REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE YAHIA FARES DE MEDEA

Faculté de **T**echnologie

Département du **G**énie **E**lectrique

Mémoire de fin d'études de Master

Filière : télécommunications

Spécialité : systèmes de télécommunications

Classification des images veines de la main par un réseau de neurone convolutif

Présentés par :

Zerkout Mohamed Lotfi

Moussaoui Mohamed

Proposés et Dirigés par :

MR: Boucherit ismail

Promotion 2022

Remerciement

Tout d'abord, nous voulons remercier Dieu de nous avoir aidés et de nous avoir donné patience et patience le courage de terminer cette lettre.

Bien sûr, nous commençons par exprimer notre plus profond respect et notre gratitude

Nous remercions notre parrain Mr. ***Boucherit*** *pour son encadrement et ses conseils faites le travail patiemment.*

Nous tenons à remercier les membres du jury de nous avoir accordé le privilège de juger ce travail.

Nous tenons également à remercier tous les enseignants du Département de génie électrique de l'Université de Médée

Nous remercions nos familles, en particulier nos parents, pour leur soutien inconditionnel, et nous n'oublions pas nos frères et amis proches qui étaient toujours là et ne se sont jamais arrêtés encourager.

Dédicaces

Je dédie ce travail

A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études.

A toute ma famille ***zerkout*** *pour leur soutien tout au long de mon parcours universitaire.*

A mon binôme ***Mohamed*** *et toute la famille* ***MOUSSAOUI****. Et à tous ceux qui ont contribué de près ou de loin pour que ce projet soit possible, je vous dis merci.*

À tous les ***professeurs*** *et* ***enseignants*** *qui m'ont suivi durant tout mon cursus scolaire et qui m'ont permis de réussir dans mes études. A tous ceux qui m'ont aidé de près ou de loin.*

MR *ZERKOUT Mohamed Lotfi*

Dédicaces

Je dédie ce travail

Je dédie ce travail à mes parents qui ont fait l'amour du savoir

véritable conviction pour leurs enfants.

A ma sœur et mon frère qui m'ont constamment encouragé et cru en moi.

Je tiens tout particulièrement à remercier Monsieur Boucherit pour ses

formidables qualités d'homme

La science suscite respect et admiration.

Je dédie également ce travail à mes défunts grands-parents (que Dieu leur fasse

miséricorde ciel).

***Mr** MOUSSAOUI Mohamed*

ملخص

في هذا العمل استعملنا طريقة تصنيف وريد اليد على أساس CNN. الخطوات الأربع الرئيسية من التعرف على وريد النخيل هي : الحصول على الصور والمعالجة المسبقة للصور واستخراج الميزات والمطابقة. لتقليل خطوات المعالجة في تصنيف صور وريد اليد استخدمنا طريقة CNN المقترحة. قد ثبت أن شبكات CNN تتمتع بمزايا رائعة وأدت أداءً جيدًا في تصنيف الصورو لتقليل معالجة الصور في المراحل المبكرة ، يتم استخدام CNN لتصنيف صور وريد اليد.

الكلمات المفتاحية :الشبكة العصبية التلافيفية ، وريد اليد

Résumé

Dans ce travail, nous avons utilisé une méthode de classification des veines de la main basée sur CNN. Les quatre étapes principales de la reconnaissance de la veine palmaire sont : l'acquisition d'images, le prétraitement d'images, l'extraction de caractéristiques et la mise en correspondance. Pour réduire les étapes de traitement dans la classification des images des veines de la main, nous avons utilisé la méthode CNN proposée.

Il a été démontré que les CNN présentent de grands avantages et fonctionnent bien dans la classification des images.Pour réduire le traitement des images dans les premiers stades, le CNN est utilisé pour classer les images des veines de la main.

Mots clés : CNN, la veine de la main

Abstract

In this work, we used a hand vein classification method based on CNN. The four main steps in palmar vein recognition are: image acquisition, image preprocessing, feature extraction, and matching. To reduce the processing steps in the classification of hand vein images, we used the proposed CNN method.

CNN have been shown to have great advantages and perform well in image classification. To reduce image processing in the early stages, CNN is used to classify images of hand veins.

Keywords : CNN, hand vein

Liste des figures

Figure I-1: Représentation des notions image et pixel ---------- 5
Figure I- 2: Pixels et niveaux de gris ---------- 6
Figure I-3: Voisinage à 4, Voisinage à 8 ---------- 9
Figure I-4: Méthode supervisée ---------- 11
Figure I-5: Courbe ROC ---------- 14

Figure II-1: schéma de l'apprentissage profond 19
Figure II-2: Les différents algorithmes supervisés de DL 20
Figure II- 3: Architecture de perceptron 22
Figure II-4: Architecture de perceptron multicouche 23
Figure II-5: Structure générale d'un réseau de neurones convolutifs (CNN) 24
Figure II-6: Couche de convolution(CONV) 25
Figure II-7: Exemple d'une opération de Maxpooling de taille 2 × 2 26
Figure II-8: les fonctions d'activation 27
Figure II- 9: Forme de la fonction ReLu 29
Figure II- 10: exemple application de RELU 29
Figure II-11: Exemples de modèles de CNN 31
Figure II-12: Architecture Alexnet 32
Figure II-13: les couches de Vgg-16, Vgg-19 33

Figure III-1: Images de veine de la main -10-classes 39
Figure III-2:Architecteur de modèle 1 (4 couches) 40
Figure III-3: Architecture de modèle 2 (6 couches) 42
Figure III-4: Précision et erreur pour le modèle 1 (10 époques) 44
Figure III-5: Nombre total des images mal et bien classées et taux d'erreur et de précision . 44
Figure III-6: Matrice de confusion pour le modèle1 (10 époques) 45
Figure III-7: Précision et erreur pour le modèle 1 (15 époques) 45
Figure III-8: Nombre total des images mal et bien classées et taux d'erreur et de précision . 46
Figure III-9: Matrice de confusion pour le modèle1 (15 époques) 47
Figure III-10: Précision et erreur pour le modèle 1 (20 époques) 47
Figure III-11: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 1 (20 époques) 48

Figure III-12: Matrice de confusion pour le modèle1 (20 époques) 48
Figure III-13: Précision et erreur pour le modèle 1 (30 époques)..... 49
Figure III-14: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 1 (30 époques)..... 49
Figure III-15: Matrice de confusion pour le modèle1 (30 époques) 50
Figure III-16: Précision et erreur pour le modèle 2 (10 époques)..... 50
Figure III-17: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 2 (10 époques)..... 51
Figure III-18: Matrice de confusion pour le modèle 2 (10 époques) 52
Figure III-19: Précision et erreur pour le modèle 2 (15 époques)..... 52
Figure III-20: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 2 (15 époques)..... 53
Figure III-21: Matrice de confusion pour le modèle 2 (15 époques) 53
Figure III-22: Précision et erreur pour le modèle 2 (20 époques)..... 54
Figure III-23: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 2 (20 époques)..... 54
Figure III-24: Matrice de confusion pour le modèle 2 (20 époques) 55
Figure III-25: Précision et erreur pour le modèle 2 (30 époques)..... 55
Figure III-26: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 2 (30 époques)..... 56
Figure III-27: Matrice de confusion pour le modèle 2 (30 époques) 57

Liste des tables

Tableau I-1: Les différents niveaux de gris---------- 8
Tableau I-2: Matrice de confusion dans le cas binaire ---------- 12

Tableau III- 1: Configuration de modèle 1 (4 couches) 41
Tableau III- 2: Configuration de modèle 2 (6 couches) 43
Tableau III- 3: Tableau de comparaison des résultats..... 57

Liste des abréviations

RVB **R**ed, **G**reen, **B**lue

PPP **P**oint **P**ar **P**ouce

DPI **D**ots **P**er **I**nch

SVM **S**upport **V**ector **M**achine

ACP **P**rincipal **C**omponent **A**nalysis

ROC **R**eceived **O**perating **C**haracteristic

CNN **C**onvolutional **N**eural **N**etwork

DNN **D**eep **N**eural **N**etwork

DL **D**eep **L**earning

RNN **R**ecurrent **N**eural **N**etwork

DBN **D**eep **B**elief **N**etworks

RBM **R**estricted **B**oltzmann **M**achines

DCT **D**iscrete **C**osine **T**ransform

DWT **D**iscrete **W**avelet **T**ransform

FC **F**ully-**C**onnected

NIN **N**etwork-**I**n-**N**etwork

VGG **V**isual **G**eometry **G**roup

API **A**pplication **P**rogramming **I**nterface

CUDA **C**ompute **U**nified **D**evice **A**rchitecture

GPU **C**entral **P**rocessing **U**nit

Sommaire

Introduction générale……..……..1

Chapitre I : classification des images

I.1 Introduction.....3

I.2 Les motivations de la Classification des images 3

I.3 Notion de base sur les images.....5

I.3.1 Définition de l'image 5

I.3.2 Les différents types de format d'image.....6

I.3.2.1 Image couleur RVB…….6

I.3.2.2 Image binaire……...6

I.3.2.3 Images à niveaux des gris……6

I.4 Caractéristiques de l'image.....7

I.4.1 Pixel 7

I.4.2 Dimension 7

I.4.3 Résolution.....7

I.4.4 Niveau de gris.....7

I.4.5 Bruit 8

I.4.6 Luminance 8

I.4.7 Voisinage.....8

I.4.8 Contrasts.....9

I.4.9 Contour et texture 10

I.5 Méthodes de classification.....10

I.5.1 Méthodes supervisées 10

I.5.2 Méthodes non supervisées.....11

I.6 Indicateurs de performance en classification 11

I.6.1 Matrice de confusion.....11

I.6.2 Courbe ROC (Received Operating Characteristic).....13

I.7 Classification des images et l'apprentissage machine 14

Sommaire

I.8 Classification des images et les réseaux de neurones 16

I.9 Conclusion 17

Chapitre II : Les réseau de neurones convolutionnels

II.1 Introduction 18

II.2 L'apprentissage en profondeur (Deep learning) 19

II.2.1 Definition 19

II.3 Les différents algorithmes de deep learning 20

II.3.1 DBN 20

II.3.2 RNN 20

II.3.3 Domains application de l'apprentissage profonde 21

II.4 Quelques Modèles de réseaux de neurones 21

II.4.1 Perceptron 21

II.4.2 Perceptron multicouche 22

II.5 Réseaux de neurone convolutifs 23

II.5.1 Definition 23

II.5.2 Architecture de réseaux CNN 23

II.5.3 Couche de convolution (CONV) 24

II.5.4 Couche de pooling (Pool) 25

II.5.5 Fonction activation 26

II.5.6 Couch de correction (ReLu) 28

II.5.7 Couch entièrement connecter (FC) 30

II.5.8 Couch de perte (LOSS) 30

II.6 Exemples de modèles de CNN 30

II.6.1 Réseau AlexNet 31

II.6.2 Réseau NiN 32

II.6.3 Réseau VGG 32

II.6.4 Réseau GoogLeNet 33

II.6.5 Réseau ResNet 33

Sommaire

II.7 Choix des paramètres 34
II.7.1 Nombre de filtres 34
II.7.2 Forme du filtres 34
II.7.3 Form du Max Pooling 34

II.8 Conclusion 35

Chapitre III : Interprétation des résultats

III.1 Introduction 36

III.2 Logiciels et librairies utilisés dans interprétation des résultats et configuration de hardware 36
III.2.1 TensorFlow 36
III.2.2 Keras 36
III.2.3 CUDA 37
III.2.4 Scikit-learn 37
III.2.5 Python 37
III.2.6 Configuration de hardware 38

III.3 La base de donnée utilisée 38

III.4 Architecture des réseaux 39
III.4.1 Réseau neuronal convolutif à 4 couches 39
III.4.2 Réseau neuronal convolutif à 6 couches 41

III.5 Résultats obtenus et discussion 43

III.6 Conclusion 58

Conclusion générale……59

Introduction Générale

Les systèmes biométriques développent des technologies , peuvent être utilisées dans des systèmes automatisés pour identifier de manière unique et efficace les individus et devenir une bonne alternative aux méthodes traditionnelles telles que les mots de passe. Les caractéristiques biométriques peuvent être divisées en deux types. Le premier type contient des caractéristiques physiques (empreinte, visage, iris, rétine, empreinte palmaire, visage, veine, etc.) et le second type contient des caractéristiques comportementales (par exemple, voix, signature, caractéristiques du mouvement (démarche, mouvement des mains ou des lèvres, appui sur une touche) , etc.)). Selon la classification donnée en premier type, les empreintes palmaires, les empreintes digitales, la transpiration des mains et le conduit auditif sont quatre méthodes biométriques qui présentent toutes les caractéristiques suivantes :

- **Universalité :** signifie que toutes les personnes devraient avoir cette qualité.
- **Unicité :** signifie que la caractéristique doit être individuelle pour toutes les personnes.
- **Durabilité :** la caractéristique ne change pas avec l'âge.

Parmi toutes les technologies biométriques, l'identification des veines de la paume est une technologie fiable pour fournir un niveau de sécurité plus élevé. De toutes les différentes caractéristiques biométriques de la main humaine qui peuvent être utilisées pour identifier une personne, comme une empreinte digitale, les veines de la paume sont peut-être les plus réussies de toutes les formes et modèles considérés comme stables et fiables, ce qui signifie qu'une fois qu'une personne atteint l'âge adulte, la structure et la composition des mains et des veines restent relativement stables tout au long de la vie d'une personne.**[1]**

La reconnaissance des empreintes digitales peut être considérée comme peu pratique en raison du besoin de connectivité et entraîner une résistance des utilisateurs. De plus, dans des conditions environnementales telles que la saleté et les blessures, la reconnaissance des empreintes digitales peut être insuffisante et les performances peuvent diminuer. En outre, l'écriture manuscrite ou les signatures peuvent être falsifiées facilement.

Pour ces raisons, la technologie de reconnaissance des veines palmaires est aujourd'hui une biométrie prometteuse et fiable.

Dans ce contexte, le processus de reconnaissance de la veine palmaire est compliqué par le prétraitement des images, l'extraction des caractéristiques et les méthodes d'appariement

Il passe par trois étapes fondamentales **:**

Tout d'abord, l'image de la veine de la paume est prétraitée pour rendre les contours de l'image plus clairs.

Deuxièmement, il faut trouver un moyen d'extraire les caractéristiques de l'image. Enfin, la correspondance des caractéristiques doit être effectuée sur des échelles de distance pour identifier l'utilisateur. Pour réduire le processus de sélection manuelle

C'est pourquoi nous avons proposé une méthode d'identification de la veine palmaire basée sur CNN qui utilise les caractéristiques de la veine palmaire pour identifier et classer les utilisateurs.

Les fonctionnalités cibles peuvent être sélectionnées automatiquement en formant le modèle. Le deep learning excelle dans les problématiques liées à la reconnaissance visuelle, à la reconnaissance vocale, au langage naturel et aux séries temporelles...etc.

Dans ce mémoire , nous allons décrit au chapitre I les concepts de base de l'image en général et les méthodes de classification des images pour leur rôle principal dans notre projet et nous avons consacré le chapitre II à expliquer les concepts de base sur les réseaux neuronaux en général et CNN en particulier et le principe de fonctionnement de ce réseau et nous avons consacré le chapitre III à l'interprétation des résultats obtenus.

Chapitre I :

Classification des images

I.1 Introduction

La classification des images consiste à répartir les images selon des catégories prédéterminées, la classification des images les rend compatibles avec une catégorie, en distinguant leur relation avec d'autres images. Le développement technologique en termes d'acquisition et de traitement d'images (caméras, capteurs, microscopes) et de stockage génère des bases de données informatives et multiplie les domaines d'application, et il devient difficile pour les humains d'analyser le grand nombre d'images. Cependant, ce n'est pas nécessairement facile pour un programme informatique dans lequel l'image est un ensemble de valeurs numériques. Le but de la classification des images est de développer un système qui peut attribuer automatiquement une catégorie à l'image.

Ainsi, ce système permet d'effectuer de nombreuses tâches de qualité et de précision en très peu de temps. Les applications de classification automatique des images sont nombreuses et se sont répandues dans tout le monde. Par exemple, je trouve des applications dans le domaine médical telles que la reconnaissance cellulaire et tumorale, la reconnaissance manuscrite pour vérifier les codes postaux. Dans le domaine de la biométrie comme la reconnaissance faciale, les empreintes digitales, les veines de la main, l'iris.

Le point commun de toutes ces applications est qu'elles nécessitent la création d'une série de traitement d'images disponibles comprenant plusieurs étapes pour prendre une décision. Pas de fonctionnalités d'étape d'extraction et de phase d'apprentissage. Cette phase d'extraction nous permet d'opérer dans un espace numérique. Il est alors nécessaire de les mettre au stade de l'apprentissage.

I.2 Les motivations de la Classification des images

La classification des images consiste à attribuer systématiquement des images selon des catégories préétablies, à classer les images pour qu'elles correspondent à une catégorie et à marquer leur parenté avec d'autres catégories image.

En général, reconnaître des images est une tâche aisée pour un humain, qui, au cours de son existence, a acquis ses connaissances lui permettent de s'adapter aux changements provoqués par les différentes conditions de cueillette. Par exemple, il peut identifier relativement facilement des objets qui sont partiellement cachés par un autre proche ou lointain dans plusieurs directions en fonction de divers éclairages.

Cependant, les progrès technologiques dans l'acquisition d'images (microscopes, caméras, capteurs) et le stockage ont produit des bases de données riches en informations et des domaines d'application accrus, l'analyse humaine d'un grand nombre d'images, le temps requis, la nature des tâches répétitives et la quantité d'attention requise. sont problématiques. Cependant, pour un programme informatique où une image est un ensemble de valeurs, ce n'est pas forcément facile.

L'objectif de la classification d'images est de développer un système capable d'attribuer automatiquement des catégories aux images. En conséquence, le système peut effectuer des tâches spécialisées qui se sont avérées coûteuses à acquérir pour les humains, notamment en raison de contraintes physiques telles que l'attention, la fatigue ou le temps requis pour de grandes quantités de données d'image.

Les applications de classification automatique d'images vont de l'analyse de documents à la médecine et à l'armée. Par conséquent, il existe certaines applications dans le domaine médical, telles que la reconnaissance cellulaire et tumorale, la reconnaissance de l'écriture manuscrite des chèques et des codes postaux. Dans les zones urbaines telles que la reconnaissance des panneaux routiers, la reconnaissance des piétons, la détection des véhicules, la reconnaissance des bâtiments, etc., pour aider au positionnement. Dans le domaine de la biométrie, comme l'identification visage, empreinte digitale, iris.

Le point commun de toutes ces applications est qu'elles doivent construire une chaîne de traitement des images disponibles en plusieurs étapes afin de fournir une décision en sortie. Chaque étape de mise en œuvre d'un tel système de classification nécessite de trouver des méthodes adaptées pour obtenir les meilleures performances globales, c'est-à-dire l'étape d'extraction de caractéristiques et l'étape d'apprentissage. Souvent, nous avons des données d'image à partir desquelles nous devons extraire des informations pertinentes et les convertir sous la forme d'un vecteur numérique. Cette étape d'extraction nous permet de travailler dans l'espace numérique. Puis une question détaillée Lors de la phase d'apprentissage, à partir de ces données initiales, une fonction de décision décide si une nouvelle donnée appartient à l'une des classes existantes. **[2]**

I.3 Notion de base sur les images

Donc ce qui suit nous démons de notion de base sur les images :

I.3.1 Définition de l'image

Nous pouvons définir une image comme un tableau à deux dimensions où chaque élément (pixel) Représente la surface de base de l'image. La disposition de ces pixels est généralement en ligne et en colonne **[3]**.

Une image est définie mathématiquement comme une fonction bidimensionnelle f(x, y), où x et y sont des coordonnées spatiales et la magnitude de tout point (x, y) correspond à l'intensité ou au niveau de niveaux de gris.

Lorsque les points (x, y) et les grandeurs sont discrétisés, on parle de nombres ou d'images numériques. Dans ce dernier cas, la fonction f est remplacée par la lettre I, et le couple (x, y) par le couple (i, j), (**Figure I-1**).

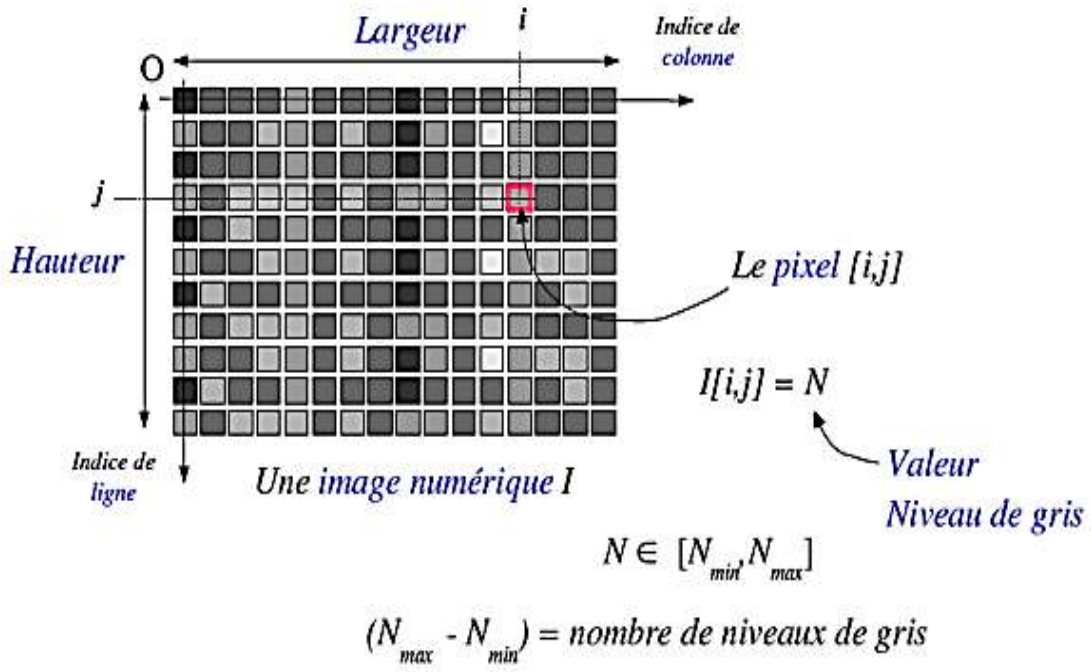


Figure I-1: Représentation des notions image et pixel **[3]**

I.3.2 Les différents types de format d'image

Les différents types de format d'image sont :

I.3.2.1 Image couleur RVB

Chaque pixel possède a une couleur décrite par la quantité de rouge (R), de vert (G) et de bleu (B). Chacune de ces trois composantes est codée sur l' 'intervalle [0 255], ce qui donne 256^3 = 16 777 216 couleurs possibles. **[3]**

I.3.2.2 Image binaire

Une image binaire est une image numérique dont les pixels ont 2 couleurs (généralement noir ou blanc). Ainsi, une image binaire peut être représentée comme une série/un tableau de 0 et de 1.

