

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

Université Mouloud Mammeri De Tizi-Ouzou

Faculté de génie électrique et informatique Département  
d'électronique



# Mémoire de fin d'études

En vue de l'obtention du Diplôme Master académique en  
Electronique des systèmes embarqués

**THEME :**

**Identification des états  
émotionnels via l'expression**

**Réalisé par :**

Alia Lounes

Mouloudj Nacera

Lazri Mourad	Professeur	à	Ummto	Encadrant
Kherchaoui Sonia	M.C.B	à	Ummto	Co-encadrant
Ouallouche Fethi	M.C.A	à	Ummto	président
Alouache Djamel	M.C.A	à	Ummto	Examineur

Année universitaire : 2023/2024.

# Remerciement

---

Nous remercions, avant tout, Allah le Tout puissant pour nous avoir donné la volonté et la patience pour accomplir ce travail.

Nous exprimons également nos sincères remerciements à notre encadreur de mémoire, Monsieur LAZRI.M, ainsi qu'à Madame KHERCHAOUIS, pour leurs précieux conseils et orientations tout au long de ce projet. Leur expertise, leur disponibilité et leurs encouragements ont été essentiels à la réalisation de ce travail.

Nous tenons aussi à remercier les membres du jury pour avoir accepté d'examiner et d'évaluer ce travail.

Nos pensées vont à tous les enseignants qui ont participé à notre formation. Nous tenons à remercier profondément tous ceux qui ont participé de loin ou de près à la réalisation de ce travail.

Je dédie ce travail à :

À mes chers parents, qui m'ont donné le meilleur  
de ce qu'ils ont, que Dieu les garde,

À ma grand-mère, que Dieu te prête longue vie,

À mes chers frères,

À tous mes chers amis,

À mes chers camarades,

À toutes les personnes qui ont contribué à la  
réalisation de ce travail de près ou de loin.

À mon binôme Nacera,

Alia Lounes.

*Dédicace:*

**À mes chers parents,**

Je tiens à vous exprimer toute ma gratitude et mon amour. Vous m'avez donné les valeurs, la force et les fondations qui m'ont guidée tout au long de mon parcours scolaire. Merci pour votre soutien indéfectible et votre amour inconditionnel

**À toute la famille MOULOUDJ « grands et petits »,**

Merci pour votre soutien et votre affection, Vous êtes une source inépuisable de bonheur et d'inspiration pour moi.

**A mon binôme Iounes.**

# Liste des figures

---

## CHAPITRE 1

<b>FIGURE I.1</b> : NOTATIONS UTILISEES POUR UNE IMAGE NUMERIQUE. ....	18
<b>FIGURE I.2</b> : IMAGE BINAIRE, AVEC LE TABLEAU DE VALEURS CORRESPONDANT. ....	19
<b>FIGURE I.3</b> : REPRESENTATION EN NIVEAUX RVB.....	20
<b>FIGURE I.4</b> : IMAGE EN NIVEAUX DE GRIS, AVEC LE TABLEAU DE VALEURS CORRESPONDANT.....	20
<b>FIGURE I.5</b> : REPRESENTATION D'UNE IMAGE EN NIVEAU DE GRIS AVEC SON HISTOGRAMME.....	21
<b>FIGURE I.6</b> : IMAGE NUMERIQUE QUI INDIQUE LE NOMBRE DE COLONNES ET DE LIGNES.....	22
<b>FIGURE I.7</b> : IMAGE AVEC DIFFERENTE RESOLUTION.....	23
<b>FIGURE I.8</b> : IMAGE D'UN VISAGE FRAGMENTE : L'ECHANTILLONNAGE DEVOILE LA CONSTRUCTION DE L'IMAGE NUMERIQUE. ....	24
<b>FIGURE I.9</b> : IMAGE AVEC ET SANS BRUIT.....	25
<b>FIGURE I.10</b> : IMAGE AVANT ET APRES DETECTION DE CONTOUR.....	26
<b>FIGURE I.11</b> : LE LISSAGE PAR MOYENNAGE.....	27
<b>FIGURE I.12</b> : LE LISSAGE PAR FILTRE GAUSSIEN.....	28
<b>FIGURE I.13</b> : LISSAGE PAR FILTRE MEDIAN.....	28
<b>FIGURE I.14</b> : EXPANSION DE LA DYNAMIQUE. ....	29
<b>FIGURE I.15</b> : IMAGE AVEC SON HISTOGRAMME EGALISE.....	30

## CHAPITRE 2

<b>FIGURE II.1</b> : LES SEPT EXPRESSIONS DE BASE DEFINIES PAR EKMAN DE LA BASE DE DONNEES JAFFE.....	34
<b>FIGURE II.2</b> : EXEMPLE DES EMOTIONS ADAPTATIVES DE LA BASE DE DONNEES JAFFE.....	35
<b>FIGURE II.3</b> : EXEMPLE DES EMOTIONS NON ADAPTATIVES DE LA BASE DE DONNEES JAFFE.....	35
<b>FIGURE II.4</b> : CARACTERISTIQUES GEOMETRIQUE. ....	39
<b>FIGURE II.5</b> : CARACTERISTIQUES D'APPARENCES.....	39
<b>FIGURE II.6</b> : SCHEMA GENERALE D'UN SYSTEME DE RECONNAISSANCE DES VISAGES. ....	40
<b>FIGURE II.7</b> : DETECTION DE VISAGE.....	42
<b>FIGURE II.8</b> : ACQUISITION ET TRAITEMENT DU VISAGE.....	42
<b>FIGURE II.9</b> : AVANT ET APRES LA NORMALISATION.....	43

### **CHAPITRE 3**

<b>FIGURE III.1</b> : LA RELATION ENTRE IA, ML, DL.....	48
<b>FIGURE III.2</b> : SEPARATION ENTRE DEUX CLASSES PAR UN HYPERPLAN.....	51
<b>FIGURE III.3</b> .....	52
<b>FIGURE III.4</b> : DIAGRAMME DU PROCESSUS D'APPRENTISSAGE AUTOMATIQUE SUPERVISE.....	52
<b>FIGURE III.5</b> : DIAGRAMME DU PROCESSUS D'APPRENTISSAGE AUTOMATIQUE NON SUPERVISE.....	53
<b>FIGURE III.6</b> : PRINCIPE D'APPRENTISSAGE PAR RENFORCEMENT.....	53
<b>FIGURE III.7</b> : MACHINE LEARNING VS DEEP LEARNING.....	54
<b>FIGURE III.8</b> : NEURONE BIOLOGIQUE VS NEURONE ARTIFICIEL.....	55
<b>FIGURE III.9</b> : SCHEMA DE FONCTIONNEMENT DES RESEAUX DE NEURONES.....	55
<b>FIGURE III.10</b> : SCHEMA D'UN PERCEPTRON.....	56
<b>FIGURE III.11</b> : ARCHITECTURE STANDARD D'UN RESEAU DE NEURONE CONVOLUTIONNEL.....	58
<b>FIGURE III.12</b> : REPRESENTATION D'UNE COUCHE CONVOLUTIVE.....	59
<b>FIGURE III.13</b> : EXEMPLE D'UN MAX POOLING.....	60
<b>FIGURE III.14</b> : OPERATION RELU.....	61
<b>FIGURE III.15</b> : LA COUCHE ENTIEREMENT CONNECTEE.....	61

### **CHAPITRE 4**

<b>FIGURE IV.1</b> : ÉCHANTILLONS DE L'ENSEMBLE D'IMAGE JAFFE.....	66
<b>FIGURE IV.2</b> : ÉCHANTILLONS DE L'ENSEMBLE D'IMAGE CK+.....	67
<b>FIGURE IV.3</b> : IMAGE DE LA BASE JAFFE APRES LA REORGANISATION.....	67
<b>FIGURE IV.4</b> : ARCHITECTURE PROPOSEE.....	71
<b>FIGURE IV.5</b> : GRAPHE DE PRECISION POUR LA BASE DONNEE CK+.....	72
<b>FIGURE IV.6</b> : GRAPHE DE PERTE POUR LA BASE DONNEE CK+.....	73
<b>FIGURE IV.7</b> : GRAPHE DE PRECISION POUR LA BASE DONNEE JAFFE.....	73
<b>FIGURE IV.8</b> : GRAPHE DE PERTE POUR LA BASE DONNEE JAFFE.....	74
<b>FIGURE IV.9</b> : MATRICE DE CONFUSION POUR 7 EXPRESSIONS POUR LA BASE DE DONNEES JAFFE.....	75
<b>FIGURE IV.10</b> : MATRICE DE CONFUSION POUR 6 EXPRESSIONS POUR LA BASE DE DONNEES CK+.....	75
<b>FIGURE IV.11</b> : INTERFACE VIDE.....	80
<b>FIGURE IV.14</b> : INTERFACE VIDE.....	81
<b>FIGURE IV.13</b> : EMOTION PREDITE (COLERE).....	82

# Liste des tableaux

---

<b>TABLEAU I :</b> LES AUs FREQUENTES DECODEES PAR LE MANUEL FACS SUR LA BASE D'IMAGES CK+. ....	36
<b>TABLEAU II :</b> COMBINAISON DES AUs CORRESPONDANT A CHAQUE EMOTION. ....	37
<b>TABLEAU III :</b> REPARTITION DES IMAGES PAR CLASSE D'EMOTION DANS LA BASE DE DONNEES JAFFE. ....	64
<b>TABLEAU IV :</b> REPARTITION DES IMAGES PAR CLASSE D'EMOTION DANS LA BASE DE DONNEES CK+. ....	65
<b>TABLEAU V :</b> REPARTITION DETAILLEE DES PERFORMANCES PAR CLASSE AVEC JAFFE. ....	75
<b>TABLEAU VI :</b> REPARTITION DETAILLEE DES PERFORMANCES PAR CLASSE AVEC CK+ ....	76
<b>TABLEAU VII :</b> COMPARAISON DES PERFORMANCES SUR LES ENSEMBLES DE DONNEES JAFFE ET CK+. ....	78

# Liste des abréviations

---

- **FER** : Facial expression Recognition.
- **FACS** : Facial Action Coding System.
- **AU** : Action Unit.
- **IA** : Intelligence Artificielle.
- **CNN** : Convolutional Neural Networks.
- **ML** : machine learning.
- **DL** : deep learning.
- **RGB**: Red, Green, Blue.
- **RVB**: Rouge, Vert, Bleu.
- **TIFF**: Tagged Image File Format.
- **PNG**: Portable Networks Graphics.
- **JPEG**: Joint Photographic Expert Group.
- **SVM** : Support Vector Machine.
- **CONV**: Convolution.
- **POOL**: Pooling.
- **RELU**: Rectified Linear Units.
- **FC**: Fully Connected.

# Table des matières

---

## INTRODUCTION

### CHAPITRE I : GENERALITES SUR LE TRAITEMENT D'IMAGE

INTRODUCTION.....	17
I.1. DEFINITION D'IMAGE: .....	17
<i>I.1.1. Types d'images :</i> .....	18
I.1.1.1. Image binaire :.....	18
I.1.1.1. Image en couleurs : .....	18
I.1.1.2. Image en niveaux de gris :.....	19
I.2. HISTOGRAMME D'UNE IMAGE : .....	20
I.3. ANALYSE DES CARACTERISTIQUES D'IMAGES EN TRAITEMENT D'IMAGE : .....	20
<i>I.3.1. Définition du traitement d'image :</i> .....	20
I.3.1.1. Historique et évolution du traitement d'image : .....	21
I.3.1.2. Applications du Traitement d'Image : .....	21
I.4. CONCEPT DE BASE : .....	21
<i>I.4.1. Pixel :</i> .....	21
<i>I.4.2. Dimension :</i> .....	22
<i>I.4.3. Résolution et échantillonnage :</i> .....	22
<i>I.4.4. Résolution :</i> .....	22
<i>I.4.5. L'échantillonnage :</i> .....	23
<i>I.4.6. Contraste :</i> .....	23
<i>I.4.7. Luminance :</i> .....	24
<i>I.4.8. Bruit (parasite) :</i> .....	24
<i>I.4.9. Contours :</i> .....	25
<i>I.4.10. Texture :</i> .....	25
I.5. PRETRAITEMENT DES IMAGES : .....	26
I.6. FILTRAGE ET LISSAGE : .....	26
<i>I.6.1. Filtre moyennneur :</i> .....	26
<i>I.6.2. Filtre gaussien :</i> .....	27
<i>I.6.3. Filtre median :</i> .....	27
I.7. LA MODIFICATION D'HISTOGRAMME : .....	28
<i>I.7.1. Expansion de dynamique :</i> .....	28
<i>I.7.2. Égalisation d'histogramme :</i> .....	29
CONCLUSION.....	29

## CHAPITRE II : ÉTATS DE L'ART SUR L'IDENTIFICATION DES EMOTIONS FACIALE

INTRODUCTION.....	31
II.1. CONTEXTE ET OBJECTIFS : .....	31
<i>II.1.1. Présentation générale du projet de (FER):</i> .....	31
<i>II.1.2. Intérêt et applications de cette technologie :</i> .....	31
II.2. ÉMOTION ET EXPRESSION FACIALE:.....	32
<i>II.2.1. Définition des émotions :</i> .....	32
<i>II.2.2. Types d'émotions :</i> .....	32
II.2.2.1. Émotions primaire : .....	33
II.2.2.2. Émotions secondaires :.....	35
II.2.2.3. Émotions sociales : .....	35
<i>II.2.3. Les expressions faciales :</i> .....	35
II.3. LE SYSTEME DE CODAGE DES (FACS) : .....	36
II.4. LES DIFFERENTES APPROCHES UTILISEES POUR UN SYSTEME (FER) .....	37
<i>II.4.1. Approche conventionnelle :</i> .....	37
II.4.1.1. Approche géométrique : .....	38
II.4.1.2. Approche basé sur les caractéristiques d'apparence :.....	38
<i>II.4.2. Approche basées sur l'apprentissage profond :</i> .....	38
II.5. METHODES D'ACQUISITIONS ET TRAITEMENT DES IMAGES : .....	39
<i>II.5.1. Acquisitions de visage :</i> .....	39
II.5.1.1. Détection de visage et prétraitement : .....	40
<i>II.5.2. La normalisation :</i> .....	42
<i>II.5.3. Extraction de caractéristiques:</i> .....	43
<i>II.5.4. Classification :</i> .....	43
<i>II.5.5. Base de données :</i> .....	44
CONCLUSION.....	44

## CHAPITRE III : RESEAUX DE NEURONES CONVOLUTIFS ET APPRENTISSAGE AUTOMATIQUE

INTRODUCTION : .....	46
III.1. TECHNIQUES D'APPRENTISSAGE AUTOMATIQUE .....	46
III.2. IMPORTANCE DE ML DANS LA REF : .....	48
<i>III.2.1. Traitement des données complexes</i> : .....	48
<i>III.2.2. Extraction des caractéristiques</i> : .....	48
<i>III.2.3. Adaptabilité et généralisation</i> : .....	48
<i>III.2.4. Amélioration de la précision</i> : .....	48
<i>III.2.5. Applications pratique</i> : .....	48
III.3. TYPES D'ALGORITHMES D'APPRENTISSAGE AUTOMATIQUE : .....	49
<i>III.3.1. Techniques d'apprentissage supervisé</i> : .....	49
III.3.1.1. Machine à vecteurs de support (SVM) (apprentissage supervisé– classification) : ....	49
III.3.1.2. Définition et fonctionnement de l'algorithme de classification SVM : .....	50
<i>III.3.2. Techniques d'apprentissage non supervisées</i> : .....	52
<i>III.3.3. Techniques d'apprentissage par renforcement</i> : .....	52
III.4. DEEP LEARNING : .....	53
III.5. RESEAUX DE NEURONES : .....	54
<i>III.5.1. Le perceptron</i> : .....	55
<i>III.5.2. Perceptron multicouche</i> : .....	56
III.6. UTILISATION DES CNN DANS L'IDENTIFICATION DES ETATS EMOTIONNELS VIA EXPRESSIONS FACIALE : .....	56
III.7. L'APPRENTISSAGE EN PROFONDEUR (DEEP LEARNING): .....	56
III.8. LES RESEAUX DE NEURONES CONVOLUTIFS .....	57
<i>III.8.1. Architecture d'un réseau de neurone</i> .....	58
III.8.1.1. Les couches de convolution : .....	58
III.8.1.2. Les couches de regroupements (pooling) : .....	59
III.8.1.3. Couche de correction (RELU) : .....	59
III.8.1.4. Couche entièrement connectée (FC) : .....	60
III.8.1.5. Couche de perte (LOSS) : .....	61
CONCLUSION .....	61

## CHAPITRE IV : IMPLEMENTATION DU CNN ET RESULTATS

INTRODUCTION : .....	63
IV.1. ENVIRONNEMENT DE TRAVAIL : .....	63
IV.1.1. <i>Environnement matériel</i> : .....	63
IV.1.2. <i>Environnement logiciel</i> : .....	63
IV.1.2.1. Utilisation de MATLAB 2021a : .....	63
IV.1.2.2. Les particularités de MATLAB : .....	64
IV.2. DESCRIPTION DES BASES DE DONNEES .....	64
IV.2.1. <i>Base de données JAFFE</i> : .....	64
IV.2.2. <i>Base de données CK+</i> : .....	65
IV.3. PRETRAITEMENT DE DONNEES : .....	66
IV.3.1. <i>Organisation des Données d'Images</i> : .....	66
IV.3.2. <i>La normalisation</i> : .....	67
IV.3.3. <i>Le redimensionnement</i> : .....	67
IV.4. L'EXTRACTION DE FONCTIONNALITE : .....	67
IV.5. PRESENTATION DU CNN : .....	67
IV.6. ENTRAINEMENT DU MODELE : .....	68
IV.7. ARCHITECTURE CNN POUR L'EXTRACTION DE FONCTIONNALITE : .....	68
IV.8. CNN CLASSIFICATION : .....	70
IV.9. EVALUATION DU MODEL : .....	70
IV.10. RESULTATS ET DISCUSSIONS : .....	71
IV.11. MATRICE DE CONFUSION : .....	74
IV.12. ANALYSE COMPARATIVE DES PERFORMANCES : .....	77
IV.13. TESTS SUR DES IMAGES N'APPARTENANT PAS A LA BASE : .....	78
IV.14. BLOC INTERFACE : .....	79
CONCLUSION .....	82
<b>CONCLUSION .....</b>	<b>83</b>

---

---

# **INTRODUCTION**

---

---

## INTRODUCTION

---

Le visage joue un rôle crucial dans la reconnaissance des émotions, car il est l'un des moyens les plus évidents pour exprimer ce que nous ressentons. En observant les expressions faciales, nous pouvons souvent déterminer si quelqu'un est heureux, en colère, triste ou surpris.

La reconnaissance des émotions faciales (FER) est un domaine de recherche en pleine expansion qui suscite un intérêt croissant ces dernières années. Son objectif est de développer des systèmes capables d'identifier et d'interpréter les émotions exprimées par les visages humains. Cette technologie est cruciale dans de nombreux domaines, notamment l'interaction homme-machine, le marketing, la médecine, la sécurité et l'éducation.

En raison de l'importance de FER, divers algorithmes d'apprentissage automatique ont été développés pour améliorer la précision et la fiabilité de ces systèmes. Parmi les techniques basées sur le machine learning, les machines à vecteurs de support (SVM) ont été largement utilisées pour analyser les caractéristiques faciales et déterminer les émotions exprimées [1]. Ainsi, les méthodes comme le WKNN et les Méthode des k plus proches voisin sont également été explorées pour ce type d'analyse. [2]

Malgré les progrès significatifs réalisés dans ce domaine, plusieurs défis persistent, notamment en ce qui concerne la précision et la fiabilité des systèmes de reconnaissance des émotions faciales. Des améliorations peuvent être envisagées, et le deep learning constitue une alternative sérieuse pour la reconnaissance des émotions faciales. En effet, cette technologie, qui s'appuie sur les réseaux de neurones, a démontré des performances impressionnantes.

