étudier et comprendre comment les émotions affectent le comportement humain. En médecine, il peut aider les médecins à diagnostiquer et à traiter les troubles de santé mentale en analysant les émotions et les expressions faciales des patients.

La reconnaissance des expressions faciales est également essentielle dans lesapplications de sécurité, telles que la surveillance, la vérification d'identité et la détection de mensonges. Dans l'interface homme-machine, il peut améliorer considérablement l'expérience utilisateur d'applications telles que la vidéoconférence, les jeux et l'informatique affective en permettant aux ordinateurs de détecter et d'interpréter les émotions humaines [12].

De plus, la reconnaissance des expressions faciales a des implications importantes dans l'animation basée sur les données, la robotique et d'autres domaines de l'interaction homme-ordinateur. Cela peut aider à créer des animations plus réalistes et des personnages générés par ordinateur qui peuvent transmettre des émotions. De plus, il peut être utilisé pour surveiller les travailleurs des industries à stress élevé, tels que les chauffeurs de camion, afin de détecter leur état émotionnel et de prévenir les accidents [13].

Enfin, la reconnaissance des expressions faciales a des applications potentielles dans l'évaluation de la douleur, la gestion et la recherche de bases de données d'images et de vidéos, et la détection polygraphique, entre autres. Dans l'ensemble, la reconnaissance des expressions faciales est un domaine très prometteur qui a le potentiel de révolutionner diverses industries et d'améliorer notre vie quotidienne de nombreuses façons [14].

I.3.3. L’émotion

Les émotions sont des phénomènes psychologiques complexes qui impliquent des sentiments subjectifs, des changements physiologiques et des réponses comportementales. Elles sont généralement définies comme un état mental et physiologique conscient qui surgit en réponse à un événement, une situation ou un stimulus spécifique.

- Une définition populaire des émotions vient du psychologue Paul Ekman, qui les décrit comme un schéma de réaction complexe, impliquant des éléments expérientiels, comportementaux et physiologiques, par lequel l'individu tente de faire face à une question ou un événement personnellement significatif [15].
- Une autre définition influente des émotions vient du psychologue Richard Lazarus, qui les décrit comme "une réponse immédiate et évaluative à un événement ou une situation personnellement significative, qui affecte les évaluations cognitives et les efforts d'adaptation en cours de l'individu [16].

D'autres chercheurs ont défini les émotions de manière légèrement différente, mais tous sontgénéralement d'accord pour dire que les émotions impliquent des expériences subjectives, des réponses physiologiques et des expressions comportementales.

I.3.3.1. Types d'émotions

Il existe différentes façons de catégoriser les émotions, mais une approche courante consiste à les classer en fonction de leur valence (positive, négative ou neutre) et de leur niveau d'activation (faible ou élevé). Sur la base de ces dimensions, les émotions peuvent êtreclassées en plusieurs catégories, notamment :

1. Émotions essentielles : Ce sont des émotions de base qui sont universellement reconnues et vécues par les personnes de toutes les cultures, telles que le bonheur, la tristesse, la colère,la peur, la surprise et le dégoût. Ils sont souvent accompagnés d'expressions faciales distinctes, de réponses physiologiques et de tendances comportementales [17].

2. Émotions d'inclinaison :

- **Émotions d'inclination sociale :** Ce sont des émotions qui surviennent en réponse à des situations et interactions sociales, telles que l'empathie, la culpabilité, la honte, l'embarras, l'envie et la jalousie. Ils impliquent de s'évaluer par rapport aux autres et peuvent motiver le comportement social et la communication [17].
- **Émotions d'inclination optionnelles :** ce sont des émotions qui ne sont pas nécessairement liées à des situations sociales mais qui peuvent survenir en réponse àdes objectifs, des préférences ou des valeurs personnelles, telles que l'espoir, la fierté, l'admiration, la crainte et la gratitude. Ils peuvent influencer la motivation, la prise de décision et le concept de soi [17].

Il convient de noter que les émotions sont des phénomènes complexes et à multiples facettes qui peuvent varier en intensité, en durée et en expérience subjective. De plus, certaines émotions peuvent se chevaucher ou se fondre dans d'autres, et leur catégorisation peut dépendre de facteurs culturels, contextuels ou individuels.

I.3.3.2. Les émotions de base de Paul Ekman

Les émotions de base de Paul Ekman sont un ensemble de six émotions primaires qui sont universellement reconnues dans différentes cultures et populations. Ces émotions incluent le bonheur, la tristesse, la colère, la peur, la surprise et le dégoût. Les recherches et les études d'Ekman ont montré que ces émotions s'expriment à travers des expressions faciales distinctes et peuvent être identifiées dans différentes cultures, indépendamment de la langue ou d'autres différences culturelles. Le concept d'émotions de base a été largement utilisé dans divers domaines, notamment la psychologie, les neurosciences et l'intelligence artificielle [18].

- Bonheur : Une émotion positive qui s'exprime par un sourire et implique des sentiments de joie, de contentement et de satisfaction.
- Tristesse : Une émotion négative qui s'exprime par un froncement de sourcils et implique des sentiments de chagrin, de perte et de déception.
- Colère : Une émotion négative qui s'exprime par un air renfrogné ou un front plissé et implique des sentiments de frustration, d'agacement ou d'hostilité.
- Peur : Une émotion négative qui s'exprime par des yeux agrandis et une bouche ouverte et qui implique des sentiments d'anxiété, d'appréhension ou de terreur.
- Dégoût : Une émotion négative qui s'exprime par une lèvre retroussée ou un nez plissé et qui implique des sentiments de répulsion, d'aversion ou de dégoût.
- Surprise : Une brève émotion qui s'exprime à travers des yeux écarquillés et une bouche ouverte et implique des sentiments d'étonnement ou d'inattendu.

La théorie d'Ekman des émotions de base suggère que ces émotions sont biologiquement et culturellement universelles, ce qui signifie qu'elles sont vécues et exprimées de manière similaire dans différentes cultures et sociétés.

Figure 2: *Un exemple d'expression pour les six Emotion de base de Paul Ekman [W3]*

I.3.3.3. Les trois éléments clés des émotions

Les trois éléments clés des émotions sont l'expérience subjective, la réponse physiologique et l'expression comportementale. Ensemble, ces éléments créent l'expérience émotionnelle complète.

I.3.3.3.1. Expérience subjective

Cela fait référence à la façon dont un individu m éprouve personnellement une émotion. C'est le sentiment intérieur qu'éprouve un individu lorsqu'il ressent une émotion [19].

I.3.3.3.2. Réponse physiologique

Cela fait référence aux changements corporels quise produisent en réponse à une émotion. Par exemple, lorsqu'une personneéprouve de la peur, son rythme cardiaque peut augmenter et ses paumes peuvent devenir moites [20].

I.3.3.3.3. Réponse comportementale

Cela fait référence à l'expression extérieure d'une émotion par le biais d'un comportement, comme les expressions faciales, le langage corporel et la communication verbale. Par exemple, lorsqu'une personne est heureuse, elle peut sourire, rire ou exprimer verbalement sa joie [21].

Ces trois éléments sont intimement liés et travaillent ensemble pour créer l'expérience complexe et multiforme des émotions. Comprendre ces éléments est essentiel pour comprendre la nature des émotions et leur impact sur le comportement et le bien-être humain.

I.4. Problèmes de reconnaissance des expressions faciales

La reconnaissance de l'expression faciale est une tâche complexe et difficile qui implique l'analyse et l'interprétation des mouvements et des caractéristiques du visage pour déduire l'état émotionnel d'un individu. Malgré les avantages potentiels de cette technologie, il existe plusieurs problèmes et difficultés associés à la reconnaissance des expressions faciales.

L'un des principaux défis de la reconnaissance des expressions faciales est la précision du système. Les algorithmes utilisés pour reconnaître les expressions faciales peuvent être très sensibles aux changements d'éclairage, d'expressions faciales et d'autres facteurs environnementaux, entraînant des erreurs de détection et d'interprétation. Cela peut entraîner des faux positifs ou négatifs et limiter l'utilité de la technologie [22].

Un autre problème important avec la reconnaissance des expressions faciales est le besoin de grands ensembles de données de haute qualité pour former les algorithmes avec précision. La collecte et l'annotation de ces données peuvent prendre du temps et coûter cher,ce qui rend difficile pour les chercheurs et les développeurs de créer des systèmes de reconnaissance d'expression faciale fiables et efficaces [23].

La confidentialité est une autre préoccupation associée à la reconnaissance des expressions faciales. L'utilisation de cette technologie soulève des questions sur la collecte, le stockage et l'utilisation des données personnelles, ainsi que sur le potentiel d'utilisation abusive. Il existe un risque que les systèmes de reconnaissance des expressions faciales soient utilisés pour surveiller des individus sans leur consentement ou pour suivre leur état émotionnel à des fins commerciales ou autres [24].

De plus, les systèmes de reconnaissance des expressions faciales peuvent ne pas représenter avec précision les émotions dans différentes cultures et données démographiques. Différentes cultures expriment les émotions de manière unique, et les systèmes de reconnaissance des expressions faciales peuvent ne pas être conçus pour tenir compte de ces variations. Par conséquent, ces systèmes peuvent ne pas être efficaces ou précis pour identifier les émotions chez des individus de cultures ou de milieux démographiques différents [25].

Enfin, il existe un risque que la technologie de reconnaissance des expressions faciales soit utilisée à mauvais escient pour discriminer certains groupes ou individus. Par exemple, le système peut être biaisé contre les personnes ayant certains tons de peau ou certains traits du visage, ce qui entraîne un traitement ou des décisions injustes [26].

En conclusion, la reconnaissance des expressions faciales est une technologie prometteuse qui a de nombreuses applications potentielles. Cependant, il existe plusieurs défis et difficultés associés à la technologie qui doivent être résolus pour garantir son utilisation efficace et éthique. Les chercheurs et les développeurs doivent s'efforcer d'améliorer la précision et la fiabilité des systèmes de reconnaissance des expressions faciales, de résoudre les problèmes de confidentialité et de tenir compte des différences culturelles etdémographiques dans l'expression émotionnelle [27].

I.5. Reconnaissance des expressions faciales

Un système automatique d'analyse d'expressions faciales a comme caractère essentiel trois phases principales qui sont : la détection de visage à étudier, l'extraction de caractéristiques de visage à étudier et enfin la classification de l'expression faciale. Dans ce chapitre, nous donnons un aperçu des différentes phases du système de reconnaissance automatique des expressions faciales ainsi que la description des différentes méthodesexistantes pour chaque phase [28].

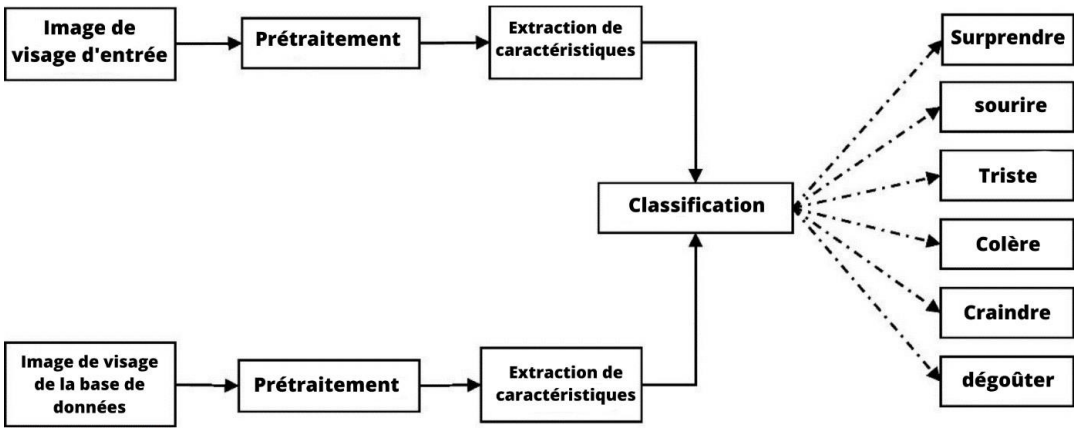

Figure 3: *Un mini architecture pour le processus de reconnaissance des émotions faciales*

I.6. Détection faciale

L'identification des visages est une innovation informatique utilisée dans un assortiment d'applications qui distinguent les visages humains dans des images informatisées. La reconnaissance faciale fait également allusion au cycle mental par lequel les gens trouvent et prennent soin des apparences dans une scène visuelle [29].

La reconnaissance faciale s'est développée des procédures de vision PC brutes à l'IA (ML) en passant par les fausses organisations neuronales raffinées. Réseau de neurones artificiels (ANN) et avancées connexes, et le résultat a été des améliorations d'exécution persistantes. Il joue actuellement un rôle important en tant que phase initiale dans quelques applications importantes, notamment le suivi du visage, l'examen du visage et la reconnaissance faciale. La reconnaissance faciale affecte fondamentalement la manière dont les activités successives sont exécutées dans une application.

Dans l'enquête sur les visages, la découverte des visages détermine les parties d'une image ou d'une vidéo sur lesquelles se concentrer pour décider de l'âge, de l'orientation et des sentiments en utilisant l'apparence. Dans un cadre de reconnaissance faciale qui cartographie numériquement les reflets faciaux d'un singulier et stocke les informations sous forme d'informations de reconnaissance faciale, une empreinte faciale unique est attendue pour les calculs qui reconnaissent quelles parties d'une image ou d'une vidéo sont censées faire une empreinte faciale. Lorsque la nouvelle empreinte digitale unique est distinguée, les empreintes faciales rangées peuvent mesurer jusqu'à décider s'il y a une correspondance [30].

Les calculs de découverte de visage commencent régulièrement par la recherche d'yeux naturels. Les yeux établissent ce que l'on appelle un lieu de vallée et sont peut-être l'élémentle moins exigeant à distinguer. Chaque fois que les yeux sont identifiés, le calcul pourrait alors s'efforcer de reconnaître les zones faciales, y compris les sourcils, la bouche, le nez, les narines et l'iris. Lorsque le calcul déduit qu'il a identifié un paramètre régional facial, il peut alors appliquer des tests supplémentaires pour déterminer s'il a reconnu un visage.

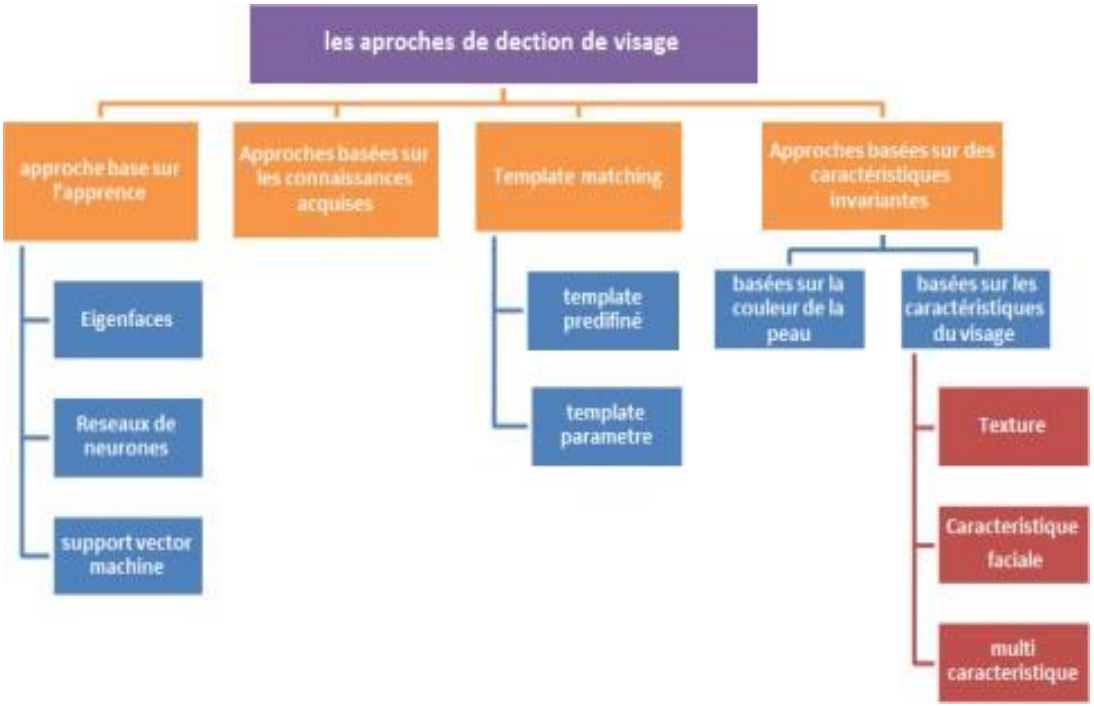

Figure 4: *Les approches de détection de visage [W4].*

I.6.1. Méthodes basées sur la connaissance

Ces techniques basées sur des normes encodent des informations humaines sur ce qui constitue un visage régulier. Normalement, les principes capturent les liens entre les reflets du visage. Ces techniques sont prévues principalement pour la limitation du visage.

Les stratégies de localisation des visages sont créées à la lumière des directives obtenues à partir des informations du spécialiste sur les apparences humaines. Il n'est pas difficile de concocter des lignes directrices de base pour représenter les éléments d'un visage et leurs connexions. Par exemple, un visage apparaît régulièrement sur une photo avec deux yeux symétriques, un nez et une bouche. Leurs distances et positions globales peuvent traiter de la relation entre les points saillants. Les éléments faciaux d'une image d'information sont supprimés en premier, et les visages qui apparaissent sont reconnus àla lumière des règles codées. Un cycle de vérification est normalement appliqué pour réduire l'identification trompeuse [31].

I.6.2. Approches invariantes des fonctionnalités :

Ces calculs consistent à observer les principaux points forts qui existent dans tous lescas, lorsque la posture, le point de vue ou les conditions d'éclairage changent, puis à les utiliser pour trouver des visages. Cette méthode est prévue principalement pour faire face à des limitations.

Contrairement à la méthodologie hiérarchique basée sur l'information, les scientifiques ont tenté de trouver des éléments invariants de visages pour la détection. La supposition de base dépend de la perception que les gens peuvent exercer reconnaît les visages et les articles dans une variété de postures et de conditions d'éclairage, il devrait doncy avoir des propriétés ou des reflets qui restent constants malgré cette variation. Diverses techniques ont été proposées pour distinguer dans un premier temps les reflets du visage et,ensuite, pour induire la présence d'un visage. Les éléments du visage tels que les sourcils, les yeux, le nez, la bouche et la racine des cheveux sont régulièrement supprimés à l'aide d'identificateurs de bord. À la lumière des reflets supprimés, un modèle mesurable est développé pour représenter leurs connexions et confirmer la présence d'un visage [31].

I.7. Extraction des points caractéristiques faciaux

Les points saillants du visage se concentrent principalement autour des éléments faciaux clés tels que les yeux, la bouche, les sourcils, le nez et le menton. La détection de ces points caractéristiques du visage commence généralement par l'utilisation d'un détecteur de visage qui identifie initialement un rectangle englobant autour du visage.

Ensuite, en extrayant des caractéristiques géométriques telles que les contours des composants faciaux, les distances entre eux, et d'autres paramètres, nous pouvons déterminer les emplacements précis ou les caractéristiques d'apparence peuvent être calculées.

Ainsi, les méthodes d'extraction des caractéristiques pour l'analyse des expressions peuvent être catégorisées en deux approches distinctes : celles basées sur les caractéristiques géométriques et celles basées sur l'apparence [32] :

I.7.1. Les caractéristiques géométriques

Elle capture la configuration et la position des éléments du visage, englobant la bouche, les yeux, les sourcils et le nez. Les composants faciaux, également appelés traits faciaux, sont isolés et transformés en un vecteur de caractéristiques qui reflète la structure géométrique du visage [33]. Cette méthode englobe divers modèles visant à atteindre la reconnaissance des expressions faciales, tels que [32] :

- Modèle de forme active (ASM)
- Modèles d'apparence active (AAM)

I.7.2. Les caractéristiques d'apparence

Cette approche concerne la capture des altérations de l'apparence du visage, telles que les rides et les sillons, qui relèvent de la texture de la peau. Ces caractéristiques d'apparence peuvent être extraites à partir de l'ensemble du visage ou de zones spécifiques de celui-ci. En fonction des méthodes d'extraction des caractéristiques employées, les effets de la rotation de la tête dans le plan et les variations d'échelle lors de la prise de vue du visage peuvent être corrigés par une normalisation préalable du visage avant l'extraction des caractéristiques ou par une représentation des caractéristiques avant la phase de reconnaissance d'expressions [33]. Pour cette approche, différentes techniques ont été développées pour obtenir la reconnaissance des expressions faciales, comme indiqué dans [32] :

- Motif binaire local (LBP)
- Quantification de phase locale (LPQ)
- Histogramme de gradient orienté (HOG)

I.8. Type D'apprentissage

L'intelligence artificielle englobe trois cadres d'apprentissage qui caractérisent ses diverses méthodes d'activité. Ceux-ci comprennent l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage semi-supervisé. Nous nous concentrerons sur l'apprentissage supervisé et non supervisé car ce sont les plus couramment utilisés.

I.8.1. L'apprentissage supervisé

L'apprentissage supervisé est le principe sous-jacent de nombreuses applications modernes fascinantes, telles que la reconnaissance faciale sur nos photos par nos smartphones et les filtres anti-spam pour les e-mails.

