

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique
Université Akli Mohand Oulhadj - Bouira -
Tasdawit Akli Muḥend Ulḥağ - Tubirett -



وزارة التعليم العالي والبحث العلمي
جامعة أكلي محمد أولحاج
- البويرة -

Faculté des Sciences et des Sciences Appliquées

كلية العلوم والعلوم التطبيقية

Référence :/MM/2021

المرجع :/م م / 2021

Mémoire de Master

Présenté au

Département : Génie Électrique

Domaine : Sciences et Technologies

Filière : Télécommunications

Spécialité : Systèmes des Télécommunication

Réalisé par :

KHELLAL Soulaf

Et

ZANOUN Nesrine

Thème

Détection de masque en utilisant Les CNN

Soutenu le : **02/07/2023**

Devant la commission composée de :

Mr : CHELBI Salim	M.C. A Univ. Bouira	promoteur
Mr : Benouicha Karim	M.C. A Univ. Bouira	Président
Mr : AliMohad Abdennour	M.C. B Univ. Bouira	Examineur



نموذج التصريح الشرفي الخاص بالالتزام بقواعد النزاهة العلمية لإنجاز بحث.

انا الممضي اسفله،

السيد(ة).....عزالدين سلال.....الصفة: طالب، استاذ، باحث.....
الحامل(ة) لبطاقة التعريف الوطنية:.....87902611716.....والصادرة بتاريخ.....2018.03.18.....
المسجل(ة) بكلية: العلوم و العلوم التطبيقية..... قسم:.....الهندسة
الكهربائية.....

والمكلف(ة) بإنجاز اعمال بحث(مذكرة، التخرج، مذكرة ماستر، مذكرة ماجستير، اطروحة دكتوراه).

عنوانها:.....Detection du nuage en
utilisant les CNN.....

تحت إشراف الأستاذ(ة):.....تشيبي سليم.....
أصرح بشرفي اني ألتزم بمراعاة المعايير العلمية والمنهجية الاخلاقيات المهنية والنزاهة الاكاديمية
المطلوبة في انجاز البحث المذكور أعلاه.

التاريخ:.....2023.07.03.....

توقيع المعني(ة).....Souley.....

رأي هيئة مراقبة السرقة العلمية:



النسبة:.....20.....%
itin:

الامضاء:



نموذج التصريح الشرفي الخاص بالالتزام بقواعد النزاهة العلمية لإنجاز بحث.

انا الممضي اسفله،

السيد(ة).....
الحامل(ة) لبطاقة التعريف الوطنية: 11 8082597 والصادرة بتاريخ 20 07 02
المسجل(ة) بكلية: العلوم و العلوم التطبيقية..... قسم: الهندسة
الكهربائية.....

والمكلف(ة) بإنجاز اعمال بحث(مذكرة، التخرج، مذكرة ماستر، مذكرة ماجستير، اطروحة دكتوراه).

عنوانها:.....
Detection de virus ou
utilisant les CNA

تحت إشراف الأستاذ(ة):.....
أصرح بشرفي اني ألتزم بمراعاة المعايير العلمية والمنهجية الاخلاقيات المهنية والنزاهة الاكاديمية
المطلوبة في انجاز البحث المذكور أعلاه.

التاريخ: 2023.07.03

توقيع المعني(ة)

رأي هيئة مراقبة السرقة العلمية:

النسبة: % 20

الامضاء:

Remerciement

Avant tout, nous exprimons notre gratitude envers Allah, le Tout-Puissant et Miséricordieux, qui nous a accordé le courage, la volonté et la santé nécessaires pour mener à bien ce travail.

Nous souhaitons adresser nos sincères remerciements à notre encadreur, M. CHELBI Salim, pour son encadrement exceptionnel, sa disponibilité constante, sa patience et sa générosité en partageant ses connaissances.

Grâce à lui, nous avons pu enrichir notre savoir. Nous lui sommes reconnaissantes pour ses qualités humaines et pédagogiques, sa rigueur et ses conseils constructifs, sans lesquels notre travail n'aurait pas été aussi abouti et réalisé.

Nous exprimons également notre vive gratitude envers les membres du jury pour l'intérêt qu'ils ont porté à notre travail, en acceptant de l'examiner et de le compléter par leurs suggestions.

Nos remerciements vont également à tous les enseignants qui ont contribué à notre formation, à qui nous exprimons notre profonde gratitude.

Enfin, nous tenons à remercier toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

Dédicaces

Au nom de Dieu, le Tout miséricordieux, le Très miséricordieux

Je souhaite exprimer ma reconnaissance en dédiant humblement ce travail à la personne la plus chère à ma vie, celle qui n'a jamais cessé de me soutenir et de m'encourager à atteindre mes objectifs, ma mère, ma source de vie.

À mon précieux don de Dieu, mon père, à qui je dois ma réussite et à qui je témoigne tout mon respect. Ta bienveillance me guide et ta présence à mes côtés a toujours été ma source de force, merci papa.

À ma moitié, ma sœur Narimane qui a toujours été présente pour moi, que Dieu la protège et lui accorde la santé.

À mes chers frères, Sofiane, Mourad, Lyes, Racim, Gaya et Safir que Dieu vous accorde santé, bonheur et réussite.

À mes chères nièces Maram, Miral, Ayssel et leur mamans, Linda et Mina.

À mon ami Tahar qui a été toujours présent pour m'aider tout au long de ce travail.

À mes meilleurs amis Rabah, Lahna, Nessrine, Dounia, Mélissa, Khalil, Rosa, Bouthaina, Amel, Sara et Sylia qui m'ont encouragée tout au long de ce projet.

À tous mes collègues de la section de STC, en particulier Achouak Amani, Amira, Idir, Yousra, Khaoula et Yasmine.

Khellal Soulaf

Dédicaces

Je dédie ce travail, Fruit de nombreuses années d'étude à :

À mes chers parents pour leur patience,

Ma chère mère. Merci pour tes conseils, tes Sacrifices, ton soutien et tes encouragements.

Papa ma Gratitude ne suffit pas à exprimer ce qu'il mérite pour

Tous ses sacrifices depuis ma naissance,

Pendant mon enfance et même à l'âge adulte

À mes frères Mohamed, Abdesslam.

À tous mes professeurs et à tous ceux qui se sont engagés

Dans ces modestes travaux.

À tous mes amis et spécialement Soulef, Melissa, Lydia, Yasmine, Amani,

Wissam, Sabrine, Kenza, Amira, Bouthaina, Yousra, Khalil, Idir .

Qui m'ont soutenu dans l'accomplissement de cet Humble travail

Et sans oublier Tahr qui nous a beaucoup aidé

À toute ma famille Zanoun.

Et À tous qui m'ont aidé de près ou de loin pour la réalisation de ce travail.

Zanoun Nesrine.

Résumé

L'apprentissage en profondeur ou le Deep Learning est une branche de l'apprentissage automatique qui est basé sur l'entraînement à partir de grandes quantités de données. Les réseaux de neurones convolutionnels ont obtenu des succès remarquables dans le domaine de la classification d'images offrant des performances exceptionnelles.

Ce projet de fin d'études présente l'utilisation des réseaux de neurones convolutionnels CNN que nous avons implémenté pour la détection du port du masque en utilisant un ensemble de données, et cela en utilisant le logiciel de programmation Matlab.

Les expériences adaptées sur le modèle proposé avec différentes architectures en variant aussi la taille de la couche convolution révèle que le modèle avec 11 couches ayant la taille des filtres respectivement (5*5,3*3) atteind une précision de 94.17%.

Mots clés : Deep Learning, masque, CNN, détection du port de masque

La table des matières

Remerciement	II
Dédicaces	V
Résumé.....	VII
La table des matières.....	VIII
Liste des figures	XI
Liste des tableaux.....	XII
Liste des Acronymes.....	XIII
Introduction générale	01

Chapitre 1 : Traitement d'image et détection du visage

1. Introduction :	18
2. Définition de l'image :	18
3. L'acquisition de l'image :	18
4. Caractéristiques de l'image :	19
4.1. Histogramme :	19
4.2. Pixel :	19
4.3. Dimension :	19
4.4. Résolution :	19
5. Type des images :	20
5.1. Les images en niveaux du gris :	20
5.2. Les images couleurs :	20
6. Formats de fichiers d'images :	20
6.1. Groupe mixte d'experts photographiques (JPEG) :	21
6.2. Format d'échange graphique (GIF) :	21
6.3. Graphique réseau portable (PNG) :	21
6.4. Carte binaire Windows (BMP) :	21
6.5. Format de fichier d'image balisé (TIFF) :	21
7. Quelques traitements d'images :	21
7.1. Binarisation :	21
7.2. Segmentation :	22
7.3. Filtrage :	22
. Les filtres linéaires :	22
. Les filtres non linéaires :	23
8. Détection du visage :	23

9. Méthodes de détection du visage :	24
9.1. Approche basée sur les connaissances acquises :	25
9.2. Approche basée sur les caractéristiques invariantes :	26
9.3. Approche basée sur l'appariement de gabarit :	26
9.4. Approche basée sur l'apparence :	26
10. Les problèmes de détection du visage :	26
10.1. La pose :	27
10.2. L'échelle :	27
10.3. L'occlusion :	27
10.4. Illumination :	28
11. Conclusion :	28

Chapitre 2 : Deep Learning

1. Introduction :	30
2. Définition :	30
3. Historique du Deep Learning:	30
4. Performances et domaines d'application du Deep Learning :	31
4.1. La Vision par ordinateur :	31
4.2. Le Traitement du langage naturel :	31
4.3. La Robotique :	31
4.4. Soins de santé :	31
4.5. La Finance :	32
5. Réseaux neuronaux convolutionnels :	32
6. Principe d'architecture d'un réseau de neurone convolutionnel :	32
7. Les blocs de construction :	33
7.1. Couche de convolution (CONV) :	33
7.2. Couche de pooling :	33
7.3. La couche de correction :	34
7.4. Couche entièrement connectée (fully-connected) :	34
8. Choix des paramètres d'un réseau CNN :	35
8.1. Le nombre de filtres :	35
8.2. La taille du filtre :	35
8.3. La forme de Max pooling :	35
9. Les modèles des réseaux CNN :	35
9.1. LeNet :	35
9.2. AlexNet :	36
9.3. VGGNet :	36

9.4. GoogLeNet :	37
9.5. ResNet :	37
10. Conception d'un réseau CNN :	37
11. Avantage du CNN :	38
12. Conclusion :	38

Chapitre 3 : Applications et résultats

1. Introduction :	40
2. Présentation d'outils du développement :	40
2.1. Choix du langage et du logiciel pour l'implémentation :	40
2.2. Environnement de travail :	40
3. Description du système :	40
3.1. La détection de visage :	41
. La méthode de viola Jones :	42
. Les images de test :	42
3.2. Classifications des images :	44
. Phase d'Entraînement :	45
.la phase de test :	45
4. Présentation du système :	46
4.1. La base de données :	46
5. L'architecture du réseau choisi :	47
6. Résultats obtenus et discussions :	48
6.1. Graphe de précision et d'erreur :	48
6.2. Matrice de confusion :	49
7. Influence nombre de couche et taille des filtres :	51
7.1. Modèle CNN à 11 couche et tailles des filtres différents :	51
. Pour une taille du filtre (11*11,5*5) :	51
.Pour une taille de filtre(5*5,3*3) :	53
8. Tableau de comparaison des résultats	54
9. Conclusion :	55
Conclusion générale	57
Références	58
Résumé	61

Liste des figures

Figure 1. 1:Histogramme d'une image au niveau du gris	19
Figure 1. 2:La résolution d'une image.....	20
Figure 1. 3:Image au niveau du gris.....	20
Figure 1. 4:Binarisation d'une image.....	22
Figure 1. 5:Détection du visage.	24
Figure 1. 6:Quelques types de méthodes de détection du visage.....	25
Figure 1. 7:Des visages typiques de la méthode basée sur les connaissances.	25
Figure 1. 8:Une variation de pose pour un visage.	27
Figure 1. 9:Différentes tailles du visage dans une seule image.	27
Figure 1. 10:Exemple d'un visage occlu par des objets.	28
Figure 1. 11:Exemple d'un visage avec une variation de luminosité.	28
Figure 2. 1 : Illustration Deep-Learning.	30
Figure 2. 2:Exemples de modèles de CNN	32
Figure 2. 3:Fonctionnement d'une couche de convolution	33
Figure 2. 4:Exemple d'une opération dans la couche pooling.....	34
Figure 2. 5:Exemple d'une correction d'une « feature Map »	34
Figure 2. 6:Le premier réseau CNN.....	36
Figure 2. 7:Le réseau AlexNet	36
Figure 2. 8:Le VGGNet.	37
Figure 2. 9:La couche proposée par GoogleNet.	37
Figure 3. 1:Description du système utilisé.....	41
Figure 3.2 : Une image de test comprenant un visage.	42
Figure 3.3 : Une image de test comportant plusieurs visages.....	42
Figure 3.4 : Résultat d'application de la méthode Viola et Jones d'une image comprenant exclusivement un visage.	43
Figure 3.5 : Résultat d'application de la méthode Viola et Jones sur une image contenant plusieurs visages.	43
Figure 3.6 : Détection du nez et la bouche dans le cas d'un seul visage.	44
Figure 3. 7:Détection du nez et la bouche dans le cas de plusieurs.	44
Figure 3. 8 : Modèle de classification.	45
Figure 3.9 : Échantillons de l'ensemble des deux classes avec masque et sans masque.....	46
Figure 3.10 : L'architecture du modèle pour 15 couches.....	48
Figure 3.11 : Le graphe de précision.....	48
Figure 3.12 : Le graphe d'erreur.	49
Figure 3.13 : Matrice de confusion résultante pour 15 couches.	50
Figure 3.14 : L'architecture du modèle pour 11couches.	51
Figure 3.15 : Graphe de précision et d'erreur pour 11 couches.	51
Figure 3.16 : Matrice de confusion pour 11 couches (11*11,5*5).....	52
Figure 3.17 : Graphe de précision et d'erreur pour une taille du filtre (5*5,3*3).....	53
Figure 3.18 : Matrice de confusion pour une taille du filtre (5*5,3*3).....	53

