

الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université –Ain Temouchent- Belhadj Bouchaib
Faculté des Sciences et de Technologie
Département Mathématiques et informatique



Projet de Fin d'Etudes
Pour l'obtention du diplôme de Master en informatique
Filière : Informatique
Spécialité : Réseaux et ingénierie des données
Thème

DETECTION ET RECONNAISSANCE DE VISAGE DANS UNE IMAGE
PAR LE DEEP LEARNING.

Présenté Par :

- 1) M^r BETAOUAF Mohamed Nadir.
- 2) M^{elle} BEN AHMED Daho Fatima Zohra.

Devant le jury composé de :

Président du jury : M^{me} Bedad Fatima Zohra
Encadreur : M^{me} Saidi Samira
Examineur : M^r Benomar mohamed Lamine

Année Universitaire 2020/2021

Remerciements

Le travail présenté dans ce mémoire a été effectué sous la direction de **M^{me} Saidi Samira** à qui nous tenons à adresser nos plus vifs remerciements, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter notre réflexion et ses encouragements lors de la réalisation de ce mémoire.

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

Dédicaces

Ce travail est dédié

A la mère de Binôme Betouaf Nadir qui est décédée trop tôt, qui sera contente de savoir que son fils a enfin terminé ses études et à la maman de Ben Ahmed Daho Fatima Zohra qui est aussi fière de sa fille.

A nos pères Betouaf Kouider et Ben Ahmed Daho Sid Ahmed qui nous a toujours poussé et motivé dans nos études, leurs prières durant ces années d'études que dieu vous protèges.

Résumé:

Les méthodes d'apprentissage en profondeur, en particulier les réseaux de neurones convolutifs (CNN), ont remporté un grand succès dans le domaine de la vision par ordinateur, et ce succès montre une grande supériorité sur les algorithmes de détection et reconnaissance des visages et lui donne la vitesse et la précision de détection, en particulier en temps réel. Le but de ce travail est d'étudier et d'appliquer ce type des réseaux pour la détection et reconnaissance des visages soumis à un apprentissage en profondeur. on a utilisé ce type de réseaux pour la classification et la reconnaissance des visages dans des images, alors on a proposé deux modèles avec différents architectures (le nombre des couches de convolutions, des couches de pooling, des couches entièrement connecté et le nombre d'époque). Les résultats obtenus ont montré que le choix du nombre d'époque et la taille de la base d'images ainsi que la profondeur du réseau ont une grande influence pour avoir des meilleurs résultats.

Mots clés:

apprentissage en profondeur, les réseaux de neurones convolutifs (CNN), classification, détection, reconnaissance, visages pooling.

الملخص :

حققت طرق التعلم العميق وخاصة الشبكات العصبية التلافيفية نجاحًا كبيرًا في مجال رؤية الكمبيوتر ، وهذا النجاح يظهر تفوقًا كبيرًا على خوارزميات اكتشاف الوجه والتعرف عليه ويمنحها سرعة ودقة في الكشف خاصة في الوقت الفعلي. الهدف من هذا العمل هو دراسة وتطبيق هذا النوع من الشبكات لاكتشاف والتعرف على الوجوه الخاضعة للتعلم المتعمق. استخدمنا هذا النوع من الشبكات لتصنيف الوجوه في الصور والتعرف عليها ، لذلك اقترحنا نموذجين ببنيات مختلفة (عدد طبقات الالتفاف وطبقات التجميع والطبقات المتصلة بالكامل وعدد العصر). أظهرت النتائج التي تم الحصول عليها أن اختيار عدد الحقب وحجم قاعدة الصورة وكذلك عمق الحزوز له تأثير كبير للحصول على أفضل النتائج.

الكلمات المفتاحية: التعلم العميق ، الشبكات العصبية التلافيفية ، التصنيف ، الكشف ، التعرف ، الوجوه ، التجميع.

Abstract:

Deep learning methods, in particular convolutional neural networks(CNN) , have achieved great success in the field of computer vision, and this success shows great superiority over face detection and recognition algorithms and gives it detection speed and accuracy, especially in real time. The aim of this work is to study and apply this type of networks for the detection and recognition of faces subjected to in-depth learning. we used this type of networks for the classification and recognition of faces in images, so we proposed two models with different architectures (the number of convolution layers, pooling layers, fully connected layers and the number of era). The results obtained showed that the choice of the number of epochs and the size of the image base as well as the depth of the grating have a great influence to have the best results.

Keywords: deep learning, convolutional neural networks, classification, detection, recognition, faces pooling.

Table des matières

<i>Remerciement</i>	2
Dédicaces	3
Résumé	4
Table des matières	5
Liste des Figures	8
Introduction générale :	9
Chapitre 1 : La détection des visages par le Deep Learning	12
1.1. Introduction :	13
1.2. Qu'est-ce que le machine Learning ? :	13
1.3 Qu'est-ce que l'apprentissage :	15
1.3.1.Types d'apprentissage :	15
1.3.1 L'apprentissage non supervisé	16
1.4.1. Concept	17
1.4.2 Avantages du Deep Learning	18
1.4.3 Les différentes Architectures du Deep Learning	19
1.4.4 Applications du Deep Learning	21
1.5 Etat de l'Art sur la détection d'objets basés sur le Deep learnin	21
1.5.1 Détecteurs à deux étages	23
1.5.2 Détecteurs à un étage	24
1.5.3 Les réseaux de neurones convolutionnels (CNN)	25
1.5.3.1 Définition des CNN	25
1.5.3.2 Architecture de réseaux de neurone convolutionnels	26
1.6. Conclusion	27
Chapitre 2 : La reconnaissance des visages par CNN	30
2.2.1 Définition d'une image	31
2.2.2 Les différents types de format d'image	32
2.2.3 Caractéristiques de l'image :	33
2.2.3.1 Pixel	33

2.2.3.2 Dimension & Résolution :	33
2.2.3.3 Voisinage	34
2.2.3.4 Niveau de gris :	35
2.2.3.5 Contraste	35
2.2.3.6 Luminance	35
2.2.3.7 Bruit	36
2.3 La classification des images & Méthodes de classification	36
2.3.1 Méthodes supervisées	37
2.3.2 Méthodes non supervisées	37
2.4 Indicateurs de performance en classification	37
2.4.1 Matrice de confusion :	37
2.4.1.1 Précision :	38
2.4.1.2 Rappel.:	38
2.4.1.3F-mesure	38
2.5 Les méthodes basées sur CNN les plus connus	39
2.5.1 L'architecture VGG16	39
2.5.2 Architecture Resnet50	40
2.6 Conclusion	41
Chapitre 3 : Conception et Implémentation	42
3.1 Introduction	43
3.2 Environnement de développement et Bibliothèques	43
3.2.1. Spyder :	43
3.2.2. OpenCV (Open Source Computer Vision Library):	43
3.2.3 TensorFlow :	43
3.2.3 Keras	44
3.2.4 Python	44
3.2.5 OS	45
3.2.6 NumPy	45
3.2.7 MLxtend	45
3.2.8 sklear	46
3.3 BDD	46
3. 4 Implémentation et résultats	47

3.5 Formation du model avec keras	48
3.6 Résultats expérimentaux	49
3.7 Conclusion	53
Conclusion générale :	55
Bibliographie :	56

Liste des figures

Figure 1.1: Le père de l'intelligence artificielle « John McCarthy »

Figure 1.2 : Le projet de L'expérience du chat par (Deep Learning)

Figure 1.3 : Un aperçu des performances récentes de détection d'objets

Figure 1.4 : Modèles des réseaux neuronaux de Deep

Figure 1.5: Erreur Top-5 sur IMAGENET

Figure 1.6 : Principales catégories de détecteurs d'objet

Figure 1.7 : Modèle Détecteurs à deux

Figure 1.8: Modèle Détecteurs à un étage

Figure 1.9 Architecture standard d'un réseau de neurone convolutionnel

Figure 1.10 : l'étape de convolution

Figure 1.11 : Max pooling (prend la plus grande valeur de chaque fenêtre)

Figure 1.12 : La partie du Couche entièrement connectée (fully connected)

Figure 2. 1: Voisinage à 4

Figure 2. 2 : Voisinage à 8

Figure 2. 3 : Modèle VGG16

Figure 2.4: architecture Resnet 50.

Figure 3. 1: images stockées dans la base de données.

Figure 3.2: Mise à plat des images.

Figure 3.3: Taux d'erreurs modèle VGG16

Figure 3.4: taux de précision modèle VGG16.

Figure 3.5: Matrice de confusion modèle VGG16 .

Figure3.6 Taux d'erreurs modèle Resnet50

Figure 3.7: Le taux de précision modèle Resnet50.

Figure 3.8: Matrice de confusion modèle Resnet50 .

Figure 3.9: détection et reconnaissance visage

Introduction général

Depuis des années la détection et reconnaissance des visages étions un sujet de recherche dans le domaine de la vision par ordinateur. Aujourd'hui, avec l'augmentation continue de l'utilisation de la détection et reconnaissance des visages dans des plusieurs applications, il est devenu nécessaire de développer des méthodes plus capables pour assurer toujours la précision et la rapidité.

La détection et reconnaissance des visages basées sur l'apprentissage en profondeur (Deep Learning) sont bonnes en termes de robustesse par rapport à la détection et reconnaissance traditionnelles. Cependant, les modèles profonds sont difficiles à former. Les travaux en cours fournissent une perspective pour la recherche dans ce domaine, ce qui ouvre la concurrence devant ses méthodes.

Le but de ce travail est la détection et reconnaissance des visages avec les réseaux de neurones convoltionnels, cela nous permet d'obtenir une méthode rapide et efficace basée sur le Deep Learning.

La détection et reconnaissance des visages et tous les systèmes de détection et reconnaissance ont deux limitations principales : Une limitation en termes de **précision**, et une limitation en termes du **temps**.

L'obtention des hautes performances pour un système de détection et reconnaissance des visages dans le monde réel est un problème ouvert pour les chercheurs depuis quelques années.

Dans son environnement quotidien, un individu a besoin de s'identifier. Donc Il est évident que nous cherchons une solution efficace et ergonomique. Cette solution est basée sur l'analyse de l'image faciale en vue de développer une application capable de sécuriser la vie privée des personnes.

La reconnaissance des visages sert à identifier ou authentifier une ou Plusieurs personnes à partir d'images fixes ou dynamique (vidéo), en utilisant une base de données de visages stockée au préalable. Dans le cas où la scène est encombrée, la détection des visages s'avère nécessaire pour segmenter et extraire des caractéristiques des régions des visages.

C'est dans ce contexte que s'insère notre projet, il s'agit de développer un programme de détection et de reconnaissance faciale qui répond à des besoins de sécurité.

Dans ce mémoire, nous allons discuter les différents travaux développés dans le domaine de la détection et reconnaissance d'une manière générale et de la modalité faciale d'une manière particulière. Nous présentons d'abord le contexte dans lequel notre projet se déroule. Par La reconnaissance des personnes. Nous présentons ensuite, les Bases majeurs à résoudre et nous terminerons par nos contributions apportées.

Ce mémoire est organisé comme suit :

Le chapitre 1 «**Détection des visages par le Deep Learning**»:

ce chapitre explique le Deep Learning brièvement et comment la détection fonctionne avec cette technologie.

Le chapitre 2 «**la reconnaissance des visages par CNN** » :

Ce chapitre introduit les bases d'images et leurs caractéristiques et explique la classification des images brièvement et décrit la reconnaissance des visage par les réseaux neurones convolutionnels en utilisant les deux modèles Qu'on a adopté (VGG16 & Resnet50).

Le chapitre 3 «**Conception et Implémentation**» : ce chapitre fournit une vision détaillée de l'implémentation de notre système, puis l'analyse des résultats obtenus par les différents tests effectués.