De manière plus formelle, dans le cadre de l'apprentissage supervisé, en partant d'un ensemble de données D décrit par un ensemble de caractéristiques X, un algorithme d'apprentissage va découvrir une fonction de correspondance entre les variables prédictives en entrée X et la variable à prédire Y. Cette fonction de correspondance, qui décrit la relation entre X et Y, est appelée un modèle de prédiction[33].

$$f(X) \rightarrow Y$$

Les caractéristiques X, qu'elles soient numériques, alphanumériques ou sous forme d'images, offrent une variété de possibilités. En ce qui concerne la variable prédite Y, elle peut être classée en deux catégories distinctes :

D'une part, il y a la catégorie des variables discrètes. Dans ce contexte, la variable à prédire peut adopter une valeur parmi un ensemble limité d'options, généralement désigné sous le nom de "classes". Par exemple, pour anticiper si un e-mail est classé comme SPAM ou non, la variable Y peut revêtir deux valeurs possibles :

Y \in \ {SPAM, NON SPAM\}

Variable continue : La variable Y a la liberté de prendre n'importe quelle valeur. Pour mieux comprendre ce concept, imaginez un algorithme qui utilise les caractéristiques d'un véhicule en entrée et s'efforce de prédire son prix (la variable Y).

Ainsi, la classification de la variable prédite Y divise l'apprentissage supervisé en deux sous-catégories distinctes :

- La classification
- La régression

I.8.1.1. Les algorithmes de classification

Lorsque la variable à prédire présente une valeur discrète, cela constitue un problème de classification. Parmi les algorithmes couramment utilisés pour résoudre de tels problèmes, on peut citer le Support Vector Machine (SVM), les Réseaux de Neurones, le Naïve Bayes, la Régression Logistique, entre autres. Chacun de ces algorithmes possède ses propres propriétés mathématiques et statistiques. Le choix de l'algorithme à utiliser dépend des données d'entraînement (Training set) ainsi que de nos caractéristiques (features). Cependant, l'objectif reste le même : être en mesure de prédire à quelle classe appartient une donnée, par exemple, déterminer si un nouvel e-mail est du spam ou non.

Lorsque l'ensemble des valeurs possibles d'une classification dépasse deux éléments, on parle de classification multi-classes (ou Multi-class Classification). L'image suivante illustre ces deux types de classifications.

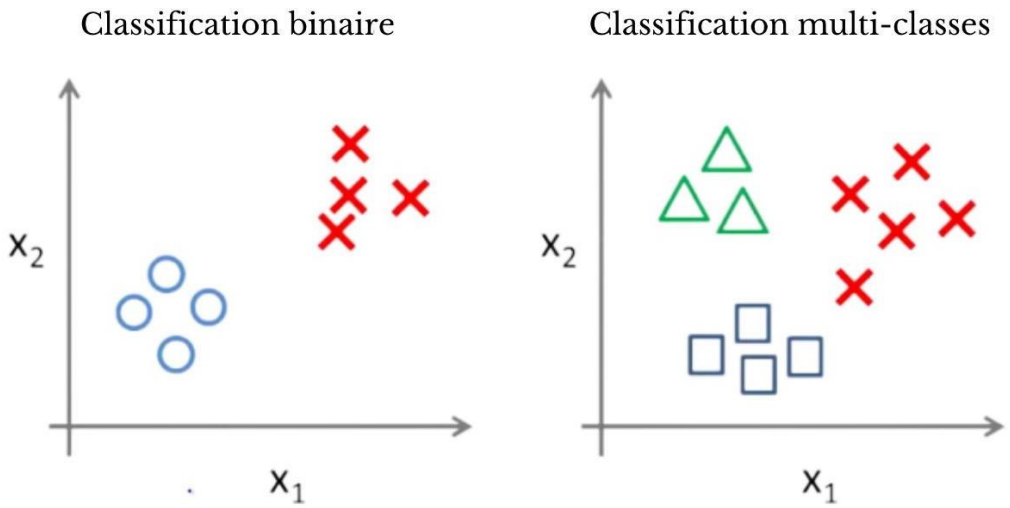

Figure 5: *illustre les deux types de classifications [W6].*

Dans l'image précédente (Figure 5), les cercles bleus symbolisent une classe (par exemple, les e-mails non-spam), tandis que les croix rouges peuvent représenter des e-mails SPAM. L'image de droite illustre une classification multi-classes, car elle comporte trois classes potentielles (les triangles, les croix et les carrés).

I.8.1.2. Les algorithmes de régression

Un algorithme de régression vise à élaborer un modèle, c'est-à-dire une fonction mathématique, en se basant sur les données d'entraînement. Ce modèle calculé sera ensuite utilisé pour fournir une estimation sur de nouvelles données qui n'ont pas été précédemment rencontrées par l'algorithme, c'est-à-dire des données qui ne faisaient pas partie de l'ensemble d'entraînement.

Les algorithmes de régression peuvent revêtir différentes formes selon le type de modèle que l'on souhaite construire. La régression linéaire représente le modèle le plus élémentaire : elle consiste à trouver la droite qui s'ajuste au mieux aux données d'apprentissage. Par conséquent, la fonction de prédiction dans ce cas prend la forme d'une ligne droite.

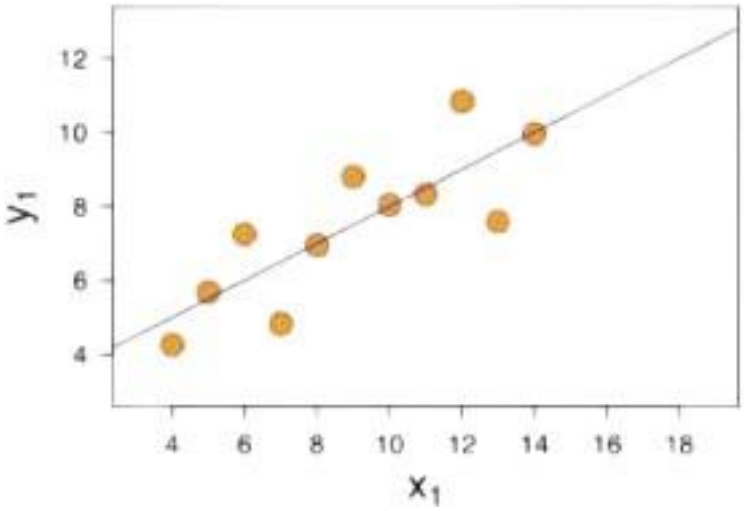

Figure 6: *l'algorithme de régression linaire [W5]*

Le modèle prédit par l'algorithme de régression linaire sera de la forme

$Y = f(X) = \alpha * X_1 + \beta$ (a et \beta sont les coefficients de la droite).

Dans la réalité, se fier uniquement à une corrélation linéaire entre les données ne suffit pas pour élaborer des modèles prédictifs robustes. Les données peuvent souvent ne pas entretenir de relation linéaire les unes avec les autres, et il peut être nécessaire d'intégrer plusieurs variables prédictives pour réaliser des prédictions réalistes. La régression polynomiale et la régression multivariée (impliquant plusieurs variables) émergent comme des outils précieux pour créer des fonctions de mappage complexes qui s'ajustent de manière plus précise aux données d'entraînement.

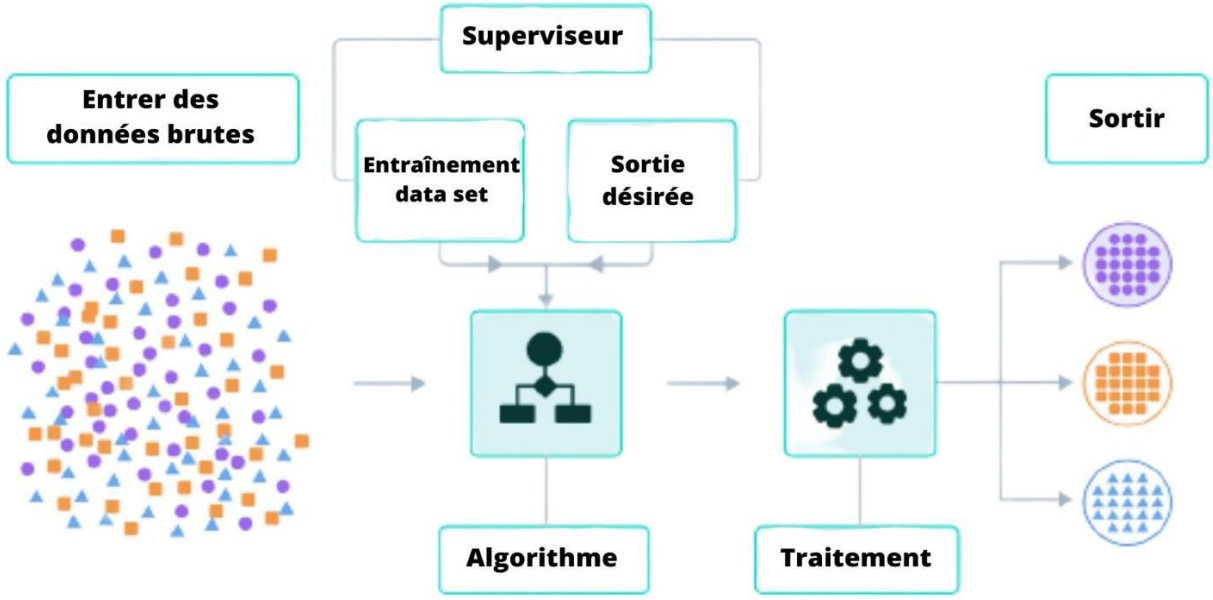

Figure 7: *Processus d'apprentissage supervisé [W6]*

I.8.2. Apprentissage non supervisé

Dans cette méthode d'activité d'apprentissage automatique, il est indéniable que, en fonction de composants prédéfinis, la tâche dépend en fin de compte de la machine pour classer les informations toute seules. Pour ce faire, le système va croiser les données qui lui sont soumises afin de regrouper dans la même classe les éléments qui présentent dessimilitudes spécifiques. Par conséquent, selon l'objectif souhaité, il incombe à l'opérateur ou à l'expert d'analyser ces résultats pour en dégager les différentes hypothèses [33].

Les modèles d'apprentissage non supervisé trouvent leur utilité principalement dans :

- Le classement des données
- Le calcul approximatif de la densité de distribution
- La réduction des dimensions

L'application de l'apprentissage non supervisé peut être regroupée en deux types de problèmes : le regroupement (clustering) et l'association.

I.8.2.1. Clustering

Un problème de regroupement (clustering) consiste à demander à la machine de regrouper des objets présents dans un ensemble de données en groupes de la manière la plus précise et efficace possible. Cette méthode, bien que parfois complexe à appréhender pour les êtres humains, trouve une application fréquente dans des domaines tels que le marketing, où elle permet de classer différents clients dans des groupes pertinents, par exemple. Un exemple couramment utilisé d'algorithme de regroupement est le K-means.

I.8.2.2. Association

Le système d'association a pour objectif de trier et de regrouper des données qui partagent des caractéristiques communes, permettant ainsi de découvrir des objets liés les uns aux autres sans qu'ils soient nécessairement identiques. À titre d'exemple, en alimentant l'algorithme avec de nombreuses images de chats et d'accessoires pour chats, l'apprentissage non supervisé ne regrouperait pas simplement tous les chats ensemble, mais pourrait identifier des associations telles qu'une pelote de laine avec un chat. Un exemple couramment utilisé d'algorithme d'association est l'algorithme Apriori.

-	**Apprentissage supervisé**	**Apprentissage non supervisé**
Données d'entrée	Données connues en entrée	Données inconnues en entrés
Complexité informatique	Complexe	Moins Complexe
Domaines d'activités	Classification et Régression	Exploitation de régles de clustering et d'association
Précision	Produit des résultats Précis	Génère des résultats modérés

Tableau 1: *comparaison entre apprentissage supervisé et non supervisé*

I.9. Classification et reconnaissance des expressions

Le dernier élément du cadre FER dépend de l'hypothèse de l'IA ; c'est certainement la tâche de regroupement. La contribution au classificateur est un ensemble d'éléments qui ontété récupérés de la zone du visage à l'étape précédente. La disposition des éléments est façonnée pour représenter le regard. La commande nécessite une préparation dirigée, de sorte que l'ensemble de préparation doit comprendre des informations marquées. Lorsque le classificateur est prêt, il peut percevoir les images d'entrée en leur attribuant un nom de classespécifique.

L'ordre de regard le plus fréquemment utilisé se fait à la fois en termes de systèmede codage d'action faciale proposé et en termes de sentiments globaux : joie, misère, indignation, choc, répulsion et terreur. Il existe différentes méthodes d'IA pour la tâche de regroupement, Pour être précis (33) :

- Réseaux de neurones **(Neural Networks, NN).**
- Machines à vecteurs de support **(Support Vector Machine, SVM).**
- Analyse Discriminante Linéaire **(Linear Discriminant Analysis, LDA).**
- K-plus proches voisins **(K Nearest Neighbors, KNN).**
- Régression logistique multinomiale **(Multinomial Logistic Regression, MRL).**
- Modèles de Markov cachés **(Hidden Markov Model, HMM).**
- Réseaux bayésiens **(Bayesian Network, BN),** et bien d'autres encore.

Choisir un bon ensemble de capacités, une méthode d'IA compétente et un ensemble de données diversifié pour l'apprentissage sont trois problèmes majeurs dans le regroupement de projets. Les ensembles de surbrillance doivent être constitués d'éléments discriminants et de marques pour une articulation spécifique. Ce type de liste de fonctionnalités reprend les procédures d'IA en général. Enfin, la base d'informations utilisée comme ensemble de préparation devrait être suffisamment grande et contenir différents types d'informations.

I.10. Conclusion

Dans ce chapitre, nous avons exposé les théories et les représentations les plus éminentes en matière de reconnaissance des expressions faciales. Nous avons abordé les techniques de codification, les approches de détection de visages, ainsi que l'extraction des caractéristiques. De plus, nous avons passé en revue les bases de données couramment utilisées dans ce domaine. Forts des connaissances acquises au cours de cette étude, le chapitre suivant se consacrera à l'exploration du Deep Learning pour la reconnaissance des expressions faciales.

CHAPITRE II

DEEP LEARNING

Chapitre II : Deep Learning

II.1. Introduction

L'apprentissage en profondeur est un domaine de pointe de l'intelligence artificielle quia révolutionné la façon dont les machines apprennent et font des prédictions. Avec sa capacitéà traiter de grandes quantités de données et à reconnaître des modèles complexes, l'apprentissage en profondeur a été appliqué à un large éventail d'applications, de la reconnaissance d'images et de la parole aux voitures autonomes et au diagnostic médical. Enapprofondissant ce sujet fascinant, vous découvrirez le fonctionnement interne des réseaux de neurones, apprendrez à construire et à former vos propres modèles et explorerez les possibilités et les limites de cette technologie passionnante. Alors, préparez-vous à plonger profondément dans le monde de l'apprentissage en profondeur

II.2. L'Apprentissage en Profondeur (Deep Learning)

Le Deep Learning représente une nouvelle frontière dans le domaine de l'apprentissage automatique (Machine Learning, ML) et a été introduit dans le but de rapprocher le ML de son objectif fondamental : l'intelligence artificielle. Ce domaine englobe des algorithmes inspirés de la structure et du fonctionnement du cerveau humain. Ces algorithmes sont capables d'apprendre des niveaux de représentation multiples afin de modéliser des relations complexes entre les données.

II.2.1. Définition

L'avènement de l'apprentissage profond découle du constat que les réseaux neuronaux devenaient de plus en plus complexes, comportant un nombre croissant de couches cachées. Cependant, lorsque le nombre de ces couches atteignait un certain seuil, cela posait des défis majeurs. En effet, au-delà d'un certain nombre de couches, les réseaux neuronaux éprouvaient des difficultés à assimiler efficacement les informations et à apprendre de manière adéquate.

Cependant, des solutions ont été élaborées pour résoudre ces problèmes, permettant ainsi aux réseaux neuronaux de disposer de multiples couches et d'apprendre de manière performante. Toutes ces variantes de réseaux neuronaux sont regroupées sous le terme global de "Deep Learning" [42].

II.2.2. Historique

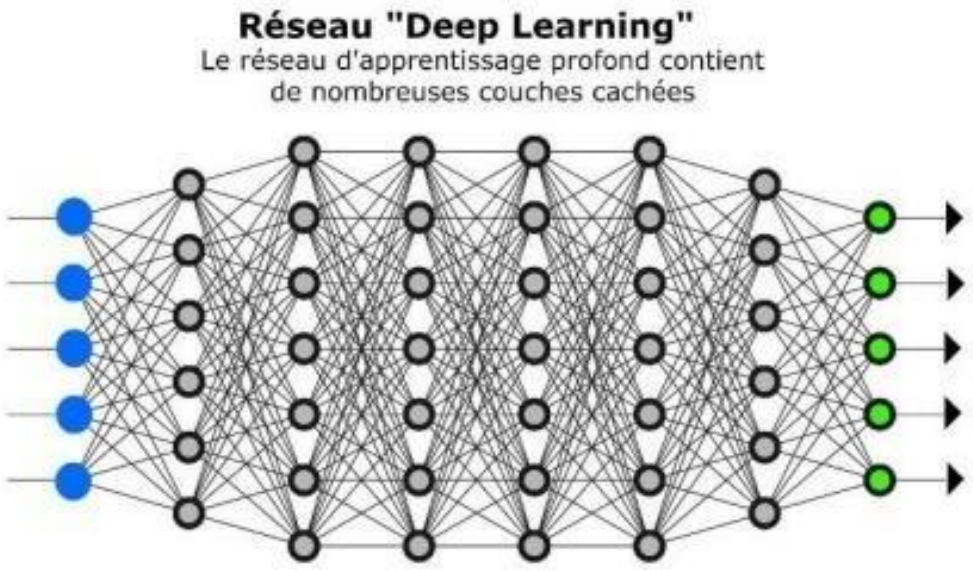

Figure 8: *Schéma illustratif d'apprentissage profond avec plusieurs couches [W6]*

Le concept de réseaux de neurones artificiels (RNA) a été introduit pour la première fois dans les années 1940 par Warren McCulloch et Walter Pitts. Ils ont proposé un modèle pour créer une machine qui pourrait imiter le comportement d'un cerveau humain. [37]

- Dans les années **1950** et **1960**, des chercheurs tels que Frank Rosenblatt et Bernard Widrow ont développé les premiers RNA pratiques, connus sous le nom de perceptrons. Cependant, les limites de la puissance de calcul et de la disponibilité desdonnées à l'époque ont entravé leur succès [38].
- Dans les années **1980**, l'algorithme de rétropropagation a été développé, ce qui a permis aux ANN d'apprendre plus efficacement des données. Cette percée a conduit au développement de perceptrons multicouches, capables d'apprendre des modèles plus complexes [39].
- Dans les années **1990**, l'apprentissage en profondeur a commencé à attirer davantagel'attention, avec l'introduction de nouvelles architectures telles que les réseaux de neurones convolutifs (CNN) et les réseaux de neurones récurrents (RNN). Cependant, le manque de données et de puissance de calcul limite encore leur succès [40].

- La percée de l'apprentissage en profondeur a eu lieu dans les années 2010, avec la disponibilité de grands ensembles de données et le développement de GPU plus puissants. Cela a conduit au développement de réseaux de neurones profonds (DNN) à plusieurs couches, capables d'apprendre des modèles très complexes et d'obtenir des résultats de pointe dans diverses tâches telles que la reconnaissance d'images et de la parole [41].

Dans l'ensemble, l'apprentissage en profondeur a une longue et riche histoire, avec de nombreuses étapes importantes et des contributions de divers chercheurs et praticiens au fildes décennies**.**

II.2.3. Pour quoi le Deep Learning

Au commencement, les divers algorithmes du Deep Learning ont émergé en réponse aux limitations de l'apprentissage automatique, qui cherchait à résoudre une multitude de problèmes en intelligence artificielle (IA). Ces nouvelles approches sont apparues dans le but :

- Afin d'améliorer le développement des algorithmes traditionnels dans de telles tâchesde 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 [42].

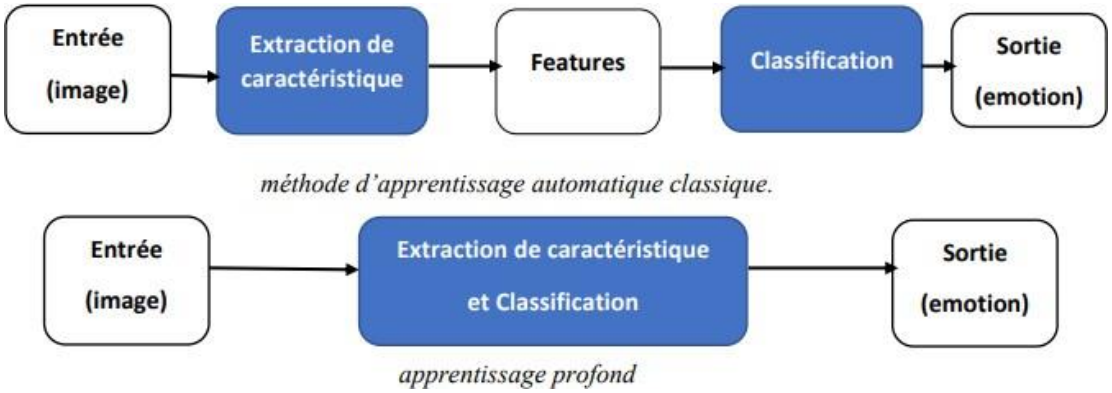

Figure 9: *Comparaison entre la machine Learning et le Deep Learning*

II.2.4. Concepts de Deep Learning

Les concepts du Deep Learning englobent une catégorie d'algorithmes d'apprentissage automatique qui se caractérisent par les éléments suivants :

- Ils utilisent plusieurs couches d'unités de traitement non linéaires pour extraire et transformer les caractéristiques des données. Chaque couche prend comme entrée la sortie de la couche précédente. Ces algorithmes peuvent être supervisés ou non supervisés, et leurs applications couvrent des domaines tels que la reconnaissance de motifs ou la classification statistique.
- Ils opèrent avec un apprentissage qui se déroule à plusieurs niveaux de détail ou de représentation des données. À travers ces différentes couches, la progression s'effectue des paramètres de bas niveau vers des paramètres de plus haut niveau.
- Ces divers niveaux correspondent à différentes strates d'abstraction des données.
- Ce domaine d'étude innovant vise à rapprocher les capacités de l'intelligence artificielle. Les architectures du Deep Learning sont désormais capables de conférer du sens à des données sous forme d'images, de sons ou de texte.