A cet effet, dans notre travail, nous proposons un modèle basé sur le deep learning pour détecter les émotions à partir d'images faciales. Pour ce faire, nous utiliserons deux bases de données, JAFFE et CK+, qui contiennent des images faciales exprimant les 7 émotions de base. Notre approche consistera à effectuer une classification afin que notre modèle puisse identifier les émotions à partir d'un ensemble de tests. Nous assistons à une utilisation massive des réseaux de neurone convolutif, qui sont structurés de plusieurs couches principales qui jouent un rôle spécifique dans l'apprentissage et l'interprétation des caractéristiques et la classification des images d'entrée. Notre objectif est de proposer des améliorations visant à accroître la précision de la reconnaissance.

## INTRODUCTION

---

Ce mémoire est structuré en quatre chapitres :

Dans le premier chapitre, nous présentons les concepts de base du traitement d'image, y compris les techniques utilisées pour améliorer, analyser et traiter des images.

Le deuxième chapitre se concentre sur les états de l'art sur l'identification des émotions faciales. Nous examinons les méthodes d'acquisition et de traitement des images qui soutiennent les systèmes de reconnaissance faciale. Nous explorons les étapes essentielles nécessaires pour comprendre et interpréter les émotions exprimées par les expressions faciales, en mettant l'accent sur des aspects tels que la détection des visages, le prétraitement des images, l'extraction de caractéristiques et la classification.

Dans le troisième chapitre, nous allons nous intéresser à l'intelligence artificielle, les machines Learning ainsi que les réseaux de neurones, spécifiquement le réseau de neurone convolutif que nous allons utiliser pour la classification de nos images.

Dans le quatrième chapitre, sera consacré à la présentation de l'implémentation du CNN sur des images de notre jeu de données. En effet, nous commencerons par expliquer les bases de données utilisées. Puis, nous exposerons le logiciel utilisé ainsi que les résultats obtenus.

---

# **CHAPITRE I : Généralités sur le traitement d'image**

---

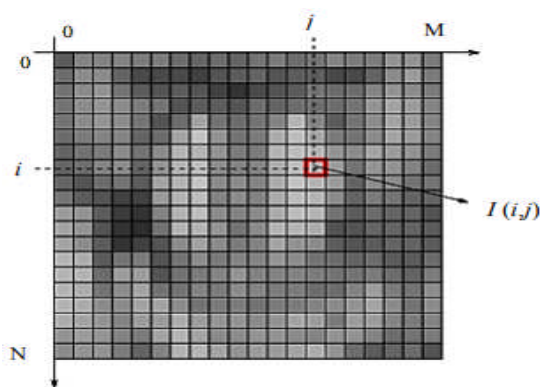
## Introduction

Dans cette section, nous abordons les bases du traitement d'image, y compris les techniques utilisées pour améliorer, analyser et traiter des images numériques. Nous discutons également des différents domaines d'application du traitement d'image et de son importance dans divers secteurs tels que la médecine, la surveillance et la vision par ordinateur. Cette introduction fournit un aperçu des concepts clés du traitement d'image et de la manière dont il peut être utilisé pour résoudre des problèmes concrets et apporter des améliorations significatives dans différents domaines.

### I.1. Définition d'image:

Une image est une représentation visuelle d'un objet ou d'une scène. En traitement d'image, une image numérique est composée de pixels, qui sont les éléments de base de l'image. Chaque pixel contient des informations sur la couleur et l'intensité de la lumière à un emplacement spécifique de l'image. Les images peuvent être en niveaux de gris, en couleurs ou en noir et blanc. Elles peuvent être capturées par des caméras numériques, des scanners ou générées par des logiciels de conception. Comprendre la nature des images est essentiel pour les manipuler efficacement dans le cadre du traitement d'image.

Projection 2D d'une scène 3D. Définie comme une fonction  $f(x, y)$ . Pour chaque  $(x, y)$ ,  $I(x, y)$  représente l'intensité de la lumière en ce point [3], tel qu'exposé dans la **Figure I.1**



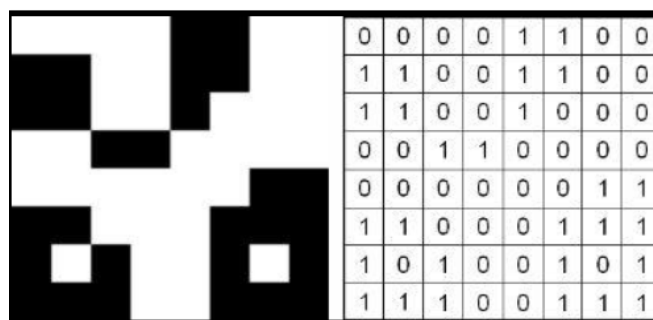
**FIGURE I.1** : Notations utilisées pour une image numérique.

### I.1.1. Types d'images :

Les images peuvent être classifiées en fonction de leur contenu et de la façon dont les informations de couleur et de luminosité sont représentées. Voici les principaux types d'images :

#### I.1.1.1. Image binaire :

Une image binaire est composée de pixels qui ne peuvent prendre que deux valeurs distinctes, généralement noir et blanc ( $I(x, y) \in \{0, 1\}$ ). Chaque pixel de l'image est représenté par un bit, ce qui signifie qu'il ne peut être que soit complètement noir (représentant la valeur 0) ou complètement blanc (représentant la valeur 1). (**Figure I.2**) Ce type d'image est couramment utilisé pour la représentation de textes et de dessins simples. Les images binaires sont également utilisées dans la reconnaissance de formes et la vision par ordinateur pour identifier clairement les contours et les formes des objets.



**FIGURE I.2** : Image binaire, avec le tableau de valeurs correspondant.

#### I.1.1.1. Image en couleurs :

Une image couleur est une image qui utilise un modèle de couleur pour représenter et afficher des couleurs. Les images couleur sont généralement composées de trois canaux de couleur: rouge, vert et bleu (RVB). Chaque pixel de l'image contient des valeurs pour ces trois canaux  $I_R(x, y)$ ,  $I_V(x, y)$ ,  $I_B(x, y)$ , ce qui permet de créer une large gamme de couleurs (**Figure I.3**). Les images couleur sont largement utilisées dans la photographie, la vidéo, la conception graphique et d'autres domaines où des couleurs précises sont nécessaires pour représenter des scènes réelles ou imaginaires. Les formats de fichier courants pour les images couleur incluent JPEG, PNG et TIFF. Ces images offrent une qualité visuelle élevée et une représentation réaliste des couleurs, ce qui les rend indispensables dans de nombreux contextes visuels.

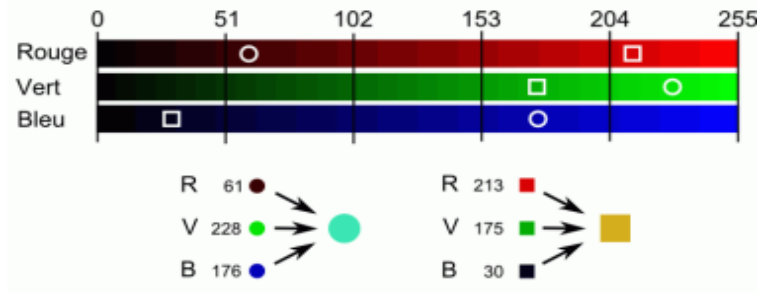


FIGURE I.3 : Représentation en niveaux RVB.

**I.1.1.2. Image en niveaux de gris :**

Une image en niveau de gris est une image composée de nuances de gris, allant du noir au blanc (La valeur zéro représente le noir (intensité lumineuse nulle) et la valeur 255 le blanc (intensité lumineuse maximale)).

Contrairement à une image binaire, qui ne comporte que des pixels noirs et blancs, une image en niveau de gris peut contenir toute une gamme de valeurs de gris. Ces images sont couramment utilisées dans des domaines tels que la photographie médicale, où les nuances de gris sont utilisées pour représenter les différentes densités des tissus et des organes. Les images en niveau de gris sont souvent utilisées pour transmettre des informations plus nuancées que ce qui est possible avec une image binaire.

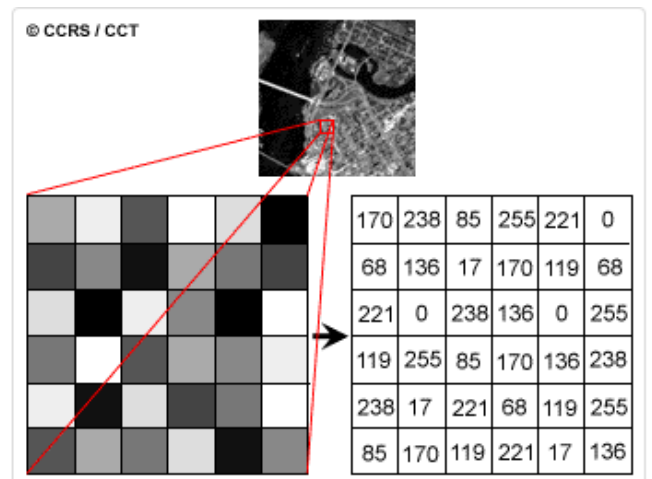
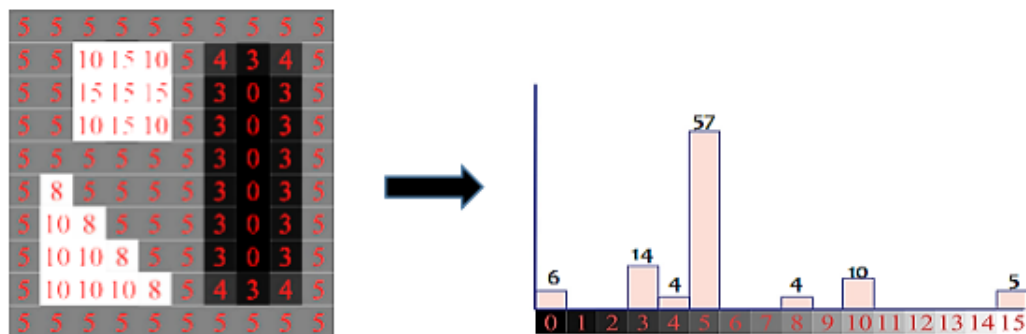


FIGURE I.4 : Image en niveaux de gris, avec le tableau de valeurs correspondant.

### I.2. Histogramme d'une image :

L'histogramme est un outil précieux qui permet de répartir les niveaux de gris en un ensemble d'intervalles de même taille. Il analyse les niveaux de gris d'une image en les regroupant par tranches de luminosité égales. On compte ensuite le nombre de pixels dans chaque tranche. Un graphique

De la (**Figure I.5**) montre la répartition des pixels en fonction de la luminosité moyenne de chaque tranche. Un pic dans le graphique indique une zone de l'image avec beaucoup de pixels de luminosité similaire.



**FIGURE I.5 :** Représentation d'une image en niveau de gris avec son histogramme.

L'histogramme de l'image I, la fonction H définie sur l'ensemble des entiers naturels par:

$$H(x) = \text{Card}\{P : I(P) = x\}$$

- H(x) = nombre d'apparitions du niveau de gris x dans l'image I.
- Outil privilégié en analyse d'images.

### I.3. Analyse des Caractéristiques d'Images en Traitement d'Image :

#### I.3.1. Définition du traitement d'image :

Le traitement d'image se réfère à l'ensemble des techniques et des algorithmes utilisés pour modifier ou améliorer des images numériques. Il englobe des opérations telles que le filtrage, la segmentation, la reconnaissance d'objets et l'analyse d'image. En somme, le traitement d'image vise à extraire des informations pertinentes des images et à les traiter de manière à en faciliter l'interprétation ou l'analyse ultérieure. C'est un domaine complexe mais passionnant qui a de nombreuses applications pratiques dans divers domaines.

**I.3.1.1. Historique et évolution du traitement d'image :**

L'histoire du traitement d'image est riche en innovations et en progrès technologiques. Des premières expérimentations en photographie aux algorithmes sophistiqués d'aujourd'hui, l'évolution du traitement d'image est marquée par des avancées majeures. Grâce aux contributions de chercheurs et d'ingénieurs passionnés, nous disposons désormais d'outils puissants pour analyser, modifier et interpréter les images. Comprendre l'histoire du traitement d'image nous permet d'apprécier tout le chemin parcouru et d'envisager les futurs développements de cette discipline fascinante.

**I.3.1.2. Applications du Traitement d'Image :**

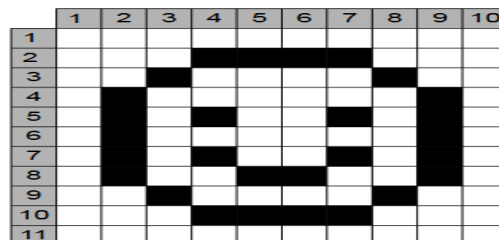
Les applications du traitement d'image sont vastes et diverses, allant de l'imagerie médicale à la reconnaissance faciale et à la vision par ordinateur. Les technologies de reconnaissance faciale sont de plus en plus utilisées pour la sécurité, tandis que l'imagerie satellite et la télédétection trouvent des applications dans la géographie et l'environnement. Le traitement d'image est également crucial dans la restauration et amélioration de la qualité des images, extraire des informations précieuses, et faciliter la prise de décisions basées sur des données visuelles, les applications industrielles pour le contrôle qualité et l'automatisation, ainsi que pour les effets spéciaux dans l'art et la créativité.

**I.4. Concept de base :**

Pour bien comprendre le traitement d'image, il est crucial de maîtriser concepts de base. Les pixels, la résolution, dimension, contours, texture, luminance, bruit

**I.4.1. Pixel :**

Une image numérique est constituée d'un ensemble de points appelés pixels (abréviation de PICTure Element) pour former une image. Le pixel représente ainsi le plus petit élément constitutif d'une image numérique



**FIGURE I.6 :** Image numérique qui indique le nombre de colonnes et de lignes.

Le nombre total de pixels dans cette image :

$$\text{Nombre total des pixels} = \text{colonnes} * \text{lignes} \dots\dots\dots (1)$$

**Exemple:** 10x11= 110 pixels au total pour l'image ci-dessus

**I.4.2. Dimension :**

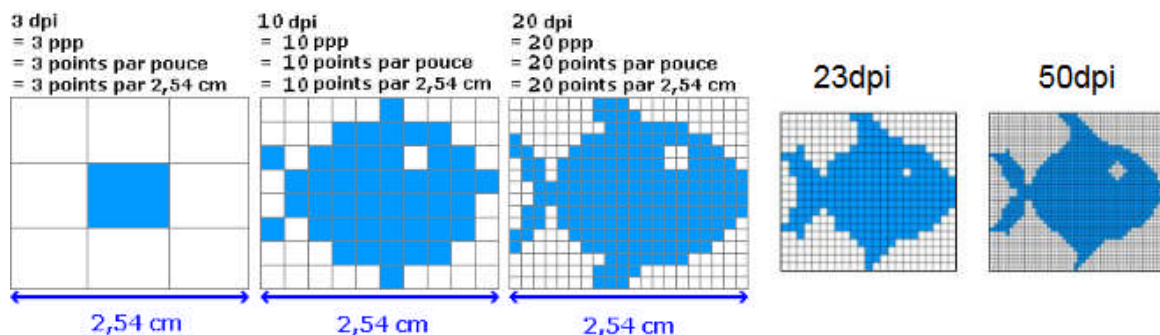
C'est la taille de l'image. Cette dernière se présente sous forme de matrice dont les éléments sont des valeurs numériques représentatives des intensités lumineuses (pixels). Le nombre de lignes de cette matrice multiplié par le nombre de colonnes nous donne le nombre total de pixels dans une image.

**I.4.3. Résolution et échantillonnage :**

La résolution et l'échantillonnage sont des aspects clés du traitement d'image qui déterminent la qualité et la taille des images, Il est essentiel de comprendre comment les ajuster pour répondre aux exigences spécifiques d'un projet. Ces paramètres influencent directement la netteté et la clarté des images traitées.

**I.4.4. Résolution :**

La résolution d'une image fait référence au nombre de pixels qu'elle contient. Plus la résolution est élevée, plus l'image sera détaillée et de meilleure qualité. La résolution est généralement exprimée en pixels par pouce (ppp) ou dots per inch (dpi)



**FIGURE I.7 :** Image avec différente résolution.

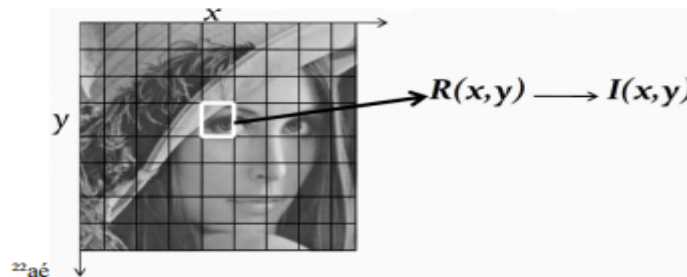
La résolution d'une image numérique établit la relation entre la quantité de pixels dans l'image et sa dimension réelle sur un support physique.

$$\text{Résolution} = \frac{\text{Quantité de pixels (largeur)}}{\text{dimension (largeur)}} \dots\dots\dots (2)$$

**I.4.5. L'échantillonnage :**

L'échantillonnage concerne la discrétisation spatiale des pixels.

L'échantillonnage consiste à associer à chaque zone rectangulaire  $R(x,y)$  d'une image continue une unique valeur  $I(x,y)$  (un pixel).



**FIGURE I.8 :** Image d'un visage fragmenté : l'échantillonnage dévoile la construction de l'image numérique.

**I.4.6. Contraste :**

Le contraste en imagerie numérique se réfère à la différence de luminosité ou de couleur entre les différents éléments d'une image. Il est crucial pour rendre une image claire et facile à interpréter. Un contraste élevé signifie des différences marquées entre les tons clairs et foncés, tandis qu'un contraste faible peut entraîner une image terne et peu détaillée. Pour améliorer le contraste d'une image, on peut ajuster les niveaux ou utiliser des techniques de retouche telles que l'égalisation d'histogramme. Comprendre et manipuler le contraste est essentiel pour obtenir des images de haute qualité dans le domaine du traitement d'images en informatique.

### I.4.7. Luminance :

La luminance est une mesure quantitative de la luminosité perçue par l'œil humain. Elle est basée sur l'intensité lumineuse d'une source et prend en compte la sensibilité de l'œil à différentes longueurs d'onde. En d'autres termes, la luminance représente à quel point une surface est lumineuse ou sombre pour un observateur donné. Dans une image, la luminance peut varier en fonction des conditions d'éclairage, des propriétés de surface et de la gamme de couleurs. Comprendre la luminance dans une image est essentiel pour ajuster l'affichage et améliorer la lisibilité des informations visuelles.

Une bonne luminance se caractérise par [4]:

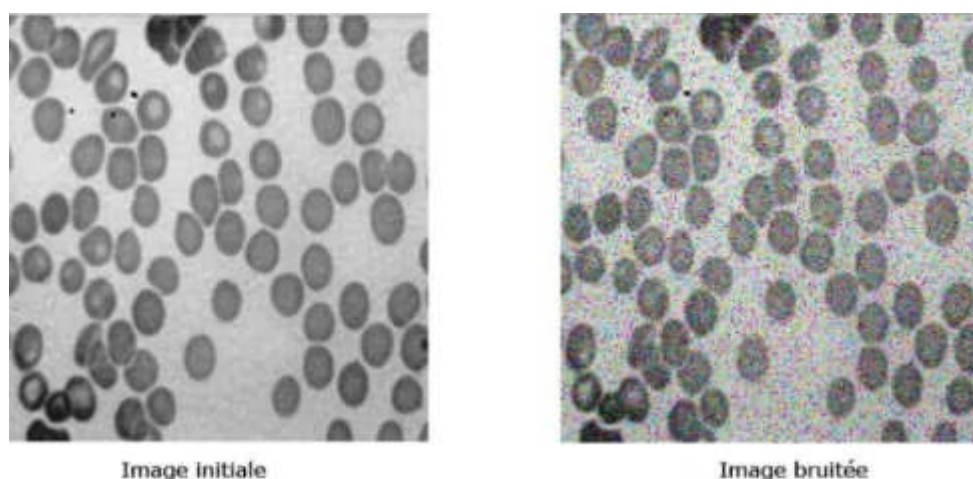
- ❖ Des images lumineuses (brillantes);
- ❖ Un bon contraste : il faut éviter les images où la gamme de contraste tend vers le blanc ou le noir; ces images entraînent des pertes de détails dans les zones sombres

Ou lumineuses.