Chapitre1

La détection des visages par le Deep Learning

1.1 Introduction

Aujourd'hui, le Deep Learning, ce programme d'apprentissage approfondi comme nous disons en français, est utilisé par des dizaines d'applications et de technologies que nous utilisons au quotidien (Google Map capable de repérer sur une image le numéro des rues, Facebook qui s'en sert pour la reconnaissance faciale sur les photos d'amis par exemple, pour comprendre la voix et adapter ses réponses aux demandes de l'utilisateur, Skype ou Google Trad, qui traduisent des conversations orales en temps réel), et il y en a bien d'autres. Si on entend beaucoup parler de Deep Learning aujourd'hui, ce n'est pas pour rien. Mais que se cache-t-il derrière ces deux mots ? À partir de ce chapitre, nous allons apprendre sur le concept de "Deep Learning" en commençant par la "machine Learning" qui est classé comme une branche majeure de l'intelligence artificielle, et nous pouvons le définir comme science ce qui permet à un ordinateur de se comporter et effectuer toutes ces tâches et d'autres efficacement sans être programmé explicitement. Ensuite, nous allons mettre en relief quelques notions de base liées à la détection de visage par Deep Learning.

1.2. Qu'est-ce que le machine Learning ?

Bien que l'apprentissage automatique ne soit pas nouveau, sa définition précise est encore confuse pour de nombreuses personnes. Le concept de Machine Learning (ML) remonte du milieu du 20ème siècle. Dans les années 1950, le mathématicien britannique Alan Turing imagine une machine capable d'apprendre. Au cours des décennies suivantes, différentes techniques de ML ont été développées pour créer des algorithmes capables

d'apprendre et de s'améliorer de manière autonome. L'apprentissage automatique est une technologie d'intelligence artificielle qui permet aux ordinateurs d'apprendre sans être explicitement programmés pour le faire. Pour apprendre et grandir, les ordinateurs ont besoin de données à analyser et à former. En effet, les données volumineuses (Big Data) est le cœur du Machine Learning, la technologie qui libère tout le potentiel des données volumineuses (Big data). [1]

L'intelligence artificielle (IA) est désormais capable d'apprendre sans l'aide d'un humain. Par exemple, l'algorithme Google DeepMind a récemment appris seul à jouer à 49 jeux vidéo Atari. Par le passé, le développement était limité par le manque d'ensembles de données disponibles, et par son incapacité à analyser des quantités massives de données en quelques secondes. Aujourd'hui, des données sont accessibles en temps réel à tout moment. Ceci permet à l'IA et au Machine Learning de passer à une approche dirigée par les données. La technologie est désormais suffisamment agile pour accéder et analyser des ensembles de données volumineuses

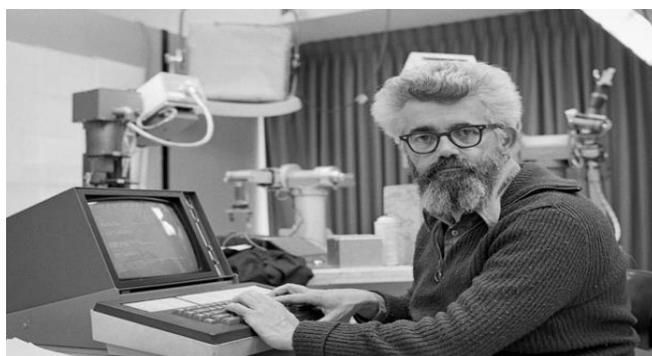


Figure 1.1: Le père de l'intelligence artificielle « John McCarthy

1.3 Qu'est-ce que l'apprentissage ?

L'apprentissage est un sous-domaine de l'intelligence artificielle (IA) qui permet aux machines de reconnaître des objets en fonction de leur expérience de détection précédente. En général, l'objectif de l'apprentissage est de comprendre la structure des données et de les intégrer dans des modèles qui peuvent être compris et utilisés par les tout le monde. [2]

1.3.1.Types d'apprentissage

Dans le domaine du Machine Learning (apprentissage automatique en français), il existe deux principaux types d'apprentissages : supervisés et non supervisés. L'apprentissage supervisé et non supervisé sont tous deux une partie importante de l'apprentissage automatique. Nous expliquons la principale différence entre les deux types comme suit :

1.3.1.1 L'apprentissage supervisé

L'apprentissage est dit supervisé lorsque les données qui entrent dans le processus sont déjà catégorisées (étiquetées) et que les algorithmes doivent s'en servir pour prédire un résultat en vue de pouvoir le faire plus tard lorsque les données ne seront plus catégorisées. Cette technique d'apprentissage est appelée « Supervised Learning » ou apprentissage supervisé. [3]

Afin de mieux comprendre ce concept, prenons un exemple : un utilisateur reçoit chaque jour un grand nombre d'e-mails, certains sont des e-mails d'entreprises importants et d'autres sont des e-mails indésirables non sollicités ou des spam. Un algorithme supervisé sera présenté avec un grand nombre d'e-mails qui ont déjà été étiquetés par l'utilisateur comme spam ou non spam. L'algorithme fonctionnera sur toutes les données étiquetées, faire des prédictions sur l'e-mail et voir si c'est un spam ou non. Cela signifie que l'algorithme examinera chaque exemple et fera une prédiction pour chacun pour savoir si

l'e-mail est un spam ou pas. La première fois, l'algorithme fonctionne sur toutes les données non étiquetées, la plupart des e-mails seront mal étiquetés car il peut fonctionner assez mal au début. Cependant, après chaque exécution, l'algorithme compare sa prédiction au résultat souhaité (l'étiquette). Au fur et à mesure, l'algorithme apprendra à améliorer ses performances et sa précision.

Dans l'exemple que nous avons utilisé, nous avons décrit un processus dans lequel un algorithme apprend à partir de données étiquetées (emails qui ont été catégorisés comme spam ou non-spam). [4]

1.3.1.2 L'apprentissage non-supervisé

La deuxième classe d'algorithmes d'apprentissage automatique est appelée apprentissage non supervisé, dans ce cas, nous n'étiquetons pas les données au préalable, nous laissons plutôt l'algorithme arriver à sa conclusion. Ce type d'apprentissage est important car il est beaucoup plus commun dans le cerveau humain que l'apprentissage supervisé. Les algorithmes d'apprentissage non supervisé sont particulièrement utilisés dans les problèmes de clustering, dans lesquels, étant donné une collection d'objets, nous voulons être en mesure de comprendre et de montrer leurs relations. Une approche standard consiste à définir une mesure de similarité entre deux objets, puis à rechercher tout groupe d'objets plus similaires les uns aux autres, par rapport aux objets des autres clusters. Par exemple, dans le cas précédent des e-mails spam/ non spam, l'algorithme peut être capable de trouver des éléments communs à tous les spam (par exemple, la présence de mots mal orthographiés). Bien que cela puisse fournir une classification meilleure qu'aléatoire, il n'est pas clair que les spam/non spam puissent être facilement séparés. [4]

1.4 Le Deep Learning

1.4.1. Concept

Le Deep Learning (L'apprentissage en profondeur) est un type de l'apprentissage automatique, en utilisant les réseaux de neurones pour saisir modèles complexes. Ce type de technologie permet aux systèmes d'Intelligence Artificielle d'exécuter des tâches humaines, «telles que la reconnaissance visuelle d'objet de la vie réelle ou comprendre la parole. » Alors que le fonctionnement de ces réseaux inspirés du cerveau reste impénétrable, leurs algorithmes de couches interconnectées donnent aux machines la possibilité d'être entraînés et d'effectuer des tâches spécifiques. Les algorithmes qui pilotent ce processus représentent un plus grand nombre des variables plus abstraites. [5] Historiquement, le Deep Learning remonte à 1943, lorsque Walter Pitts et Warren McCulloch ont créé un modèle informatique basé sur les réseaux neuronaux du cerveau humain. Ils ont utilisé une combinaison d'algorithmes et de mathématiques qu'ils ont appelé «logique de seuil» pour imiter le processus de pensée. Depuis lors, le Deep Learning a évolué régulièrement, avec seulement deux pauses significatives dans son développement. En 2011, la vitesse des GPU avait considérablement augmenté, ce qui permettait d'entraîner des réseaux de neurones convolutifs «sans» la pré-formation couche par couche. Avec l'augmentation de la vitesse de calcul, il est devenu évident que le Deep Learning présentait des avantages significatifs en termes d'efficacité et de vitesse. Un exemple est AlexNet, un réseau de neurones convolutifs dont l'architecture a remporté plusieurs concours internationaux en 2011 et 2012.

Toujours en 2012 avec la grande reprise, Google Brain (est un projet de recherche de Deep Learning conduit par Google) a publié les résultats d'un projet inhabituel appelé L'expérience du chat (voir figure 1.2). Le projet à l'esprit libre a exploré les difficultés de «l'apprentissage non supervisé». Le Deep Learning utilise «l'apprentissage supervisé», ce

qui signifie que le réseau neuronal convolutif est formé à l'aide de données étiquetées (pensez aux images d'ImageNet). Actuellement, le traitement du Big Data et l'évolution de l'intelligence artificielle dépendent tous deux du Deep Learning. Le Deep Learning est toujours en évolution et a besoin d'idées créatives.[6]

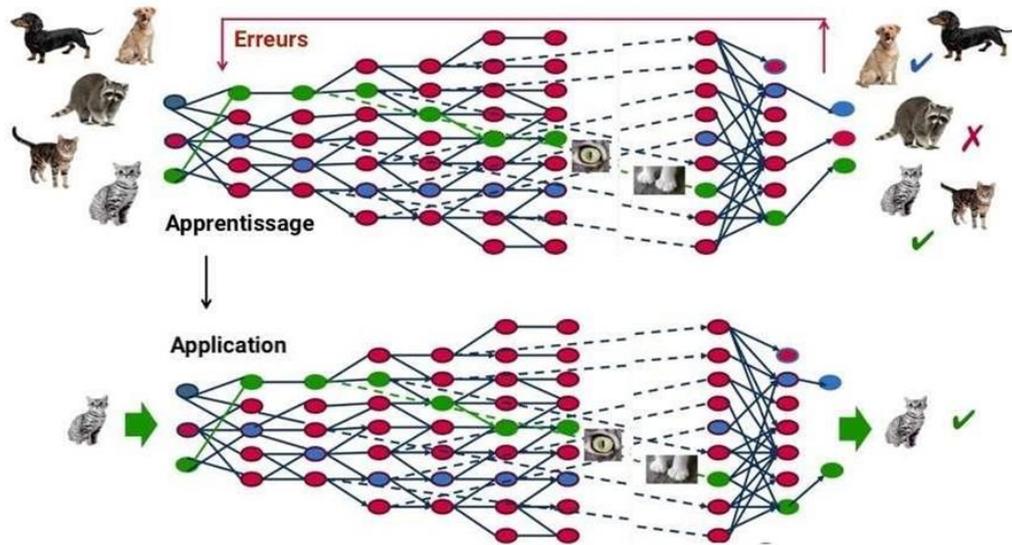


Figure 1.2 : Le projet de L'expérience du chat par (Deep Learning) [8]

1.4.2 Avantages du Deep Learning

L'un des principaux avantages du Deep Learning est l'analyse et l'apprentissage des quantités massives de données, Ce qui lui a permis de pouvoir dans tous les domaines. Plus que ça on observe une amélioration significative des performances **MAP** (mesurées de précision moyenne) depuis l'arrivée du Deep Learning en 2012. **(a)** résultats de la compétition de détection d'objet gagnantes aux concours VOC2007-2012, et **(b)** résultats de la compétition de détection d'objet les plus élevé dans ILSVRC2013-2017 (les résultats dans les deux panels utilisent uniquement les données d'entraînement fournies) [7]