Liste des tableaux

Tableau 3. 1:Tableau de comparaison des résultats. 54

Liste des Acronymes

BMP:	Bitmap
CNN:	Convolutional Neural Network
CMYK:	Cyan,Magenta,yellow,Key
DL:	Deep Learning
GIF:	Graphics Interchange Format
GPU:	Graphics Processing Unit
IA:	Intelligence Artificielle
JPEG:	Joint Photographic Expert Group
KNN:	K-Nearest Neighbor
ML :	Machine Learning
PNG:	Portable Network Graphics
ReLU:	Rectifiedlinear
RGB :	Red,Green and Blue
SVM :	Support Vector Machine
TIFF:	Tagged Image File Format

Introduction

Générale

Introduction générale

Le Covid-19 est une souche de coronavirus qui est apparue à Wuhan le 16 novembre 2019. Depuis lors, elle s'est répandue à l'échelle mondiale. Ce virus peut se propager via de minuscules particules de sécrétions respiratoires appelées aérosols, qui peuvent contaminer les personnes par inhalation. Afin de réduire le risque de contamination, il est vivement recommandé de suivre des mesures d'hygiène strictes, telles que se laver régulièrement les mains, utiliser des désinfectants à base d'alcool, maintenir une distance sociale adéquate et porter un masque chirurgical.

Plusieurs méthodes sont utilisées pour combattre le COVID-19. Parmi ces méthodes, la détection du port du masque à l'aide du traitement d'images. Cette technique vise à contribuer à la prévention et à la lutte contre la propagation de la maladie en identifiant les individus qui ne respectent pas le port un masque.

La détection du visage est une étape primordiale et essentielle avant d'entamer la reconnaissance faciale, de manière générale, les méthodes de reconnaissance faciale peuvent être regroupées en trois catégories : les méthodes globales qui utilisent des techniques linéaires et non linéaires, les méthodes locales qui se basent sur des approches telles que le KNN et le SVM, et les méthodes hybrides qui imitent le fonctionnement d'un système de perception humain.

Au cours des dernières années, le "Deep Learning", considéré comme l'une des techniques d'apprentissage les plus sophistiquées au monde. Il a démontré des performances supérieures aux algorithmes classiques pour résoudre divers problèmes tel que le domaine de la détection faciale ou il a obtenu d'excellents résultats. Il se base sur l'apprentissage profond avec un traitement de grandes quantités de données afin d'extraire efficacement le maximum d'informations pour prendre des décisions.[1]

Notre recherche se concentrera sur l'un des algorithmes qui se démarque par sa performance dans le domaine du Deep Learning : les réseaux de neurones convolutionnels (CNN). Ces derniers représentent un modèle de programmation intense qui excelle notamment dans la reconnaissance d'images, attribuant automatiquement à chaque donnée fournie en entrée une étiquette correspondant à sa classe d'appartenance.

Introduction générale

Notre travail est divisé en trois parties suivante :

- Dans le premier chapitre, nous présenterons en premier lieu quelques notions du traitement d'images ensuite nous aborderons les méthodes de la détection des visages, ainsi les problèmes qui empêchent ces techniques d'avoir des meilleurs résultats.
- Le deuxième chapitre se concentrera sur l'apprentissage profond ou bien le Deep Learning, exactement les réseaux de neurones convolutionnels (CNN) qui sont l'objet de ce travail, avec une explication de leurs architectures via différentes couches.
- Le troisième et dernier chapitre sera dédié à la partie pratique, nous présenterons le processus que nous avons implémenté en utilisant le langage de programmation Matlab.

Chapitre 1

Traitement d'image et détection du visage.

1. Introduction :

Le traitement d'image est un domaine très vaste, connu depuis longtemps et encore reconnu aujourd'hui, au cours des dernières décennies, il a connu une évolution significative. Il est un ensemble de méthodes et de techniques permettant de modifier une image afin d'améliorer ou d'extraire des informations.

Les systèmes d'analyse faciale, y compris la détection et la reconnaissance faciale, sont actuellement l'une des applications les plus pertinentes. En effet, elles sont utilisées en vidéosurveillance, biométrie, robotique, indexation d'images ou de vidéo.

Dans ce chapitre nous allons présenter dans la première partie l'image numérique en générale, ses caractéristiques et ses différents traitement effectués. Dans la deuxième partie nous expliquons le principe et les différentes méthodes de détection du visage.

2. Définition de l'image :

Une image est une présentation visuelle d'un objet, d'un paysage, d'un concept ou d'une idée. Elle peut être produite en utilisant divers médias tels que la peinture, la photographie, le dessin, ...etc. L'image est couramment utilisée pour transmettre des idées, des sentiments ou des informations dans un contexte personnel, artistique ou professionnel.

Une image numérique est une composition de pixels (ou points) contenant des informations sur la couleur et la luminosité des points exacts de l'image et peut être facilement modifiée, retouchée, transformée et distribuée à l'aide d'un logiciel.

3. L'acquisition de l'image :

Est le processus de transformation d'une source visuelle en une représentation numérique de l'image, qui peut être utilisée pour une variété de fins telles que la visualisation, l'analyse et la modification.

En reconnaissance faciale par exemple, on peut utiliser des capteurs 3D pour résoudre les problèmes d'installation. Le processus de reconnaissance de visage 3D commence par l'acquisition d'une image tridimensionnelle du visage de la personne à l'aide d'un capteur de profondeur, tel qu'un scanner 3D ou une caméra infrarouge [1].

4. Caractéristiques de l'image :

Il existe 4 éléments essentielles :

4.1. Histogramme :

Un histogramme représente la répartition des pixels selon le niveau de gris. Il fournit diverses informations telle que les statistiques d'ordre, l'entropie et peut être utilisé pour détecter des objets [2].

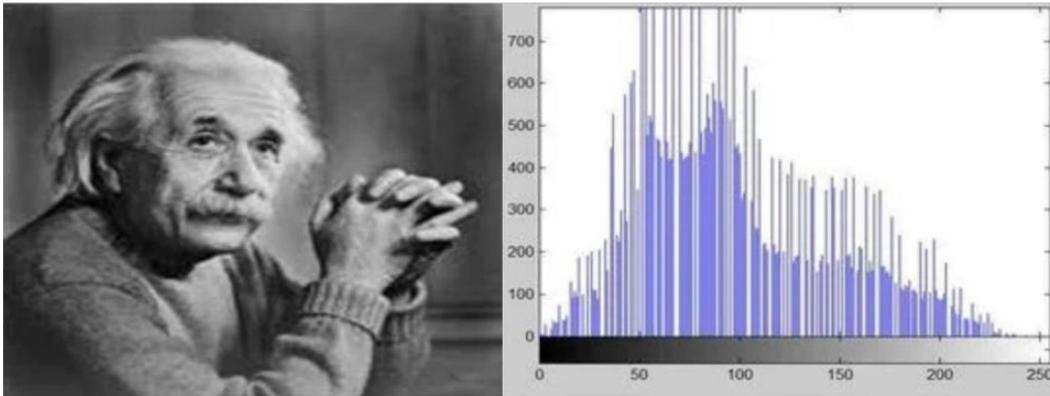


Figure 1. 1:Histogramme d'une image au niveau du gris.

4.2. Pixel :

Le pixel est l'unité de base de l'image numérique, il correspond le plus petit point sur l'écran. Chaque pixel contient des informations sur la couleur, la luminosité et la position sur l'image. Plus il y a de pixels dans une image, plus celle-ci est détaillée et précise.

4.3. Dimension :

La dimension d'une image numérique correspond à sa taille en pixels. Elle est déterminée par le nombre de pixels sur la largeur et la hauteur de l'image. Par exemple, une image de 1920 pixels de largeur sur 1080 pixels de hauteur est dite avoir une dimension de "1920x1080" pixels...

4.4. Résolution :

La résolution d'une image correspond à la densité de pixels présents dans cette image, exprimée en pixels par pouce. Sa détermination est basée sur les dimensions de l'image et sa qualité d'impression obtenue.



Figure 1. 2: La résolution d'une image.

5. Type des images :

On désigne deux types d'images :

5.1. Les images en niveaux du gris :

Les images en niveaux de gris, également appelées images monochromes, sont des images qui ne contiennent que des nuances de gris, sans aucune couleur. Les niveaux de gris représentent l'intensité de la lumière dans chaque pixel de l'image. Ces images contiennent généralement 256 niveaux de gris. Par convention la valeur zéro représente le noir et la valeur 255 représente le blanc et le nombre 256 représente la quantification de l'image.



Figure 1.3 : Image au niveau du gris.

5.2. Les images couleurs :

Les images en couleur sont des images qui contiennent des informations de couleur pour chaque pixel de l'image, elles sont souvent représentées en utilisant un modèle de couleur, tel que le modèle RGB (rouge, vert, bleu) ou le modèle CMYK (cyan, magenta, jaune, noir). Dans le modèle RGB, chaque pixel est défini par trois valeurs de couleur, représentant la quantité de rouge, de vert et de bleu présente dans ce pixel.

6. Formats de fichiers d'images :

Les formats d'image sont des représentations informatiques d'images. Il existe plusieurs formats :

6.1. Groupe mixte d'experts photographiques (JPEG) :

Le format JPEG est largement utilisé pour les images photographiques en raison de sa capacité à compresser les fichiers tout en conservant une qualité d'image raisonnable.

6.2. Format d'échange graphique (GIF) :

Le format GIF est utilisé pour les images animées ou les petits graphiques. Il utilise une palette de couleurs limitée et peut donc être compressé pour une taille de fichier plus petite.

6.3. Graphique réseau portable (PNG) :

Est un format de fichier d'image qui est largement utilisé sur le web. Il a été développé pour remplacer le format GIF qui a des limitations de licence et de compression. Le format PNG est un format d'image compressé sans perte, ce qui signifie qu'il peut réduire la taille du fichier sans perte de qualité, il prend également en charge la transparence, ce qui signifie que les parties d'une image qui ne sont pas nécessaires peuvent être supprimées pour laisser voir l'arrière-plan. Cela est particulièrement utile pour les logos ou les images avec des formes complexes [3].

6.4. Carte binaire Windows (BMP) :

Est un format de fichier d'image non compressé qui stocke chaque pixel de l'image, ce qui le rend adapté aux images de haute qualité.

6.5. Format de fichier d'image balisé (TIFF) :

Est un format de fichier d'image non compressé qui prend en charge la gestion de couleur en haute qualité et des images sur plusieurs pages.

7. Quelques traitements d'images :

Il existe différents types de traitement d'image, nous allons présenter quelques-uns :

7.1. Binarisation :

La binarisation est une procédure qui transforme une image en une version binaire, où chaque pixel de l'image est représenté par un seul bit, 0 ou 1.

Le processus implique le seuillage de l'image, ce qui signifie définir une valeur seuil qui sépare les pixels en deux classes en fonction de leurs valeurs d'intensité. Les pixels dont les valeurs d'intensité sont supérieures au seuil sont définis sur 1, et ceux dont les valeurs d'intensité sont inférieures ou égales au seuil sont définis sur 0.

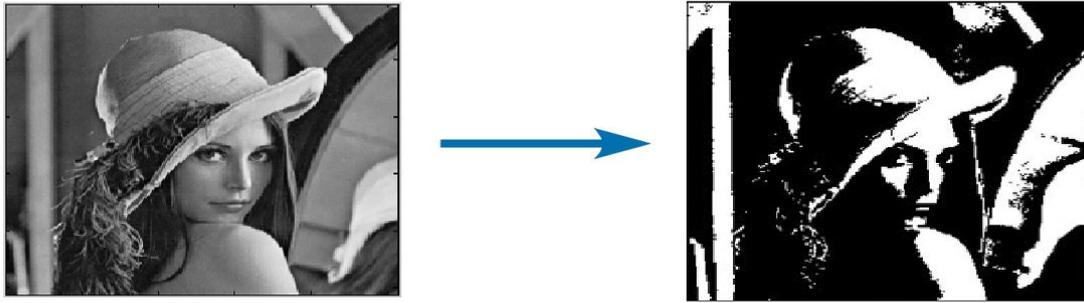


Figure 1.4 : Binarisation d'une image.

7.2. Segmentation :

La segmentation d'image est une étape importante dans de nombreuses applications de traitement d'images consistant à détecter et rassembler les pixels suivant des critères prédéfinis, tels que la couleur, l'intensité ou la forme. Son objectif est de simplifier une image et de la rendre plus facile à analyser, en regroupant des pixels similaires en régions significatives [4].

7.3. Filtrage :

Le filtrage d'image est une opération de traitement effectuée sur une image, qui implique l'application d'un filtre ou d'un noyau à une image pour obtenir une version modifiée de l'image originale. Les filtres sont souvent utilisés pour améliorer la qualité d'une image, supprimer le bruit ou les artefacts, améliorer la netteté ou les contours de l'image.