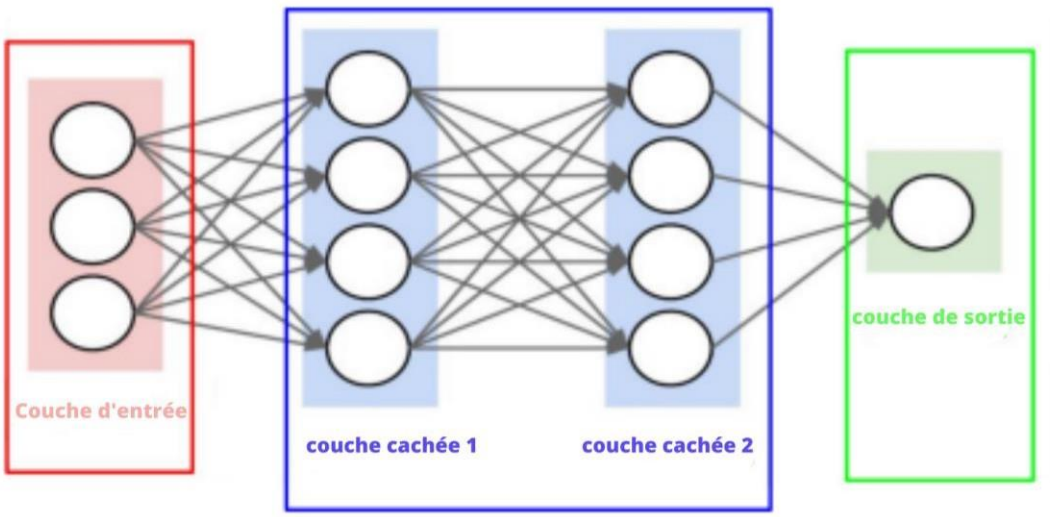

Figure 10: *Une structure générale d'un réseau de neurones profonds*

Le système de Deep Learning repose sur des réseaux de neurones artificiels composés d'une série de couches cachées. L'adjectif "profond" (d'où le terme "apprentissage profond") tire son origine du nombre de ces couches cachées, comme illustré dans la figure. La distinction clé entre un perceptron classique et un système de Deep Learning réside dans le fait que, dans le premier cas, les entrées du réseau sont les caractéristiques de l'image, tandis que, dans le second cas, ce sont les pixels bruts de l'image d'origine. En réalité, dans un système de Deep Learning, chaque couche est considérée comme une strate d'abstraction de l'image.

La première couche est capable d'extraire des caractéristiques de niveau d'abstraction inférieur à celui de la deuxième couche, tandis que la troisième couche est encore plus abstraite. À partir de ces caractéristiques, le système est en mesure de reconnaître des objets avec un taux d'erreur minime [43].

II.2.5. Les avantages de Deep Learning

Le Deep Learning offre de nombreux avantages distinctifs, parmi lesquels :

- Une grande robustesse pour la compréhension et l'utilisation de nouvelles données.
- La gestion à un niveau d'abstraction bien plus élevé que les réseaux de neurones classiques.
- L'obtention de résultats plus rapides, avec un apprentissage qui se fait progressivement au fil du temps plutôt que de manière instantanée.
- La capacité à traiter de vastes volumes de données avec un coût d'apprentissage considérablement réduit pour les réseaux de taille réduite.

II.2.6. Quelques types d'algorithmes de Deep Learning

Il existe de nombreux algorithmes de deep Learning, chacun ayant ses propres particularités et avantages en fonction de l'application. Voici quelques-uns des algorithmes les plus couramment utilisés en deep Learning :

- **Les réseaux de neurones convolutifs (CNN) :** utilisés pour la reconnaissance d'images, de vidéos et de sons, et pour le traitement de la parole [43].

- **Les réseaux de neurones récurrents (RNN) :** utilisés pour le traitement de séquences de données, comme le langage naturel et la traduction automatique [44].
- **Les réseaux de neurones adverses génératifs (GAN)** : utilisés pour la génération de données, comme les images et les vidéos[45].
- **Les réseaux de neurones auto-encodeurs (AE) :** utilisés pour la compression et la reconstruction de données, comme les images et les vidéos [46].
- **Les réseaux de neurones résiduels (ResNet)** : utilisés pour la reconnaissance d'images et pour la classification de données complexes [47].

Ces algorithmes ne sont pas exhaustifs et il existe de nombreux autres types de réseaux de neurones profonds utilisés dans diverses applications de Deep Learning.

II.2.7. Les différentes Architectures du Deep Learning

Malgré la profusion de variantes d'architectures profondes, il n'est pas toujours envisageable de comparer les performances de toutes ces architectures, car elles ne sont pas toutes évaluées sur les mêmes ensembles de données. Le domaine du Deep Learning est en constante évolution, avec l'émergence de nouvelles architectures, variantes et algorithmes presque chaque semaine.

II.2.7.1. L e s réseaux de neurones convolutifs (CNN)

Les CNN (Convolutional Neural Networks), ou réseaux de neurones convolutionnels, sont un type de réseau de neurones spécialement conçu pour traiter des données ayant une structure semblable à une grille. Ils ont démontré une grande efficacité dans des domaines tels que la reconnaissance et la classification d'images et de vidéos. Les CNN ont réussi à identifier des éléments tels que des visages, des objets, des panneaux de signalisation et même à être utilisés dans des applications de conduite autonome. Plus récemment, les CNN se sont également révélés performants dans plusieurs tâches de traitement du langage naturel, notamment la classification de phrases [48].

Dans le domaine de l'apprentissage automatique, un réseau convolutionnel est une variante de réseau de neurones à propagation avant qui trouve son inspiration dans les processus biologiques. Le CNN se compose de quatre opérations fondamentales, comme démontré ci-dessous:

- La couche convolution.
- La couche Rectified Linear Unit.

- La couche Pooling.
- La couche entièrement connectée.

II.3. Réseau de neurones récurrents

L'idée sous-jacente aux Réseaux de Neurones Récurrents (RNN) est d'exploiter des informations séquentielles. Dans un réseau neuronal classique, nous supposons que toutes les entrées (et sorties) sont mutuellement indépendantes. Cependant, cette approche se révèle peu adaptée à de nombreuses tâches. Par exemple, pour prédire le prochain mot dans une phrase, il est essentiel de connaître les mots précédents. Les RNN sont qualifiés de "récurrents" car ils exécutent une tâche similaire pour chaque élément d'une séquence, avec la sortie dépendant des calculs antérieurs [48].

On peut également concevoir les RNN comme ayant une "mémoire" qui retient l'information générée jusqu'à présent. En théorie, les RNN peuvent exploiter des informations dans des séquences de longueur arbitraire. Cependant, en pratique, ils sont souvent restreints à examiner seulement quelques étapes en arrière. Ils trouvent leur utilisation dans :

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

Modèle génératif

Alors que les modèles discriminatifs tels que les CNN et les RNN sont employés pour prédire les données en se basant sur les entrées et les étiquettes, les modèles génératifs, quant à eux, se concentrent sur la manière de générer ces données. Ils apprennent et effectuent des prédictions en utilisant la loi de Bayes [49].

Cependant, les modèles génératifs sont capables de bien plus que de simples classifications, ils sont par exemple aptes à créer de nouvelles observations. Voici quelques exemples de modèles génératifs :

- Boltzmann Machines
- Restricted Boltzmann Machines
- Deep Belief Networks
- Deep Boltzmann Machines
- Générative Adversarial Networks

Exemple de l'application de modèles generatifs

Les applications du Deep Learning se retrouvent dans une variété de secteurs, de la conduite autonome aux dispositifs médicaux [49]. Grâce à l'apprentissage profond, nous sommes désormais capables de :

- Coloriser des images en noir et blanc.
- Ajouter des sons à des films silencieux.
- Effectuer des traductions automatiques.
- Classer des objets dans des photographies.
- Générer de l'écriture automatique.
- Créer des légendes pour des images.
- Jouer automatiquement à des jeux.

Réseaux de neurones convolutifs

Au cours de cette étape, nous nous concentrerons exclusivement sur le CNN, qui constitue l'approche centrale de notre système de reconnaissance des émotions.urant cette phase nous baserons uniquement sur le CNN qui est l'approche utilisée dans notre système de reconnaissance des émotions.

II.3.1. Présentation

Les réseaux de neurones convolutionnels (CNN) sont actuellement les modèles les plus performants dans des domaines tels que la reconnaissance et la classification d'images. Ils ont démontré leur capacité à identifier des éléments tels que des visages, des objets, des panneaux de signalisation et même à être utilisés dans la conduite autonome. Au cours de cette phase de formation, notre focalisation sera exclusivement sur le CNN, abrégé par l'acronyme CNN (Convolutional Neural Network)[50]

Le terme "réseau de neurones convolutif" fait référence à l'utilisation d'une opération mathématique appelée convolution. La convolution est une opération linéaire spécifique. Les réseaux convolutifs, en fait, sont des réseaux de neurones qui utilisent la convolution à la place de la multiplication matricielle dans au moins une de leurs couches. Ils se composent de deux parties principales.

Lorsque nous fournissons une image en entrée, celle-ci est représentée sous forme d'une matrice de pixels à deux dimensions pour les images en niveaux de gris et à trois dimensions pour les images en couleur (avec les canaux Rouge, Vert et Bleu). Cette image traverse la première partie d'un CNN, qui est la partie convolutive. Cette partie agit comme un extracteur de caractéristiques d'images.

L'image passe successivement à travers une série de filtres ou de noyaux de convolution, qui la transforment en de nouvelles images appelées cartes de convolution ou "feature maps". Certains de ces filtres intermédiaires peuvent réduire la résolution de l'image en utilisant une opération de maximum local. Finalement, les cartes de convolution sont aplaties et combinées en un vecteur de caractéristiques appelé le "code CNN".

Le résultat en sortie de la partie convolutive est ensuite utilisé comme entrée pour une deuxième partie du réseau, composée de couches entièrement connectées (similaire à un perceptron multicouche). Cette partie a pour rôle de combiner les caractéristiques extraites de l'ensemble du réseau pour classer l'image. La couche de sortie contient généralement un neurone par classe, fournissant des valeurs numériques normalisées entre 0 et 1, qui représentent la distribution de probabilité sur les différentes classes [51].

Les CNN sont particulièrement adaptés au traitement de grandes quantités de données telles que les images pour plusieurs raisons :

- Les images présentent une corrélation spatiale, ce qui signifie que les valeurs des pixels voisins sont généralement très similaires. Les couches de convolution sont capables de capturer ces liens spatiaux entre les données.
- Les couches de convolution sont souvent suivies de couches de sous-échantillonnage (pooling), ce qui permet de réduire significativement la taille des données [52].

II.3.2. Différents modules d'un réseau de neurones convolutif

Les CNN sont constitués de trois types de couches distinctes : les couches convolutionnelles, les couches de regroupement et les couches entièrement connectées. Lorsque ces couches sont empilées en séquence, elles forment l'architecture d'un CNN. Chacune de ces couches joue un rôle spécifique :

- La couche de convolution (CONV) traite les données à l'intérieur d'un champ récepteur donné.
- La couche de pooling (POOL) a pour fonction de compresser l'information en réduisant généralement la taille de l'image intermédiaire, souvent en utilisant une opération de sous-échantillonnage.
- La couche de correction (Relu), fréquemment désignée par "Relu" en référence à la fonction d'activation utilisée, qui est l'Unité de Rectification Linéaire.
- La couche "entièrement connectée" (FC), qui ressemble à une couche de perceptron classique.
- Enfin, la couche de perte (LOSS) est responsable de l'évaluation de la différence entre les prédictions du modèle et les valeurs réelles [53].

Figure 11: *Architecture générale d'un CNN.*

II.3.2.1. La convolution

La convolution est l'élément central des réseaux de neurones convolutionnels. À l'origine, la convolution est un concept mathématique largement utilisé en traitement d'images, car elle permet d'extraire des caractéristiques importantes à partir des images d'entrée, ce qui permet d'appliquer des filtres appropriés. En pratique, la convolution fonctionne en prenant une image en entrée et un filtre (qui est essentiellement une autre image), effectue un calcul spécifique, puis produit une nouvelle image en sortie (généralement de taille réduite).

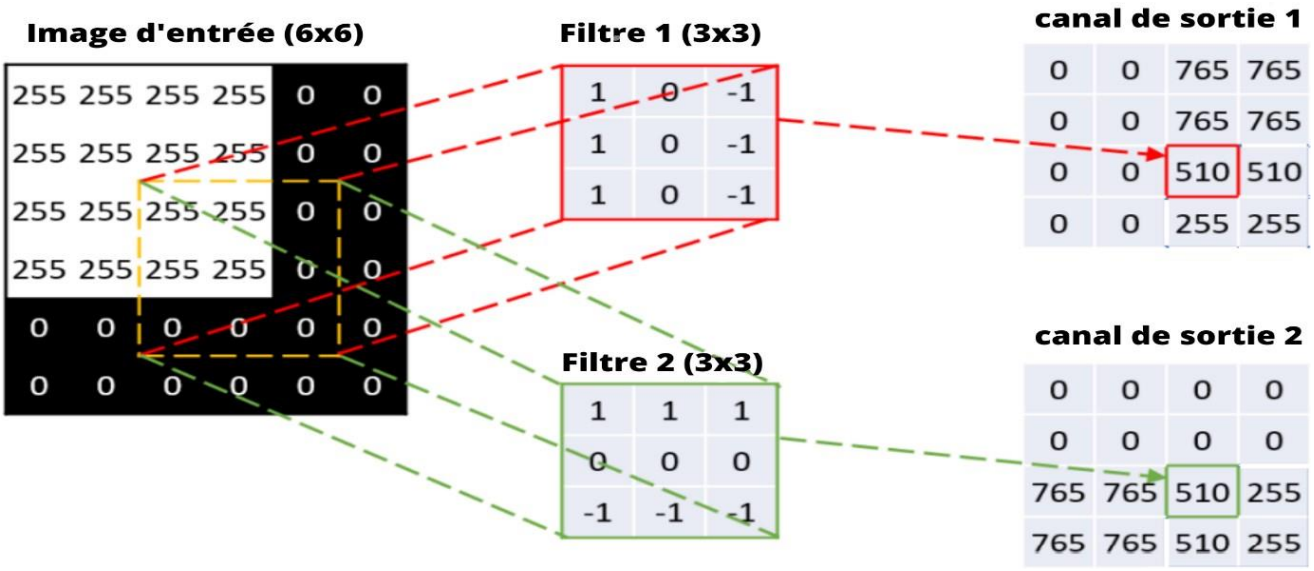

Figure 12: *Illustration d'une opération de convolution entre deux couches [W7]*

La couche de convolution, également connue sous le nom de volume de sortie, est dimensionnée en fonction de trois hyperparamètres principaux : la profondeur, le pas et la marge [52].

- **La profondeur de la couche** représente le nombre de noyaux de convolution, ou le nombre de neurones associés à un même champ récepteur.
- **Le pas (stride)** contrôle le chevauchement des champs récepteurs. Plus le pas est petit, plus les champs récepteurs se chevauchent, ce qui augmente la taille du volume de sortie.
- **La marge (zero-padding),** souvent réglée à 0, consiste à ajouter des zéros autour des bords du volume d'entrée. La taille de ce "zero-padding" constitue le troisième hyperparamètre.

L'utilisation de cette marge permet de contrôler la dimension spatiale du volume de sortie. En certains cas, il est souhaitable de conserver la même surface que celle du volume d'entrée [52].

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.

II.3.2.2. Les différentes convolutions

Il existe plusieurs types de convolutions, bien que la convolution classique soit généralement la plus couramment utilisée. Il peut cependant être utile de connaître les différentes options à notre disposition [50].

- La convolution classique : Elle consiste en un décalage du noyau entre chaque calcul, et le padding détermine la manière dont on peut « déborder » de l'image pour appliquer la convolution.
- La dilated convolution : Elle est similaire à la convolution classique, mais avec la particularité que le noyau est étendu (par exemple, en sautant un pixel sur deux pour calculer la convolution). Un paramètre supplémentaire, le taux de dilation, indique le nombre de pixels à ignorer.
- La transposed convolution : Elle construit la sortie comme si on inversait une convolution sur l'image.
- La séparable convolution : Elle représente une convolution décomposable en convolutions plus simples.

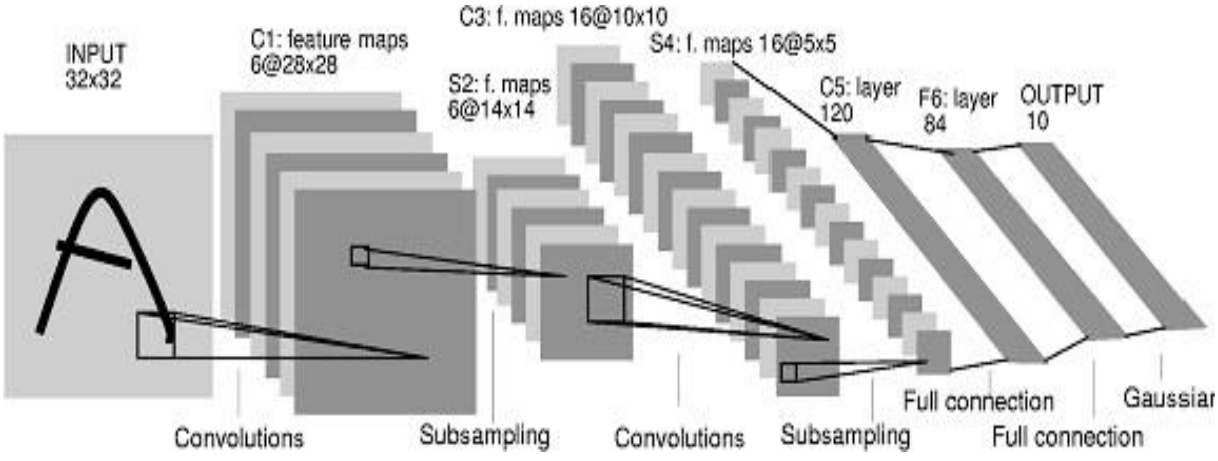

Figure 13: *Les différentes couches d'un réseau de neurones convolutif standard [W8]*

II.3.2.3. Le Pooling

Ce type de couche est généralement inséré entre deux couches de convolution : elle prend en entrée plusieurs cartes de caractéristiques et applique l'opération de pooling à chacune d'entre elles. L'opération de pooling (ou sous-échantillonnage) a pour objectif de réduire la taille des images tout en préservant leurs caractéristiques essentielles.

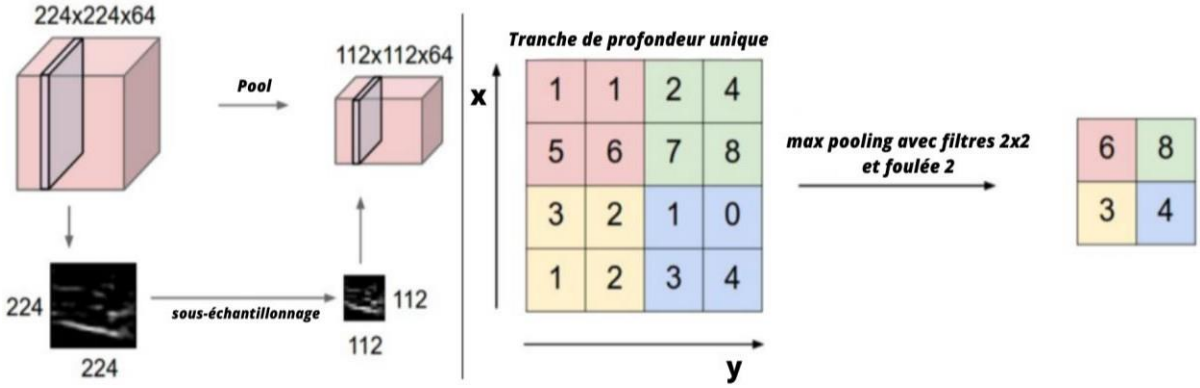

Figure 14: *Illustration du Max Pooling [W9]*

Pour cela, l'image est subdivisée en cellules régulières, et à l'intérieur de chaque cellule, seule la valeur maximale est conservée. En pratique, on utilise souvent des cellules carrées de petite taille pour minimiser la perte d'informations. Les options les plus courantes sont des cellules carrées de 2 x 2 pixels sans chevauchement ou des cellules de 3 x 3 pixels avec un chevauchement de 2 pixels. En sortie, le nombre de cartes de caractéristiques reste le même, mais elles sont considérablement réduites en taille. La couche de pooling permet de réduire le nombre de paramètres et de calculs dans le réseau, améliorant ainsi son efficacité et prévenant le surapprentissage.