- ❖ L'absence de parasites.

### I.4.8. Bruit (parasite) :

Dans le domaine du traitement du signal et de l'image, le bruit correspond à un phénomène aléatoire qui se surajoute à l'image idéale. [5] considéré comme une brusque variation de l'intensité d'un pixel par rapport à ses voisins, il provient de l'éclairage des dispositifs optiques et électroniques du capteur. la (**Figure I.9**) représente une image avec (original) et sans bruit :



**FIGURE I.9 :** Image avec et sans bruit.

### I.4.9. Contours :

Dans le domaine du traitement d'image, la détection de contours est essentielle pour identifier les limites et les formes des objets dans une image.

Les contours définissent les transitions de luminosité, les changements brusques de couleur, de texture ou de texture qui permettent de distinguer les différentes parties de l'image (comme montré sur la figure ci-dessous).



**FIGURE I.10 :** Image avant et après détection de contour

En utilisant des techniques telles que la détection de contours par gradient ou par différence de luminosité, il est possible d'extraire des informations clés pour la segmentation et l'analyse ultérieure de l'image.

### I.4.10. Texture :

Le terme «texture» est largement connu et admis dans notre langage, il est difficile de construire une définition formelle.

La texture en traitement d'image se réfère à la qualité visuelle et tactile de la surface d'un objet représenté dans une image. Elle se caractérise par des motifs répétitifs ou aléatoires qui peuvent être lisse, fins ou grossiers. La définition de texture en traitement d'image consiste à décrire ces motifs de manière quantitative pour permettre une analyse et une interprétation automatique des images.

Comprendre la texture d'une image peut aider à identifier des régions d'intérêt, à diagnostiquer des maladies médicales et à améliorer la qualité des produits.

### I.5. Prétraitement des Images :

Le prétraitement des images est une étape essentielle dans le domaine du traitement d'images. Il s'agit de préparer les images brutes en effectuant des opérations telles que la correction des défauts, et la normalisation. Ce processus permet d'optimiser la qualité des données et d'améliorer la performance des algorithmes de traitement d'images.

Les techniques de pré-traitements les plus courantes qu'on va présenter sont :

- ❖ La modification d'histogramme,
- ❖ La réduction du bruit par filtrage.

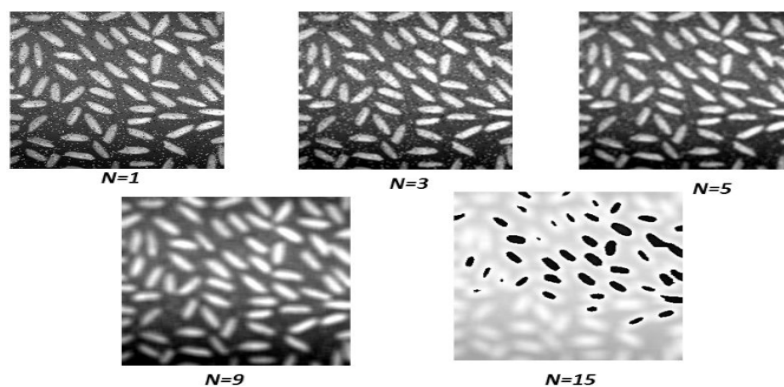
### I.6. Filtrage et Lissage :

Le filtrage et le lissage sont des techniques utilisées pour éliminer le bruit et les imperfections des images. Le filtrage consiste à appliquer des filtres pour supprimer les éléments indésirables, tandis que le lissage vise à adoucir les contours et à rendre l'image plus homogène.

Voici quelques exemples de filtres réducteurs de bruit :

#### I.6.1. Filtre moyenneur :

Le filtre moyenneur est un type de filtre linéaire largement utilisé en traitement d'image pour lisser une image en réduisant le bruit. En moyennant les valeurs des pixels voisins, ce filtre permet d'atténuer les variations brusques de luminosité tout en conservant les contours et les structures importantes de l'image. Son principal avantage réside dans sa simplicité et son efficacité pour améliorer la qualité des images en éliminant le bruit parasite de manière uniforme. (**Figure I.11**) montre le lissage par moyenneur avec des densités différentes :



**FIGURE I.11** : Le lissage par moyennage.

Plus  $N$  est grand, plus le lissage sera important, et plus l'image filtrée perd les détails de l'image originale.

### I.6.2. Filtre gaussien :

Le filtre gaussien est un outil essentiel en traitement d'image pour réduire le bruit et lisser une image tout en préservant les contours et les détails. (**Figure I.12**) représente l'application du filtre gaussien sur des images bruitées avec des densités différentes :

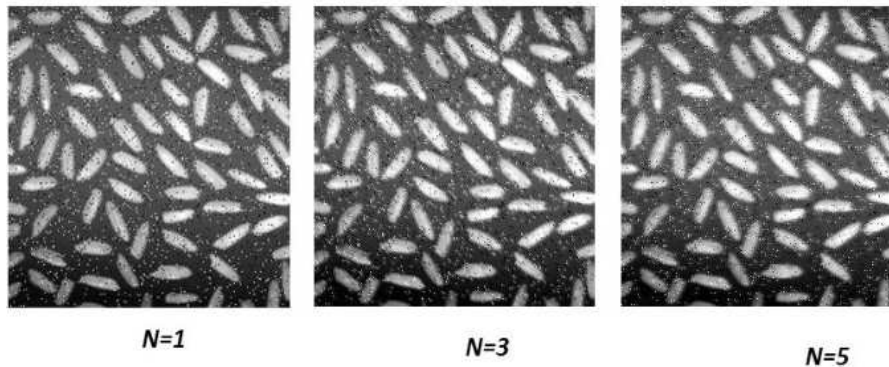


FIGURE I.12 : Le lissage par filtre gaussien.

### I.6.3. Filtre median :

Le filtre médian est une méthode de filtrage d'images qui consiste à remplacer la valeur de chaque pixel par la médiane des valeurs des pixels autour de lui. Cette technique est particulièrement efficace pour éliminer le bruit salt-and-pepper présent dans les images comme montre la (**Figure I.13**) :

En comparaison avec d'autres filtres, le filtre médian permet de conserver les contours et les détails de l'image tout en lissant les variations de luminosité.

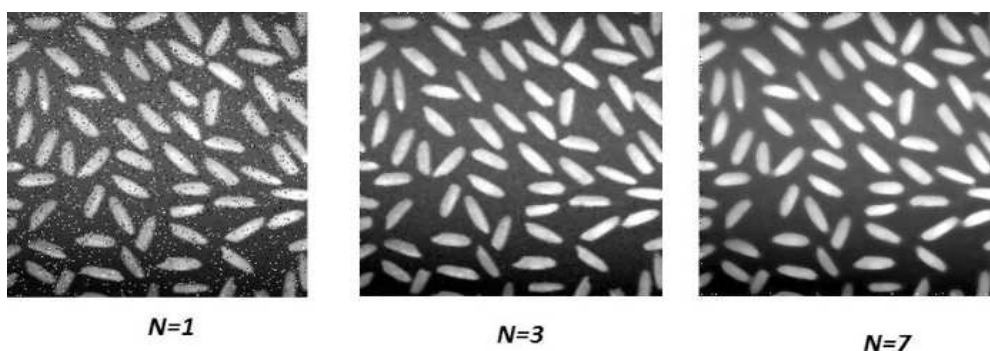


FIGURE I.13 : Lissage par filtre median.

## I.7. La modification d'histogramme :

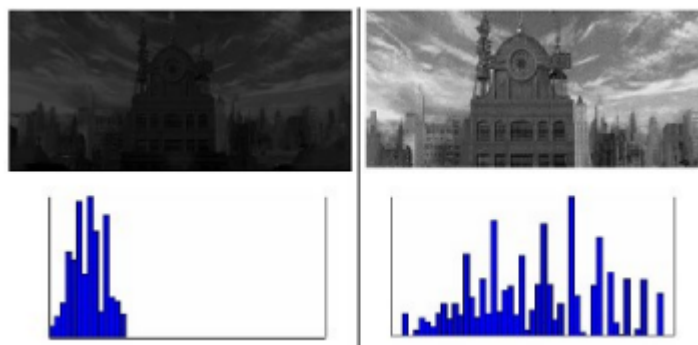
La modification d'histogramme est une technique essentielle en traitement d'image qui permet de rééquilibrer la répartition des niveaux de gris d'une image. Cette méthode permet d'obtenir des images plus nettes, plus claires et plus fidèles à la réalité.

Dans ce qui suit, on va présenter deux types de modification d'histogramme :

- ❖ Expansion de dynamique.
- ❖ Egalisation d'histogramme.

### I.7.1. Expansion de dynamique :

L'expansion de dynamique dans le traitement d'images vise à améliorer le contraste des images en accentuant les différences de luminosité entre les zones sombres et claires. Cela permet de rendre les détails plus visibles et de mieux distinguer les différents éléments de l'image. L'une des méthodes les plus courantes pour l'expansion de dynamique est la transformation de fonctions, où les niveaux de gris de l'image sont ajustés pour étendre la plage de luminosité. Comme indiqué dans la (**Figure I.14**).



**FIGURE I.14 :** Expansion de la dynamique.

Cette technique est particulièrement utile pour améliorer la qualité des images mal exposées ou présentant un faible contraste, et peut être appliquée de manière automatique ou manuelle pour obtenir des résultats optimaux.

### I.7.2. Égalisation d'histogramme :

L'égalisation de l'histogramme consiste à équilibrer le mieux possible la distribution des pixels dans la dynamique. L'idéal est d'obtenir un histogramme plat où l'on affecte le même nombre de pixels pour chaque niveau de gris (ceci étant impossible à réaliser pour les images numérique à cause de la nature discrète de l'histogramme et de la quantification des niveaux de gris (qui nous limite à un nombre fini de niveaux)).

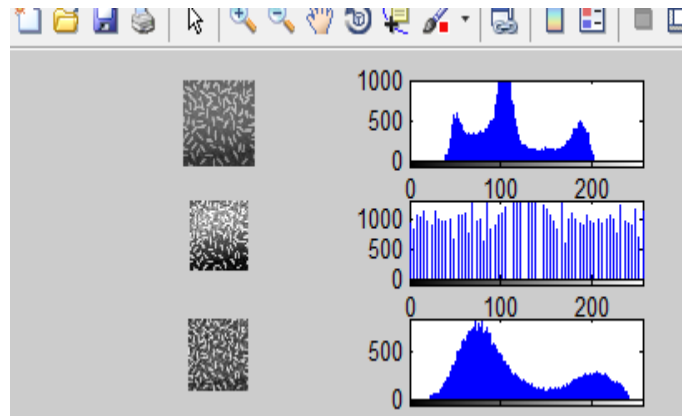


FIGURE I.15 : Image avec son histogramme égalisé.

## Conclusion

Nous avons introduit dans ce chapitre les notions de base qui servent de fondement à la compréhension de différentes techniques de traitement d'images. Plusieurs méthodes classiques de traitement ont été proposées dans la littérature, nous avons présenté quelques unes qui nous semblent les plus courantes dans le processus du traitement et analyse d'image.

En partant des bases, nous avons défini ce qu'est une image numérique et exploré ses différentes formes, de l'image binaire à l'image en couleurs. Nous avons également discuté des concepts clés tels que les pixels, la résolution, le contraste, la luminance, le bruit, les contours et la texture, qui sont tous cruciaux pour une compréhension approfondie et une manipulation efficace des images.

---

## **CHAPITRE II : États de l'art sur l'identification des émotions faciale.**

---

## **Introduction**

Dans ce chapitre, nous explorons les méthodes d'acquisition et de traitement des images qui sous-tendent les systèmes de reconnaissance faciale. De la capture initiale des visages à travers diverses technologies à l'extraction minutieuse des caractéristiques distinctives du visage, nous examinons les étapes essentielles nécessaires pour comprendre et interpréter les émotions exprimées par les expressions faciales. En mettant l'accent sur des aspects tels que la détection des visages, le prétraitement des images, l'extraction de caractéristiques et la classification, nous dévoilons les fondements techniques qui alimentent les avancées continues dans le domaine de la reconnaissance des émotions faciales (**FER**).

### **II.1. Contexte et objectifs :**

#### **II.1.1. Présentation générale du projet de (FER):**

Le projet de reconnaissance des émotions faciales est une initiative visant à développer des systèmes capables d'identifier et d'interpréter les émotions exprimées par les visages humains. Cette technologie repose souvent sur l'utilisation de l'intelligence artificielle et de l'apprentissage automatique pour analyser les caractéristiques faciales telles que les expressions, les mouvements musculaires, afin de déterminer l'état émotionnel d'une personne.

#### **II.1.2. Intérêt et applications de cette technologie :**

La reconnaissance des émotions faciales présente un intérêt significatif dans de nombreux domaines d'application. Tout d'abord, dans le domaine de la publicité et du marketing, elle peut être employée pour évaluer les réactions des consommateurs face à des produits ou des publicités, aidant ainsi les entreprises à mieux comprendre et répondre aux besoins de leur clientèle. De plus, dans le secteur de la santé mentale, la reconnaissance des émotions faciales peut être utilisée pour évaluer l'état émotionnel des patients et aider les professionnels de la santé à diagnostiquer et à traiter des troubles tels que la dépression, l'autisme...

Dans le domaine de la sécurité, cette technologie peut être utilisée pour détecter des comportements suspects ou menaçants en analysant les expressions faciales des individus.

## **II.2. Émotion et expression faciale:**

L'importance du visage et de ses expressions dans la communication non verbale est telle qu'au sein des sciences psychologiques de l'émotion, la question de l'expression faciale a toujours constitué une spécialité à part entière. [6]

La reconnaissance faciale est en fait l'un des nombreux types d'analyse d'images du domaine de la vision par ordinateur qui permet de constituer et de reconnaître automatiquement les personnes à partir d'images de leurs visages. La reconnaissance faciale utilise des algorithmes qui analysent les visages existants Images ou vidéos pour extraire un ensemble de caractéristiques distinctives.

### **II.2.1. Définition des émotions :**

Le mot « émotion » provient du mot français « émouvoir ». Elle a d'abord une manifestation interne et génère une réaction extérieure. Elle est provoquée par la confrontation à une situation et à l'interprétation de la réalité. [7]

Une émotion est une réaction physiologique et psychologique du corps face à une situation, qui le met en action. Elle agit un peu comme un réflexe : on détecte les émotions avant même d'en avoir conscience. Il n'y a donc ni émotions négatives ou positives, simplement des réactions de notre corps face à des stimuli. [8]

### **II.2.2. Types d'émotions :**

Les émotions sont importantes dans notre vie. On les divise souvent en trois types : les émotions de base, les émotions plus complexes et les émotions qui concernent nos relations avec les autres. Cette classification nous aide à mieux comprendre nos réactions et nos relations avec le monde qui nous entoure.

### II.2.2.1. Émotions primaire :

Les émotions primaires sont les premières réactions émotionnelles naturelles que nous ressentons en réponse à des événements. Elles sont directement liées à la situation a ce qui ce passe autour de nous. Selon Paul Ekman, les six émotions primaires universelles sont la colère, le dégoût, la peur, le bonheur, la tristesse et la surprise. Tel qu'illustré dans la (**Figure II.1**) :

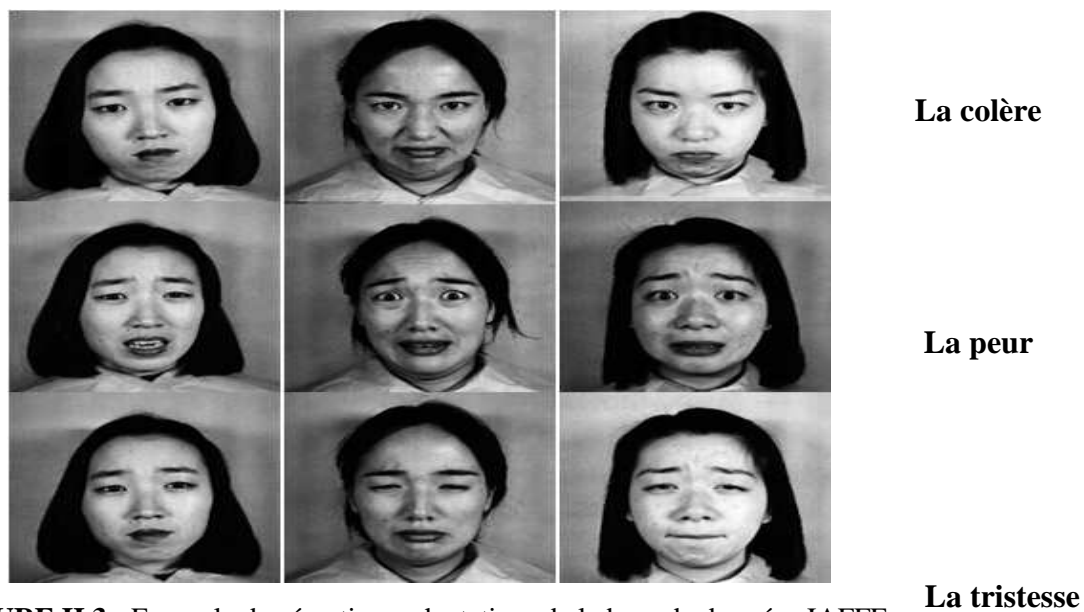


**FIGURE II.1** : Les sept expressions de base définies par Ekman de la base de données JAFFE.

Cependant, ces émotions peuvent être subdivisées en deux groupes les émotions primaires adaptatives et les émotions primaires non adaptatives :

#### ➤ Les émotions adaptatives :

Les émotions adaptatives jouent un rôle crucial dans notre survie et notre bien-être. Elles nous aident à réagir rapidement et de manière appropriée aux différents stimuli de notre environnement. Par exemple, la peur face à un danger, la colère face à une injustice, la tristesse face à une perte. La (**Figure II.2**) illustre les émotions adaptatives à savoir la colère, la peur et la tristesse :



**FIGURE II.3 :** Exemple des émotions adaptatives de la base de données JAFFE.

➤ **Les émotions non adaptatives :**

Les émotions non adaptatives peuvent nuire à notre santé mentale. Par exemple, être trop bouleversé ou stressé peut être mauvais pour nous. (Figure II.3) montre les exemples des émotions non adaptatives :



**FIGURE II.4 :** Exemple des émotions non adaptatives de la base de données JAFFE.

### **II.2.2.2. Émotions secondaires :**

Ce sont des émotions qui se produisent après les émotions primaires et qui sont un peu plus compliquées, elles sont souvent influencées par notre façon de penser et notre expérience.

Les émotions secondaires varient d'une personne à l'autre et d'une culture à l'autre, incluant des sentiments tels que la culpabilité, la frustration et la confusion.

Par exemple, la culpabilité est une émotion secondaire qui dérive de la tristesse primaire. Quand on fait quelque chose de mal, on ressent d'abord de la tristesse, puis cette tristesse se transforme en culpabilité, une émotion plus complexe qui implique un jugement moral sur notre comportement.

### **II.2.2.3. Émotions sociales :**

Ce sont les émotions liées à la façon dont nous interagissons avec les autres. Elles peuvent créer un sentiment collectif qui s'impose à tous, en nous mettant tous sur la même longueur d'onde.

Sont souvent influencées par notre compréhension des normes sociales, des attentes sociales et des relations interpersonnelles.

Par exemple, ressentir ce que quelqu'un d'autre ressent (empathie), ou le respect envers les autres. Ces émotions sont importantes dans nos relations avec les autres et dans la société en général.