I.3.2.3 Images à niveaux des gris

De manière générale, une image en niveaux de gris contient 256 nuances de gris et, par convention, une valeur de 0 représente le noir (intensité lumineuse nulle) et une valeur de 255 représente le blanc (intensité lumineuse maximale). Le nombre 256 est lié à la quantification de l'image. En effet, tout entier représentant un niveau de gris est codé sur 8 bits. Il est donc compris entre 0 et 2^8-1=255. C'est la quantification la plus courante. L'image peut être encodée en niveau de gris sur 16 bits ($0 \leq n \leq 2^{16}-1$) ou 2 bits : dans ce dernier cas (niveau de gris) est égal à 0 ou 1 : c'est un problème avec les images binaires (noir et blanc). **[4]**

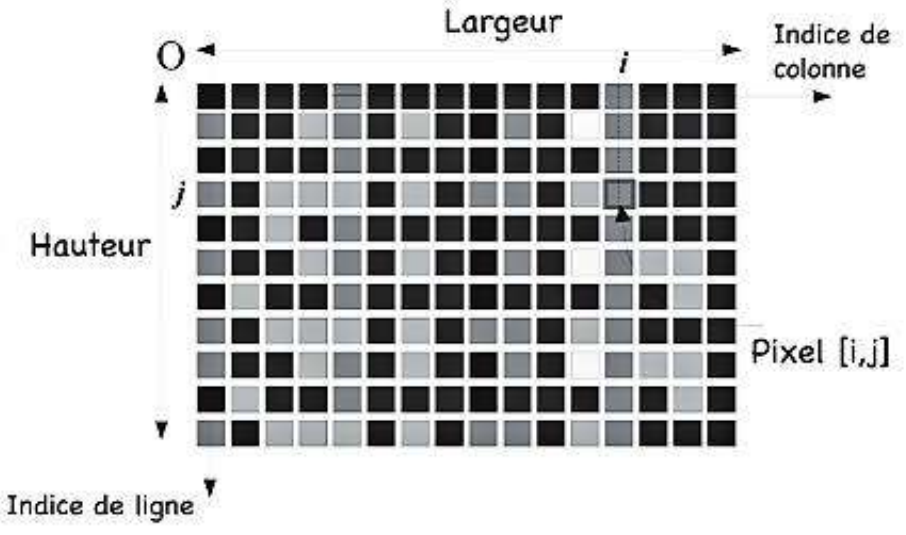


Figure I- 2: Pixels et niveaux de gris **[4]**

I.4 Caractéristiques de l'image

Une image est un ensemble structuré d'informations avec les paramètres suivants :

I.4.1 Pixel

Un pixel est une unité de mesure. Il s'appelle px, qui signifie élément d'image. Par exemple, il est utilisé pour mesurer la qualité des images matricielles numériques sur un ordinateur ou un téléviseur. Les pixels ont généralement une forme rectangulaire ou carrée. Il contient des informations sur la nuance de couleur, la quantité et la transparence ou la texture. Plus une image ne contient de pixels et de couleurs, plus sa résolution et sa qualité sont élevées. Lorsque nous zoomons sur une image, nous voyons tous les pixels qui la composent. Les pixels occupent un espace lorsqu'ils sont affichés, de sorte que les pixels sur les écrans en noir et blanc n'occupent qu'un octet d'espace. En revanche, aujourd'hui, la résolution la plus élevée peut avoir jusqu'à 32 bits de pixels. 16,8 millions d'écrans couleur ont des pixels de 24 bits, mais ils prennent parfois plus d'espace pour encoder et afficher la transparence. Plus un pixel a d'espace mémoire, plus il y a de couleurs dans l'image et plus l'image est nette.

I.4.2 Dimension

C'est la taille de l'image, elle se présente sous la forme d'une matrice dont les éléments sont des valeurs numériques, et le nombre de lignes de cette matrice multiplié par le nombre de colonnes est le nombre total de pixels de l'image**. [3]**

I.4.3 Résolution

La résolution d'une image est le nombre de pixels par unité de longueur de l'image, le plus souvent exprimée en PPP (Point Par Pouce) ou DPI (dots per inch), parfois en unités de 3 points par centimètre, la résolution définit la qualité de la image, résolution Plus le taux est élevé, plus les détails de l'image sont précis. **[5]**

I.4.4 Niveau de gris

Le niveau de gris est la valeur d'intensité lumineuse d'un point. La couleur d'un pixel peut aller du noir au blanc en passant par un nombre limité de couches intermédiaires. Ainsi, pour représenter une image en niveaux de gris, nous pouvons attribuer à chaque pixel de l'image une valeur qui correspond à la quantité de lumière réfléchie. Par exemple, la valeur peut être comprise entre 0 et 255. Ainsi chaque pixel n'est plus représenté par des bits, mais par des octets. Pour ce faire, le matériel utilisé pour afficher l'image doit pouvoir générer

différents niveaux de gris correspondants. Le nombre de niveaux de gris dépend du nombre de bits utilisés pour décrire la "couleur" de chaque pixel de l'image. Plus le nombre est élevé, plus il y a de niveaux possibles. **[6]**

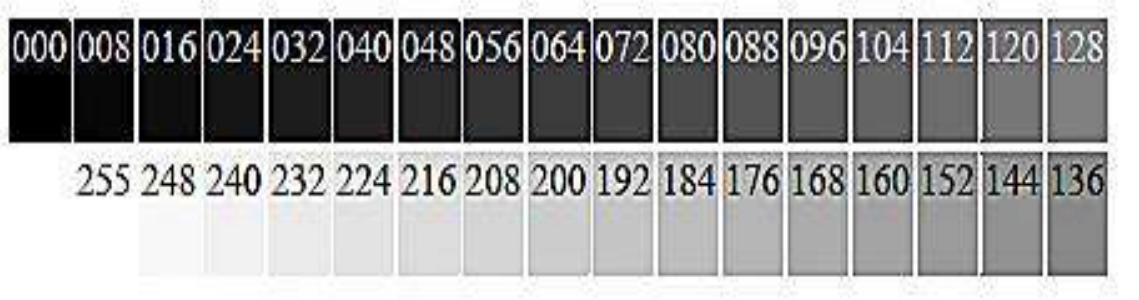


Tableau I-1: Les différents niveaux de gris **[6]**

I.4.5 Bruit

Dans les images numériques, le bruit numérique est toute fluctuation ou dégradation parasite subie par l'image de l'acquisition à l'enregistrement. Le bruit numérique est un concept général pour tout type d'image numérique, quel que soit le type de capteur à partir duquel elle a été acquise (appareil photo numérique, scanner, imageur thermique, etc.). Les sources de bruit numérique sont multiples, certaines sont physiquement liées à la qualité de l'éclairage, à la scène, à la température du capteur, à la stabilité du capteur d'image lors de l'acquisition, etc., d'autres apparaissent dans le processus de numérisation de l'information. **[7]**

I.4.6 Luminance

La luminosité est l'intensité lumineuse d'une source lumineuse s'étendant dans une direction donnée divisée par la surface apparente de cette source lumineuse dans la même direction. L'unité d'intensité lumineuse est la candela par mètre carré, symbole cd/m2**. [3]**

I.4.7 Voisinage

Le plan image est divisé en rectangles ou en hexagones, permettant ainsi utilisation du concept de voisinage **(voir Figure I-3).** Le voisinage d'un pixel est formé par tous les pixels situés autour du même pixel. Nous définissons également une dalle comme un ensemble de pixels qui définit le voisinage considéré autour d'un pixel.

Il existe deux types de quartiers :

Voisinage à 4 : nous ne considérons que les pixels qui partagent une arête commune avec le pixel considéré.

Voisinage à 8 : on considère tous les pixels qui ont au moins un point associé au pixel considéré.

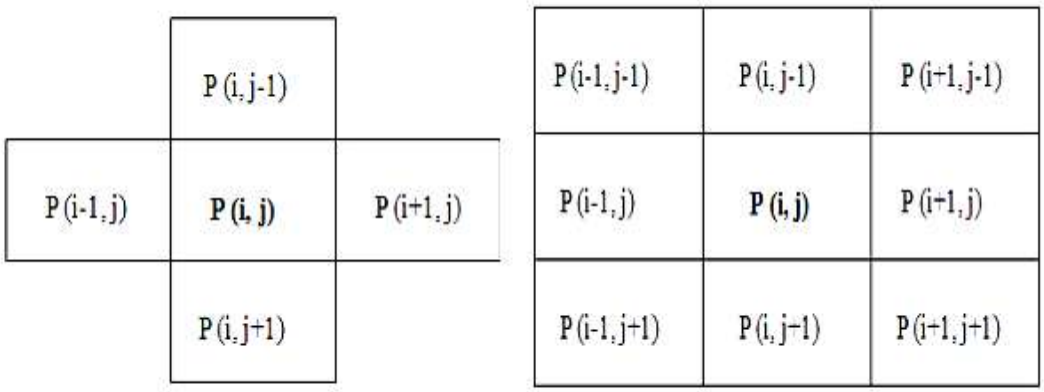


Figure I-3: Voisinage à 4, Voisinage à 8

I.4.8 Contrasts

C'est l'opposition apparente entre deux zones d'une image, plus précisément entre les zones sombres et claires de cette image. Le contraste est défini en fonction de la luminosité de deux zones de l'image.

Le niveau de gris est la valeur d'intensité lumineuse d'un point. Avec un nombre limité de couches intermédiaires, la couleur d'un pixel peut prendre des valeurs allant du noir au blanc. Ainsi, pour représenter une image en niveaux de gris, nous pouvons attribuer à chaque pixel de l'image une valeur qui correspond à la quantité de lumière réfléchie. Par exemple, la valeur peut être comprise entre 0 et 255. Ainsi, chaque pixel n'est plus représenté par des bits, mais par des octets. Pour ce faire, le matériel utilisé pour afficher l'image doit pouvoir générer différents niveaux de gris correspondants. Le nombre de niveaux de gris dépend du nombre de bits utilisés pour décrire la "couleur" de chaque pixel de l'image. Plus le nombre est élevé, plus il y a de niveaux possibles**. [8]**

I.4.9 Contour et texture

Les contours représentent les limites entre les objets de l'image ou entre deux les niveaux de gris représentent des pixels sensiblement différents. description des textures la structure de celles-ci. L'extraction de contour consiste à identifier des points dans une image Séparez les deux textures différentes**. [9]**

I.5 Méthodes de classification

Les méthodes de classification les plus courantes peuvent être divisées en deux catégories :

Les méthodes de classification supervisées et les méthodes de classification non supervisées.

I.5.1 Méthodes supervisées

Dans la classification supervisée, le réseau reçoit des étiquettes de classe dans le cadre de la formation (souvent appelées étiquettes de vérité terrain). Il apprend ensuite les caractéristiques importantes pour la classification et améliore progressivement l'identification et la mise en correspondance de ces caractéristiques avec la classe appropriée. Les méthodes de supervision comprennent :

K-plus proches voisins, arbres de décision, réseaux de neurones, machines à vecteurs de support **(SVM)** et classificateurs bayésiens.

Quel que soit le type de classification, nous sommes tous confrontés à des problèmes différents. Dans le cas supervisé, un problème important peut être le manque de données pour apprendre ou des données insuffisantes, telles que l'incertitude et l'imprécision, ce qui empêche de construire un modèle correct. Pour la classification non supervisée, la séparation des frontières entre les classes n'est pas toujours simple et identifiable.

Quel que soit le type de classification, les données multidimensionnelles ou la dépendance de la méthode de classification aux paramètres initiaux (comme le nombre de classes) peuvent poser des problèmes. **[10]**

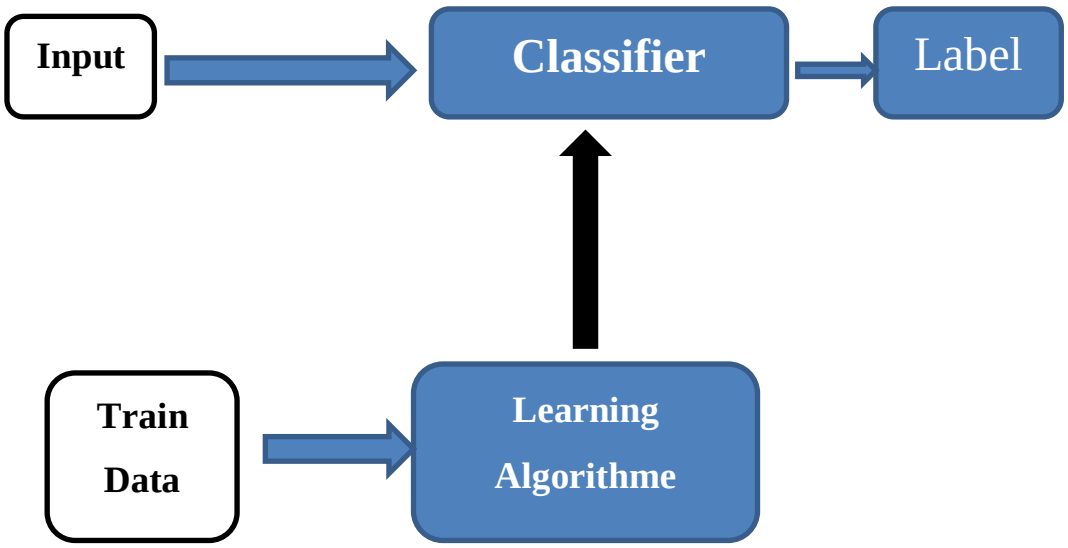


Figure I-4: Méthode supervisée

I.5.2 Méthodes non supervisées

Faites-le dans le sens inverse. Autrement dit, ils ne nécessitent aucun apprentissage et aucune tâche d'étiquetage manuel préalable. Il comprend des nuages de points qui représentent n'importe quel espace dans un ensemble de groupes appelés clusters. Cela implique généralement le domaine de l'analyse de données, comme l'ACP. Un "cluster" est une collection d'objets qui sont "similaires" les uns aux autres mais "différents" des objets appartenant à d'autres groupes**. [11]**

I.6 Indicateurs de performance en classification

I.6.1 Matrice de confusion

La première façon d'évaluer un classificateur est de comparer les valeurs observées de la variable dépendante avec les valeurs prédites fournies par le modèle. L'outil de choix est la matrice de confusion. Une matrice de confusion est un tableau de contingence qui compare la classe obtenue (colonnes) et la classe attendue (lignes) d'un échantillon. Donc sur la

diagonale principale on retrouve des valeurs bien classées hors Les lignes diagonales sont des éléments mal classés **(Tableau I-2).**

Classe Réelle	Classe prédite	
	C1	C2
C1	VP	FN
C2	FP	VN

Tableau I-2: Matrice de confusion dans le cas binaire **[10]**

- Un VP est un vrai positif, une observation qui a été classée comme positive et qui est effectivement positive.
- Les PF sont des faux positifs, c'est-à-dire des individus classés comme positifs mais en réalité négatifs.
- De même, FN est un faux négatif et VN est un vrai négatif.

Cette matrice peut dériver les paramètres suivants :

- Précision : Combien de ceux que l'algorithme a pu classer étaient corrects.

$$\textbf{Précision} = \frac{V_P}{V_P + F_P} \qquad (I.1)$$

- Rappel : Parmi ceux qui existent, combien d'algorithmes peut-on trouver.

$$\textbf{Rappel} = \frac{V_P}{V_P + F_N} \qquad (I.2)$$

- F-mesure : la moyenne harmonique entre la précision et le rappel.

$$\textbf{F-mesure} = \frac{2*(\text{précision}*\text{rappel})}{\text{rappel}+\text{précision}} \qquad \text{(I.3)}$$

La mesure F décrit la qualité de la classification en termes de catégories, mais Ne tient pas compte des éventuels déséquilibres entre les classes (ne tient pas compte des faux négatif**). [10]**

I.6.2 Courbe ROC (Received Operating Characteristic)

Les modèles de classification impliquent souvent des mesures quantitatives Lequel est déclaré positif et lequel est inférieur à celui qui est déclaré négatif. En changeant ce seuil s, la sensibilité et la spécificité changent.

Graphiquement, la courbe ROC (Receviez Operating Caractéristique) est représenté sous la forme suivante Tracé du taux de vrais positifs (sensibilité) en fonction du taux de faux positifs (1-spécificité). Il est obtenu en modifiant le seuil s : plus il augmente, plus la sensibilité diminue et la spécificité augmente.

En principe, la courbe ROC se situe au-dessus de la première bissectrice, correspondant au classement aléatoire des observations. Point (0; 1) Le graphique représente la situation idéale où toutes les observations positives sont bien classées et aucune observation négative n'est mal classée.

La métrique composite liée à la courbe ROC est la surface sous la courbe, l'AUC (Fois sous la courbe). Un modèle de classification est valide si l'AUC est proche de 1. A l'inverse, si l'AUC est proche de 0,5, le modèle de classification ne discrimine pas. **(Figure I-5)**. **[12]**

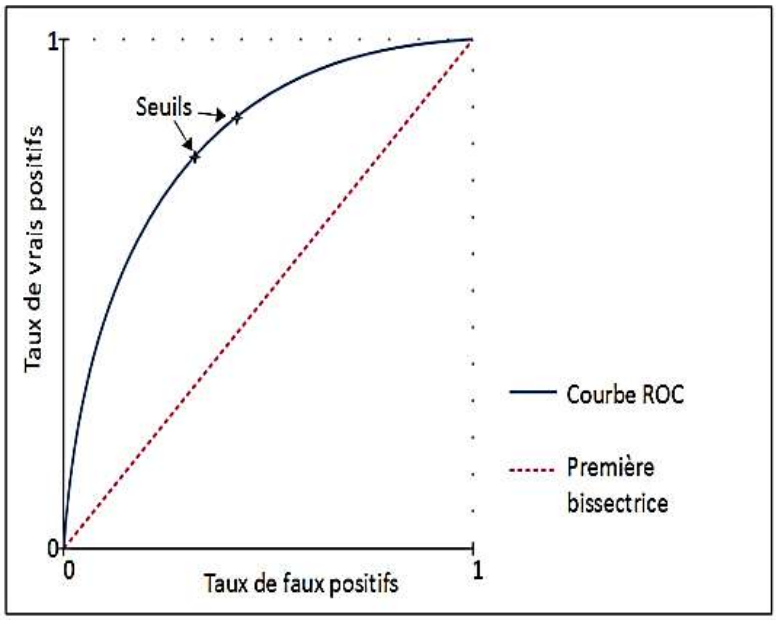


Figure I-5: Courbe ROC **[12]**

I.7 Classification des images et l'apprentissage machine

Les méthodes manuelles se sont avérées difficiles à appliquer à des tâches apparemment très simples telles que la classification d'images, la reconnaissance d'objets dans des images ou la reconnaissance vocale. Les données du monde réel (échantillons sonores ou pixels d'une image) sont complexes, variables et polluées par le bruit.

Pour une machine, une image est un tableau de nombres représentant la luminosité (ou la couleur) de chaque pixel. Signal acoustique Série de chiffres indiquant la pression atmosphérique à chaque instant.

Comment une machine peut-elle retranscrire une séquence de chiffres d'un signal sonore en une séquence de mots, tout en ignorant le bruit ambiant, l'accent du locuteur et les particularités de sa voix ? Comment une machine Peut-elle identifier un chien ou une chaise dans un schéma numérique d'une image alors que l'apparence du chien ou de la chaise et des objets qui l'entourent peuvent varier à l'infini ?

Il est presque impossible d'écrire un programme qui fonctionne de manière robuste en toutes circonstances. C'est là qu'intervient l'apprentissage automatique (également appelé apprentissage automatique). Ce qui motive les systèmes de toutes les grandes sociétés Internet, c'est l'apprentissage. Ils l'utilisent depuis longtemps pour filtrer le contenu

indésirable, trier les réponses de recherche, faire des suggestions ou sélectionner des informations qui intéressent chaque utilisateur.

Un système entraînable peut être considéré comme une boîte noire qui a une entrée, telle qu'une image, un son ou un texte, et une sortie qui peut représenter la classe de l'objet dans l'image, le mot parlé ou le texte du sujet parlé. . Ceux-ci sont appelés systèmes de classification ou de reconnaissance de formes.

Dans sa forme la plus courante, l'apprentissage automatique est supervisé : vous montrez à la machine une photo d'un objet en entrée, comme une voiture, et vous lui donnez la sortie dont la voiture a besoin. Il verrait alors une image d'un chien avec la sortie souhaitée pour le chien. Après chaque exemple, la machine ajuste ses paramètres internes pour rapprocher sa sortie de la sortie souhaitée. Après avoir montré à la machine des milliers ou des millions d'exemples étiquetés avec sa classe, la machine a pu classer correctement la plupart d'entre eux. Mais plus intéressant, elle a également pu classer correctement des images de voitures ou de chiens qu'elle n'avait jamais vues pendant la phase d'apprentissage. C'est ce qu'on appelle la capacité de généralisation.