Il existe plusieurs types de pooling, dont les principaux sont les suivants :

- **Max pooling :** qui consiste à sélectionner la valeur maximale dans la région. C'est le type le plus couramment utilisé en raison de sa rapidité de calcul et de son efficacité pour simplifier l'image.
- **Mean pooling :** qui calcule la moyenne des valeurs dans la région. Il s'agit de la somme de toutes les valeurs divisée par le nombre de valeurs, produisant ainsi une valeur intermédiaire pour représenter ce groupe de pixels.
- **Sum pooling :** qui est similaire au mean pooling, mais sans la division par le nombre de valeurs, se contentant de calculer la somme totale des valeurs.

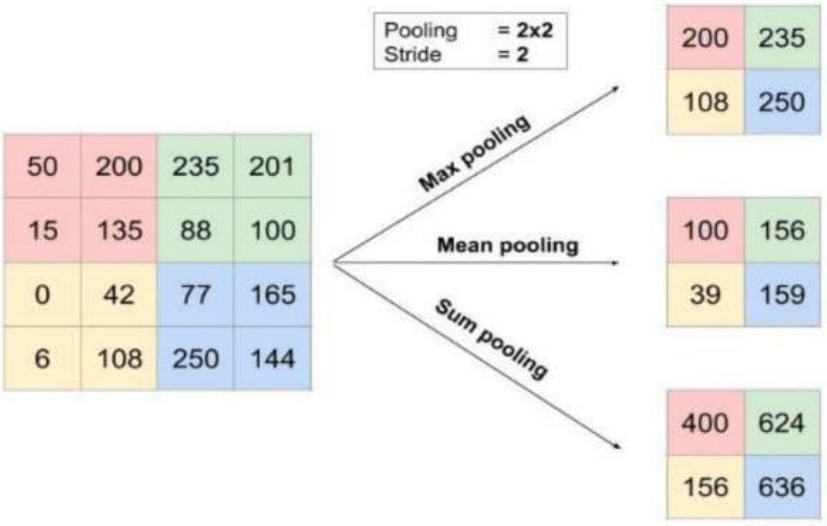

Figure 15: *les différents types de Pooling avec un filtre 2x2 et un pas de 2 [W9]*

II.3.2.3. Les fonctions d'activations

Pour améliorer l'efficacité d'un réseau CNN, il est possible d'insérer des fonctions qui agissent comme des couches de correction entre les couches de traitement. Ces fonctions introduisent des non-linéarités pour permettre au réseau d'apprendre des systèmes complexes qui ne sont pas linéaires. Plusieurs fonctions d'activation sont utilisées pour introduire la non-linéarité dans différentes couches des CNN. Parmi les plus connues, on trouve le sigmoïde (logistique), la tangente hyperbolique et la fonction ReLU (Rectified Linear Unit).

- **RELU :**

Il s'agit d'une couche qui vise à améliorer l'efficacité du traitement en insérant une fonction mathématique, appelée fonction d'activation, entre les différentes couches de traitement. Une des fonctions d'activation couramment utilisées est la fonction ReLU (Rectified Linear Unit), définie comme F(x) = max(0, x)[51]. Cette fonction contraint les neurones à produire des valeurs positives. En général, les fonctions d'activation sont non linéaires. Leur rôle principal est de permettre aux réseaux de neurones d'apprendre des fonctions plus complexes que celles pouvant être obtenues par une simple régression linéaire, car la multiplication des poids dans une couche cachée représente essentiellement une transformation linéaire [54].

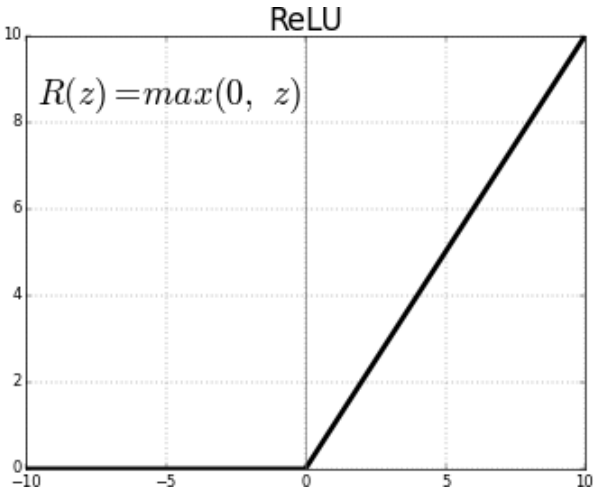

Figure 16: *Fonction d'activation de RELU [W10].*

La fonction Rectified Linear Unit (ReLU) est couramment appliquée après chaque opération de convolution, où toutes les valeurs de pixels négatives sont converties en zéro. L'objectif de ReLU est d'introduire de la non-linéarité dans notre CNN, car la plupart des caractéristiques du monde réel, lorsqu'elles sont appliquées à l'une des cartes d'entrée, produisent une carte de sortie également appelée carte de caractéristiques rectifiées. [55].

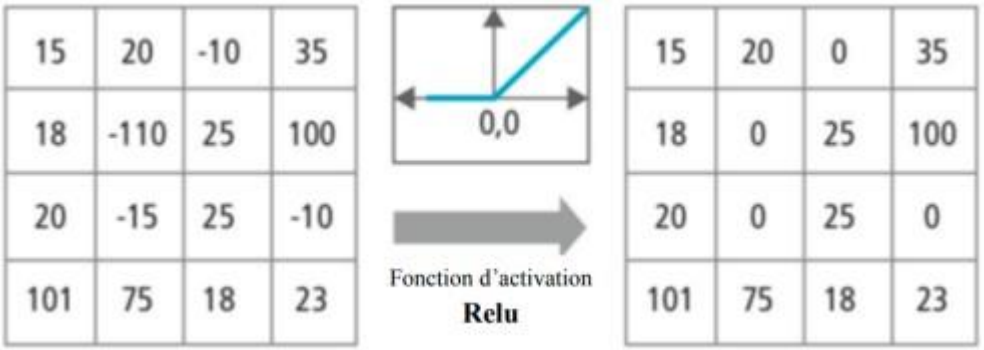

Figure 17: *Illustration du fonctionnement d'une couche ReLU. (Dans la case de gauche, tous les nombres négatifs ont été convertis en zéro après l'application de la fonction d'activation, tandis que toutes les autres valeurs sont restées inchangées).*

II.3.2.4. Couche entièrement connectée (Fully Connected Layer (FC))

Après les étapes de convolution et de pooling, le raisonnement de haut niveau dans le réseau neuronal se réalise grâce à des couches entièrement connectées. Dans les réseaux de neurones convolutifs, chaque couche agit comme un filtre pour détecter la présence de caractéristiques spécifiques ou de motifs dans les données d'origine. Les premières couches identifient des caractéristiques facilement reconnaissables et interprétables, tandis que les couches ultérieures découvrent des caractéristiques de plus en plus abstraites. La dernière couche du réseau convolutif est capable d'effectuer une classification ultra-précise en combinant toutes les caractéristiques spécifiques détectées par les couches précédentes dans les données d'entrée.

Les couches totalement connectées accomplissent des tâches similaires à celles des réseaux de neurones artificiels standard (ANN) en produisant des scores de classe à partir des activations, qui sont ensuite utilisés pour la classification. Il est également recommandé d'utiliser la fonction ReLU entre ces couches pour améliorer les performances. L'objectif de la couche entièrement connectée est de classer l'image d'entrée dans différentes catégories en fonction de l'ensemble de données d'apprentissage. [50].

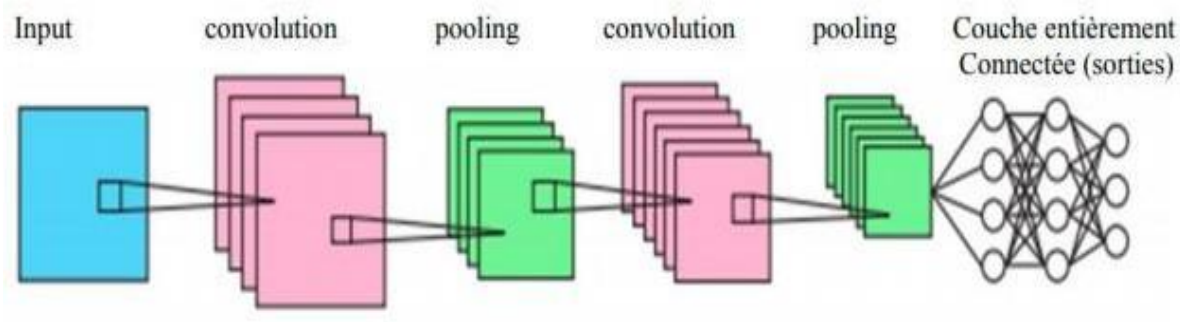

Figure 18: *Architecture standard d'un réseau convolutifs [W11]*

II.3.3. Outils d'optimisation des réseaux convolutifs

Les réseaux de neurones convolutifs (CNN) se démarquent des réseaux de neurones multicouches classiques en utilisant une variété de paramètres d'optimisation. Dans cette section, nous explorerons différentes méthodes visant à améliorer l'efficacité de l'optimisation des CNN.

II.3.3.1. La batch normalisation

Dans le contexte de l'apprentissage automatique, l'ordre dans lequel les échantillons sont présentés revêt une grande importance, car les paramètres du réseau sont mis à jour après chaque passage d'un échantillon d'apprentissage. Cependant, il n'existe pas de méthodes efficaces pour prédire à l'avance quel échantillon apportera le plus d'informations, à moins de réaliser une recherche exhaustive, ce qui est coûteux [38]. Une méthode simple consiste donc à mélanger l'ordre de présentation des échantillons après chaque passage de l'ensemble d'apprentissage. Ainsi, les échantillons successifs ne proviennent pas tous de la même classe, ce qui peut potentiellement apporter davantage d'informations [33].

Une technique plus récente, introduite en 2015 [38], vise à accélérer et à améliorer l'apprentissage des CNN. Elle repose sur l'observation que, pendant l'apprentissage, la distribution des données d'entrée aux différentes couches du réseau change à chaque itération. Cette variation constante induit une adaptation continue des paramètres du CNN à ces différentes distributions, ce qui rallonge le temps nécessaire à l'apprentissage. La batch normalization (normalisation par lots) est une idée novatrice qui consiste à normaliser les données d'entrée de chaque couche de manière à ce que leurs distributions aient une moyenne nulle et une variance unitaire. Pendant l'apprentissage, les couches de batch normalization [32] utilisent des paramètres (un facteur d'échelle et un biais) pour ajuster cette normalisation.

Ces paramètres permettent d'appliquer une transformation à la distribution normalisée. En d'autres termes, pendant l'apprentissage, si le réseau estime que la distribution normalisée n'est pas adaptée à une couche particulière, il apprend les paramètres nécessaires pour l'ajuster.

II.3.3.2. Les fonctions de perte

La fonction (ou couche) de perte détermine comment le réseau pénalise l'écart entre la prédiction du réseau et la vérité terrain. Plusieurs fonctions de perte (ou fonctions d'objectif) sont disponibles pour l'entraînement des réseaux de neurones, et le choix dépend de la tâche spécifique que le réseau doit accomplir, que ce soit la classification, la régression, etc. Voici une liste des fonctions de perte les plus couramment utilisées :

- La fonction de perte Softmax : Elle est employée avec une couche Softmax pour convertir les scores logistiques calculés dans la couche dense en une distribution de probabilité en sortie. Cette distribution n'est pas uniforme, mais plutôt exponentielle, ce qui signifie qu'elle peut accentuer les différences entre les probabilités attribuées aux différentes classes. Par exemple, elle peut rapprocher la probabilité d'un résultat de 1 tout en éloignant celle d'un autre résultat de 0.

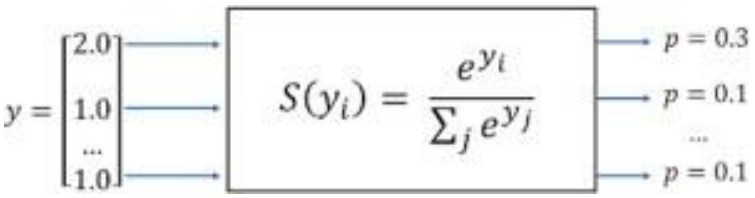

Figure 19: *La fonction de perte Softmax*

Ainsi, le modèle peut représenter la répartition de probabilité des expressions faciales..

- La fonction de perte par entropie croisée (sigmoïde) : Elle permet de réaliser une régression basée sur des probabilités.
- La perte quadratique (squared loss) : Cette fonction évalue les carrés des écarts entre la valeur prédite par un modèle pour un exemple étiqueté et la valeur réelle de l'étiquette. Une autre fonction associée à ce type de perte est l'erreur quadratique moyenne (Mean Squared Error (MSE)), qui se calcule en divisant la perte quadratique par le nombre d'exemples.

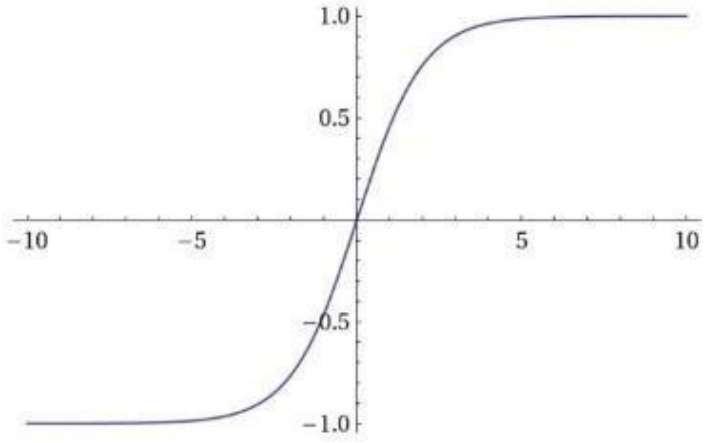

Figure 20: *fonctions d'activation SoftMax*

II.3.3.3. Méthodes de régularisation

Le nombre d'entrées et de sorties d'un réseau est généralement déterminé par les données d'apprentissage. Cependant, le nombre total de neurones dans les couches intermédiaires et le nombre de ces couches sont des paramètres qui doivent être ajustés pour éviter à la fois le surapprentissage (overfitting) et le sous-apprentissage (underfitting) [39].

Dans le domaine de l'apprentissage automatique, la régularisation est une technique visant à améliorer les performances de généralisation d'un algorithme d'apprentissage. En d'autres termes, elle cherche à réduire l'erreur sur de nouvelles données si elles suivent les mêmes tendances que les données d'apprentissage. La régularisation est introduite pour améliorer la capacité de généralisation sans perturber l'erreur d'apprentissage. Diverses méthodes de régularisation peuvent être envisagées [39] [40]:

- **Dropout :** Pour prévenir le surapprentissage, une technique appelée dropout a été introduite. Elle est utilisée pendant la phase d'apprentissage pour désactiver de manière aléatoire certains neurones lors des différentes itérations. Cette approche permet d'apprendre des paramètres plus généraux qui ne se concentrent pas uniquement sur les détails spécifiques de l'ensemble d'apprentissage. Une fois l'apprentissage terminé, tous les neurones sont réactivés [55].

- **Arrêt précoce (Early stopping) :** L'arrêt précoce consiste à entraîner le réseau en utilisant à la fois un ensemble d'entraînement et un ensemble de test, et à interrompre l'entraînement lorsque l'erreur sur l'ensemble de test commence à augmenter à nouveau.

- **Augmentation de données :** Cette méthode vise à augmenter la taille de l'ensemble d'apprentissage en ajoutant des données générées par des transformations (ajout de bruit, transformations géométriques, etc.) des données d'origine.

- **Régularisation L1 :** Cette forme de régularisation réduit spécifiquement le poids des entrées aléatoires et faibles tout en augmentant le poids des entrées considérées comme "importantes". Cela rend le système moins sensible au bruit [32].

- **Régularisation L2 (norme euclidienne) :** Cette méthode de régularisation réduit le poids des entrées fortes et encourage le neurone à accorder plus d'attention aux entrées de faible poids [32] [33].

II.3.4. Les architectures neuronales convolutifs

- **LeNet :** Les premières réussites des réseaux de neurones convolutifs ont été développées dans les années 1990 par Yann LeCun. L'architecture la plus célèbre de cette époque est LeNet, qui a été utilisée pour la reconnaissance de codes postaux, de chiffres, et d'autres applications.

- **AlexNet** : Le premier réseau convolutif qui a popularisé l'utilisation de ces réseaux dans la vision par ordinateur est AlexNet, développé par Alex Krizhevsky, Ilya Sutskever, et Geoff Hinton. En 2012, ce CNN a été soumis au défi d'ImageNet et a largement dépassé ses concurrents. AlexNet avait une architecture similaire à LeNet, mais il était plus profond, plus grand, et comportait des couches convolutives empilées les unes sur les autres (contrairement à l'approche précédente qui consistait à avoir une seule couche convolutive suivie immédiatement d'une couche de pooling).

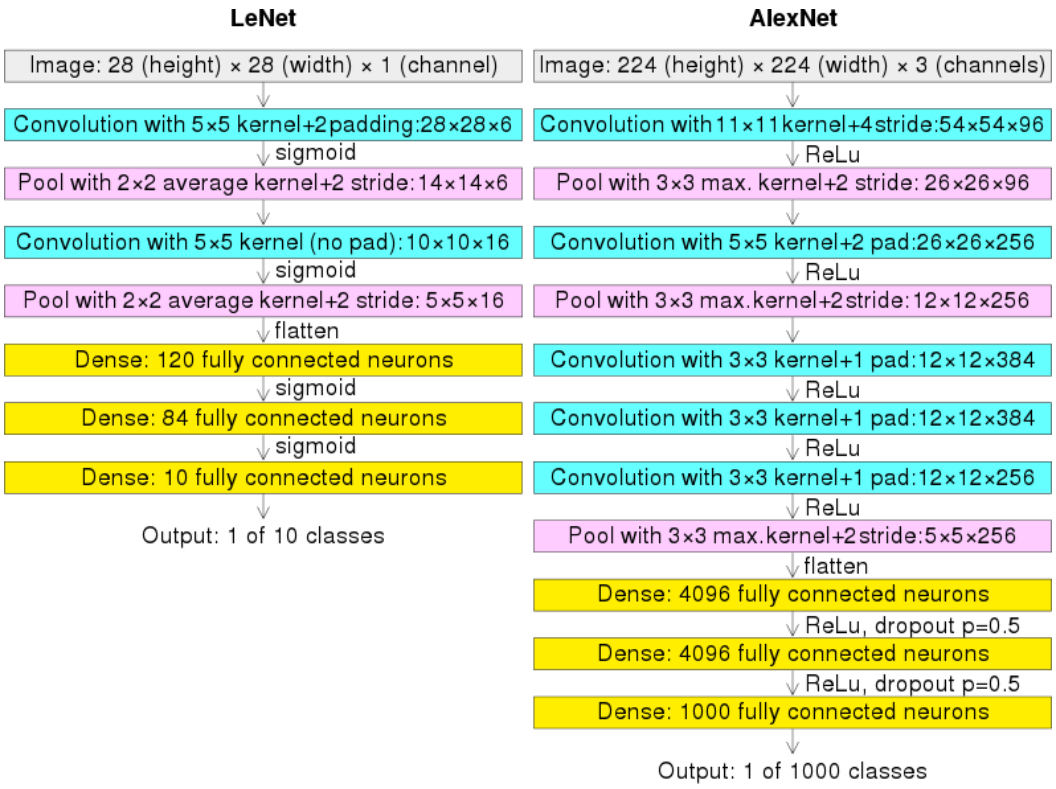

Figure 21: *LeNet et AlexNet Architectures*

- **ZFnet :** Cette architecture est une amélioration d'AlexNet qui ajuste les hyperparamètres, notamment en élargissant la taille des couches convolutionnelles et en réduisant la taille du noyau sur la première couche.

- **GoogLeNet :** Ce modèle, développé par Google, a introduit le concept de module Inception, qui a considérablement réduit le nombre de paramètres dans le réseau (passant de 60 millions pour AlexNet à 4 millions). De plus, ce module utilise le global average pooling, ce qui supprime de nombreux paramètres. Il existe plusieurs versions de GoogLeNet, dont Inception-v4 et Xception.

- **ResNet :** Les réseaux résiduels, développés par Kaiming He et son équipe, ont remporté la compétition ILSVRC en 2015. Ils se caractérisent par l'utilisation de sauts de connexion et une forte utilisation de la batch normalization. De plus, ils utilisent le global average pooling à la fin du réseau.

- **VGG Net :** Cette architecture a été développée par le Visual Geometry Group d'Oxford, dirigé par Andrea Vedaldi et Andrew Zisserman, en 2017.

- **MobileNet :** Cette classe de réseaux neuronaux convolutionnels profonds légers se distingue par leur taille réduite et leurs performances élevées par rapport à de nombreux autres modèles populaires. Ils utilisent des convolutions séparables en profondeur, ce qui signifie qu'ils effectuent une convolution distincte pour chaque canal de couleur plutôt que de les combiner en un seul [56].
- **MobileNetV2 :** Cette version améliorée de MobileNet, appelée MobileNetV2, conserve les avantages de la version précédente tout en ajoutant des blocs résiduels inversés avec des fonctionnalités de goulot d'étranglement. MobileNetV2 présente un nombre de paramètres considérablement inférieur à celui de MobileNet d'origine. Ces modèles sont capables de prendre en charge des tailles d'entrée d'image supérieures à 32 x 32, avec de meilleures performances pour des images plus grandes [57].