### **II.2.3. Les expressions faciales :**














L'expression faciale est un ensemble de signes du visage qui traduisent un sentiment de changement dans le visage (mimique faciale chargée de sens), perceptible visuellement, dû à l'activation (volontaire ou non) de l'un ou de plusieurs muscles composant le visage.[9]

Les expressions faciales sont très importantes pour pouvoir connaître l'état de la personne. Grâce à ces expressions, il est possible de faire plusieurs déductions et récupération de plusieurs informations comme :

- L'état affectif que ce soit les émotions (peur, colère, joie, surprise, tristesse, Dégoût) ou bien certaines humeurs.
- L'activité cognitive comme la concentration, l'ennui ou la perplexité.
- Le tempérament et la personnalité. [10]

### II.3. Le système de codage des (FACS) :

Le visage humain est un vecteur d'expression complexe et fascinant, qui nous permet de communiquer une multitude d'émotions sans prononcer un seul mot. Le FACS (**Facial Action Coding System**) est utilisée dans le domaine de l'intelligence artificielle. En outre, le FACS sert à améliorer les modèles de reconnaissance faciale. Le FACS II a été développé par les psychologues Paul Ekman et Wallace V. Friesen en 1978, c'est un système de classification et de mesure des mouvements faciaux. Il repose sur l'analyse des muscles du visage et de leurs contractions, appelées "unités d'action" (AU). Le FACS compte 46 unités d'action, qui sont combinées pour former des expressions faciales complexes (Chaque AU est attribué un numéro et un nom). [11]

AU	Nom	Exemple
1	Remontée de la partie interne des sourcils	
2	Remontée de la partie externe des sourcils	
4	Abaissement et rapprochement des sourcils	
5	Ouverture entre la paupière supérieure et les sourcils	
6	Remontée des joues	
7	Tension de la paupière	
9	Plissement de la peau du nez vers le haut	
12	Étirement du coin des lèvres	
15	Abaissement des coins externes des lèvres	
16	Ouverture de la lèvre inférieure	
20	Étirement externe des lèvres	
23	Tension refermante des lèvres	
26	Ouverture de la mâchoire	

**Tableau I:** Les AUs fréquentes décodées par le manuel FACS sur la base d'images CK+.

Emotion	Action Units
Joie	6+12
Tristesse	1+4+15
Surprise	1+2+5+16
Peur	1+2+4+5+20+26
Colère	4+5+7+23
Dégout	9+15+26

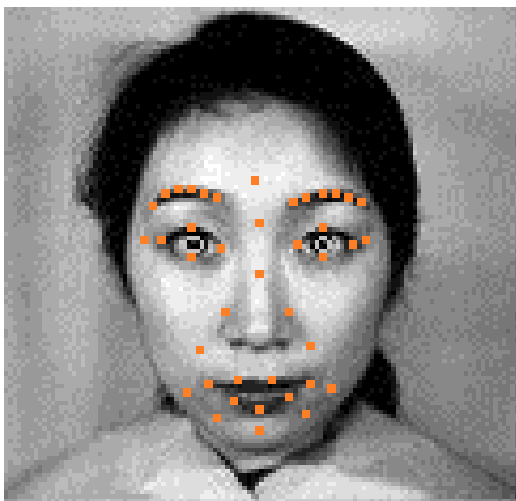
**Tableau II :** Combinaison des AUs correspondant à chaque émotion.

#### **II.4. Les différentes approches utilisées pour un système (FER)**

Différentes méthodes ont été employées pour développer des systèmes de Reconnaissance Faciale d'Émotions (**REF**) automatiques, regroupées généralement en trois approches principales : conventionnelle, géométrique et basée sur les caractéristiques d'apparence, Les points communs de ces approches sont la détection de la région faciale et l'extraction des caractéristiques :

##### **II.4.1. Approche conventionnelle :**

Pour les systèmes (**REF**) automatiques conventionnelles Ils passent par trois étapes principales. Premièrement, une image de visage est détectée à partir d'une image d'entrée. Deuxièmement, l'extraction des caractéristiques soit l'utilisation des caractéristiques géométriques présentent dans la (**Figure II.4**) ou d'apparence présentent dans la (**Figure II.5**) ou un hy-bride des caractéristiques géométriques et d'apparence sont détectés à partir de la région de visage. Troisièmement, les classifieurs des émotions tels que les réseaux de neurones conventionnels, sont utilisés pour produire les résultats de la reconnaissance à l'aide des caractéristiques extraites. [12]

**FIGURE II.5 :** Caractéristiques géométrique.**FIGURE II.6 :** Caractéristiques d'apparences.

#### **II.4.1.1. Approche géométrique :**

Dans cette approche, Les distances entre les points distincts entre deux images, l'une neutre et l'autre expressive, sont calculées afin de détecter les mouvements. [13] Représentent la forme et l'emplacement des composants du visage (y compris la bouche, les yeux, les sourcils et le nez). Les composants faciaux ou les traits faciaux sont extraits pour former un vecteur de caractéristiques représentant la géométrie du visage.

#### **II.4.1.2. Approche basé sur les caractéristiques d'apparence :**

Ces approches appliquent généralement des techniques d'apprentissage automatique. Ainsi, les modèles sont appris à partir d'un ensemble d'images représentatives de la variabilité de l'aspect facial. Ces modèles sont alors employées pour la détection. [13]

Le principe de ces méthodes est d'utiliser toute la surface du visage ou des régions spécifiques du visage pour représenter les changements d'apparence. [14]

#### **II.4.2. Approche basées sur l'apprentissage profond :**

L'approche basée sur l'apprentissage profond consiste à utiliser des algorithmes d'apprentissage automatique pour analyser de grandes quantités de données et apprendre à identifier les états émotionnels à partir de l'expression faciale. Cette approche permet d'obtenir des résultats plus précis et fiables.

## II.5. Méthodes d'acquisitions et traitement des images :

Les systèmes de reconnaissance faciale passent par plusieurs étapes de traitement, depuis l'acquisition de l'image jusqu'à l'identification de la personne. Voici les principales étapes illustrées dans la (Figure II.6) suivante :

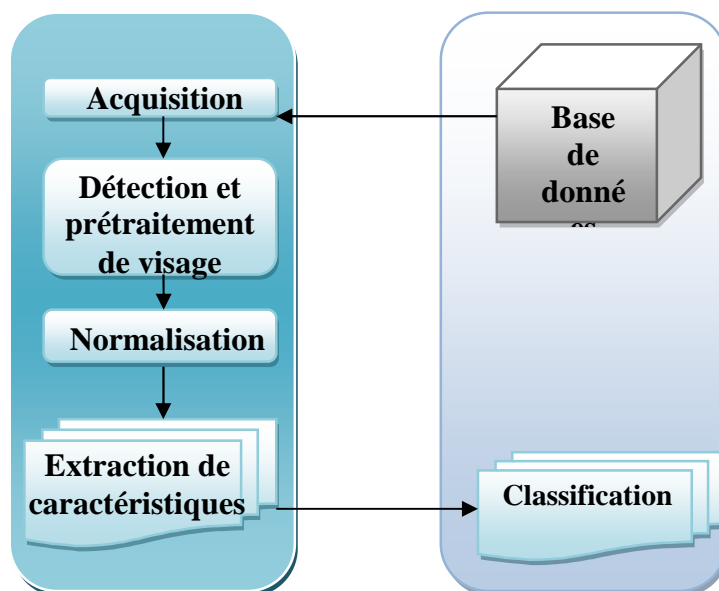


FIGURE II.7 : Schéma générale d'un système de reconnaissance des visages.

### II.5.1. Acquisitions de visage :

L'acquisition de visage est une étape cruciale dans la compréhension de ce domaine, elle englobe l'ensemble des processus permettant de capturer des images à partir de différents dispositifs. Il faut réussir à capter l'information pertinente sans bruit. Dans la reconnaissance de visage on peut utiliser les capteurs 3D par exemple pour s'affranchir des problèmes de pose. [15]

### II.5.1.1. Détection de visage et prétraitement :

#### ➤ **Détection :**

On peut diviser les approches de détection en quatre catégories : les méthodes basées sur la connaissance où on code la connaissance humaine du visage, les méthodes de correspondance de masques, les méthodes à caractéristiques invariables où on utilise la couleur, les textures et les contours et finalement les méthodes les plus répandues et qui sont ceux basées sur l'apprentissage ou les statistiques. [15]

Le défi de détecter et enregistrer les visages implique de repérer la présence de visages dans une image et de déterminer leurs positions et tailles. La précision de cette détection et enregistrement est cruciale, surtout dans des conditions réalistes où la présence et la localisation des visages ne sont pas préalablement connues. Cependant, ce processus est confronté à plusieurs défis :

- **Pose :** Les visages peuvent être partiellement cachés ou déformés en raison de l'angle ou de la position de la tête par rapport à la caméra.
- **Occlusion :** Les parties du visage peuvent être cachées par des accessoires comme des lunettes, des barbes, ou des moustaches. De plus, le maquillage peut altérer l'apparence naturelle du visage, rendant certaines régions difficiles à détecter.
- **Expressions faciales :** Les visages présentent une grande variabilité dans leur forme selon les expressions faciales. Certaines caractéristiques peuvent être masquées ou modifiées selon l'expression du visage.
  
- **Conditions d'acquisition d'image :** Les variations d'éclairage et les caractéristiques de la caméra peuvent altérer la chrominance du visage. Les zones du visage peuvent être ombrées ou éclairées de manière inégale, ce qui rend la détection des caractéristiques faciales plus difficile.

Pour obtenir une détection et un enregistrement de visage précis, Comme le démontre la (**Figure II.7**). Les systèmes doivent être robustes face à ces défis et être capables de reconnaître les visages malgré les variations de pose, d'occlusion, d'expression et de conditions d'acquisition d'image. [16]

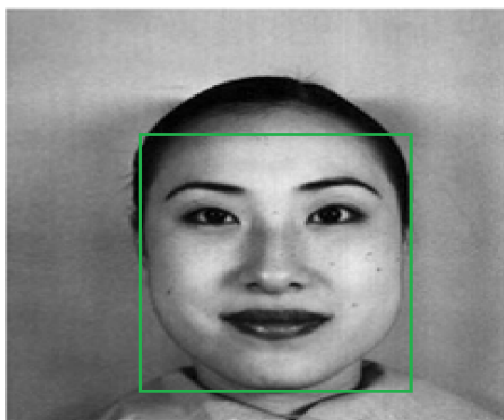


FIGURE II.8 : Détection de visage.

➤ **Prétraitement :**

Dans le monde physique, il y a trois paramètres à considérer : l'éclairage, la variation de posture et l'échelle. La variation de l'un de ces trois paramètres peut conduire à une distance entre deux images du même individu, supérieure à celle séparant deux images de deux individus différents. Le rôle de cette étape est d'éliminer les parasites causés par la qualité des dispositifs optiques ou électroniques lors de l'acquisition de l'image en entrée, dans le but de ne conserver que les informations essentielles et donc préparer l'image à l'étape suivante. Elle est indispensable car on ne peut jamais avoir une image sans bruit à cause du fond (background) et de la lumière qui est généralement inconnue. Il existe plusieurs types de traitement et d'amélioration de la qualité de l'image, telle que : la normalisation, l'égalisation d'histogramme. [15].

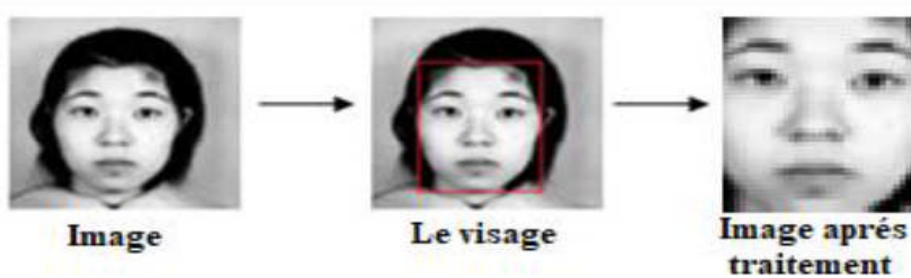


FIGURE II.9 : Acquisition et traitement du visage.

### II.5.2. La normalisation :

Le but de la normalisation dans la reconnaissance faciale est de rendre les données d'images faciales cohérentes et comparables, indépendamment de facteurs tels que l'éclairage, l'angle de la caméra, l'expression faciale et les variations dans les caractéristiques individuelles. La normalisation vise à réduire les variations non pertinentes dans les images faciales, ce qui facilite ensuite la comparaison et la reconnaissance des visages.

Il existe plusieurs techniques de normalisation utilisées dans la reconnaissance faciale, par exemple :

- **Correction de l'éclairage** : Ajuster les niveaux de luminosité et de contraste pour rendre les images plus uniformes en termes d'éclairage (égalisation d'histogramme).
- **Normalisation de l'échelle** : Mettre les visages à la même taille dans les images, ce qui élimine les variations dues à la distance entre la caméra et le visage.

En normalisant les images faciales, les systèmes de reconnaissance faciale peuvent mieux extraire les caractéristiques distinctives des visages et améliorer la précision de la reconnaissance. (**Figure II.9**) illustre un exemple d'une image avant (originale) et après l'ajustement :



FIGURE II.10 : Avant et après La normalisation.

### II.5.3. Extraction de caractéristiques:

L'extraction de caractéristiques est une étape fondamentale du traitement de l'image tels que la reconnaissance faciale, l'analyse des émotions, implique l'application de calculs mathématiques sur les pixels d'une image numérique afin d'extraire des propriétés visuelles utilisées pour des traitements ultérieurs, comme la classification ou la reconnaissance.

Dans le contexte de l'extraction de caractéristiques du visage, le but est d'identifier les traits distinctifs d'un visage humain à partir des images ou de vidéos, et à les représenter sous forme de caractéristiques numériques (vecteur caractéristique). Ce processus peut se faire de différente méthode, comme les réseaux de neurones convolutifs (CNN).

Il existe généralement deux approches pour traiter le problème :

- **L'approche globale** : C'est une méthode relativement récente qui analyse le visage dans sa globalité, elle se base principalement sur la mise en correspondance de modèles de visage, et les caractéristiques plus subtiles comme les plis et les ridules.
- **L'approche par composantes** : Contrairement à l'approche globale, cette méthode se concentre sur les différentes parties du visage, telles que les yeux, le nez, la bouche, etc.

### II.5.4. Classification :

La classification est un processus qui consiste à organiser des éléments, des objets, des données ou des informations en différentes catégories ou classes en fonction de leurs caractéristiques communes (selon des critères spécifiques).

Son rôle consiste à analyser les différents mouvements et contractions des muscles du visage pour déterminer quelle émotion une personne est en train de ressentir, en analysant la configuration précise des différents éléments du visage (sourcils, yeux, bouche...), il est donc possible de déterminer quelle émotion est exprimée.

### **II.5.5. Base de données :**

Les données faciales (anciennement connues sous le nom de photo de référence) sont une photo d'une personne utilisée comme référence par le système de reconnaissance faciale pour identifier une personne. Elles sont stockées dans une base de données (c'est la "bibliothèque" de visages de référence.).

La base de données joue un rôle central, lors de la comparaison, les caractéristiques extraites du visage de test sont comparées à celles stockées dans la base de données pour trouver une correspondance.

### **Conclusion**

À travers ce chapitre, nous avons exploré en profondeur les concepts fondamentaux liés à la reconnaissance des émotions faciales, nous avons constaté que c'est un langage universel qui nous permet de communiquer nos émotions sans même parler.

Cette technologie révolutionnaire, vise à identifier et interpréter les émotions humaines en analysant les caractéristiques du visage à l'aide de l'intelligence artificielle.

Les émotions peuvent être classées en catégories primaires, secondaires et sociales, chacune influençant nos interactions avec le monde. Différentes approches, telles que conventionnelles, géométriques et basées sur l'apparence, sont utilisées pour la reconnaissance des émotions à travers les expressions faciales. Le système comporte plusieurs étapes depuis l'acquisition du visage jusqu'au traitement de l'image, en passant par la détection, le prétraitement, l'extraction des caractéristiques, la normalisation et la classification, afin de déterminer avec précision l'émotion exprimée par l'individu.

---

# **CHAPITRE III : Réseaux de neurones convolutifs et Apprentissage automatique**

---

## **Introduction :**

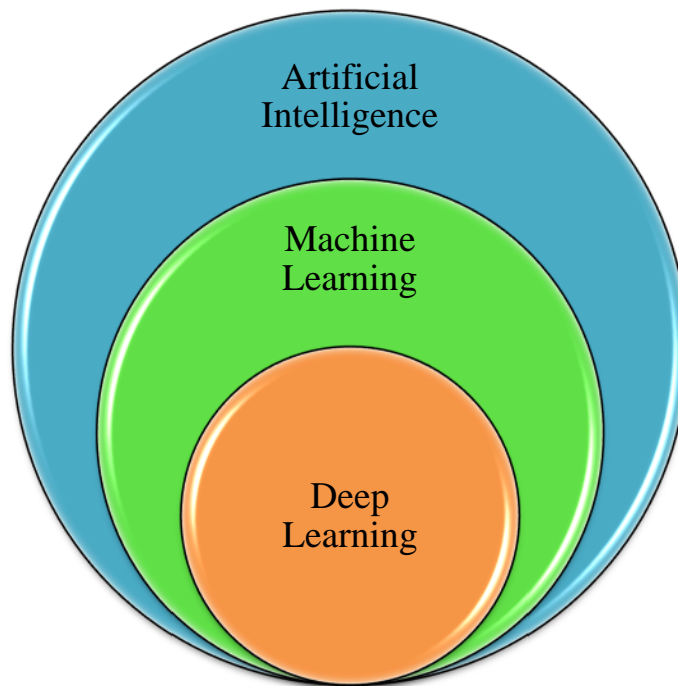
Dans ce chapitre, nous explorons les concepts fondamentaux de l'apprentissage automatique (Machine Learning), un champ d'étude qui permet aux ordinateurs de se former à résoudre des problèmes sans être explicitement programmés pour cela. Au lieu de fournir des instructions détaillées sur comment résoudre un problème, l'apprentissage automatique permet aux machines de découvrir des relations et des modèles à partir des données fournies, ce qui leur permet de prendre des décisions et de résoudre des problèmes de manière autonome. Les algorithmes d'apprentissage automatique sont divisés en trois catégories : supervisés, non supervisés et par renforcement, chacun adapté à des contextes spécifiques. Nous examinons également l'apprentissage en profondeur (Deep Learning), qui est un sous-domaine de l'apprentissage automatique qui utilise des réseaux de neurones pour extraire des caractéristiques à partir des données brutes. Les réseaux de neurones convolutionnels (CNN) sont particulièrement efficaces pour classer des images, en utilisant des filtres pour extraire des caractéristiques et des couches de regroupement pour réduire la taille des données.

### **III.1. Techniques d'apprentissage automatique**

Plutôt que d'expliquer à un ordinateur avec précision comment résoudre un problème, le Machine Learning (ou apprentissage automatique) permet de lui apprendre à résoudre un problème par lui-même. Ce champ d'étude comporte des dizaines d'algorithmes [14]. Chacun visant à permettre à la machine d'acquérir des connaissances et des compétences pour prendre des décisions et résoudre des problèmes de manière autonome qui se concentre sur l'élaboration de modèles capables de représenter certaines caractéristiques du monde qui nous entoure, d'apprendre certaines propriétés statistiques des distributions des données qu'ils traitent, afin d'accomplir diverses tâches. Le rapport à l'intelligence vient de la capacité de ces modèles à généraliser, c'est-à-dire à extraire l'information pertinente de données étudiées au fil d'un processus de mise(s)-à-jour appelé entraînement, et de savoir la réutiliser avec efficacité sur de nouvelles données jamais rencontrées auparavant.

