

REPUBLICQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj

Faculté des Sciences et de la technologie

Département d'Electronique

Mémoire

Présenté pour obtenir

LE DIPLOME DE MASTER

FILIERE : TELECOMMUNICATION

Spécialité : Systèmes des Télécommunications

Par

➤ **ZITOUNI YASMINE**

Intitulé

Reconnaissance des expressions faciales basée sur l'apprentissage profond

Soutenu le : 2 juillet.

Devant le Jury composé de :

<i>Nom & Prénom</i>	<i>Grade</i>	<i>Qualité</i>	<i>Etablissement</i>
<i>Mme. MESSALI .Zoubaida</i>	<i>Professeur</i>	<i>Président</i>	<i>Univ-BBA</i>
<i>Mr. HACINE GHARBI ABDENOUR</i>	<i>MCA</i>	<i>Encadreur</i>	<i>Univ-BBA</i>
<i>Mr. BENGUEDDOUDJ ABDALLAH</i>	<i>MCB</i>	<i>Co.Encadreur</i>	<i>Univ-BBA</i>
<i>Mr. TALBI MOHAMED LAMIN</i>	<i>MCA</i>	<i>Examineur</i>	<i>Univ-BBA</i>

Année Universitaire 2022/2023

Remerciements

Avant tout, nous remercions **Allah** le tout puissant qui nous a guidé tout au long de notre vie, il nous a donné le courage et la patience pour passer tous les moments difficiles, nous a permis d'achever ce travail et de pouvoir le mettre entre vos mains aujourd'hui.

Un travail de recherche, nécessite le concours d'un certain nombre de personnes. Ce mémoire est aujourd'hui l'occasion de remercier toutes les personnes qui ont collaboré à ce travail.

nous tenons à remercier l'encadreur **Dr. HACINE GHARBI ABDENOUR** en le remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur, son aide pratique et sa disponibilité durant notre préparation de ce mémoire.

et **Dr. BENGUEDDOUDJ ABDALLAH** pour son aide pratique

Nous remercions vivement les membres du **jury** qui nous ont fait l'honneur de juger notre travail.

Nos remerciements aussi au département d'électronique de l'université de Mohamed el Bachir el Ibrahim BBA, enseignants et étudiants.

Nous remercions les plus sincères à toutes les personnes qui ont contribué de près ou de loin à l'élaboration de ce mémoire.

Merci à tous

Dédicace



Au tout puissant Allâh

A toi la louange, à la lumière des cieux, de la terre et de ce qu'ils renferment. Gloire à toi de nous avoir assisté de ta lumière et en toute circonstance matin et soir.

Je dédie ce Modest travail de fin d'études à ma famille au sens large et à tout mon Entourage mais tout particulièrement à :

Ma chère maman et mon cher père, pour leur patience, conseils, aident et aussi de m'encourager à la réalisation de ce modeste travail.

Je vous remercie mes parents

Aussi mes frères et ma sœur

Spéciale dédicace à mon fiancé et futur mari Je t'aime de tout mon cœur.



Résumé

L'intérêt de la Reconnaissance automatique des Expressions Faciales (REF) est en hausse en raison de l'émergence de nombreuses applications, telles que l'interaction homme-machine, la sécurité et la mesure de la satisfaction client. La reconnaissance en temps réel de l'état émotionnel est cruciale pour ces applications. Un système de REF fonctionne généralement en trois étapes : la détection du visage, l'extraction des caractéristiques et la classification. L'étape d'extraction consiste à convertir chaque image en un ensemble de caractéristiques discriminantes. L'étape de classification utilise généralement des algorithmes d'apprentissage automatique pour classer cet ensemble de caractéristiques en une classe d'expression faciale. Notre objectif est de mettre en œuvre un système REF performant en utilisant l'algorithme d'apprentissage profond CNN (réseaux de neurones convolutifs). Notre contribution principale consiste à chercher la bonne configuration du système REF proposé permettant d'obtenir de bonnes performances en termes de précision et de complexité. Les résultats nous ont montrés de bonnes performances de l'algorithme de classification CNN avec une précision de validation de 85%, une précision de test de 89% et une précision d'apprentissage de 95%, évaluées sur la base de données JAFFE qui comprend 7 classes d'expressions faciales.

الملخص

يتزايد الاهتمام بالتعرف التلقائي على تعبيرات الوجه (REF) بسبب ظهور العديد من التطبيقات ، مثل التفاعل بين الإنسان والحاسوب والأمن وقياس رضا العملاء. يعد التعرف على الحالة العاطفية في الوقت الفعلي أمرًا بالغ الأهمية لهذه التطبيقات. يعمل نظام REF عادةً في ثلاث خطوات: اكتشاف الوجه ، واستخراج الميزات ، والتصنيف. تتكون خطوة الاستخراج من تحويل كل صورة إلى مجموعة من الميزات المميزة. تستخدم خطوة التصنيف عادةً خوارزميات التعلم الآلي لتصنيف هذه المجموعة من الميزات في فئة تعبيرات الوجه. هدفنا هو تنفيذ نظام REF عالي الأداء باستخدام خوارزمية التعلم العميق CNN (الشبكات العصبية التلافيفية). تتمثل مساهمتنا الرئيسية في البحث عن التكوين الصحيح لنظام REF المقترح للحصول على أداء جيد من حيث الدقة والتعقيد. أظهرت النتائج أداءً جيدًا لخوارزمية تصنيف CNN مع دقة تحقق تبلغ 85% ودقة اختبار 89% ودقة تعلم 95%، والتي تم تقييمها على قاعدة بيانات JAFFE التي تضم 7 فئات من التعبيرات الوجهية.

Abstract

Interest in Automatic Facial Expression Recognition (REF) is on the rise due to the emergence of numerous applications, such as human-computer interaction, security, and customer satisfaction measurement. Real-time emotional state recognition is crucial for these applications. A REF system typically works in three steps: face detection, feature extraction, and classification. The extraction step consists of converting each image into a set of discriminating features. The classification step typically uses machine learning algorithms to classify this set of features into a facial expression class. Our goal

is to implement a high-performance REF system using the CNN (convolutional neural networks) deep learning algorithm. Our main contribution is to seek the right configuration of the proposed REF system to obtain good performance in terms of accuracy and complexity. The results showed us good performance of the CNN classification algorithm with a validation accuracy of 85%, a test accuracy of 89% and a learning accuracy of 95%, evaluated on the JAFFE database which includes 7 classes of facial expressions.

Table des matières

Introduction générale	1
1. Chapitre 1 : Généralités sur la reconnaissance des expressions faciales.	
1.1. Introduction.....	2
1.2. Expression faciale émotionnelle	2
1.2.1. Expression faciale.....	2
1.2.2. Émotions.....	2
1.2.3. Les expressions faciales universelles.....	3
1.3. L'importance de choisir le visage.....	4
1.4. La reconnaissance d'expressions faciales.....	5
1.4.1. Architecture générale d'un système de reconnaissance d'expressions faciale...5	
1.4.2. Les méthodes de la reconnaissance des expressions faciale.....	6
1.4.3. Les phases d'un système de la reconnaissance des expressions faciales.....	7
1.5. Extraction des caractéristiques.....	8
1.5.1. Les méthodes d'extraction des caractéristiques.....	8
1.6. Comparaison des caractéristiques (classification) et décision.....	11
1.7. Evaluation des performances	11
1.8. Conclusion.....	12
2. Chapitre 2 : Reconnaissance des expressions faciales basé sur l'apprentissage profond.	
2.1. Introduction.....	11
2.2. L'apprentissage profond (deep Learning).....	12
2.2.1. Définition de l'apprentissage automatique (machine Learning).....	12
2.2.2. Définition de l'apprentissage profond	12
2.2.3. La différence entre l'apprentissage profond et l'apprentissage automatique...13	
2.2.4. Applications de l'apprentissage profond.....	14
2.2.5. Méthodes de l'apprentissage profond	14
2.3. Reconnaissance des expressions facial basées sur (CNN).....	15
2.3.1. Réseaux de neurones artificiels.....	15

2.3.2.	Réseaux de neurones convolutifs (CNN).....	18
A).	La couche de convolution.....	19
B).	Couche d'échantillonnage (Pooling).....	21
C).	Couche complètement connectée.....	22
2.3.4.	Quelques travaux utilisant l'algorithme CNN.....	23
2.4.	Conclusion.....	24
3.	Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.	
3.1.	Introduction.....	26
3.2.	Description de la base de données JAFFE.....	27
3.3.	Système REF basé sur KNN (related Work).....	28
3.4.	Système REF basé sur ANN.....	30
3.5.	Système REF proposé basé sur CNN.....	31
3.6.	Expériences et résultats.....	31
3.6.1.	Expérience 1 : Performances du système REF basé sur le classificateur KNN combiné avec le descripteur HOG	31
3.6.2.	Expérience 2 : Performances du système REF basé sur le classificateur ANN combiné avec le descripteur HOG.....	33
3.6.3.	Expérience 3 : Performances du système REF basé sur le classificateur CNN.....	37
3.6.4.	Expérience 4 : L'application de classificateur CNN sur les images LBP.....	37
3.7.	Conclusion.....	40
	Conclusion générale.....	41

Liste des Figures

Figure 1. 1 : Les six expressions faciales universelles.

Figure 1.2 : Système de la reconnaissance des expressions faciales.

Figure 1.3 : Les phases d'un système de la reconnaissance des expressions faciales.

Figure 1. 4 : Construction d'une forme binaire et calcul du code LBP pour un pixel central.

Figure 1. 5 : Formation du HOG d'une image : (a) découpage en cellules et formation de l'histogramme par bloc, (b) vote d'un pixel selon l'orientation de ses gradients.

Figure 2.1 : La relation entre l'intelligence artificielle, le ML et le DL.

Figure 2.2 : Comparaison entre l'apprentissage automatique et l'apprentissage profond.

Figure 2.3 : Caractérisation d'un neurone artificiel.

Figure 2.4 : Exemple d'un réseau de neurones artificiel.

Figure 2.5 : Deux types de perceptron.

Figure 2.6 : Schéma de fonctionnement d'un réseau de neurones convolutifs.

Figure 2.7 : Calcul des valeurs de sortie d'une convolution discrète. Pour les sous-graphes, les carrés bleu foncé, indiquent les neurones dans la région de filtre. Les carrés vert foncé indiquent le neurone de sortie pour lequel l'activation totale de l'entrée est calculée.

Figure 2.8 : Avec un rembourrage nul, l'effet de la réduction de la taille de sortie est contrecarré pour maintenir la taille d'entrée à la sortie.

Figure 2.9 : Convolution avec rembourrage nul et stride > 1 , avec taille du filtre est de 3×3 .

Figure 2.10 : Exemples de calcul du pooling sur une image 4×4 .

Figure 3.1 : Schéma de fonctionnement de classificateur KNN.

Figure 3.2 : Schéma de fonctionnement du classificateur ANN.

Figure 3.3 : Schéma de système REF basé sur l'algorithme CNN.

Figure 3.4 : Images faciales de la base de données JAFFE.

Figure 3.5 : Précision de validation : 85%.

Figure 3.6 : Les pixels des images LBP.

Figure 3.7 : Précision de validation : 58%.

Liste des Tableaux

Tableau 1.1 : Avantages et inconvénients de la reconnaissance faciale.

Tableau 2.1 : Tableau données quelques travaux qui utilisées le CNN.

Tableau 3.1 : Répartition détaillée des données en base d'apprentissage et base de test.

Tableau 3.2 : Taux de reconnaissance correspond aux différentes tailles de décompositions de l'image en cellules (C_x , C_y).

Tableau 3.3 : Présente le taux de reconnaissance TR correspondant à différentes tailles de décomposition en cellules (C_x , C_y) du descripteur HOG extrait de la région du visage (V).

Abréviations

AG : Approche Globale.

AI : Artificial intelligence.

ANN : Réseau de Neurones Artificiels.

CNN : Convolutional Neural Networks.

DL : Deep Learning.

HOG : Histogram of Oriented Gradients.

JAFFE : Japanese Female Facial Expression.

K-NN : k-Nearest Neighbor.

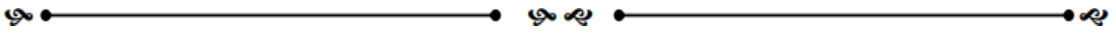
LBP : Local Binary Patterns.

ML : Machine Learning.

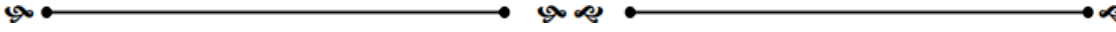
REF : Reconnaissance automatique des Expressions Faciales.

TR : Taux de Reconnaissance.

V : Visage.



Introduction générale



Introduction générale

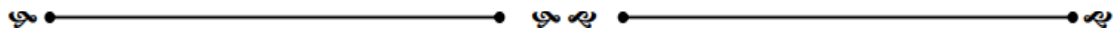
L'expression faciale est l'un des moyens naturels les plus puissants pour communiquer les émotions et les intentions d'un être humain [1]. Des études ont montré l'existence de sept expressions émotionnelles universelles : le dégoût, la colère, le bonheur, la tristesse, la surprise et la peur, ainsi que la neutralité. La reconnaissance en temps réel de l'état émotionnel basée sur l'image faciale est importante pour des applications telles que l'interaction homme-machine, aide au diagnostic médical, la sécurité et la mesure de la satisfaction du client en Marketing [2].

La conception d'un système de reconnaissance d'expressions faciales (REF) se base sur une phase d'apprentissage permettant la modélisation des différentes classes d'expressions faciales, ainsi qu'une base de test permettant d'évaluer les performances du système REF en utilisant une base de données d'images faciales. Les systèmes REF basés sur les algorithmes d'apprentissage automatique exigent une étape d'extraction de caractéristiques permettant de convertir chaque image faciale en un ensemble de caractéristiques discriminantes des différentes classes d'expressions faciales. La phase de test se base sur une étape de reconnaissance de chaque image faciale d'entrée en utilisant des algorithmes d'apprentissage automatique tels que : KNN, ANN, GMM, SVM, etc.

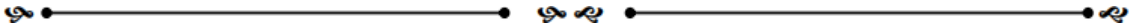
Ces dernières années, les algorithmes d'apprentissage profond (deep learning), en particulier les réseaux de neurones convolutifs (convolutional neural network : CNN), ont été largement utilisés pour résoudre des problèmes complexes de machine learning.

Notre objectif est d'implémenter un système REF basé sur l'algorithme CNN permettant d'obtenir les meilleures performances comparées avec d'autres algorithmes d'apprentissage automatique tels que : KNN, ANN. Ces performances seront évaluées sur la base de données des images faciales JAFFE [3] [4].

Ce manuscrit est organisé en trois chapitres, donnés respectivement comme suit: Généralités sur la reconnaissance des expressions faciales, Reconnaissance des expressions faciales basée sur l'apprentissage profond, Implémentation du système REF basé sur l'algorithme CNN. Nous terminons ce manuscrit par une conclusion générale incluant des perspectives.



*1. Chapitre 1 : Généralités sur la
reconnaissance des expressions
faciales.*



1.1. Introduction

Ces dernières années, la vision par ordinateur a suscité un grand intérêt dans l'analyse et la reconnaissance automatique des expressions faciales dans les images animées et fixes. De nombreuses recherches se sont concentrées sur la classification des expressions faciales en grandes catégories émotionnelles telles que la joie, la tristesse, la colère, la surprise, la peur et le dégoût [5].