II.4. Contexte émotion- Deep Learning

Diverses approches ont été conçues pour la reconnaissance des expressions faciales, et elles peuvent être classées en deux catégories principales : les méthodes globales (ou holistiques) et les méthodes locales.

II.4.1. Méthodes globales

Ces approches s'appuient sur des méthodes d'analyse statistique bien établies. Elles ne nécessitent pas l'identification de points caractéristiques spécifiques du visage, tels que les centres des yeux, les narines ou le centre de la bouche, à l'exception de la normalisation des images. Dans ces méthodes, les images faciales, qui peuvent être considérées comme des matrices de valeurs de pixels, sont traitées de manière globale et transformées en vecteurs, ce qui facilite leur manipulation.

L'avantage principal des méthodes globales réside dans leur mise en œuvre relativement rapide et leur complexité de calcul modérée. Cependant, elles sont sensibles aux variations d'éclairage, de pose et d'expression faciale. En effet, la moindre modification des conditions environnementales entraîne des variations inévitables dans les valeurs des pixels traitées directement. Ces méthodes se concentrent principalement sur l'analyse de sous-espaces faciaux, ce qui a considérablement contribué aux avancées de la technologie de reconnaissance faciale.

Il existe deux types de techniques parmi les méthodes globales : les techniques linéaires et les techniques non linéaires. Les techniques linéaires projettent les données depuis un espace de grande dimension, comme l'espace de l'image originale, vers un sous-espace de dimension inférieure. Cependant, elles sont limitées dans leur capacité à capturer les variations non convexes des caractéristiques géométriques des visages, ce qui complique la distinction entre différentes formes et individus.

La technique linéaire la plus connue est l'Analyse en Composantes Principales (PCA), également appelée Transformée de Karhunen-Loève. Initialement utilisée pour représenter efficacement les images de visages humains, ces méthodes globales linéaires basées sur l'apparence ont évité les instabilités des premières méthodes géométriques. Cependant, elles ne sont pas suffisamment précises pour décrire les subtilités des variations géométriques présentes dans l'espace des images d'origine.

Cela s'explique par leur incapacité à gérer la non-linéarité inhérente à la reconnaissance faciale, où les déformations des variétés non linéaires peuvent être atténuées et les concavités remplies, entraînant des résultats peu satisfaisants. Pour résoudre ce problème de non-linéarité dans la reconnaissance des expressions faciales, ces méthodes linéaires ont été étendues à des techniques non linéaires basées sur le concept mathématique de noyau (kernel), telles que le Kernel PCA et le Kernel LDA [58].

II.4.2. Méthodes locales :

Ces méthodes sont basées sur des modèles qui intègrent des connaissances a priori sur la morphologie faciale et reposent généralement sur l'identification de points caractéristiques du visage. Par exemple, Kanade a présenté l'un des premiers algorithmes de ce type en détectant des points ou des traits distinctifs du visage, puis en les comparant à des paramètres extraits d'autres visages. Ces approches offrent une manière différente de traiter la non-linéarité en construisant un espace de caractéristiques local et en appliquant des filtres d'image appropriés pour atténuer l'impact des variations.

Pour atteindre cet objectif, diverses méthodes ont été utilisées, notamment les approches bayésiennes, les machines à vecteurs de support (SVM), la méthode des modèles actifs d'apparence (AAM), ainsi que la méthode "local binary pattern" (LBP). Toutes ces techniques présentent l'avantage de pouvoir modéliser plus efficacement les variations de pose, d'éclairage et d'expression par rapport aux méthodes globales. Cependant, elles requièrent généralement un effort plus important, car un nombre considérable de points caractéristiques doivent souvent être placés manuellement sur le visage. En revanche, les méthodes globales se limitent à la connaissance de la position des yeux pour normaliser les images, une tâche pouvant être effectuée de manière automatique et relativement fiable grâce à des algorithmes de détection[59].

II.4.3. Méthodes hybrides :

Ces approches combinent à la fois des éléments des méthodes globales et locales en associant la détection de caractéristiques géométriques (ou structurales) à l'extraction de caractéristiques locales d'apparence. Cette combinaison vise à renforcer la stabilité des performances de reconnaissance face aux variations de pose, d'éclairage et d'expressions faciales [60].

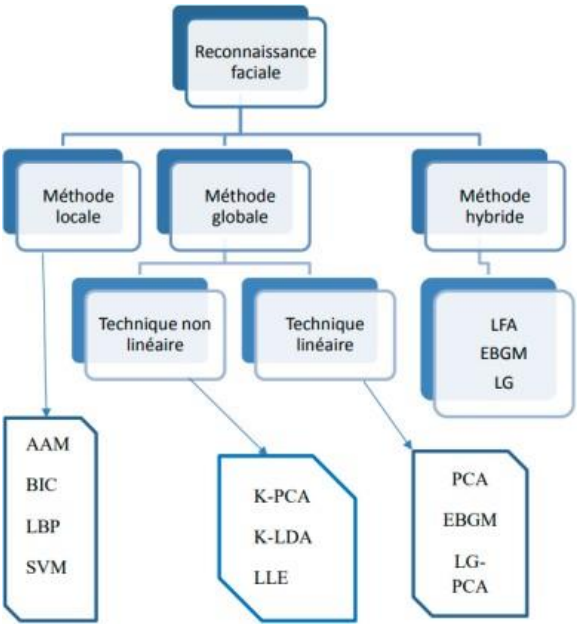

Figure 22: *Classification des algorithmes principaux utilisés en reconnaissance faciale*

II.4.4. La classification

La phase finale d'un système de reconnaissance automatique des émotions implique la classification des émotions en fonction des caractéristiques faciales extraites, se déclinant en deux catégories distinctes : la reconnaissance d'expressions à partir d'images fixes et la reconnaissance d'expressions dans des séquences vidéo. Diverses méthodes de classification ont été utilisées pour la reconnaissance des émotions, dont [61].

• Réseaux de neurones (Neural Networks, NN)."

II.5. Applications possibles et les avantages de la reconnaissance d'émotions

La détection et la classification des expressions faciales émotionnelles ont une multitude d'applications potentielles dans divers aspects de notre vie quotidienne. Voici une liste non exhaustive des domaines susceptibles de bénéficier de cette capacité :

- **Interaction homme-machine :** Les expressions faciales sont l'un des moyens de communication parmi d'autres, tels que le langage vocal. La détection des émotions est une compétence naturelle pour les humains, mais elle représente un défi de taille pour les machines.

- **Interaction homme-robot :** Les robots sociaux doivent également être capables de reconnaître différentes expressions faciales et d'agir en conséquence pour établir des interactions efficaces. Pour parvenir à une telle interaction homme-robot, il est essentiel que le robot comprenne les expressions faciales des humains.
- **Neuromarketing :** La mesure automatisée des préférences des consommateurs à partir de leurs expressions faciales en réponse à des publicités produits pourrait avoir un impact significatif sur les études de marché. Cela permettrait aux entreprises de mieux comprendre les consommateurs et de proposer de nouveaux produits et services adaptés à leurs besoins. Des travaux ont déjà montré qu'il était possible de déterminer si les individus appréciaient certaines publicités en analysant leurs réactions faciales.
- **Psychologie :** La détection des expressions faciales revêt une importance considérable dans l'analyse de la psychologie humaine. La reconnaissance de l'incapacité d'une personne à exprimer certaines expressions faciales peut contribuer au diagnostic précoce de troubles psychologiques.

Ces applications démontrent la pertinence et le potentiel des systèmes de reconnaissance des expressions faciales émotionnelles dans divers domaines de la vie quotidienne.

II.6. SVM (Support vector machine)

Les machines à vecteurs de support, également connues sous le nom de séparateurs à vaste marge ou SVM (Support Vector Machines) sont une famille de techniques d'apprentissage supervisé. Elles ont été développées dans les années 1990, basées sur les travaux théoriques de Vladimir Vapnik concernant une théorie statistique de l'apprentissage, connue sous le nom de théorie de Vapnik-Chervonenkis. Les SVM sont une généralisation des classifieurs linéaires et se sont rapidement imposées grâce à leurs caractéristiques uniques.

Les SVM sont appréciées pour leur capacité à traiter des données de grande dimension, leur faible nombre d'hyperparamètres, leurs garanties théoriques solides, ainsi que leurs performances pratiques impressionnantes. Elles ont trouvé des applications dans de nombreux domaines, notamment la bio-informatique, la recherche d'information, la vision par ordinateur, la finance, et bien d'autres encore.

Dans certaines situations, les performances des machines à vecteurs de support sont comparables, voire supérieures, à celles des réseaux de neurones ou des modèles de mélanges gaussiens. Cette polyvalence et cette efficacité ont contribué à faire des SVM un outil précieux

dans le domaine de l'apprentissage automatique[60].

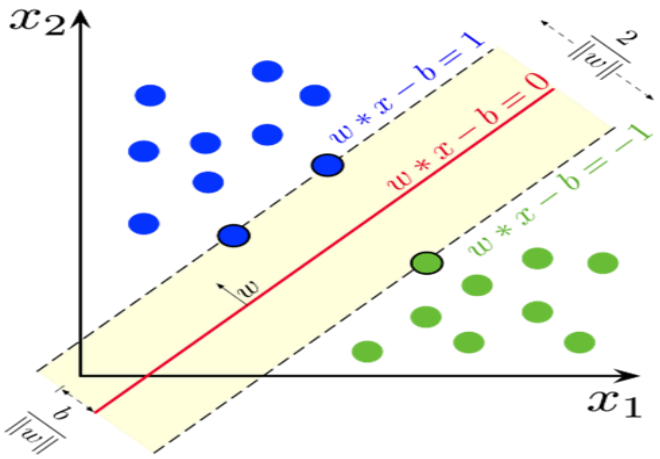

Figure 23: *L'algorithme SVM (Support Vector Machine) fonctionne [W14]*

II.7. Matrice de confusion et matrice d'évaluation des modèles de deep learning

La matrice de confusion fournit des valeurs pour les quatre combinaisons de valeurs vraies et prédites, vrai positif (TP), vrai négatif (TN), faux positif (FP) et faux négatif (FN). La précision, le rappel et le score F sont calculés à l'aide de TP, FP, TN, FN. TP est la prédiction correcte d'une émotion, FP est la prédiction incorrecte d'une émotion, TN est la prédiction correcte d'une émotion incorrecte et FN est la prédiction incorrecte d'une émotion incorrecte. Prenons une image de la classe heureuse. La matrice de confusion pour cet exemple est illustrée. La section rouge a la valeur TP car l'image Angry est prédite comme étant Angry.

La section bleue a des valeurs FP car l'image est prévue pour être Dégoût, Peur, Heureux, Triste,Surprise ou Neutre. La section jaune a des valeurs TN car l'image n'est pas Dégoût, Peur, Heureux, Triste, Surprise ou Neutre, mais le modèle l'a prédit. La section verte a des valeurs FN car l'image n'est pas en colère mais a été prédite comme étant en colère. [18]

II.7.1. Rappel

Le rappel nous donne le taux de vrais positifs (TPR), qui est le rapport des vrais positifs à tout ce qui est positif.

Le rappel est donné par :

$$Recall = \frac{TP}{TP + FN}$$

II.7.2. Précision

Lorsque nous avons un déséquilibre de classe, la précision peut devenir une mesure peu fiable pour mesurer nos performances.

La précision est donnée par :

$$Precision = \frac{TP}{TP + FP}$$

II.7.3. Score F1

Le score F1 est la moyenne harmonique de la précision et du rappel, où un score F1 atteint sa meilleure valeur à 1 (précision et rappel parfaits) et la pire à 0.

Le score F est la moyenne consonantique du rappel et de la précision et est donné par :

$$F - Score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

II.8. Conclusion

Dans cette partie, Dans cette section, nous avons exploré en profondeur le concept de l'émotion, en mettant en lumière les concepts essentiels qui lui sont associés, tels que l'expression faciale et les types d'émotions de base. De plus, nous avons plongé dans les avancées de la reconnaissance faciale, en mettant en évidence l'approche novatrice du Deep Learning et en la comparant aux méthodes traditionnelles d'apprentissage automatique. Nous avons également examiné en détail les trois familles principales de modèles : les réseaux convolutifs (CNN), les réseaux récurrents et les modèles génératifs .

CHAPITRE III

IMPLEMENTATION ET RESULTATS EXPERIMENTAUX ET ETUDE COMPARATIVE

Chapitre III : Implémentation et résultats expérimentaux et étude comparative.

III.1. Introduction

L'objectif de ce chapitre est de vous guider à travers les étapes de mise en œuvre de notre approche au sein d'un système de reconnaissance des émotions. Nous débuterons en présentant les ressources, le langage de programmation, et l'environnement de développement que nous avons employés. Ensuite, nous explorerons en détail les différentes étapes de la création de notre modèle. Pour conclure, nous partagerons les résultats des tests que nous avons réalisés, y compris une comparaison avec d'autres modèles tels que RES-Net, VGG et CNN-SVM, dans le cadre de notre rapport sur les performances de l'année précédente.

III.1.1. Architecteur globale

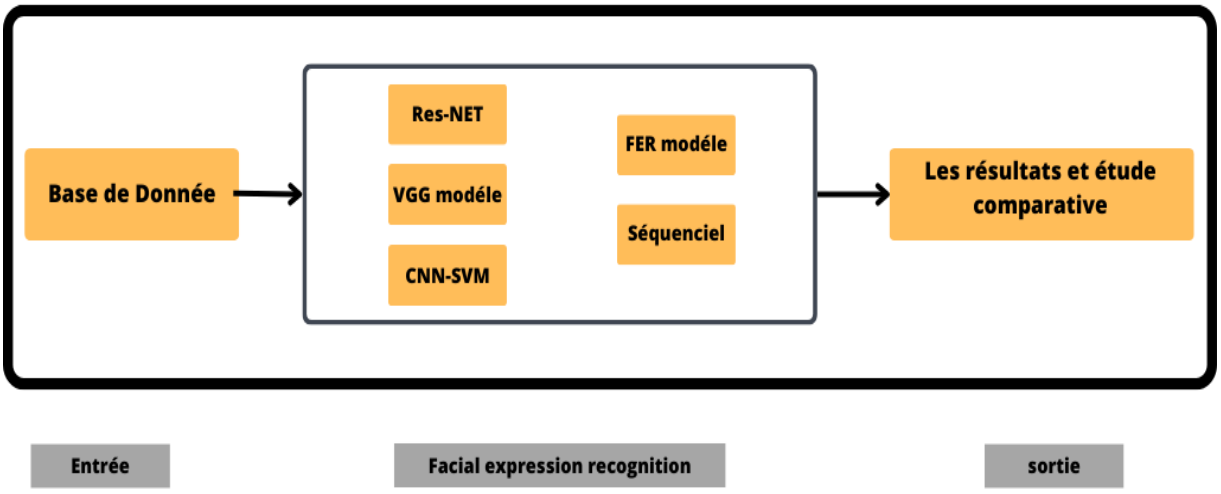

Figure 24: *Architecture global d'un CNN.*

III. 2. Environnement de développement

III.2.1. Python

Python, conçu par Guido van Rossum et introduit en 1991, se distingue comme un langage de programmation de premier plan utilisé pour un large éventail d'applications. Fondé sur une philosophie de conception qui met l'accent sur la clarté du code grâce à l'utilisation d'une mise en forme soignée, Python offre un ensemble de fonctionnalités qui simplifient la programmation, qu'il s'agisse de projets de petite ou de grande envergure. Il se caractérise par son typage dynamique et sa gestion automatique de la mémoire. De plus,

Python offre une flexibilité impressionnante en prenant en charge divers styles de programmation, notamment l'orienté objet, l'impératif, le fonctionnel et le procédural. En outre, Python est accompagné d'une bibliothèque standard étendue et complète qui facilite la réalisation de nombreuses tâches. Il convient également de noter que Python est compatible avec de multiples systèmes d'exploitation, ce qui en fait un choix polyvalent pour les développeurs [W18].

III.2.2. Google Colab

Google Colab, ou Colaboratory, se révèle être un service en ligne gratuit proposé par Google. Il repose sur la technologie du cahier Jupyter et a pour objectif de soutenir la formation et la recherche dans le domaine de l'apprentissage automatique. Ce service vous permet de former des modèles d'apprentissage automatique directement dans le cloud, éliminant ainsi la nécessité d'installer des logiciels supplémentaires sur votre ordinateur, à l'exception d'un simple navigateur. C'est une solution pratique, n'est-ce pas ? Mais avant de plonger dans les détails de ce service exceptionnel, rappelons brièvement ce qu'est un cahier Jupyte [W20].

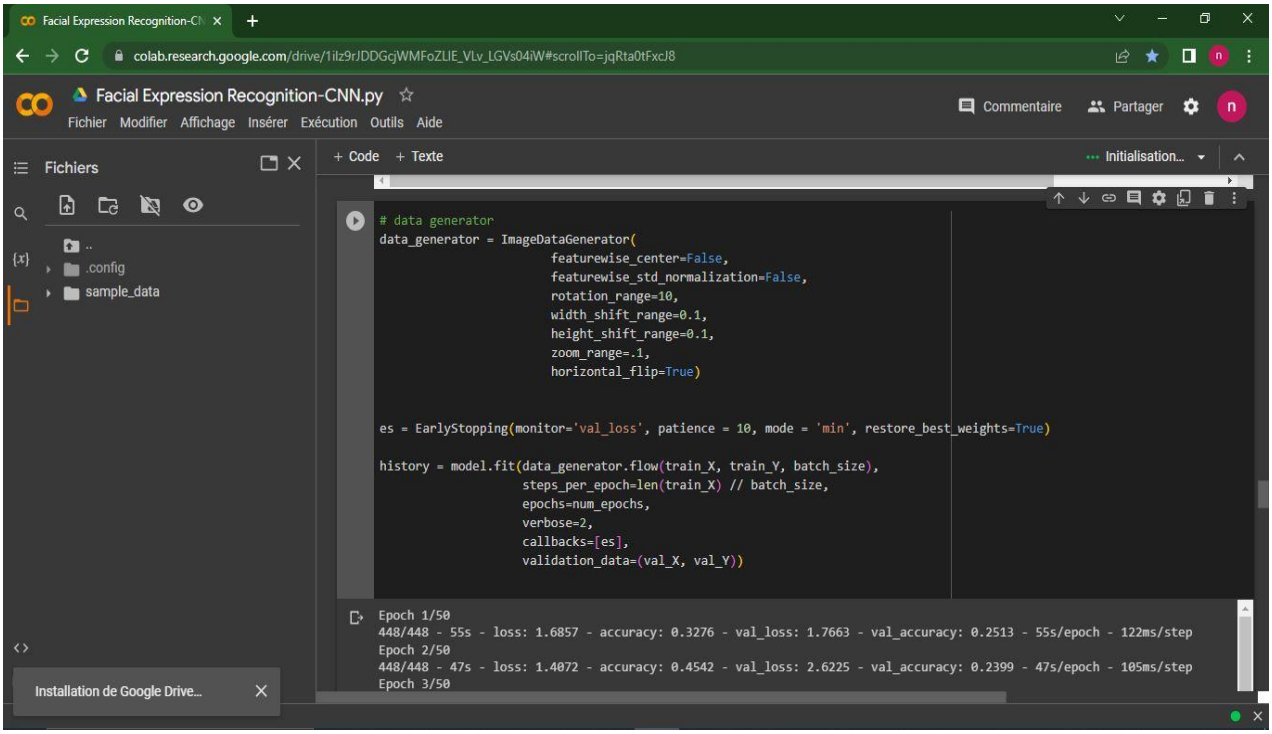

Figure 25: *Google colab*

III.3. Bibliothèques utilisées

III.3.1. OpenCV (Open Source Computer Vision Library)

Est une bibliothèque proposant un ensemble de plus de 2500 algorithmes de vision par ordinateur spécialisé dans le traitement d'images, accessible au travers d'API pour les langages C, C++, et Python. Elle est distribuée sous une licence BSD (libre) pour les plateformes Windows,

GNU/Linux, Android et MacOS [W21], nous avons utilisé cette bibliothèquepour la détection du visage à partir des images introduites.

III.3.2. Numpy

NumPy est un atout essentiel pour exécuter des calculs numériques avec Python. Cette bibliothèque simplifie la manipulation de tableaux de nombres, propose des fonctionnalités avancées telles que la diffusion, et offre même la possibilité d'intégrer du code en langage C, C++, et Fortran pour une performance accrue [W22].

III.3.3. Keras

Keras est un précieux outil open source, développé en Python, qui facilite l'interaction avec les puissants algorithmes de réseaux de neurones profonds et d'apprentissage automatique, notamment avec des frameworks tels que TensorFlow et Theano. À l'origine, cette bibliothèque a été créée par François Chollet pour simplifier ces tâches complexes. [W23].

III.3.4. Pandas

Pandas, une bibliothèque open source avec une licence BSD, offre des structures de données performantes et conviviales ainsi que des outils d'analyse de données pour le langage de programmation Python. Il est important de noter que le projet Pandas est soutenu par NumFOCUS, ce qui favorise son développement en tant que projet open source de renommée mondiale et permet d'apporter des contributions essentielles au projet. [W24].