L'idée générale est d'être capable de reproduire une relation grâce à une fonction  $f$ . Cette fonction est identifiée par un algorithme d'apprentissage. Pour aboutir à la décision ou à la prédiction  $Y$ , l'algorithme utilise en entrée un jeu de données  $X$  afin de déterminer un modèle de classification. «  $Y = f(x)$  ». [14]

Parmi les algorithmes les plus courants, on trouve les SVM (Support Vector Machine), réseaux de neurones etc... Ils opèrent dans des contextes variés : supervisé, ou non-supervisé, par renforcement. (Figure III.1) montre la relation entre Intelligence artificiel (IA), apprentissage automatique (ML) et apprentissage profond (DL). [14]



**FIGURE III.1** : La relation entre IA, ML, DL.

### **III.2. Importance de ML dans la REF :**

L'apprentissage automatique joue un rôle crucial dans la reconnaissance des émotions faciales. Voici quelques raisons pour lesquelles son importance est si significative :

#### **III.2.1. Traitement des données complexes :**

Les expressions faciales sont des données complexes et multidimensionnelles. L'apprentissage automatique permet de traiter ces données de manière efficace en identifiant les modèles et les caractéristiques qui correspondent à différentes émotions.

#### **III.2.2. Extraction des caractéristiques :**

L'apprentissage automatique peut extraire automatiquement les caractéristiques pertinentes des images faciales, ce qui permet de capturer les nuances subtiles des expressions émotionnelles qui peuvent être difficiles à détecter pour les humains.

#### **III.2.3. Adaptabilité et généralisation :**

Des algorithmes d'apprentissage automatique peuvent être entraînés sur de grandes quantités de données provenant de diverses sources, ce qui leur permet de s'adapter à différentes conditions d'éclairage, d'angle de vue et de variations faciales et de généraliser leur capacité à reconnaître les émotions dans des situations réelles.

#### **III.2.4. Amélioration de la précision :**

Avec suffisamment de données d'entraînement et d'optimisation des algorithmes, l'apprentissage automatique peut atteindre des niveaux élevés de précision dans la reconnaissance des émotions faciales, souvent surpassant les performances des méthodes traditionnelles basées sur des règles ou des heuristiques.

#### **III.2.5. Applications pratique :**

La reconnaissance des émotions faciales est utilisée dans de nombreuses applications pratiques, telles que les interfaces homme-machine, la détection des mensonges, la surveillance de la santé mentale, l'analyse du sentiment des consommateurs, etc... L'apprentissage automatique rend ces applications possibles en fournissant des outils efficaces pour analyser et interpréter les expressions faciales.

### **III.3. Types d'algorithmes d'apprentissage automatique :**

Il existe différents types d'algorithmes à utiliser, chacun selon la tâche à accomplir et selon le type de données disponibles. Ces algorithmes se divisent en trois catégories : supervisés, non supervisés et par renforcement.

#### **III.3.1. Techniques d'apprentissage supervisé :**

L'apprentissage supervisé, aussi appelé machine learning supervisé .Il consiste à entraîner un algorithme à l'aide de données d'entrée et de sortie connues et étiquetées.

Les données étiquetées sont utilisées pour la formation et les tests.

Le but est que l'algorithme apprenne les associations entre les entrées et les sorties pour qu'il puisse faire des prédictions précises sur de nouvelles données pour lesquelles la sortie est inconnue.

Concrètement, l'algorithme reçoit un jeu de données d'entraînement contenant des exemples annotés. Chaque exemple est constitué d'une entrée et d'une sortie correspondante souhaitée.

Dans le cas d'un problème de classification d'images, les entrées peuvent être des images étiquetées par la classe à laquelle elles appartiennent (chat, chien, etc.).

L'algorithme analyse alors les données d'entraînement pour trouver des motifs et des règles permettant de mapper les entrées aux sorties. Il optimise un modèle mathématique en minimisant l'erreur entre les prédictions et les vraies étiquettes. Une fois entraîné, le modèle est utilisé pour faire des prédictions sur de nouvelles données inconnues. [17]

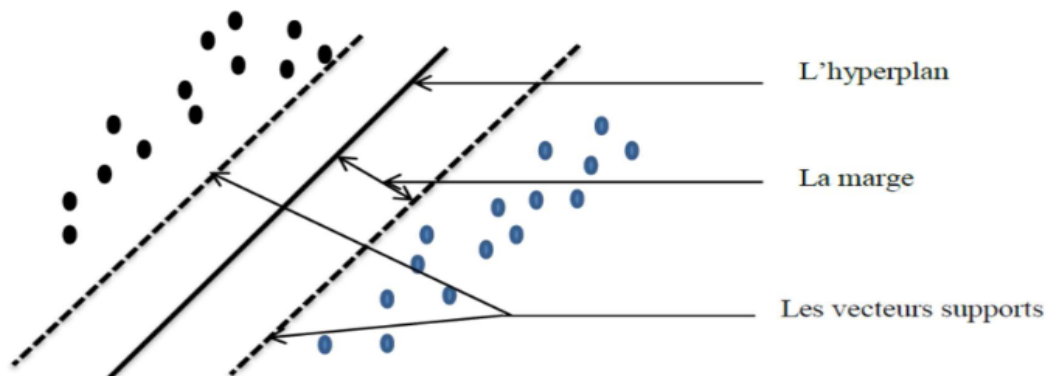
Voici un exemple d'algorithme d'apprentissage supervisé :

##### **III.3.1.1. Machine à vecteurs de support (SVM) (apprentissage supervisé–classification) :**

Les machines à vecteur support (SVM) sont des modèles puissants utilisés en apprentissage automatique pour la classification, la régression et la détection d'anomalies. Elles ont été introduites dans les années 1990 par V. Vapnik et sont devenues populaires en raison de leur capacité à gérer efficacement des ensembles de données de grande dimension. Les SVM sont basées sur le concept de séparation maximale des données en utilisant des hyperplans dans un espace de grande dimension.

**III.3.1.2. Définition et fonctionnement de l'algorithme de classification SVM :**

La machine à vecteur support (SVM) est un algorithme d'apprentissage supervisé qui est utilisé pour effectuer des tâches de classification et de régression. Pour comprendre son fonctionnement, il est important de mentionner les éléments clés tels que les vecteurs supports, l'hyperplan de séparation et le noyau (**Figure III.2**). Les SVM travaillent en classant les données en différentes catégories à l'aide d'hyperplans de séparation. Les vecteurs supports sont les points de données les plus proches des hyperplans de séparation, et ils jouent un rôle crucial dans la définition de l'hyperplan optimal. L'objectif de l'algorithme est de maximiser la marge entre les différentes catégories, ce qui permet d'avoir une meilleure généralisation (**Figure III.3**). Dans le cas des SVM à noyau, une fonction de transformation est utilisée pour projeter les données dans un espace de dimension supérieure, ce qui permet de traiter des problèmes non linéaires. Le fonctionnement de la machine à vecteur support est donc basé sur ces concepts clés et peut être utilisé pour résoudre une variété de problèmes de classification et de régression avec une grande précision.



**FIGURE III.2 :** Séparation entre deux classes par un hyperplan.



**FIGURE III.3.** A gauche, c'est la maximisation de la distance entre la frontière et les points d'entraînement. Au centre, une frontière non-optimale, qui passe très près des points d'entraînement. Cette frontière classe de façon erronée le rond rouge clair comme un carré bleu ! Au contraire, à droite, la frontière optimale (qui maximise la distance) classe bien le rond rouge clair comme un rond rouge.

Il existe deux types d'algorithmes d'apprentissage supervisé :

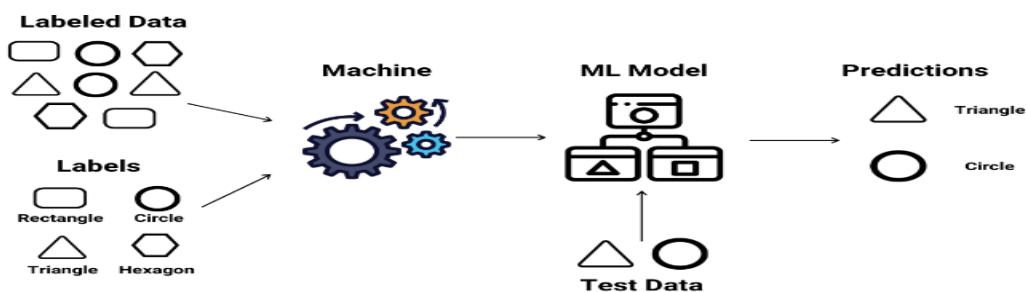
➤ **La classification :**

L'apprentissage supervisé par la classification est utilisé lorsque la sortie ou le résultat peut être catégorisé avec deux classes ou plus.

**Exemple :** Prenons l'image d'une personne avec un large sourire, les yeux ouverts et les coins de la bouche relevés. L'algorithme, grâce à son apprentissage sur les données d'entraînement, pourrait identifier cette image comme exprimant la joie avec un haut degré de certitude.

➤ **La régression :**

La régression est un autre type de méthode d'apprentissage supervisé qui utilise un algorithme pour comprendre la relation entre les variables dépendantes et indépendantes. Les modèles de régression sont utiles pour prédire des valeurs numériques sur la base de différents points de données.



**FIGURE III.4 :** Diagramme du processus d'apprentissage automatique supervisé.

**III.3.2. Techniques d'apprentissage non supervisées :**

Les algorithmes d'apprentissage automatique non supervisées sont utilisés lorsque l'information utilisée pour entraîner le modèle n'est ni classifiée ni étiquetée, et que le nombre de classes et leur nature n'ont pas été prédéterminés. Aucun expert n'est requis. L'algorithme doit découvrir par lui-même la structure plus ou moins cachée des données. [15].

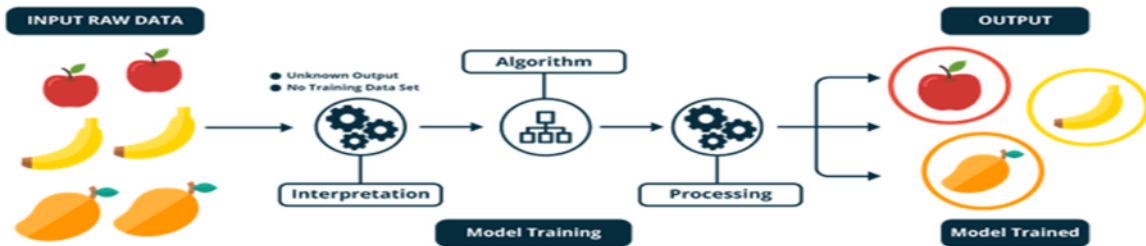


FIGURE III.5 : Diagramme du processus d'apprentissage automatique non supervisé.

**III.3.3. Techniques d'apprentissage par renforcement :**

L'apprentissage par renforcement (RL pour Reinforcement Learning) fait référence à une classe de problèmes d'apprentissage automatique, dont le but est d'apprendre, à partir d'expériences successives, ce qu'il convient de faire de façon à trouver la meilleure solution.[17]

Dans un tel problème, on dit qu'un « agent » interagit avec « l'environnement » pour trouver la solution optimale. Les machines essaient différentes situations pour déterminer les actions les plus avantageuses, plutôt que de simplement recevoir des instructions sur les actions à entreprendre. Ce qui distingue l'apprentissage par renforcement des autres techniques, ce sont l'apprentissage par essais et erreurs et la récompense différée.

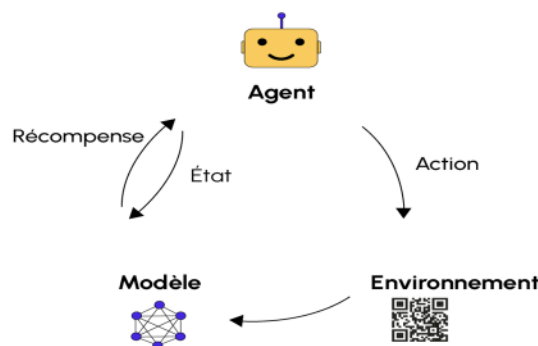


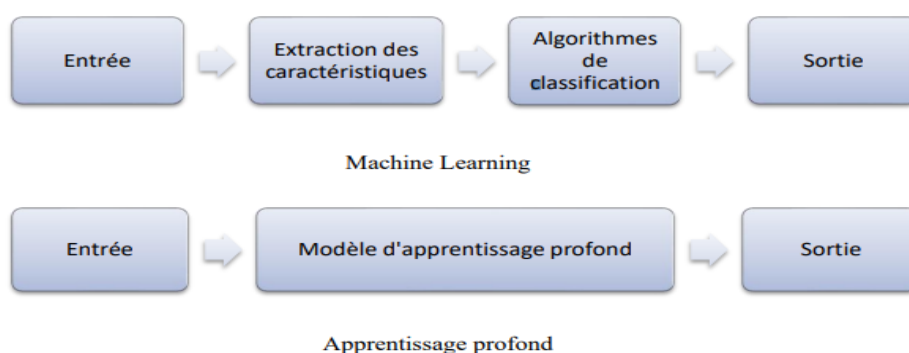
FIGURE III.6 : Principe d'apprentissage par renforcement.

### III.4. Deep learning :

Le Deep Learning est une sous- domaine du machine learning qui repose sur l'utilisation de réseaux de neurones artificiels.

Ces réseaux, inspirés du fonctionnement du cerveau humain, permettent à la machine d'apprendre à partir de données et d'effectuer des tâches complexes sans être explicitement programmée pour les accomplir. [17]

(Figure III.7) montre la différence entre le deep learning est les méthodes conventionnelles du machine learning.



**FIGURE III.7 :** Machine learning Vs Deep learning.

Le deep learning, donne la capacité à la machine d'apprendre comment réaliser des tâches de classification directement à partir des données brutes (image, texte ou d'audio).

Le principe du deep learning est de construire un modèle composé de plusieurs couches, chacune effectuant des transformations sur les données d'entrée. Ces transformations se font une par une, en suivant un ordre de haut en bas. Chaque couche utilise les résultats de la couche précédente pour faire ses calculs. C'est comme si chaque couche construisait sur ce que la précédente a appris.

### III.5. Réseaux de neurones :

Les réseaux de neurones (Artificial Neural Networks ANN) sont des modèles informatiques inspirés par le fonctionnement des neurones biologiques, tel qu'indiqué dans (Figure III.8). Tout comme les neurones dans le cerveau humain interagissent pour traiter l'information, les neurones artificiels dans un réseau de neurones travaillent ensemble pour recevoir des entrées, les traiter et générer des sorties. Donc, un réseau de neurones est une abstraction informatique qui simule le comportement des réseaux de neurones biologiques.

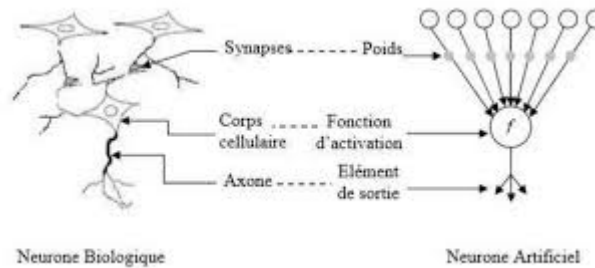


FIGURE III.8: Neurone biologique vs neurone artificiel

Le neurone est une unité qui est exprimée généralement par une fonction sigmoïde.

$$F(X) = \frac{1}{1+e^{-X}} \dots\dots\dots (3)$$

Les réseaux de neurones sont entraînés avec une multitude de données d'entrées couplée à leurs données de sortie respectives. Ils calculent ensuite la donnée de sortie, ils la comparent à la donnée de sortie réelle connue et se mettent à jour en permanence pour améliorer les résultats (si nécessaire).

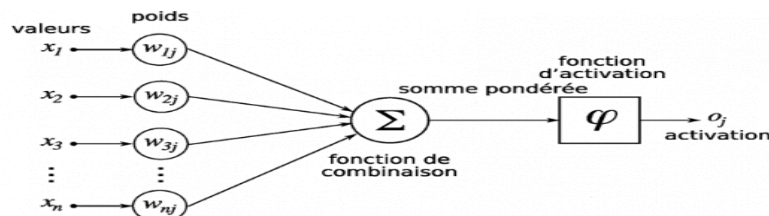


FIGURE III.9 : Schéma de fonctionnement des réseaux de neurones.

Le concept de réseau de neurones repose sur trois étapes principales:

- Pour chaque neurone dans une couche, multiplier la valeur d'entrée par le poids qui est processus connu sous le nom de rétro propagation, afin de minimiser l'erreur entre la sortie du réseau et la sortie souhaitée.
- Ensuite, pour chaque couche, additionner toutes les pondérations des neurones et ajouter un biais (qui est un paramètre ajustable qui permet de contrôler la sortie d'un neurone même lorsque toutes ses entrées sont nulles.).
- Enfin, appliquer la fonction d'activation sur cette valeur pour calculer une nouvelle sortie.

### III.5.1. Le perceptron :

Le perceptron est un modèle simple de réseau de neurones, composé d'une seule couche de neurones. Chaque neurone reçoit des signaux d'entrée, les pondère et les passe à travers une fonction d'activation pour produire une sortie. Il est principalement utilisé pour la classification binaire et peut apprendre à séparer linéairement des données. Cependant, il a des limitations en termes de capacité à modéliser des relations non linéaires. La règle d'apprentissage du perceptron consiste à ajuster les poids des connexions en fonction des erreurs commises lors de la prédiction. Bien que simple, le perceptron a jeté les bases du développement ultérieur des réseaux de neurones en tant que modèle de base.

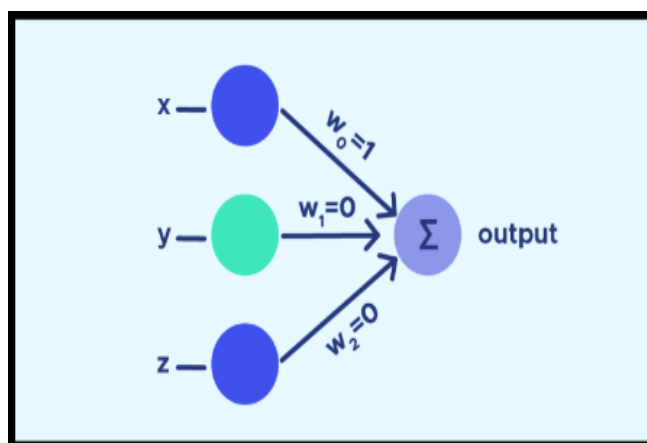


FIGURE III.10 : Schéma d'un perceptron.

### **III.5.2. Perceptron multicouche :**

L'architecture du perceptron multicouche est composée de plusieurs couches de neurones interconnectées, y compris une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. Chaque couche est composée de plusieurs neurones, qui sont des unités de traitement. Les connexions entre les neurones sont associées à des poids qui sont ajustés lors de l'apprentissage. Le perceptron multicouche est capable d'apprendre des représentations complexes en utilisant des algorithmes d'optimisation comme la rétropropagation du gradient. Les réseaux multicouches sont largement utilisés dans des applications telles que la reconnaissance de motifs, la vision par ordinateur et le traitement du langage naturel.

### **III.6. Utilisation des CNN dans l'identification des états émotionnels via expressions faciale :**

Les CNN sont d'une importance capitale dans l'identification des émotions à partir des expressions faciales car ils permettent une analyse approfondie des traits du visage et des nuances émotionnelles. Leur capacité à extraire des caractéristiques clés des visages et à les interpréter pour prédire les états émotionnels est essentielle pour améliorer la précision des classifications. Grâce à leur architecture complexe et à leur capacité d'apprentissage profond, les CNN peuvent identifier des émotions subtiles et complexes avec une grande précision, ce qui en fait un outil de choix pour les recherches en psychologie et en sciences cognitives

### **III.7. L'apprentissage en profondeur (Deep Learning):**

L'apprentissage en profondeur (Deep learning) est un ensemble d'algorithmes d'apprentissage automatique qui tentent d'apprendre à plusieurs niveaux. Il a la capacité d'extraire des caractéristiques à partir des données brutes grâce aux multiples couches de traitement composé de multiples transformations linéaires et non linéaires et apprendre sur ces caractéristiques petites à petit à travers chaque couche avec une intervention humaine minimale. [16].