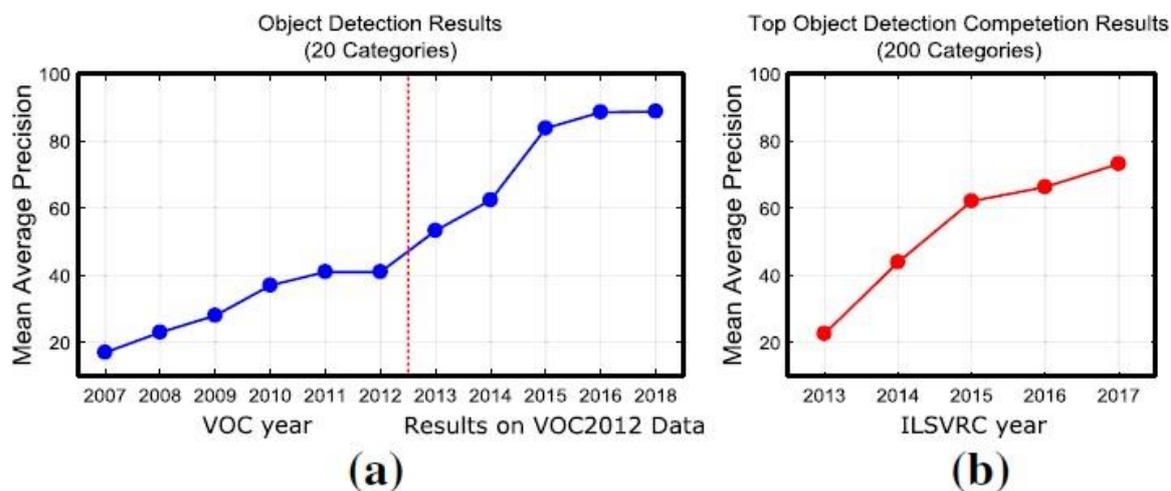


Figure 1.3 : Un aperçu des performances récentes de détection d'objets [7]

1.4.3 Les différentes Architectures du Deep Learning

Dans les réseaux neurones de Deep Learning les informations sont transmises à travers chaque couche, la sortie de la couche précédente fournissant une entrée pour la couche suivante. La première couche d'un réseau est appelée couche d'entrée, tandis que la dernière est appelée couche de sortie. Toutes les couches entre les deux sont appelées couches cachées (**Hidden layers**). Plus le nombre de neurones est élevé, plus le réseau est « profond ». Chaque couche est généralement un algorithme simple et uniforme contenant un type de fonction d'activation. [6] Comme le montre la figure ci-dessous.

A mostly complete chart of Neural Networks

©2016 Fjodor van Veen - asimovinstitute.org

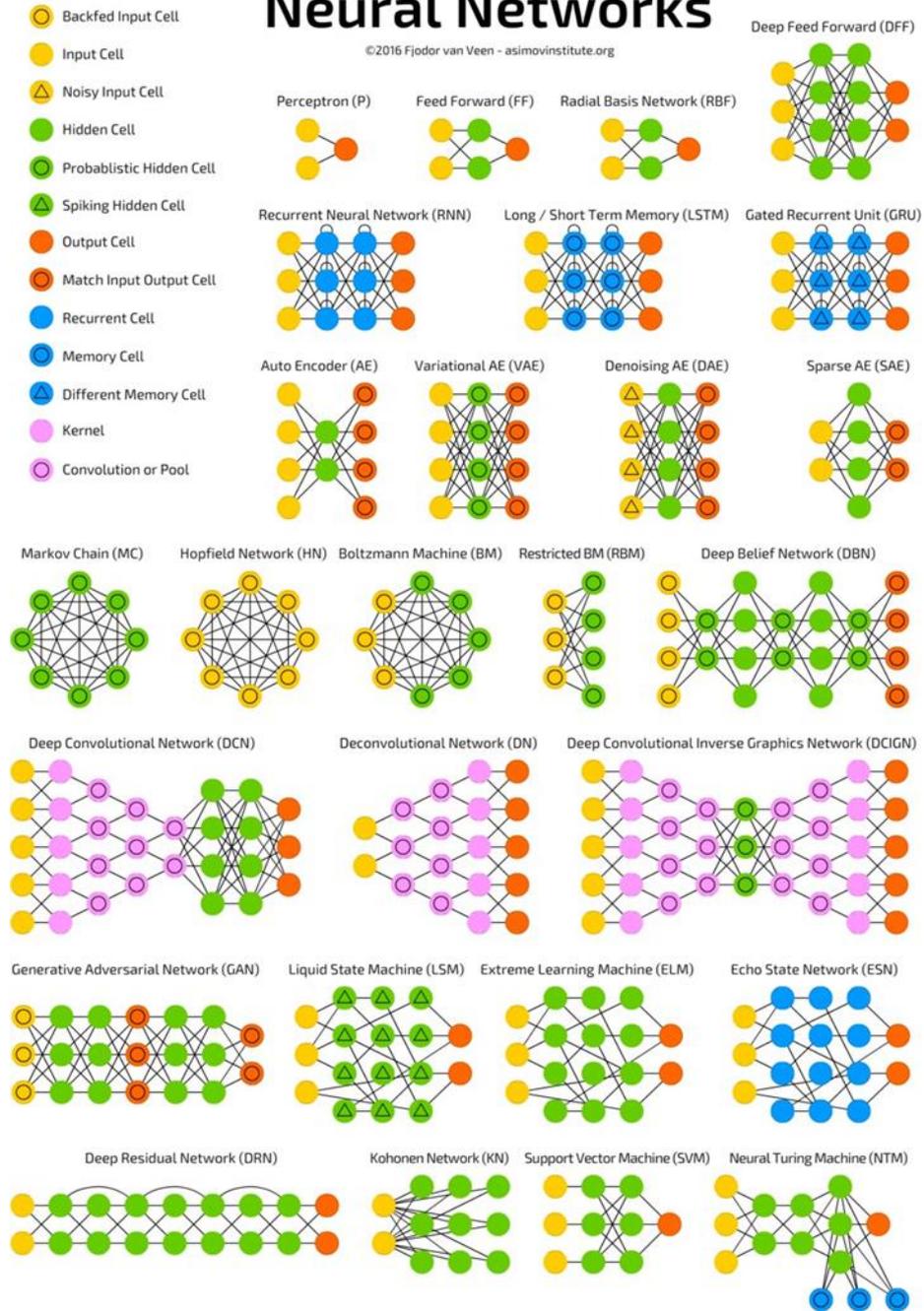


Figure 1.4 : Modèles des réseaux neuronaux de Deep Learning [9]

1.4.4 Applications du Deep Learning

Le Deep Learning est utilisé dans de nombreux domaines comme :

- Reconnaissance d'image,
- Traduction automatique,
- Voiture autonome,
- médicale, Recommandations personnalisées,
- Modération automatique des sociaux, Prédiction financière et commerce automatisé, Identification de pièces défectueuses,
- Détection de malwares ou de fraudes,
- Chatbots (agents conversationnels),
- Exploration spatiale,
- Robots intelligents.

C'est aussi grâce au Deep Learning que l'intelligence artificielle de Google Alpha Go a réussi à battre les meilleurs champions de Go en 2016. Le moteur de recherche du géant américain est lui-même de plus en plus basé sur l'apprentissage par Deep Learning plutôt que sur des règles écrites.[8] Aussi, il est utilisé pour la reconnaissance faciale de Facebook pour identifier automatiquement vos amis sur les photos. C'est également cette technologie qui permet aussi à la reconnaissance faciale Face ID de l'iPhone X d'Apple de s'améliorer au fil du temps. [10] Aujourd'hui, le Deep Learning est même capable de « créer » tout seul des tableaux de Van Gogh ou de Rembrandt, d'inventer aussi un langage totalement nouveau pour communiquer entre deux machines. [8]

1.5 Etat de l'Art sur la détection d'objets basés sur le Deep learning

L'apprentissage en profondeur, également connu sous le nom de Deep Learning ou de techniques basées sur l'apprentissage structuré profond, a récemment remporté un énorme succès dans le traitement d'images numériques pour la détection et la classification d'objets. En conséquence, ils gagnent rapidement en popularité et en

attention auprès de la communauté de recherche en vision par ordinateur. Cette croissance des données d'image a conduit au besoin de détection et de classification automatiques à l'aide de classificateurs basés sur un réseau neuronal profond. Depuis 2012, nous constatons une diminution significative d'erreur comme la montre la figure ci-dessous. Les meilleurs classificateurs d'images se sont trompés de 27% de leurs suppositions en 2010, le Deep learning l'a réduit à 3,5% en 2015

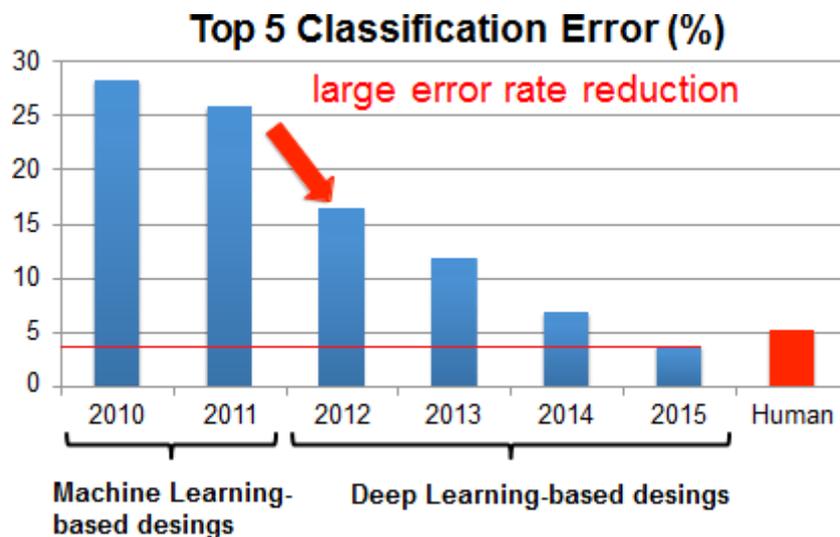


Figure 1.5: Erreur Top-5 sur IMAGENET

Actuellement, les cadres de détection d'objets basés sur l'apprentissage approfondi peuvent être principalement divisés en deux familles (figure 1.6) : **Détecteurs à deux étages**, tels que le CNN régional (R-CNN) et ses variantes, et **Détecteurs à un étage**, tels que YOLO et ses variantes.

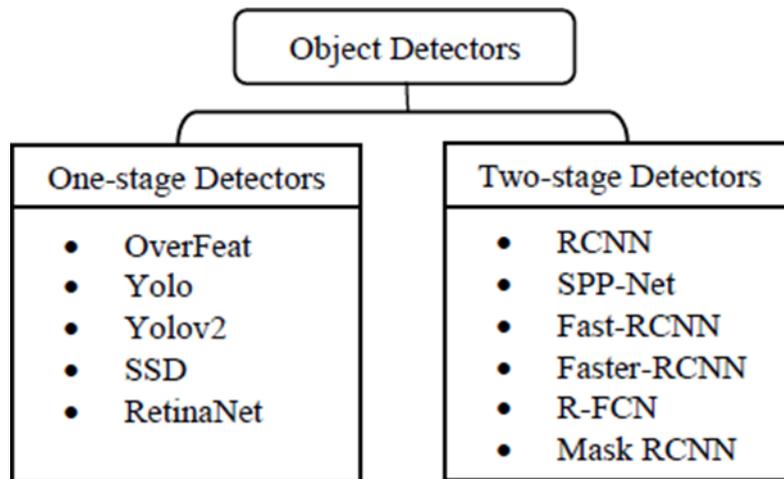


Figure 1.6 : Principales catégories de détecteurs d'objet [11]

1.5.1 Détecteurs à deux étages

Les détecteurs à deux étages ont donné une précision plus élevée avec meilleures performances. Ils rapportent des résultats idéaux mieux que les détecteurs à un étage dans la détection d'objets, mais ils sont généralement plus lents que les détecteurs à un étage car ils ont deux étapes (figure 1.7): La première étape utilise un réseau de proposition de région pour générer des régions d'intérêt qui ont une forte probabilité d'être soit un arrière-plan, soit un objet. La deuxième étape, un modèle basé sur l'apprentissage en profondeur est utilisé pour effectuer la classification finale et la régression par boîte englobante [11], ce qui les rend plus chronophages que les détecteurs à un étage qui utilisent directement un réseau pour la proposition de région et la classification finale. [12]