III.3.5. Matplotlib

Matplotlib se présente comme une bibliothèque de création de graphiques conçue pour le langage de programmation Python, en collaboration avec son extension mathématique numérique NumPy. Elle offre une interface orientée objet qui facilite l'intégration de graphiques dans diverses applications, en s'appuyant sur des kits d'outils d'interface graphique polyvalents tels que Tkinter, Python, Qt ou GTK+.

III.3.6. TensorFlow (GPU version 2.7.0)

Le groupe Google Brain, composé de scientifiques et d'architectes, est à l'origine de TensorFlow, qui se distingue comme la bibliothèque de programmation la plus répandue dans le domaine de l'apprentissage profond. Ce qui a particulièrement contribué à la renommée de TensorFlow, c'est sa polyvalence en matière de création de modèles d'apprentissage profond, avec un support étendu pour différents langages tels que Python, C++ et R.

À un niveau plus fondamental, TensorFlow peut être considéré comme une bibliothèque Python qui permet aux utilisateurs de définir des calculs complexes sous forme de graphique de flux de données[W25].

III.4. Implémentation de Base de données FER2013

L'ensemble de données est **FER2013** [W26] téléchargé depuis Kaggle etil se compose d'images de visages en niveaux de gris de 48x48 pixels. Les visages ont été automatiquement enregistrés afin que le visage soit plus ou moins centré et occupe à peu près la même quantité d'espace dans chaque image.

La tâche consiste à classer chaque visage en fonction de l'émotion montrée dans l'expression faciale dans l'une des sept catégories (0 = en colère, 1 = dégoût, 2 = peur, 3 = heureux, 4 = triste, 5 = surprise, 6 = neutre).

Elle contient 35887 images expressions faciales.

l'ensemble d'apprentissage compose de 28 709 exemples et l'ensemble de test public se compose de 7 178 exemples[65].

Emotion	**Numbers**
Angry	**4953**
Fear	**5121**
Sad	**6077**
Neutral	**6198**
Happy	**8989**
Surprise	**4002**
Digest	**547**

Tableau 2: *FER2013 Dataset par émotion*

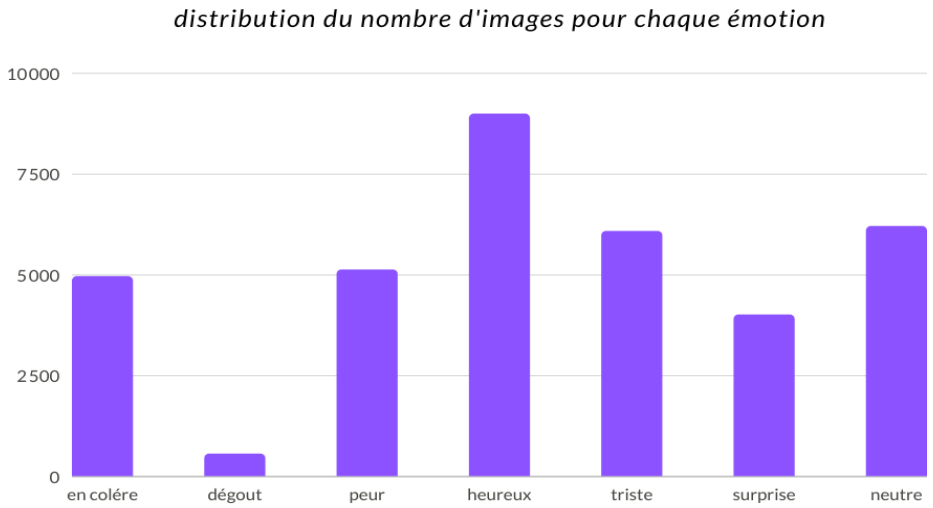

Figure 26: *Répartition de la base de données FER2013 par émotion*

Pour les valeurs d'usage on a 80% training, 10% validation et 10% pour le test.

Entrainment	**Test Public**	**Test Privée**
28709	**3589**	**3589**

Tableau 3: *FER2013 dataset*

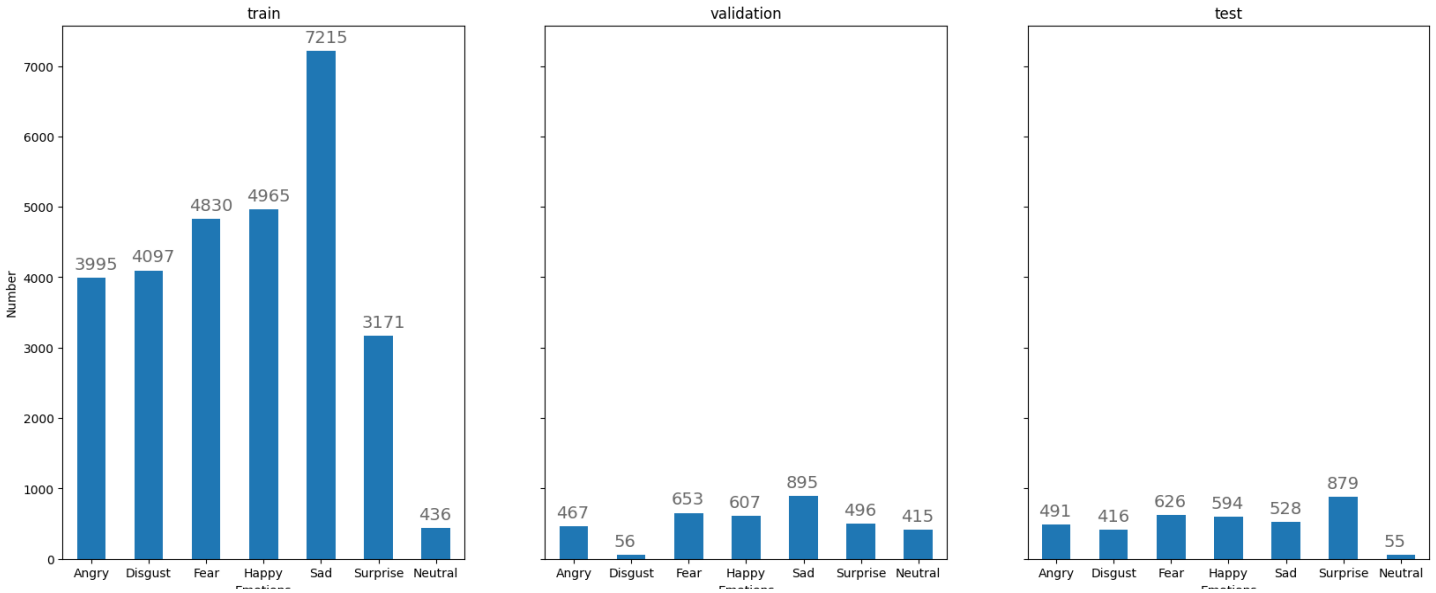

Figure 27: *Data set FER2013.*

III.5. Implémentation et Résultats des expériences

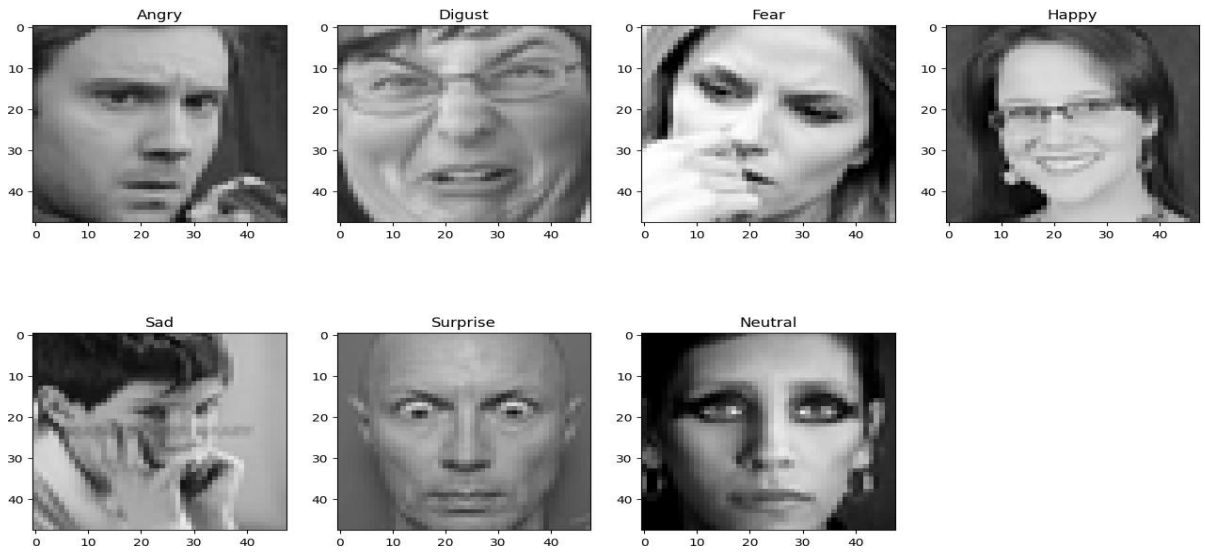

Figure 28: *Échantillons de FER2013 Émotions [W27]*

Cette section présente les expériences effectuées sur notre système de reconnaissance des émotions faciales, en utilisant quatre modèles différents, à savoir le modèle Séquentiel, Fer-modèle, Res-NET modèle et VGG modèle**.** L'objectif était d'évaluer la performance de l'information obtenue à partir de ces modèles.

III.5.1. Modèle Séquentiel

Dans cette partie, nous discuterons du résultat de l'apprentissage de tous lesmodèles précédents sur le jeu de données FER2013.

Dans notre modèle, nous avons atteint 0,9265 de perte et 66-67 % de précision, voirles figures 20 et 21.

III.5.1.1. Le code source en python :

```
model = Sequential()

#module 1

model.add(Conv2D(2*2*num_features, kernel_size=(3, 3), input_shape=(width, height, 1), data_format='channels_last'))

model.add(BatchNormalization())

model.add(Activation('relu'))

model.add(Conv2D(2*2*num_features, kernel_size=(3, 3), padding='same'))

model.add(BatchNormalization())

model.add(Activation('relu'))
```

```
model.add(MaxPooling2D(pool_size=(2, 2), strides=(2, 2)))
#module 2
model.add(Conv2D(2*num_features, kernel_size=(3, 3), padding='same'))
model.add(BatchNormalization())
model.add(Activation('relu'))
model.add(Conv2D(2*num_features, kernel_size=(3, 3), padding='same'))
model.add(BatchNormalization())
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2), strides=(2, 2)))
#module 3
model.add(Conv2D(num_features, kernel_size=(3, 3), padding='same'))
model.add(BatchNormalization())
model.add(Activation('relu'))
model.add(Conv2D(num_features, kernel_size=(3, 3), padding='same'))
model.add(BatchNormalization())
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2), strides=(2, 2)))
#flatten
model.add(Flatten())
#dense 1
model.add(Dense(2*2*2*num_features))
model.add(BatchNormalization())
model.add(Activation('relu'))
#dense 2
model.add(Dense(2*2*num_features))
model.add(BatchNormalization())
model.add(Activation('relu'))
#dense 3
model.add(Dense(2*num_features))
```

```
model.add(BatchNormalization())

model.add(Activation('relu'))

#output layer

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

model.compile(loss='categorical_crossentropy',

              optimizer=Adam(lr=0.001, beta_1=0.9, beta_2=0.999, epsilon=1e-7),

              metrics=['accuracy'])

import matplotlib.pyplot as plt

def plot_model_graph(model):

    plt.figure(figsize=(10, 6))

    tf.keras.utils.plot_model(model, to_file='model_graph.png', show_shapes=True)

    plt.show()

plot_model_graph(model)

model.summary()
```

III.5.1.2. Architectures de modèle séquentiel

Notre architecture finale (modèle séquentiel) avait une précision de test d'environ **67-66 %**

L'architecture est une combinaison de ces 3 blocs :

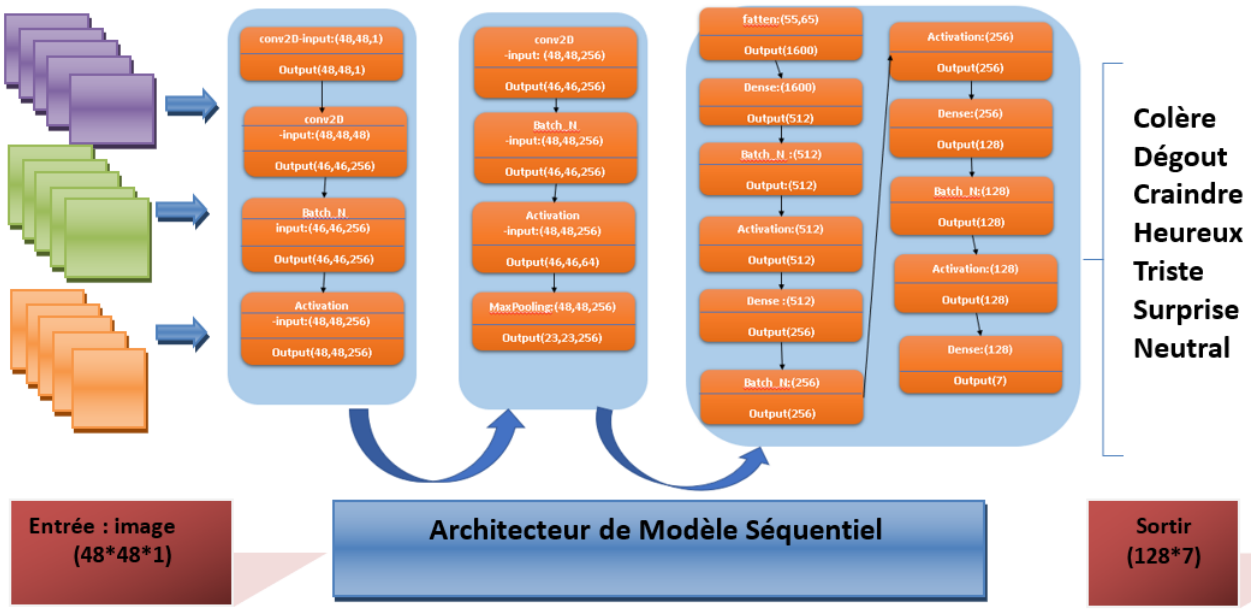

Figure 29: *Architectures de modèle séquentiel*

Se fier exclusivement à la précision et à la perte d'un modèle entraîné ne permet pas toujours d'obtenir une vision exhaustive de ses performances.

III.5.1.3. Entraînement

Dans cette section, nous examinons les résultats obtenus grâce à l'approche que nous avons développée. Une fois que le modèle est entraîné, nous allons explorer certains de ses paramètres:

Paramètres	**Utilisation**
Epoch	Désigne le nombre d'itération dans notre basede données.
Loss	Désigne le taux d'erreur.
Accuracy	Désigne le taux de précision.

Tableau 4: *les paramètres et ses utilisations*

III.5.1.4. Résultats expérimentaux et analyse de performance :

Total params	**Trainable params**	**Non-trainable params**
2,137,991	**2,134,407**	**3,584**

Tableau 5: *Taille des paramètres du modèle Séquentiel*

Après 210/210 Epoch :

Epoch	**Loss**	**Accuracy**	**Val_loss**	**Val_accuracy**
210/210	**1.298**	**0.5024**	**0.8450**	**0.6601**

Tableau 6: *Résultats de Loss, Accuracy, Val_Loss et Val_Accuracy pour le modèle Séquentiel*

Dans cette partie, nous présentons les résultats obtenus lors de nos expérimentations sur la reconnaissance des émotions faciales. Nous allons visualiser la précision (accuracy) ainsi que le taux d'erreur (loss) du modèle séquentiel :

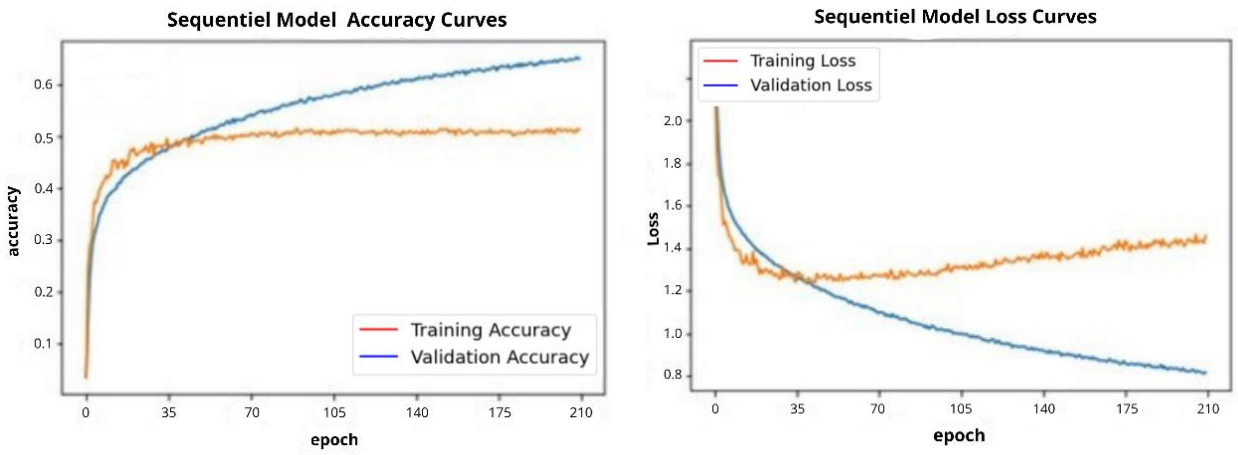

Figure 30: *Les courbes de précision et de perte pour le modèle Séquentiel*

Des métriques plus avancées sont à notre disposition, telles que le score F1 que nous avons choisi d'adopter.

Le score F1 repose sur deux métriques préalablement calculées : la précision et le rappel. Ces mesures se basent sur les prédictions de vrais positifs, de faux positifs et de faux négatifs, dont la compréhension est facilitée par l'utilisation de la matrice de confusion.

Comme notre modèle vise à reconnaître les 7 émotions faciales universelles, tandis que l'ensemble de données FER2013 incluait une 8e classe pour les émotions de « mépris », nous avons choisi d'incorporer tous les exemples de la classe « mépris » à la classe « Happy » au lieu de les exclure.

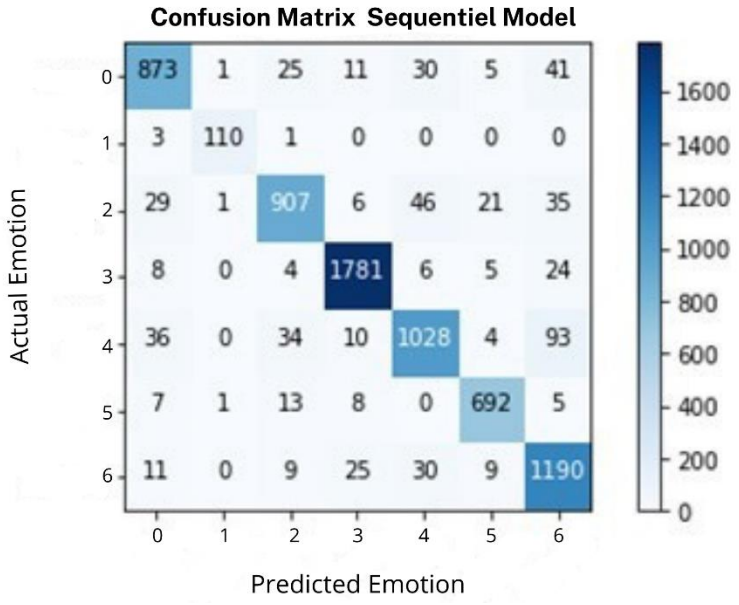

Figure 31: *Matrice de confusion pour le modèle Séquentiel*

III.5.2. Fer-modèle

Toujours avec le même data set « Fer2013.cvs » et changer le modèle séquentiel

Avec le Fer-modèle.