Jusqu'à récemment, les systèmes de reconnaissance d'images classiques se composaient de deux modules : un extracteur de caractéristiques et un simple classificateur entraînable. L'extracteur de caractéristiques est programmé "à la main" pour convertir le tableau de nombres représentant l'image en une série de nombres, un vecteur de caractéristiques, chaque nombre représentant la présence ou l'absence de motifs simples dans l'image. Ce vecteur est envoyé à un classificateur, dont un type courant est un classificateur linéaire. Ce dernier calcule une somme pondérée de caractéristiques : chaque nombre est multiplié par un poids (positif ou négatif) avant d'être additionné. Si la somme est supérieure au seuil, la classe est identifiée. Les poids forment une sorte de "prototype" de la classe à laquelle le vecteur de caractéristiques est comparé. Les chargements sont différents à les classifieurs pour chacun classe et ce sont eux lequel sont modifiés dans pour l’apprentissage Les premières méthodes pour taxinomie affin entraînable datent pour les issues incontinentes années cinquante et sont toujours considérablement utilisées aujourd’hui Elles prennent les doucereux noms pour perceptron soit dégénérescence approvisionnement. **[13]**

I.8 Classification des images et les réseaux de neurones

Le problème avec les méthodes classiques de reconnaissance d'images est qu'il est difficile de construire un bon extracteur de caractéristiques et qu'il doit être repensé pour chaque nouvelle application.

C'est là qu'intervient l'apprentissage en profondeur ou l'apprentissage en profondeur en anglais. Il s'agit d'une classe de méthodes dont les principes sont connus depuis la fin des années 1980, mais dont l'usage ne s'est vraiment généralisé qu'en 2012, env.

L'idée est simple : un système entraînable se compose d'une série de modules, chaque module représentant une étape de traitement. Chaque module peut être formé avec des paramètres réglables similaires aux poids de classificateur linéaire. Entraînez le système de bout en bout : dans chaque cas, ajustez tous les paramètres de tous les modules afin que la sortie produite par le système soit plus proche de la sortie souhaitée. Le qualificatif profond vient de l'agencement de ces modules en couches successives.

Pour pouvoir entraîner le système de cette manière, il est nécessaire de savoir dans quel sens et de combien ajuster chaque paramètre de chaque module. Pour ce faire, le gradient doit être calculée, c'est-à-dire, pour chaque paramètre réglable, la quantité dont l'erreur de sortie augmentera ou diminuera lorsque le paramètre change d'une quantité donnée. Ce gradient est calculé à l'aide de la méthode de rétro propagation utilisée depuis le milieu des années 1980.

Dans son implémentation la plus courante, une architecture profonde peut être considérée comme un réseau multicouche d'éléments simples, similaire à un classificateur linéaire, interconnectés par des poids entraînables. C'est ce qu'on appelle un réseau neuronal multicouche.

Pourquoi les nerfs ? Un modèle extrêmement simplifié de neurones cérébraux les traite comme des neurones informatiques pondérer les sommes et activer leurs sorties lorsqu'un seuil est dépassé. L'apprentissage modifie l'efficacité des synapses, le poids des connexions entre les neurones. Les réseaux de neurones ne sont pas des modèles précis des circuits cérébraux, mais sont considérés comme des modèles conceptuels ou fonctionnels. Les réseaux de neurones s'inspirent du cerveau, tout comme les avions s'inspirent des oiseaux.

La force des architectures profondes est leur capacité à apprendre à représenter le monde de manière hiérarchique. Étant donné que toutes les couches peuvent être entraînées, il n'est pas nécessaire de créer manuellement des extracteurs d'entités. La formation résoudra ce

problème. De plus, la première couche. Des caractéristiques simples (présence de contours) seront extraites et les couches suivantes se combineront pour former des concepts de plus en plus complexes et abstraits : les contours sont combinés en motifs, les motifs sont combinés en parties d'objets, les parties d'objets sont combinées en objets, etc. **[14]**

I.9 Conclusion

Ce chapitre est consacré aux notion de base sur les image et motivations les plus importantes pour la reconnaissance et la classification d'images, puis nous nous concentrons sur ces méthodes Le classement le plus important. Après cela, nous introduisons les concepts et les idées les plus importants de l'apprentissage automatique profond pour la classification d'images basée sur les réseaux de neurones. Dans le chapitre 2, nous expliquerons que les réseaux de neurones consolatifs sont actuellement les méthodes de classification d'images les plus efficaces.

Chapitre II :
Les réseaux de neurones convolutionnels (CNN)

II.1 Introduction

Les réseaux de neurones convolutifs sont des réseaux d'apprentissage profond inspirés du cortex visuel [Hubel & Wiesel 1962]. Ces réseaux ont été utilisés dans les systèmes de recommandation [Ying et al. 2018], le traitement du langage naturel [Kim 2014] et la vision par ordinateur [Krizhevsky et al. 2012]. Ils ont été développés avec beaucoup de succès en vision par ordinateur en raison de leurs caractéristiques inspirées des systèmes visuels naturels.

En vision par ordinateur, le processus de classification des méthodes d'apprentissage classiques repose sur deux étapes principales : l'extraction de caractéristiques et l'apprentissage. Ces fonctionnalités sont Les éléments sont considérés comme fabriqués à la main en raison de la nécessité d'un effort humain dans l'étude des attributs distinctifs. mode d'emploi Ces caractéristiques sont extraites de manière non supervisée. Cette séparation entre le module d'extraction et le module de classification peut nuire à la tâche de classification si certaines propriétés discriminantes sont ignorées lors de l'étape d'extraction.

Contrairement aux méthodes d'apprentissage classiques, les CNN implémentent implicitement le processus d'extraction de caractéristiques à travers des couches convolutifs, où la première couche représente une seule caractéristique. Puis combinez ces fonctionnalités D'autres, plus complexes, se forment dans des couches plus profondes. Cette particularité fait des CNN un bon outil pour classer les données non structurées telles que les images et le texte. Malgré ces avantages, les CNN profonds souffrent d'un sur ajustement sur des quantités limitées de données, et en raison du sur ajustement, ils sont mieux adaptés pour de grandes quantités de données.

Régions profondes classiques en termes de complexité temporelle et spatiale. Ces réseaux se caractérisent par leur stratégie de partage de paramètres. Contrairement aux DNN où les couches adjacentes sont fortement connectées, dans les CNN, chaque neurone de la couche actuelle n'est connecté qu'à un sous-ensemble de neurones de la couche précédente. Les CNN sont considérés comme des méthodes structurellement stables qui réduisent les problèmes de sur ajustement en optimisant le nombre de paramètres.

Dans ce chapitre, nous expliquerons superficiellement l’apprentissage profond et un réseau CNN en détail.

II.2 L'apprentissage en profondeur (Deep learning)

II.2.1 Definition

Apprentissage profond (DL) est une branche de l'apprentissage automatique qui est basée sur les réseaux de neurones. Les réseaux DL ont été largement exploités dans différents domaines grâce à leur capacité dans la résolution des problèmes complexes. Ces réseaux ont prouvé leur efficacité par rapport aux méthodes d'apprentissage classique dans plusieurs applications récentes. Malgré leurs avantages, ils ont plusieurs défis liés aux problèmes de sur-apprentissage, de dégradation de gradient, et de la complexité temporelle élevée. Afin de réduire ces problèmes, plusieurs architectures optimisées ont été proposées en apprentissage supervisé et non supervisé. Le but de ce chapitre est de présenter l'évolution des réseaux de neurones aux réseaux d'apprentissage profond, d'expliquer le processus d'apprentissage dans un réseau de neurones, de détailler les architectures connues de type DL, et de présenter les problèmes de chaque architecture avec les solutions proposées. **[15]**

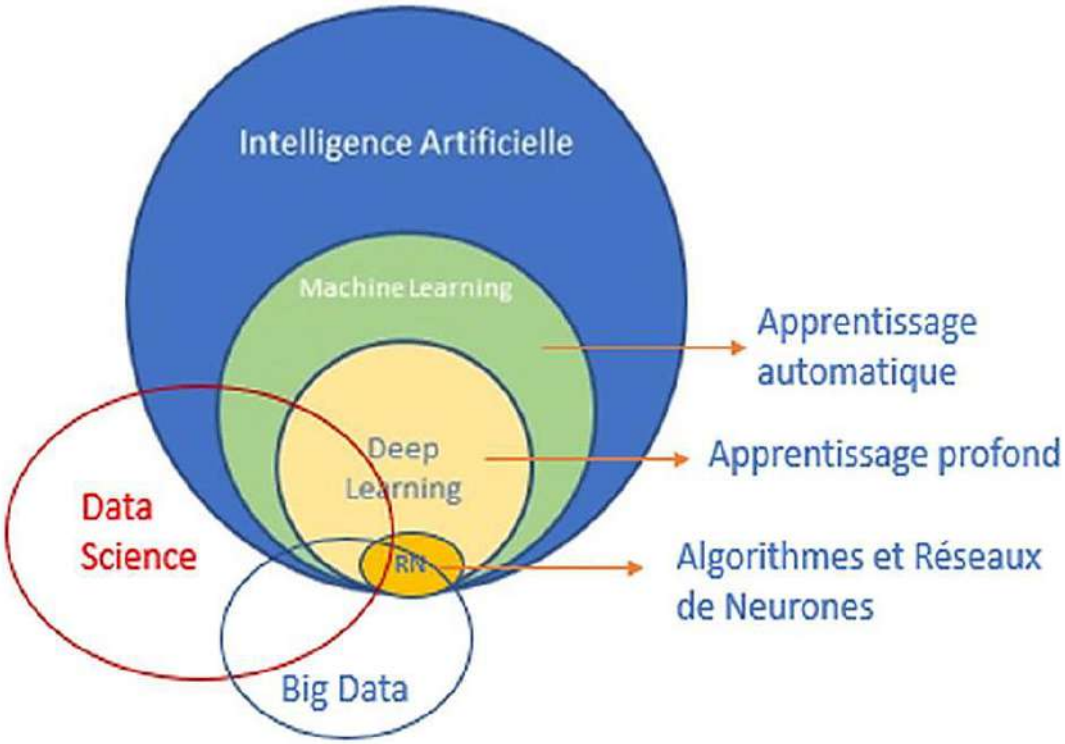


Figure II-1: schéma de l'apprentissage profond

II.3 Les différents algorithmes de deep learning

Plusieurs types d'algorithmes sont utilisés dans le deep learning. Chaque algorithme a ses propres spécificités et applications.

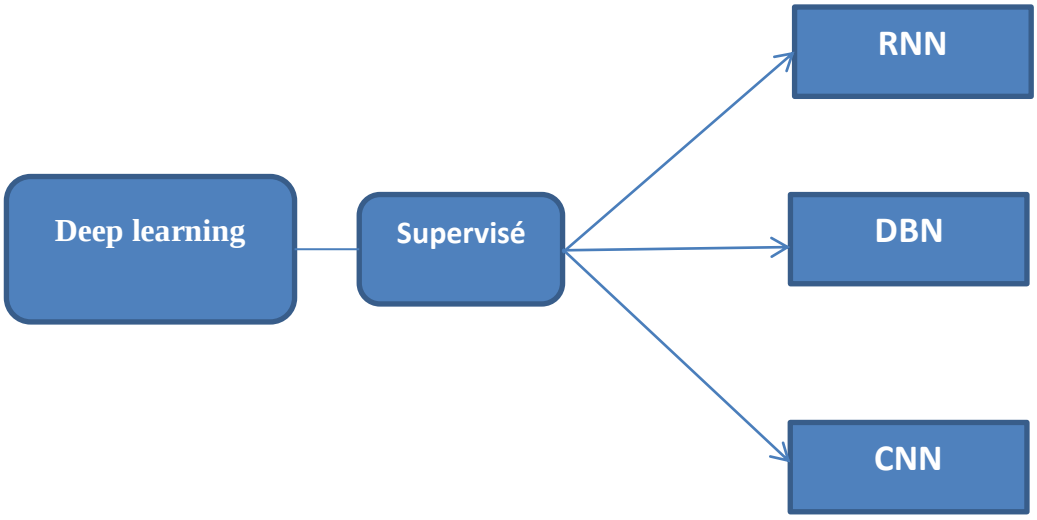


Figure II-2: Les différents algorithmes supervisés de DL

II.3.1 DBN

Deep Belief Networks [Hin06] [LeC15] est un modèle probabiliste génératif créé par plusieurs couches cachées. Chaque couche est associée à des dépendances d'ordre supérieur Les méthodes génératives incluent la modélisation de la distribution conjointe des variables observées .Un critère basé sur la log-vraisemblance est généralement utilisé. Cette méthode fait Aucune donnée étiquetée n'est requise. Le DBN utilisé pour former nos données est Il se compose d'une couche visible et de deux couches cachées. En fait, DBN est une machine .Le Boltzmann restreint (RBM : Restricted Boltzmann Machine) apprend les fonctionnalités (étape de prétraitement) puis les propage pour améliorer le réseau .Après l'étape de prétraitement, le réseau DBN agit comme un perceptron multicouche (Perceptron multicouche). Enfin, après avoir ajouté une couche de sortie, nous pouvons insérer un La régression logistique comme classifier. **[16]**

II.3.2 RNN

Un réseau de neurones récurrent [Pineda 1987] est un réseau Des systèmes d'apprentissage en profondeur conçus pour apprendre à partir des données en ordre. Ils

représentent un moyen de partage du poids temps. Contrairement aux MLP, dans les RNN, les activations dans les couches cachées dépend des entrées actuelles et passées. Toute connexion permet à une ou plusieurs informations prédites à l'étape précédente d'être prises en compte dans l'étape en cours, où les mêmes éléments Traité différemment. En conséquence, le résultat L'étape t - 1 affecte la décision de l'étape suivante t. Ces propriétés font de RNN dans Le traitement de texte, car un mot peut avoir plusieurs sens, selon sa place dans la phrase. **[15]**

II.3.3 Domains application de l'apprentissage profonde

Le deep Learning est utilisé dans de nombreux domaines :

- Reconnaissance d'image.
- Traduction automatisée.
- Voiture autonome.
- Diagnostic médical.
- Recommandations personnelles.
- Modération automatique des réseaux sociaux.
- Prévisions financières et négociation automatisée.
- Identifier les pièces défectueuses.
- Détection de logiciels malveillants ou de fraude.
- Robots de clavardage (agents de clavardage).
- Exploration spatiale.
- Robots intelligents.

II.4 Quelques Modèles de réseaux de neurones

II.4.1 Perceptron

Le perceptron est un réseau neuronal simple, proposé par Frank Rosenblatt en 1958. Il s'agit d'un réseau monocouche linéaire avec seulement deux couches, l'une représentant l'entrée du système et l'autre représentant la sortie. La connexion entre ces deux couches est modifiable et bidirectionnelle **[17].** Le processus d'apprentissage est supervisé et le réseau est capable de résoudre des opérations logiques simples.

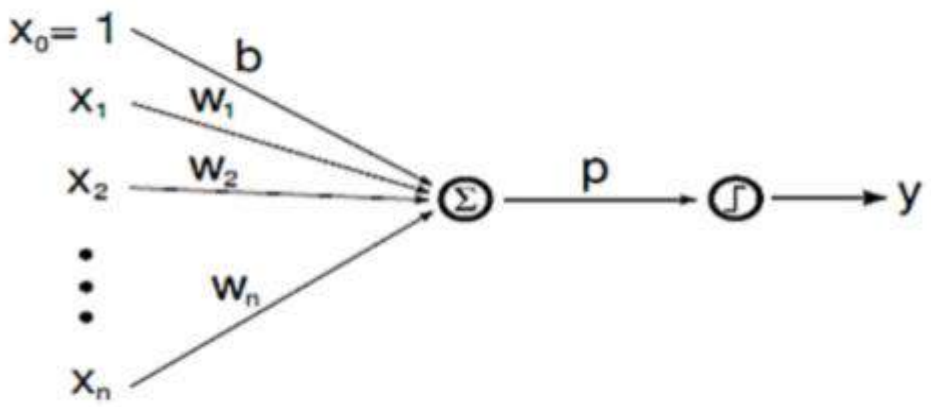


Figure II- 3: Architecture de perceptron **[18]**

Le perceptron prend en entrée n valeurs X1, X2,…..Xn et calcule une sortie O qui dépend de la somme de ces composantes, pondérée par le poids réel W et le seuil θ La sortie O de ce réseau aura la forme : O =1 si wX ≥ θ 0 si wX < θ Les algorithmes d'apprentissage sont des règles d'apprentissage Hebbian qui affectent le changement Pondéré en multipliant l'entrée d'un neurone par sa sortie et le taux d'apprentissage du réseau. **[18]**

II.4.2 Perceptron multicouche

Le perceptron multicouche est une amélioration du perceptron Contient une ou plusieurs couches cachées, faisant des réseaux MLP un outil puissant tâches complexes. Largement utilisé dans la prise de décision reconnaissance de visage. Les réseaux MLP sont généralement des réseaux matures Lié. Les neurones de la première couche reçoivent des vecteurs d'entrée, ils calculent Leurs sorties sont transmises aux neurones de la deuxième couche, qui calculent leur propre Leurs sorties etc. vont de couche en couche jusqu'à en produire une **[19].** Dans un réseau MLP, il n'y a pas de connexions entre les unités d'une même couche.

Capteur Plusieurs couches sont utilisées avec un apprentissage supervisé et également avec des techniques de rétro propagation pour la correction d'erreurs. **[19]**

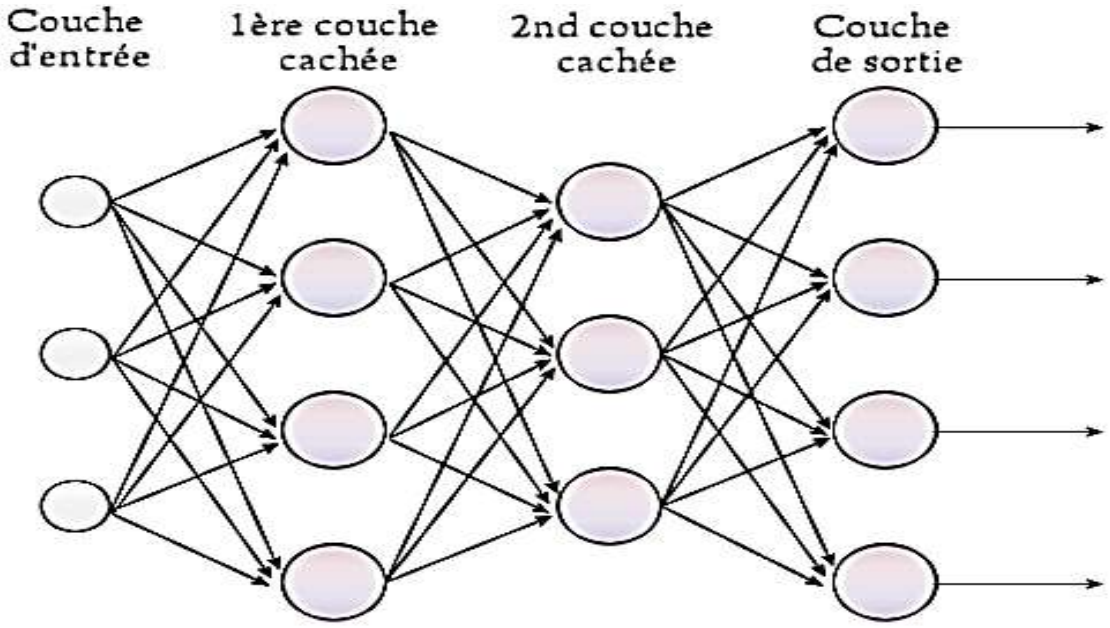


Figure II-4: Architecture de perceptron multicouche **[19]**

II.5 Réseaux de neurone convolutifs

Donc ce qui suit nous démons réseaux de neurone convolutifs :

II.5.1 Definition

Un réseau de neurones convolutifs (CNN) est un algorithme d'apprentissage en profondeur. Ce réseau a été largement utilisé en vision par ordinateur pour la classification et la détection d'objets en raison de sa fonctionnalité inspirée du cortex visuel. Contrairement aux réseaux d'apprentissage en profondeur classiques, les CNN comportent des couches de convolution et de mise en commun. Ces couches introduisent un chaînage partiel pour réduire le nombre de paramètres et renforcer le partage des caractéristiques communes. Malgré ces avantages, ils souffrent de plusieurs défis liés au problème de sur ajustement avec un volume de données limité et une complexité de calcul élevée. Ces architectures sont basées sur des blocs convolutifs optimisés qui peuvent générer des structures plus profondes et moins exigeantes en termes de capacité de calcul et de stockage. Le but de ce chapitre est de détailler la structure générale de CNN et de comparer les architectures courantes de type CNN dans la classification, la détection d'objets et la segmentation**. [15]**

II.5.2 Architecture de réseaux CNN

Les réseaux de neurones convolutifs tels qu'illustrés à la (**figure II-5)** intègrent généralement plusieurs sofas de convolution, de regroupement (pooling) et de

standardisation qui extraient des caractéristiques de bas niveau à haut niveau de l'entrée pendant le processus d'inférence en aval. Ces caractéristiques peuvent être classées en un nombre fini de classes de foray standard les longes chair d'ordre finales telles que le perceptron multicouche ou les sofas entièrement connectées. **[20], [21]**

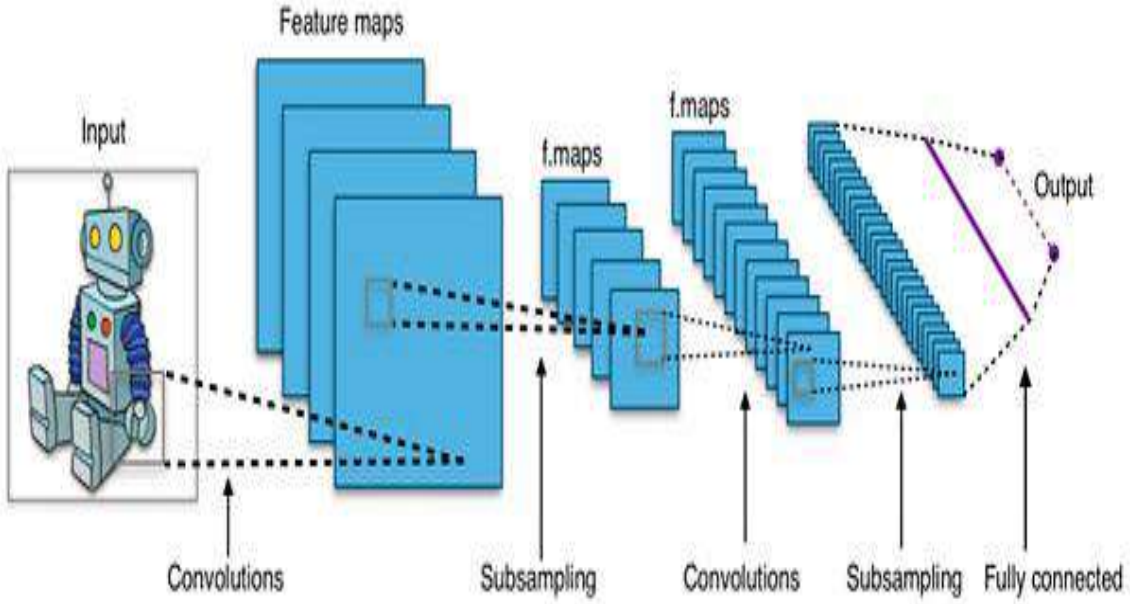


Figure II-5: Structure générale d'un réseau de neurones convolutifs (CNN)

Une architecture CNN est formée par un empilement de couches de traitement :

- La couche de convolution (CONV) qui traite les données d'un champ récepteur
- La couche de pooling (POOL), qui permet de compresser l'information en réduisant la taille de l'image intermédiaire (souvent par sous-échantillonnage).
- La couche de correction (ReLU), souvent appelée par abus 'ReLU' en référence à la fonction d'activation (Unité de rectification linéaire)
- La couche "entièrement connectée" (FC), qui est une couche de type perceptron.
- La couche de perte (LOSS).