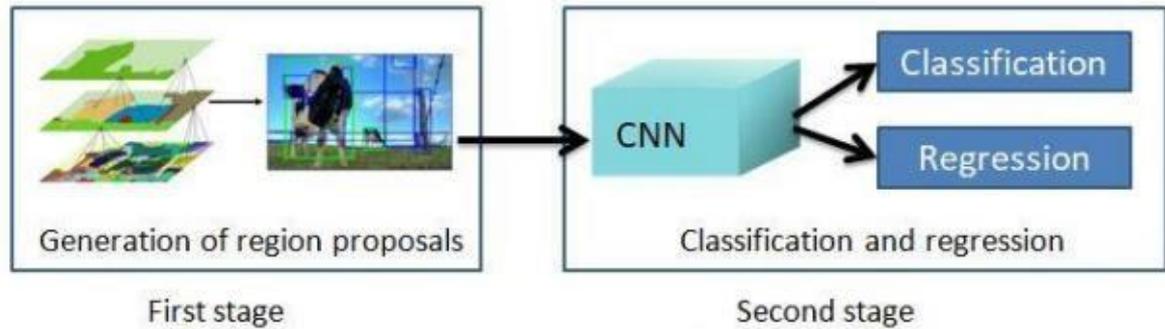


Figure 1.7 : Modèle Détecteurs à deux étages [13]

1.5.2 Détecteurs à un étage

Les détecteurs à un étage sont beaucoup plus rapides et plus recherchés pour les applications de détection d'objets en temps réel, mais ont des performances relativement médiocres par rapport aux détecteurs à deux étages [11]. Car un détecteur à un étage n'a pas d'étape séparée pour la génération de proposition (ou l'apprentissage d'une génération de proposition). Dans ce cas, les étapes de pré-détection de régions d'intérêt et de classification sont fusionnées en une seule étape de détection gérée par un seul réseau neuronal. [12]

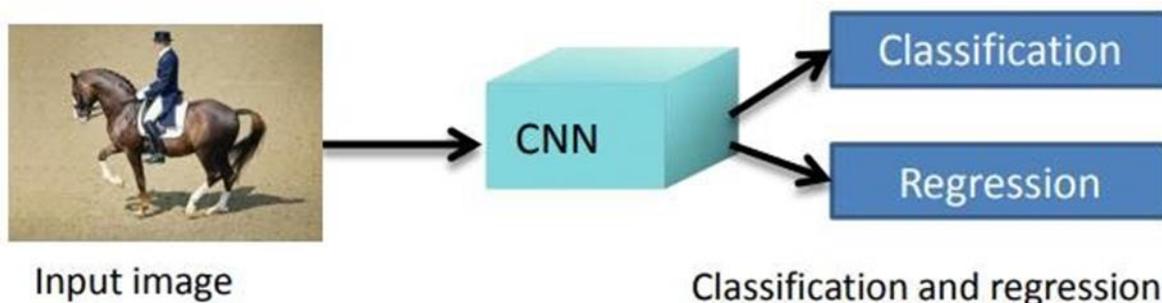


Figure 1.8: Modèle Détecteurs à un étage[13]

1.5.3 Les réseaux de neurones convolutionnels (CNN)

1.5.3.1 Définition des CNN

Les réseaux de neurones convolutionnels sont à ce jour les modèles les plus performants pour classer des images. Désignés par l'acronyme CNN, de l'anglais Convolutional Neural Network, ils comportent deux parties bien distinctes. En entrée, une image est fournie sous la forme d'une matrice de pixels. Elle a 2 dimensions pour une image en niveaux de gris. La couleur est représentée par une troisième dimension, de profondeur 3 pour représenter les couleurs fondamentales [Rouge, Vert, Bleu].

La première partie d'un CNN est la partie convolutive à proprement parler. Elle fonctionne comme un extracteur de caractéristiques des images. Une image est passée à travers une succession de filtres, ou noyaux de convolution, créant de nouvelles images appelées cartes de convolutions. Certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum local. Au final, les cartes de convolutions sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN.

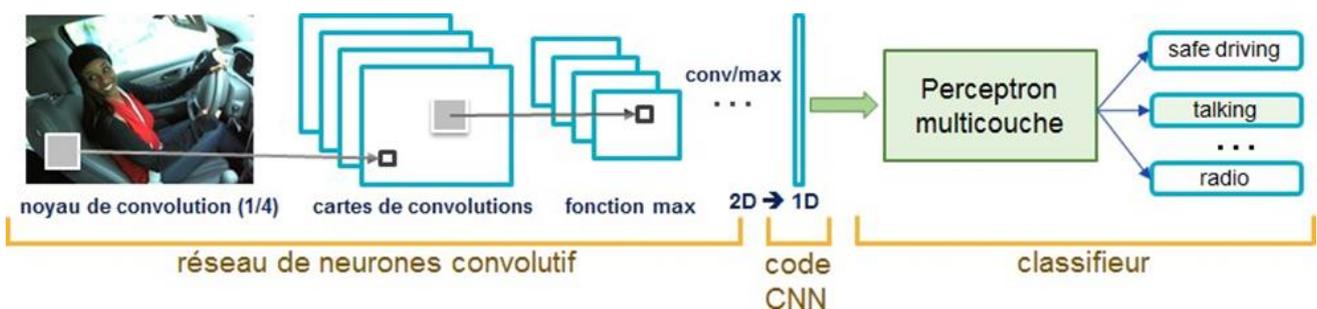


Figure 1.9 Architecture standard d'un réseau de neurone convolutionnel.

Ce code CNN en sortie de la partie convolutive est ensuite branché en entrée d'une deuxième partie, constituée de couches entièrement connectées (perceptron multicouche). Le rôle de cette partie est de combiner les caractéristiques du code CNN pour classer l'image.

La sortie est une dernière couche comportant un neurone par catégorie. Les valeurs numériques obtenues sont généralement normalisées entre 0 et 1, de somme 1, pour produire une distribution de probabilité sur les catégories.

1.5.3.2 Architecture de réseaux de neurone convolutionnel

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

- 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).

a. Couche de convolution

La couche de convolution est le bloc de construction de base d'un CNN. Le but principal de la convolution dans le cas d'un ConvNet est d'extraire des caractéristiques de l'image d'entrée. La convolution préserve la relation spatiale entre les pixels en apprenant les caractéristiques de l'image à l'aide de petits carrés de données d'entrée. Nous n'entrerons pas ici dans les détails mathématiques de la Convolution, mais nous essayerons de comprendre comment cela fonctionne sur les images.

Chaque image peut être considérée comme une matrice de valeurs de pixels. Prenons une image (5 x 5) dont les valeurs de pixel ne sont que de 0 et 1 (notez que pour une image en niveaux de gris, les valeurs de pixel vont de 0 à 255). La matrice verte ci-dessous (voir figure 1.10) est un cas particulier où les valeurs de pixel ne sont que de (0 et 1), et le carré jaune est appelé un «filtre» ou un «noyau». Les filtres agissent comme des détecteurs de caractéristiques de l'image d'entrée d'origine. Maintenant, lorsque nous glissons la matrice de **filtre** sur la matrice (5 x 5) en partant du coin supérieur gauche, on obtient le résultat indiqué dans la figure ci-dessous

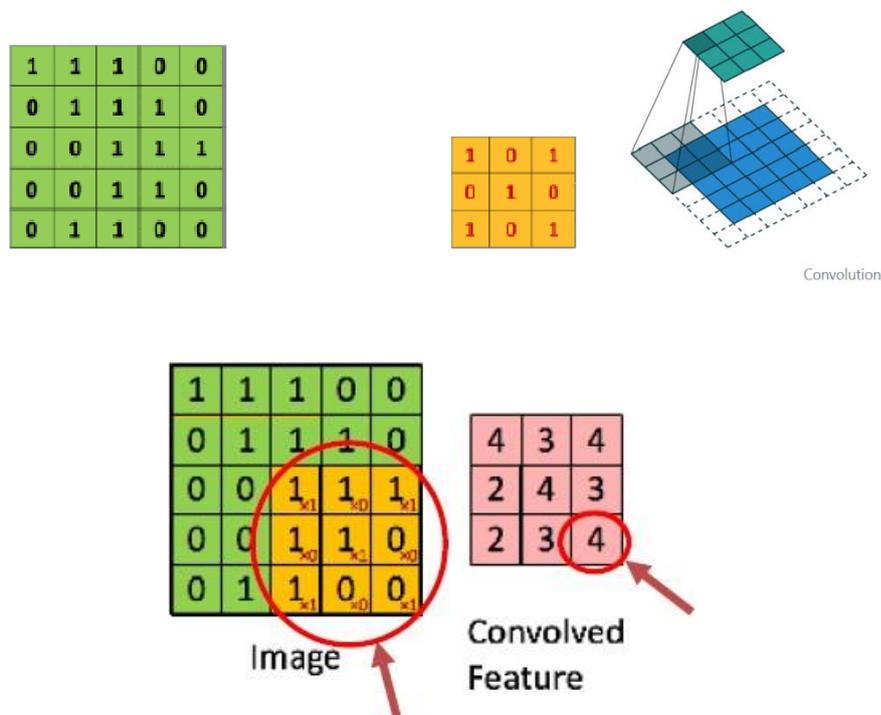


Figure 1.10 : l'étape de convolution. [14]

b. Couche de Pooling

La couche de pooling (POOL) est une opération de sous-échantillonnage typiquement appliquée après une couche convolution. Sa fonction est de réduire progressivement la

taille de la carte des caractéristiques (matrice convolue) pour réduire la quantité des paramètres et de calculs dans le réseau, tout en conservant les informations importantes.[15]

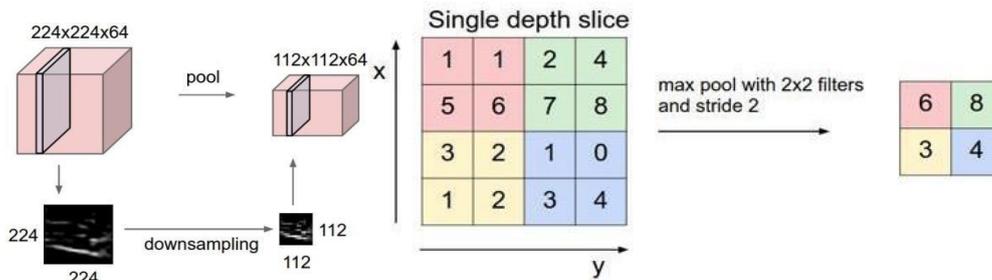


Figure 1.11 : Max pooling (prend la plus grande valeur de chaque fenêtre)

c. La couche entièrement connectée (*Fully-Connected*)

Enfin, après les couches de convolution et pooling, le raisonnement de haut niveau dans le réseau neuronal se fait via des couches totalement connectées. Dans les réseaux de neurones convolutifs, chaque couche agit comme un filtre de détection pour la présence de caractéristiques spécifiques ou des motifs présents dans les données d'origine. Les premières couches d'un réseau convolutif détectent des caractéristiques qui peuvent être reconnues et interprétées facilement. Les couches ultérieures détectent de plus en plus des caractéristiques plus abstraites. La dernière couche du réseau convolutif est capable de faire une classification ultra-spécifique en combinant toutes les caractéristiques spécifiques détectées par les couches précédentes. [16]

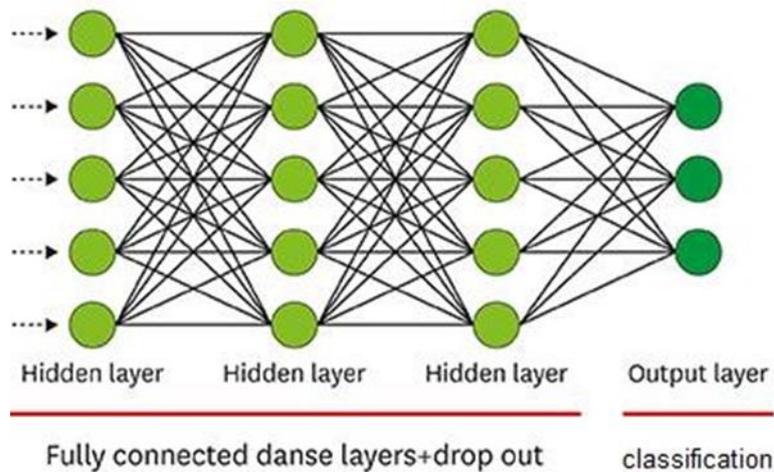


Figure 1.12 : La partie du Couche entièrement connectée (fully connected) [17]

1.6 Conclusion

Dans ce chapitre, nous avons présenté le principe de la détection d'objets et ces différents domaines d'applications. Ensuite, nous avons présenté un état de l'art sur la détection d'objets visant à montrer la diversité des approches proposées dans ce domaine. Ces approches sont divisées en deux catégories : approches basées sur l'apprentissage automatique (Machine Learning) et approches basées sur l'apprentissage en profondeur (Deep Learning). Enfin, nous avons discuté le principe et les différentes étapes des méthodes de détection les plus courantes dans les deux catégories : Viola & Jones et CNN.