Dans ce chapitre nous présenterons des généralités sur les expressions faciales émotionnelles.

1.2. Expression faciale émotionnelle

Les expressions faciales émotionnelles sont les changements du visage qui reflètent les états émotionnels, les intentions ou les communications sociales d'une personne. La reconnaissance d'une expression faciale (REF) se concentre sur la classification des mouvements et des déformations du visage en classes abstraites basées sur des informations visuelles. Cependant, la reconnaissance des émotions humaines est un processus plus complexe qui prend en compte divers facteurs tels que la voix, la posture, les gestes, la direction du regard et l'expression faciale. La correspondance entre une expression faciale et une émotion implique la compréhension des catégories d'émotions humaines associées aux expressions faciales. Cela représente un défi dans la recherche et la vision par ordinateur [6].

Lors de l'interprétation des expressions faciales, ce sont ces déformations que nous observons, à partir desquelles nous devons déduire l'émotion sous-jacente [7].

1.2.1. Expression faciale

Une expression faciale est un ensemble de signes faciaux qui reflètent une sensation faciale altérée, perceptible à l'œil nu, provoquée par l'activation d'un ou plusieurs des 44 muscles qui composent le visage, il existe 250 000 expressions possibles.

Les expressions faciales sont très importantes pour connaître l'état d'une personne [8].

1.2.2. Émotion

L'émotion est un concept vague et très difficile à définir. De ce fait, il existe plusieurs

définitions du mot « émotion ». Ces états internes peuvent être positifs agréables comme la joie ou négatifs désagréables, comme la colère. Les scientifiques Ekman, Izard, Plutchik, Tomkins et MacLean ont développé la théorie des émotions de bases ou fondamentales, mais seules 6 émotions de base sont communes aux divers auteurs [8].

Les émotions ne sont pas des états subjectifs, intérieurs à une personne, ou du moins pas en première instance [9]. Et il est possible de mimer une expression représentant une émotion sans pour autant ressentir cette émotion.

1.2.3. Les expressions faciales universelles

Les expressions faciales universelles ont été étudiées par le psychologue américain Paul Ekman, qui a identifié six émotions de base qui sont communément exprimées à travers le visage dans toutes les cultures humaines. Ces six émotions universelles sont :

- ✚ **Joie** : La joie se manifeste par un sourire et des yeux plissés.
- ✚ **Tristesse** : La tristesse se manifeste par des yeux humides et un visage tombant.
- ✚ **Peur** : La peur se manifeste par des yeux écarquillés et une bouche ouverte.
- ✚ **Colère** : La colère se manifeste par un front plissé, des sourcils froncés et une mâchoire serrée.
- ✚ **Dégoût** : Le dégoût se manifeste par un renflement de la lèvre supérieure et une contraction de la narine.
- ✚ **Surprise** : La surprise se manifeste par des yeux grands ouverts et une bouche ouverte.

Ces expressions faciales universelles sont considérées comme innées et instinctives, et peuvent être reconnues par les gens de toutes les cultures, indépendamment de leur origine ou de leur langue [10].



Figure 1. 1 : Les six expressions faciales universelles.

1.3. L'importance de choisir le visage

Le choix du visage comme méthode de reconnaissance reste la plus préférée car elle correspond à ce que les humains utilisent naturellement dans l'interaction visuelle. Comparée aux autres modalités, la reconnaissance faciale présente plusieurs avantages. Tout d'abord, c'est une méthode non intrusive, ce qui signifie qu'elle ne nécessite pas la coopération du sujet. De plus, les capteurs utilisés pour la reconnaissance faciale sont généralement peu coûteux [11].

Voici un tableau présentant les avantages et inconvénients de la reconnaissance faciale.

Tableau 1.1 : Avantages et inconvénients de la reconnaissance faciale [11].

Avantage	Inconvénients
<ul style="list-style-type: none">• Cette technologie jouit d'une forte acceptation de la part du public.• Cette technologie est peu intrusive car elle ne nécessite aucune action de la part de l'utilisateur.• Cette technologie n'implique aucun contact physique.• Cette technique est économiquement abordable	<ul style="list-style-type: none">• La technologie de reconnaissance faciale est sensible à l'environnement (éclairage, position, expression faciale).• Elle peut également être affectée par des changements tels que la présence de barbe, de moustache, de lunettes, de piercing ou de chirurgie esthétique.

1.4. La reconnaissance d'expressions faciales

La reconnaissance faciale est une technologie utilisée pour identifier ou vérifier l'identité d'une personne en analysant les caractéristiques distinctives de son visage. Elle repose sur l'utilisation d'algorithmes informatiques capables de détecter, extraire et comparer les traits faciaux tels que la forme du visage, les yeux, le nez, la bouche, etc. Ces algorithmes peuvent être entraînés à reconnaître des visages à partir d'une base de données préalablement constituée, ce qui permet de les utiliser dans divers domaines, tels que la sécurité, la surveillance, le contrôle d'accès, la gestion des identités, etc. La reconnaissance faciale peut être effectuée à l'aide de caméras ou d'images déjà enregistrées, et elle peut être utilisée de manière automatisée ou manuelle. Cependant, il convient de noter que la reconnaissance faciale soulève également des questions liées à la protection de la vie privée et aux préoccupations éthiques, en raison de la collecte et de l'utilisation potentiellement intrusive des données biométriques des individus [12].

1.4.1. Architecture générale d'un système de reconnaissance d'expressions faciales

Un système de la reconnaissance faciale peut être structuré comme suit :

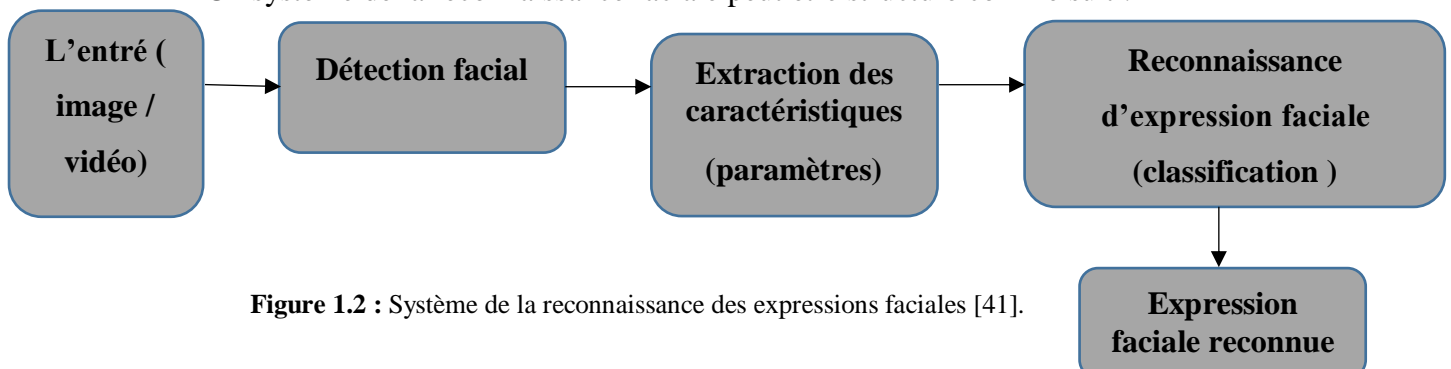


Figure 1.2 : Système de la reconnaissance des expressions faciales [41].

1.4.2. Les méthodes de la reconnaissance d'expressions faciales

Différentes approches ont été développées dans le domaine de la reconnaissance des expressions faciales. Elles peuvent être divisées en trois catégories, à savoir : les méthodes locales et les méthodes globales [10].

Approche locale

Les approches locales pour la reconnaissance des expressions faciales consistent à utiliser des modèles qui traitent séparément différentes régions de l'image faciale. Ces modèles sont basés sur les connaissances préalables que nous avons de la morphologie des visages. Dans la plupart des cas, cela implique la détection et l'extraction de caractéristiques faciales spécifiques à des régions particulières [13].

Approches globales

Les approches globales de reconnaissance faciale prennent en compte l'ensemble de l'image du visage et utilisent couramment des techniques d'analyse statistique. Elles effectuent souvent une projection de l'image vectorisée dans un espace de dimension réduite pour faciliter la reconnaissance. Ces méthodes sont rapides à implémenter grâce à des opérations matricielles simples, mais elles sont sensibles aux variations de luminosité, de pose et d'expression faciale. Elles reposent sur l'observation selon laquelle les visages se trouvent dans un sous-espace de l'espace de l'image d'entrée. Les méthodes globales peuvent être de nature linéaire ou non linéaire [13].

Approches hybrides

Les approches hybrides sont le résultat de la combinaison des méthodes locales et globales. Elles intègrent à la fois la détection de caractéristiques spécifiques à des régions du visage et l'extraction de caractéristiques globales. Ces techniques cherchent à exploiter les avantages des deux types de méthodes mentionnées précédemment. En utilisant cette approche, on cherche à bénéficier des meilleures caractéristiques locales et globales pour améliorer la reconnaissance faciale [13].

1.4.3. Les phases d'un système de la reconnaissance des expressions faciales

Chaque système REF comporte deux phases :

- La phase d'apprentissage.
- La phase de test.

Les phases de fonctionnement d'un système sont illustrées sur la figure suivante.

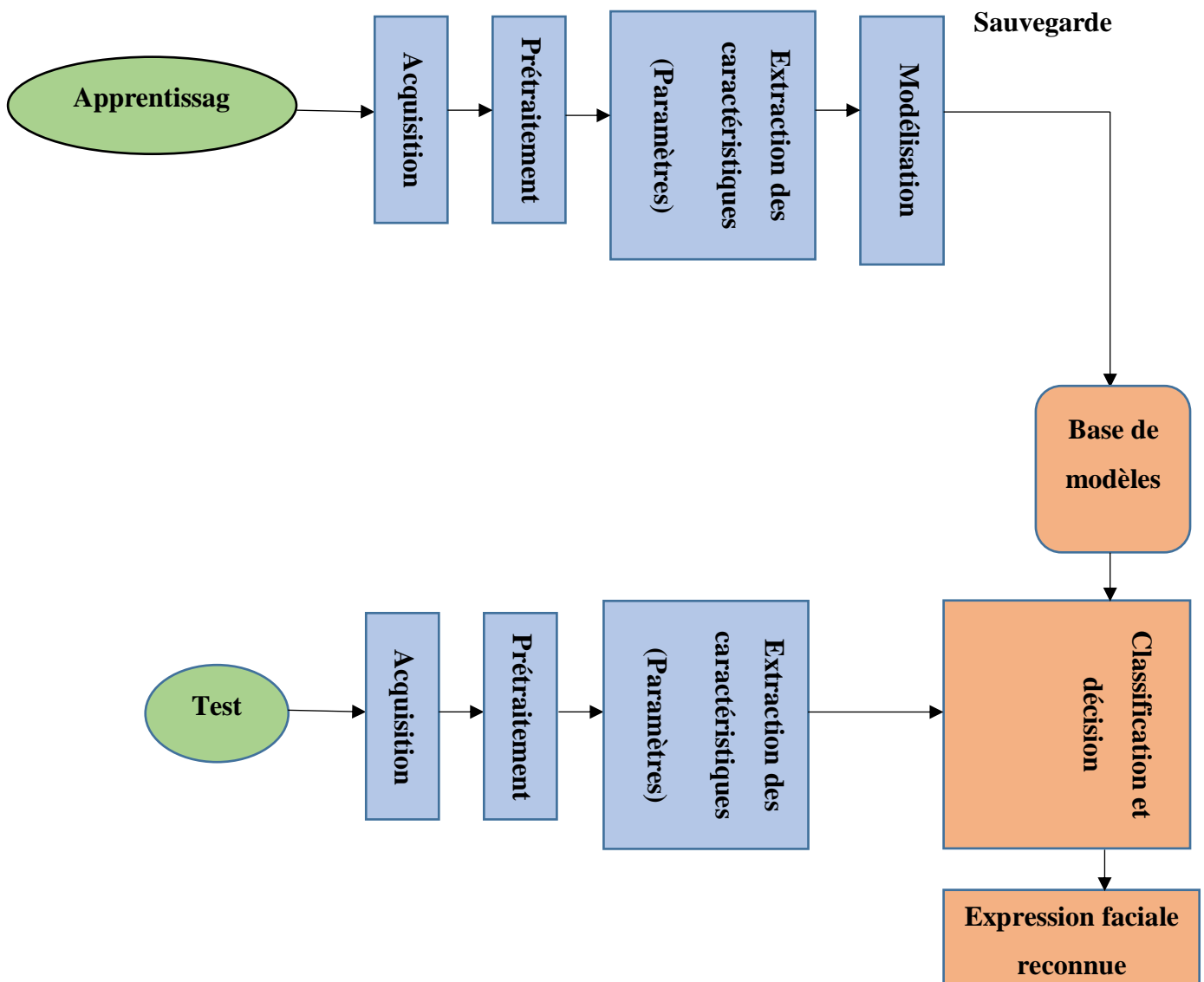


Figure 1.3 : Les phases d'un système de la reconnaissance des expressions faciales [11].

1.5. Extraction des caractéristiques

Dans la reconnaissance des expressions faciales, l'étape d'extraction des caractéristiques vise à identifier et à extraire les caractéristiques discriminantes des différentes classes d'expressions faciales [14]. Les descripteurs de caractéristiques couramment utilisés dans les systèmes REF sont les descripteurs LBP et HOG.

1.5.1. Les méthodes d'extraction des caractéristiques

✚ LBP (Local Binary Patterns)

La technique de motifs binaires locaux LBP [15] consiste à transformer une image en échelle de gris en une image (matrice) LBP dont chaque pixel est représenté par un mot de code LBP obtenu comme illustre la figure suivante.

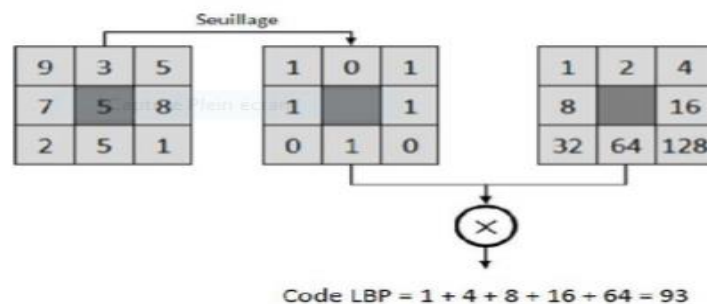


Figure 1.4 : Construction d'une forme binaire et calcul du code LBP pour un pixel central [4].

Le code LBP possède une caractéristique essentielle : il reste invariable face aux variations globales d'illumination. Cette invariance est rendue possible par le fait que le code LBP d'un pixel ne dépend que des écarts entre son niveau de gris et ceux de ses voisins.

L'image considérée peut être ensuite représentée par un vecteur constitué des valeurs de l'histogramme de la matrice LBP.

✚ HOG (Histogramme de Gradients Orientés)

HOG a été proposé par les chercheurs d'Inria de Grenoble Navneet Dalal et Bill Triggs lors de la conférence CVPR de juin 2005.

L'histogramme de gradient directionnel est un descripteur de caractéristiques utilisé dans la vision par ordinateur et le traitement d'image pour la détection d'objets.