Il existe différents types de filtres utilisés pour le filtrage d'image, ceux-ci peuvent être regroupés en deux grandes catégories, à savoir : filtres linéaires et non linéaires.

. Les filtres linéaires :

Les filtres linéaires sont des outils de traitement de signal et d'image qui permettent d'effectuer une convolution entre une image ou un signal.

Les filtres linéaires sont dits "linéaires" car l'opération de convolution est une opération linéaire, c'est-à-dire que la sortie est une combinaison linéaire des entrées. Cela signifie que les propriétés de linéarité peuvent être utilisées pour simplifier les calculs et les analyses.

Il existe plusieurs filtres linéaires, dont certains sont présentés ici :

a. Les filtres passe-bas :

Sont utilisés pour réduire les hautes fréquences de l'image, ce qui conduit à une image floue. Ils sont souvent utilisés pour réduire le bruit dans l'image ou pour réduire les détails fins dans l'image.

b. Les filtres passe-haut :

Sont utilisés pour ajouter les hautes fréquences de l'image, ce qui conduit à une image plus nette. Ils sont souvent utilisés pour améliorer les contours ou les détails fins dans l'image.

. Les filtres non linéaires :

Ils sont conçus pour résoudre les problèmes rencontrés par les filtres linéaires, notamment la perte de détails dans les contours. Leur fonctionnement est similaire à celui des filtres linéaires, car il nécessite le remplacement de la valeur de chaque pixel par la valeur d'une fonction calculée dans son voisinage. La distinction réside dans le fait que cette fonction n'est plus linéaire, mais plutôt arbitraire ou de nature quelconque.

On peut mentionner parmi les filtres non linéaires les plus connus :

a. Le filtre médian :

Ce filtre est utilisé pour la suppression du bruit dans l'image. Ils remplacent chaque pixel de l'image par la médiane des pixels voisins.

b. Le filtre maximum :

Ce filtre remplace la valeur de pixel de milieu par la valeur maximale des pixels voisins.

c. Le filtre minimum :

Ce filtre remplace la valeur de pixel de milieu par la valeur minimum des pixels voisins [5].

8. Détection du visage :

La détection du visage est la principale étape de la reconnaissance faciale, c'est une technique qui permet de localiser et identifier la région du visage sur une image ou une vidéo. Des différentes applications qui utilisent ce genre des systèmes telles que le codage vidéo, la récupération d'image basée sur le contenu, la vidéo-conférence, la surveillance et les interfaces homme-machine et la photographie. Plusieurs méthodes ont été proposées et cela en utilisant des techniques de haut niveau de reconnaissance faciale [6].



Figure 1.5 : Détection du visage [7].

9. Méthodes de détection du visage :

La première étape de la détection du visage se résume à prendre une capture d'image à partir d'une vidéo ou sur une photo et les caractéristiques du visage sont identifiées et extraites selon la méthode choisie. Dans les paragraphes suivants, nous allons examiner diverses méthodes de la détection du visage à partir d'une image. Ces méthodes sont généralement classées en deux grandes catégories :

La première catégorie inclue à savoir : Les différentes approches qui sont basées sur les connaissances acquises, sur l'appariement de gabarit et sur l'approche basée sur les caractéristiques invariantes (analyse bas niveau), dite approches basées sur les traits du visage.

La deuxième catégorie regroupe les approches basées sur l'image, également appelées approches basées sur l'apparence [8].

En résumé, il existe quatre principales techniques de détection du visage, comme illustré dans la figure ci-dessous :

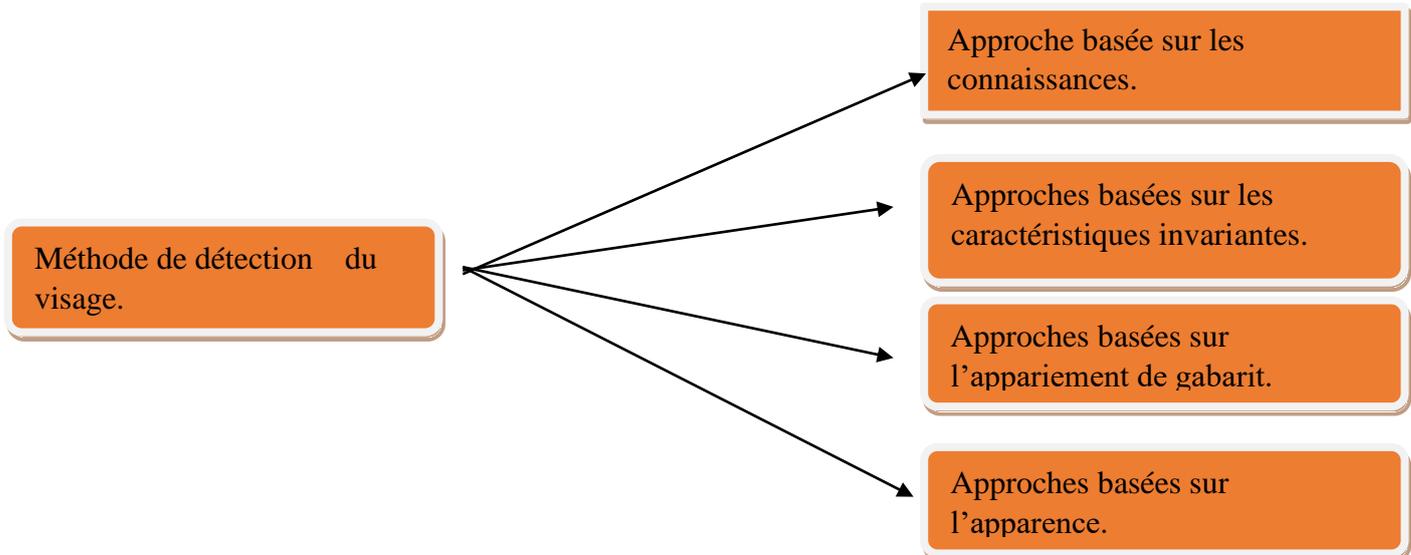


Figure 1.6 : Quelques types de méthodes de détection du visage.

9.1. Approche basée sur les connaissances acquises :

Ces techniques sont élaborées à partir des règles fondamentales déduites de la connaissance des visages humains, ce qui permet de créer des règles simples pour décrire les traits du visage et leur relation. Par exemple, un visage de face se compose d'un nez, d'une bouche et de deux yeux symétriques dans une image. Les relations entre ces traits peuvent être définies en termes de position et de distance relative. Ces méthodes impliquent d'abord l'extraction des caractéristiques faciales dans l'image d'entrée, puis l'identification des visages potentiels en se basant sur les règles préétablies [9].

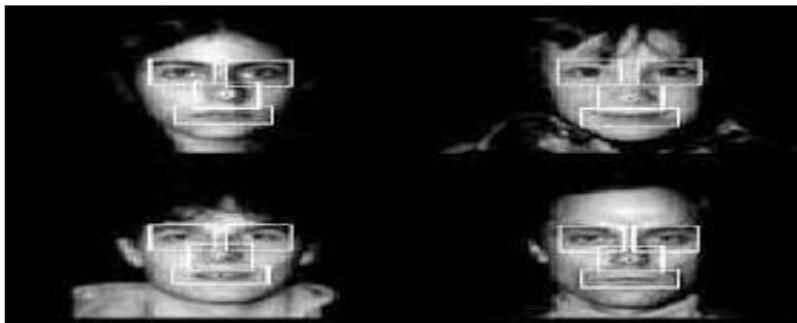


Figure 1.7: Des visages typiques de la méthode basée sur les connaissances [10].

L'application de cette approche est confrontée à plusieurs contraintes, notamment la difficulté de créer un ensemble de règles rigoureuses qui puissent être appliquées à toutes les formes sous différentes conditions, ainsi que la complexité de la détection du visage lorsque celui-ci est placé sur un arrière-plan complexe. En outre, cette technique est sujette à des fausses détections multiples.

9.2. Approche basée sur les caractéristiques invariantes :

L'objectif de cette méthode est de détecter les caractéristiques structurelles du visage, même lorsqu'il est présenté sous différentes conditions d'éclairage positions ou angles de vue, la qualité de l'image avec cette technique peut être diminuée en raison de facteurs tels que le bruit, ou l'éclairage.

9.3. Approche basée sur l'appariement de gabarit :

C'est une méthode qui consiste à apprendre à partir d'exemples standard de visages ou d'images frontales contenant des visages. La procédure implique la corrélation entre les images d'entrée et les exemples enregistrés (gabarits), et le résultat final permet de déterminer la présence ou l'absence d'un visage. Il existe deux types de corrélation selon le type de gabarits utilisés, à savoir : les gabarits de visages prédéfinis et les modèles déformables [11].

9.4. Approche basée sur l'apparence :

En général, ces méthodes utilisent des techniques d'apprentissage automatique pour la détection, elles considèrent que le problème de la détection de visage comme un problème de classification binaire (visage ou non-visage). Bien que ces méthodes soient rapides à exécuter, elles nécessitent un temps d'entraînement considérable.

Les méthodes faisant partie de cette approche ont obtenu des résultats satisfaisants en comparant avec les trois autres types de procédés cités précédemment. Quelques exemples de ces méthodes :

- La méthode de Rowley et al. Qui s'appuie sur les réseaux de neurones.
- L'algorithme de Viola et Jones, développé par Paul Viola et Michael Jones [12].
- La méthode des "Eigenfaces" (visages propres).
- La méthode de Schneiderman et Kanade qui est basée sur un classifieur de Bayes naïf.

10. Les problèmes de détection du visage :

La détection de visage peut se révéler complexe en raison de divers facteurs tels que les parties du visage cachées par des occlusions, les variations d'expression faciale, les variations de luminosité, les changements de pose et d'angle de vue, la présence d'accessoires comme les lunettes, les barbes ou le maquillage etc....

On va expliquer quelques-uns de ces facteurs, en commençant par :

10.1. La pose :



Figure 1.8 : Une variation de pose pour un visage [13].

Les capacités d'un système de détection de visages peuvent être significativement réduites en présence de variations de pose.

10.2. L'échelle :

Les visages peuvent apparaître à différentes tailles (échelles) dans une image.



Figure 1. 9:Différentes tailles du visage dans une seule image [14].

10.3. L'occlusion :

La détection de visages est confrontée à un autre problème appelé occlusion. En effet, la présence de lunettes, de barbes, ou d'écharpes peut modifier l'apparence du visage.



Figure 1.10: Exemple d'un visage occlusion par des objets [15].

10.4. Illumination :

Les variations causées par les différences de luminosité peuvent être plus significatives que les variations entre les caractéristiques faciales individuelles.



Figure 1.11: Exemple d'un visage avec une variation de luminosité [16].

11. Conclusion :

Dans ce chapitre, nous avons abordé plusieurs concepts liés à l'image numérique, tels que les types d'images, leurs caractéristiques, ainsi que quelques formats de fichiers couramment utilisés. Nous avons également expliqué et présenté quelques traitements d'images effectués.

Dans la deuxième partie de ce chapitre, nous avons abordé l'objectif de notre étude qui est la détection de visage. Nous avons cité de manière générale les différentes méthodes utilisées pour la détection ainsi que les principales contraintes rencontrées lors de la détection qui conditionne le résultat attendu à savoir : la disposition du visage, l'occlusion, la variation de l'échelle et la luminosité du visage.

Dans le prochain chapitre, nous aborderons le Deep Learning et les réseaux de neurones convolutionnels.

Chapitre 2

Deep Learning

1. Introduction :

Ces dernières années, les articles scientifiques ont été inondés de nouveaux dictionnaires liés à l'émergence de l'intelligence artificielle dans notre société, nous pouvons citer parmi eux le Deep Learning qui est largement utilisé dans différentes applications. Le Deep Learning permet de concevoir des applications et des systèmes informatiques qui réalisent plusieurs tâches en imitant le comportement humain, telles que la reconnaissance vocale, faciale et le langage naturel arrivant à des très bons résultats [17].

2. Définition :

L'apprentissage profond ou le Deep Learning est l'une des principales technologies de l'intelligence artificielle (IA), qui est conçue comme solution pour résoudre des problèmes complexes utilisés pour plusieurs applications à savoir : la reconnaissance faciale et vocale, le traitement du langage...etc. Son principe repose sur la composition et la formation de réseaux neurones artificiels qui imitent le comportement du cerveau humain, il se compose de plusieurs couches de nœuds interconnectés, nommés neurones [18], [19].



Figure 2.1 : Illustration Deep-Learning [20].

3. Historique du Deep Learning:

Les débuts des réseaux de neurones remontent aux années 1940 et 1950 par les recherches en neurosciences, en effet, la première modélisation d'un neurone par McCulloch et Pitts en 1943 [21]. Ces réseaux ont rapidement été abandonnés car les ordinateurs de l'époque n'étaient pas assez puissants pour gérer de tels réseaux. Il faudra cependant attendre le début des années 2000 pour que

le Deep Learning prenne son essor, grâce à l'augmentation de la puissance de calcul et à l'émergence de nouvelles techniques d'entraînement des réseaux de neurones.

En 2006, le chercheur canadien Geoffrey Hinton a publié un article sur un nouvel algorithme d'apprentissage en profondeur appelé Restricted Boltzmann Machines (RBM). Cette percée a permis la construction de réseaux de neurones plus profonds et plus efficaces qui peuvent résoudre des problèmes complexes tels que la reconnaissance d'images et la compréhension du langage naturel [21].