### III.8. Les réseaux de neurones Convolutifs

Les réseaux de neurones convolutionnels sont à ce jour les modèles les plus performants pour classer des images. Désignés par l'acronyme CNN, de l'anglais Convolutional Neural Network, ils comportent deux parties bien distinctes. En entrée, une image est fournie sous la forme d'une matrice de pixels. Elle a 2 dimensions pour une image en niveaux de gris. La couleur est représentée par une troisième dimension, de profondeur 3 pour représenter les couleurs fondamentales [Rouge, Vert, Bleu]. La première partie d'un CNN est la partie convolutive à proprement parler. Elle fonctionne comme un extracteur de caractéristiques des images. Une image est passée à travers une succession de filtres, ou noyaux de convolution, créant de nouvelles images appelées cartes de convolutions. Certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum local.

Au final, les cartes de convolutions sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN. Comme démontré dans (Figure III.11) :

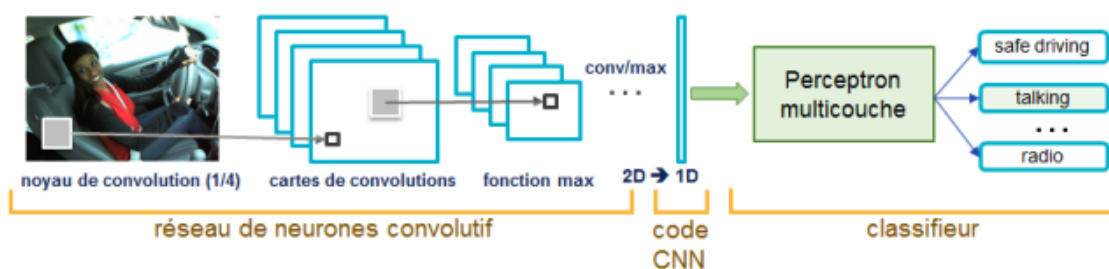


FIGURE III.11 : Architecture standard d'un réseau de neurone convolutionnel.

Ce code CNN en sortie de la partie convolutive est ensuite branché en entrée d'une deuxième partie, constituée de couches entièrement connectées (perceptron multicouche). Le rôle de cette partie est de combiner les caractéristiques du code CNN pour classer l'image. La sortie est une dernière couche comportant un neurone par catégorie. Les valeurs numériques obtenues sont généralement normalisées entre 0 et 1, de somme 1, pour produire une distribution de probabilité sur les catégories. [18]

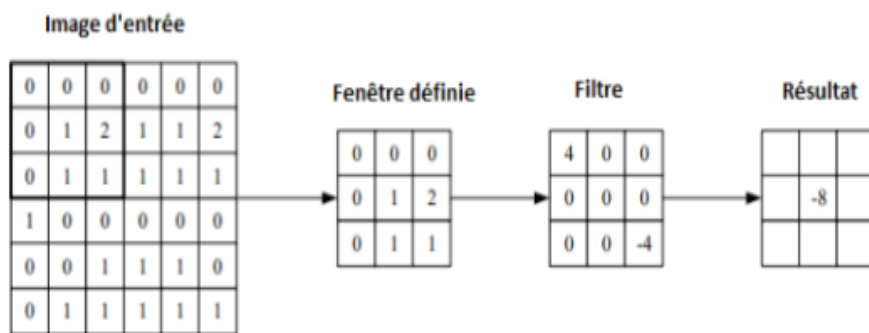
### III.8.1. Architecture d'un réseau de neurone

Les réseaux à couches sont les modèles de traitement de données les plus couramment utilisés. Leur architecture, organisée en couches successives, comprend une couche d'entrée, une couche de sortie et une ou plusieurs couches intermédiaires appelées couches cachées, car elles ne sont pas vues de l'extérieur. Chaque couche est composée d'un certain nombre de neurones. Les connexions sont établies entre les neurones appartenant à des couches successives mais les neurones d'une même couche ne peuvent pas communiquer entre eux dans le cas des réseaux à couches. [19]

Les CNN généralement sont composés de plusieurs couches, chacune avec sa propre fonctionnalité. Voici les principales couches d'un CNN :

#### III.8.1.1. Les couches de convolution :

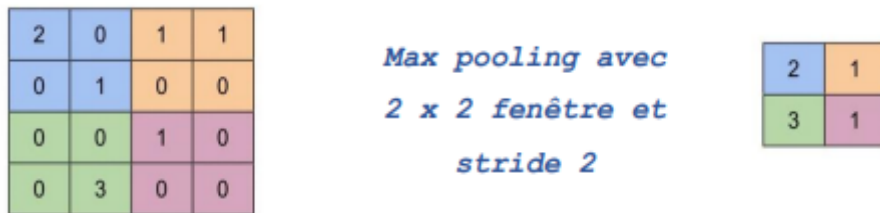
C'est la première couche d'un CNN qui joue un rôle principal pour son fonctionnement, cette couche utilise un ensemble de filtres pour extraire les caractéristique (features) de l'image d'entrée (input image). Chaque filtre est une petite matrice de poids (**Figure III.12**) qui parcourt l'image d'entrée, en effectuant un produit scalaire entre les entrées et ces filtres et au même temps une carte d'activation est produite et elles sont modulées par une fonction d'activation, généralement une unité linéaire rectifiée (communément appelée Relu). La taille et le nombre de ces filtres sont dénis à priori. [20]



**FIGURE III.12 :** Représentation d'une couche convolutive

### III.8.1.2. Les couches de regroupements (pooling) :

Cette couche a pour objectif de diminuer la taille de la carte de caractéristique produite par la couche de convolution. Elles fonctionnent indépendamment des différentes profondeurs du réseau et ne requièrent pas de poids à entraîner. Le regroupement maximal est la technique la plus utilisée pour diminuer la taille d'une image. Elle consiste à réduire une fenêtre donnée à une valeur unique (**Figure III.13**) en sélectionnant la plus grande valeur parmi les éléments de la fenêtre. [21]



**FIGURE III.13 :** Exemple d'un max pooling.

### III.8.1.3. Couche de correction (RELU) :

Un élément important dans l'ensemble du processus est l'Unité linéaire rectifiée ou Relu. Les mathématiques derrière ce concept sont assez simples encore une fois : chaque fois qu'il y a une valeur négative dans un pixel, on la remplace par un 0. Ainsi, on permet au CNN de rester en bonne santé (mathématiquement parlant) en empêchant les valeurs apprises de rester coincer autour de 0 ou d'exploser vers l'infinie. C'est un outil fondamental car sans lequel le CNN ne produirait pas vraiment les résultats qu'on lui connaît.

$$\text{Relu}(X) = \max(0, X) \dots \dots \dots (4)$$

Cela signifie que pour toute valeur d'entrée X, si X est négatif, la sortie est 0. Si X est positif ou nul, la sortie est X lui-même.

Le résultat d'une couche Relu est de la même taille que ce qui lui est passé en entrée, avec simplement toutes les valeurs négatives éliminées.

Le résultat d'une couche Relu est de la même taille que ce qui lui est passé en entrée, avec simplement toutes les valeurs négatives éliminées. [22]



FIGURE III.14 : Opération RELU.

**III.8.1.4. Couche entièrement connectée (FC) :**

La couche entièrement connectée dans un réseau de neurones convolutif (CNN) est une partie cruciale de l'architecture. Elle est chargée de classifier l'image en entrée en produisant un vecteur de sortie. Cette couche applique une combinaison linéaire aux valeurs d'entrée, suivie éventuellement d'une fonction d'activation, pour générer un vecteur de taille N, où N représente le nombre de classes dans le problème de classification d'images. Chaque élément de ce vecteur indique la probabilité que l'image appartienne à une classe spécifique.

Dans la couche entièrement connectée, chaque nœud de la couche de sortie se connecte directement à un nœud de la couche précédente.

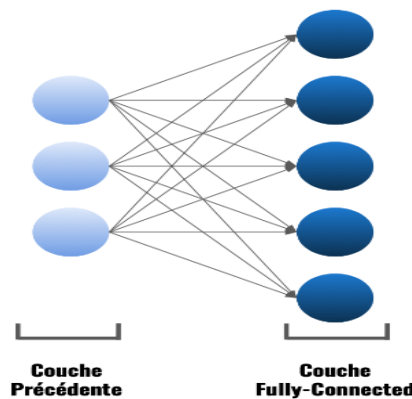


FIGURE III.15 : La couche entièrement connectée.

### III.8.1.5. Couche de perte (LOSS) :

La couche de perte suit la couche entièrement connectée et gère les ajustements de poids sur le réseau. Avant que l'entraînement du réseau ne commence, les poids des couches convolution et entièrement connectées reçoivent des valeurs aléatoires. Ensuite, pendant l'entraînement, la couche de perte vérifie en permanence les suppositions de la couche entièrement connectée par rapport aux valeurs réelles dans le but de minimiser autant que possible la différence entre l'estimation et la valeur réelle. La couche de perte effectue cela en ajustant les poids à la fois dans la couche de convolution et dans les couches entièrement connectées. Maintenant, nous devons classer ces fonctionnalités extraites. Nous avons donc besoin d'un algorithme capable de classer ces entités extraites et de nous dire si les entités extraites sont les caractéristiques d'une personne A, ou d'une autre personne.

## Conclusion

L'intelligence artificielle et le deep learning sont à la base de tout apprentissage informatique et représentent l'avenir des processus décisionnels complexes. Nous avons mis en lumière dans ce chapitre les principales techniques et leurs applications spécifiques. Nous avons vu que le Machine Learning permet aux machines de découvrir des relations et des modèles à partir de données, ce qui leur confère la capacité de prendre des décisions et de résoudre des problèmes de manière autonome.

---

---

# **CHAPITRE IV : Implémentation du CNN et résultats**

---

---

## **Introduction :**

Dans ce chapitre, nous appliquons les réseaux de neurones convolutifs pour la reconnaissance des émotions en utilisant deux bases de données bien établies : JAFFE et CK+. Ces bases de données sont très utilisées et considérées comme des bases de données de référence. À travers ce projet, nous explorons les techniques de prétraitement des données, l'architecture du CNN employé, ainsi que les résultats obtenus et les analyses comparatives entre les performances sur ces deux ensembles de données distincts.

### **IV.1. Environnement de travail :**

Dans cette section, nous présenterons les environnements matériels et logiciels utilisés pour réaliser l'application. Les caractéristiques sont les suivantes :

#### **IV.1.1. Environnement matériel :**

Afin de mener à bien ce projet, il a été mis à notre disposition un ordinateur DELL avec les caractéristiques suivantes :

- Processeur : Intel Core i5-4310U CPU @ 2.00 GHz 2.60 GHz.
- RAM : 8,00 Go.
- Système d'exploitation : Windows 10, 64 bits.

#### **IV.1.2. Environnement logiciel :**

##### **IV.1.2.1. Utilisation de MATLAB 2021a :**

Il existe plusieurs langages de programmation parmi ces langages on a choisi d'utiliser le langage de programmation MATLAB R2021a (Matrix Lonction MATLAB ABoratory) dans notre application car il représente un langage de haut niveau doublé d'un environnement de travail. Il est principalement utilisé dans les calculs scientifiques et les problèmes d'ingénierie parce qu'il permet de résoudre des problèmes numériques complexes en moins de temps requis par les langages de programmation courants, et cela grâce à une multitude de fonctions intégrées et à plusieurs programmes outils testés et regroupés selon usage dans des dossiers appelés boîtes à outils ou "toolbox".

### IV.1.2.2. Les particularités de MATLAB :

MATLAB propose une plateforme complète pour les applications d'apprentissage profond, fournissant une gamme d'outils et de ressources pour faciliter le développement et le déploiement de réseaux de neurones profonds.

Le Deep Learning Toolbox de MATLAB propose des commandes simples pour créer et connecter des couches de réseaux de neurones profonds, la rendant accessible même aux utilisateurs sans connaissances avancées des réseaux de neurones ou des algorithmes de vision par ordinateur.

## IV.2. Description des bases de données

Les bases de données des images des expressions faciales que nous avons utilisées pour l'entraînement et les tests sont : JAFFE et CK+ :

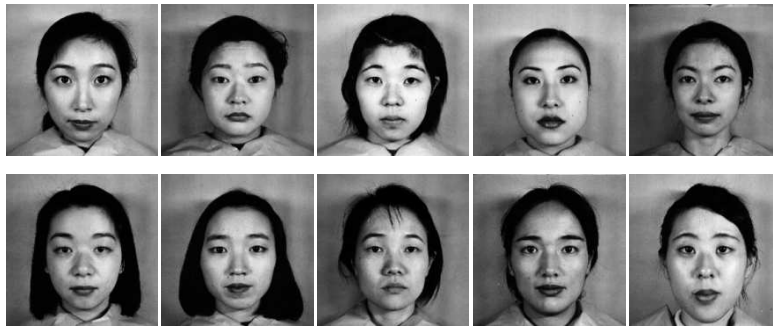
### IV.2.1. Base de données JAFFE :

Cette base a été conçue et assemblée par Michael Lyons, Miyuki Kamachi et Jiro Gyoba. Les photos ont été prises au département de psychologie de l'université de Kyushu, largement utilisé pour évaluer les systèmes de reconnaissance des expressions faciales (FER).

La base de données JAFFE (Japanese Female Facial Expression) est une collection d'images d'expressions faciales de femmes japonaises, elle contient 213 images en niveaux de gris dont chaque image est de taille 256 x256, prise à partir de dix femmes exposées trois fois pour chaque expression faciale de base (colère, dégoût, peur, bonheur, tristesse, neutre et surprise).

Classe	Nombre d'image par classe
Anger	30
Disgust	29
Fear	32
Happiness	31
Neutral	30
Sadness	30
Surprise	30

**Tableau III :** Répartition des images par classe d'émotion dans la base de données JAFFE.



**FIGURE IV.1 :** Échantillons de l'ensemble d'image JAFFE.

#### IV.2.2. Base de données CK+ :

La base de données Cohn-Kanade (CK+) est la base la plus utilisée en FER elle offre une gamme variée d'expressions faciales, d'âges et d'origines ethniques, avec un total de 4009 images. Les participants sont des adultes âgés de 18 à 50 ans et sont aussi de 81% Euro-américains, 13% afro-américains et 6% autres groupes.

Classe	Nombre d'image par classe
Anger	611
Disgust	458
Fear	428
Happiness	1099
Sadness	541
Surprise	872

**Tableau IV :** Répartition des images par classe d'émotion dans la base de données CK+.



FIGURE IV.2 : Échantillons de l'ensemble d'image CK+.

### IV.3. Prétraitement de données :

Le prétraitement de données est une étape cruciale dans le processus d'apprentissage automatique. Elle consiste à nettoyer, transformer et intégrer les données pour les rendre plus fiables et plus efficaces pour l'analyse :

#### IV.3.1. Organisation des Données d'Images :

Au début de l'étude, toutes les images étaient stockées dans un seul dossier, comprenant un total de 213 images représentant différentes émotions faciales. Pour faciliter l'entraînement et l'évaluation du modèle CNN, il était nécessaire de réorganiser ces images en fonction de l'émotion qu'elles représentaient. Pour ce faire, nous avons créé un projet nommé "DataClean" afin de réorganiser les images.

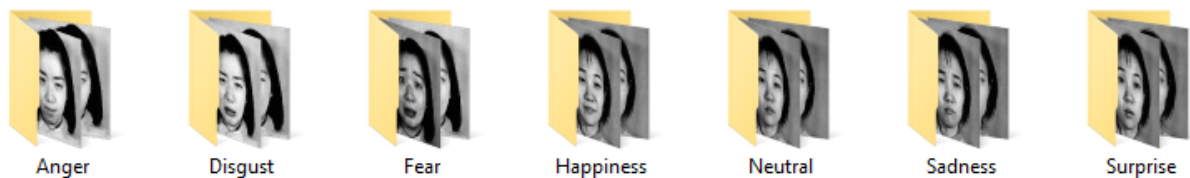


FIGURE IV.3 : Image de la base jaffe après la réorganisation.

### **IV.3.2. La normalisation :**

Lorsque nous normalisons les images en divisant les valeurs des pixels par 255, elle consiste à réduire l'écart de valeurs entre les différentes caractéristiques des données pour éviter que certaines caractéristiques ne dominent les autres. Dans notre code, nous avons normalisé les images en divisant les valeurs par 255 pour les ramener entre 0 et 1. Cela signifie que les valeurs des pixels les plus faibles seront proches de 0, tandis que les valeurs des pixels les plus fortes seront proches de 1.

L'écart (ou scale) dans le contexte de la normalisation des données fait référence à la plage de valeurs qui peuvent prendre les caractéristiques (ou feature) des données.

Les images sont représentées sous forme de matrices de valeurs de pixels, où chaque valeur représente l'intensité de la couleur d'un pixel donné. Les valeurs des pixels peuvent varier considérablement entre les différentes images, ce qui peut affecter l'entraînement du modèle

### **IV.3.3. Le redimensionnement :**

Le redimensionnement est une technique de prétraitement qui consiste à modifier la taille des images pour les adapter aux exigences du modèle de machine learning. Dans notre cas nous avons utilisé la fonction `resize()`, pour redimensionner les images faciales (jaffe et ck+).

Cette opération est importante car elle permet de standardiser la taille des images et de réduire leur dimensionnalité, ce qui facilite l'apprentissage automatique et permet de réduire le temps de traitement.

### **IV.4. L'extraction de fonctionnalité :**

Dans notre cadre d'identification des émotions faciales, le processus d'extraction des caractéristiques visuelles essentielles des images est automatisé. À travers les couches convolutionnelles et de pooling du CNN, les caractéristiques significatives telles que les textures, les contours et les motifs spécifiques aux expressions faciales sont extraites directement à partir des pixels des images d'entrée.

### **IV.5. Présentation du CNN :**

Un modèle CNN est un type de réseau de neurones artificiels conçu pour traiter les données de notre base données. Il est composé de plusieurs couches qui permettent d'extraire des caractéristiques de plus en plus complexes dans les images et de les classifiées.

## IV.6. Entraînement du Modèle :

Le modèle de réseau de neurones convolutifs (CNN) que nous avons utilisé pour l'identification des émotions faciales a été entièrement entraîné à partir de zéro. Cela signifie que les poids et les biais des couches du modèle ont été initialisés de manière aléatoire au début de l'entraînement.

L'entraînement à partir de zéro présente plusieurs avantages et défis :

- **Avantages** : Permet une personnalisation complète du modèle aux données spécifiques de notre tâche, garantissant ainsi que le modèle apprend directement à partir des caractéristiques propres à notre jeu de données.
- **Défis** : Nécessite une grande quantité de données d'entraînement et un temps de calcul considérable pour atteindre une performance optimale.

L'entraînement de notre modèle à partir de zéro nous a permis de construire un modèle entièrement adapté à notre tâche de reconnaissance des émotions faciales, en apprenant directement à partir de notre jeu de données spécifique.

## IV.7. Architecture CNN pour l'extraction de fonctionnalité :

Chaque couche de cette architecture CNN a un rôle précis et essentiel dans le processus de reconnaissance des émotions à partir des expressions faciales.

Notre code dispose de :

- **Trois couches de convolution :**

La première couche de convolution utilise «8 filtres de taille 3x3 avec une fonction d'activation ReLU (Rectified Linear Unit). Cette couche de convolution prend en entrée l'image d'origine et applique 8 filtres de taille 3x3 pour extraire des caractéristiques locales. Le padding 'ame' permet de conserver la taille de l'image en sortie

La deuxième couche de convolution utilise 16 filtres de taille 3x3 avec une fonction d'activation ReLU. Cette couche de convolution prend en entrée la sortie de la première couche de convolution et applique 16 filtres de taille 3x3 pour extraire des caractéristiques plus complexes. Le padding 'ame' permet de conserver la taille de l'image en sortie.