Cette méthode est similaire aux histogrammes orientés bords, aux descripteurs de caractéristiques invariants à l'échelle et aux contextes de forme, mais diffère en ce qu'elle est calculée sur une grille dense de cellules régulièrement espacées.

Étant donné que le descripteur d'histogramme de pente directionnelle fonctionne sur des parcelles localisées, cette méthode maintient l'invariance aux transformations géométriques et photométriques, qui n'apparaîtront que dans de grandes régions spatiales [16].

➤ Les étapes pour calcul le HoG

✚ Etape 1

L'image (x, y) est subdivisée en N cellules régulières, constituées de petites régions et de blocs. Les valeurs de gradient sont calculées pour chaque pixel en utilisant un filtre dérivatif 1-D centré dans les directions horizontales et verticales. A cet effet, les masques suivants sont employés [4].

$$D_x = [-1 \ 0 \ 1] \quad (1.1)$$

$$D_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad (1.2)$$

$$\begin{cases} G_x(x, y) = I(x, y) * D_x \\ G_y(x, y) = I(x, y) * D_y \end{cases} \quad (1.3)$$

✚ Etape 2

Le gradient d'un pixel dans une image numérique est représenté par un vecteur qui possède une amplitude et une direction. L'amplitude est directement liée à la quantité de variation des niveaux de gris locaux. La direction du gradient est perpendiculaire à la frontière qui passe par le point en question [4].

L'amplitude et l'orientation du gradient sont calculées pour chaque pixel (x, y) en utilisant les équations (1.4) et (1.5) respectivement.

$$|G(x, y)| = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (1.4)$$

$$\alpha(x, y) = \arctan\left(\frac{G_x(x, y)}{G_y(x, y)}\right) \quad (1.5)$$

$G_x(x, y)$ et $G_y(x, y)$ sont le gradient horizontal et le gradient vertical, au pixel (x, y) , respectivement.

✚ Etape 3

L'image est subdivisée en plusieurs petites cellules (Figure a), et pour chaque cellule, un histogramme est calculé. Chaque pixel dans une cellule contribue à un vote pour une orientation qui peut varier de 0 à 180 degrés dans le cas non signé (Figure b), ou de 0 à 360 degrés dans le cas signé [4].

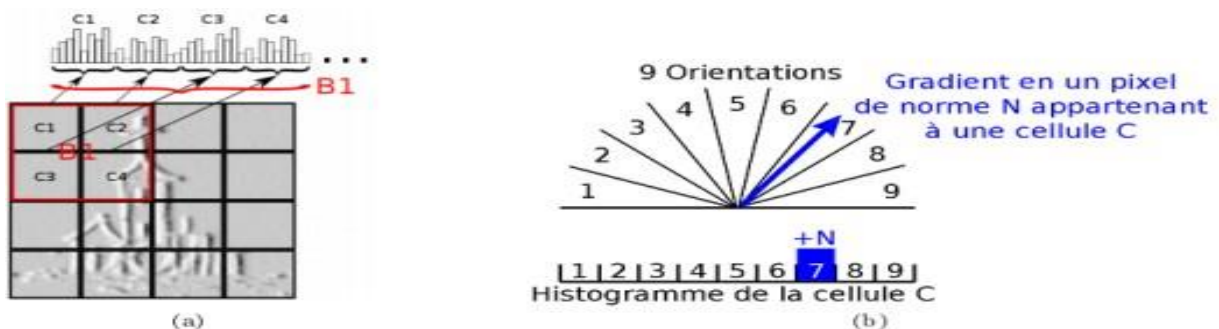


Figure 1. 5 : Formation du HOG d'une image : (a) découpage en cellules et formation de l'histogramme par bloc, (b) vote d'un pixel selon l'orientation de ses gradients.

✚ Etape 4

Pour former le vecteur caractéristique de chaque bloc (c), on concatène les vecteurs d'histogramme de toutes les cellules présentes dans ce bloc. Finalement, le vecteur caractéristique HOG est formé en concaténant les vecteurs de caractéristiques de tous les blocs pour une image donnée [4].

$$V_n = \frac{v}{\sqrt{\|V\|_2^2 + \varepsilon^2}} \quad (1.6)$$

Où

V : est le vecteur non normalisé contenant tous les histogrammes d'un bloc donné.

V_2 : est la norme L_2 du vecteur (v) du descripteur.

s : est une petite constante.

Il existe d'autre facteur de normalisation comme L1-norme et L2-sqrt, définies par :L1-norme :

$$V_n = \frac{v}{\sqrt{\|V\|_1 + \varepsilon^2}} \quad (1.7)$$

L2-norme donne de bons résultats par rapport à L1-norme, mais toutefois bien meilleurs que l'absence de normalisation.

1.6. Comparaison des caractéristiques (classification) et décision

Le processus consiste à représenter des paramètres extraits des caractéristiques communes d'une expression ou un ensemble d'expressions faciales. Un modèle est un ensemble d'informations pertinentes, distinctives et non redondantes qui caractérisent une ou plusieurs expressions présentant des similitudes, et ceux-ci sont regroupés dans la même catégorie, les catégories variant en fonction du type de décision. Les algorithmes de comparaison diffèrent en fonction des caractéristiques précédemment extraites. Plusieurs approches sont disponibles, la plus simple consistant à calculer la distance pour trouver des similarités. D'autres méthodes reposent sur la classification des entités à l'aide de classificateurs tels que les machines à vecteurs de support (SVM), les réseaux neuronaux (NN), l'analyse discriminante linéaire (LDA), le k-plus proche voisin (KNN), la régression logistique multinomiale (MRL), les modèles de Markov cachés (HMM), les réseaux bayésiens (BN), etc., ou utilisent des techniques spécifiques comme Adaboost. L'apprentissage consiste à mémoriser les représentations calculées lors de la phase d'analyse pour les expressions connues. Généralement, les étapes d'analyse et d'apprentissage sont combinées en une seule étape [11].

1.7. Evaluation des performances

L'évaluation des performances d'un système de reconnaissance est une étape importante dans son processus de conception, qui permet de mesurer sa précision, sa rapidité et sa complexité. Par conséquent, cette étape nécessite de définir certains critères d'évaluation, tels que le taux de reconnaissance, le temps de calcul, l'espace de stockage, etc. Dans ce travail, nous nous

intéressons au taux de reconnaissance comme mesure de précision donnée par la formule suivante [4] :

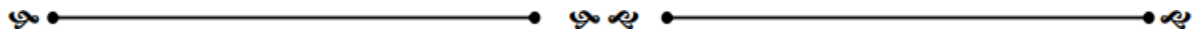
$$TR = \frac{H}{N} * 100\%$$

Où : H est le nombre des images faciales reconnues, N est le nombre total des images faciales de la base de test.

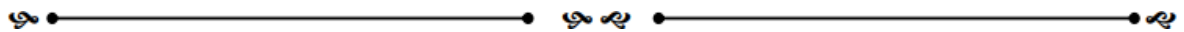
1.8. Conclusion

Dans ce chapitre intitulé « Généralités sur la reconnaissance des expressions faciales » nous avons présenté les caractéristiques générales des expressions faciales émotion. Ensuite, nous avons montré également la relation entre les expressions faciales et sentiment. Nous avons ensuite décrit la représentation et les propriétés des expressions visages et émotions. En plus de cela, nous avons mentionné certaines méthodes d'extraction des caractéristiques et certains classificateurs que les gens utilisaient pour classifier les images.

Dans le chapitre suivant nous présenterons l'apprentissage profond en citant les différents types de réseaux de neurones et en détaillant les réseaux de neurones convolutifs.



*2. Chapitre 2 : Reconnaissance des
expressions faciales basée sur
l'apprentissage profond.*



2.1. Introduction

L'apprentissage profond, fait partie de l'apprentissage automatique, se focalise sur les algorithmes qui s'inspirent de la structure et du fonctionnement du cerveau. Ces algorithmes sont capables d'apprendre différents niveaux de représentation afin de modéliser des relations complexes entre les données. Récemment, les progrès réalisés grâce au Deep Learning ont considérablement amélioré les performances des systèmes d'analyse des expressions faciales. Plus particulièrement, les réseaux de neurones convolutifs (Convolutional Neural Networks, CNN) se sont révélés particulièrement efficaces dans ce domaine [17]. Dans le deuxième chapitre, nous examinerons plus en détail l'apprentissage profond (Deep Learning) en se basant sur les réseaux CNN.

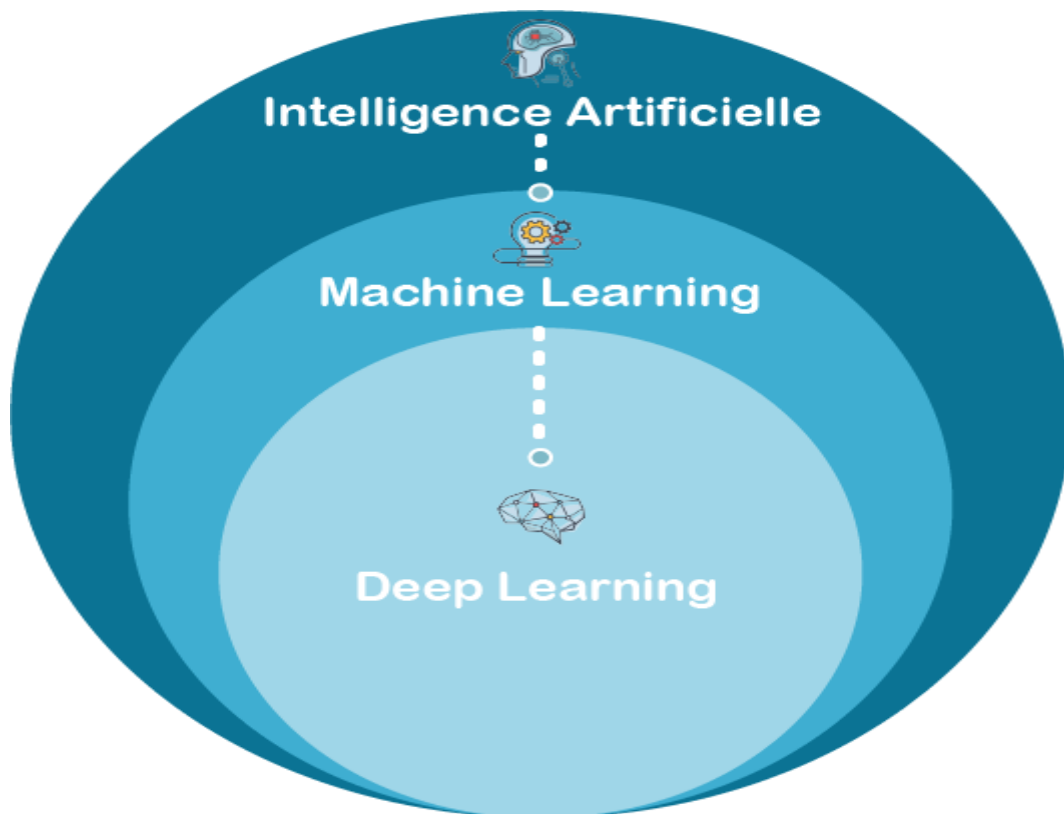


Figure 2.1 : La relation entre l'intelligence artificielle, le ML et le DL [18].

2.2. L'apprentissage profond (Deep Learning)

2.2.1. Définition de l'apprentissage automatique (Machine Learning)

Le ML est un domaine de l'IA qui donne aux ordinateurs la capacité d'apprendre à partir d'un ensemble d'observations appelé ensemble d'apprentissage.

Voici quelques exemples utilisant le ML :

- Classification (image, texte, vidéo, son, ...).
- Détection de fraude [19].
- Vision par ordinateur [20].
- Diagnostic médical [21].
- Les publicités ciblées [22].

2.2.2. Définition de l'apprentissage profond

L'apprentissage profond se réfère à un ensemble d'algorithmes d'apprentissage automatique qui cherchent à acquérir des connaissances à différents niveaux d'abstraction, en utilisant plusieurs couches de traitement. Ces algorithmes ont la capacité d'extraire des caractéristiques à partir de données brutes grâce à des transformations linéaires et non linéaires effectuées à travers ces couches. De plus, ils sont capables d'apprendre progressivement à partir de ces caractéristiques, avec une intervention humaine minimale [23].

En deep Learning (L'apprentissage profond), en particulier pour la détection d'images, le réseau de neurones est capable de découvrir automatiquement les composants nécessaires, en évoluant d'un niveau d'abstraction bas vers un niveau plus élevé, couche par couche. Le deep Learning est couramment utilisé pour le traitement du langage, de la parole, du bruit, de l'écriture et des images. Il trouve également d'autres applications dans les outils d'aide à la prise de décision [24].

2.2.3. La différence entre L'apprentissage profond et l'apprentissage automatique

Les algorithmes ML fonctionnent bien pour une variété de problèmes. Cependant, ils n'ont pas réussi à résoudre certains des gros problèmes d'IA comme la reconnaissance vocale et reconnaissance d'objets [23]. Tout d'abord, les différents algorithmes d'apprentissage profond n'entrent en jeu que lorsque l'apprentissage automatique ne parvient pas à résoudre de nombreux problèmes d'intelligence artificielle (IA) [24] :

- Pour développer une grande quantité de données comme le big data.
- Améliorer le développement des algorithmes traditionnels dans des tâches telles que l'intelligence artificielle.
- Extraire automatiquement les fonctionnalités.
- S'adapter à tout type de problème.

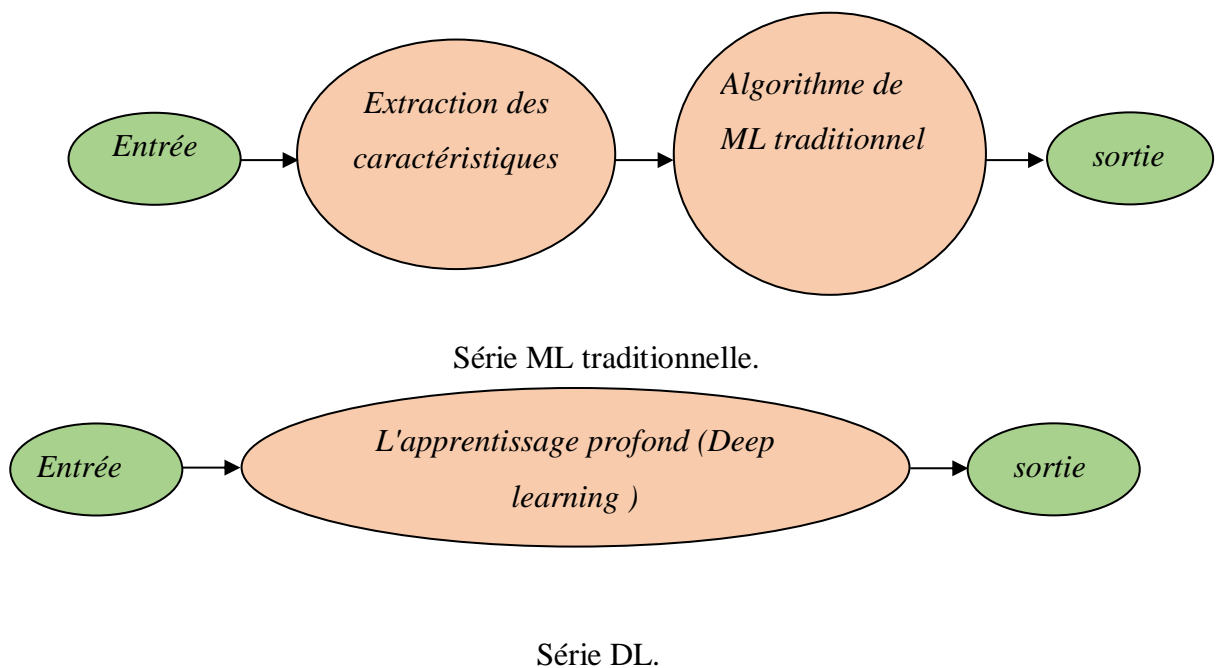


Figure 2.2 : Comparaison entre l'apprentissage automatique et l'apprentissage profond [9].