Depuis lors, le Deep Learning a connu une croissance extraordinaire avec la disponibilité de vastes ensembles de données et l'amélioration constante des algorithmes et des techniques de formation. Révolutionnant ainsi de nombreux domaines tels que la reconnaissance faciale et vocale.

4. Performances et domaines d'application du Deep Learning :

Le Deep Learning s'adapte efficacement aux données, et ses performances augmente avec la quantité de données fournies, tout en automatisant l'extraction de caractéristiques qui était une étape manuelle dans les algorithmes traditionnels, et qui nécessite l'intervention des spécialistes [22].

Ce type d'algorithme forme des réseaux de neurones artificiels à partir de grandes quantités de données pour automatiser des tâches complexes dans différents domaines à savoir [23], [24] :

4.1. La Vision par ordinateur :

Est utilisé pour la reconnaissance d'objets, la segmentation d'images, la détection de visages, la reconnaissance d'écriture manuscrite...etc.

4.2. Le Traitement du langage naturel :

Est utilisé pour la reconnaissance vocale, la traduction automatique, la classification de texte, la génération de texte...etc.

4.3. La Robotique :

Est utilisé pour la planification de trajectoires, la reconnaissance de gestes, la détection d'obstacles...etc.

4.4. Soins de santé :

Est utilisé pour la détection précoce des maladies, la recherche de médicaments, l'étude d'images médicales ...etc.

4.5. La Finance :

Est utilisé pour la prévision des prix des actions, la détection de fraude, la classification de documents financiers...etc.

5. Réseaux neuronaux convolutionnels :

Sont des réseaux de neurones artificiels spécialement conçus pour le traitement de données, ils fonctionnent en appliquant des filtres de convolution sur l'ensemble de l'image pour extraire le maximum de caractéristiques spécifiques, qui sont utilisées par la suite pour effectuer la classification ou la prédiction sur les données d'entrée [25].

La plupart des utilisations des réseaux de neurones convolutionnels se limitent aux données d'image, alors que ces réseaux peuvent également être appliqués à d'autres types de données, tels que les données temporelles, spatiales et spatio-temporelles [26].

Les ConvNets, abréviation de réseaux neuronaux convolutionnels, sont des réseaux de neurones particulièrement adaptés aux tâches de vision par ordinateur, notamment la reconnaissance d'images et d'objets [27].

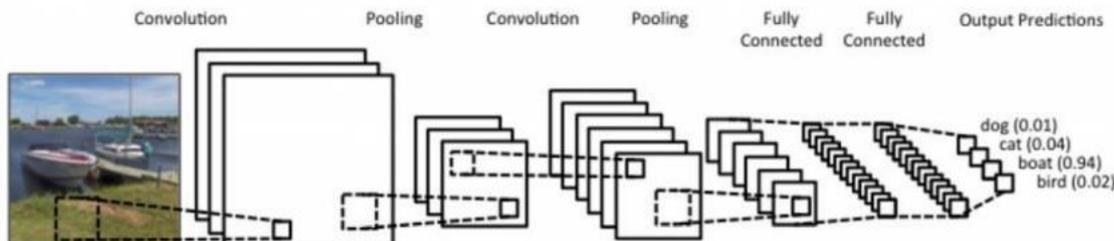


Figure 2.1 : Exemples de modèles de CNN.

6. Principe d'architecture d'un réseau de neurone convolutionnel :

Le principe fondamental de l'architecture d'un réseau de neurones convolutionnel est basé sur l'utilisation de couches de convolution pour extraire les caractéristiques significatives des données d'entrée, chaque couche de convolution contient plusieurs filtres appliquant l'opération de la convolution sur les données d'entrée, produisant ainsi une carte de caractéristiques. Ces dernières, sont ensuite soumises à des couches de sous-échantillonnage pour réduire leur dimension.

En alternant, les couches de convolution et de sous-échantillonnage, une architecture en cascade. Les couches entièrement connectées sont ensuite utilisées pour classifier les données d'entrée en

fonction des caractéristiques extraites. Tout le réseau est entraîné en utilisant un algorithme d'apprentissage supervisé pour minimiser la perte entre les prédictions et les étiquettes de classes réelles [28].

7. Les blocs de construction :

Une architecture de réseau de neurones convolutionnels (CNN) est constituée d'un empilement de couches de traitement qui fonctionnent de manière indépendante afin d'extraire les caractéristiques de l'image et de les classifier :

7.1. Couche de convolution (CONV) :

La couche de convolution joue un rôle fondamental et crucial dans les réseaux de neurones, elle traite les images en entrée pour extraire des caractéristiques [29].

Cette couche produit une carte de caractéristiques qui permet d'identifier les éléments importants de l'image.

Cette étape est réalisée en appliquant un filtre sur l'image, et en effectuant une opération de convolution pour chaque portion de l'image par un balayage. Cette opération implique une multiplication matricielle par éléments suivie d'une somme sur la matrice résultante.

La figure ci-dessous montre le principe de fonctionnement de cette couche :

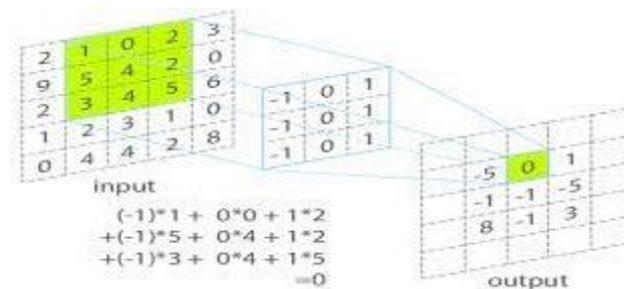


Figure 2.2 : Fonctionnement d'une couche de convolution [30].

7.2. Couche de pooling :

La phase de Pooling permet de réduire la dimension de chaque « Feature Map » ou bien la carte de caractéristiques tout en préservant les informations les plus importantes. Les deux types les plus couramment utilisés sont le «Max Average», qui sélectionne la valeur la plus élevée de chaque «Feature Map» et « Average Pooling » qui consiste à faire la moyenne des valeurs de chaque région.

Le Pooling permet de réduire la taille des représentations d'entrée, ce qui les rend plus faciles à gérer. Il contribue également à la réduction du nombre de paramètres et de calculs dans le réseau, permettant ainsi de limiter le risque de sur-apprentissage.

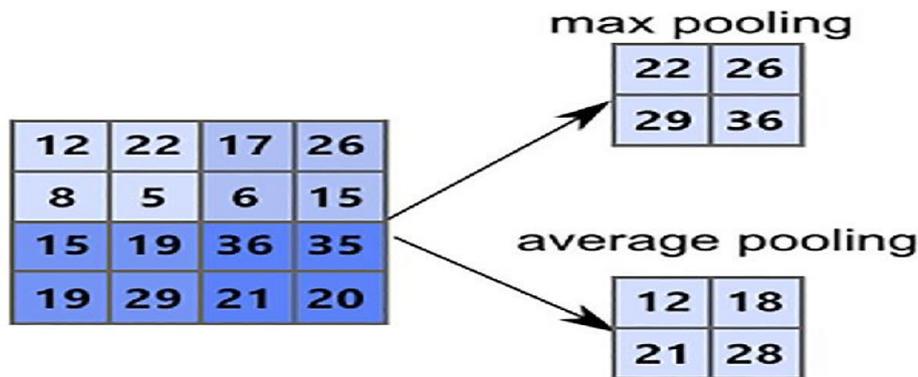


Figure 2.3 : Exemple d'une opération dans la couche pooling [31].

7.3. La couche de correction :

Après chaque opération de convolution, une étape supplémentaire appelée ReLU est appliquée, cette étape signifie Unité Linéaire Rectifiée. C'est une fonction qui s'applique sur chaque pixel de la « Feature map », et remplace toutes les valeurs de pixels négatifs par zéro [19]. La fonction ReLU est donnée par l'équation suivante :

$$F(x) = \max(0, x) \quad [22].$$

L'objectif de l'utilisation de la couche ReLU est d'introduire la notion de non-linéarité dans le modèle.

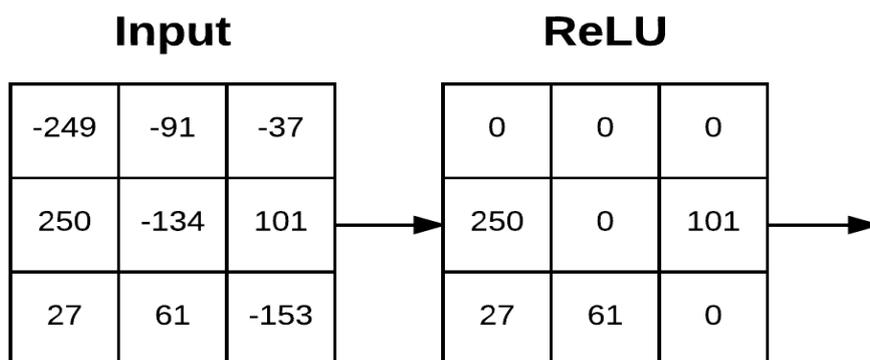


Figure 2. 4 : Exemple d'une correction d'une « feature Map » [32] .

7.4. Couche entièrement connectée (fully-connected) :

La couche « Fully Connected » est nommée ainsi, car elle est entièrement connectée à la sortie de la couche précédente. Cette couche prend en compte tous les neurones de la couche précédente (qu'il s'agisse d'une couche entièrement connectée, de Pooling ou de Convolution), et les connecte à

chacun de ses neurones [19].

8. Choix des paramètres d'un réseau CNN :

Afin d'améliorer les performances du CNN, il est indispensable de spécifier manuellement une série de paramètres, tels que :

8.1. Le nombre de filtres :

Lors de la conception d'un CNN, certains paramètres, tels que le nombre de filtres, doivent être spécifiés manuellement avant l'entraînement. Plus le nombre de filtres est élevé, plus les caractéristiques des images sont extraites avec précision, ce qui améliore les performances du réseau en termes d'extraction de caractéristiques et de classification d'images.

8.2. La taille du filtre :

La taille des filtres dans un réseau de neurones convolutionnel (CNN) est souvent ajustée en fonction du jeu de données utilisé, pour les images de faible résolution, les filtres les plus efficaces ont tendance à être de taille (5x5) pour la première couche. En revanche, pour l'ensembles de données contenant des images naturelles, des filtres plus grands tels que (11x11) ou (15x15) sont couramment utilisés.

8.3. La forme de Max pooling :

Les dimensions du pooling courantes sont généralement de 2x2. Dans les premières couches, pour les entrées de très grande taille, un pooling de 4x4 peut être justifié. Toutefois, l'utilisation de dimensions de pooling plus importantes aura pour conséquence de réduire la dimension du signal, ce qui pourrait conduire à la perte d'informations.

9. Les modèles des réseaux CNN :

Il existe différents modèles de réseaux de neurones convolutionnel (CNN), leurs efficacités dépendent de la tâche à accomplir, du nombre de couches de convolution ainsi que de la structure spécifique adoptée. Nous pouvons citer :

9.1. LeNet :

Ce modèle a été employé pour la reconnaissance de l'écriture manuscrite, il comporte 7 couches, en excluant la couche d'entrée, et utilise des images d'entrée de 32x32 pixels. La capacité à traiter

des images de plus haute qualité nécessite des couches plus grandes et plus profondes en termes de convolutions. Ce qui rend cet algorithme limité par les ressources informatiques disponible [33].

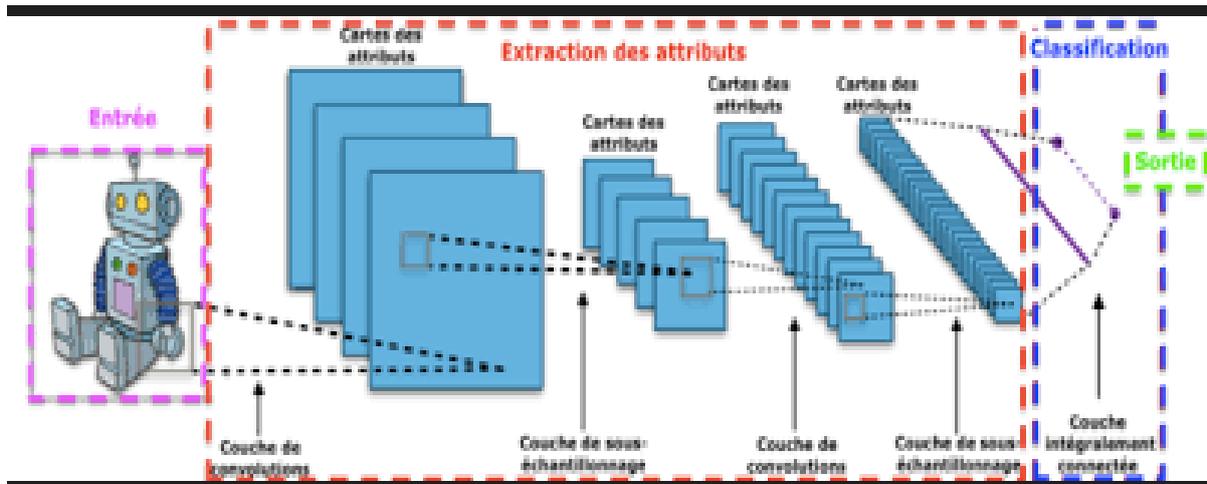


Figure 2.5 : Le premier réseau CNN.