La troisième couche de convolution utilise 32 filtres de taille 3x3 avec une fonction d'activation ReLU. Cette couche de convolution prend en entrée la sortie de la deuxième couche de convolution et applique 32 filtres de taille 3x3 pour extraire des caractéristiques encore plus complexes. Le padding 'same' permet de conserver la taille de l'image en sortie.

➤ **Trois couches de pooling :**

Les couches de pooling (MaxPooling1, MaxPooling2 et MaxPooling3) permettent de réduire la taille de l'image d'entrée de moitié, ce qui se fait de manière systématique pour simplifier les informations traitées par le réseau de neurones. Concrètement, chaque couche utilise un filtre de taille 2x2 pour parcourir l'image. Pour chaque région 2x2, elle sélectionne la valeur maximale, ce qui permet de conserver les caractéristiques les plus importantes tout en éliminant les détails moins significatifs.

- le dropout améliore la capacité du modèle à généraliser sur de nouvelles données, augmentant ainsi sa performance et sa fiabilité dans la reconnaissance des émotions faciales. une couche de dropout (**dropoutLayer**) :

Dans notre projet d'identification des états émotionnels via l'expression faciale, la couche de dropout est utilisée pour éviter le surapprentissage (overfitting), en désactivant aléatoirement une fraction des neurones pendant chaque itération d'entraînement. Cela empêche notre modèle de s'adapter trop bien aux données d'entraînement (JAFPE, CK+), perdant ainsi sa capacité à généraliser sur de nouvelles données.

Le dropout améliore donc la capacité du modèle à généraliser sur de nouvelles données, augmentant ainsi sa performance et sa fiabilité dans la reconnaissance des émotions faciales.

➤ **Couche entièrement (connectéeFullyConnectedLayer « FC ») :**

Cette couche joue un rôle crucial dans l'architecture d'un réseau de neurones convolutifs (CNN) utilisé pour des tâches de classification des émotions à partir de notre jeu de données. Elle se situe à la fin du réseau et a pour fonction de combiner toutes les caractéristiques extraites par les couches de convolution et de pooling précédentes. Spécifiquement, la couche (FC) utilise un nombre de neurones égal au nombre de classes d'émotions, soit 7 dans notre cas, correspondant aux 7 états émotionnels identifiés dans les bases de données CK+ et JAFPE. Chaque neurone de cette couche est connecté à tous les neurones de la couche précédente, permettant ainsi de prendre en compte toutes les caractéristiques pertinentes pour la classification finale.

Les sorties de cette couche sont ensuite passées à travers une fonction d'activation softmax, qui convertit les scores en probabilités, facilitant la détermination de l'émotion.

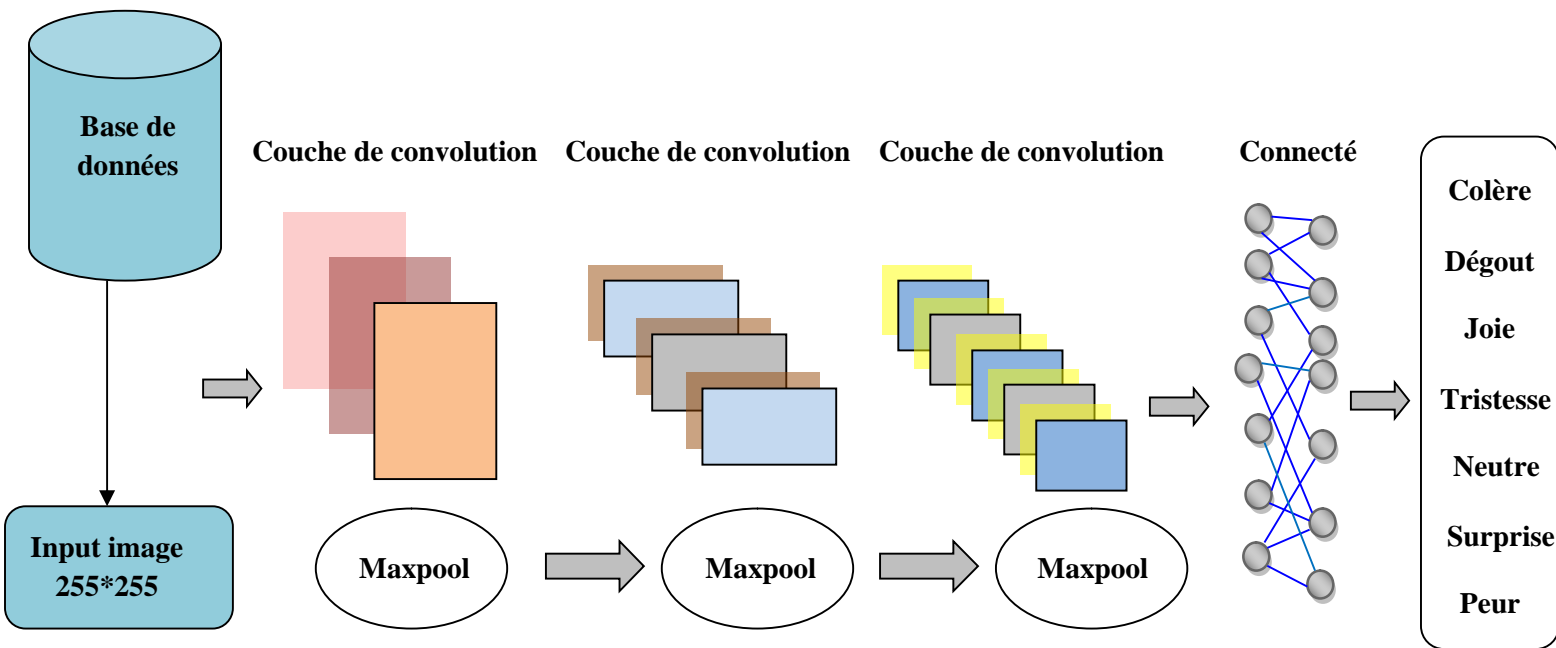


FIGURE IV.4: Architecture proposée.

#### IV.8. CNN classification :

Cette couche finale prend les probabilités produites par la couche fully connected, qui a sept neurones correspondant aux sept états émotionnels présents dans les bases de données CK+ et JAFFE, et attribue une classe spécifique à chaque image d'entrée. En comparant les prédictions avec les étiquettes réelles des données d'entraînement, elle utilise une fonction de perte pour calculer l'erreur et ajuste les poids du réseau via rétropropagation. Cela permet au modèle d'apprendre efficacement et d'améliorer sa précision dans la reconnaissance des émotions faciales.

#### IV.9. Evaluation du modèle :

Pour évaluer les performances de notre modèle de reconnaissance des émotions faciales, nous avons utilisé la relation de taux de bonne classification. Cette relation est calculée comme la proportion de prédictions correctes parmi le nombre total de prédictions (le résultat produit par le modèle lorsqu'il reçoit une entrée). La formule que nous avons utilisée divise le nombre de prédictions correctes par le nombre total d'images dans l'ensemble de teste, on obtient ce la précision du modèle. :

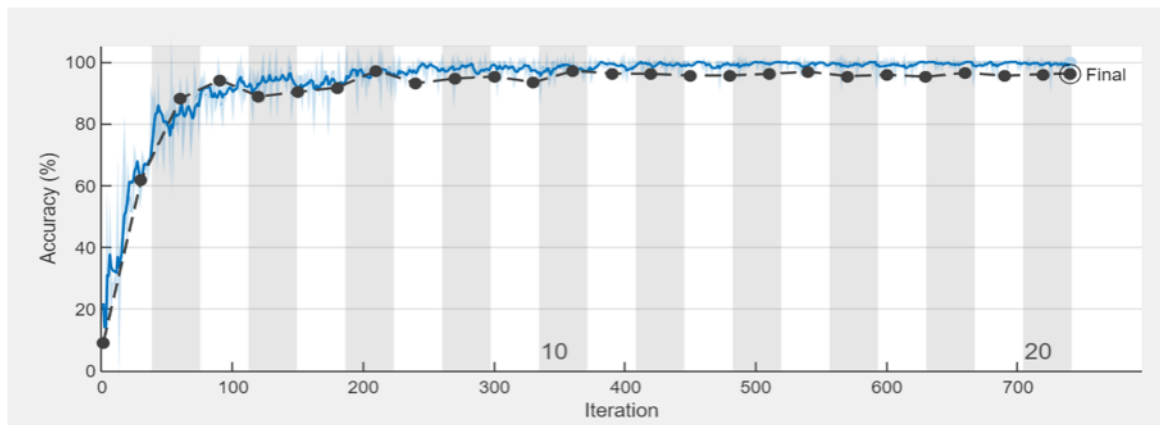
$$\text{Taux de bonne classification} = \frac{\text{nombre de prediction correcte}}{\text{normbre totale de prediction}} \dots\dots\dots (5)$$

#### IV.10. Résultats et discussions :

Dans cette section, nous procédons à une analyse détaillée des graphes illustrant les performances de notre modèle CNN pour la reconnaissance des émotions faciales lors de son entraînement, évalué sur deux ensembles de données distincts : JAFFE et CK+.

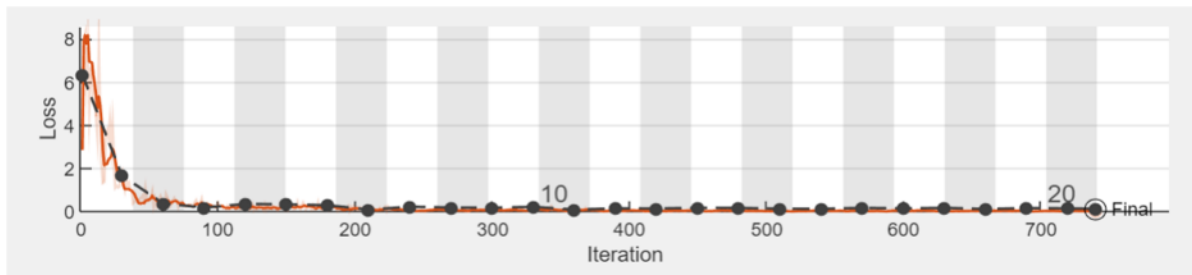
Dans un premier temps, nous avons choisi de travailler avec le dataset CK+ qui contient six états émotionnels distincts pour notre étude. Nous avons utilisé un total de **400** images par classe pour entraîner le modèle, totalisant ainsi **2400** images, avec les **1609** images restantes réservées à la validation (test). Cette répartition équilibrée garantit une représentation uniforme de chaque émotion dans notre ensemble de données.

Nous présentons ci-dessous les graphes représentant l'évolution de la précision et de l'erreur en fonction du nombre d'itérations d'apprentissage.



**FIGURE IV.5** : Graphe de précision pour la base donnée CK+.

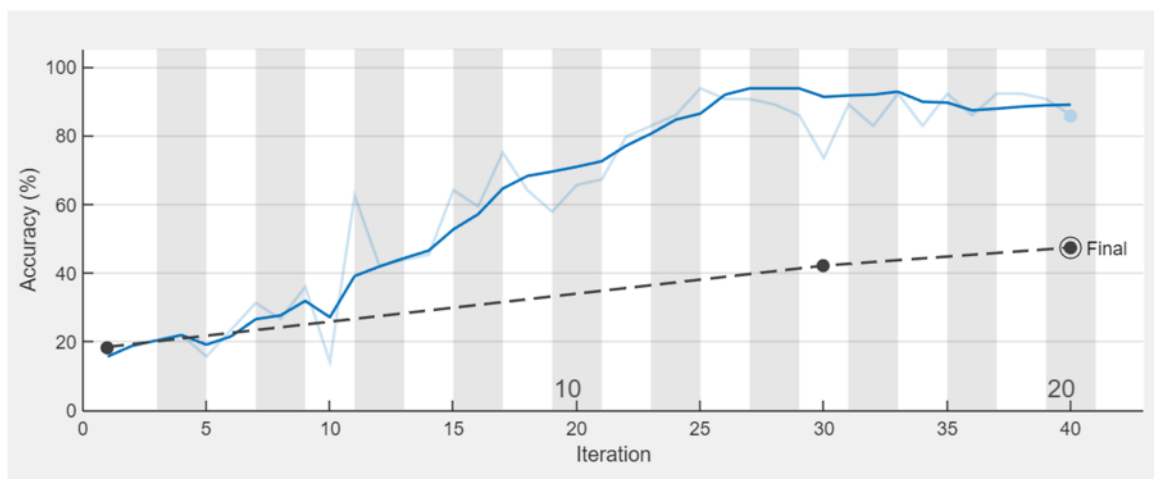
On observe une augmentation progressive de la précision tout au long du processus d'apprentissage, atteignant un maximum de **96,21%** après **700** itérations. Cela indique que le modèle apprend efficacement à reconnaître les différentes émotions à partir des images du dataset.



**FIGURE IV.6 :** Graphe de perte pour la base donnée CK+.

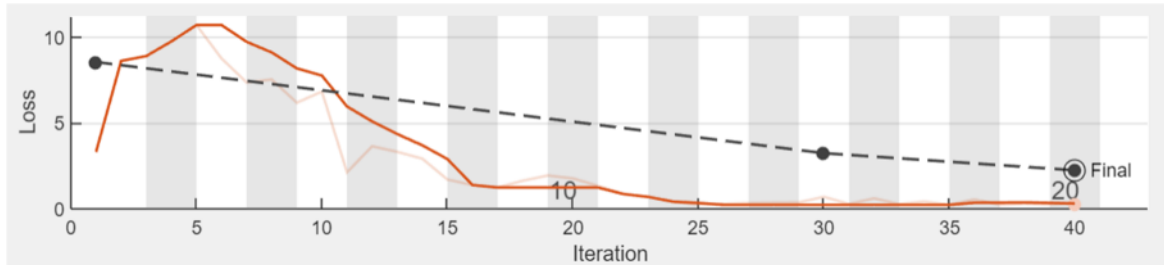
Parallèlement, ce graphe représente l'évolution de l'erreur (Loss) de classification en fonction du nombre d'itérations. L'erreur de classification, (définie comme la différence entre les prédictions du modèle et les valeurs réelles), diminue progressivement au cours du processus d'apprentissage. Cela indique que la capacité du modèle à prédire les expressions faciales s'améliore avec le temps et que moins d'erreurs se produisent à mesure que l'apprentissage progresse.

Maintenant, nous avons utilisé une autre base de données, « **JAFFE** » qui contient sept états émotionnels distincts pour notre étude. Nous avons utilisé un total de **25** images par classe pour l'apprentissage, totalisant ainsi **175** images, et les **38** images restantes pour l'évaluation (test). Cette répartition équilibrée garantit une représentation uniforme de chaque émotion dans notre ensemble de données.



**FIGURE IV.7 :** Graphe de précision pour la base donnée JAFFE.

Ce graphe fourni révèle que le modèle d'apprentissage a atteint une précision de validation de **47,37%** après environ **40** itérations d'entraînement, la précision modérée suggère une marge d'amélioration. D'après ces résultats, on dit que la précision du modèle relativement faible.



**FIGURE IV.8 :** Graphe de perte pour la base donnée JAFFE.

D'après ce graphe, nous constatons que la courbe de perte diminue constamment au cours de l'entraînement, ce qui indique que le modèle s'améliore en permanence.

L'observation d'une erreur d'entraînement élevée au début de la courbe, suivie d'une diminution progressive jusqu'à un niveau bas, est un phénomène courant dans l'apprentissage des modèles de réseaux de neurones artificiels, et ce pour plusieurs raisons:

➤ **Initialisation des paramètres :**

Au début de l'apprentissage, les paramètres du modèle, tels que les poids des connexions entre les neurones, sont initialisés avec des valeurs aléatoires. Ces valeurs aléatoires peuvent conduire à un comportement erratique du modèle et à une erreur d'entraînement élevée.

➤ **Apprentissage des caractéristiques:**

Au fur et à mesure que l'apprentissage progresse, le modèle commence à identifier les caractéristiques importantes dans les données d'entraînement. Cela lui permet d'améliorer ses prédictions et de réduire l'erreur d'entraînement.

**IV.11. Matrice de confusion :**

Pour évaluer le modèle que nous avons implémenté, nous avons utilisé la matrice de confusion. Les paramètres d'évaluation utilisés sont calculés en utilisant l'équation ....

Les résultats des matrices de confusion obtenues sont illustrés dans la **Figure IV.9** et **Figure IV.10**. Elles fournissent un aperçu des performances d'un modèle de classification des émotions (présente le nombre d'images correctement et incorrectement classées pour chacune des sept émotions : Colère, Dégoût, Peur, Bonheur, Neutre, Tristesse et Surprise) utilisant les dataset JAFFE et CK+ comme suivant :

**Matrice de Confusion pour la Validation**

Anger	3					2	60.0%	40.0%
Disgust	2	1	1				25.0%	75.0%
Fear	1	2	2			2	28.6%	71.4%
Happiness				6			100.0%	
Neutral				3	2		40.0%	60.0%
Sadness			1	1		4	66.7%	33.3%
Surprise	1		1	2		1		100.0%

**FIGURE IV.9 :** Matrice de confusion pour 7 expressions Pour la base de données JAFFE.

**Matrice de Confusion pour la Validation**

ANGRY	197	6	4	1	3		93.4%	6.6%
DISGUST		58					100.0%	
FEAR			28				100.0%	
HAPPINESS	5	6	23	654	9	2	93.6%	6.4%
SADNESS	1			1	139		98.6%	1.4%
SURPRISE						472	100.0%	

**FIGURE IV.100 :** Matrice de confusion pour 6 expressions Pour la base de données CK+.

Ces matrices de confusion sont composées de lignes et de colonnes. Chaque ligne correspond à une classe d'étiquette réelle, et chaque colonne correspond à une classe d'étiquette prédite par le modèle.

Les valeurs dans les cellules rouges de la matrice représentent le nombre d'images qui ont été prédites comme appartenant à une classe donnée, alors qu'elles appartiennent en réalité à une autre classe, et les valeurs dans les cellules bleues représentent le nombre d'images qui ont été correctement classées.

Pour mieux comprendre les performances du modèle, il est utile de visualiser les données de ces deux matrices de confusion sous forme de tableaux illustré comme suivants :

Classe	Nombre d'image d'entre	Nombre d'image Correctement prédites	Précision	Erreurs	Répartition des erreurs
Anger	5	3	60%	40%	Sadness : 40%
Disgust	4	1	25%	75%	Fear : 25% Anger : 50%
Fear	7	2	28,6%	71,4%	Anger : 28,6% Disgust : 14,2 % Sadness : 28,6%
Happiness	6	6	100%	0%	/
Neutral	5	2	40%	60%	Happiness : 60%
Sadness	6	4	67,7%	33,3%	Fear : 16,7% Happiness : 16,7%
Surprise	5	0	0%	100%	Fear : 20% Anger : 20% Happiness : 40% Sadness : 20%

**Tableau V** : Répartition détaillée des performances par classe avec JAFFE.

A partir du tableau V, nous avons observé des variations de la précision selon les différentes classes émotionnelles. Par exemple, certaines émotions ont été identifiées avec une précision parfaite (**100%**) comme **Happiness**, tandis que d'autres comme **Surprise** ont montré une incapacité totale à être identifiées correctement avec une précision de **0%** (une confusion entre les émotions fréquente, comme le montre la répartition des erreurs).

La précision pour les autres émotions est **faible**, en particulier pour la **Fear (28,6%)** et le **Disgust (25%)**.

Classe	Nombre d'image d'entre	Nombre d'image Correctemen prédites	Précision	Erreurs	Répartition des erreurs
Anger	211	197	93,4%	6,6%	Disgust : 2,8 % Fear : 1,9 % Happiness : 0.5 % Sadness : 1,4 %
Disgust	58	58	100%	0%	0%
Fear	28	28	100%	0%	0%
Happiness	699	654	93,6%	6,4%	Disgust : 0,9 % Fear : 3,3 % Sadness : 1,3 % Anger : 0,7 % Happiness : 0,3 %
Neutral	141	139	98,6%	1,4%	Anger : 0,7 % Happiness : 0,7 %
Sadness	472	472	100%	0%	0%

**Tableau VI** : Répartition détaillée des performances par classe avec CK+

Les résultats indiquent que le modèle est particulièrement efficace pour identifier les émotions comme le **Dégoût**, la **Peur** atteignant une précision parfaite de **100%**. Cela indique une forte capacité à distinguer ces émotions des autres, la **Tristesse** est également très bien identifiée avec une précision de **98,6%**, démontrant une bonne précision du modèle à cette émotion subtile.