II.5.3 Couche de convolution (CONV)

Les couches convolutives forment le cœur des réseaux convolutifs. Les couches consistent en une grille rectangulaire de neurones avec un petit champ récepteur qui s'étend

sur toute la profondeur du volume d'entrée. Ainsi, une couche convolutive n'est qu'une convolution d'image de la couche précédente, où les poids spécifient les filtres convolutifs. La couche convolutive déterminera la sortie des neurones connectés à la région En calculant leurs poids et en se connectant à Entrez le volume. ReLu est conçu pour appliquer une fonction d'activation "de base" (telle qu'une fonction sigmoïde) à la sortie des activations produites par la couche précédente**. [22] [23]**

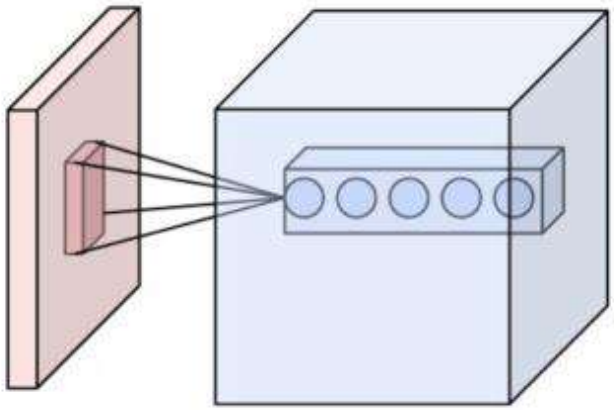


Figure II-6: Couche de convolution(CONV)

II.5.4 Couche de pooling (Pool)

La couche d'échantillonnage est chargée de réduire la taille spatiale des cartes d'entités, mais elle conserve les informations les plus importantes. Il existe différents types d'échantillonnage comme max ( ), moyen ( ), la norme L2 ou les transformées spectrales comme **DWT** "Discrete Wavelet Transform" et **DCT** "Discrete Cosine Transform".

L'échantillonnage consiste à appliquer le noyau à la matrice des cartes d'activation en utilisant une taille de pas préalablement définie (généralement la taille de pas est égale à la taille du noyau pour éviter les phénomènes de chevauchement). Max Pooling renvoie la valeur maximale de la partie de l'image couverte par le noyau. Au lieu de prendre la valeur maximale, on peut prendre la moyenne de tous les éléments couverts par le noyau, ce qui est assuré par la mutualisation moyenne. En pratique, Max Pooling est plus utilisé que moyen Pooling car il fonctionne mieux. La forme la plus courante est une couche de regroupement avec une taille de noyau de 2 × 2 et une foulée de 2, ce qui réduit la dimensionnalité de l'entrée par 2 et élimine 75% des activations et gardez la dimension de profondeur la même. Dans la Figure 2-6, le nombre d'activations passe de 16 à 4 en appliquant la mise en commun maximale**. [24]**

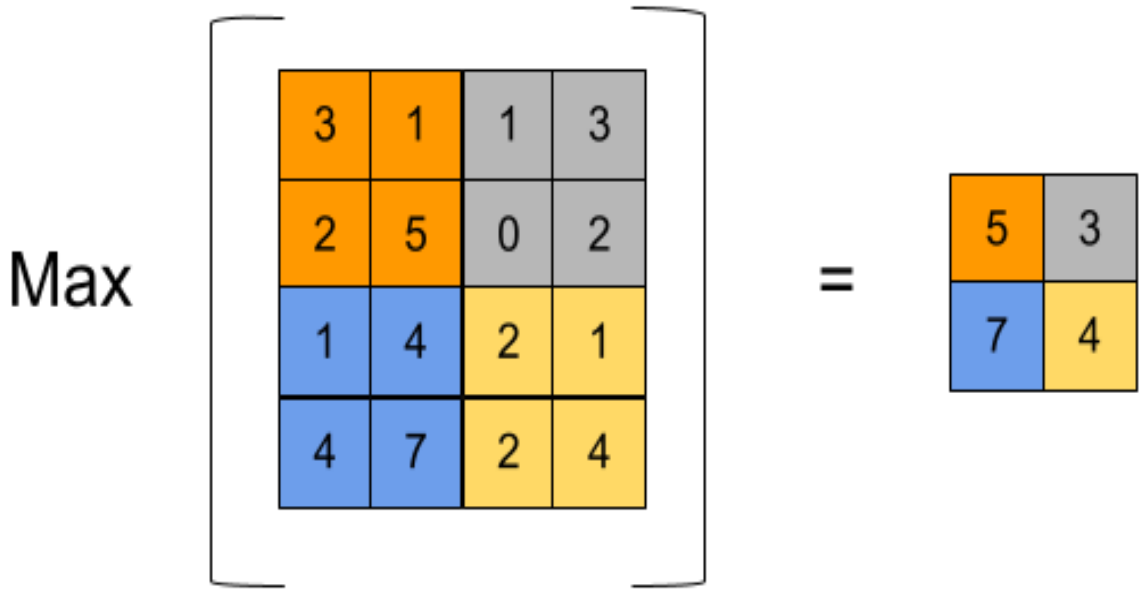


Figure II-7: Exemple d'une opération de Maxpooling de taille 2 × 2

L'échantillonnage a pour effet de réduire progressivement la taille Des données d'entrée. En effet, il permet **:**

- Rendre la représentation d'entrée (carte des caractéristiques) plus petite, et Plus facile à gérer.
- Réduire la puissance de calcul nécessaire au traitement des données en réduisant taille, de sorte qu'il contrôle le sur ajustement.
- Extraction des caractéristiques dominantes invariantes en rotation, position, maintenant ainsi l'efficacité du processus d'apprentissage du modèle.
- Aide à obtenir une représentation qu'invariante de l'image. c'est très puissant Parce qu'il peut détecter des objets dans les images, peu importe où ils se trouvent Trouver.

II.5.5 Fonction activation

Une fonction d'activation est une fonction mathématique utilisée pour le signal émis par un neurone artificiel. Le terme "fonction d'activation" est dérivé du terme biologique "potentiel d'activation", un seuil de stimulation qui, s'il est atteint, provoque Neurones.

La fonction d'activation est généralement une fonction non linéaire. Utiliser la fonction L'activation non linéaire est essentielle pour une raison simple

Les fonctions linéaires ne fonctionnent qu'avec une seule couche neuronale. en augmentant Un biais, le neurone calcule le produit scalaire entre le signal d'entrée et le vecteur poids. Le résultat est utilisé pour activer la fonction. Après l'opération de produit scalaire, Les fonctions d'activation sont utilisées pour introduire des opérations non linéaires. Cette non-linéarité permet plusieurs changements d'état interne sur la même classe d'objets **[17].**

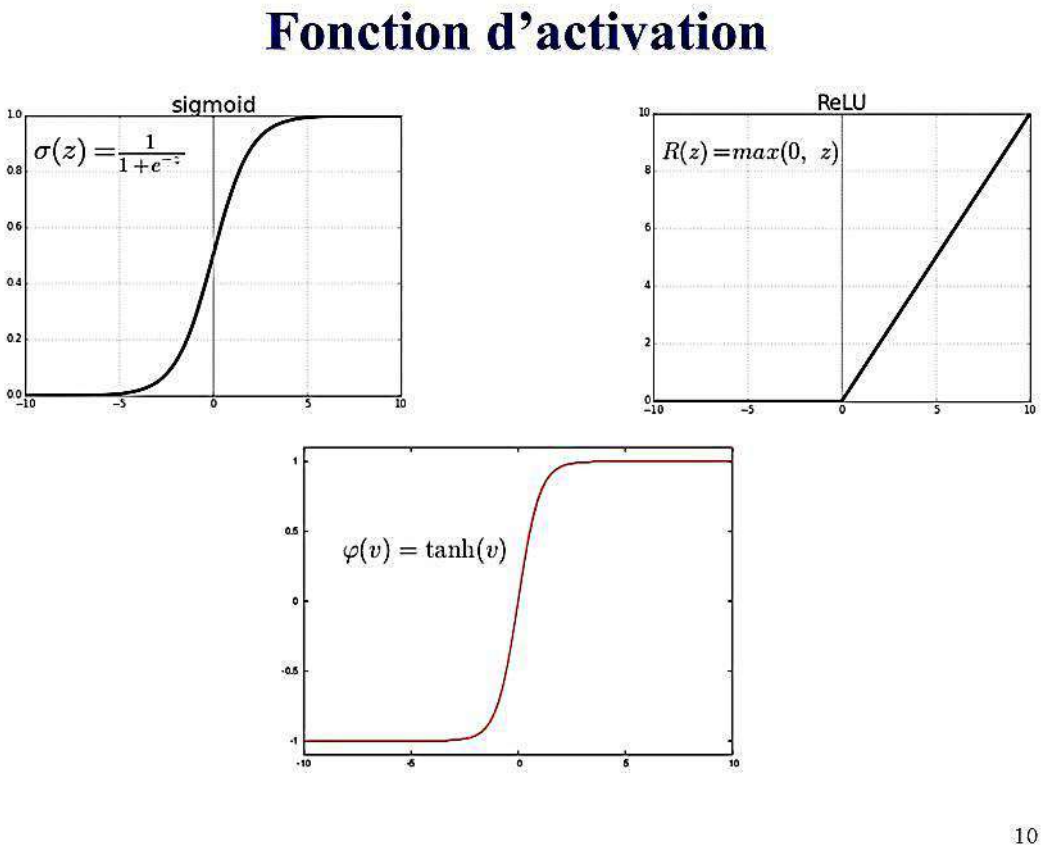


Figure II-8: les fonctions d'activation **[17]**

Or, le choix de la fonction d'activation s'avère être un élément clé des réseaux de neurones.

Cette fonctionnalité a un impact significatif sur les performances du réseau. dans un réseau neurones, il faut donc choisir le bon type de fonction d'activation. Différentes sortes J'utilise ces fonctionnalités depuis des années. Seulement le plus important

Cette section.

- **Sigmoid** : $\boldsymbol{\varphi(X) = \frac{1}{1+e^{-x}}}$. (II.1)
- **Tanh** : $\boldsymbol{\varphi(X) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}}$. (II.2)
- **Relu** : $\boldsymbol{\varphi(X) = \max(0, x)}$. (II.3)

La fonction **(ReLU)** est le plus couramment utilisé

II.5.6 Couch de correction (ReLu)

La sortie de la couche convolutive est essentiellement couche d'entrée. Comme un réseau de neurones normal, cette sortie doit passer fonction d'activation. On trouve que dans la couche convolutive, la fonction max (0, x) est plus efficace que la fonction sigmoïde en termes d'activation **[25].** Fonction max (0, x) est fondamentalement une fonction linéaire corrigée car il ne s'agit que d'une correction ou d'une réfraction La moitié négative de y = x est 0 **(Figure II-9).** C'est pourquoi on l'appelle souvent Les redresseurs et les nœuds activés par des redresseurs sont également appelés unités linéaires redressées ou ReLU. L'activation ReLU est généralement implémentée en tant que couche séparée après la couche convolutive.

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

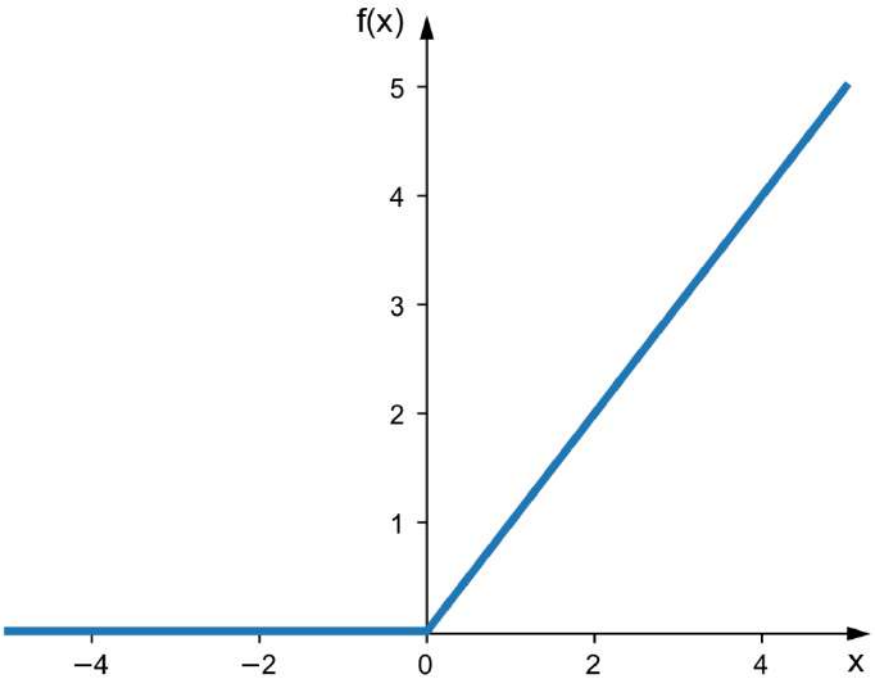


Figure II-9: Forme de la fonction ReLu

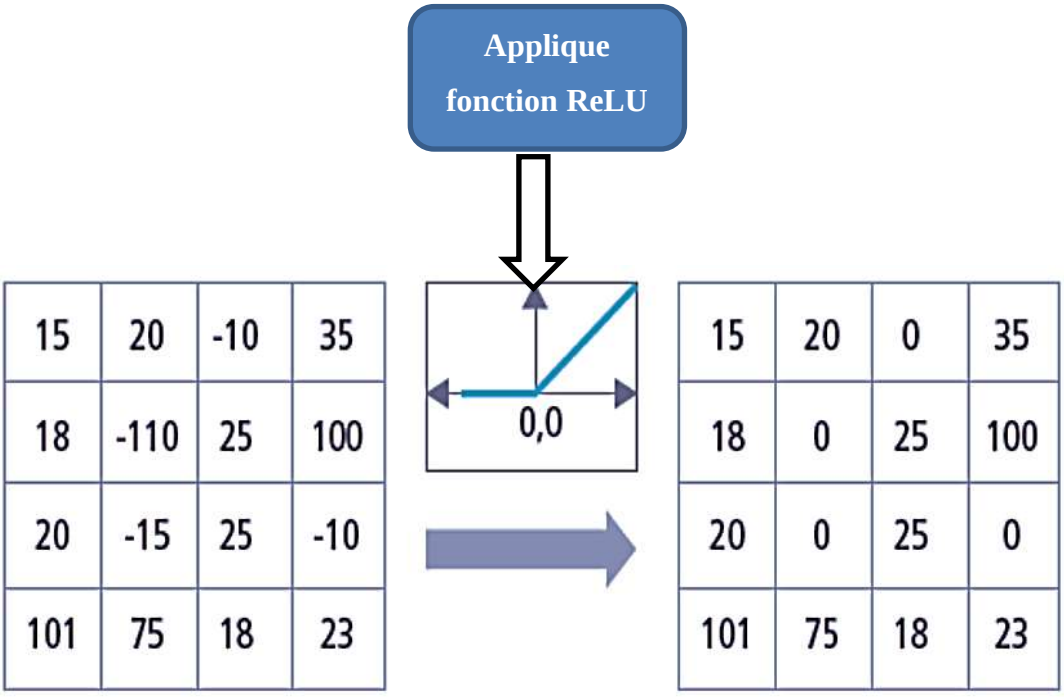


Figure II-10: exemple application de RELU

II.5.7 Couch entièrement connecter (FC)

Après plusieurs couches de convolution et de regroupement maximal, l'inférence avancée dans les réseaux de neurones se fait via des couches entièrement connectées. Les neurones d'une couche entièrement connectée ont des connexions à toutes les sorties de la couche précédente (comme on le voit souvent dans les réseaux de neurones réguliers). Par conséquent, leurs fonctions d'activation peuvent être calculées par multiplication matricielle et décalage de polarisation.

II.5.8 Couch de perte (LOSS)

La couche de perte spécifie comment le réseau est formé pour pénaliser Signal attendu et signal réel. Souvent, elle est le dernier canapé du réseau. Il est Diverses fonctions de perte adaptées à des tâches spécifiques sont disponibles. Soft max est une fonction qui calcule la distribution de probabilité de la classe de sortie. **[25]**

II.6 Exemples de modèles de CNN

La forme la plus courante d'architecture de réseau neuronal convolutif consiste à empiler plusieurs couches Conv-ReLU, à utiliser une couche Pool après elles et à répéter ce modèle jusqu'à ce que l'entrée s'effondre dans un espace suffisamment petit. À un moment donné, une couche entièrement connectée (FC) est généralement placée. La dernière couche entièrement connectée est connectée à la sortie. Voici quelques architectures courantes de réseaux de neurones convolutifs qui suivent ce modèle :

- **INPUT-> CONV-> RELU-> FC**
- **INPUT** -> [**CONV** -> **RELU** -> **POOL]** * **2** -> **FC** -> **RELU** -> **FC** Ici, il y a une couche de **CONV** unique entre chaque couche **POOL**.
- **INPUT -> [CONV -> RELU -> CONV -> RELU -> POOL] * 3** ->
  [FC -> RELU] * 2 -> **FC** Ici, il y a deux couches **CONV** empilées avant chaque couche **POOL**.

L'empilement des couches CONV avec de petits filtres de regroupement permet un traitement plus puissant avec moins de paramètres. Cependant, l'inconvénient est qu'il nécessite plus de puissance de calcul (y compris tous les résultats intermédiaires de la couche CONV). **[26]**

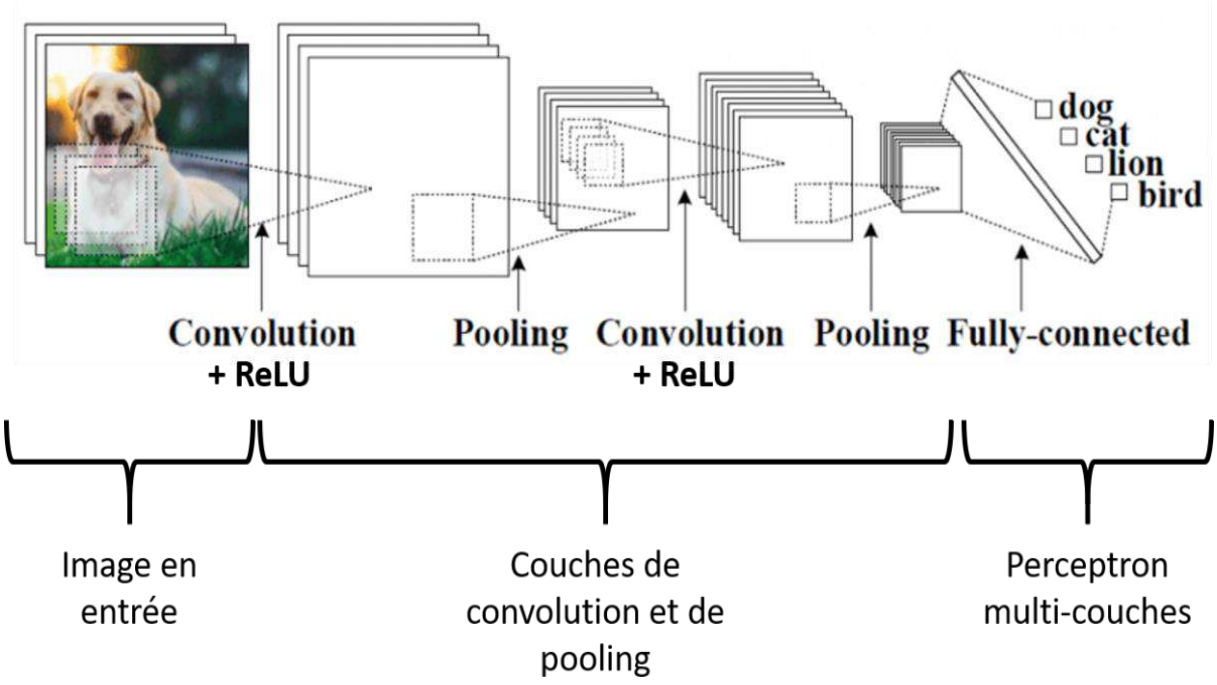


Figure II-11: Exemples de modèles de CNN **[26]**

II.6.1 Réseau AlexNet

Ce réseau, proposé par Krizhevsky et al **[25]**, améliore considérablement la précision de la classification par rapport aux méthodes traditionnelles d'apprentissage automatique basées sur des fonctionnalités artisanales, et a été le premier modèle CNN à remporter le concours ILSVRC 2012. AlexNet se compose de 5 couches convolutionnelles et de 3 couches FC, il a 61 millions de poids de noyau (paramètres) et nécessite environ 1,4 milliard d'opérations de multiplication et d'addition. Sur la base de données ImageNet Russakovsky **[28]**, le top 5 de la précision du modèle AlexNet Caffe est de 80,2 %. Il existe trois tailles de noyau différentes (Nkx Nky) dans les couches convolutives d'AlexNet : 11×11, 5×5 et 3×3.