Ainsi, Deep Learning utilise des réseaux de neurones pour apprendre des représentations utiles des caractéristiques directement à partir des données [14].

2.2.4. Applications de l'apprentissage profond

Le Deep Learning a montré ses performances sur différents domaines tels que la détection des fraudes et la restauration des pixels. L'apprentissage profond est un secteur en pleine expansion, qui trouve ses applications dans les différents domaines d'activité [2] :

- Voitures autonomes.
- Agrégation d'actualités et détection d'actualités frauduleuses.
- Traitement du langage naturel.
- Assistants virtuels.
- Divertissement.
- Reconnaissance visuelle.
- Détection de fraude.
- Soins de santé.
- Personnalisations.
- Détecter un retard de développement chez les enfants.
- Correction des couleurs des photos en noir et blanc.
- Ajout de sons aux films muets.
- Traduction automatique.
- Génération automatique d'écriture manuscrite.
- Jeu automatique.
- Traductions linguistiques.
- Restauration de pixels.
- Descriptions des photos.
- Prédictions démographiques et électorales.

2.2.5. Types de l'apprentissage profond

IL y a plusieurs méthodes de l'apprentissage profond, telles que :

1. Réseaux de neurones artificiels (Artificial Neural Networks - ANN).
2. Réseaux de neurones convolutifs (Convolutional Neural Networks - CNN).
3. Réseaux de neurones récurrents (Recurrent Neural Networks - RNN).

4. Réseaux de neurones récurrents bidirectionnels (Bidirectional Recurrent Neural Networks - BiRNN).
5. Réseaux de neurones générateurs adverses (Generative Adversarial Networks - GAN).
6. Réseaux de neurones récurrents à mémoire à court terme (Long Short-Term Memory - LSTM).
7. Réseaux de neurones récurrents à portes (Gated Recurrent Units - GRU).
8. Autoencodeurs (Autoencoders).

2.3. Reconnaissance des expressions faciales basées sur (CNN)

2.3.1. Réseaux de neurones artificiels

a). Principe de fonctionnement

Le neurone artificiel, qui a été défini en 1959 [25], est l'unité fondamentale de calcul dans un réseau de neurones artificiels. Ce neurone reçoit des entrées provenant d'autres neurones ou d'une source externe, sous la forme de valeurs numériques x_1, x_2, \dots, x_n , avec lesquelles il est connecté par des synapses. Il calcule ensuite une sortie y . Chaque entrée x_i est associée à un poids w_i , qui est attribué en fonction de son importance relative par rapport aux autres entrées. La valeur d'entrée x d'un neurone est obtenue en faisant la somme pondérée de ses entrées, en ajoutant une autre entrée avec un poids b appelé biais. Ensuite, le neurone applique une fonction f à cette somme, [26] comme illustré dans la Figure 2.3.

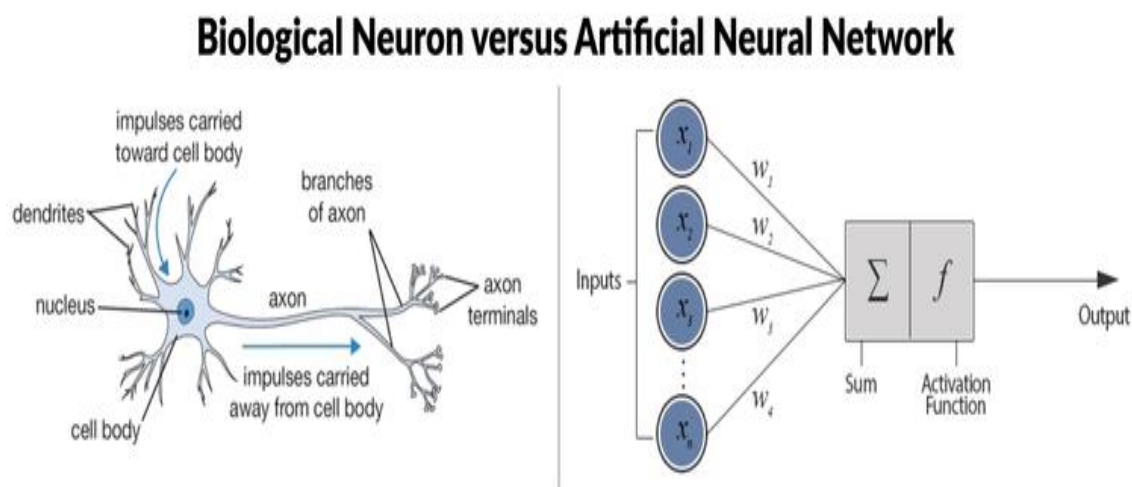


Figure 2.3 : Caractérisation d'un neurone artificiel [27].

✚ **Y** est la sortie du neurone obtenue par l'application de la formule suivante :

$$y = f\left(\sum_{i=1}^n W_i X_i + b\right)$$

✚ **La fonction f** : appelée fonction d'activation, est une fonction non linéaire. Son but est d'introduire une non-linéarité dans la sortie des neurones. Ceci est important car la plupart des données du monde réel sont non linéaires et les neurones apprendront donc ces représentations non linéaires. Cette fonction doit également être différentiable pour pouvoir ensuite appliquer l'algorithme de rétro-propagation. Chaque fonction d'activation prend un numéro unique et effectue des calculs en dessus. Il existe plusieurs fonctions d'activation utilisées dans la pratique comme suit [26] :

- la fonction Sigmoidé.
- la fonction Tangente hyperbolique.

Un exemple de réseau de neurones à propagation avant est présenté sur la figure 2.4 :

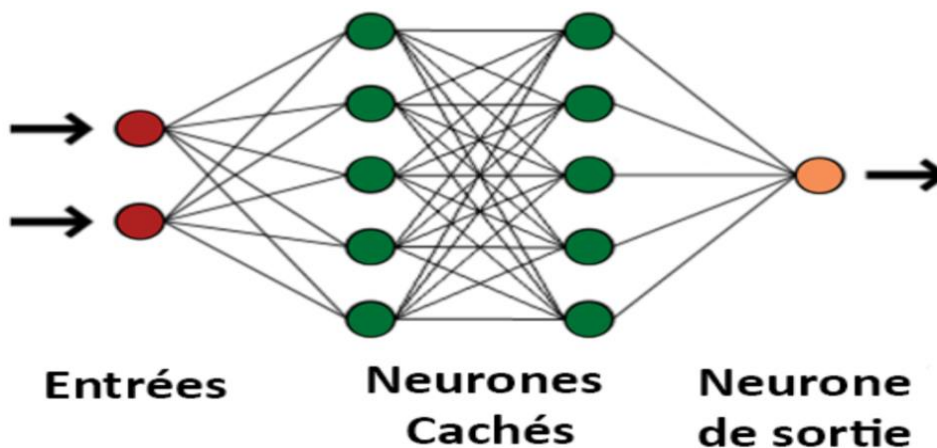


Figure 2.4 : Exemple d'un réseau de neurones artificiel [26].

Un réseau de neurones à propagation directe peut être constitué de trois types de neurones :

✚ **Neurones d'entrée:** les neurones d'entrée fournissent des informations provenant du monde extérieur au réseau et forment ensemble une « couche d'entrée ». Aucun calcul n'est effectué dans les nœuds d'entrée. Ils transmettent simplement des informations aux nœuds cachés [26].

✚ **Neurones cachés:** les neurones cachés qui n'ont aucun lien direct avec le monde extérieur (d'où le nom "caché"). Ils effectuent des calculs et transfèrent des informations des neurones d'entrée vers les neurones de sortie. Un ensemble de neurones cachés forme une "couche cachée" [26].

✚ **Neurones de sortie :** les neurones de sortie forment ensemble une "couche de sortie". Ils sont responsables du calcul et du transfert des informations du réseau vers le monde extérieur [26].

Dans un réseau à propagation directe, l'information est partagée dans une seule direction, en partant des neurones d'entrée, en passant par les neurones cachés (s'ils existent) et vers les neurones de sortie. Un réseau de neurones peut être assimilé à un graphe orienté cycliquement. Deux exemples de réseaux de propagation vers l'avant [26] sont illustrés sur la Figure 2.5 :

- **perceptron mono-couche** : Il s'agit du réseau de neurones à propagation directe le plus simple contenant une couche cachée.

- **perceptron multi-couche** : Il a une ou plusieurs couches cachées, le cas des perceptrons multicouches, est plus utile que les perceptrons mono-couches dans les applications actuelles.

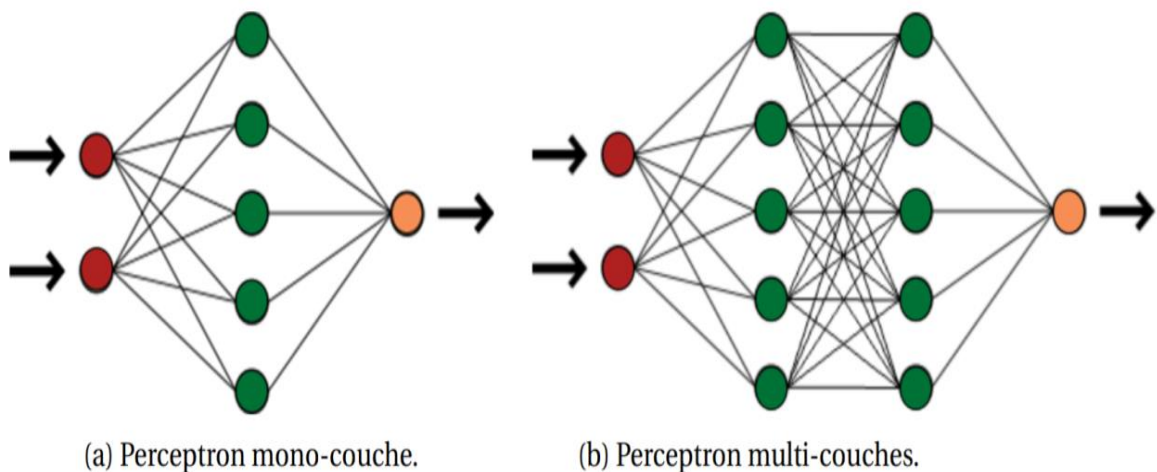


Figure 2.5 : Deux types de perceptron [26].

2.3.2. Réseaux de neurones convolutifs (CNN)

Les réseaux de neurones convolutifs (CNN) sont appliqués dans divers domaines de la reconnaissance d'images et ont des performances élevées montrant que la structure CNN peut identifier les visages, les individus, les panneaux de signalisation et de nombreux autres aspects des données visuelles. L'architecture générale type de CNN est illustrée à la Figure 2.8 [28].

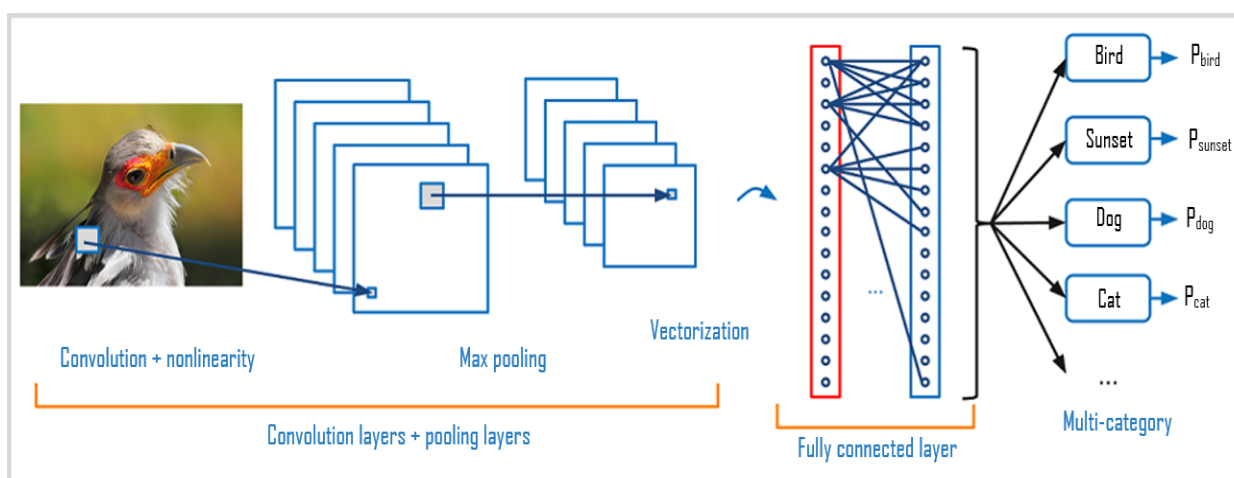


Figure 2.6 : Schéma de fonctionnement d'un réseau de neurones convolutifs [29].

A). La couche de convolution

La couche de convolution est une composante essentielle des réseaux de neurones convolutifs. Son rôle principal est de détecter la présence de caractéristiques spécifiques dans les images en entrée. Pour ce faire, elle applique un processus de filtrage par convolution. Le principe consiste à faire glisser une fenêtre représentant la caractéristique sur l'image et à calculer le produit de convolution entre la caractéristique et chaque partie de l'image balayée. Cela permet de générer une carte de caractéristiques en sortie [17]. La figure 2.7 montre un exemple de la façon dont une convolution est calculée avec un filtre sur les trois premières positions de filtre dans une image [30].

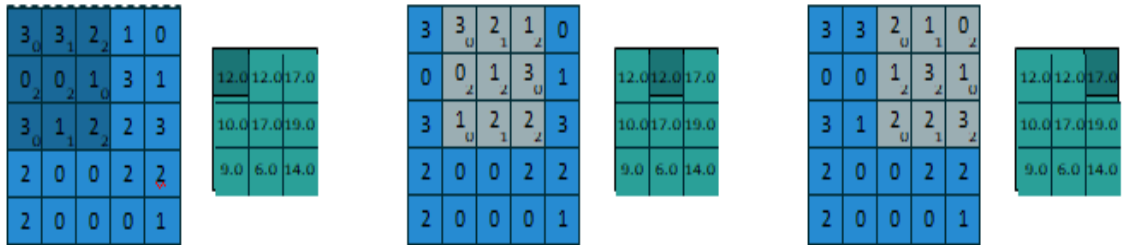


Figure 2.7 : Calcul des valeurs de sortie d'une convolution discrète. Pour les sous-parcelles, les carrés bleu foncé indiquent les neurones dans la région du filtre. Les carrés vert foncé indiquent le neurone de sortie pour lequel l'activation d'entrée totale est calculée [30].