La **Colère** et la **Joie** sont classées avec des précisions respectives de **93,4%** et **93,6%**. Bien que ces pourcentages soient élevés, ils indiquent que le modèle peut parfois confondre ces émotions (erreurs de classification minimales) avec d'autres, indiquant que le modèle les distingue généralement avec précision.

#### **IV.12. Analyse comparative des performances :**

La différence de performance observée entre ces deux modèles s'explique par la taille et la diversité des datasets utilisés.

Le choix d'un petit dataset homogène comme « **JAFFE** » est susceptible de subir un surapprentissage, où le modèle peut apprendre spécifiquement les données d'entraînement et est moins adaptable à de nouvelles données. Cela s'explique en partie par la difficulté pour le modèle à saisir la variabilité et la diversité des expressions faciales.

En revanche, les modèles entraînés sur un grand dataset diversifié tel que « **CK+** » permettent à notre modèle d'apprendre une plus grande variabilité des expressions faciales. Il apprend à reconnaître les caractéristiques qui sont pertinentes pour classifier des données, quelle que soit l'origine ethnique, grâce au grand nombre d'exemples d'apprentissage. Cela aide à améliorer les performances du modèle sur de nouvelles données.

Le tableau ci-dessous présente un résumé des résultats de classification pour les deux bases de données :

	Nombre d'image pour l'apprentissage	Nombre d'image entrainer	Images bien classés	Images mal classés	Taux de précision	Taux d'erreur	Temps d'exécution
<b>JAFFE</b>	<b>175</b>	<b>38</b>	<b>18</b>	<b>20</b>	<b>47,37%</b>	<b>52,63%</b>	<b>1 min 40 sec</b>
<b>CK+</b>	<b>2400</b>	<b>1609</b>	<b>61</b>	<b>1548</b>	<b>96,21%</b>	<b>3,79%</b>	<b>42 min 03 sec</b>

**Tableau VII :** Comparaison des performances sur les ensembles de données JAFFE et CK+.

Il y a une différence notable entre l'utilisation d'un dataset large et d'un dataset petit. Les datasets larges, en raison de leur grand volume et de la diversité des données qu'ils contiennent, nécessitent des périodes d'entraînement plus longues pour les modèles d'apprentissage. En revanche, les datasets petits nécessitent moins de temps pour le prétraitement en raison de leur taille réduite et de la simplicité de leurs données.

#### **IV.13. Tests sur des images n'appartenant pas à la base :**

Pour évaluer les performances de notre réseau de reconnaissance des émotions, nous avons téléchargé des images complètement différentes depuis Google. Ces images, n'appartenant pas aux bases de données CK+ et JAFFE utilisées pour l'entraînement, ont servi à tester la robustesse et la capacité de généralisation du modèle. C'est pour cette raison que nous avons créé une interface utilisateur dédiée, permettant de faciliter l'importation et l'analyse de ces images inédites.

#### IV.14. Bloc interface :

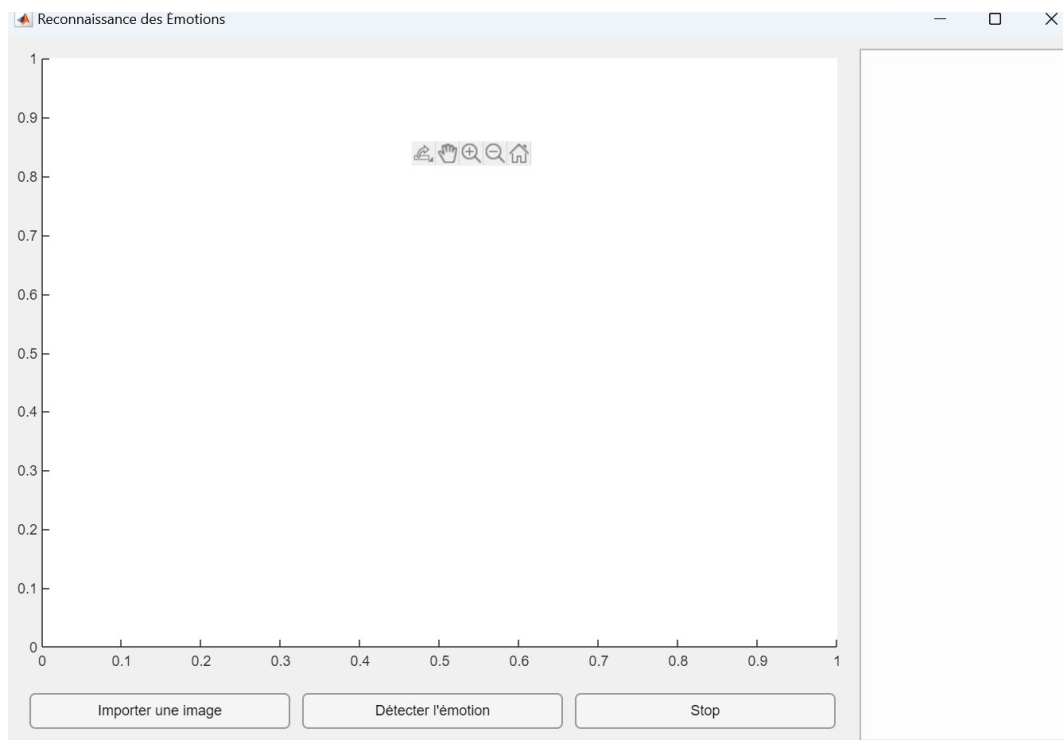
L'interface que nous avons créée est structurée autour de plusieurs éléments essentiels :

- Une zone d'affichage de l'image : permettant la visualisation des images importées.
- Une zone de texte : dédiée pour afficher les émotions détectées avec les pourcentages associés à chaque classe émotionnelle.

Ainsi qu'une série de boutons pour contrôler les principales fonctionnalités de l'application :

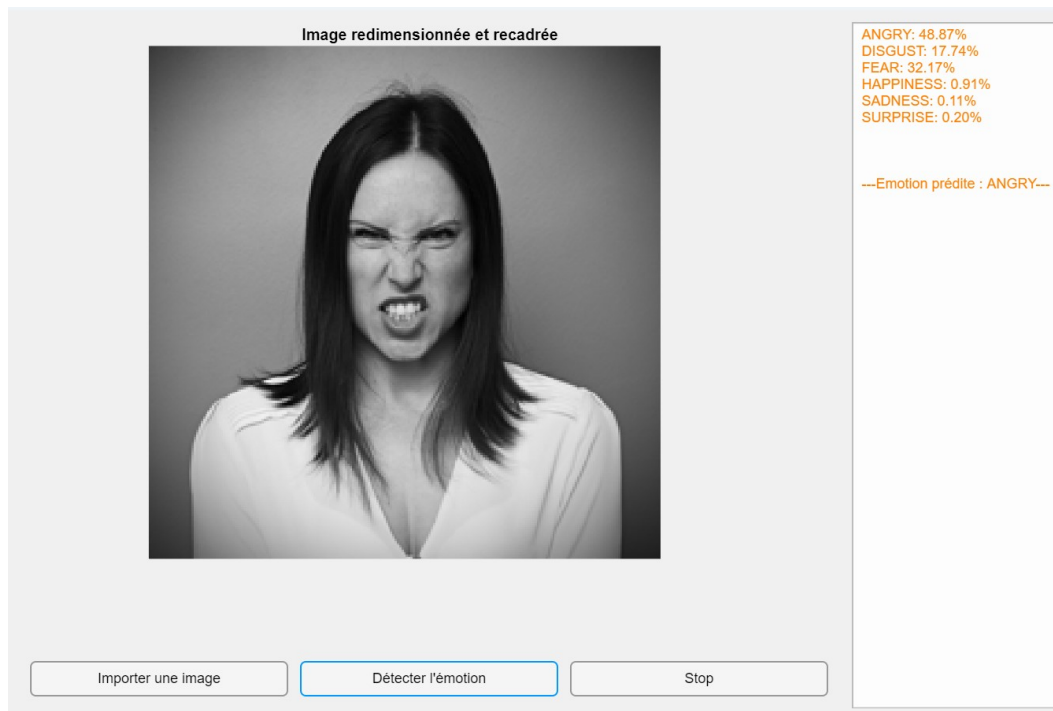
- Le bouton d'importation d'image facilite le chargement de nouvelles images,
- Le bouton de détection d'émotion initie le processus de reconnaissance des émotions.
- Le bouton d'arrêt permet de mettre fin à l'exécution de l'application.

Ces éléments facilitent l'utilisation de l'interface avec le modèle et permettent une évaluation précise des résultats obtenus. Nous pouvons interagir facilement avec les fonctionnalités pour importer des images, détecter les émotions, et arrêter l'application selon nos besoins.



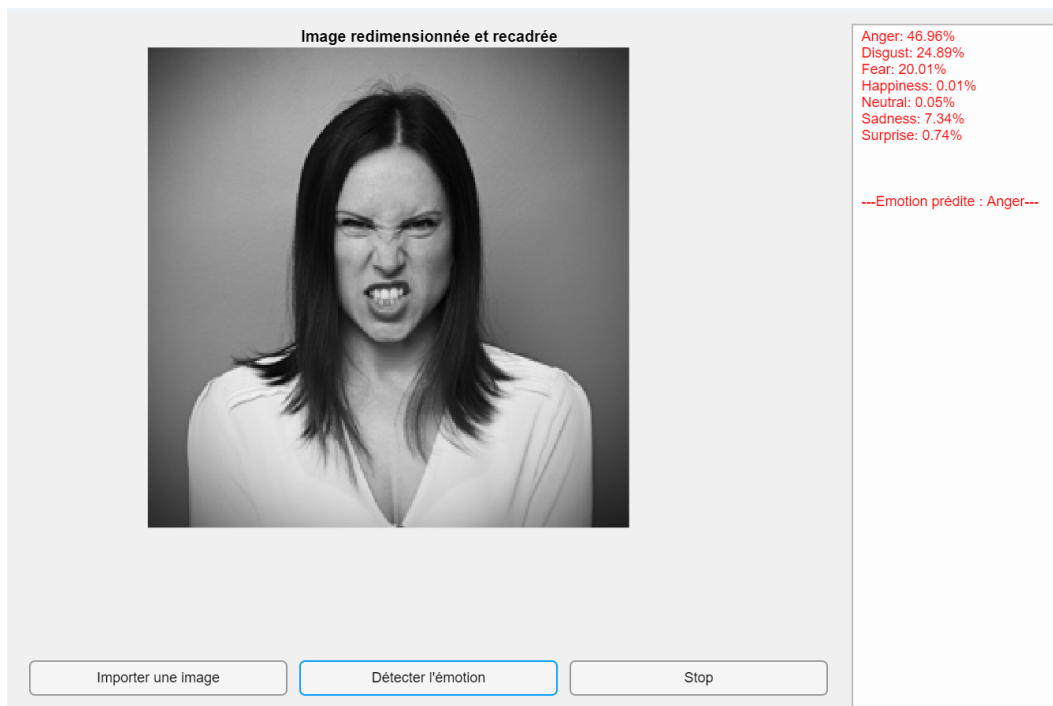
**FIGURE IV.111 :** interface vide.

## ❖ Test du modèle CNN entraîné avec la base CK+ :



**FIGURE IV.122** : émotion prédite (colère) entraîné sur la base de données CK

Afin d'évaluer les performances de notre modèle CNN entraîné sur la base de données **CK+**, nous avons téléchargé une image et l'avons analysée. La (**FIGURE IV.12**) montre les résultats de la détection d'émotions indiquant que le modèle a prédit la femme sur l'image exprimait de la colère avec un niveau de confiance de **48,87%**.

❖ **Test du modèle CNN entraîné avec la base JAFFE :**

**FIGURE IV.133 :** émotion prédite (colère) entraîné sur la base de données JAFFE

Dans ce cas on a testé notre modèle CNN entraîné avec le jeu de données **JAFFE**. Nous avons téléchargé une autre image avec un état émotionnelle pour l'analysée, (**FIGURE IV.13**) montre les résultats de la détection d'émotions indiquent que le modèle a prédit que la femme sur l'image exprimait de la colère avec un niveau de confiance de **64,22%**.

Après avoir testé les performances de notre réseau CNN entraîné avec les deux bases de données (JAFFE et CK+), les résultats se sont révélés satisfaisants pour les deux modèles : celui entraîné avec CK+ et celui entraîné avec JAFFE. Bien que le jeu de données JAFFE est très petit comparé à CK+, les performances obtenues ont été remarquables. Cela démontre que, bien qu'une grande quantité de données soit généralement nécessaire pour un réseau performant, notre modèle a tout de même bien réussi avec les données limitées de JAFFE.

## **Conclusion**

Dans ce chapitre, nous avons présenté la partie pratique de notre travail, où nous avons entraîné un réseau de neurones convolutionnels sur deux bases de données distinctes, JAFFE et CK+.

L'entraînement du réseau avec ces bases de données a permis d'analyser les résultats en termes de précision et d'erreur. La reconnaissance des états émotionnels, basée sur les six états de CK+, a atteint une précision supérieure à 90%, tandis que celle réalisée avec JAFFE n'a pas dépassé 60%. Cette comparaison a mis en évidence l'importance de la quantité et de la qualité des données d'entrée pour obtenir de meilleurs résultats.

Nous avons constaté que le modèle montre d'excellents pourcentages d'évaluation en l'entraînant avec la base CK+.

Nous avons constaté que le modèle montre d'excellentes performances lorsqu'il est entraîné avec un jeu de données diversifié et volumineux.

---

---

# CONCLUSION

---

---

## CONCLUSION

---

Les systèmes de (FER) sont de plus en plus performants grâce aux avancées technologiques. Ces systèmes peuvent identifier avec précision les sept émotions faciales fondamentales : la joie, la tristesse, la colère, la peur, la surprise, le dégoût et le neutre. Ce travail nous a permis de relever des défis dans plusieurs domaines différents, par exemple en combinant des aspects de l'automatique, de l'informatique et de la psychologie, tout en enrichissant nos connaissances en apprentissage automatique.

Dans ce mémoire, nous avons abordé plusieurs aspects essentiels de la reconnaissance des émotions faciales. Nous avons commencé par introduire les concepts fondamentaux du traitement d'image et discuté des expressions faciales ainsi que de leur rôle crucial dans la communication non verbale. Ensuite, nous avons exploré les techniques de Machine Learning et de Deep Learning, qui se révèlent particulièrement efficaces pour la reconnaissance faciale et d'objets. Ces méthodes ont démontré une grande efficacité dans l'identification des expressions faciales, prouvant qu'elles peuvent être appliquées avec succès dans ce domaine.

Nous avons appliqué des réseaux de neurones convolutifs (CNN) à la reconnaissance des expressions faciales. Nous avons présenté une étude comparative des résultats obtenus lors de l'apprentissage de notre réseau avec deux bases de données : JAFFE et CK+. Afin d'évaluer la technique utilisée, nous avons opté pour l'utilisation de métriques d'évaluations telles que Le taux de bonne classification. Ces métriques nous permettent d'évaluer les résultats du modèle obtenu. Les résultats obtenus sont satisfaisants. Pour la base de données CK+, nous avons atteint un taux de bonne classification (précision) de 96,21 %. En opposition pour la base de données JAFFE, le taux de bonne classification est de 47,37 %. Ces chiffres montrent une performance remarquable sur la base CK+, tandis que les résultats pour JAFFE indiquent des défis potentiels liés à la taille et à la diversité limitée de cette base de données.

Tous ces résultats ont montré le potentiel des techniques basées sur le Deep Learning pour améliorer la précision et la robustesse des systèmes de reconnaissance des émotions faciales. Comme perspective pour ce travail, nous suggérons d'élargir l'apprentissage du modèle CNN pour inclure d'autres bases de données. Ceci permettra au modèle de se généraliser pour n'importe quel type de visage.

---

---

# **BIBLIOGRAPHIE**

---

---

# Bibliographie

---

[1] : KHERCHAOUI Sonia, “identification des états émotionnels par les expressions faciales “, these de doctorat en science .UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE HOUARI BOUMEDIENE.

[2] : GUEMIDI Abdelwahab -KAZITANI Noredine, Détection et classification des émotions des personnes, master en Electronique des systèmes embarqués, 2019.

[3]TRAITEMENT D’IMAGES INTRODUCTION Département d’Informatique et de Recherche Opérationnelle.Http : //www.iro.umontreal.ca/~mignotte/ift6150 E-mail : [mignotte@iro.umontreal.ca](mailto:mignotte@iro.umontreal.ca), mai 2024

[4] : R.C. Gonzalez, R.E. Woods and S.L. Eddins, Digital Image Processing Using Matlab, Pearson Prentice Hall, New Jersey, 2004.

[5] : Les fondamentaux – Chapitre 2 – Caractéristiques d’une image médicale, E. Durand et E. Blondiaux, <https://www.cnp-mn.fr/les-fondamentaux-chapitre-2-caracteristiques-dune-image-medicale/>

[6] : Bernard Guévorts, Comment reconnaître les expressions faciales des émotions, mai 2024

[7] : <https://fr.wikipedia.org/wiki/%C3%89motion>, mai 2024

[8] : <https://antoinepeze.com/emotions/> comprendre les émotions, mai 2024

[9] : OUAKKAF, Hassiba et BERKANE, M. Reconnaissance automatique des expressions faciales par support vector machine. 2017.theory. American Journal of Psychology, 39 :10–124, 1927, juin 2024

[10]: Ouakkaf, H., & Berkane, M. (2017). Reconnaissance automatique des expressions faciales par support vector machine, juin 2024

[11]: Valentin Bisson. Septembre, 2012 Algorithmes d’apprentissage pour la recommandation, juin 2024

[12] :Byoung Ko. A brief review of facial emotion recognition based on visual information. sensors, 18(2):401, 2018, juin 2024

[13] : LEROY, Bertrand, CHOUAKRIA, Ahlame, HERLIN, Isabelle, et al. Approche, juin 2024

[14] : ZERROUGUI IBTISSEM, HAMITOUCHE IMENE, Réalisation d'un Système de Reconnaissance des Emotions Basé sur des Images Faciales 2021/2022, juin 2024

[15] Docteur de l'Université Clermont Auvergne Spécialité : "Électronique et Systèmes" Khoulood DAHMANE le 17 juin 2020 Analyse d'images par méthode de Deep Learning appliquée au contexte routier en conditions météorologiques dégradées, juin 2024

[16] <https://dataanalyticspost.com/lhistoire-de-la-datascience-timeline-sciences/>

[17] : <https://www.salesforce.com/fr/resources/definition/apprentissage-supervise/> , juin 2024

[18] : <https://www.lepont-learning.com/fr/deep-learning-applications-concretes/>, juin 2024

[19] Mokri Mohammed Zakaria, classification des images avec les réseaux de neurones convolutionnels (2017) Université Abou Bakr Belkaid Tlemcen, juin 2024

[20] 2008 -David -RECHAL Analyse du potentiel des classifications de type SVM pour l'identification des parcelles arborées dans les images de télédétection à très haute résolution spatiale <https://rfia2012.wordpress.com/wp-content/uploads/2012/02/svm.pdf> , juin 2024

[21] SARAH MADELAINE, Présentation du réseau de neurones convolutifs appliqué aux images médicales, 202, juin 2024

[22] : [https://staff.univ-batna2.dz/sites/default/files/merzougui\\_ghalia/files/deep\\_learning-cou\\_presentation3.pdf](https://staff.univ-batna2.dz/sites/default/files/merzougui_ghalia/files/deep_learning-cou_presentation3.pdf) , juin 2024