Chapitre2

La reconnaissance des visages par CNN

2.1 Introduction:

La classification automatique des images consiste à attribuer automatiquement une classe à une image à l'aide d'un système de classification. On retrouve ainsi la classification d'objets, de scènes, de textures, la reconnaissance de visages, d'empreintes digitale et de caractères. Il existe deux principaux types d'apprentissage l'apprentissage supervisé et l'apprentissage non-supervisé. Dans l'approche supervisée, chaque image est associée à une étiquette qui décrit sa classe d'appartenance. Dans l'approche non supervisée les données disponibles ne possèdent pas d'étiquettes. Dans ce chapitre on va voir tout d'abord les notions de base des images et les méthodes de classification en et la reconnaissance par les CNN .

2.2.1 Définition d'une image

Une image est une représentation planaire d'une scène ou d'un objet situé en général dans un espace tridimensionnel, elle est issue du contact des rayons lumineux provenant des objets formants la scène avec un capteur (caméra, scanner, rayons X, ...). Il ne s'agit en réalité que d'une représentation spatiale de la lumière.

L'image est considérée comme un ensemble de points auquel est affectée une grandeur physique (luminance, couleur). Ces grandeurs peuvent être continues (image analogique) ou bien discrètes (images digitales). Mathématiquement, l'image représente une fonction continue IF , appelée fonction image, de deux variables spatiales représentée par $IF(x, y)$

mesurant la nuance du niveau de gris de l'image aux coordonnées (x, y) . [18]

La fonction Image peut se représenter sous la forme suivante :

$IF : \mathbb{R}^2 \rightarrow \mathbb{R}$ Avec \mathbb{R} : ensemble des réelles. $(x,y) \rightarrow IF(x, y)$ x et y : Deux variables réelles.

2.2.2 Les différents types de format d'image

- **Image couleur RVB** : L'œil humain analyse la couleur à l'aide de trois types de cellules photo 'les cônes'. Ces cellules sont sensibles aux basses, moyennes, ou hautes fréquences (rouge, vert, bleu). Pour représenter la couleur d'un pixel, il faut donc donner trois nombres, qui correspondent au dosage de trois couleurs de base : Rouge, Vert, Bleu. On peut ainsi représenter une image couleur par trois matrices chacune correspondant à une couleur de base.

- **Image d'intensités** : C'est une matrice dans laquelle chaque élément est un réel compris entre 0 (noir) et 1 (blanc). On parle aussi d'image en niveaux de gris, car les valeurs comprises entre 0 et 1 représentent les différents niveaux de gris.

- **Image binaire** : Une image binaire est une matrice rectangulaire dans l'élément valent 0 ou 1. Lorsque l'on visualise une telle image, les 0 sont représentés par du noir et les 1 par du blanc. [18]

2.2.3 Caractéristiques de l'image

L'image est un ensemble structuré d'information caractérisé par les paramètres suivants :

2.2.3.1 Pixel

Le pixel est l'abréviation du mot « Picture élément », est une unité de surface permettant de définir la base d'une image numérique. Il matérialise un point donné (x, y) du plan de l'image. L'information présentée par le pixel est le niveau de gris (ou la couleur) prélevée à l'emplacement correspondant dans l'image réelle. La différence entre image monochrome et image couleur réside dans la quantité d'informations contenue dans chaque pixel, par exemple dans une image couleur (RVB : Rouge, Vert, Bleu) la valeur d'un pixel est représentée sur trois octets pour chaque couleur.

2.2.3.2 Dimension & Résolution

La dimension est la taille de l'image. Elle se présente sous forme d'une matrice dont les éléments sont des valeurs numériques représentatives des intensités lumineuses (pixels).

Le nombre de lignes de cette matrice multiplié par le nombre de colonnes nous donne le nombre total de pixels dans une image.

Par contre, la résolution est la clarté ou la finesse de détails atteinte par un moniteur ou une imprimante dans la production d'images. Sur les moniteurs d'ordinateur, la résolution est exprimée en nombre de pixels par unité de mesure (pouce ou centimètre). On utilise aussi le mot résolution pour désigner le nombre total de pixels horizontaux et verticaux

sur un moniteur. Plus ce nombre est grand, plus la résolution est meilleure.

2.2.3.3 Voisinage

Le plan de l'image est divisé en termes de formes rectangulaires ou hexagonales

permettant ainsi l'exploitation de la notion de voisinage (voir figure 2.1). Le voisinage

d'un pixel est formé par l'ensemble des pixels qui se situent autour de ce même pixel. On

définit aussi l'assiette comme étant l'ensemble de pixels définissant le voisinage pris en

compte autour d'un pixel.

On distingue deux types de voisinage :

Voisinage à 4 : On ne prend en considération que les pixels qui ont un coté commun avec

le pixel considéré.

Voisinage à 8 : On prend en compte tous les pixels qui ont au moins un point en liaison avec le pixel considéré.

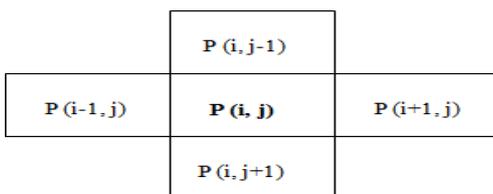


Figure 2. 1: Voisinage à 4

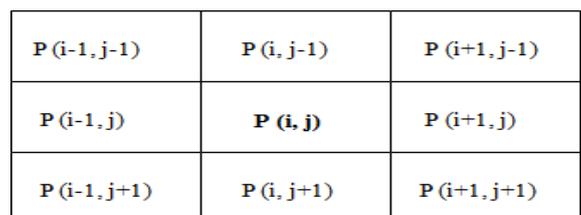


Figure 2. 2 : Voisinage à 8

2.2.3.4 Niveau de gris

C'est la valeur d'intensité lumineuse d'un pixel. Cette valeur peut aller du noir (0) jusqu'au blanc (255) en passant par les nuances qui sont contenues dans l'intervalle [0, 255]. Elle correspond en fait à la quantité de la lumière réfléchie.

Pour 8 bits, on dispose de 256 niveaux de gris dont 40 sont reconnus à l'œil nu. Plus le nombre de bits est grand plus les niveaux sont nombreux et plus la représentation est fidèle.

2.2.3.5 Contraste

C'est l'opposition marquée entre deux régions d'une image. Une image contrastée présente une bonne dynamique de la distribution des valeurs de gris sur tout l'intervalle des valeurs possibles, avec des blancs bien clairs et des noirs profonds. Au contraire une image peu contrastée a une faible dynamique, la plupart des pixels ayant des valeurs de gris très proches.

Si L_1 et L_2 sont les degrés de luminosité respectivement de deux zones voisines A_1 et A_2 d'une image, le contraste est défini par le rapport : $C = \frac{L_1 - L_2}{L_1 + L_2}$

2.2.3.6 Luminance

C'est le degré de luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface, pour un observateur lointain, le mot luminance est substitué au mot brillance, qui correspond à l'éclat d'un objet.

Une bonne luminance se caractérise par :

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

2.2.3.7 Bruit

Un bruit (parasite) dans une image est considéré comme un phénomène de brusque variation de l'intensité d'un pixel par rapport à ses voisins, il provient de l'éclairage des dispositifs optiques et électroniques du capteur. C'est un parasite qui représente certains défauts (poussière, petits nuages, baisse momentanée de l'intensité électrique sur les capteurs, ...etc.). Il se traduit par des taches de faible dimension et dont la distribution sur l'image est aléatoire. [19]

2.3 La classification des images & Méthodes de classification

Une tâche très courante est la classification d'images. Pour pouvoir utiliser les images pour la cartographie ou pour des analyses complémentaires, il est souvent important de traduire l'information de fréquence contenue dans les images en information thématique portant sur l'occupation du sol ou la couverture végétale. Vous avez généralement le choix entre deux approches : la classification supervisée et non-supervisée. De nombreuses méthodes classiques ont été consacrées, elles peuvent être séparées en deux grandes catégories : les méthodes de classification supervisée et les méthodes de classification non supervisée.

2.3.1 Méthodes supervisées

L'objectif de la classification supervisée est principalement de définir des règles permettant de classer des objets dans des classes à partir de variables qualitatives ou quantitatives caractérisant ces objets. On dispose au départ d'un échantillon dit d'apprentissage dont le classement est connu. Cet échantillon est utilisé pour l'apprentissage des règles de classement.

Il est nécessaire d'étudier la fiabilité de ces règles pour les comparer et les appliquer, évaluer les cas de sous apprentissage ou de sur apprentissage (complexité du modèle). On utilise souvent un deuxième échantillon indépendant, dit de validation ou de test.

2.3.2 Méthodes non supervisées

Procède de la façon contraire. C'est à dire ne nécessitent aucun apprentissage et aucune tâche préalable d'étiquetage manuel. Elle consiste à représenter un nuage des points d'un espace quelconque en un ensemble de groupes appelé Cluster. Il lié généralement au domaine de l'analyse des données comme l'ACP. Un «Cluster» est une collection d'objets qui sont «similaires» entre eux et qui sont «dissemblables» par rapport aux objets appartenant à d'autres groupes. [20]

2.4 Indicateurs de performance en classification

2.4.1 Matrice de confusion

Prenons l'exemple d'un classifieur binaire, c'est-à-dire, qui prédit 2 classes notées classe 0 et classe 1.

Pour mesurer les performances de ce classifieur, il est d'usage de distinguer 4 types d'éléments classés pour la classe voulue :

1. Vrai positif VP : Elément de la classe 1 correctement prédit
2. Vrai négatif VN : Elément de la classe 0 correctement prédit

3. Faux positif FP : Elément de la classe 1 mal prédit

4. Faux négatif FN : Elément de la classe 0 mal prédit

Ces informations peuvent être rassemblés et visualisés sous forme de tableau dans une matrice de confusion. Dans le cas d'un classifieur binaire, on obtient :

		Classe prédite	
		Classe 0	Classe 1
Classe réelle	Classe 0	VN	FN
	Classe 1	FP	VP

En particulier, si la matrice de confusion est diagonale, le classifieur est parfait.