9.2. AlexNet :

Ce réseau est considéré comme l'un des premiers à avoir popularisé les réseaux de neurones convolutionnel, il présentait une architecture très similaire à celle du premier réseau mais plus profond et plus grand.

Il possède cinq couches de convolution et trois couches de pooling., la taille des noyaux de convolution variait en fonction de la couche considérée avec des valeurs allant de 11x11 à 3x3. Le réseau AlexNet contient 60 millions paramètres et 650000 neurones [34].

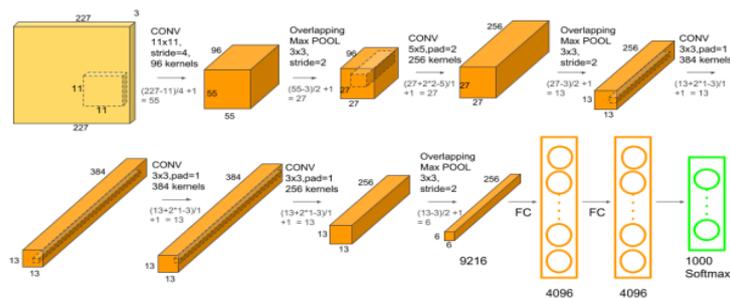


Figure 2.6 : Le réseau AlexNet [34].

9.3. VGGNet :

VGGNet est un modèle de 16 couches convolutives, connu pour son architecture uniforme et sa taille de filtres plus petite (3x3), qui en fait un choix populaire pour l'extraction de caractéristiques d'images. Bien qu'il soit similaire à AlexNet, il est considéré comme plus attrayant en raison de sa structure (uniforme) [35].

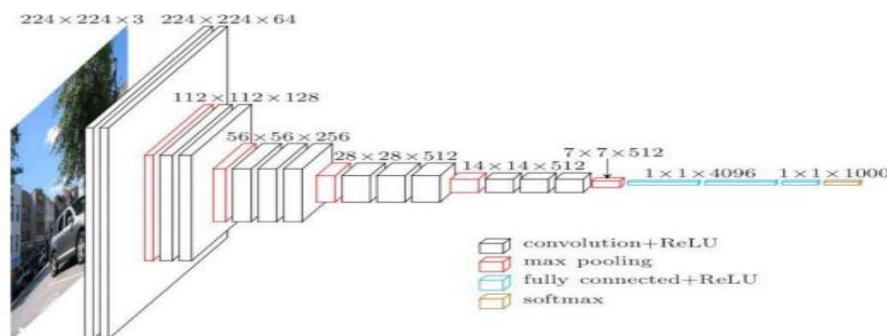


Figure 2.7 : Le VGGNet [34] .

9.4. GoogLeNet :

Un modèle qui a remporté le défi ILSVRC de 2014 a proposé un module Inception basé sur la moyenne globale (AVG pooling) qui a permis de réduire considérablement le nombre de paramètres nécessaires (4 millions contre 60 millions pour AlexNet et 138 millions pour VGG). La version la plus récente de ce réseau est Inception V4 [36].

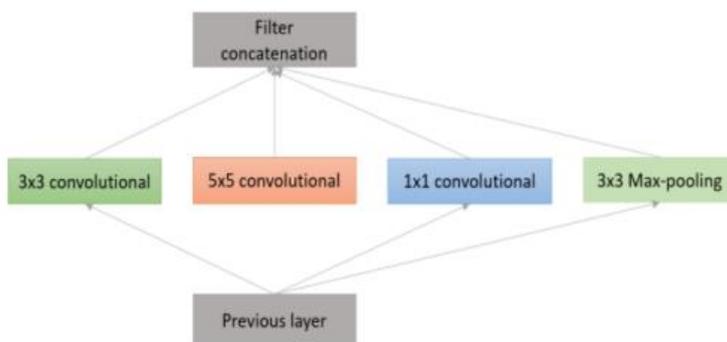


Figure 2.8 : La couche proposée par GoogleNet [34].

9.5. ResNet :

Ce réseau a été couronné vainqueur du défi ILSVRC de 2015. Il se distingue par l'emploi de sauts de connexions, une utilisation intensive de la normalisation par lots et du pooling global AVG [37].

10. Conception d'un réseau CNN :

Créer un réseau de neurones convolutionnel est une tâche difficile et coûteuse, car cela nécessite du matériel et une grande quantité de données. La première étape consiste à concevoir l'architecture du réseau, en déterminant le nombre de couches, leur taille et les opérations matricielles qui les relient. Ensuite, l'entraînement du réseau vise à optimiser les paramètres afin de réduire l'erreur de

classification. Les meilleurs réseaux CNN peuvent prendre plusieurs jours pour s'exécuter, car les unités de traitement graphique fonctionnent sur des centaines de milliers d'images.

11. Avantage du CNN :

Les réseaux de neurones convolutionnels sont très efficaces dans le traitement des images, des vidéos et des données en 3D, ils peuvent prendre des caractéristiques visuelles à partir des données d'entrée sans nécessiter de prétraitement manuel des données. Ces algorithmes réduisent la quantité de mémoire requise d'une manière considérable et améliore les performances du modèle et le rendant plus efficace. [33]

Les CNN peuvent apprendre de manière automatique des caractéristiques applicables à partir des données d'entrée, ce qui peut réduire la nécessité d'une ingénierie manuelle des fonctionnalités. CNN offre la possibilité de calculer automatiquement des cartes d'entités, évitant ainsi aux utilisateurs d'effectuer [19] calculs d'entités lourds.

12. Conclusion :

Au cours de ce chapitre, nous avons abordé le sujet du Deep Learning et ses réseaux de neurones convolutionnels, nous avons donné un aperçu sur les domaines d'applications et leurs avantages. Nous avons ensuite expliqué son architecture avec ses différentes couches à savoir : couche convolution avec ces différents filtres, couche pooling, couche de correction RLU, et enfin la couche entièrement connectée.

A ce stade, nous avons acquis toutes les connaissances nécessaires, dont nous avons besoin pour procéder à une application pratique du concept des réseaux de neurones convolutionnels dans la détection de masques qui sera l'objet du prochain chapitre.

Chapitre 3

Applications et résultats

1. Introduction :

Après avoir examiné les principales techniques de détection du visage dans les chapitres précédents, ainsi que l'apprentissage profond basé sur les réseaux de neurones convolutionnels, le troisième chapitre se concentre sur l'aspect expérimental du travail théorique. Plus précisément, il se focalise sur l'utilisation d'un réseau de neurones convolutionnel (CNN) capable de classifier des images avec des bonnes performances en termes de temps et d'efficacité, à l'aide du logiciel MATLAB.

2. Présentation d'outils du développement :

2.1. Choix du langage et du logiciel pour l'implémentation :

Parmi les différents langages de programmation disponibles, nous avons sélectionné MATLAB (Matrix Laboratory) pour notre application en raison de sa combinaison de langage de haut niveau et d'environnement de travail intégré.

MATLAB est un logiciel de calcul numérique et de programmation développé par MathWorks. Il est principalement utilisé dans les domaines de l'ingénierie et des mathématiques, pour la simulation, l'analyse et la visualisation de données.

MATLAB est doté d'un langage de programmation haut niveau qui permet une manipulation aisée de données, de matrices et de fonctions mathématiques. Il dispose également d'une interface utilisateur graphique pour faciliter l'interaction avec l'utilisateur [38].

2.2. Environnement de travail :

- Un ordinateur portable DELL DESKTOP-P2IMR6L.
- Windows 10 Professionnel.
- Processus Intel(R) Core (TM) i7-6650U CPU @ 2.20GHz 2.21 GHz.
- Mémoire installée (RAM) 8,00 Go.
- Type du système : Système d'exploitation 64 bits, processeur x64.

3. Description du système :

Notre travail vise principalement à atteindre l'objectif principal suivant : détecter les visages et de classer les images en fonction de deux catégories : avec masque, sans masque. Pour ce faire, notre travail est divisé en deux parties distinctes :

- ✓ La première partie consiste à détecter des visages en utilisant l'algorithme de Viola-Jones et extraire les caractéristiques correspondantes.
- ✓ La deuxième partie consiste à classifier les visages en utilisant l'approche CNN.

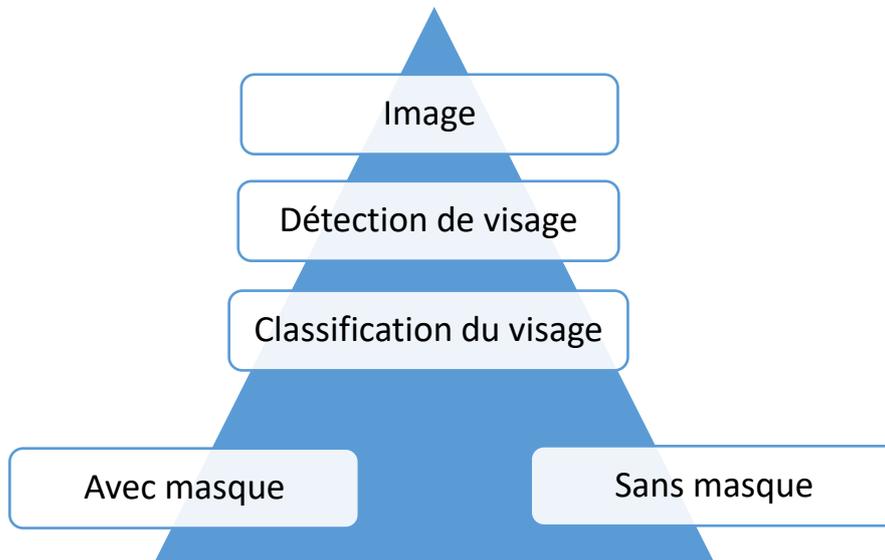


Figure 3. 1:Description du système utilisé.

Pour assurer l'efficacité du système de détection de visage, la méthode utilisée pour localiser le visage dans une image est primordiale. Nous avons choisi d'utiliser l'algorithme Viola-Jones pour réaliser cette phase de détection. Si le visage est correctement détecté, le système passe à l'étape d'extraction des caractéristiques. Enfin, la classification est réalisée à l'aide de la technologie de réseau de neurones convolutionnels (CNN) dans le but d'obtenir une décision en fonction des caractéristiques extraites.

3.1. La détection de visage :

L'objectif principal de l'algorithme de détection faciale, est d'identifier et de localiser un ou plusieurs visages présents dans une image.

Afin de réaliser cette tâche vitale de détection de visages, nous avons opté pour l'approche simplifiée de l'algorithme Viola-Jones, qui utilise le classificateur "**vision.CascadeObjectDetector**" disponible dans MATLAB. Cependant, il est important de noter que cette méthode a tendance à détecter principalement les visages orientés vers l'avant. Par conséquent, il est préférable d'obtenir la meilleure prise de vue possible de l'image afin d'optimiser les résultats de détection.

. La méthode de Viola Jones :

Est une technique de détection d'objets largement utilisée dans le domaine de l'informatique visuelle, en particulier pour la détection de visages. Elle a été développée par Paul Viola et Michael Jones en 2001 et a été révolutionnaire à l'époque pour sa capacité à détecter rapidement et efficacement des objets dans des images en temps réel [12].

Le concept de la méthode de Viola-Jones réside dans l'utilisation de caractéristiques simples mais efficaces pour représenter des objets, qui sont connues sous le nom de "carreaux de Haar", ce dernier est essentiellement une région rectangulaire de l'image où les différences de niveaux de gris entre les pixels sont calculées. Ces caractéristiques sont fondées sur les différences de niveaux de gris entre les régions adjacentes de l'image.

. Les images de test :

Les images sélectionnées pour la phase de détection :



Figure 3.2 : Une image de test comprenant un visage [39].



Figure 3.3 : Une image de test comportant plusieurs visages [40].

Les résultats obtenus grâce à la méthode de détection de visage de Viola et Jones sont présentés ci-dessous :

➤ **Cas d'image à un seul visage :**

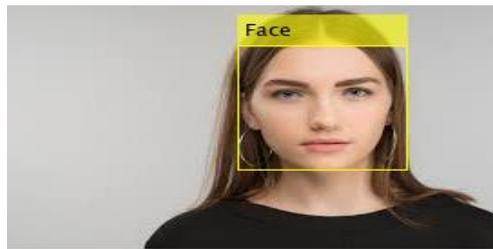


Figure 3.4 : Résultat d'application de la méthode Viola et Jones d'une image contenant exclusivement un visage.