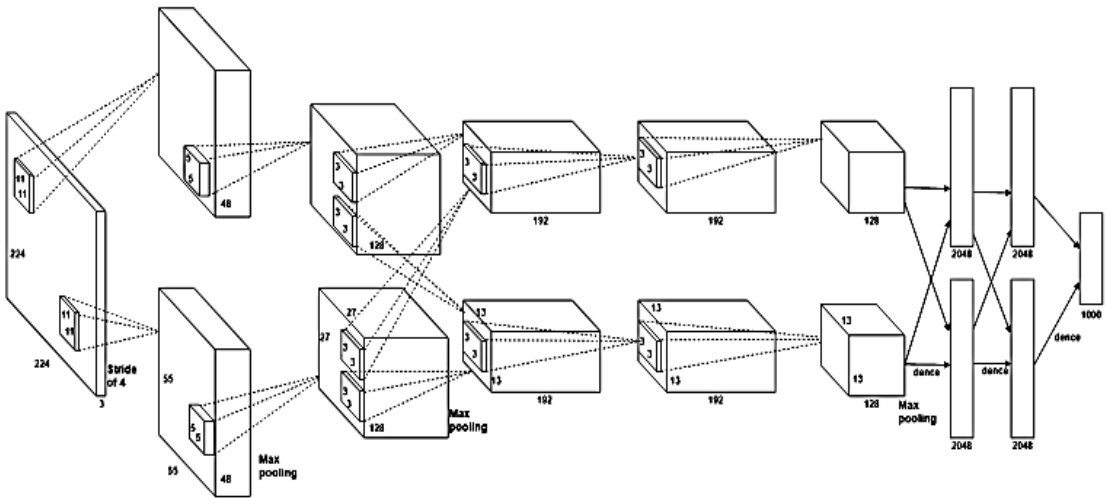


Figure II-12: Architecture Alexnet

II.6.2 Réseau NiN

Le réseau NiN « réseau en réseau » a été proposé par Lin et al. **[29]**. Il remplace la couche FC par une couche de type global average pooling à la fin pour économiser beaucoup de poids tout en étant capable d'atteindre le même niveau de précision qu'AlexNet sur la base de données ImageNet. NiN se compose de petites couches Les convolutions empilées avec une taille de noyau 1× 1 peuvent mieux extraire les fonctionnalités locales. Comme AlexNet, NiN a des dimensions de base : 11 × 11,5 × 5 et 3 × 3.

II.6.3 Réseau VGG

Le réseau VGG, proposé par Simonyan et Zisserman**,** a une structure très régulière avec seulement 3×3 couches convolutives et 2×2 couches max-pooling. Cependant, les trois couches FC nécessitent environ 123 millions de poids, et les couches convolutives dans VGG-16 nécessitent également environ 15 millions de poids. VGG-16 global nécessite environ 31 milliards d'opérations pour traiter l'image dans le processus d'anticipation. Le top 5 de la précision du modèle VGG-16 validé par ImageNet est de 88,7 %. **[30]**

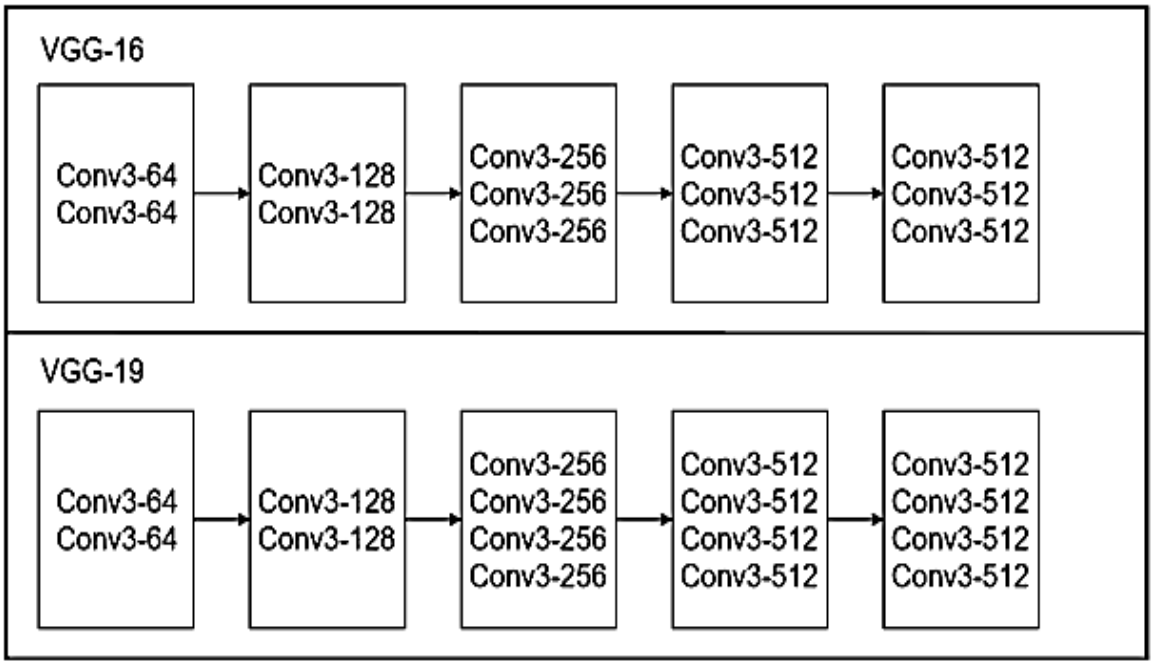


Figure II-13: les couches de Vgg-16, Vgg-19

II.6.4 Réseau GoogLeNet

Le réseau VGG a été proposé par Szegedy et al. **[31],** qui atteint le même niveau de précision que VGG-16, avec 89,0 % de précision dans le top 5 sur ImageNet, tout en ne nécessitant que 6,1 millions de poids et 3,2 milliards d'opérations, ce qui est obtenu par des schémas plus complexes. Un module de démarrage se compose de six couches Convolutions avec des tailles de noyau de 1 × 1, 3 × 3 et 5 × 5, et des couches de regroupement maximum dans quatre branches parallèles.

II.6.5 Réseau ResNet

Le réseau ResNet a été proposé par He et al. ont abordé le problème de la formation de CNN très profonds avec des centaines de couches tout en obtenant une précision convaincante avec plus de 92,2 % de précision dans le top 5 sur ImageNet. Les réseaux très profonds sont difficiles à former car les gradients disparaissent après avoir propagé trop de couches. ResNet incorpore des connexions de saut (\raccourci") entre deux couches non adjacentes afin que les gradients puissent être plus facilement récupérés sur les couches précédentes pendant la formation. ResNet-50 a 53 couches convolutives, une normalisation de mise en commun maximale et une couche FC, où chaque convolution est suivie par une couche de normalisation de Bach. Les tailles de noyau utilisées dans ResNet sont 7 × 7,

3×3 et 1×1. Le nombre de poids dans ResNet-50 est d'environ 26 millions, le nombre d'opérations est d'environ 7,7 milliards. **[32]**

II.7 Choix des paramètres

Les CNN utilisent plus de paramètres que les MLP standard. Bien que les règles habituelles pour les taux d'apprentissage et les constantes de régularisation s'appliquent toujours, il est nécessaire de Considérez le concept de nombre de filtres, de forme et de forme maximale collection. **[18]**

II.7.1 Nombre de filtres

Comme la taille des images intermédiaires diminue avec la profondeur de traitement, les couches plus proches de l'entrée ont tendance à avoir moins de filtres, tandis que les couches plus proches de la sortie peuvent en avoir plus. Afin d'équilibrer le calcul de chaque couche, le produit du nombre d'entités et du nombre de pixels traités est généralement choisi pour être approximativement constant d'une couche à l'autre. Afin de préserver les informations d'entrée, il est nécessaire de faire en sorte que le nombre de sorties intermédiaires (le nombre d'images intermédiaires multiplié par le nombre d'emplacements de pixels) augmente (au sens large) d'une couche à l'autre. Le nombre d'images intermédiaires contrôle directement la puissance du système, en fonction du nombre d'exemples disponibles et de la complexité de traitement.

II.7.2 Forme du filtres

Les formes de filtre dans la littérature varient considérablement. Ils sont généralement choisis en fonction de l'ensemble de données. Les meilleurs résultats pour les images MNIST (28x28) se situent généralement dans la plage 5x5 de la première couche, tandis que les ensembles de données d'images naturelles (souvent des centaines de pixels dans chaque dimension) ont tendance à utiliser des filtres de première couche 12x12 plus grands, Même 15x15. Le défi consiste donc à trouver le bon niveau de granularité pour créer La juste échelle d'abstraction et s'adaptant à chaque situation.

II.7.3 Form du Max Pooling

La valeur typique est 2x2. Une très grande taille d'entrée peut justifier une mise en commun 4x4 dans la première couche. Cependant, le choix d'une forme plus grande peut réduire considérablement la taille du signal et entraîner la perte d'une trop grande quantité d'informations.

II.8 Conclusion

L'apprentissage profond est devenu une méthode dominante dans une variété de tâches complexes, telles que la classification des images et la découverte de choses à travers les réseaux neuronaux, qui ont montré un grand succès et d'excellents résultats dans le domaine de la classification des images, et cela nous a amenés à utiliser cette technologie dans notre travail. Dans ce chapitre, nous avons donné un aperçu des concepts d'apprentissage profond, axés sur les réseaux de dérivation neuronale, et expliqué l'infrastructure du réseau, ses composants et la façon de classer les images. CNN s'appuie sur le processus de torsion. La structure de ce type de réseau est basée sur deux parties, la partie torsadée et la partie connectée entière. La partie torsion dépend principalement des séquences de trois processus de base, à savoir, torsion, correction ReLu, et l'assemblage. Le segment entier se termine par la couche soft max, qui donne la probabilité de l'image d'entrée appartenant à chacune des catégories possibles. Dans le chapitre suivant, nous allons présenter notre modèle, puis expliquer les résultats obtenus lors des phases d'apprentissage et de test, les discuter et les critiquer.

Chapitre III :
Interprétation des résultats

III.1 Introduction

Dans ce chapitre, nous allons nous concentrer sur notre conception et montrer comment implémenter notre classeur et créer notre modèle. Ensuite, nous évaluerons et créerons deux modèles par conséquent, pour différentes architectures, nous appliquerons ces modèles basés sur des images veine de la main. Pour cela nous utiliserons le langage de programmation Python des bibliothèques Tensorflow et Keras pour la formation, la classification et le but pour améliorer les performances du modèle, nous utiliserons quelques techniques simples et efficace .

III.2 Logiciels et librairies utilisés dans interprétation des résultats et configuration de hardware

Donc ce qui suit nous démons Logiciels et librairies utilisés dans interprétation des résultats et configuration de hardware:

III.2.1 TensorFlow

TensorFlow est une bibliothèque de logiciels open source pour le calcul numérique haute performance. Son architecture flexible permet une variété de plates-formes (CPU, GPU, TPU), des ordinateurs de bureau aux clusters de serveurs en passant par les appareils mobiles. Développé à l'origine par des chercheurs et des ingénieurs de l'équipe Google Brain au sein de l'organisation AI de Google, il s'appuie sur l'apprentissage automatique et l'apprentissage en profondeur. **[33]**

III.2.2 Keras

Keras est une API de réseau neuronal de haut niveau, écrite en Python, capable de fonctionner sur TensorFlow ou Theano. Il a été développé en mettant l'accent sur l'expérimentation rapide. Pouvoir passer de l'idée au résultat dans les plus brefs délais est la clé d'une bonne recherche. Il a été développé dans le cadre des travaux de recherche du projet ONEIROS (Open Neuroelectronic Intelligent Robot Operating System), dont l'auteur principal et mainteneur est l'ingénieur Google François Chollet.

En 2017, l'équipe TensorFlow de Google a décidé de prendre en charge Keras dans la bibliothèque principale TensorFlow. Chollet a expliqué que Keras a été conçu comme une interface plutôt qu'un framework apprentissage de bout en bout. Il dispose d'un ensemble

d'abstractions de niveau supérieur et plus intuitif qui facilite la configuration de réseaux de neurones indépendants des bibliothèques informatiques principales. Microsoft travaille également sur l'ajout d'un backend CNTK à Keras**. [34]**

III.2.3 CUDA

CUDA est une plate-forme informatique parallèle et un modèle de programmation développé par NVIDIA pour l'informatique à usage général sur les unités de traitement graphique (GPU). Avec CUDA, les développeurs peuvent exploiter la puissance des GPU pour accélérer considérablement les applications informatiques. Dans les applications accélérées par GPU, la partie séquentielle de la charge de travail fonctionne sur le processeur, optimisé pour les performances à un seul thread, et la section

Les applications à forte intensité de calcul s'exécutent en parallèle sur des milliers de cœurs de GPU. Lors de l'utilisation de CUDA, les développeurs programment dans des langages populaires tels que C, C++, Fortran, Python et MATLAB, et via quelques extensions de formulaire de mots clés de base. **[35]**

III.2.4 Scikit-learn

Scikit-learn est une bibliothèque Python gratuite dédiée au machine learning. Il est développé par de nombreux contributeurs, notamment dans le milieu universitaire par l'Institut Français.Dans l'enseignement supérieur et la recherche, comme Inria et Télécom ParisTech.Il comprend spécifiquement des méthodes d'estimation de forêts aléatoires, de régression logistique.Algorithmes de classification et machines à vecteurs de support. Il est conçu pour se coordonner avec d'autres bibliothèques Python gratuites, notamment NumPy et SciPy.

III.2.5 Python

Python est un langage de programmation interprété multi-paradigmes. Il favorise la programmation impérative structurée et orientée objet. Il dispose d'un typage dynamique puissant, d'une gestion automatique de la mémoire avec récupération de place et d'un système de gestion des exceptions ; il est donc similaire à Perl, Ruby, Scheme, Smalltalk et Tcl.

Le langage Python est placé sous une licence libre proche de la licence BSD et peut fonctionner sur la plupart des plates-formes informatiques, des supercalculateurs aux

ordinateurs centraux, de Windows à Unix, en passant par Linux et MacOS, en passant par Java et même .NET. Il est conçu pour optimiser la productivité des programmeurs en fournissant des outils avancés et une syntaxe facile à utiliser. Il est également apprécié des éducateurs, qui trouvent que sa syntaxe est clairement séparée de la mécanique de bas niveau, permettant une initiation plus facile aux concepts de base de la programmation.

Python est un langage qui peut être utilisé dans de nombreuses situations, et comme il existe des bibliothèques spécialisées pour chaque processus, il peut être adapté à tout type d'utilisation. Cependant, il est surtout utilisé comme langage de script pour automatiser des tâches simples mais fastidieuses, comme récupérer la météo sur Internet ou s'intégrer dans des logiciels de conception assistée par ordinateur pour automatiser certaines séquences répétitives de scripts d'action. Il peut également être utilisé comme langage de prototypage jusqu'à ce qu'une application fonctionnelle soit requise, qui peut ensuite être optimisée à l'aide de langages de niveau inférieur. Il est particulièrement répandu dans la communauté scientifique et possède de nombreuses extensions pour des applications numériques. **[36]**

III.2.6 Configuration de hardware

La configuration du matériel utilisé dans notre implémentation est :

- Un PC portable Dell i5 CPU 2 GHZ
- Carte graphique Nvidia GeForce GT525M
- RAM de taille 8 GO
- Disque dur de taille 500 GO

III.3 La base de donnée utilisée

De nombreuses bases de données contiennent des informations permettant d'évaluer les systèmes de reconnaissance des veines de la main. Cependant, ces bases de données sont généralement conçues pour répondre aux besoins d'un algorithme de reconnaissance spécifique.

Notre base de données se compose de 100 images infrarouges des veines de la main de 10 personnes dans 10 classes, chacune ayant pris 10 images différentes. Il y a 80 images d'entraînement et 20 images un test.

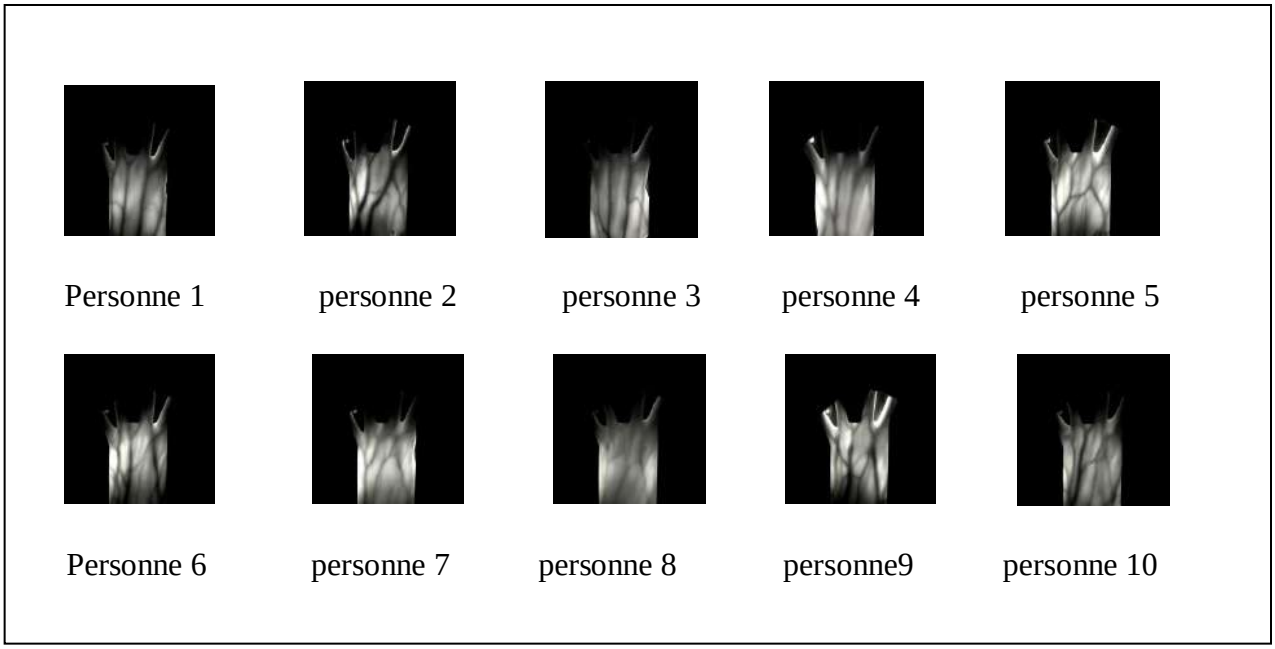


Figure III-1: Images de veine de la main -10-classes

III.4 Architecture des réseaux

Dans nos expériences, nous avons créé deux architectures de modèles différentes, où la base de données d'images a été appliquée aux deux modèles, et pour chaque modèle, nous avons évalué le nombre d'époques. Nous décrivons ci-dessous la structure des deux modèles et les résultats obtenus à chaque fois :

III.4.1 Réseau neuronal convolutif à 4 couches

Le premier modèle que nous proposons consiste en quatre couches convolutives, deux couches de mise en commun maximale et deux couches entièrement connectées.

La taille de l'image d'entrée est de 50*50, et chaque couche convolutive se compose de plusieurs filtres (32 pour les 2 premières couches), chacun de taille 3*3, chaque passage à travers la couche de déconvolution utilise une fonction d'activation ReLU qui force le neurone à renvoyer une valeur positive.

Maxpooling est ensuite appliqué pour réduire la taille de l'image. à la fin de la seconde couches convolutives, nous utilisons Dropout, qui est une technique de régularisation réduire le sur ajustement dans les réseaux de neurones.

Dans les troisième et quatrième couches, nous avons modifié certains paramètres, tels que le filtre devient 64 au lieu de 32, la fonction d'activation reste la même (ReLu) et Il en va de même pour Maxpooling, puis nous appelons à nouveau la fonction Dropout.

Après les couches de convolution et de l'agglomération, notre partie classification se compose de plusieurs couches entièrement connectées. Cependant, ces couches ne peuvent accepter que des données unidimensionnelles. Pour convertir nos données 3D en 1D, nous utiliserons la fonction Flatten, qui organise essentiellement notre volume 3D en un vecteur 1D.

Les dernières couches du réseau neuronal convolutif sont complètement connecté. ils sont avec niveau supérieur.

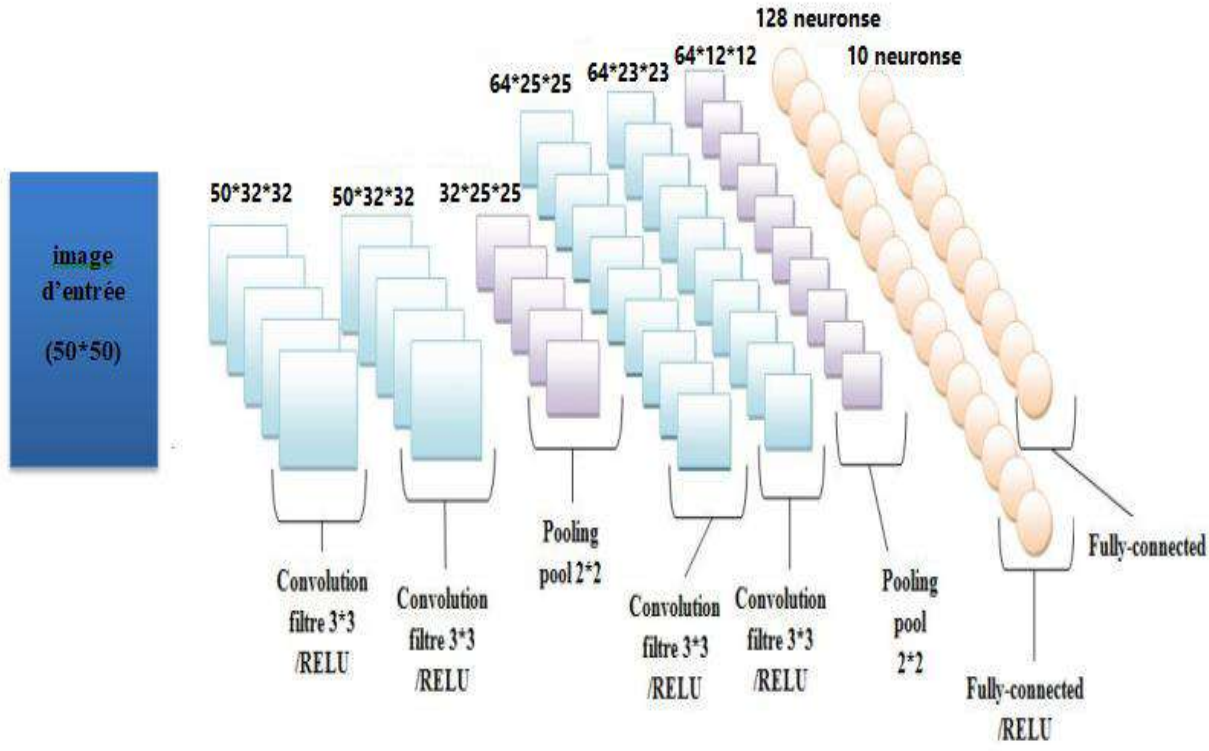


Figure III-2: Architecteur de modèle 1 (4 couches)

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 32)	896
conv2d_1 (Conv2D)	(None, 46, 46, 32)	9248
max_pooling2d (MaxPooling2D)	(None, 23, 23, 32)	0
conv2d_2 (Conv2D)	(None, 21, 21, 64)	18496
conv2d_3 (Conv2D)	(None, 19, 19, 64)	36928
max_pooling2d_1 (MaxPooling2D)	(None, 9, 9, 64)	0
flatten (Flatten)	(None, 5184)	0
dense (Dense)	(None, 128)	663680
dense_1 (Dense)	(None, 10)	1290

Total params: 730,538
Trainable params: 730,538
Non-trainable params: 0

Tableau III-1: Configuration de modèle 1 (4 couches)

III.4.2 Réseau neuronal convolutif à 6 couches

Sur le deuxième modèle, nous avons modifié le nombre de couches convolutives et max-pooling, et maintenant notre architecture se compose de six couches convolutives et trois couches max-pooling et les deux dernières couches entièrement connectées.