Notons que la matrice de confusion est aussi généralisable lorsqu'il y a $k > 2$ classes à prédire. [21]

Il est possible de calculer plusieurs indicateurs résumant la matrice de confusion. Par exemple si nous souhaitons rendre compte de la qualité de la prédiction sur la classe 1, on définit :

2.4.1.1 Précision. Proportion d'éléments bien classés pour une classe donnée :

$$Precision_{de\ la\ classe\ 1} = \frac{VP}{VP + FP}$$

2.4.1.2 Rappel. Proportion d'éléments bien classés par rapport au nombre d'éléments de la classe à prédire :

$$Rappel_{de\ la\ classe\ 1} = \frac{VP}{VP + FN}$$

2.4.1.3 F-mesure. Mesure de compromis entre précision et rappel :

$$F - \text{mesure de la classe } i = \frac{2 * (\text{Précision} * \text{Rappel})}{\text{Précision} + \text{Rappel}}$$

Il est possible de calculer tous ces indicateurs pour chaque classe. La moyenne sur chaque classe de ces indicateurs donne des indicateurs globaux sur la qualité du classifieur.

$$\text{Précision} = \frac{1}{k} \sum_{i=1}^k \frac{VP_i}{VP_i + FP_i}$$

$$\text{Rappel} = \frac{1}{k} \sum_{i=1}^k \frac{VP_i}{VP_i + FN_i}$$

$$F - \text{mesure} = \frac{2 * (\text{Précision} * \text{Rappel})}{\text{Précision} + \text{Rappel}}$$

2.5 Les méthodes basées sur CNN les plus connus

Plusieurs architectures de reconnaissance faciale basée sur les réseaux convolutifs existent, les plus utilisées sont :

2.5.1 L'architecture VGG16

VGG16 est un modèle de réseau de neurones convolutif proposé par Simonyan et al . de l'Université d'Oxford dans l'article intitulé «Réseaux convolutionnels très profonds pour la reconnaissance d'images à grande échelle». Le modèle atteint une précision de 92,7% dans le top 5 des tests dans ImageNet, qui est un jeu de données de plus de 14 millions d'images appartenant à 1000 classes. C'était l'un des fameux modèles soumis à la conférence ILSVRC-2014.

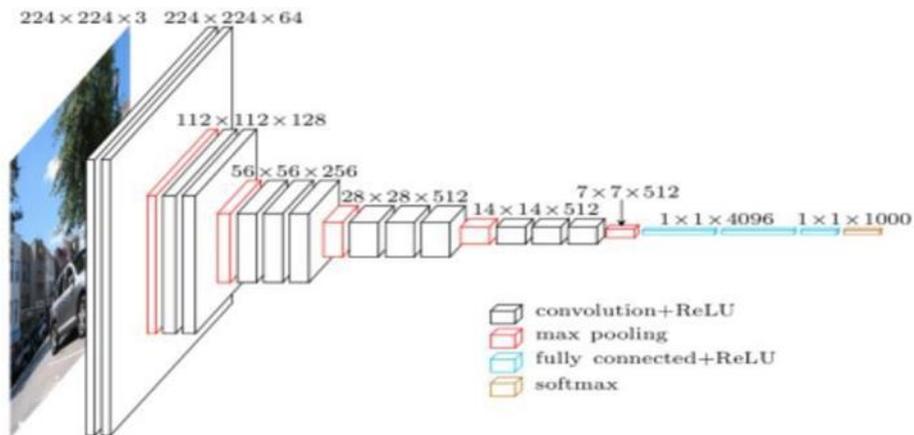


Figure 2. 3: Modèle VGG16

2.5.2 Architecture Resnet50

Pour les réseaux plus profonds, nous utilisons la structure de goulot d'étranglement, en utilisant d'abord une convolution 1×1 pour la réduction de dimensionnalité, puis une convolution 3×3 , et enfin en utilisant la dimensionnalité 1×1 pour restaurer la dimension d'origine.

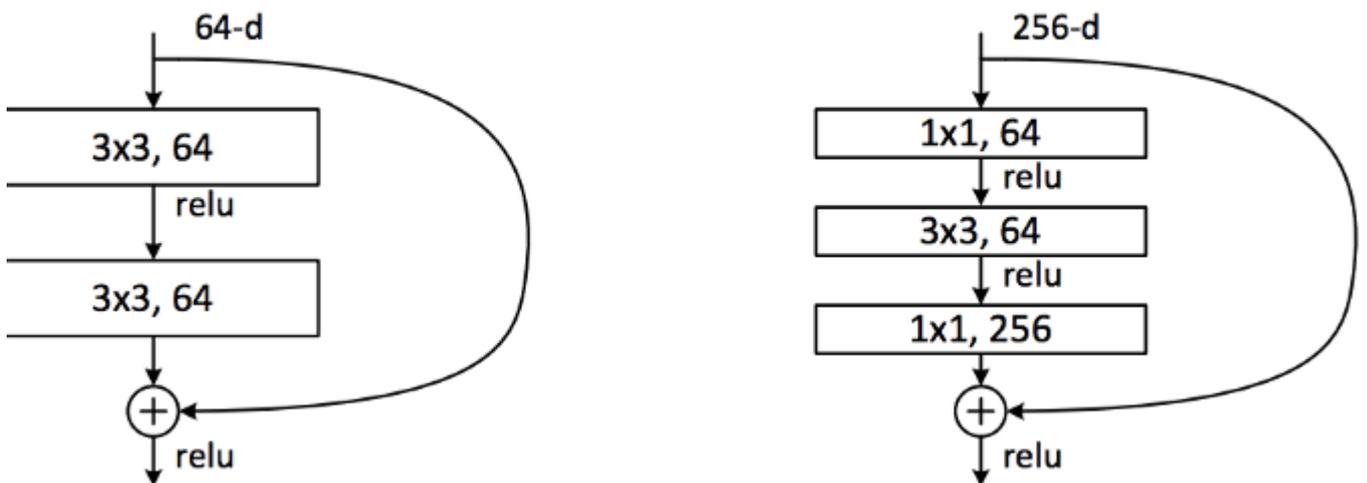


Figure 2.4: architecture Resnet 50.

Cela équivant à réduire le nombre de paramètres pour le même nombre de couches, afin de pouvoir l'étendre à des modèles plus profonds. Par conséquent, l'auteur a proposé ResNet avec 50, 101 et 152 couches, et non seulement il n'a pas eu de problèmes de dégradation, mais aussi le taux d'erreur a été considérablement réduit et la complexité de calcul est devenue très faible.

2.6 Conclusion

Ce chapitre On a bien compris les réseaux neurones convolutionnels et on a montré les Les modèles des CNN qu'on a utilisé Vgg16 et Resnet50 et avant ça on a pu savoir les bases d'images et les motivations de classification des images modèles Dans la détection et reconnaissance des visage. La structure et les différentes couches de ces Modèles et la comparaison entre eux et leur implémentation seront détaillées dans le chapitre 3.

Chapitre 3

Conception et Implémentation

3.1 Introduction

L'objectif de ce chapitre est de présenter les étapes de l'implémentation de l'approche proposée dans le Cadre d'un système de la détection et reconnaissance d'un visage basé sur le Deep Learning. Nous commençons tout d'abord par la présentation des différents outils d'implémentation, du langage et de L'environnement de développement que nous avons utilisé. Puis la comparaison entre les deux modèles VGG16 et ResNet50 qui sont deux modèles de CNN utilisées pour la détection et la reconnaissance et on termine par les tests effectués. Ce chapitre en fait est composé de deux parties, l'implémentation du système et les résultats expérimentaux des tests.

3.2 Environnement de développement et Bibliothèques

3.2.1. Spyder

Spyder est un environnement de développement pour Python. Libre (Licence MIT) et multiplateforme (Windows, Mac OS, GNU/Linux), il intègre de nombreuses bibliothèques d'usage scientifique : Matplotlib, NumPy, SciPy et IPython. Créé et développé par Pierre Raybaut en 2008, Spyder est maintenu, depuis 2012, par une communauté de développeurs qui ont pour point commun d'appartenir à la communauté Python scientifique.

3.2.2. OpenCV (Open Source Computer Vision Library)

Est une bibliothèque proposant un ensemble de plus de 2500 algorithmes de vision par ordinateur spécialisé dans le traitement d'images, accessible au travers d'API pour les langages C, C++, et Python. Elle est distribuée sous une licence BSD (libre) pour les plateformes Windows, GNU/Linux, Android et MacOS [W8], nous avons utilisé cette bibliothèque pour la détection du visage à partir des images introduites.

3.2.3 TensorFlow

TensorFlow est un framework de programmation pour le calcul numérique qui a été rendu Open Source par Google en Novembre 2015. Depuis son release, TensorFlow n'a cessé de gagner en popularité, pour devenir très rapidement l'un des frameworks les plus utilisés pour le Deep Learning et donc les réseaux de neurones. Son nom est notamment inspiré

du fait que les opérations courantes sur des réseaux de neurones sont principalement faites via des tables de données multi-dimensionnelles, appelées Tenseurs (Tensor). Un Tensor à deux dimensions est l'équivalent d'une matrice. Aujourd'hui, les principaux produits de Google sont basés sur TensorFlow: Gmail, Google Photos, Reconnaissance de voix.

3.2.3 Keras

Keras est une API de réseaux de neurones de haut niveau, écrite en Python et capable de fonctionner sur TensorFlow ou Theano. Il a été développé en mettant l'accent sur l'expérimentation rapide. Être capable d'aller de l'idée à un résultat avec le moins de délai possible est la clé pour faire de bonnes recherches. Il a été développé dans le cadre de l'effort de recherche du projet ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System), et son principal auteur et mainteneur est François Chollet, un ingénieur Google.

En 2017, l'équipe TensorFlow de Google a décidé de soutenir Keras dans la bibliothèque principale de TensorFlow. Chollet a expliqué que Keras a été conçue comme une interface plutôt que comme un cadre d'apprentissage end to end. Il présente un ensemble d'abstractions de niveau supérieur et plus intuitif qui facilitent la configuration des réseaux neuronaux indépendamment de la bibliothèque informatique de backend. Microsoft travaille également à ajouter un backend CNTK à Keras aussi.

3.2.4 Python

Python est un langage de programmation de haut niveau interprété (il n'y a pas d'étape de compilation) et orienté objet avec une sémantique dynamique. Il est très sollicité par une large communauté de développeurs et de programmeurs. Python est un langage simple, facile à apprendre et permet une bonne réduction du cout de la maintenance des codes.

Les bibliothèques (packages) python encouragent la modularité et la réutilisabilité des codes. Python et ses bibliothèques sont disponibles (en source ou en binaires) sans charges pour la majorité des plateformes et peuvent être redistribués gratuitement.

3.2.5 OS

Ce module fournit une manière portable d'utiliser les fonctionnalités dépendantes du système d'exploitation (interagir avec votre système d'exploitation ; créer, supprimer, déplacer des dossiers et parfois de modifier le répertoire de travail). Il est installé par: *pip install os-sys*

3.2.6 NumPy

NumPy est une extension du langage de programmation Python, bibliothèque logicielle libre et open source, destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux. Installer par : *pip install numpy*. La figure 5.6 montre le logo de NumPy.

3.2.7 MLxtend

MLxtend est une bibliothèque qui implémente une variété d'algorithmes et d'utilitaires de base pour la machine, l'apprentissage et l'exploration de données. L'objectif principal de MLxtend est de créer des outils couramment utilisés, accessible aux chercheurs universitaires et aux scientifiques des données dans les industries se concentrant sur les API conviviales et intuitives et la compatibilité avec les bibliothèques d'apprentissage automatique existantes, telles que comme scikit-learn, le cas échéant. Alors que MLxtend implémente une grande variété de fonctions, les points forts incluent les algorithmes de sélection de caractéristiques séquentielles (Pudil, Novovicova et Kittler 1994), des implémentations de la généralisation empilée (Wolpert 1992) pour la classification et régression, et des algorithmes pour l'exploration de

motifs fréquents (Agrawal et Ramakrishnan 1994).

3.2.8 sklearn

scikit-learn est livré avec quelques ensembles de données standard, par exemple les ensembles de données sur l'iris

et les chiffres pour la classification et l'ensemble de données sur le diabète pour la régression.

3.3 BDD

La base de données d'images des visages que nous avons utilisés est une base de données collectée et enregistrée par notre code .elle contient 565 visages qui sont réparties en deux parties à savoir :

- 45 images représentant les images de test
- 520 représentent les images d'apprentissage.
- Toutes ces images ont été déjà résolues et contiennent chacune une taille de 224x224 pixels (les dimensions utilisées dans la base de données ImagNet).La détection et la reconnaissance se font en temps réel et l'extension est JPEG.