➤ **Cas d'image à plusieurs visages :**

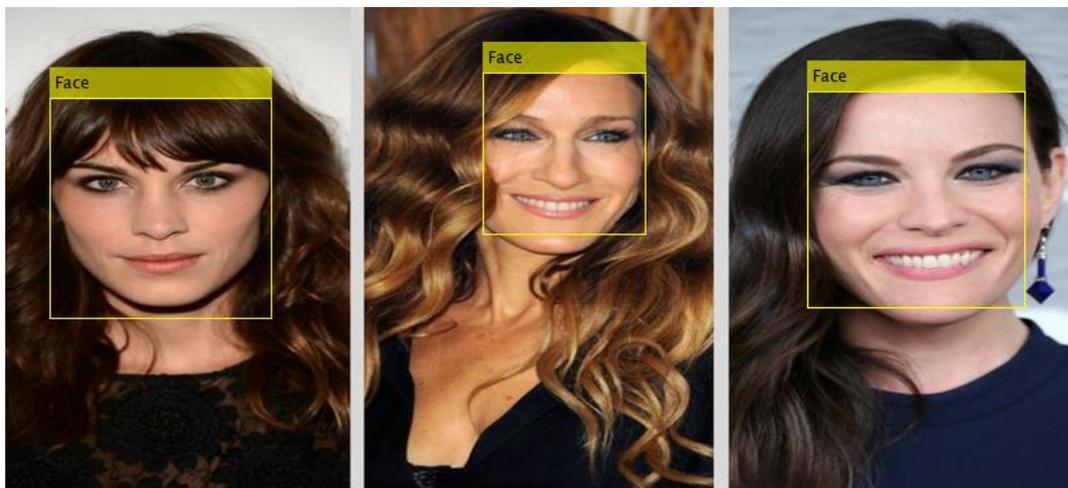


Figure 3.5 : Résultat d'application de la méthode Viola et Jones sur une image contenant plusieurs visages.

Les résultats obtenus illustrent la capacité et l'efficacité de la méthode utilisée. En appliquant l'algorithme Viola et Jones, nous parvenons à détecter de manière précise les visages sur toutes les images, que ce soit un seul visage ou plusieurs visages.

L'algorithme de Viola-Jones adopte une méthode logique et simple pour détecter les visages en évaluant la présence ou l'absence de ses différentes parties.

Pour accomplir cette tâche, nous avons implémenté les fonctions spécifiques de l'algorithme de Viola-Jones :

- **NoseDetect** = `vision.CascadeObjectDetector('Nose','MergeThreshold',50)` pour la détection du nez.
- **MouthDetect** = `vision.CascadeObjectDetector('Mouth','MergeThreshold',60)` pour la détection de la bouche.

Ce modèle traite les images en les extrayant directement de la base de données.

Les résultats obtenus sont illustrés dans les figures présentées ci-dessous :



Figure 3.6 : Détection du nez et la bouche dans le cas d'un seul visage.



Figure 3. 7: Détection du nez et la bouche dans le cas de plusieurs.

Nous pouvons constater que l'algorithme Viola-Jones permet de détecter de manière efficace certaines caractéristiques clés telles que le nez et la bouche.

3.2. Classifications des images :

Notre méthode de classification est composée de deux étapes, comme illustré dans le schéma ci-dessous :

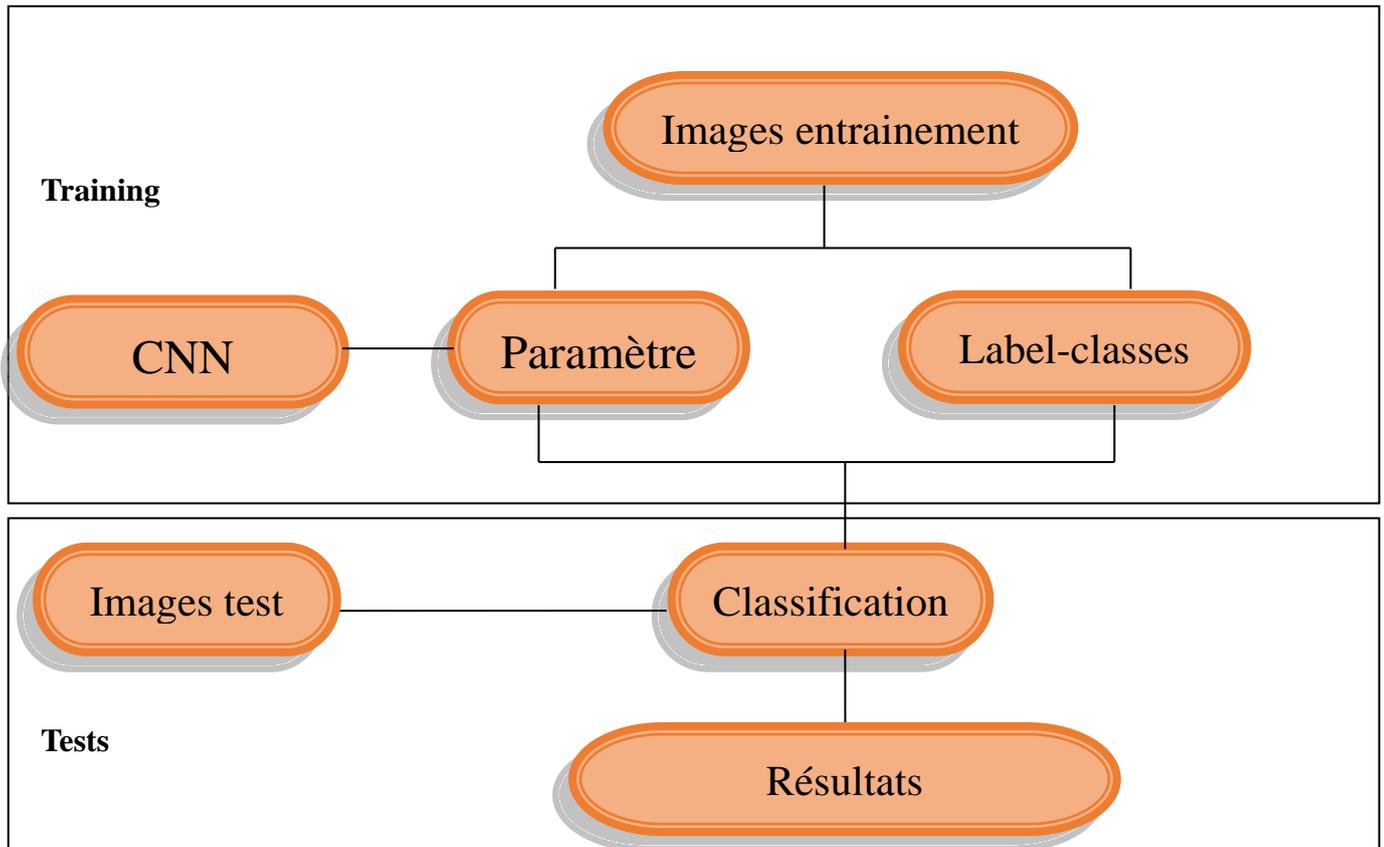


Figure 3. 8 : Modèle de classification.

La figure ci-dessous montre bien que le processus proposé comprend deux phases à savoir :

. Phase d'Entraînement :

Cette étape est cruciale car elle permet la création d'un modèle de classification en utilisant des configurations précises.

- **Images d'entraînement** : Elles s'agissent d'une collection d'images regroupées par classe.
- **Labels-classes** : Il s'agit d'un fichier texte qui répertorie les noms des classes présentes dans notre base de données.
- **Réseau CNN** : Nous appliquons notre base de données à un algorithme CNN afin de générer un modèle, qui sera ensuite utilisé lors de la phase de test.

. La phase de test :

Nous identifions les éléments suivants :

- **L'image test** : C'est la donnée d'entrée utilisée pour les tests, pouvant être soit une seule image ou plusieurs.

- **La classification** : C'est la comparaison entre les caractéristiques correspondante à l'image test et les caractéristiques du modèle proposé pour attribuer à la fin l'étiquette de classification.
- **L'affichage de la classification** : Son nom reflète sa fonction, qui consiste à afficher le résultat obtenu par le modèle, représentant le nom d'une classe.

4. Présentation du système :

Nous allons maintenant exposer les résultats obtenus à partir des expériences menées en appliquant notre méthode de classification d'images.

Notre système suit la procédure suivante :

1. Création des classes.
2. Apprentissage et classification.
3. Prise de décision.

4.1. La base de données :

Cette base de données est utilisée pour entraîner le modèle CNN. Dans ce contexte, nous souhaitons reconnaître deux classes distinctes : la présence du masque ou l'absence du masque. En effet chaque image de notre base de données doit être associée à l'une de ces deux classes.

Pour constituer cette base de données destinée au classifieur CNN, nous avons utilisé des images de nos propres visages ainsi que d'autres récupérées sur Google.

La figure ci-dessous illustre quelques échantillons de notre base de données utilisées :

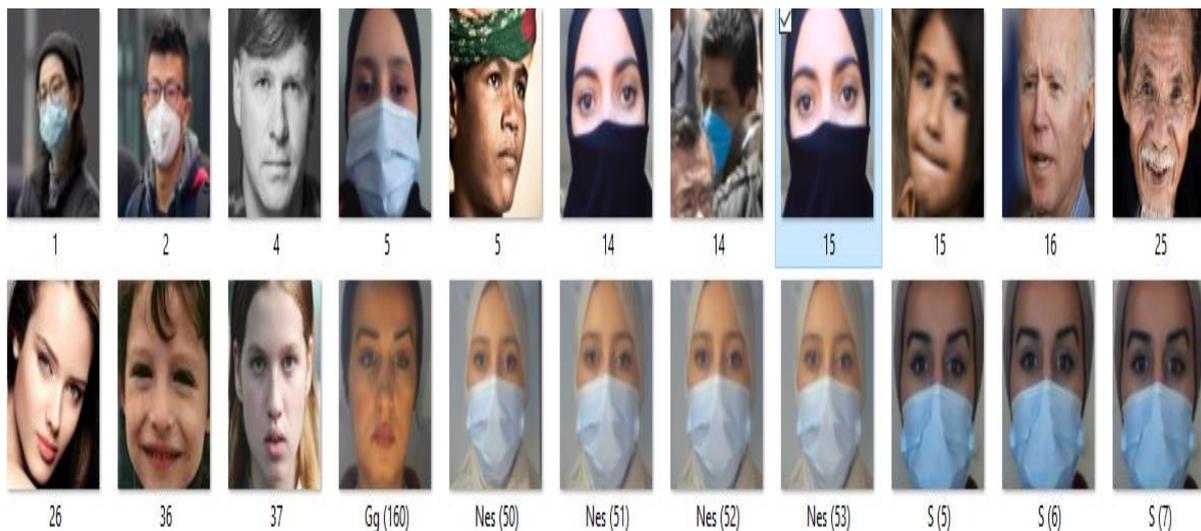


Figure 3.9 : Échantillons de l'ensemble des deux classes avec masque et sans masque.

5. L'architecture du réseau choisi :

- Nous avons effectué plusieurs essais afin de déterminer la configuration optimale d'une architecture de CNN pour obtenir le meilleur taux de classification. Pour ce faire, nous avons modifié le nombre de couches et nous avons finalement choisi un modèle composé de 15 couches, comprenant des couches telles que Imageinput, Convolutionlayer, Max Pooling layer, ReLu layer et SoftMax (fonction d'activation sigmoïde) layer pour la classification et la réduction des erreurs.
- Le modèle que nous avons implémenté, présenté dans la figure ci-dessous, est constitué de trois couches de convolution, deux couches de max pooling et une couche entièrement connectée (fully connected). L'image d'entrée a une taille de 227×227 . Tout d'abord, elle est soumise à une première couche de convolution de taille 11×11 . Chaque couche de convolution est suivie d'une fonction d'activation ReLU, qui permet aux neurones de générer des valeurs positives. Cette opération de convolution produit des cartes de caractéristiques.
- Les cartes de caractéristiques obtenues sont ensuite utilisées en entrée de la deuxième couche de convolution, qui a une taille de 5×5 . Une fonction d'activation ReLU est appliquée à cette couche de convolution, suivie d'une opération de Maxpooling pour réduire la taille de l'image, ainsi que le nombre de paramètres et de calculs. À la sortie de cette couche, des cartes de caractéristiques de taille réduite sont obtenues.
- Le même processus est répété avec la troisième couche de convolution, mais cette fois avec une taille de 3×3 . La fonction d'activation ReLU est appliquée à chaque convolution de cette couche. À la sortie de cette couche, des cartes de caractéristiques encore plus réduites que celles de la couche de convolution précédente sont obtenues. Ces cartes de caractéristiques, également appelées cartes de convolution, sont ensuite concaténées en un seul vecteur contenant ces caractéristiques.
- Enfin, une couche de Softmax est appliquée pour calculer la distribution de probabilité des classes suivantes : "avec masque" et "sans masque".

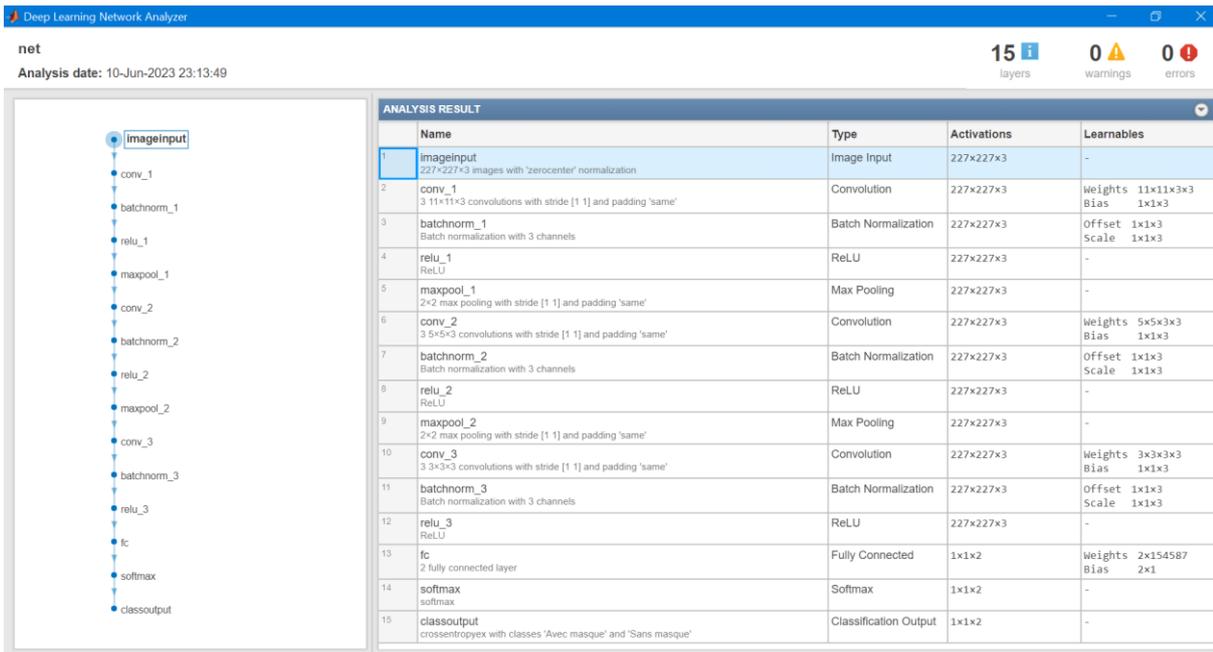


Figure 3.10 : L'architecture du modèle pour 15 couches.