Pour le premier modèle, la taille de l'image d'entrée est de 50*50, et chaque couche convolutive se compose de plusieurs filtres (32 pour les 2 premières couches, 64 pour les 2 dernières couches et 128 pour les 2 dernières couches), avec un size of six La taille de chaque filtre des couches convolutives est constante (elle est de 3*3), la fonction d'activation utilisée par toutes les couches convolutives est ReLU, et à la fin des deux couches convolutives, on utilise maxpooling pour réduire l'image Taille.

Après chaque couche de convolution et de clustering, nous utiliserons Dropout pour réduire le sur ajustement d'un pourcentage de 20 % à chaque fois, et notre partie de classification se compose de deux couches entièrement connectées. Cependant, ces couches ne peuvent accepter que des données unidimensionnelles. Nous allons convertir les données en 1D via la fonction Flatten.

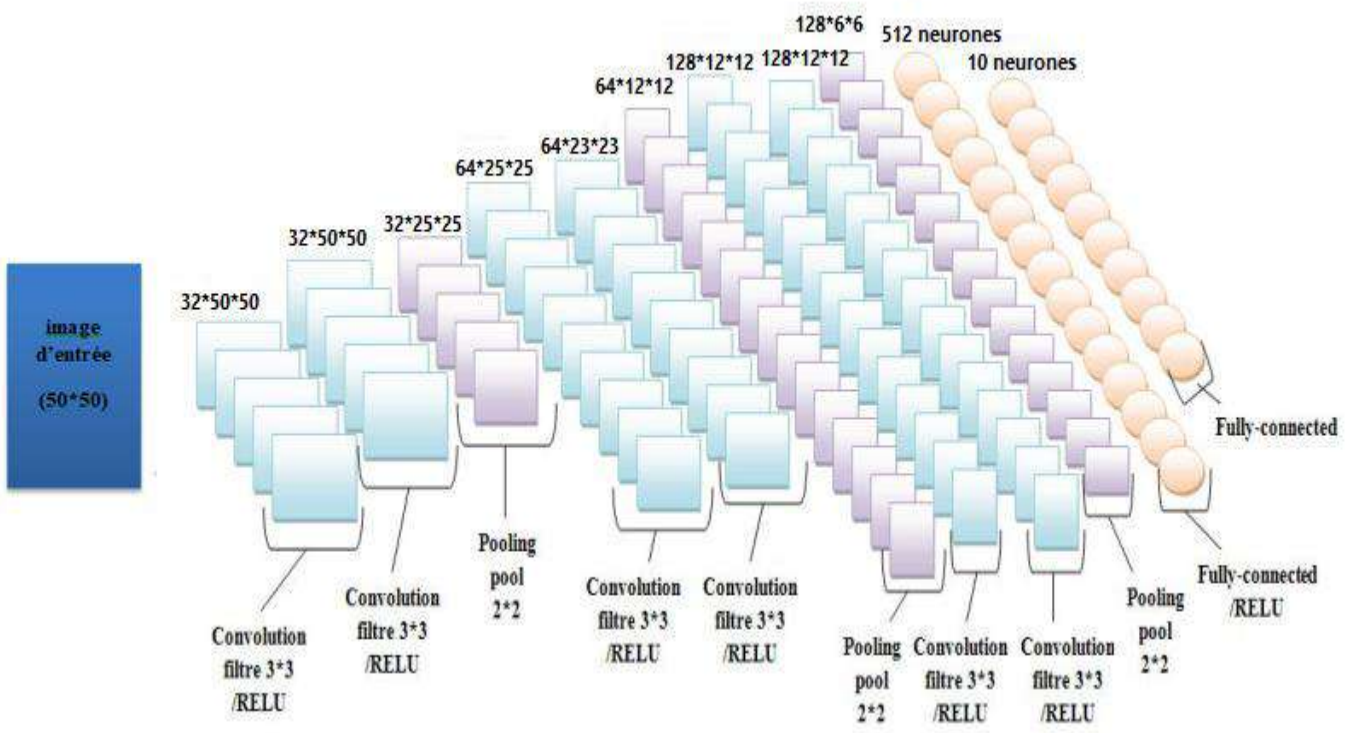


Figure III-3: Architecture de modèle 2 (6 couches)

Layer (type)	Output Shape	Param #
conv2d_4 (Conv2D)	(None, 48, 48, 32)	896
conv2d_5 (Conv2D)	(None, 46, 46, 32)	9248
max_pooling2d_2 (MaxPooling 2D)	(None, 23, 23, 32)	0
conv2d_6 (Conv2D)	(None, 21, 21, 64)	18496
conv2d_7 (Conv2D)	(None, 19, 19, 64)	36928
max_pooling2d_3 (MaxPooling 2D)	(None, 9, 9, 64)	0
conv2d_8 (Conv2D)	(None, 7, 7, 128)	73856
conv2d_9 (Conv2D)	(None, 5, 5, 128)	147584
max_pooling2d_4 (MaxPooling 2D)	(None, 2, 2, 128)	0
flatten_1 (Flatten)	(None, 512)	0
dense_2 (Dense)	(None, 512)	262656
dense_3 (Dense)	(None, 10)	5130

Tableau III-2: Configuration de modèle 2 (6 couches)

III.5 Résultats obtenus et discussion

Pour montrer les résultats de ces deux modèles, nous donnons dans l'exemple ci-dessous les résultats en termes de précision et de matrice d'erreur et de confusion nombre d'époques (une époque décrit le nombre de fois que l'algorithme parcourt l'ensemble données) pour chaque modèle.

A. Résultats obtenus pour le modèle 1 (4 couches)

A.1. Nombre d'époques = 10

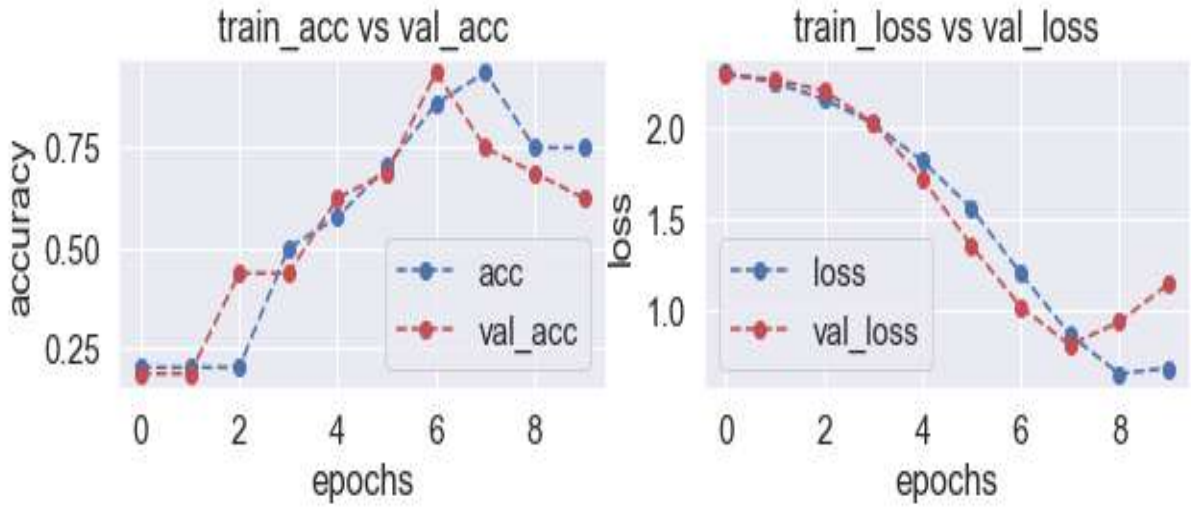


Figure III-4: Précision et erreur pour le modèle 1 (10 époques)

À partir de la Figure III-4, La précision de l'apprentissage et de la validation augmente à mesure que le nombre d'époques augmente, jusqu'au nombre d'époques où la validation commence à diminuer (6).

De même, les erreurs d'apprentissage et de validation diminuent avec le nombre d'époques jusqu'au nombre d'époques (7) où la validation commence à augmenter.

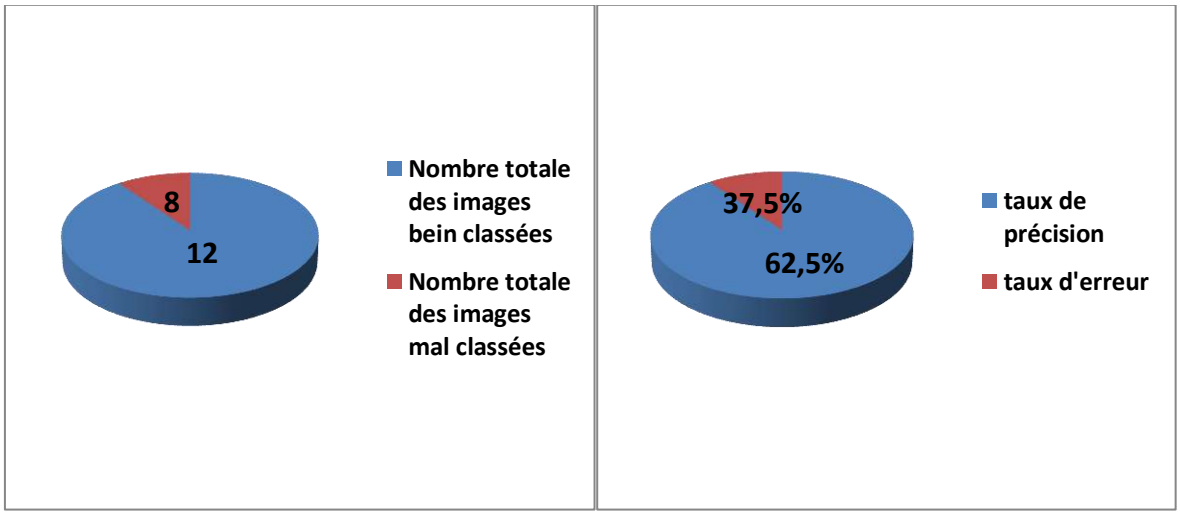


Figure III-5: Nombre total des images mal et bien classées et taux d'erreur et de précision

Sur la Figure III-5, nous remarquons que le nombre total d'images mal classées est de 8 images, le taux d'erreur est de 37,5% et le nombre total d'images bien classées est de 12, donc Le taux de précision est de 62,5 %.

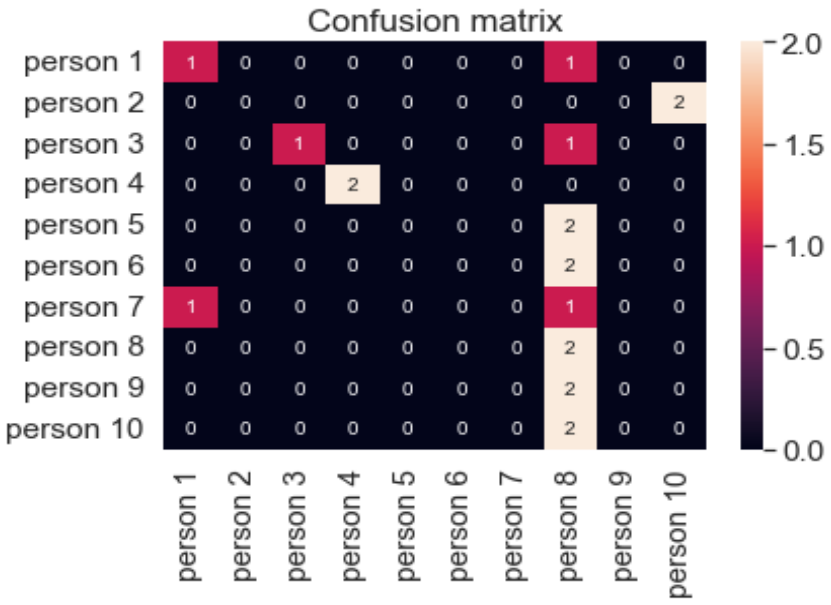


Figure III-6: Matrice de confusion pour le modèle1 (10 époques)

A.2. Nombre d'époques = 15

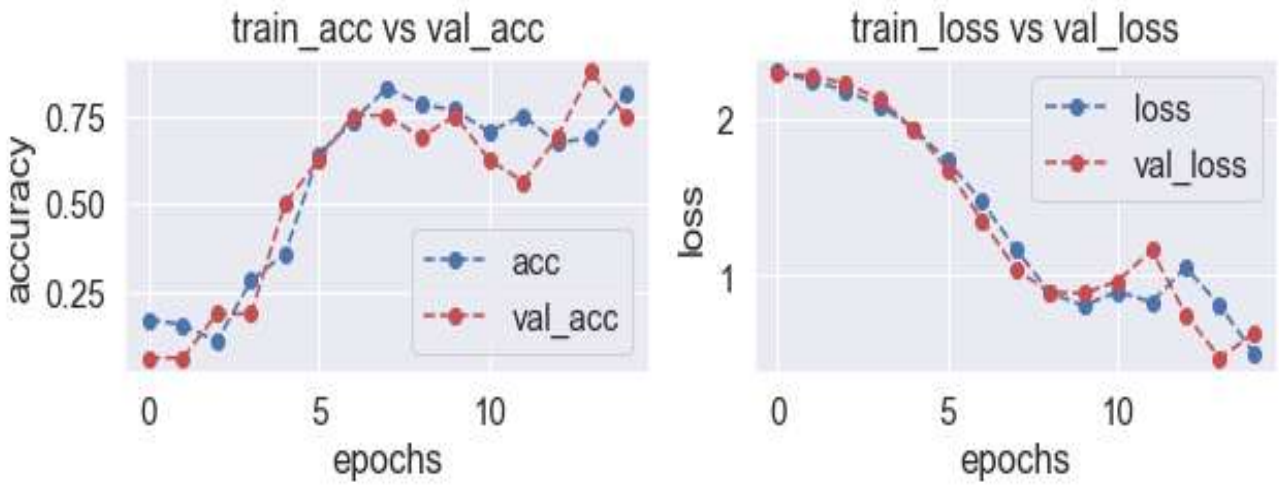


Figure III-7: Précision et erreur pour le modèle 1 (15 époques)

À partir de la Figure III-7, La précision de l'apprentissage et de la validation augmente dans l'intervalle du nombre d'époques [0-7] et est relativement stable dans la période [6-9] et dans la période [9-11] la vérification diminue pour revenir au plus haut dans la période [11 - 13] pour s'installer au niveau de (0,75).

De même, les erreurs d'apprentissage et de validation diminue dans l'intervalle de nombre d'époques [0-8], puis augmente dans la période [9-11], puis redescend pour rester à (0,25).

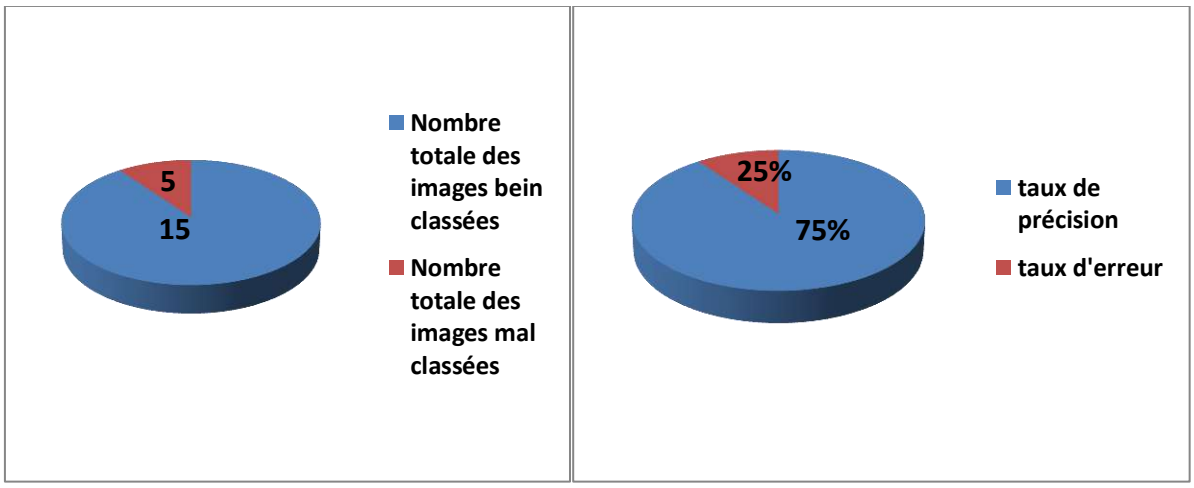


Figure III-8: Nombre total des images mal et bien classées et taux d'erreur et de précision

Sur la Figure III-8, nous remarquons que le nombre total d'images mal classées est de 5 images, le taux d'erreur est de 25% et le nombre total d'images bien classées est de 15, donc Le taux de précision est de 75 %

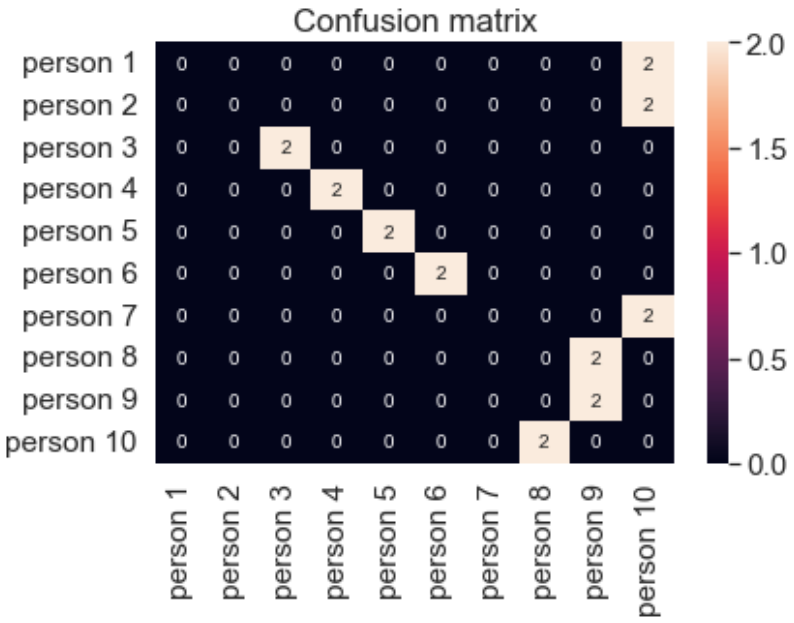


Figure III-9: Matrice de confusion pour le modèle1 (15 époques)

A.3. Nombre d'époques = 20

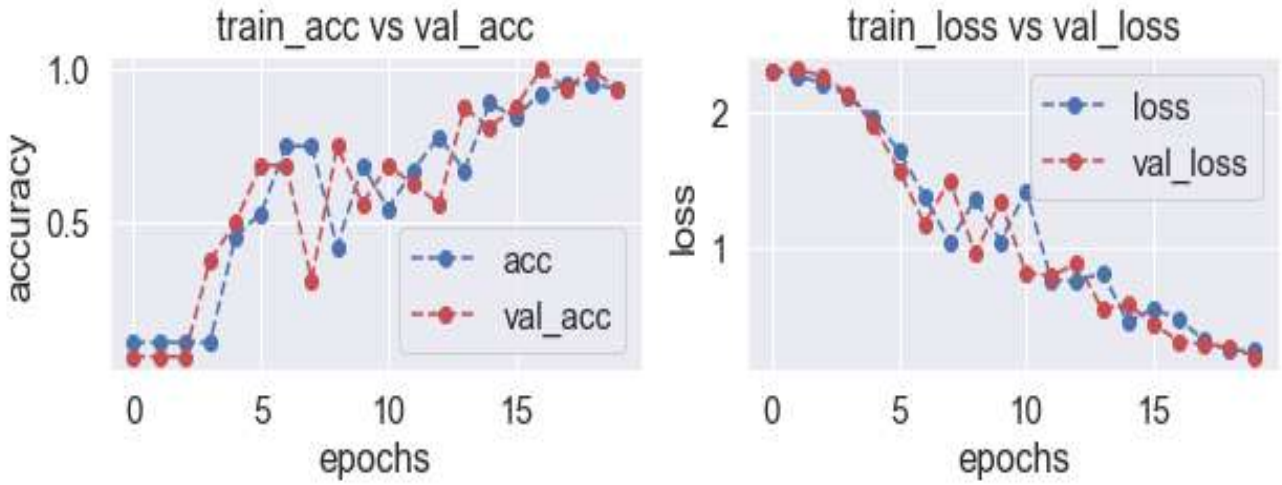


Figure III-10: Précision et erreur pour le modèle 1 (20 époques)

À partir de la Figure III-10, La précision de l'apprentissage et de la validation augmente dans l'intervalle du nombre d'époques [2-6], et dans la période [6-12], la validation sont à une diminution et une augmentation relatives, et la validation augmente dans la période [12-20] pour s'installer au niveau de (0,9).