Figure 3. 1: images stockées dans la base de données.

3. 4 Implémentation et résultats

Cette section décrit les expériences réalisées et les résultats expérimentaux de notre système sur la détection et la reconnaissance faciale. Elle contient deux parties pour examiner la performance de l'information.

Dans la première expérience, notre système est testé en utilisant notre base de données.

Dans la seconde expérience, notre système est testé en utilisant des vidéos externes capturées par la Webcam afin de vérifier l'efficacité de notre système.

L'algorithme proposé est principalement réalisé en Quatre étapes comme ci-dessous :

1. Capturer et enregistrer les images dans la base de données .
2. Redimensionner les images d'entrée au format 244x244 ;
3. Construire une structure CNN avec trois couches constituées respectivement de :

filtre convolutif, ReLu et de regroupement maximum .Cette couche est répétée trois fois.

Après avoir extrait toutes les fonctionnalités, nous allons utiliser :

Flatten : elle permet d'aplatir les images .

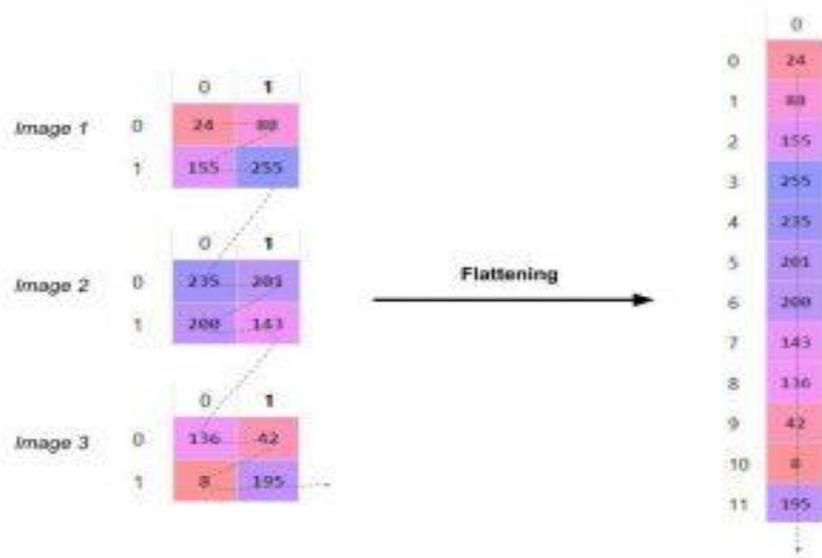


Figure 3.2: Mise à plat des images.

Dense : désigne le nombre que le veut sur une couche, elle prend toujours en entrée le résultat de l'image aplatis de la couche précédente.

Softmax : qui est un classifieur qui désigne la distribution de la probabilité ou encore que la somme de toute la probabilité de la sortie doit être égale à 1.

3.5 Formation du model avec keras

La première de la chose à faire consiste à tout d'abord configurer l'environnement

Ensuite, installer toutes les bibliothèques citées ci-dessus

Importer les bibliothèques installées

Ensuite, nous allons importer le type de modèle séquentiel de Keras. Il s'agit tout simplement d'un empilement linéaire de couches de réseaux de neurones à travers la

de la commande : **from keras. models import Model**

Nous importerons les couches "principales" de Keras.

Voici les couches utilisées dans presque tous les réseaux de neurones :

From keras. Layers import Dense, Flatten

Notre modèle contient:

(1, 244, 244) qui correspond à la (profondeur, largeur, hauteur) de chaque image de notre Base de donnée qui sera redimensionnée au format 0x244x244.

Nous avons utilisé une matrice à deux dimensions.

3.6 Résultats expérimentaux :

Dans cette partie nous nous intéressons aux résultats d'exécution de l'approche développée des CNN Resnet50 et VGG16 et on va comparer les résultats obtenus pour la même base de donnée et le même nombre d'epoches = 5 (nombre d'itération dans notre base de données) .

Si nous faisons une comparaison entre les epoches nous constatons que le taux d'erreur baisse tant dis que la précision augmente cela signifie que notre modèle a été bien entraîné et réponds d'ailleurs à la définition des réseau de neurones qui disait que plus le réseau de neurones est profond plus que ses performances sont meilleurs.

Résultats obtenus par le modèle VGG6:

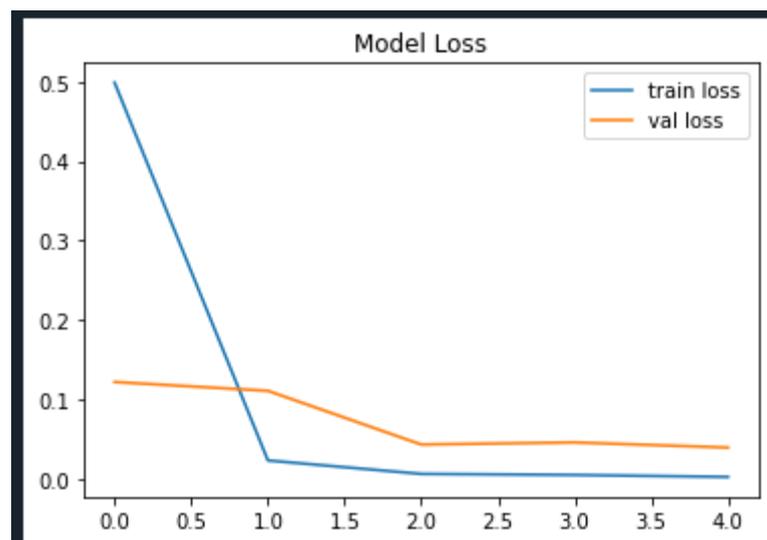


Figure 3.3: Taux d'erreurs modèle VGG16

Les résultats obtenus sur le taux d'entraînement de la perte représentés dans la figure ci-dessous nous ont donné **0.0021**

Et les résultats sur la perte nous ont donné un taux de **0.0391** .

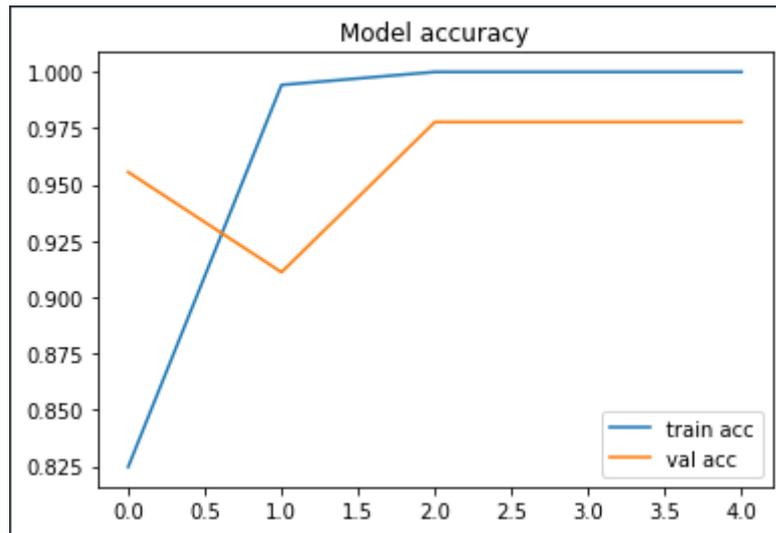


Figure 3.4: taux de précision modèle VGG16.

Les résultats obtenus sur les données tests représentées dans le tableau ci-dessous nous ont donné un taux de : **1.0000** .

Et les résultats sur les données d'entraînement nous ont donné un taux de **0.9778**.

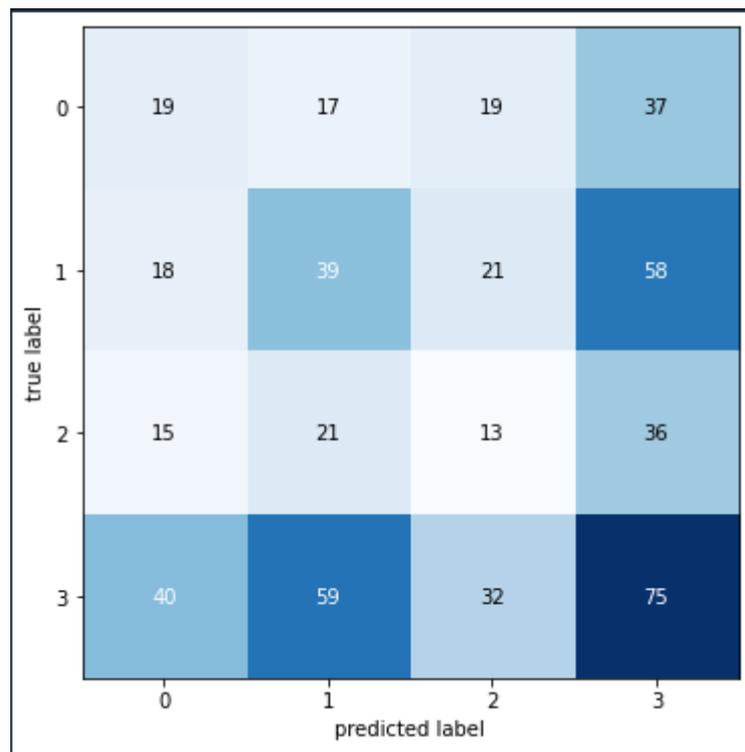


Figure 3.5: Matrice de confusion modèle VGG16 .

Résultats obtenus par le modèle Resnet50

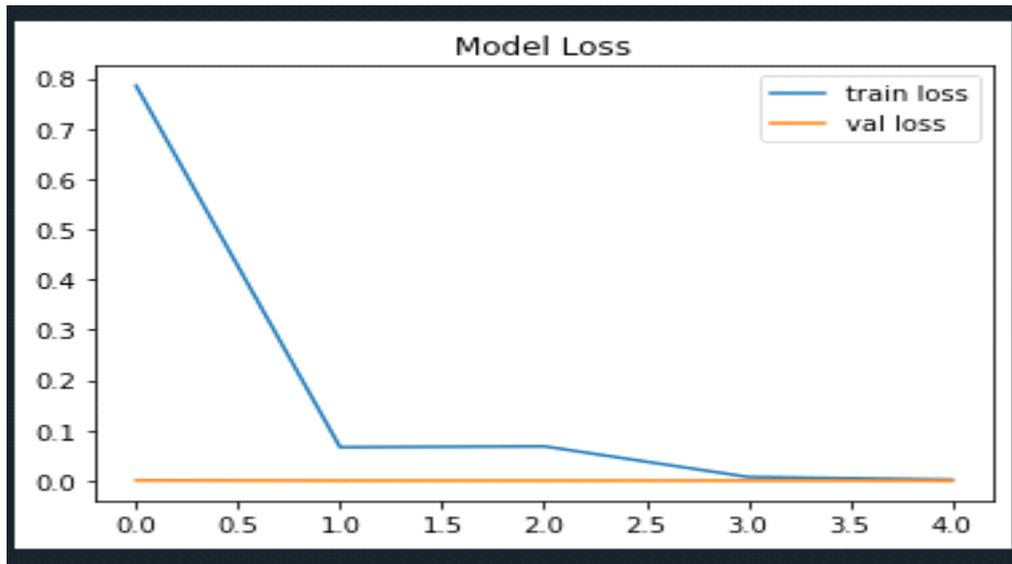


Figure3.6 Taux d'erreurs modèle Resnet50

Les résultats obtenus sur le taux d'entraînement de la perte représenté dans le tableau ci-dessous nous ont donné **0.0020**

Et les résultats sur la perte nous ont donné un taux de **3.6292e-07**

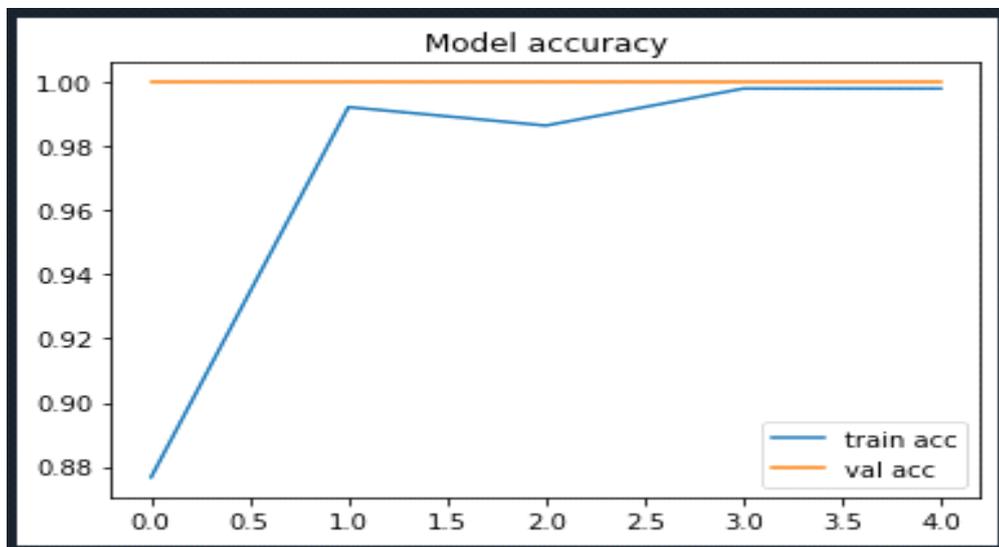


Figure 3.7: Le taux de précision modèle Resnet50.