Avant d'effectuer la convolution, il est nécessaire de déterminer certains paramètres qui contribuent au calcul de la taille des images en sortie, c'est-à-dire la carte des caractéristiques. Ces paramètres incluent :

✚ Rembourrage (Padding)

Le padding consiste à ajouter des pixels sur les bords de l'image afin de permettre l'application de la convolution aux bordures de l'image (voir Figure 2.8). Le nombre de lignes et de colonnes ajoutées dépend de la taille du filtre de convolution utilisé [31].

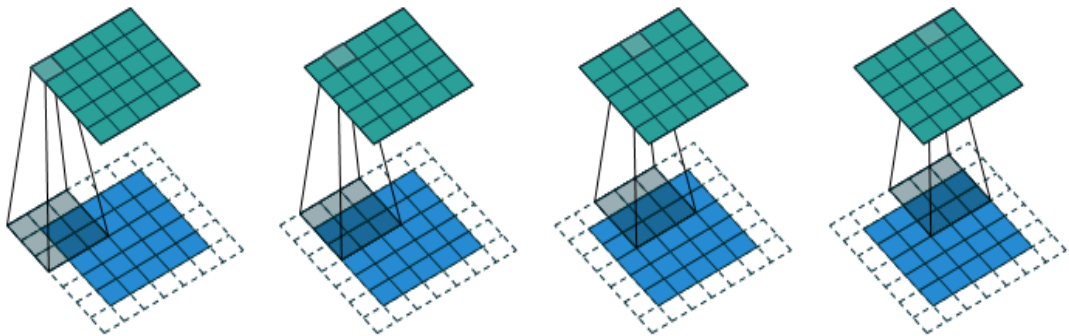


Figure 2.8 : Avec un rembourrage nul, l'effet de la réduction de la taille de sortie est contrecarré pour maintenir la taille d'entrée à la sortie [30].

✚ Le pas (Stride)

Le pas fait référence au nombre de pixels par lesquels le filtre se déplace à chaque itération lors de l'application de la convolution sur une matrice d'entrée. Lorsque le pas est fixé à 1, le filtre se déplace d'un pixel à la fois. Si le pas est de 2, le filtre se déplace de deux pixels à la fois, et ainsi de suite [17].

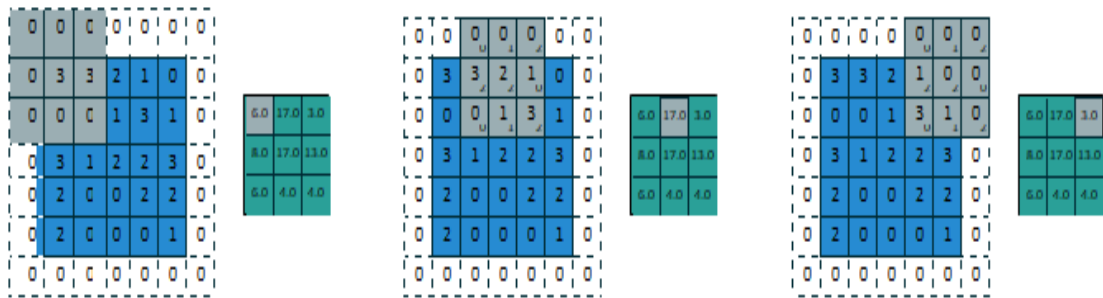


Figure 2.9 : Convolution avec rembourrage nul et stride > 1, avec taille du filtre est de 3×3 [30].

✚ Arithmétique convolutive

Lors de la conception d'un réseau, il est nécessaire de connaître l'impact des paramètres sur les dimensions de sortie. Nous utiliserons la variable N comme la longueur d'un côté de la couche d'activation actuelle. P représente la profondeur du rembourrage. F est la longueur d'un côté du filtre. S représente la longueur du pas. Les équations ci-dessous décrivent la relation entre la taille de la sortie, la taille de l'entrée, la taille du filtre, le rembourrage et le pas.

$$L_{\text{output}} = (N - F)/S + 1 \quad (1)$$

$$L_{\text{output}} = (N + 2P - F)/S + 1 \quad (2)$$

L'équation 1 calcule la longueur d'un côté d'une couche d'activation après une convolution sans rembourrage de zéro. L'équation 2 tient compte du rembourrage de zéro [30].

B). Couche d'échantillonnage (Pooling)

La couche de pooling est une méthode mathématique utilisée pour réduire la taille d'une image sans perdre les informations les plus importantes. Son principe consiste à diviser l'image en petites matrices, par exemple 2x2, et à sélectionner la valeur maximale de chaque matrice

pour obtenir une image réduite. Cela permet de réduire la dimension spatiale de l'image tout en préservant les caractéristiques essentielles [17]. Il y'a Beaucoup de types de pooling [32] mais le type le plus utilisée est le « max pooling » :

✚ somme maximale (max pooling)

C'est le type le plus utilisé, car il est rapide à calculer (immédiat), et permet de simplifier efficacement l'image.

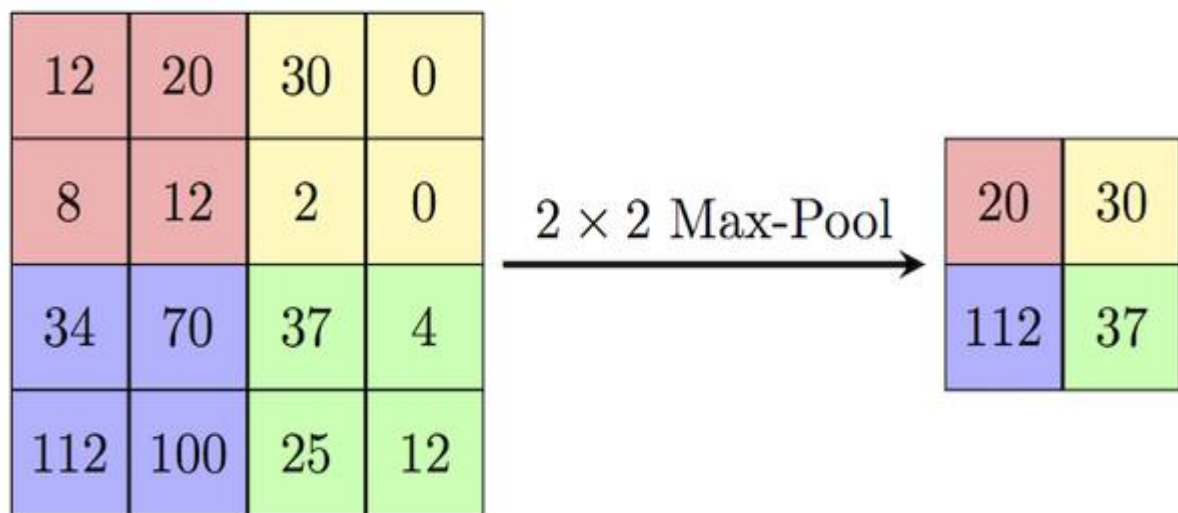


Figure 2.10 : Exemples de calcul du pooling sur une image 4x4 [33].

✚ Couche de correction (ReLU)

Il est souvent possible d'améliorer l'efficacité du traitement en insérant une couche qui applique une fonction mathématique, appelée fonction d'activation, sur les signaux de sortie entre les différentes couches de traitement. La couche de correction ReLu (Rectified Linear Unit) remplace toutes les valeurs négatives reçues en entrée par des zéros, tandis que les valeurs positives sont conservées sans modification. Cela permet d'introduire une non-linéarité dans le réseau neuronal et d'améliorer la capacité du modèle à apprendre des motifs complexes [17].

Contrairement à l'opération de convolution, l'opération d'agrégat n'utilise aucun poids et fait glisser une fenêtre sur son entrée et prend simplement une valeur de cette fenêtre en fonction du type d'agrégat.

C). Couche complètement connectée

La couche entièrement connectée est un type de perceptron multicouche traditionnel qui, en général, utilise la fonction d'activation softmax dans sa couche de sortie. La fonction softmax garantit que la somme des valeurs de sortie reste égale à un, tout en limitant également les valeurs de sortie individuelles dans la plage de 0 à 1, selon la formule mathématique suivante.

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{i=1}^K e^{z_i}}$$

Avec : z : vecteur d'entrée de k nombres réels. $Z = \{ z_1, z_2, \dots, z_k \}$.

(\mathbf{z}) : vecteur de sortie de k nombres réels strictement positifs et de somme 1.

L'expression « complètement connectée » signifie que chaque neurone dans la couche précédente est relié à tous les neurones de la couche suivante. Les sorties des couches de convolution et de pooling capturent les caractéristiques de haut niveau de l'image d'entrée. L'objectif de la couche entièrement connectée est d'utiliser ces caractéristiques pour classifier l'image d'entrée dans différentes classes, en se basant sur l'ensemble de données d'apprentissage [32].

2.3.4. Quelques travaux utilisant l'algorithme CNN

Voici un tableau avec des sources d'articles sur l'expression faciale à l'aide de CNN :

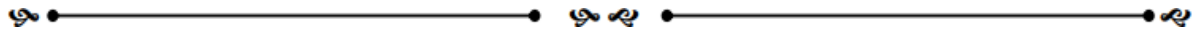
Tableau 2.1 : Tableau données quelques travaux qui utilisées le CNN.

Auteur(s)	Titre de l'article	Source	Année
Liu, L. et al. [34]	"Deep convolutional neural network for facial expression recognition"	Neural Networks	2016
Mollahosseini, A. et al. [35]	"Going deeper in facial expression recognition using	Proceedings of the IEEE Conference on Computer Vision	2016

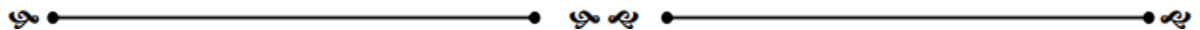
	deep neural networks"	and Pattern Recognition (CVPR)	
Khorrami, P. et al. [36]	"Deep convolutional neural networks for emotion recognition from facial expressions"	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	2015
Parkhi, O. M. et al. [37]	"Deep face recognition"	British Machine Vision Conference (BMVC)	2015

2.4. Conclusion

Dans ce chapitre, nous avons décrit brièvement le Deep learning et ses notions de base. Nous avons présenté particulièrement les concepts principaux du réseau de neurone artificiel (ANN) et les réseaux de neurones convolutifs (CNN). L'implémentation de ces réseaux sera décrite dans le chapitre 3. Plus particulièrement, nous allons décrire et implémenter un système REF basé sur l'algorithme CNN. Également, une étude comparative avec d'autres classificateurs tels que ANN et KNN, sera menée.



*Chapitre 3 : Implémentation du
système de reconnaissance
d'expressions faciales basé sur CNN.*



3.1. Introduction

La reconnaissance automatique d'expressions faciales (REF) est très utilisée dans plusieurs domaines d'applications tels que : Marketing, santé, description de photos, prédictions démographiques et électorales...etc. La conception d'un système REF s'effectue en une phase d'apprentissage permettant de modéliser les différentes classes d'expressions faciales et une phase de test permettant d'évaluer les performances du système en termes de précision et de complexité (temps de calcul et d'espace mémoire). Ainsi, une base d'apprentissage et une base de test sont nécessaires pour l'évaluation des systèmes REF. Généralement, dans les systèmes REF, ces deux phases exigent une étape d'extraction de paramètres (caractéristiques, attributs) permettant de convertir l'image faciale en autre espace de paramètres discriminants des différentes classes d'expressions tels que LBP, HOG, LPQ, ...etc. Ces paramètres peuvent être extraits en utilisant différentes approches à savoir l'approche globale, l'approche locale ou l'approche hybride. La première approche utilise l'image entière du visage, tandis que la deuxième méthode détecte les traits de certaines régions locales du visage (Yeux, bouche, nez, etc.), puis extrait les paramètres locaux de ces régions. La troisième approche combine les paramètres globaux et locaux. Dans notre travail, on utilise l'approche globale.

Dans la phase de test, la reconnaissance de la classe d'expression d'une image faciale d'entrée du système s'effectue en utilisant des algorithmes de classification tels que ANN, KNN, GMM, SVM,... etc. [4] [38] [39] Ces algorithmes appliqués sur les paramètres obtenus durant l'étape d'extraction de paramètres appartiennent à l'ensemble des algorithmes de 'machine Learning'. Dans notre travail, on s'intéresse à appliquer l'apprentissage profond (Deep Learning) pour la conception d'un système REF. Plus particulièrement, on applique l'algorithme CNN pour la classification des images faciales en 7 expressions faciales, à savoir : Joie, Peur, Dégoût, Surprise, Tristesse et Colère. Une étude comparative sera menée pour comparer les performances des systèmes REF basés sur l'algorithme CNN avec des algorithmes ML tels que KNN et ANN. A cet effet, on propose d'utiliser la base de données JAFFE (Japanese Female Facial Expression) [4] [3]. pour évaluer les performances de ses systèmes REF. Cette base de données a été largement utilisée comme base de référence pour la validation des travaux scientifiques [40] [4].

3. *Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.*

Plusieurs questions peuvent se poser durant la conception de notre système basée sur CNN telles que :

1. Quelle est la configuration optimale du système REF basé sur le classificateur CNN appliqué sur les images brutes ?
2. Quelle est la configuration optimale du système REF basé sur le classificateur CNN appliqué sur les images LBP ?
3. Les systèmes REF basés sur le classificateur CNN sont-ils plus performants que ceux basés sur les classificateurs ML?

Dans les sections suivantes, nous décrivons les différentes étapes qui composent notre système ainsi que les différentes expériences réalisées. Le système proposé est implémenté sous l'environnement MATLAB 2020.

3.2. Description de la base de données JAFFE

La base de données JAFFE contient 213 images de 10 femmes Japonaises, réparties en 7 classes d'expressions faciales, à savoir : 'Colère', 'dégoût', 'peur', 'joie', 'neutre', 'tristesse' et 'surprise', ainsi que la classe d'expression neutre. Dans cette base, 3 à 4 images exemplaires de la même expression faciale ont été prises pour chaque personne. Toutes les images sont enregistrées à une résolution de 256x256 pixels. Pour évaluer les performances du système REF à concevoir, cette base est divisée en une base d'apprentissage de 138 images (équivalent presque à 65% de la base de données), constituée des deux premières images exemplaires pour modéliser les différentes classes d'expressions faciales et une base de test de 75 images (équivalent à 35%), constituée des exemplaires restants pour évaluer le taux de reconnaissance. Cette répartition nous conduit à concevoir un système REF dépendant de la personne puisque chaque personne contribuant dans la phase d'apprentissage, contribue également dans la phase de test.

3. Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.

Tableau 3.1 : Répartition détaillée des données en base d'apprentissage et base de test [4].