III.5.2.1. Le code source en python

```
def FER_Model(input_shape=(48,48,1)):

    # first input model

    visible = Input(shape=input_shape, name='input')

    num_classes = 7

    #the 1-st block

    conv1_1 = Conv2D(64, kernel_size=3, activation='relu', padding='same', name = 'conv1_1')(visible)

    conv1_1 = BatchNormalization()(conv1_1)

    conv1_2 = Conv2D(64, kernel_size=3, activation='relu', padding='same', name = 'conv1_2')(conv1_1)

    conv1_2 = BatchNormalization()(conv1_2)

    pool1_1 = MaxPooling2D(pool_size=(2,2), name = 'pool1_1')(conv1_2)

    drop1_1 = Dropout(0.3, name = 'drop1_1')(pool1_1)#the 2-nd block

    conv2_1 = Conv2D(128, kernel_size=3, activation='relu', padding='same', name = 'conv2_1')(drop1_1)

    conv2_1 = BatchNormalization()(conv2_1)

    conv2_2 = Conv2D(128, kernel_size=3, activation='relu', padding='same', name = 'conv2_2')(conv2_1)

    conv2_2 = BatchNormalization()(conv2_2)

    conv2_3 = Conv2D(128, kernel_size=3, activation='relu', padding='same', name = 'conv2_3')(conv2_2)

    conv2_2 = BatchNormalization()(conv2_3)

    pool2_1 = MaxPooling2D(pool_size=(2,2), name = 'pool2_1')(conv2_3)

    drop2_1 = Dropout(0.3, name = 'drop2_1')(pool2_1)#the 3-rd block

    conv3_1 = Conv2D(256, kernel_size=3, activation='relu', padding='same', name = 'conv3_1')(drop2_1)

    conv3_1 = BatchNormalization()(conv3_1)

    conv3_2 = Conv2D(256, kernel_size=3, activation='relu', padding='same', name = 'conv3_2')(conv3_1)

    conv3_2 = BatchNormalization()(conv3_2)

    conv3_3 = Conv2D(256, kernel_size=3, activation='relu', padding='same', name = 'conv3_3')(conv3_2)

    conv3_3 = BatchNormalization()(conv3_3)
```

```
conv3_4 = Conv2D(256, kernel_size=3, activation='relu', padding='same', name = 'conv3_4')(conv3_3)

conv3_4 = BatchNormalization()(conv3_4)

pool3_1 = MaxPooling2D(pool_size=(2,2), name = 'pool3_1')(conv3_4)

drop3_1 = Dropout(0.3, name = 'drop3_1')(pool3_1)#the 4-th block

conv4_1 = Conv2D(256, kernel_size=3, activation='relu', padding='same', name = 'conv4_1')(drop3_1)

conv4_1 = BatchNormalization()(conv4_1)

conv4_2 = Conv2D(256, kernel_size=3, activation='relu', padding='same', name = 'conv4_2')(conv4_1)

conv4_2 = BatchNormalization()(conv4_2)

conv4_3 = Conv2D(256, kernel_size=3, activation='relu', padding='same', name = 'conv4_3')(conv4_2)

conv4_3 = BatchNormalization()(conv4_3)

conv4_4 = Conv2D(256, kernel_size=3, activation='relu', padding='same', name = 'conv4_4')(conv4_3)

conv4_4 = BatchNormalization()(conv4_4)

pool4_1 = MaxPooling2D(pool_size=(2,2), name = 'pool4_1')(conv4_4)

drop4_1 = Dropout(0.3, name = 'drop4_1')(pool4_1)

#the 5-th block

conv5_1 = Conv2MM D(512, kernel_size=3, activation='relu', padding='same', name = 'conv5_1')(drop4_1)

conv5_1 = BatchNormalization()(conv5_1)

conv5_2 = Conv2D(512, kernel_size=3, activation='relu', padding='same', name = 'conv5_2')(conv5_1)

conv5_2 = BatchNormalization()(conv5_2)

conv5_3 = Conv2D(512, kernel_size=3, activation='relu', padding='same', name = 'conv5_3')(conv5_2)

conv5_3 = BatchNormalization()(conv5_3)

conv5_4 = Conv2D(512, kernel_size=3, activation='relu', padding='same', name = 'conv5_4')(conv5_3)

conv5_3 = BatchNormalization()(conv5_3)

pool5_1 = MaxPooling2D(pool_size=(2,2), name = 'pool5_1')(conv5_4)

drop5_1 = Dropout(0.3, name = 'drop5_1')(pool5_1)#Flatten and output

flatten = Flatten(name = 'flatten')(drop5_1)

ouput = Dense(num_classes, activation='softmax', name = 'output')(flatten)# create model

model = Model(inputs =visible, outputs = ouput)

# summary layers
```

```
    print(model.summary())

    return model

model = FER_Model()
```

III.5.2.2. Architecteur de Fer-modèle

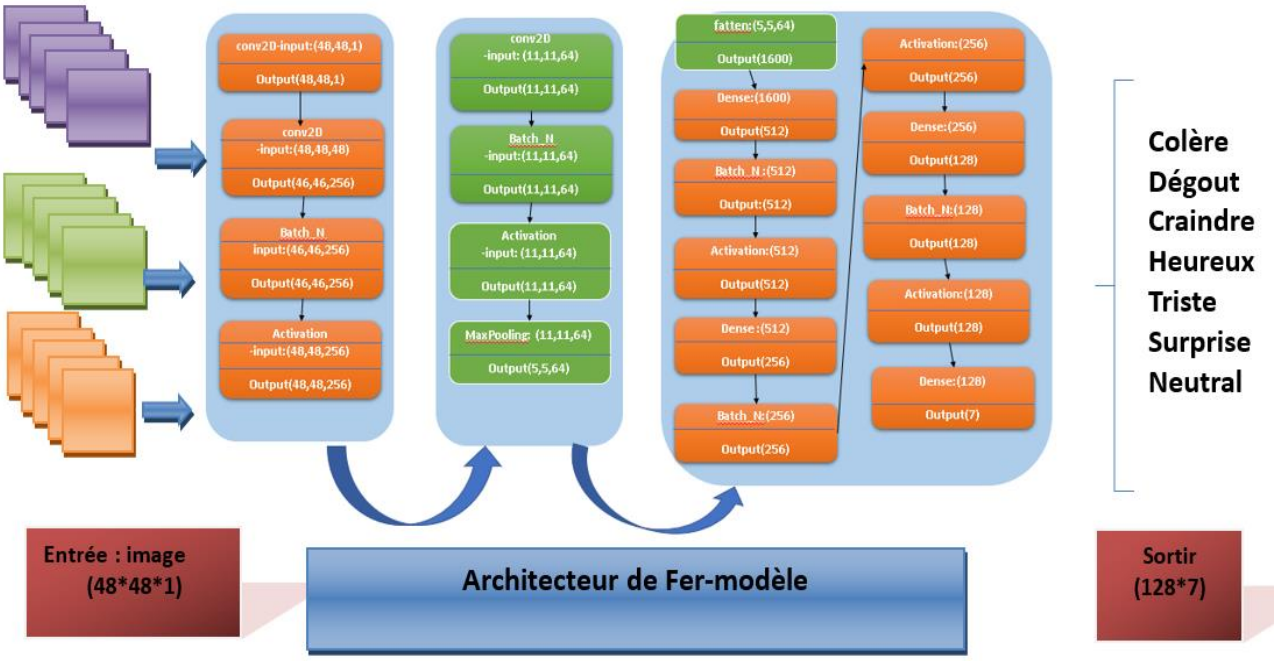

Figure 32: *Architecteur de Fer-modèle*

III.5.2.3. Résultats expérimentaux et analyse de performance

Cependant, dépendre uniquement de la précision (accuracy) et de la perte (loss) du modèle entraîné ne donne pas toujours une compréhension complète des performances du modèle.

Total params	Trainable params	Non-trainable params
13,111,367	13,103,431	7,936

Tableau 7: *Taille des paramètres du modèle FER*

Après avoir compilé le modèle, nous ajustons ensuite les données pour l'entraînement et

La validation. Ici, nous prenons la taille du batch à 64 avec 210 époques () :

Et 210/210 d'Epoch :

Epoch	Loss	Accuracy	Val_loss	Val_accuracy
210/210	**0.8281**	**0.7296**	**0.8450**	**0.6980**

Tableau 8: *Résultats de Loss, Accuracy, Val_Loss et Val_Accuracy pour le modèle FER*

Cette section décrit les résultats expérimentaux de notre système sur la reconnaissance des émotions faciales, Tracer la précision « « accuracy » et Désigne le taux d'erreur « loss » du Fer-modèle :

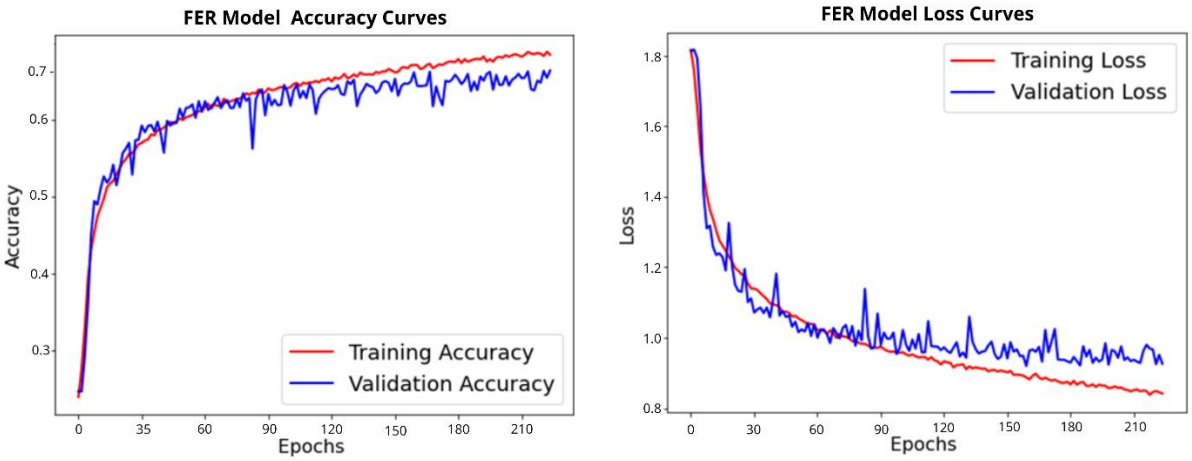

Figure 33: *Les courbes de précision et de perte pour le FER-Modèle.*

Il existe des mesures plus avancées qui peuvent être utilisées comme le score F1 que nous avons décidé d'utiliser. Le score F1 est calculé à l'aide de deux métriques pré-calculées :

La précision et le rappel. Ces deux mesures utilisent les exemples prédits vrais positifs, faux positifs et faux négatifs qui sont mieux visualisés à l'aide de la matrice de confusion. :

Puisque nous avons conçu notre modèle pour reconnaître les 7 émotions faciales universelles et que l'ensemble de données FER2013 avait une 8e classe pour les émotions de « contempt », nous avons décidé d'ajouter tous les exemples de classe de mépris à la classe « Happy » plutôt que de jeter ces données.

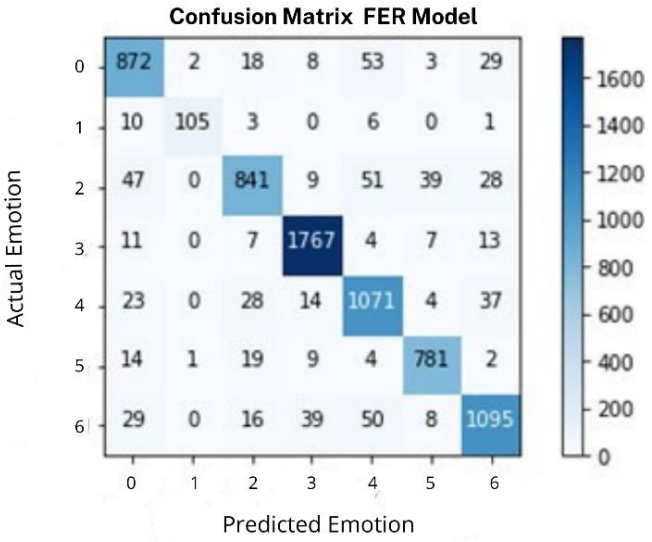

Figure 34: *Matrice de confusion pour le modèle FER*

Pour mieux tester l'efficacité du modèle développé dans ce projet, nous allons Essayer de comparer entre certain nombre de modèle avec le travail qui nous effectué.

Le 1er modèle c'est le RES-Net et 2éme VGG et 3éme c'est CNN-SVM de mémoire de l'année passe « Benzina Yacine & Loucif Kamel ».

III.5.3. RES-Net : (Residual Network)

Le modèle ResNet est une architecture de réseau neuronal convolutif introduite par Microsoft Research lors de la compétition ImageNet 2015.

Il a été conçu pour résoudre le problème de la dégradation de la performance avec l'augmentation de la profondeur du réseau.

L'idée clé des modèles ResNet est l'utilisation de blocs résiduels, qui permettent à un réseau d'apprendre des fonctions résiduelles plutôt que des fonctions directes.

Les blocs résiduels utilisent des connexions "skip" (connexions sautées) pour ajouter des raccourcis qui permettent au gradient de se propager plus facilement à travers le réseau.

Les modèles ResNet sont connus pour leur profondeur extrême, allant jusqu'à ResNet-152 avec 152 couches.

Le modèle ResNet a montré une performance supérieure sur de nombreux ensembles de données et a été largement utilisé dans des applications de vision par ordinateur.

III.5.3.1. Architecture global RES-NET

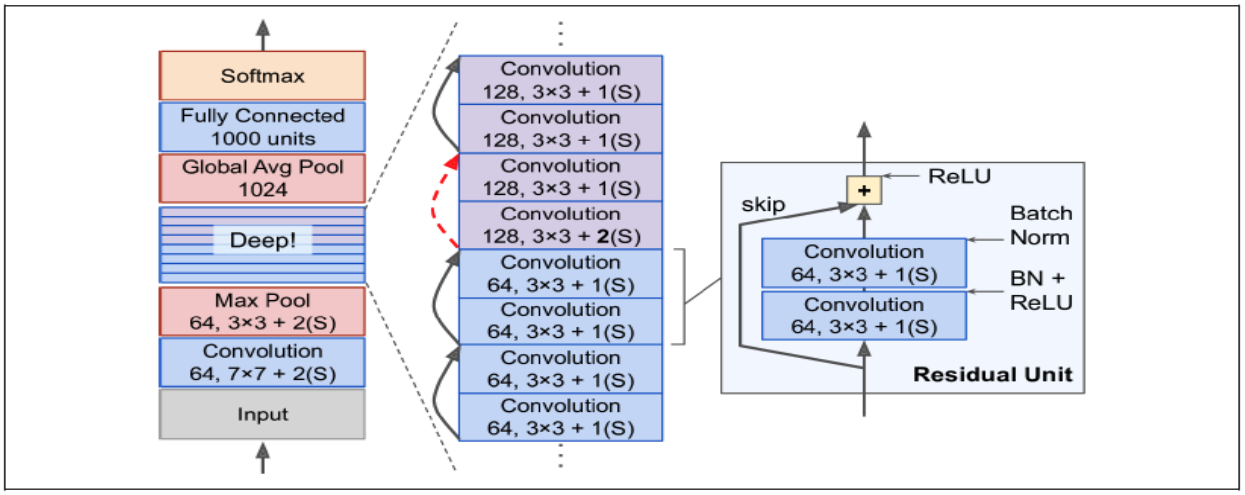

Figure 35: *Architecteur Global de Modéle Res-Net.[W28]*

III.5.3.2. Résultats expérimentaux et analyse de performance

III.5.3.2.1. Total de « parms »

Total params	Trainable params	Non-trainable params
21,303.943	21,288,71	15,232

Tableau 9: *Taille des paramètres du modèle Res-Net*

Après 210/210 Epoch :

Epoch	Loss	Accuracy	Val_loss	Val_accuracy
210/210	1.2103	0.4877	0.9265	0.6311

Tableau 10: *Résultats de Loss, Accuracy, Val_Loss et Val_Accuracy pour le modèle Res-Net*

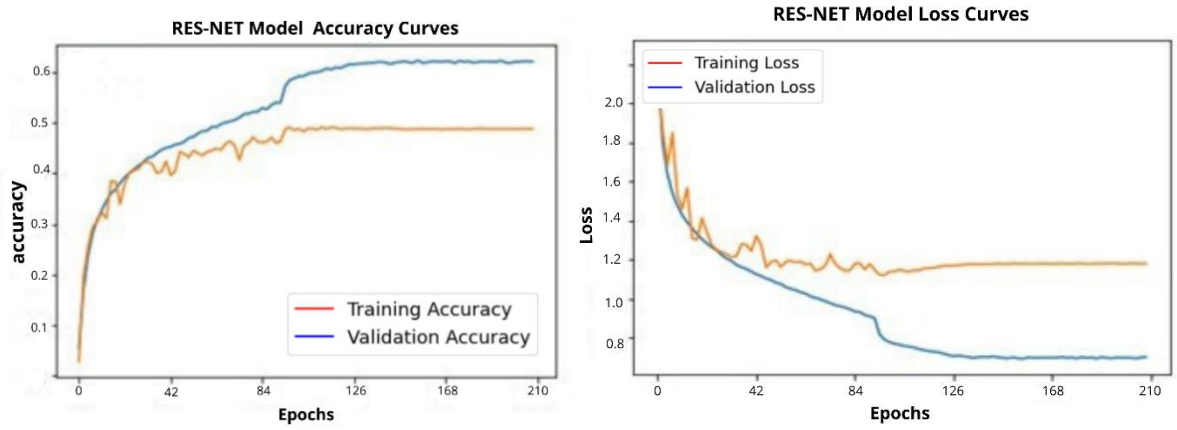

Figure 36: Les courbes de précision et de perte pour le modèle RES-NET

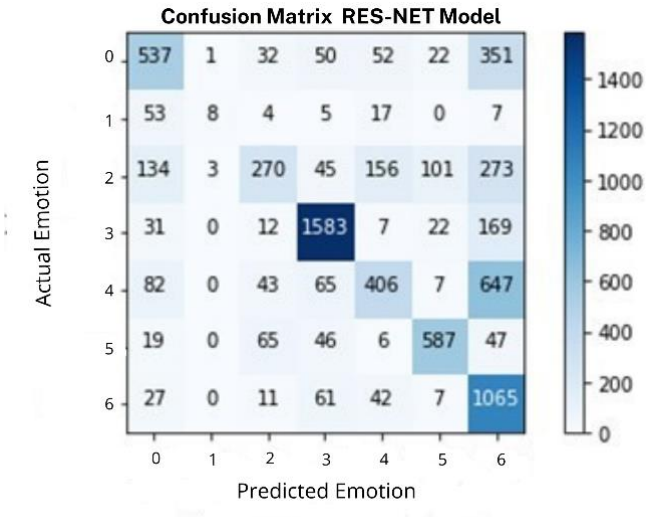

Figure 37: *Matrice de confusion pour le modèle RES-NET*

III.5.4. VGG (Visual Geometry Group)

Le modèle VGG est une architecture de réseau neuronal convolutif (CNN) proposée par le groupe Visual Geometry de l'Université d'Oxford.

Il est caractérisé par des couches de convolutions très profondes et une structure simple et uniforme.

L'architecture VGG comprend principalement des couches de convolutions avec des filtres de petite taille (3x3) et des couches de MaxPooling pour réduire les dimensions spatiales.

Le modèle VGG est connu pour avoir une profondeur variable avec différentes variantes : VGG16 avec 16 couches (13 convolutions et 3 couches entièrement connectées) et VGG19 avec 19 couches (16 convolutions et 3 couches entièrement connectées).

Le modèle VGG a une grande capacité d'apprentissage, mais il est également coûteux en termes de calcul et nécessite plus de mémoire en raison de sa profondeur.

III.5.4.1. Architecture VGG

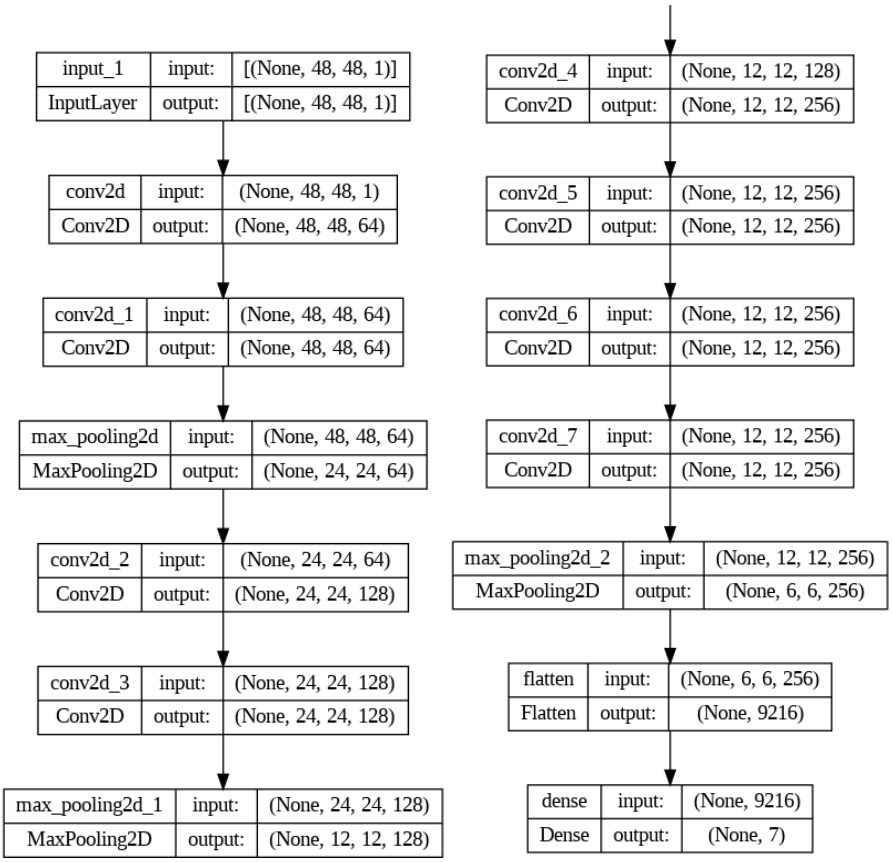

Figure 38 : *Les composons de L'architecture VGG Modèle*

III.5.4.2. Résultats expérimentaux et analyse de performance

Total params	Trainable params	Non-trainable params
2 ,388,935	**2,388,935**	**0**

Tableau 11: *Taille des paramètres du modèle VGG*

Epoch	Loss	Accuracy	Val_loss	Val_accuracy
210/210	**0.8292**	**0,6632**	**1.0698**	**0.5961**

Tableau 12: *Résultats de Loss, Accuracy, Val_Loss et Val_Accuracy pour le modèle VGG*

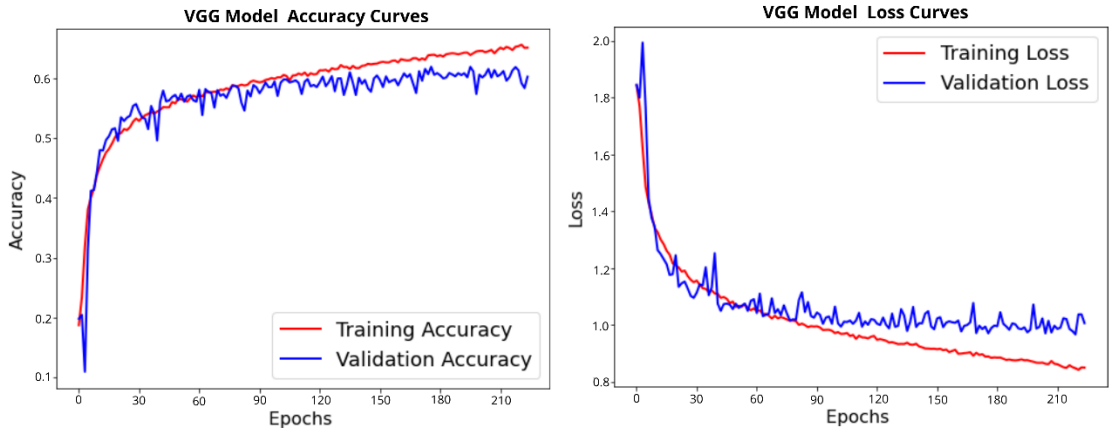

Figure 39: *Les courbes de précision et de perte pour le modèle VGG*

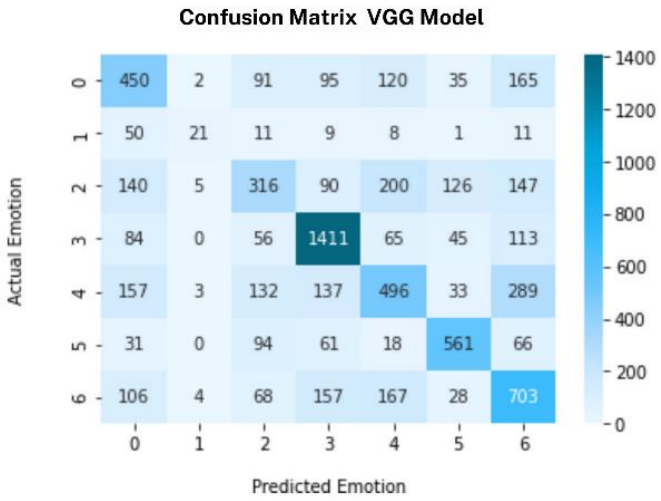

Figure 40: *Matrice de confusion pour le modèle VGG*

III.5.5. CNN-SVM (Loucif Kamel & Benzina Yacine) [61]

Dans ce modèle, combiné CNN avec SVM la même architecture du modèle FER précédent seulement quelques changements et en a parlé.