Les résultats obtenus sur les données tests représentées dans le tableau ci-dessous nous ont donné un taux de : **0.9981** .

Et les résultats sur les données d'entraînement nous ont donné un taux de **1.0000**.

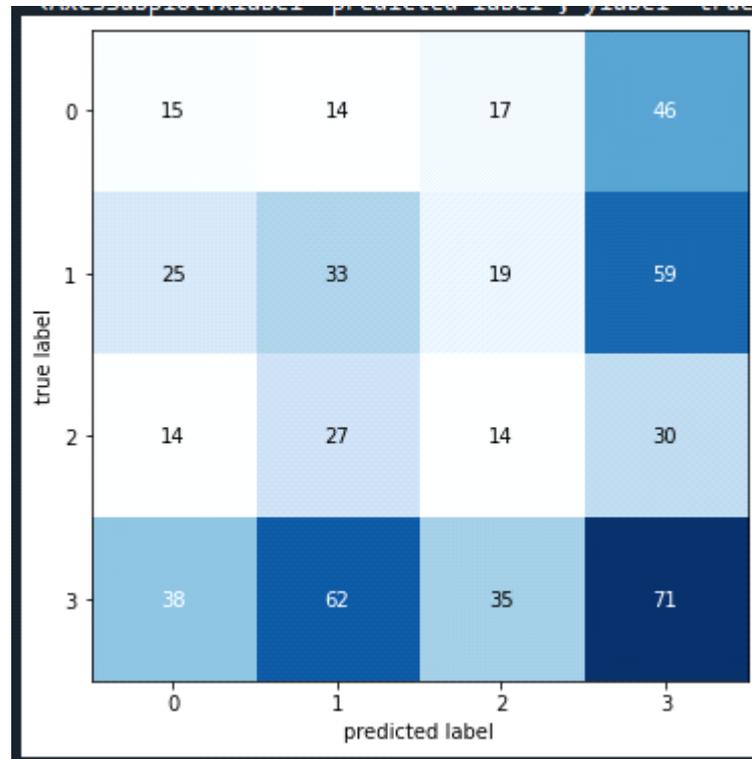


Figure 3.8: Matrice de confusion modèle Resnet50 .

Nous remarquons que le modèle VGG16 a été adapté avec notre base de données plus que le modèle Resnet50. Les performances obtenues sont relativement acceptables prenons en compte la taille réduite de l'échantillon d'images de la BDD. Pour mieux tester l'efficacité du modèle développé dans ce projet, nous allons essayer d'introduire des images externes par un capteur vidéo.

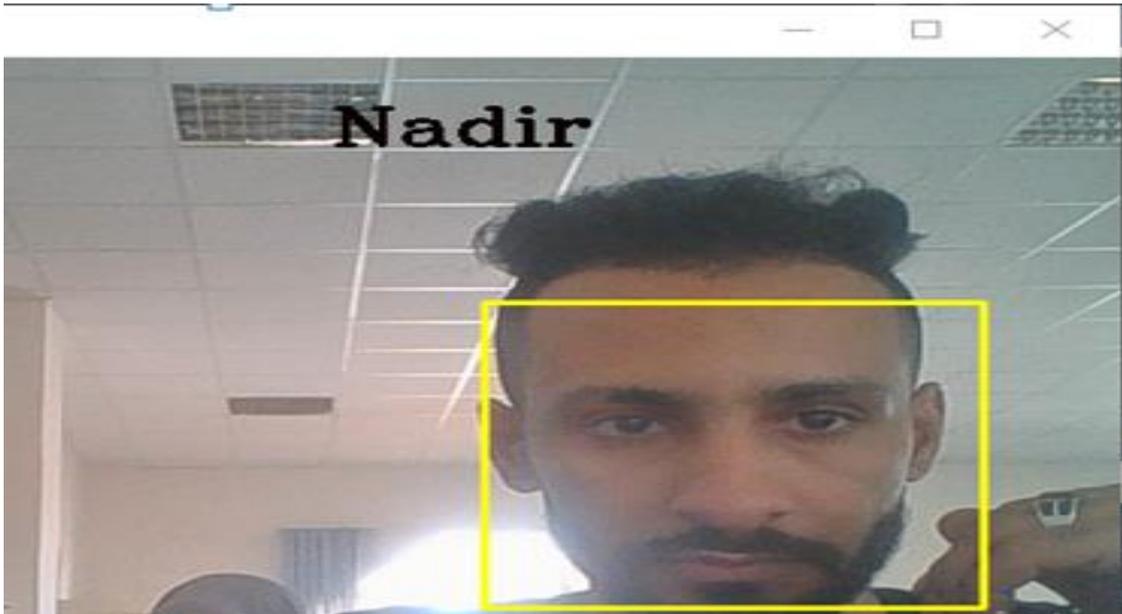


Figure 3.9: détection et reconnaissance visage

3.7 Conclusion

Nous avons présenté dans ce chapitre l'implémentation de l'approche de détection, classification et reconnaissance des images. reconnaissance faciales basée sur les réseaux de neurones convolutionnels, L'approche proposée pour la reconnaissance faciale fonctionne pour de Moyenne résolutions (les images sont de taille 244×244), intéressantes dans le cadre de futures extensions de la technique. L'approche utilise une architecture particulière de réseau de neurones convolutionnels. , la reconnaissance proprement dite est effectuée. Les tests effectués sur une base de données créée par le code.

Conclusion générale

L'objectif de ce mémoire est d'implémenter un programme de détection et de reconnaissance faciale en temps réel qui répond à des besoins de sécurité en général. L'application en question doit répondre à des exigences de rapidité et de robustesse des résultats.

L'application implémentée est basée sur le modèle Resnet50 pour la détection des visages et leurs alignements et la reconnaissance de visage et le modèle VGG-16 aussi pour les mêmes objectifs. Les raisons de ce choix est double. D'une part, les modèles exploités dans notre programme sont imposés par l'encadreur, et d'autre part, les articles dans la littérature ont bien démontré l'avantage de ces modèles par rapport à d'autres modèles. La combinaison et la comparaison des modèles Resnet50 et VGG-16 reprend les avantages de réseaux classiques tout en améliorant grandement sa vitesse.

Le programme ainsi développée répond aux objectifs du projet et ses performances en termes de taux de reconnaissance restent à améliorer

Ce mémoire de master 2 nous a permet de :

- ❖ Approfondir nos connaissances théoriques et pratique déjà acquises, maitriser les nouvelles techniques et compléter notre initiale pour atteindre ainsi un niveau de perfection supérieur et de pouvoir apprendre d'autres nouveautés dans les différents domaines de la science en générale et de l'informatique en particulier.
- ❖ Construire des savoirs et des savoir-faire dans le domaine de traitement d'images, la détection et la reconnaissance.

❖ Comprendre les différentes étapes pour créer des algorithmes de deep learning

Un prolongement de ce travail pourrait s'axer sur plusieurs points à savoir :

- Améliorer l'architecture afin de la rendre plus légère
- Effectuer le modèle de reconnaissance

Enfin, concevoir un programme de reconnaissance d'un visage hétérogène ou non contrôlée. En effet, Aujourd'hui, les techniques conventionnelles de reconnaissance faciale ont atteint une performance quasi-parfaite dans un environnement fortement contraint où la pose, l'éclairage, l'expression faciale et d'autres sources de variation sont contrôlées. Cependant, ces méthodes sont souvent destinées aux domaines d'application limités parce que les environnements de prise des visages ne sont pas idéals dans les cas pratiques

BIBLIOGRAPHIE

- [1] <https://www.lebigdata.fr/machine-Learning-et-big-data>
- [2] international University,, <https://www.supinfo.com/articles/single/6041-machine-Learning-introduction>
- [3] <https://medium.com/@bobkrc/machine-learning-apprentissage-supervis%C3%A9-ou-non-supervis%C3%A9-bced5be4fd7f>
- [4] Mèmoire de Master 2, OULMI Mehdi - KALOUNE Salim, Classification d'objets avec le Deep Learning, Université Akli Mohand Oulhadj de Bouira 2017/2018
- [5] Par Michelle Knight le 30 juillet 2020, <https://www.dataversity.net/what-is-Deep-Learning/>
- [6] Par Keith D. Foote ,7 février 2017 <https://www.dataversity.net/brief-history-Deep-Learning/#>
- [7] Li Liu^{1,2} · Wanli Ouyang³ · Xiaogang Wang⁴ · Paul Fieguth⁵ · Jie Chen² · Xinwang Liu¹ · Matti Pietikäinen² International Journal of Computer Vision Deep Learning for Generic Object Detection Accepted: 26 September 2019
- [8] <https://www.futura-sciences.com/tech/definitions/intelligence-artificielle-Deep-Learning-17262/>
- [9] <https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464>
- [10] Yann LeCun^{1,2}, Yoshua Bengio³ & Geoffrey Hinton Université de Montréal Article in Nature · May 2015
- [11] <https://www.researchgate.net/publication/332781107>
- [12] Li Liu^{1,2} · Wanli Ouyang³ · Xiaogang Wang⁴ · Paul Fieguth⁵ · Jie Chen² · Xinwang Liu¹ · Matti Pietikäinen² International Journal of Computer Vision Deep Learning for Generic Object Detection Accepted: 26 September 2019
- [13] Farhana Sultana¹ , Abu Sufian¹ , Paramartha Dutta², Review of Object Detection Algorithms using CNN, ICCDC 2019,HIT,Haldia
- [14] <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>
- [15] <http://cs231n.github.io/convolutional-networks>
- [16] <https://medium.com/@junehaoching/beginners-guide-to-understanding-convolutional-neural-networks-5209e5d9f717>

[17] https://jpis.org/ViewImage.phpType=F&aid=612629&id=F1&afn=1150_JPIS_48_2_114 &fn=jpis-48-114-g001_1150JPIS

[18] Naciri H., Chaoui N, Conception et Réalisation d'un système automatique d'identification des empreintes digitales, Mémoire de PFE, Université de Tlemcen, 2003.

[19] Rafael C.Gonzalez & Richard E.Woods , Digital Image Processing, Pearson Education Inc, 2008.

[20] Shi, Yichun et Jain, Anil K,' DocFace +: ID Document à Selfie Matching ',Transactions IEEE sur la biométrie, le comportement et la science de l'identité', 1, IEEE ,(2019).

[21] Adel Oulefki, ' Defense of accreditation to supervise research 'Summary of research and pedagogical activities' ,(2018), https://www.researchgate.net/publication/324991823_Defense_of_accreditation_to_supervise_research_Summary_of_research_and_pedagogical_activities.

[22]: Rouse,M 'access control', (2018)