Classes Bases	<i>Colère</i> Anger (AN)	<i>dégoût</i> Disgust (DI)	<i>peur</i> Fear (FE)	<i>Joie</i> Happiness (HA)	<i>neutre</i> Neutral (NE)	<i>tristesse</i> Sadness (SA)	<i>surprise</i> Surprise (SU)	<i>Total</i>
<i>Base d'apprentissage</i>	20	18	20	20	20	20	20	138
<i>Base de test</i>	10	11	12	11	10	11	10	75

3.3. Système REF basé sur KNN (Related Work)

La méthode des k plus proches voisins, également connue sous le nom k-NN ou KNN, est une technique d'apprentissage supervisé utilisée en intelligence artificielle, plus précisément en apprentissage automatique. Dans [4], l'auteur a appliqué cette technique de classification pour la tâche REF évaluée sur la base de données JAFFE. Cette technique exige premièrement une étape permettant de stocker l'ensemble des observations d'apprentissage des différentes classes d'expressions faciales dans une matrice. Chacune de ces observations stockée dans une ligne est un vecteur de caractéristiques représentant une image faciale de la base d'apprentissage. L'étape de classification consiste à chercher les k observations de la base d'apprentissage les plus proches de l'observation d'entrée en utilisant une mesure de similarité ou de distance. La classe reconnue correspond à la classe majoritaire de l'ensemble des k observations retenues. Dans [4], les meilleures performances sont obtenues en choisissant k égal à 1 avec la distance euclidienne appliquée sur des vecteurs de caractéristiques HOG. Les images utilisées sont des images de la base de données JAFFE de tailles différentes représentant que la partie du visage. Dans notre travail, on a fixé la taille des images à (256×256) pour comparer les performances du système REF basé sur KNN et le descripteur HOG extrait sur la région globale de l'image, avec celles du système CNN qui doit fonctionner sur des images brutes de mêmes tailles.

La figure 3.1 illustre les différentes étapes de fonctionnement du système REF basé sur le classificateur KNN combiné avec le descripteur HOG.

3. Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.

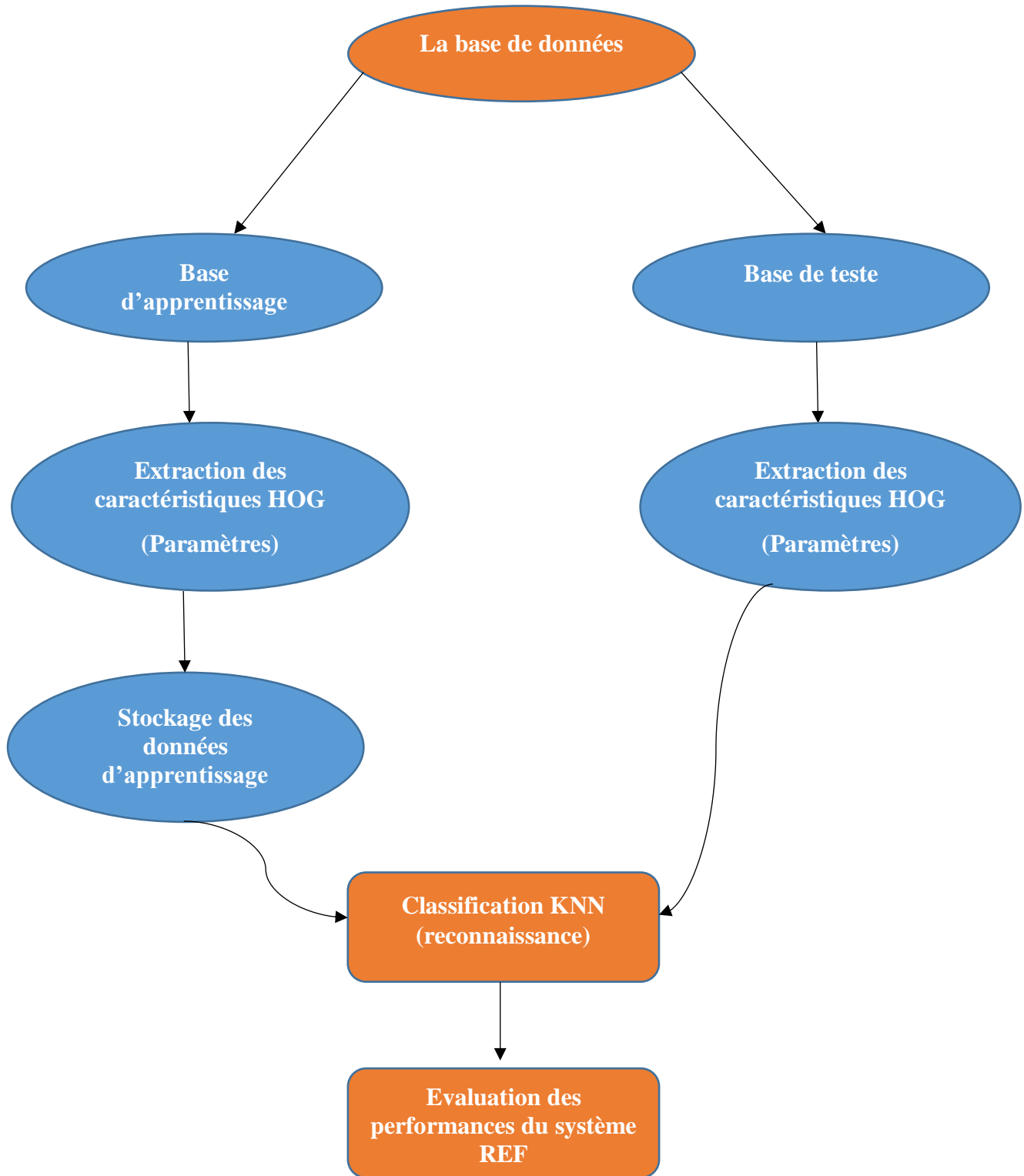


Figure 3.1 : Schéma de fonctionnement de classificateur KNN [4].

3.4. Système REF basé sur ANN

Le classificateur ANN (Artificial Neural Network, ou Réseau de Neurones Artificiels en français) est un type d'algorithme de classification utilisé dans des applications de reconnaissances de formes telles que la reconnaissance des expressions faciales pour détecter et classifier différentes expressions émotionnelles sur un visage [26]. Dans notre travail, on s'intéresse à appliquer cet algorithme sur la base de données JAFFE.

Les différentes étapes du système implémenté basé le classificateur ANN sont illustrées sur la figure 3.2.

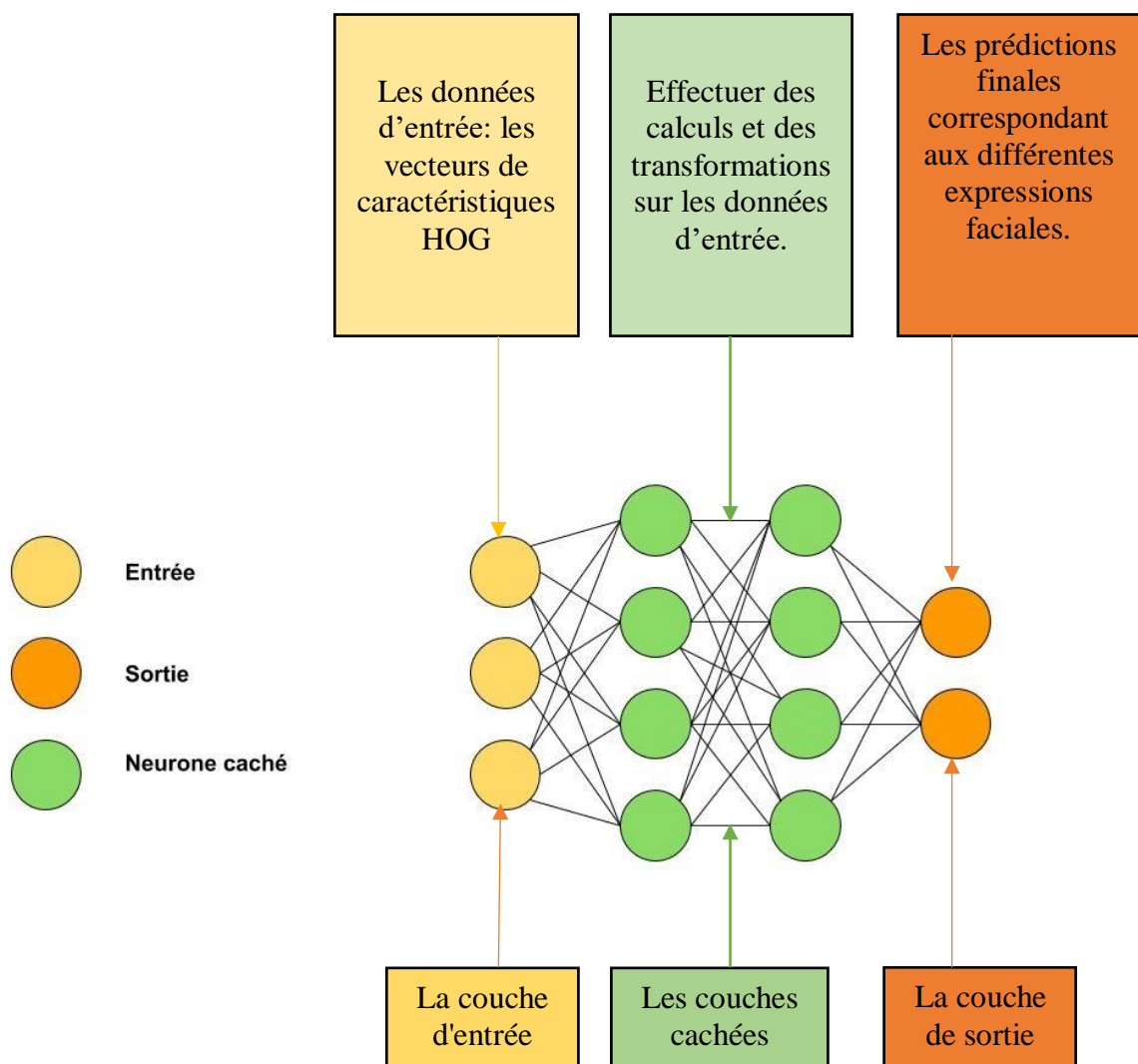


Figure 3.2 : Schéma de fonctionnement du classificateur ANN.

3.5. Système REF proposé basé sur CNN

Le classificateur CNN (Convolutional Neural Network) est un type d'algorithme utilisé dans la reconnaissance des expressions faciales pour détecter et classifier les différentes expressions émotionnelles sur un visage.

Les différentes étapes du système implémenté basé le classificateur CNN sont illustrées sur la figure 3.3.

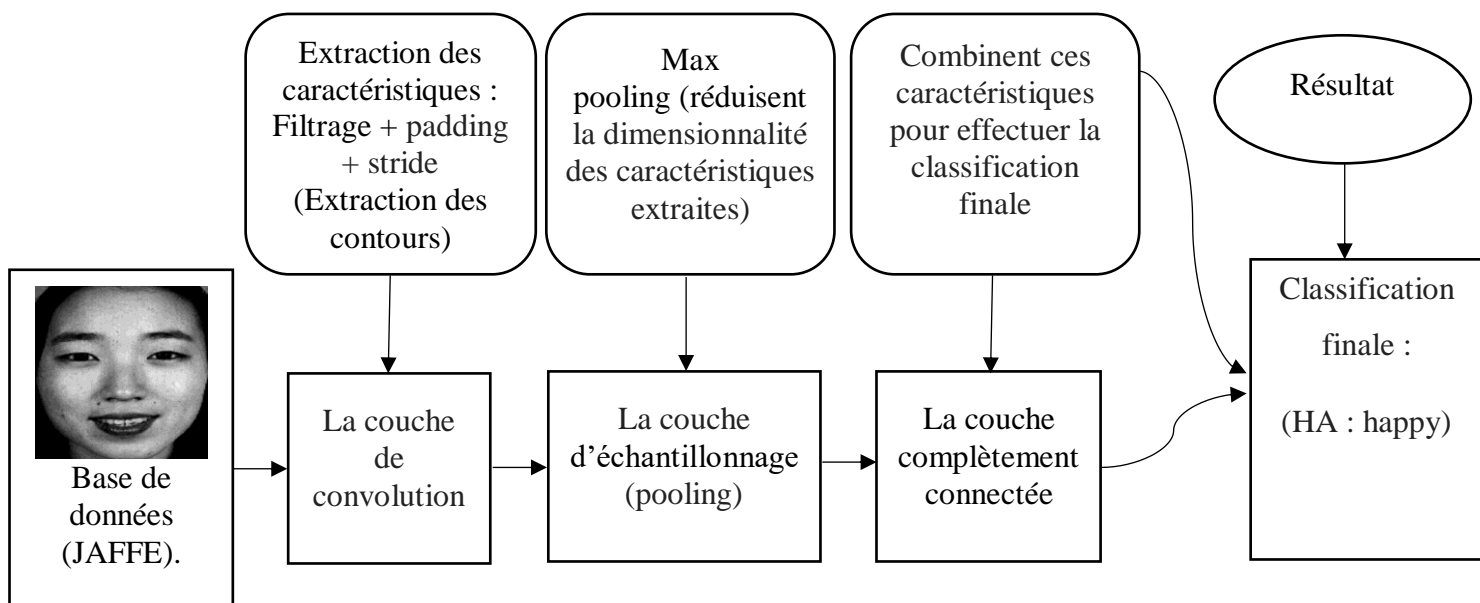


Figure 3.3 : Schéma de système REF basé sur l'algorithme CNN.

3.6. Expériences et résultats

Cette section a pour objectif de répondre à différentes questions posées dans l'Introduction en menant diverses expériences. Nous décrivons ainsi le but de chaque expérience ainsi que les résultats obtenus et leur interprétation.

3.6.1. Expérience 1 : Performances du système REF basé sur le classificateur KNN combiné avec le descripteur HOG

Cette expérience a pour objectif d'illustrer les résultats du système REF basé le classificateur KNN combiné avec l'extraction de paramètres HOG estimé sur la région globale et également

3. Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.

sur des régions locales de chaque image faciale. Chaque image est décomposée horizontalement en Cx cellules et verticalement Cy cellules pour extraire des caractéristiques HOG.

Le but de cette expérience est de trouver la taille optimale de décomposition en cellules (Cx, Cy) pour obtenir le meilleur compromis entre la précision et la complexité.

Le tableau 3.2 montre le taux de reconnaissance TR du système REF basé sur l'algorithme KNN, correspondant à la décomposition cellulaire (Cx, Cy)

Tableau 3.2 : Taux de reconnaissance correspond aux différentes tailles de décompositions de l'image en cellules (Cx, Cy) [4].

classificateur	Taille (Cx,Cy)	1.1	2.2	3.3	4.4	5.5	6.6	7.7	8.8	9.9	10.10
	Paramètres										
K-NN	TR	52	84	90.66	88	90.66	90.66	93.33	90.66	89.33	92
Nombre de caractéristiques		9	36	81	144	225	324	441	576	729	900

➤ Les résultats du tableau 3.5 nous permettent de remarquer les points suivants :

- Le découpage de la région globale du visage en plusieurs cellules améliore largement le taux de reconnaissance. Cependant, il augmente la complexité du point de vue de temps de calcul et de l'espace de mémoire. Le choix de la taille (7,7) a donné le meilleur taux de reconnaissance. Cependant, le choix de la taille de cellules (3.3) permet d'obtenir un bon compromis entre le TR et la complexité.

- L'extraction de caractéristiques HOG sur la région globale a donné la plus faible du taux TR.

On peut voir que l'extraction de descripteurs HOG basée sur la décomposition de toute la région du visage en cellules a donnée de bonnes performances, mais nécessite plus de temps de calcul et d'espace mémoire.

3. Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.

3.6.2. Expérience 2 : Performances du système REF basé sur le classificateur ANN combiné avec le descripteur HOG

Cette expérience illustre les résultats du système REF basé sur l'algorithme ANN. Dans cette expérience, nous avons repris la même décomposition de l'image faciale, utilisée précédemment. Le tableau 3.3 montre le taux de reconnaissance TR du système REF basé sur l'algorithme ANN, correspondant à la décomposition cellulaire (Cx, Cy).