De même, les erreurs d'apprentissage et de validation diminue dans l'intervalle de nombre d'époques [0-6], en suite dans la période [6-12], l'erreur de validation augmentera et diminuera et diminuera dans la période [12-20] pour se stabiliser au niveau (0.1).

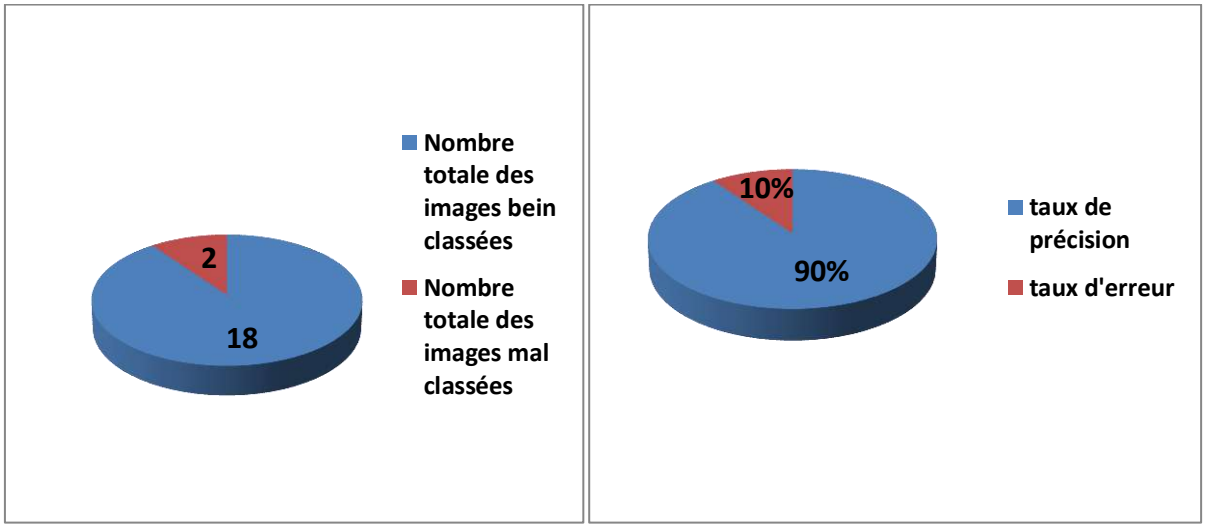


Figure III-11: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 1 (20 époques)

Sur la Figure III-11, nous remarquons que le nombre total d'images mal classées est de 2 images, le taux d'erreur est de 10% et le nombre total d'images bien classées est de 18, donc Le taux de précision est de 90 %

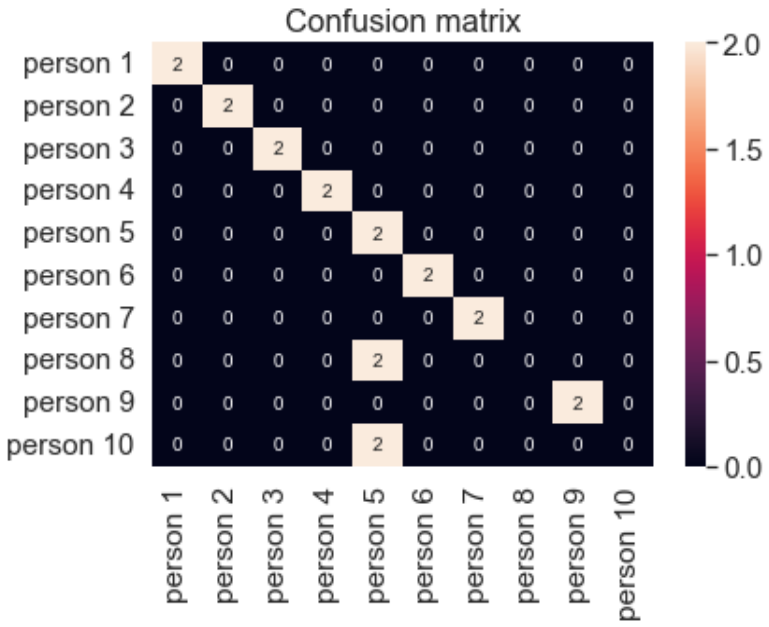


Figure III-12: Matrice de confusion pour le modèle1 (20 époques)

A.4. Nombre d'époques = 30

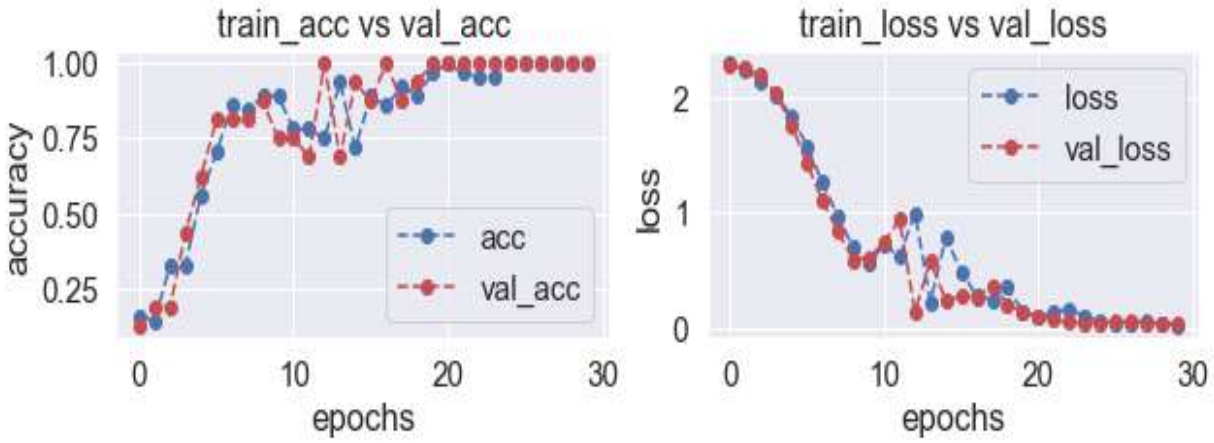


Figure III-13: Précision et erreur pour le modèle 1 (30 époques)

À partir de la Figure III-13, La précision de l'apprentissage et de la validation augmente dans l'intervalle du nombre d'époques [0-8], et dans la période [8-18], la validation sont à une diminution et une augmentation relatives, mais dans l'intervalle [18-30] nous avons remarqué que la précision de l'apprentissage se stabilise au niveau (1) et la validation reste stable dans le même niveau.

De même, les erreurs d'apprentissage et de validation diminue dans l'intervalle de nombre d'époques [0-8], en suite dans la période [8-18], l'erreur de validation augmentera et diminuera, nous avons remarqué que la précision de l'apprentissage se stabilise au niveau (0) et celle de la validation reste stable aussi dans le même niveau.

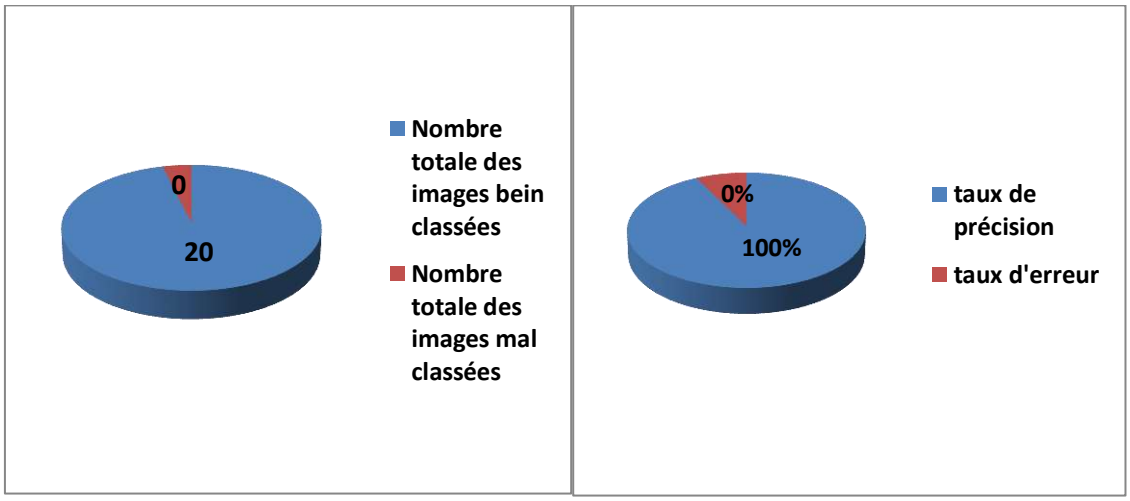


Figure III-14: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 1 (30 époques)

Sur la Figure III-14, nous remarquons que le nombre total d'images mal classées est de 0 images, le taux d'erreur est de 0% et le nombre total d'images bien classées est de 20, donc Le taux de précision est de 100 %

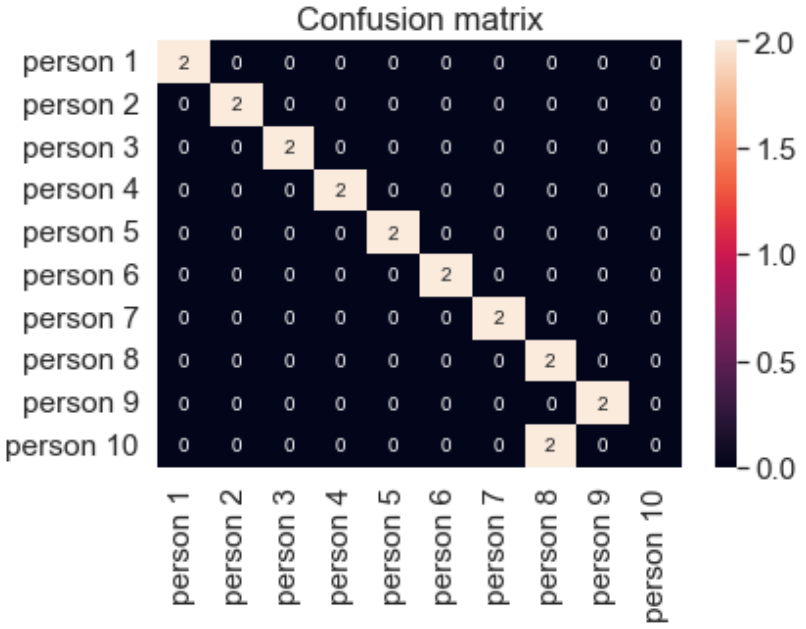


Figure III-15: Matrice de confusion pour le modèle1 (30 époques)

B. Résultats obtenus pour le modèle 2 (6 couches)

B.1. Nombre d'époques = 10

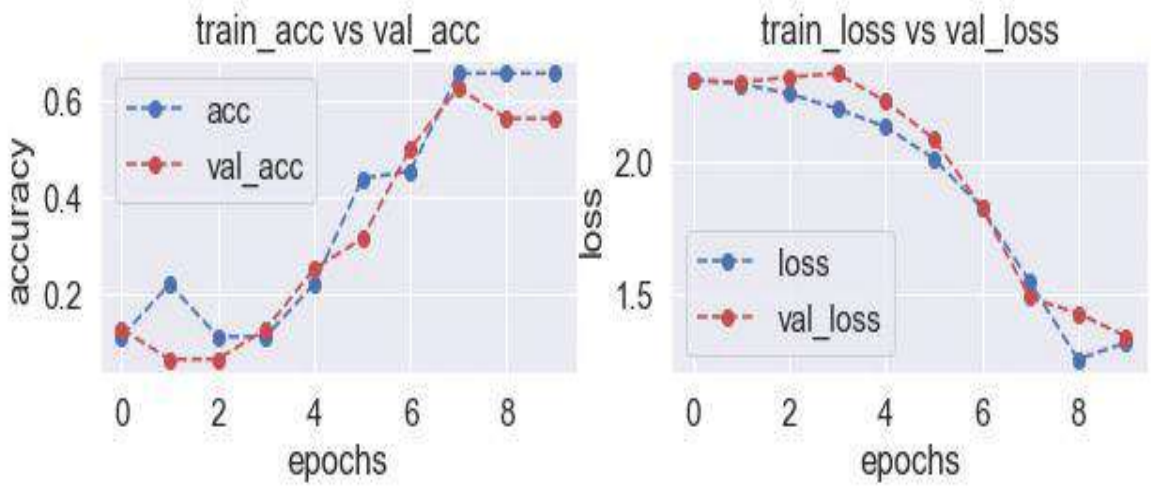


Figure III-16: Précision et erreur pour le modèle 2 (10 époques)

À partir de la Figure III-16, La précision de l'apprentissage et de la validation augmente à mesure que le nombre d'époques augmente, jusqu'au nombre d'époques où la validation commence à diminuer (7).

De même, les erreurs d'apprentissage et de validation diminuent avec le nombre d'époques jusqu'au nombre d'époques (9).

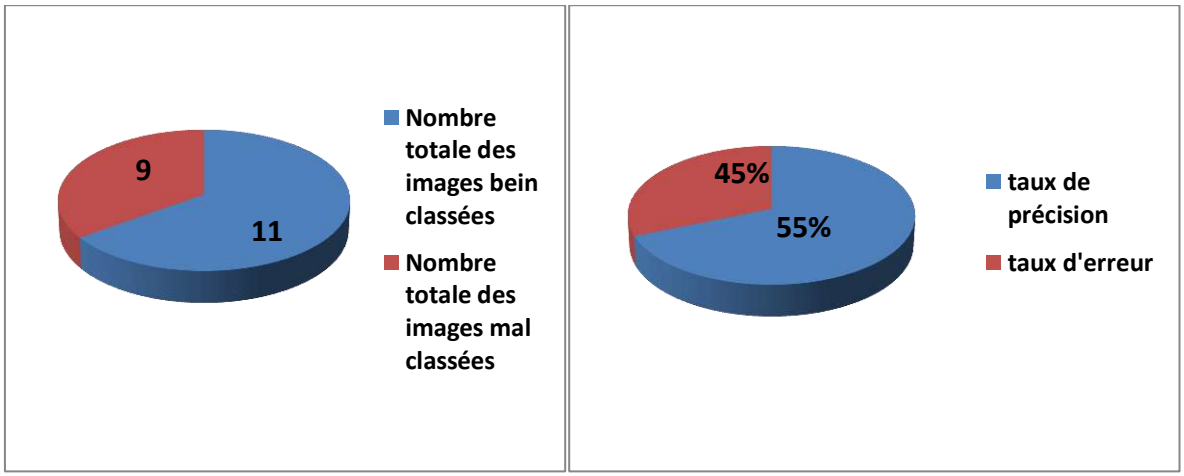


Figure III-17: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 2 (10 époques)

Sur la Figure III-17, nous remarquons que le nombre total d'images mal classées est de 9 images, le taux d'erreur est de 44% et le nombre total d'images bien classées est de 11, donc Le taux de précision est de 56 %

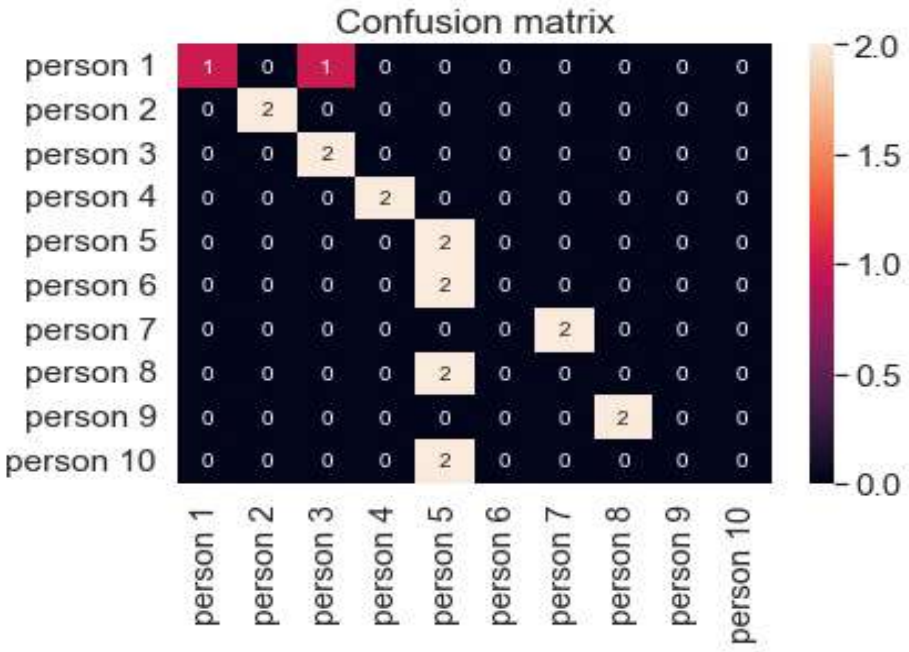


Figure III-18: Matrice de confusion pour le modèle 2 (10 époques)

B.2. Nombre d'époques = 15

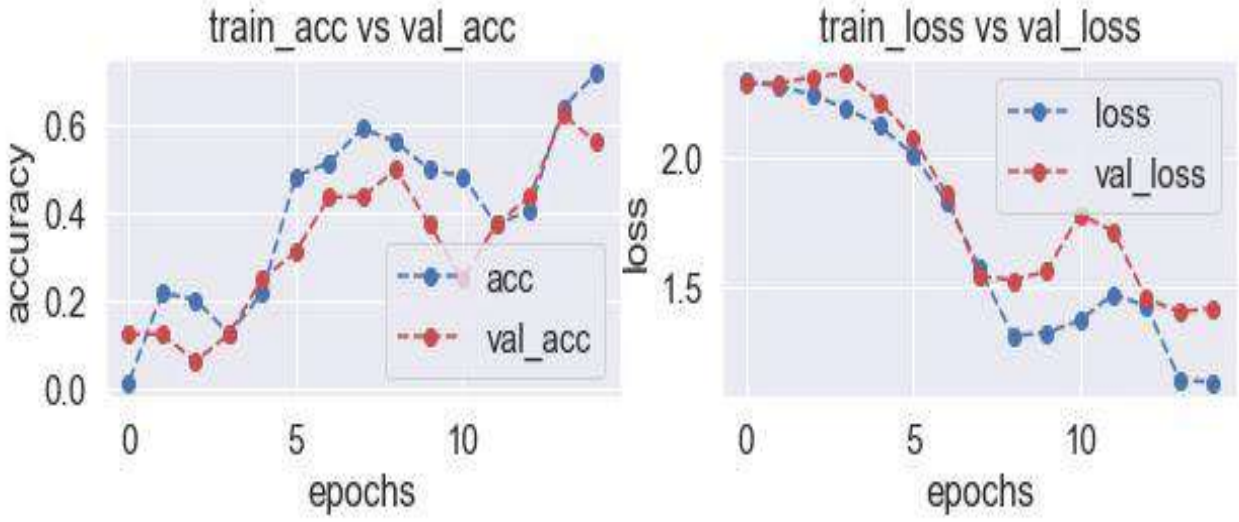


Figure III-19: Précision et erreur pour le modèle 2 (15 époques)

À partir de la Figure III-19, La précision de l'apprentissage et de la validation augmente dans l'intervalle du nombre d'époques [2-8] et est relativement stable dans la période [6-9] et dans la période [9-11] la vérification diminue pour revenir au plus haut dans la période [11 - 13] pour s'installer au niveau de (0,62).

De même, les erreurs d'apprentissage et de validation diminue dans l'intervalle de nombre d'époques [3-8], puis augmente dans la période [9-10], puis redescend pour rester à (0,37).