III.5.5.1. Résultats expérimentaux et analyse de performance

Dans la dernière couche dense de sortie, ajout d'un régularisateur l2 de 0.01 et d'une fonction de perte de charnière au carré.

Epoch	Loss	Accuracy	Val_loss	Val_accuracy
140/140	**1.0453**	**0,6421**	**1.0582**	**0.6244**

Tableau 13: *Résultats de Loss, Accuracy, Val_Loss et Val_Accuracy pour le modèle CNN-SVM (Loucif kamel et Benzina Yacine) [61].*

Avec SVM, le résultat obtenu était de **1,0582** de perte et de **62,44 %** de précision

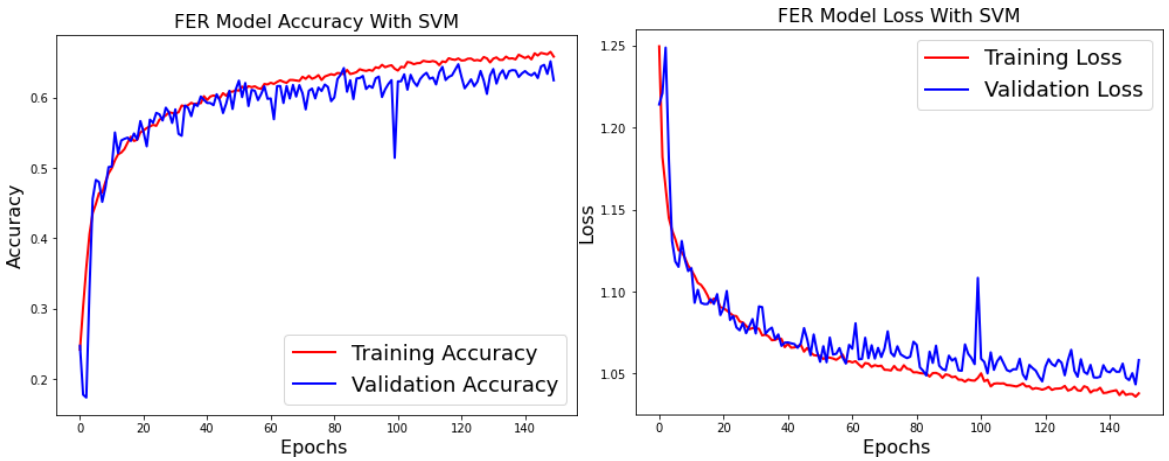

Figure 41: *The accuracy and loss curves for the FER model with SVM [61]*

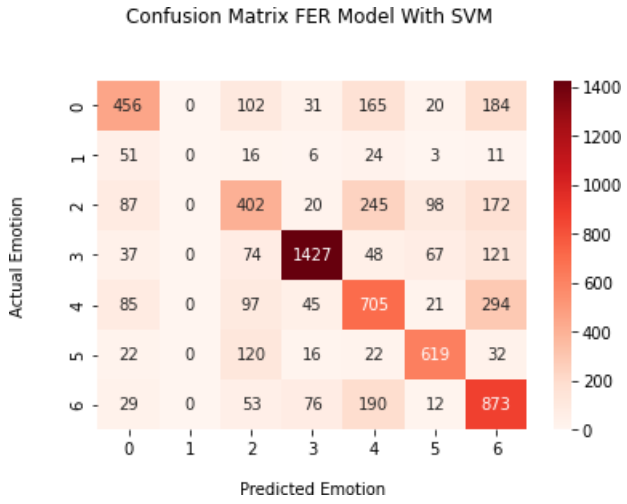

Figure 42: *Matrice de confusion pour le modèle FER CNN avec SVM [61]*

ID	Face Emotion	Precision	Recall	F1 Score
0	Angry	0.59452412	0.47599165	0.52869565
1	Disgust	0	0	0
2	Fear	0.46527778	0.39257812	0.42584746
3	Happy	0.88032079	0.80439684	0.84064801
4	Sad	0.50393138	0.56535686	0.53287982
5	Surprise	0.73690476	0.74488568	0.74087373
6	Neutral	0.51748666	0.7080292	0.59794521

Tableau 14: *Résultats de précision, de rappel et de score F1 pour le modèle FER avec SVM.[61]*

III.6. La comparaison

Dans cette partie spécifique, nous comparerons nos différents modèles FER [**Tableau 15**]

Model	**Accuracy %**	**Loss**
VGG	**59.61%**	**1.0698**
CNN-SVM	**62.44%**	**1.0582**
RES-Net	**63.11%**	**0.9265**
Séquentiel	**66.01%**	**0.8967**
Fer-modèle	**69.80%**	**0.8450**

Tableau 15: *La différence de précision et les résultats de perte des modèles proposés*

En comparant ces modèles, nous pouvons noter les observations suivantes :

- Le modèle Fer-modèle présente la meilleure précision parmi tous les modèles, atteignant 69.80%. Cela indique que ce modèle est le plus performant en termes de classification sur la tâche spécifique.
- Le modèle Séquentiel est également très performant avec une précision de 66.01%, ce qui en fait le deuxième meilleur modèle de la liste.
- RES-Net suit de près avec une précision de 63.11%. Il est légèrement moins performant que le modèle séquentiel mais reste solide.
- Les modèles CNN-SVM et VGG ont des précisions légèrement inférieures, 62.44% et 59.61% respectivement, ce qui les place en bas de la liste en termes de précision.
- En ce qui concerne la perte, le modèle Fer-modèle présente également la plus faible perte, ce qui suggère une meilleure capacité à minimiser les erreurs de classification.

En résumé, le modèle Fer-modèle est le leader en termes de précision et de perte, suivi du modèle Séquentiel et de RES-Net, tandis que les modèles CNN-SVM et VGG sont légèrement moins performants. Le choix du modèle dépendra de l'importance que vous accordez à la précision et à la perte dans votre application spécifique.

III.7. Consultions

Ce chapitre a couvert les principales conclusions de la thèse, qui est la reconnaissance des émotions faciales à l'aide de techniques d'apprentissage en profondeur. Les expériences ont été détaillées et expliquées. Les résultats ont été illustrés sous forme de figures, de tableaux et de commentaires. Chaque section donne des comparaisons des résultats obtenus.

Conclusion Générale

Dans ce travail présenté nous nous sommes concentrés sur la classification automatique de la reconnaissance des expressions faciales afin de développer des applications utiles pour identifier les émotions d'une personne en utilisant des approches d'apprentissage profond.

Nous avons mené de nombreuses expériences en utilisant et en comparant différentes algorithmes d'apprentissage profond : les réseaux de neurones convolutifs, les réseaux de neurones artificiels, des méthodes d'apprentissage par transfert avec différentes architectures RES-NET, VGG, ainsi que le modèle FER proposé et le modèle séquentiel. Les expériences ont été menées en ajustant les hyperparamètres tels que la taille du lot, le nombre d'époques, le nombre de couches, d'unités, et le type de couches (convolutives, d'abandon, BatchNormalization, MaxPooling) pour améliorer la précision et éviter le surajustement ou le sous-ajustement.

Nous avons utilisé l'ensemble de données FER2013, qui contient les sept émotions de base, pour construire et tester le modèle d'apprentissage profond proposé. Il s'agit d'un réseau de neurones convolutif profond qui apprend automatiquement des caractéristiques à partir des images d'origine.

Les expériences ont montré que, en utilisant notre architecture proposée, les résultats de précision surpassent toutes les autres approches proposées (RES-NET, VGG, SVM) en termes de performances d'apprentissage profond. Notre proposition a atteint une précision de 69.80% dans le test du modèle FER en utilisant les sept émotions, ce qui dépasse les autres architectures utilisées.

Dans les travaux futurs, nous proposerons d'autres méthodes pour améliorer ces résultats, comme l'utilisation d'autres techniques de détection des caractéristiques du visage et le calcul de certaines équations qui pourraient contribuer à rendre le problème plus précis.

Bibliographie

[1] Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), Nebraska Symposium on Motivation (Vol. 19, pp. 207-283). University of Nebraska Press.

[2] Bledsoe, W.W., Chan Wolf, H.P. et Bisson, L.F. (1966). Reconnaissance des visages humains. Dans Actes de la conférence informatique conjointe d'automne du 27 au 29 décembre 1966, partie I (pp. 394-409).

[3] Kanade, T. (1973). Traitement d'image par complexe informatique et reconnaissance de visages humains. Thèse de doctorat, Université de Kyoto.

[4] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition pages 586–591

[5] Turk, M., & Pentland, A. (1991). Eigenfaces pour la reconnaissance. Journal des neurosciences cognitives, 3(1), 71-86.

[6] A. Boughida, M. N. Kouahla, and F. Z. Bouhlaci et al. Emad : Un systéme d'apprentissage humain adaptatif à base d'émotions. ISKO Maghreb, pages 25– 26

[7] Angwin, J., Larson, J., Mattu, S. et Kirchner, L. (2016). Biais de la machine. ProPublica.

[8] F. Khalfi. Reconnaissance automatique des émotions par données multimodales : expressions faciales et des signaux physiologiques. PhD thesis, Université Paul Verlaine de Metz, France, 2010.

[9] Ali Mollahosseinil, David Chan2, and Mohammed H. Mahoor1 Going deeper in Facial Expression Recognition using Deep Neural Network, Department of Electrical and Computer Engineering, Department of Computer Science, University of Denver, Denver, CO, 5 January 2017.

[10] Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. Journal of personality and social psychology, 17(2), 124-129

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

[12] E. Couzon and F. Dorn. Les émotions : développer son intelligence émotionnelle. ESF editeur,2011

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

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

[15] Ekman, P. (1992). An argument for basic emotions. Cognition & Emotion, 6(3-4), 169-200

[16] Lazarus, R. S. (1991). Emotion and adaptation. Oxford University Press.

[17] E. Couzon and F. Dorn. Les émotions : développer son intelligence émotionnelle. ESF editeur,2011.

[18] Paul Ekman, Universal Facial Expressions of Emotion, Calif. Ment. Heal. Res. Dig. Vol. 8, no.4, pp.151-158, 1970

[19] Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. Psychological Review, 69(5), 379-399.

[20] Levenson, R. W. (1994). Human Emotion: A Functional View. In P. Ekman & R. J. Davidson (Eds.), The Nature of Emotion: Fundamental Questions (pp. 123-126). Oxford University Press.

[21] Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. Emotion Review, 3(4), 397-405.

[22] Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2005). Automatic recognition of facial actions in spontaneous expressions. Journal of Multimedia, 1(6), 22-35.

[23] Deng, J., Guo, J., & Xue, N. (2019). Deep learning for image-based facial expression recognition: A comprehensive review. Neural Computation, 31(5), 885-938.

[24] Kaltwang, S., Todorovic, S., & Pantic, M. (2012). Emotion recognition from facial expressions using multilevel HMMs. IEEE Transactions on Multimedia, 14(3), 677-690.

[25] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10) (pp. 807-814).

[26] Shahid, N., & Chiaberge, M. (2019). Facial expression recognition: A brief review of the technology and its applications. IEEE Instrumentation and Measurement Magazine, 22(2), 16-23.

[27] Valstar, M. F., Mehu, M., Jiang, B., Pantic, M., & Scherer, K. (2014). Meta-analysis of the first facial expression recognition challenge. IEEE Transactions on Affective Computing, 5(3), 242-251.

[28] R.Collobert et J.Weston : A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning (pp. 160- 167). 2008, July.

[29] Maalej, A., Amor, B. B., et M.Daoudi : Analyse locale de la forme 3D pour la reconnaissance d'expressions faciales. 2011, June.

[30] M.Quinn, G.Sivesind, G.Reis," Real-time Emotion Recognition From Facial Expressions", Stanford University 2015.

[31] Detection of emotions from video in a non-controlled environment, Rizwan Ahmed Khan.

[32] BELHADJ Mahdi ,Etude et simulation d'un syst`eme de reconnaissance des expressions faciale.universit´e de Biskra 2019.

[33] Khadija Lekdioui. Reconnaissance d''états ´émotionnels par analyse visuelle du visage et apprentissage machine. Synthèse d'image et réalité virtuelle [cs.GR]. Université Bourgogne

Franche-Comté ; Université Ibn Tofail. Faculté des sciences de Kenitra, 2018.

[34] McCulloch, W.S., & Pitts, W. (1943). Calcul logique des idées immanentes à l'activité nerveuse. Le Bulletin de biophysique mathématique, 5(4), 115-133.

[35] Rosenblatt, F. (1958). Le perceptron : un modèle probabiliste pour le stockage et l'organisation de l'information dans le cerveau. Revue psychologique, 65(6), 386-408.

[36] Rumelhart, D.E., Hinton, G.E. et Williams, R.J. (1986). Apprentissage des représentations par rétro-propagation des erreurs. Nature, 323(6088), 533-536.

[37] LeCun, Y., Bengio, Y. et Hinton, G. (2015). L'apprentissage en profondeur. Nature, 521(7553), 436-444.

[38] Goodfellow, I., Bengio, Y., & Courville, A. (2016). L'apprentissage en profondeur. Presse du MIT.

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

[40] A. Fathallah, L. Abdi et A. Douik : Facial expression recognition via deep learning. In 2017 IEEE/ACS14th International Conference on Computer Systems and Applications (AICCSA) (pp. 745-750), 2017.

[41] Y. LeCun, Y. Bengio, et G. Hinton, « Deep learning », Nature, vol. 521, no 7553, p. 436-444, mai 2015.

[42] A. Graves, « Generating sequences with recurrent neural networks », arXiv preprint arXiv:1308.0850, 2013.

[43] I. Goodfellow et al., « Generative adversarial nets », Advances in neural information processing systems, p. 2672-2680, 2014.

[44] P. Vincent et al., « Extracting and composing robust features with denoising autoencoders », Proceedings of the 25th international conference on Machine learning, p. 1096-1103, 2008.

[45] K. He et al., « Deep residual learning for image recognition », Proceedings of the IEEE conference on computer vision and pattern recognition, p. 770-778, 2016.

[46] M.Lyons et S.Akamatsu et M.Kamachi et J.Gyoba : Coding facial expressions with gabor wavelets. InProceedings Third IEEE international conference on automatic face and gesture recognition (pp. 200-205).(1998, April).

[47] Fathallah, L. Abdi et A. Douik : Facial expression recognition via deep learning. In 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA) (pp. 745-750), 2017.

[48] Kh.Lekdioui : Reconnaissance d'états émotionnels par analyse visuelle du visage et apprentissagemachine. Synthèse d'image et réalité virtuelle [cs.GR]. Université Bourgogne Franche-Comté; Université Ibn Tofail. Faculté des sciences de Kénitra, 2018.

[49] C.Qi et M.Li et Q.Wang, H.Zhang et J.Xing et Z.Gao et H.Zhang : Facial expressions recognition based on cognition and mapped binary patterns, 2018.

[50] A. Fathallah, L. Abdi et A. Douik : Facial expression recognition via deep learning. In 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA) (pp. 745-750), 2017

[51] Kh.Lekdioui : Reconnaissance d'états émotionnels par analyse visuelle du visage et apprentissage

[52] C.Qi et M.Li et Q.Wang, H.Zhang et J.Xing et Z.Gao et H.Zhang : Facial expressions recognition based on cognition and mapped binary patterns, 2018.

[53] A. Fathallah, L. Abdi et A. Douik : Facial expression recognition via deep learning. In 2017 IEEE/ACS

[54] C.Padgett et G.Cottrell : Representing face images for emotion classification. In Advances in neural information processing systems, pages 894–900, 1997

[56] V.Perlibakas : Face recognition using principal component analysis and loggabor filters. arXiv preprint cs/0605025 evolution for feature selection in facial expression recognition systems. Expert Systems with Applications, 89, 129-137. 2006.

[57] Dung Nguyen M.Sc., B.Sc : Multimodal Emotion Recognition Using Deep Learning Techniques,School of Electrical Engineering and Computer Science Science and Engineering Faculty Queensland University of Technology 2020.

[58] U.Mlakar, I.Fister, J.Brest et B.Potočnik: Multi-objective differential 2017

[59] I.Cohen, N.Sebe, A.Garg, L. S.Chen et T. S.Huang : Facial expression recognition from video sequences: temporal and static modeling. Computer Vision and image understanding, 91(1-2), 160-187. 2003.

[60] F.Davoine, B.Abboud et V. M. Dang,: Analyse de visages et d'expressions faciales par modèle actif d'apparence. traitement du signal, 1(3). 2004.

[61] K. LOUCIF, Y. BENZINA. Facial Emotion Recognition Based on Deep Learning. Master's thesis, Université de Mohamed El Bachir El Ibrahimi Bordj Bou Arréridj, Algerie, 2022.

[W1] https://www.verdict.co.uk/computer-vision-timeline/

[W2] https://www.eiagroup.fr/domaines-expertise/expressions-faciales-et-micro-expressions/

[W3] http://emotionresearcher.com/an-audio-interview-with-paul-ekman/

[W4] https://en.wikipedia.org/wiki/Face_detection

[W5] https://www.techtarget.com/searchenterpriseai/definition/face-detection

[W6] https://magazine.comunicazionestrategica.it/les-micro-expressions-faciales-une-langue-universelle/

[W7] https://dspace.univ-guelma.dz/jspui/bitstream/

[W8] https://mrmint.fr/apprentissage-supervise-machine-learning

[W9] https://fr.linedata.com/quest-ce-que-lapprentissage-supervise

[W10] http://archives.univbiskra.dz/bitstream/123456789/18930/1/ALLAOUA Youcef.pdf

[W11] https://www.researchgate.net/figure/Illustration-of-convolution-operation-for-2Dgrids-left-and-3D-volumes-right-Complete_fig3_331165618

[W12] https://cedric.cnam.fr/vertigo/cours/ml2/tpDeepLearning3.html

[W13] https://leonardoaraujosantos.gitbook.io/artificial-inteligence/machine_learning/deep_learning/pooling_layer

[W14] https://www.researchgate.net/figure/Representation-graphique-des-fonctions-dactivation-sigmoide-et-ReLU-7_fig9_327882341

[W15] https://www.researchgate.net/figure/Architecture-classique-dun-reseau-de-neurones-convolutif-Une-image-est-fournie-en_fig5_330995099

[W16] https://towardsdatascience.com/

[W17] https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

[W18] https://www.anaconda.com

[W19] https://www.spyder-ide.org/

[W20] https://www.python.org./about/

[W21] https://opencv.org/about/

[W22] https://numpy.org/

[W23] Keras, available at: http://www.keras.com

[W24] https://pandas.pydata.org/

[W25] Tensorflow, available at: http://www.tensorflow.com.

[W26] https://www.kaggle.com/datasets/msambare/fer2013.

[W27] https://www.projectpro.io/article/facial-emotion-recognition-project-using-cnn-with-source-code/570

[W28] https://jason-adam.github.io/resnet50/