Tableau 3.3 : Présente le taux de reconnaissance TR correspondant à différentes tailles de décomposition en cellules (Cx, Cy) du descripteur HOG extrait de la région du visage (V).

classificateur	Taille (Cx,Cy)	1.1	2.2	3.3	4.4	5.5	6.6	7.7	8.8	9.9	10.10
	Paramètres										
A-NN	TR	30.66	52	68	82.66	81.33	85.33	90.66	85.33	88	92
Nombre de caractéristiques		9	36	81	144	225	324	441	576	729	900

➤ Les résultats du tableau 3.3 mettent en évidence plusieurs points importants :

- Le choix de la taille (10,10) a donné le meilleur taux de reconnaissance. Cependant, le choix de la taille de cellules (7.7) permet d'obtenir un bon compromis entre le TR et la complexité.

3.6.3. Expérience 3 : Performances du système REF basé sur le classificateur CNN

Nous avons utilisé le classificateur CNN pour la reconnaissance des expressions faciales en utilisant entièrement l'image faciale brute comme entrée du système REF. Nous avons mis en œuvre notre modèle à l'aide de l'application MATLAB 2020 en suivant les étapes suivantes :

1. Chargement du jeu de données d'images :

Nous avons chargé la base de données d'apprentissage en utilisant la fonction "imageDatastore" de MATLAB. Cette base est répartie préalablement en sous dossiers incluant chacun des images faciales de mêmes classes.

3. *Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.*

2. **Division du jeu de données :**

Nous avons divisé le jeu de données de cette base en ensembles d'apprentissage, de validation en utilisant la fonction "splitEachLabel". Dans notre cas, nous avons utilisé une proportion de 70% pour l'apprentissage et 30% pour la validation, avec un mélange aléatoire des données.

3. **Définition des couches du CNN :**

Nous avons défini les couches du CNN en utilisant une architecture spécifique. Les couches comprennent des couches de convolution, de normalisation par lots, de ReLU, de max pooling et de couches entièrement connectées. La dernière couche est une couche de classification softmax pour prédire les classes des expressions faciales.

4. **Définition des options d'entraînement :**

Nous avons défini les options d'entraînement pour le CNN en utilisant la fonction "trainingOptions". Les options comprennent l'optimiseur Adam, le nombre maximal d'époques, le taux d'apprentissage initial, la taille des mini-lots, le mélange des données à chaque époque, les données de validation et la fréquence de validation.

5. **Entraînement du CNN :**

Nous avons entraîné le CNN en utilisant la fonction "trainNetwork" avec les données d'apprentissage, les couches du CNN et les options d'entraînement définies précédemment.

➤ Les résultats obtenus par notre classificateur CNN dans la reconnaissance des expressions faciales sont les suivants :

- Précision de validation : 85%

Cela signifie que notre modèle a réussi à classer correctement 85% des images dans l'ensemble de validation. Une précision de validation de 85% indique que notre modèle généralise bien sur de nouvelles données.

3. Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.

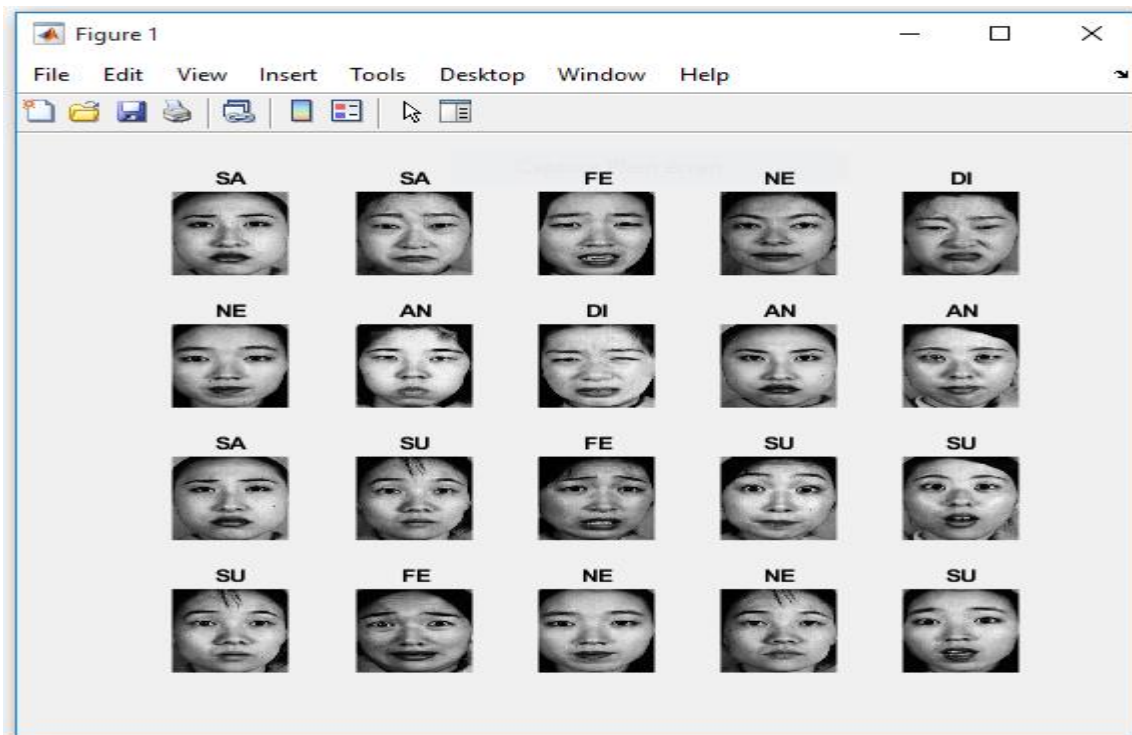
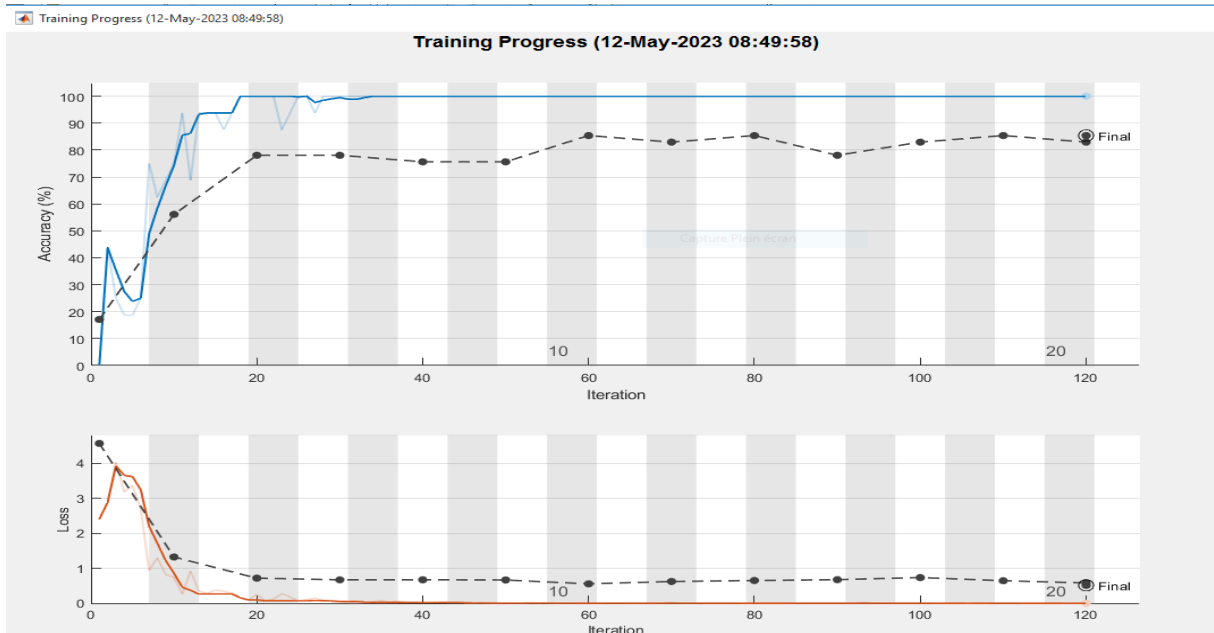


Figure 3.4 : Images faciales de la base de données JAFFE.

3. Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.



Results

Validation accuracy: 85.37%
 Training finished: Reached final iteration

Training Time

Start time: 12-May-2023 08:49:58
 Elapsed time: 15 min 0 sec

Training Cycle

Epoch: 20 of 20
 Iteration: 120 of 120
 Iterations per epoch: 6
 Maximum iterations: 120

Validation

Frequency: 10 iterations

Other Information

Hardware resource: Single CPU
 Learning rate schedule: Constant
 Learning rate: 0.001

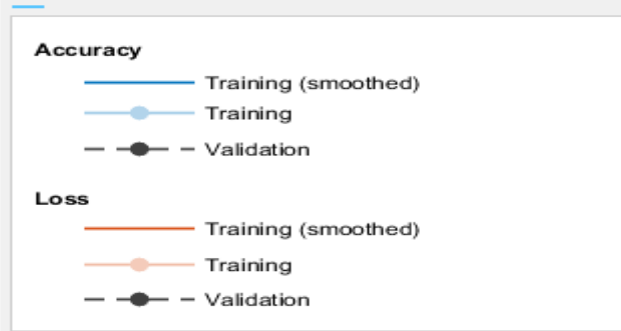


Figure 3.5 : Précision de validation : 85%.

3. Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.

- Précision de test : 89%

Lorsque nous avons évalué notre modèle sur l'ensemble de test, il a réussi à classer correctement 89% des images faciales. Cette précision de test supérieure à la précision de validation est un bon indicateur de la capacité de notre modèle à généraliser sur des données inconnues.

- Précision d'apprentissage : 95%

Lors de l'entraînement de notre modèle sur l'ensemble de données d'apprentissage, nous avons obtenu une précision d'apprentissage de 95%. Cela indique que notre modèle a pu apprendre efficacement les motifs et les caractéristiques des expressions faciales présentes dans l'ensemble d'apprentissage.

En conclusion, notre classificateur CNN basé sur une approche globale a donné des résultats prometteurs dans la reconnaissance des expressions faciales.

3.6.4. Expérience 4 : L'application de classificateur CNN sur les images LBP

Nous avons suivi les mêmes étapes de classificateur CNN sur les images bruit aussi avec les images LBP.

➤ Les résultats obtenus comme suit :

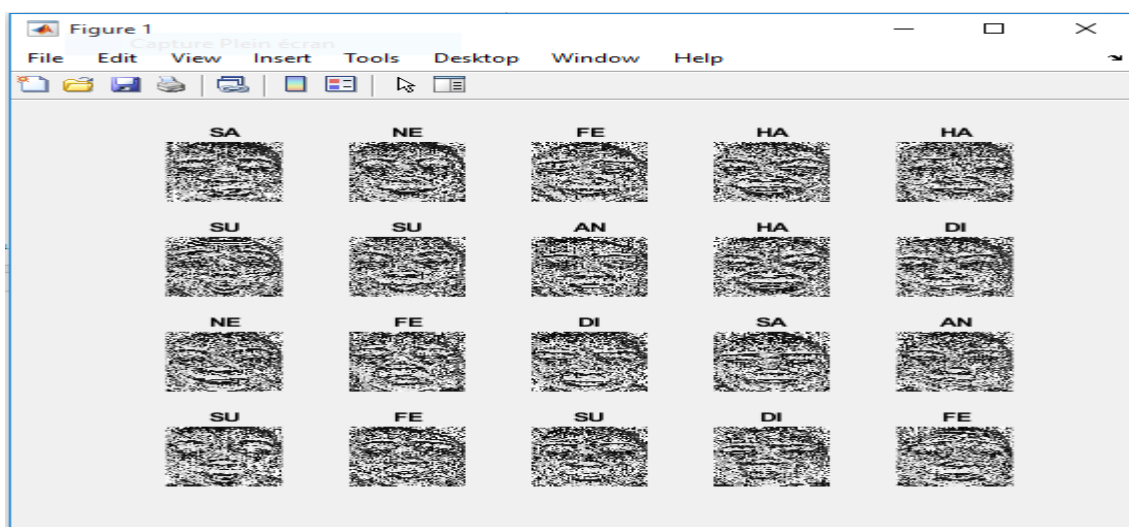


Figure 3.6 : Les pixels des images LBP.

3. Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.

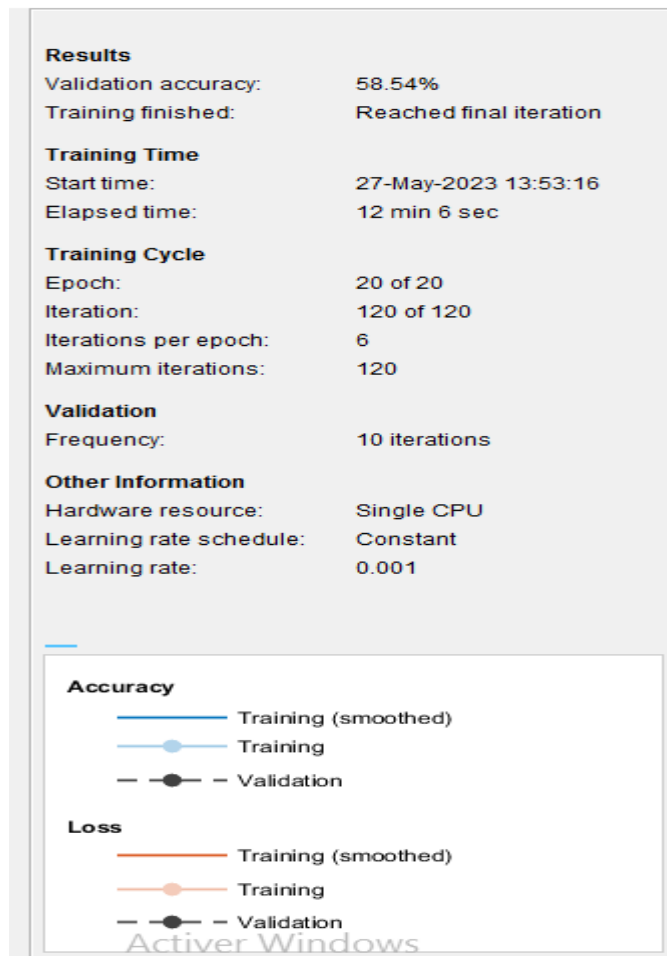
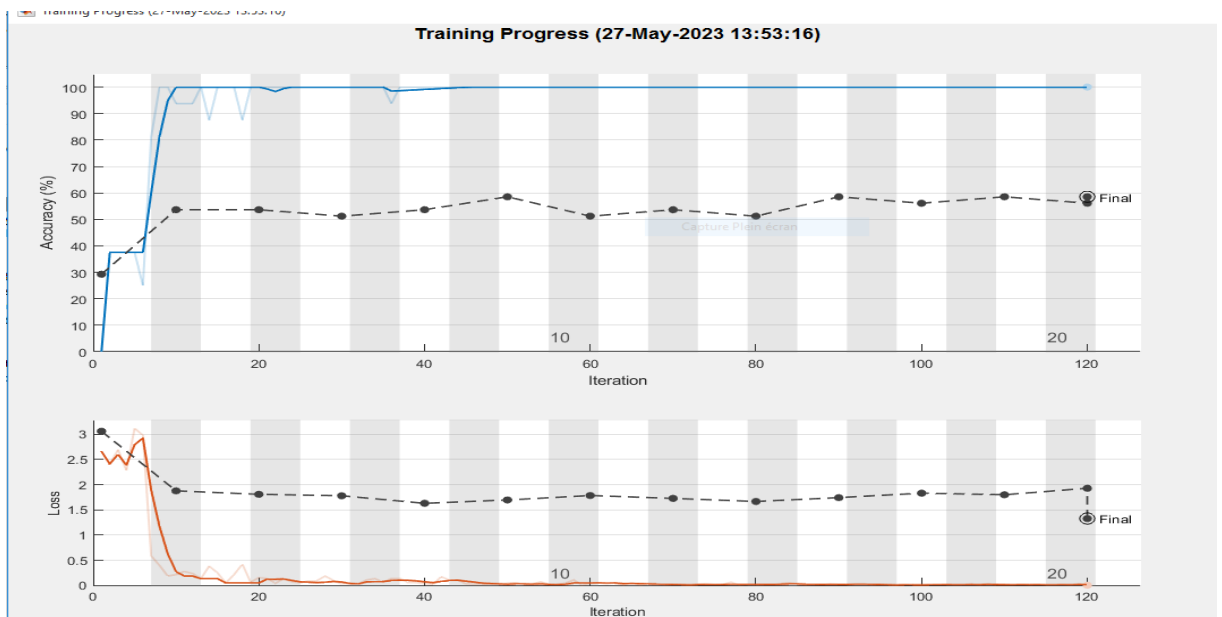


Figure 3.7 : Précision de validation : 58%.