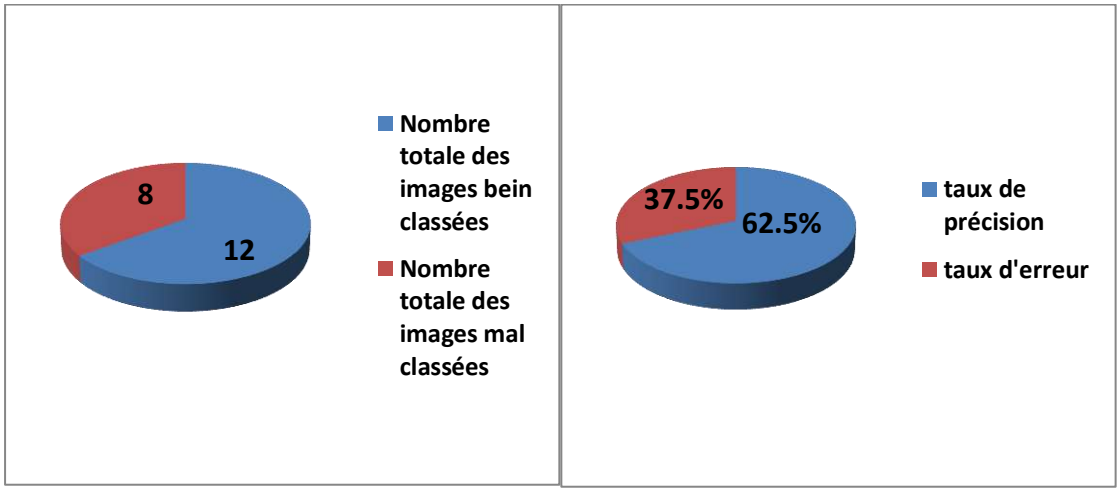


Figure III-20: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 2 (15 époques)

Sur la Figure III-20, nous remarquons que le nombre total d'images mal classées est de 8 images, le taux d'erreur est de 38% et le nombre total d'images bien classées est de 12, donc Le taux de précision est de 62 %

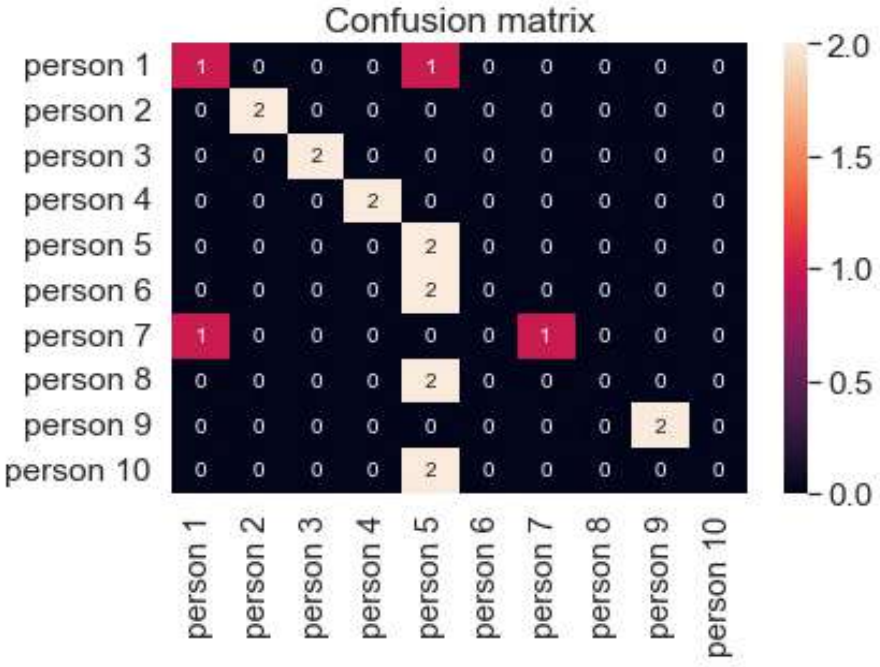


Figure III-21: Matrice de confusion pour le modèle 2 (15 époques)

B.3. Nombre d'époques = 20

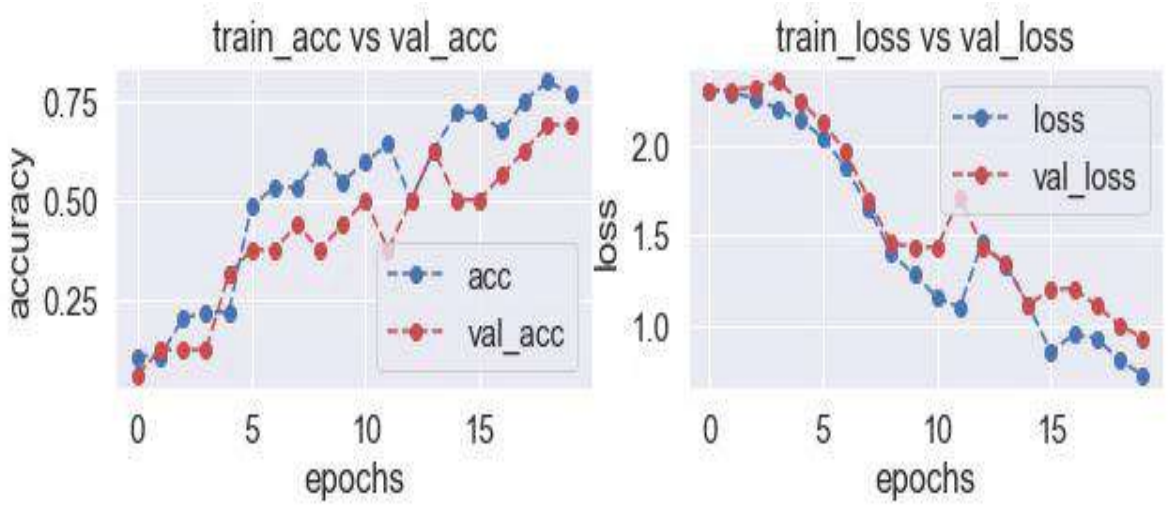


Figure III-22: Précision et erreur pour le modèle 2 (20 époques)

À partir de la Figure III-22, les précisions de formation et de validation augmentent à mesure que le nombre d'époques augmente, jusqu'au nombre d'époques où la validation commence à diminuer (10).

De même, les erreurs d'apprentissage et de validation diminuent avec le nombre d'époques jusqu'au nombre d'époques (10) où la validation commence à augmenter.

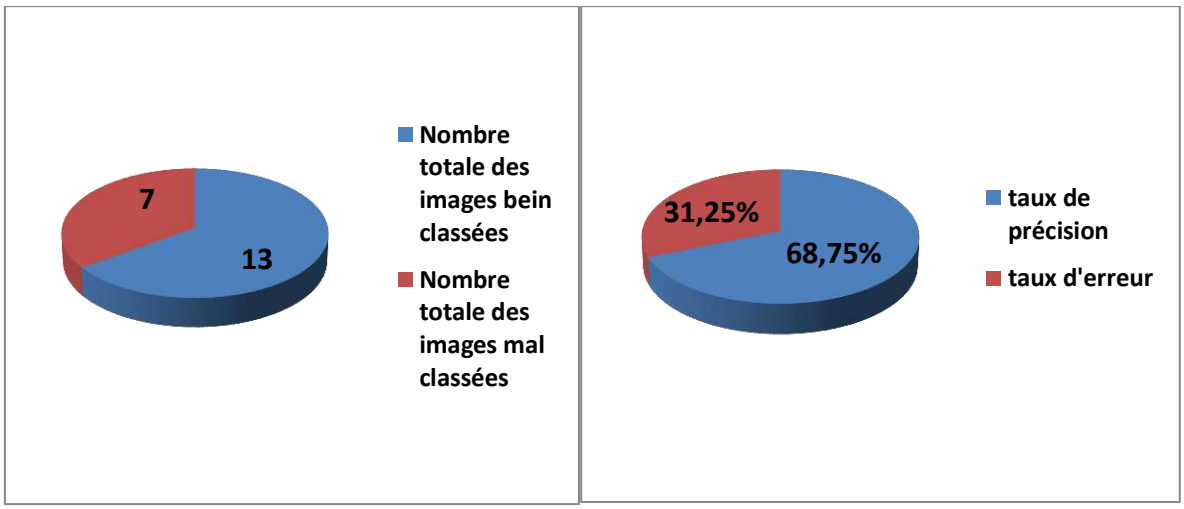


Figure III-23: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 2 (20 époques)

Sur la Figure III-23, nous remarquons que le nombre total d'images mal classées est de 7 images, le taux d'erreur est de 31.25% et le nombre total d'images bien classées est de 13, donc Le taux de précision est de 68.75 %

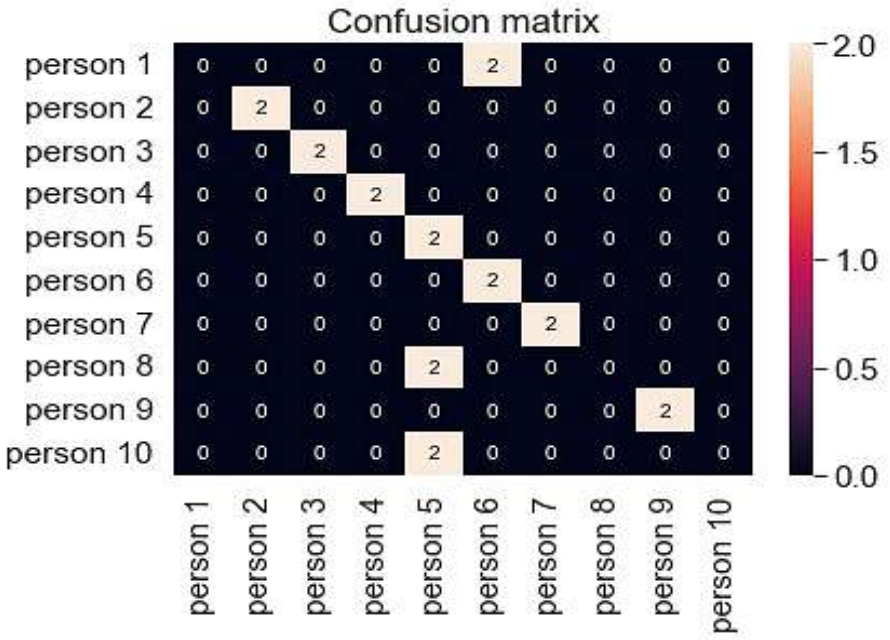


Figure III-24: Matrice de confusion pour le modèle 2 (20 époques)

B.4. Nombre d'époques = 30

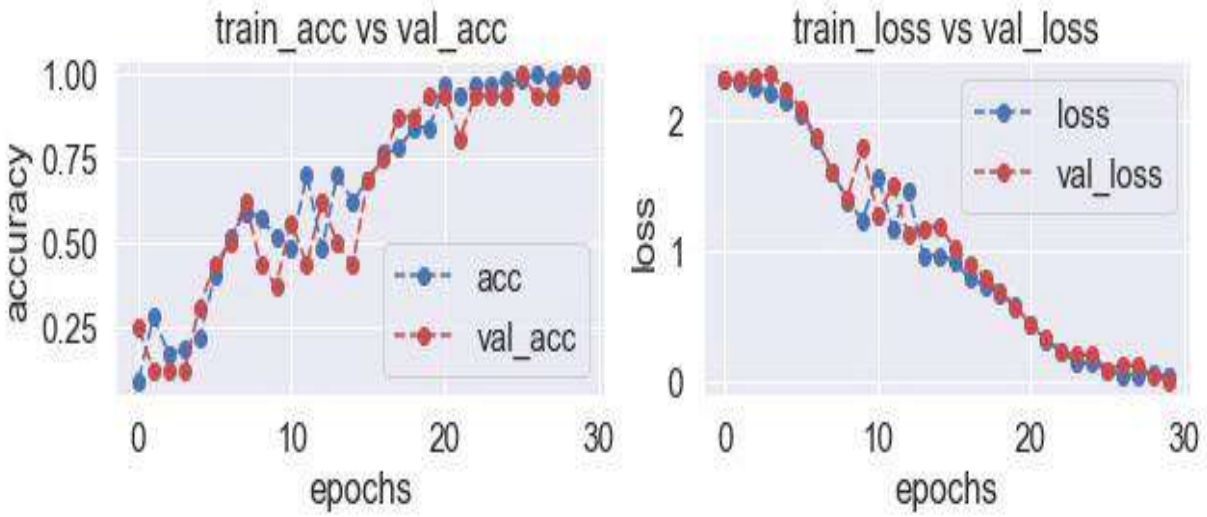


Figure III-25: Précision et erreur pour le modèle 2 (30 époques)

À partir de la Figure III-25, La précision de l'apprentissage et de la validation augmente dans l'intervalle du nombre d'époques [0-7], et dans la période [7-14], la validation sont à une diminution et une augmentation relatives, mais dans l'intervalle [18-30] nous avons remarqué que la précision de l'apprentissage se stabilise au niveau (1) et la validation reste stable dans le même niveau.

De même, les erreurs d'apprentissage et de validation diminue dans l'intervalle de nombre d'époques [0-8], en suite dans la période [8-12], l'erreur de validation augmentera et diminuera, nous avons remarqué que la précision de l'apprentissage se stabilise au niveau (0) et celle de la validation reste stable aussi dans le même niveau.

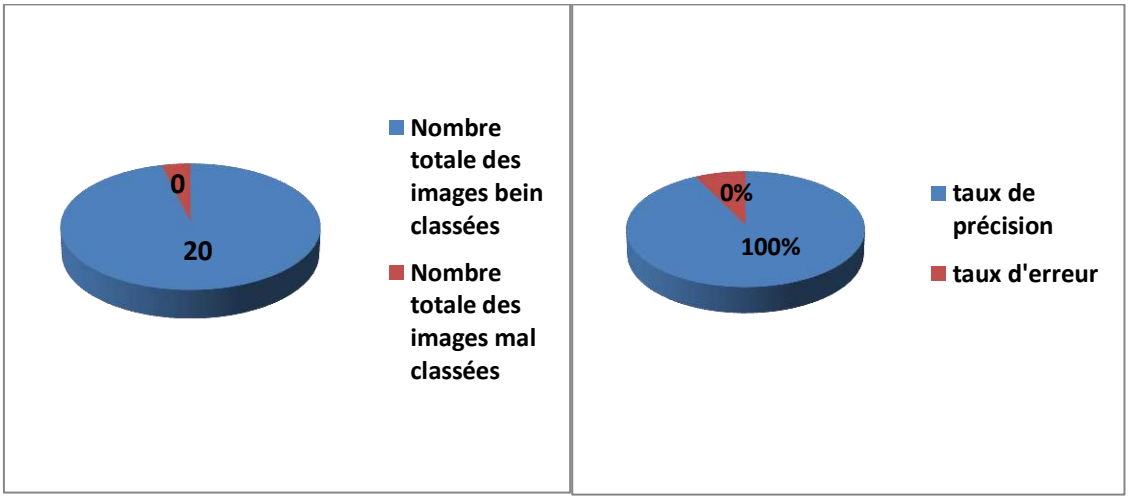


Figure III-26: Nombre total des images mal et bien classées et taux d'erreur et de précision de modèle 2 (30 époques)

Sur la Figure III-26, nous remarquons que le nombre total d'images mal classées est de 0 images, le taux d'erreur est de 0% et le nombre total d'images bien classées est de 20, donc Le taux de précision est de 100 %

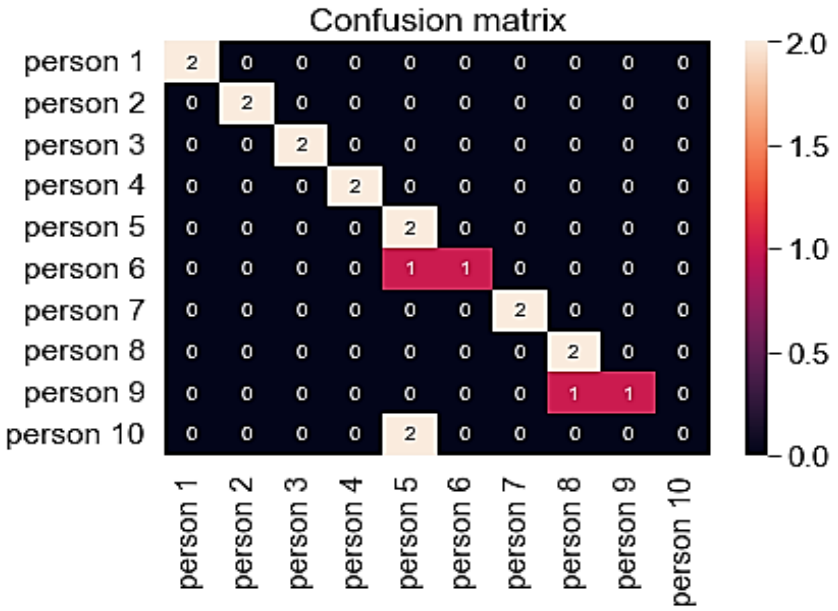


Figure III-27: Matrice de confusion pour le modèle 2 (30 époques)

C. Tableau de comparaison des résultats

Le tableau ci-dessous montre les différents résultats obtenus sur les deux modèles :

	Architecture utilisé			Nombre d'époques	Précision	Erreur
	couche de convolution	couche de pooling	couche de fully connected			
Modèle 1 (4couches)	4	2	2	10	62.75 %	37.25%
				15	75%	25%
				20	90%	10%
				30	100%	0%
Modèle 2 (6couches)	6	3	2	10	55%	45%
				15	62.5%	37.5%
				20	68.75%	31.25%
				30	100%	0%

Tableau III-3: Tableau de comparaison des résultats

Le tableau montre l'architecture utilisée dans chaque modèle et le nombre d'époques. Les résultats obtenus sont exprimés en termes de précision d'apprentissage, de validation, d'essais et d'erreurs et de temps d'exécution final. Le temps d'exécution est trop cher. Cela se résume à la grande taille de la base, nécessitant l'utilisation d'un GPU au lieu d'un CPU. Le modèle 2 présente les meilleurs résultats. Le nombre d'époques et de couches convolutives reflète ces bons résultats, mais le temps d'exécution est très coûteux (à cause du nombre d'époques)

En général, les réseaux de neurones convolutifs larges et profonds produisent de bons résultats, et si les couches convolutives sont supprimées, les performances de notre réseau se détérioreront.

Alors que nous approfondissons le réseau et augmentons numéro d'époque. La base d'apprentissage est également un élément décisif du réseau de communication. Les neurones convolutifs nécessitent une grande base d'apprentissage pour atteindre meilleurs résultats.

III.6 Conclusion

Dans ce chapitre, nous avons introduit une méthode de classification basée sur les réseaux de neurones convolutifs, nous avons donc utilisé le nombre d'époques de deux modèles avec des structures différentes, puis expliqué les différents résultats en termes de précision et d'erreur. La comparaison des résultats montre que le nombre d'époques et la profondeur du réseau sont des facteurs importants pour de meilleurs résultats.

Conclusion générale

La classification des images est une tâche importante dans le domaine de la vision par ordinateur, de la reconnaissance des objets et de l'apprentissage automatique. Grâce au deep learning, l'avenir de l'intelligence artificielle évolue très rapidement.

Pour effectuer notre travail de classification, nous avons utilisé CNN parce qu'il a plusieurs caractéristiques avancées autres que les anciennes méthodes de classification Parmi eux, CNN combine les phases d'extraction et de classification. Dans le chapitre III, nous présentons les résultats que nous avons obtenus en appliquant ce réseau à un ensemble de données pour la veine manuelle pour différents échantillons de personnes et nous avons fait plusieurs tests différents en changeant chaque fois dans les paramètres de CNN afin d'obtenir un bon résultat et nous avons fait une comparaison de ces tests et dans ce dernier nous avons obtenu de bons résultats selon nos possibilités.

Dans l'évolution d'aujourd'hui dans divers domaines, la reconnaissance de la veine de cessation est un domaine avec un avenir prometteur et plus sûr, en particulier en développant des systèmes biométriques qui peuvent être utilisés dans des systèmes automatisés pour identifier les individus de manière unique et efficace et devenir une bonne alternative aux méthodes traditionnelles telles que les mots de passe et autres.

Bibliographie

a. Bibliographie :

[1] Babalola, Felix Olanrewaju, Önsen Toygar, and Yiltan BİTİRİM. "A Palm Vein Recognition Approach by Multiple Convolutional Neural Network Models." Avrupa Bilim ve Teknoloji Dergisi 29 (2021): 237-242.

[2] Chesner Desir, Classification Automatique d'Images, Application à l'Imagerie du Poumon Profond, Université de Rouen, 2013

[3] R. Gonzalez and E. Woods, "Digital Image Processing", 1992

[4] Bergounioux, Maïtine. Introduction au traitement mathématique des images-méthodes déterministes. Vol. 76. Springer, 2015.

[5] Idoui, Rabah, Ahmed Himeur, and SabrinaEncadreurBouatmane. Reconnaissance d'images par les réseaux de neurones convolutionnels (CNN). Diss. Université de Jijel, 2020.

[6] Habri, Amira, and Kamel Houari. "Un Système d'indexation et recherche d'images par le contenu basée sur la classification." (2011).

[7] Julliand, Thibaut, Vincent Nozick, and Hugues Talbot. "Image noise and digital image forensics." International Workshop on Digital Watermarking. Springer, Cham, 2015.

[9] Ali, Khabou, and Kerouane Samir. Filtrage numérique d'image par les transformées orthogonales. Diss. Université Mouloud Mammeri, 2012.

[10] Idoui, Rabah, Ahmed Himeur, and Sabrina Encadreur Bouatmane. Reconnaissance d'images par les réseaux de neurones convolutionnels (CNN). Diss. Université de Jijel, 2020.

[13] Yann LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998..

[15] Dif, Nassima. L'apprentissage profond pour le traitement des images. Diss. Djillali Liabes University, 2020.

[16] (Kaabi, Rabeb. Apprentissage profond et traitement d'images pour la détection de fumée. Diss. Toulon, 2020.

Bibliographie

[17] BENKADDOUR, M. K. (2020). Reconnaissance des formes et classification automatique : Application à l'identification biométrique (Doctoral dissertation).

[18] Chollet, Francois. Deep learning with Python. Simon and Schuster, 2021.

[19] Buyssens, Pierre. Fusion de différents modes de capture pour la reconnaissance du visage appliquée aux e_transactions. Diss. Université de Caen, 2011.

[20] Yufei Ma, « Hardware Acceleration of Deep Convolutional Neural Networks on FPGA »,Doctor of Philosophy thesis, Arizona State University, USA, 2018.
[21] Ilja Ivanovs, « Implementation Analysis of Convolutional Neural Networks on FPGAs »,Master Thesis, University of technology Eindhoven, 2019.

[22] O'SHEA Keiron, NASH Ryan. An introduction to convolutional neural networks. 2015.
[23] WANG Peng, XU Jiaming, XU Bo, LIU Chenglin, ZHANG Heng, WANG Fangyuan HAO Hongwei. Semantic clustering and convolutional neural networks for short text categorization. 2015

[24] Khouloud Dahmane, « Analyse d'images par méthode de Deep Learning appliquée au contexte routier en conditions météorologiques dégradées », thèse de doctorat, l'Université Clermont Auvergne, 2020.

[25] Krizhevsky, A., I. Sutskever and G. E. Hinton, « Imagenet classi_cation with deep convolutional neural networks », in Conf. on Neural Information Processing Systems (NIPS)", 2012.

[28] Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei, « ImageNet Large Scale Visual Recognition Challenge », International Journal of Computer Vision (IJCV), 2015

[29] Lin, M., Q. Chen and S. Yan, « Network In Network », CoRR URL http://arxiv. org/abs/1312.4400, 2013.

[30] Simonyan, K. and A. Zisserman, « Very deep convolutional networks for large-scale image recognition », URL http://arxiv.org/abs/1409.1556, 2014.

[31] Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, « Going deeper with convolutions », in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)", 2015.

Bibliographie

[32] He, K., X. Zhang, S. Ren and J. Sun, « Deep residual learning for image recognition », in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016

b. Webgraphie :

[8] https://www.rapport-gratuit.com/definition-du-traitement-dimages/

[11] https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-Intro-ApprentStat.pdf

[12] http://www.jms-insee.fr/2018/S08_3_ACTE_TARAYOUN_JMS2018.pdf.

[14] http://deeplearning.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/

[26]http://dspace.univ-tlemcen.dz/bitstream/112/12235/1/Classification-des-images-avec-les-reseaux-de-neurones.pdf

[27] https://fr.wikipedia.org/wiki/R%C3%A9seau_neuronal_convolutif

[33] https://www.tensorflow.org/

[34] https://keras.io/

[35] https://developer.nvidia.com/

[36] https://www.techno-science.net/glossaire-definition/Python-langage.html