6. Résultats obtenus et discussions :

Pour présenter les résultats de notre modèle, nous allons maintenant exposer les données concernant la précision et l'erreur en fonction du nombre d'itérations, ainsi que la matrice de confusion. Ces éléments nous permettront d'évaluer les performances de notre modèle.

6.1. Graphe de précision et d'erreur :

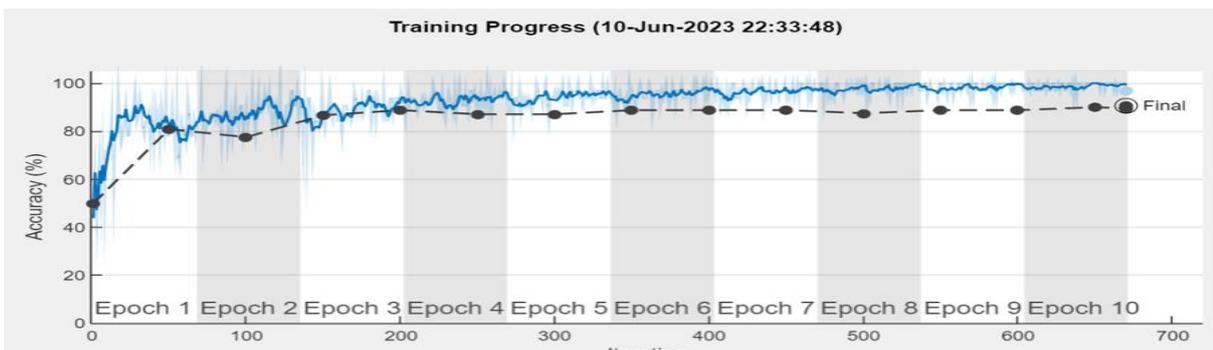


Figure 3.11 : Le graphe de précision.

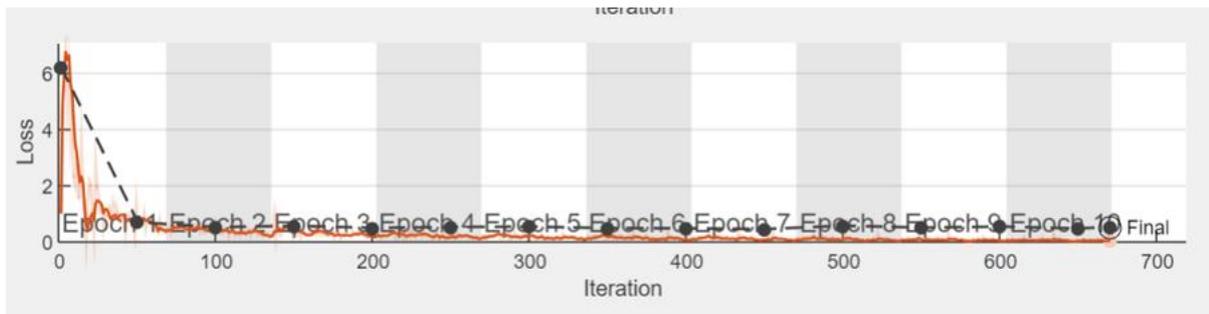


Figure 3.12 : Le graphe d'erreur.

- Les graphes présentés ci-dessus illustrent la courbe de précision et de perte d'entraînement du modèle CNN sur 10 epochs. Nous pouvons observer que la précision d'entraînement augmente à chaque itération, atteignant finalement une précision de 90,83 %. Cela indique que le modèle acquiert de plus en plus d'informations sur les images à chaque itération, ce qui améliore la précision et les performances globales du modèle.
- Par ailleurs, la Figure 3.12 démontre que l'erreur de classification diminue progressivement à chaque itération. Cela signifie que le modèle devient plus précis dans la classification des images au fil du temps.

6.2. Matrice de confusion :

La matrice de confusion joue un rôle essentiel dans l'évaluation des performances de notre modèle, car elle présente les métriques de vrai positif, vrai négatif, faux positif et faux négatif. La figure illustre la position de ces métriques pour chaque classe, ce qui nous permet de visualiser et d'analyser les performances de manière dynamique.

L'avantage de ces métriques est leur simplicité de lecture et de compréhension. Elles offrent une visualisation claire des données et des statistiques, ce qui facilite l'analyse des performances du modèle.

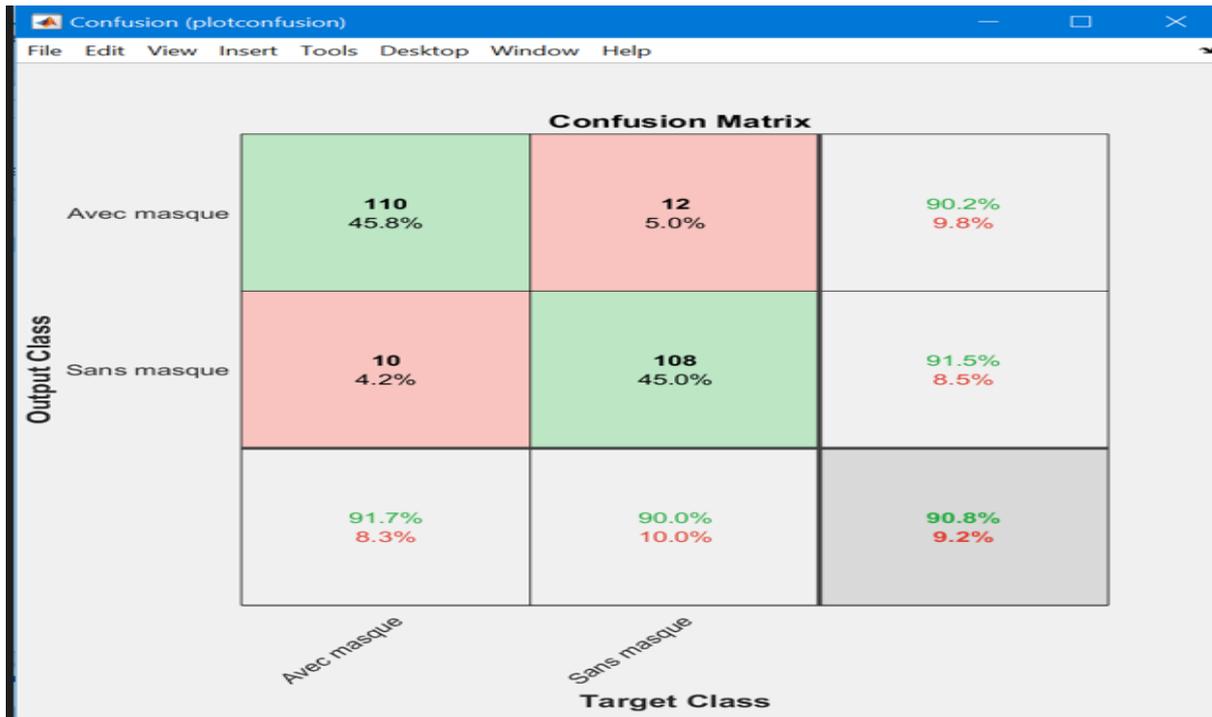


Figure 3.13 : Matrice de confusion résultante pour 15 couches.

La matrice de confusion présentée dans la figure récapitule les prédictions des images de test pour les deux classes : la classe « masque » et la classe « sans masque » .

- Pour la classe « masque », 110 images sont correctement classées, tandis que 12 images sont mal classées. Cela correspond à un taux de précision de (90.2%) et un taux d'erreur de (9.8)%.
- Pour la classe « sans masque » : 108 images sont bien classées et 10 images mal classées ce qui donne un taux de précision de (91.5%) et un taux d'erreur de (8.5%).

Les résultats montrent que le taux de la bonne classification est de (90.8%), et le taux de classification erroné est de (9.2%), ce qui reflète la performance et l'efficacité de notre modèle.

7. Influence nombre de couche et taille des filtres :

7.1. Modèle CNN à 11 couche et tailles des filtres différents :

. Pour une taille du filtre (11*11,5*5) :

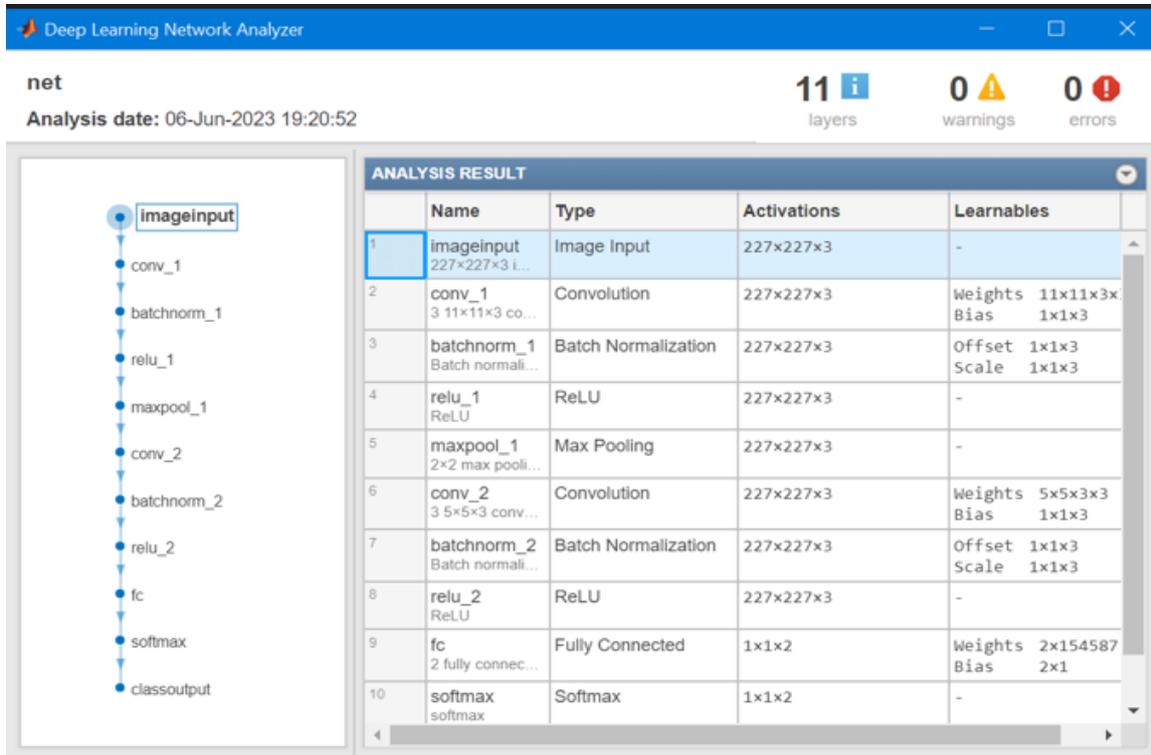


Figure 3.14 : L'architecture du modèle pour 11couches.



Figure 3.15 : Graphe de précision et d'erreur pour 11 couches.