3. Chapitre 3 : Implémentation du système de reconnaissance d'expressions faciales basé sur CNN.

- Précision de validation : 58%
- Précision de test : 73%
- Précision d'apprentissage : 87%

Ces résultats indiquent que le modèle a obtenu une précision relativement faible dans la reconnaissance des expressions faciales. Une précision de validation de 58% signifie que le modèle n'a réussi à classer correctement que 58% des images dans l'ensemble de validation. De même, une précision de test de 73% indique que le modèle a réussi à classer correctement 73% des images dans l'ensemble de test. Enfin, la précision d'apprentissage de 87% indique que le modèle a réussi à classer correctement 87% des images de l'ensemble d'apprentissage utilisé pour l'entraînement.

Comparé aux résultats précédents obtenus avec des images complètes, il est clair que l'utilisation des pixels d'images LBP n'a pas donné des performances aussi élevées.

Cela peut être s'expliquer par plusieurs facteurs :

✓ **Perte d'information** : Les pixels d'images LBP sont des représentations basées sur des motifs locaux et des opérateurs de texture. En utilisant uniquement ces caractéristiques locales, certaines informations globales et fines nuances des expressions faciales peuvent être perdues, ce qui affecte négativement la précision du modèle.

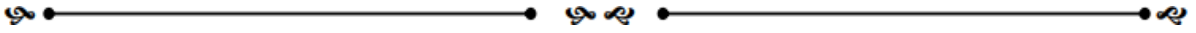
✓ **Spécificité de la méthode LBP** : LBP est une méthode qui se concentre principalement sur les motifs de texture et ne prend pas en compte les informations de couleur, de forme ou de structure globale du visage. Cela peut entraîner une perte d'informations discriminantes pour la reconnaissance des expressions faciales.

✓ **Adaptation du modèle** : Les modèles CNN sont généralement conçus pour traiter des images en niveaux de gris ou en couleur avec des dimensions spécifiques. L'utilisation des pixels d'images LBP peut nécessiter des modifications supplémentaires de l'architecture du modèle ou des techniques de prétraitement spécifiques pour s'adapter à ces caractéristiques.

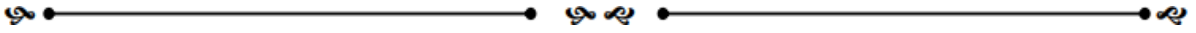
En conclusion, les résultats obtenus en utilisant des pixels d'images LBP montrent une précision relativement faible dans la reconnaissance des expressions faciales.

3.7. Conclusion

Dans ce chapitre, nous avons présenté le système de reconnaissance des expressions faciales proposé, incluant la colère, le dégoût, la peur, la joie, la neutralité, la tristesse et la surprise. Nous avons décrit en détail l'architecture générale de ce système ainsi que les différentes étapes de son fonctionnement et son implémentation. Nous avons abordé plusieurs questions relatives à sa conception et les différentes expériences menées afin d'y répondre. Les résultats obtenus ont démontré que le système REF basé sur le classificateur CNN, a donné des bonnes performances témoignant du succès de ce classificateur dans la reconnaissance des expressions faciales. Avec une précision de validation de 85%, une précision de test de 89% et une précision d'apprentissage de 95%, notre modèle démontre sa capacité à reconnaître avec précision les différentes expressions faciales. Ces résultats ouvrent des perspectives prometteuses pour son application dans divers domaines tels que la détection des émotions, l'interaction homme-machine et bien d'autres, ce qui est essentiel pour sa fiabilité dans des applications du monde réel.



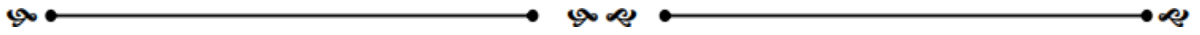
Conclusion générale



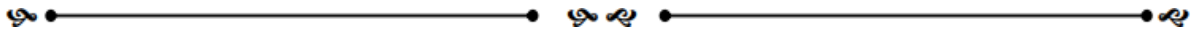
Conclusion générale

Notre travail consiste à implémenter un système de reconnaissance d'expressions faciales basé sur le classificateur C-NN. Le fonctionnement de reconnaissance se déroule en trois étapes à savoir : la détection du visage, l'extraction des caractéristiques et la classification, Ainsi, nous avons donné dans le chapitre 1 des généralités sur les systèmes de reconnaissance d'expressions faciales et les différents types de descripteurs et classificateurs. Dans le chapitre 2, nous avons donné des notions de bases sur l'apprentissage profond et plus particulièrement sur l'algorithme CNN. Notre objectif consiste à concevoir et implémenter un système REF basé sur l'algorithme CNN appliqué sur des images brutes et des images LBP. Le système implémenter sous l'environnement Matlab est évalué en utilisant la base de données JAFFE incluant sept classes d'expressions faciales (colère, dégoût, peur, joie, tristesse et surprise, plus le neutre). Les performances de ce systèmes sont comparées avec celles des systèmes basés sur des algorithmes d'apprentissage automatique tels que : KNN, ANN, combinés avec la méthode d'extraction de caractéristiques de type HOG. Ces derniers systèmes exigent une étape d'extraction des caractéristiques, alors le système proposé applique directement l'algorithme CNN sur les images faciales brutes sans appliquer des méthodes d'extraction standards (HOG, LBP, etc). Plus particulièrement, cet algorithme s'applique sur la région globale de l'image faciale. Les résultats nous ont montré que les systèmes REF basés sur l'extraction des caractéristiques HOG utilisant des régions locales combinée avec les classificateur KNN et ANN donnent des performances comparables avec celles du système proposé. Plus particulièrement le système proposé présente de meilleures performances avec un taux de classification de 89%, par rapport aux systèmes basés sur les classificateurs précédents combinés avec l'extraction de caractéristiques sur des régions globales.

Enfin, comme perspectives à ce travail, nous proposons d'appliquer des matrices de caractéristiques telles que HOG comme des entrées à l'algorithme d'apprentissage profond CNN.



Bibliographies



Bibliographie

- [1] R. P. L. M.-F. Gosselin Pierre, Le développement de la reconnaissance des expressions faciales émotionnelles du répertoire humain, In: *Enfance*, n°4, 1995. pp. 379-396.
- [2] D. D. F. L. M. A. W. Bhiksha Raj, *Deep Learning Applications, Volume 3*, Singapour: Springer Nature Singapore, 2021.
- [3] B. AYMEN, automatisation de la sélection des paramètres optimisés de filtre de Gabor pour la détection des expressions faciales, mémoire master, Algérie : Université de 8 Mai 1945 -Guelma, 2020.
- [4] B. R. Hamadouche Linda, Reconnaissance des expressions faciales basée sur le descripteur HOG, mémoire master, Algérie: Université de Mohamed El-Bachir El-Ibrahimi-Bordj Bou Arreridj, 2018-2019.
- [5] R. A. S. M. A. M. M. a. S. D. P. Lisa Feldman Barrett, "Emotional Expressions [Reconsidered]: Challenges to Inferring Emotion From Human Facial Movements", *Psychological Science in the Public Interest*, July 2019, Vol.
- [6] K. LEKDIOUI, 'Reconnaissance d'états émotionnels par analyse visuelle du visage et apprentissage machine'. Thèse doctorat, Kenitra-Maroc : BOURGOGNE FRANCHE-COMTE, 2018.
- [7] J.-J. J. Lien, «Automatic Recognition of Facial Expressions Using Hidden Markov Models and Estimation of Expression Intensity». CMU-RI-TR-98-31 p13.
- [8] O. H. BELGASMI, OURECONNAISSANCE AUTOMATIQUE DES EXPRESSIONS FACIALES PAR SUPPORT VECTOR MACHINE, mémoire master, Algérie : Université Larbi Ben M'hidi-Oum El Bouaghi, 2016-2017.
- [9] J. C. e. A. Tcherkassof, *Les émotions : cognition, langage et développement*, Pierre Mardaga, 2003.
- [10] s. m. Saad fawzi, reconnaissance des expressions faciales en temps réel, mémoire master, algérie: université saad dahlab-blida, 2019-2020.

- [11] <http://thesis.univ-biskra.dz/944/4/Chap%201%20Syst%C3%A8me%20RV%20sept%2012.pdf>. [2023].
- [12] S. KAOUTHAR, Reconnaissance Facial dans une séquence d'images, mémoire master, Algérie, Université Mohamed Khider-BISKRA, 2021-2022.
- [13] H. O. e. M. Benatia, Identification De Reconnaissance Faciale Avec Des Expressions, mémoire master, Algérie, Université Mohamed Khider-Biskra, 2012.
- [14] C. HADJER, Montage d'un Système de Reconnaissance des Expressions Faciales avec le Deep Learning, mémoire master, Algérie, Université Mohamed Khider-BISKRA, 2020.
- [15] M. T. a. A. Pentland, Eigenfaces for recognition, *Journal of cognitive neuroscience*, 3(1) :71–86, 1991.
- [16] N. e. B.Triggs, Histograms of Oriented Gradients for Human Detection, INRIA RhoneAlps, *Computer Vision and Pattern Recognitio. version 1* (20-12-2010).
- [17] A. I. A. AMIRA, Développement d'un système de reconnaissance faciale, Mémoire de Master, Algérie, Université saad dahleb-blida, 2020.
- [18] <https://www.bial-r.com/2019/05/22/comprendre-le-machine-learning-et-le-deep-learning/>.
- [19] P. K. C. a. S. J. Stolfo, Toward scalable learning with non-uniform class and cost distributions : A case study in credit card fraud detection, in *KDD*, vol. 1998, pp. 164–168.
- [20] A. K. a. V. P. G. Bradski, Learning-based computer vision with intel's open source computer vision library, *Intel Technology Journal*, vol. 9, 2005.
- [21] I. Kononenko, Machine learning for medical diagnosis : history, state of the art and perspective, *Artificial Intelligence in medicine*, vol. 23, no. 1. 2001, pp. 89–109.
- [22] W. H. G. E. J. H. J. T. G. B. A. B. S. T. D. G. W. F. T. J. G. a. O. H.-H. X. D. Huang, Web-based targeted advertising in a brick-and-mortar retail establishment using online customer information, *US Patent App. 11/427/764.*, 2006.
- [23] M. D. Youcef, Deep Learning pour la classification des images, Algérienne : Université Abou Bakr Belkaid-Tlemcen, 2016-2017.

- [24] K. N. Azzoune Ikram, Une approche IA pour la reconnaissance des expressions faciale, mémoire master, Algérie, Université Akli Mohand Oulhadj-Bouira, 2019-2020.
- [25] H. M. W. M. a. W. P. J. Lettvin, What the Frog's Eye Tells the Frog's Brain. Proceedings of the IRE, 47(11) :1940–1951, Nov. 1959.
- [26] K. DAHMANE, Analyse d'images par méthode de Deep Learning appliquée au contexte routier en conditions météorologiques dégradées. Thèse doctorat, France Université Clermont Auvergne, 2020.
- [27] Investigating Keystroke Dynamics as a Two-Factor Biometric Security - Scientific Figure on ResearchGate. Available from : https://www.researchgate.net/figure/Biological-Neuron-versus-Artificial-Neural-Network_fig4_325870973 [accessed 17 Jun, 2023].
- [28] S.kom, Deep Learning Bible - 2. Classification, Part F. Convolutional Neural Networks, Available from : <https://wikidocs.net/164365>. Korean, Last Edit: Jun 14, 2023 4:22 PM.
- [29] N.Belaidi, Réseaux convolutifs (CNN) : comment ça marche ?, Available from : <https://blent.ai/blog/a/cnn-comment-ca-marche>. [21 juin 2022].
- [30] J. Murphy, An Overview of Convolutional Neural Network Architectures for Deep Learning, micro way, 2016.
- [31] I. C. e. S. Madene, « Détection visuelle d'objets statiques et dynamiques dans un environnement de type route et classification en exploitant l'apprentissage profond », mémoire master, Algérie, Université des Sciences et de la Technologie-Houari Boumediene, 2019.
- [32] M. M. NOUIOUA Naouel, Reconnaissance Automatique des Expressions Faciales, mémoire master, Algérie, Université Abderrahmane Mira-Bejaia, 2019-2020.
- [33] A.Prakash, Available from : <https://www.quora.com/What-is-pooling-in-a-convolutional-neural-network>. [2023].
- [34] L. e. a. Liu, "Deep convolutional neural network for facial expression recognition", Neural Networks, 2016.

- [35] A. e. a. Mollahosseini, "Going deeper in facial expression recognition using deep neural networks", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [36] P. e. a. Khorrami, "Deep convolutional neural networks for emotion recognition from facial expressions", IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , 2015.
- [37] O. M. e. a. Parkhi, "Deep face recognition", British Machine Vision Conference (BMVC), 2015.
- [38] M. C. a. eddine, CLASSIFICATION DES MASSES MAMMAIRE UTILISANT LOCAL TERNARY PATTERN, mémoire master, Algérie, UNIVERSITÉ BADJI MOKHTAR-ANNABA, 2020-2021.
- [39] C. I. Peter Corcoran, Hidden Markov Models in Automatic Face Recognition, 27 juillet 2011.
- [40] C. R. e. R. KHELIF, Reconnaissance automatique des expressions faciales par le modèle de Markov caché (MMC), mémoire master, Algérie, Université de LARBI BEN MHIDI-OUM EL BOUAGHI, 2016-2017.
- [41] M. A. Zhanatkyzy, reconnaissance des expressions faciales basée sur l'apprentissage, Mémoire master, Kazakhstan: Université technique nationale de recherche kazakhe nommée d'après KI SATPAEV, 2020.
- [42] G. Sonia, 'Reconnaissance des émotions par traitement d'images'. Thèse doctorat, Université d'Orléans, 2016.