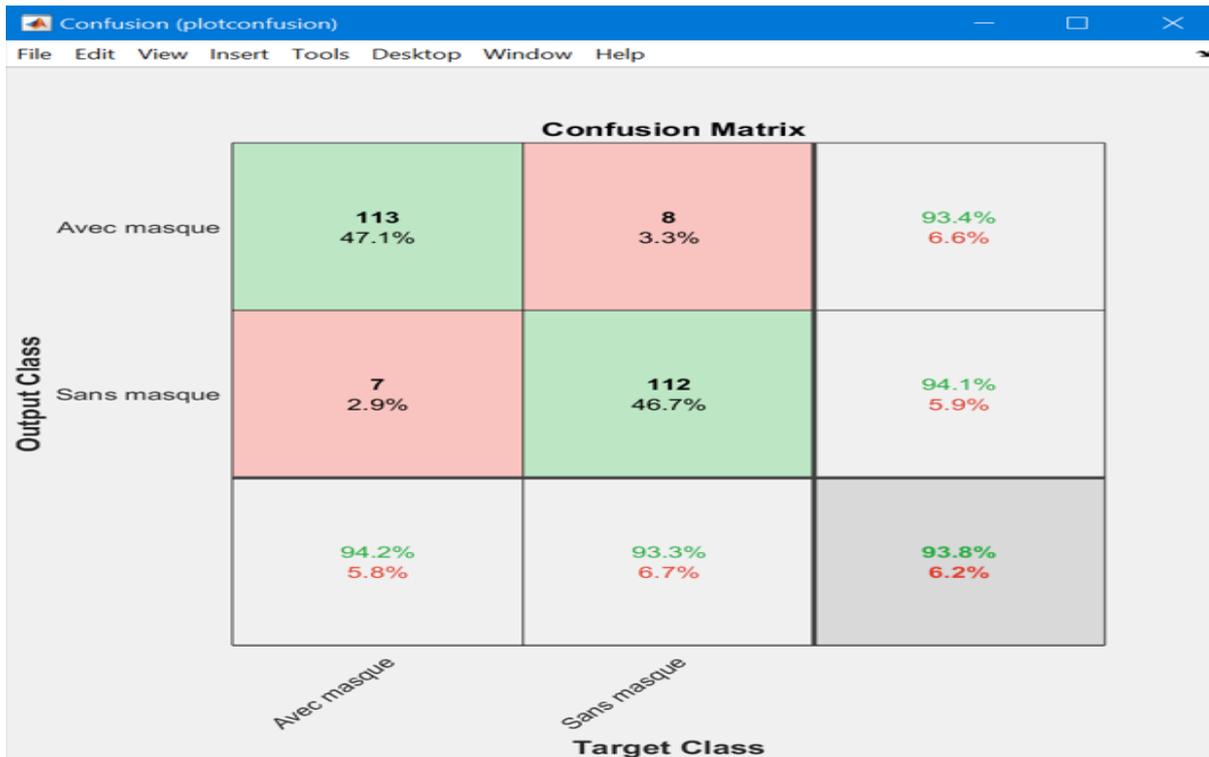


Figure 3.16 : Matrice de confusion pour 11 couches (11*11,5*5).

- La Figure 3.15 présente les courbes de précision et de perte d'entraînement du modèle CNN à 11 couches, avec un nombre d'epochs de dix. Selon ce graphe, nous observons que la précision augmente progressivement pour atteindre (93.75%), tandis que la courbe d'erreur diminue au fur et à mesure des epochs.
- La Figure 3.16 illustre la matrice de confusion pour les deux classes lorsque le nombre de couches est réduit à 11. Nous constatons que 225 images sont correctement classées, ce qui correspond à la diagonale de la matrice, tandis que 15 images sont mal classées.

Nous remarquons une augmentation de la précision de classification du modèle (93.8%) et du taux d'erreur (6.2%) par rapport au premier modèle, où la précision était de (90,8%) et le taux d'erreur était de (9.2%).

.Pour une taille de filtre(5*5,3*3) :

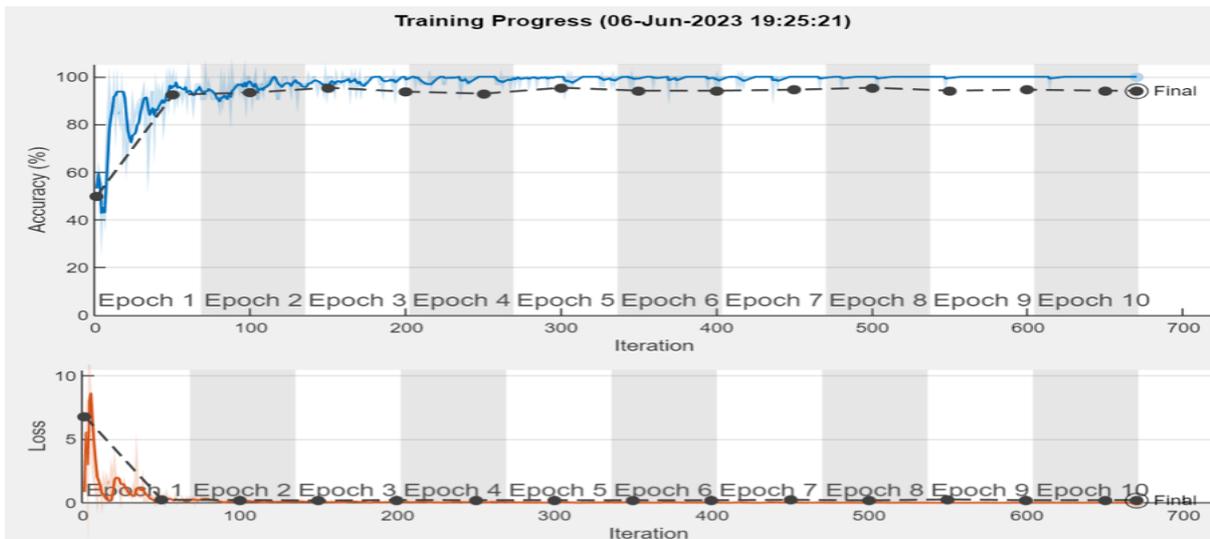


Figure 3.17 : Graphe de précision et d’erreur pour une taille du filtre (5*5,3*3).

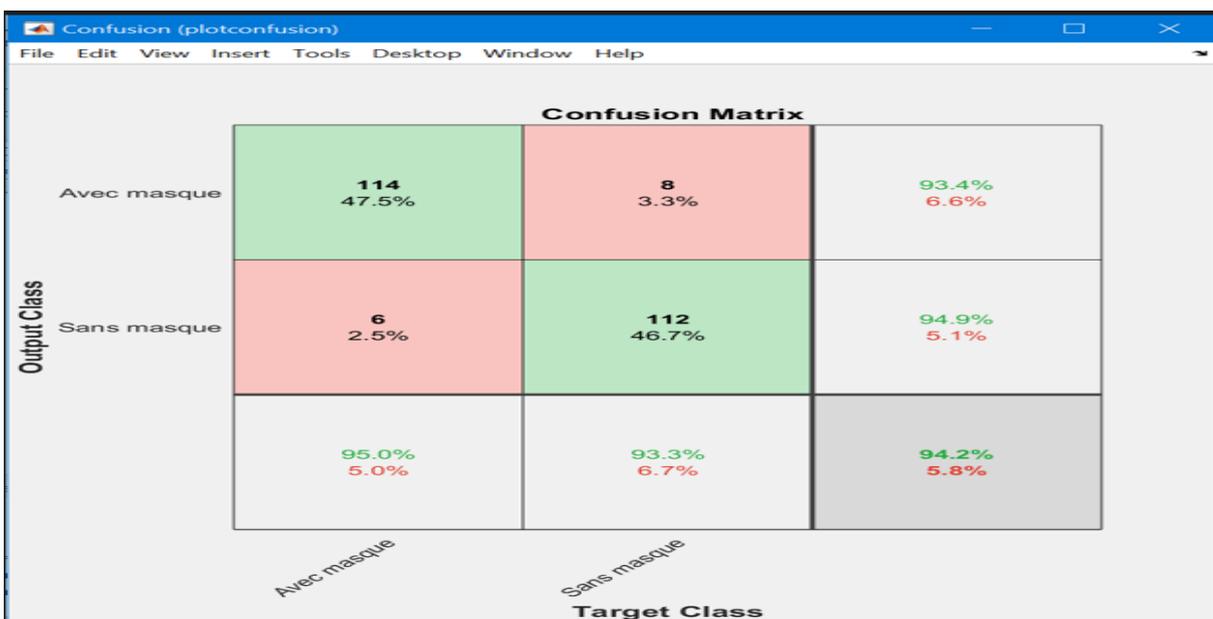


Figure 3.18 : Matrice de confusion pour une taille du filtre (5*5,3*3).

- La Figure 3.17 présente les courbes de précision et de perte d'entraînement du modèle CNN à 11 couches, avec un nombre d'epochs de dix. Selon ce graphe, nous observons que la précision augmente progressivement pour atteindre (94.17%), tandis que la courbe d'erreur diminue au fur et à mesure des epochs.
- La Figure 3.18 illustre la matrice de confusion pour les deux classes lorsque le nombre de couches est réduit à 11. Nous constatons que 226 images sont correctement classées, ce qui correspond à la diagonale de la matrice, tandis que 14 images sont mal classées.

Nous remarquons une augmentation de la précision de classification du modèle (94.2%) et du taux d'erreur (5.8%) par rapport au premier modèle avec une taille du filtre(11*11.5*5), où la précision était de (93,8%) et le taux d'erreur était de (6.2%).

8. Tableau de comparaison des résultats :

Nombre de couches	Taille de filtre	Images bien classées	Images mal classées	Taux de précision	Taux d'erreur	Temp d'exécution
11	(11 11),(5 5)	225	15	93.8%	6.2%	31min47sec
	(5 5),(3 3)	226	14	94.2%	5.8%	10min46sec
15	(11 11),(5 5) (3 3)	218	22	90.8%	9.2%	38min41sec

Tableau 3.1: Tableau de comparaison des résultats.

Le tableau 3.1, illustre l'impact de l'augmentation du nombre de couches de convolution dans notre étude. Les résultats sont exprimés en termes de précision d'apprentissage, de tests, d'erreurs et de temps d'exécution, nous remarquons qu'avec le modèle à 15 couches, nous avons obtenu un taux de précision de (90.83 %). En revanche avec le modèle à 11 couches, nous parvenons à atteindre un taux de précision de 93.75 % pour un filtre de taille (11*11,5*5) et (94.17%) et pour un autre de taille (5*5,3*3).

Nous pouvons constater que :

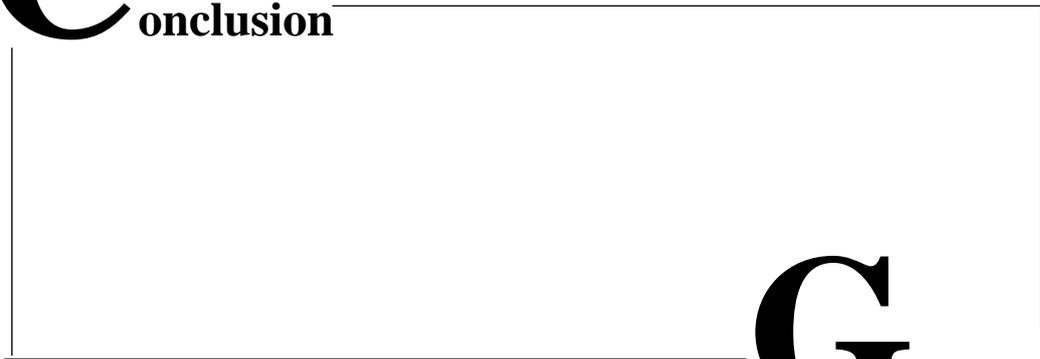
- Le nombre de couches de convolution semble influencer ces bons résultats, le modèle à 11 couches présente les meilleurs résultats.
- La taille des filtres est également un élément déterminant dans les réseaux de neurones CNN.
- Le temps varie en fonction de la taille de la base de données utilisée.
- Les résultats obtenus se sont améliorés à mesure que nous approfondissions notre réseau.
- Il est essentiel d'avoir une base d'apprentissage de grande taille (plusieurs visages à tester) pour obtenir de meilleurs résultats.

9. Conclusion :

Dans ce chapitre, nous avons mis en œuvre notre modèle de classification d'images basée sur les réseaux de neurones convolutionnels CNN avec logiciel Matlab. Nous avons utilisé deux modèles d'architectures, l'un avec 15 couches et l'autre avec 11 couches, tout en variant la taille des filtres dans ce dernier afin d'obtenir des résultats différents.

La comparaison des résultats obtenus a révélé que la profondeur du réseau et la taille des filtres sont des facteurs importants pour obtenir de meilleur taux de précisions et d'erreurs ainsi le temps d'exécution.

Conclusion



Générale

Conclusion générale

La détection de masque en utilisant les CNN est une approche prometteuse et efficace pour relever le défi de la surveillance et du respect des normes sanitaires, notamment pendant les périodes de pandémie comme celle que nous avons récemment traversée avec le COVID-19.

L'objectif de notre travail est de développer un modèle capable de classifier un ensemble d'images en différentes classes. Raison pour lequel on a opté pour les réseaux de neurones convolutionnels CNN, car ils permettent d'extraire un nombre très important des caractéristiques visuelles à partir d'images et les classifier de manière plus rapide et précise comparant ces réseaux par les autres méthodes classiques utilisées dans la classification des images. Ce qui les rend particulièrement adaptés à la tâche de détection de masque. Grâce à leur capacité à apprendre des motifs et des structures complexes dans les images, les CNN peuvent identifier efficacement la présence ou l'absence de masques faciaux sur des individus.

Nous avons utilisé ces réseaux de neurones convolutionnels CNN en présentant les différents types de couches utilisées dans la classification. Nous avons présenté différents modèles avec des architectures différentes afin d'obtenir des meilleurs résultats en termes de précision et d'erreur, le meilleur modèle a atteint une précision de 94.17%.

L'utilisation des réseaux de neurones convolutionnels permettra d'extraire un nombre très important de caractéristiques des images en entrée, agrandir la base de données pour assurer une meilleure précision d'apprentissage, exécuter les calculs avec une station de travail ou avec des PC plus performants permettra de minimiser le temps d'apprentissage. Et à la fin tous ces points feront que la précision de classification sera meilleure.

En conclusion, les résultats obtenus lors de la phase de test confirment l'efficacité de notre approche, mais notre travail reste ouvert à des travaux de comparaison avec d'autres méthodes de classification car il est dans sa version initiale.

Références

Références

- [1] S. Meivel, K. Indira Devi, S. Uma Maheswari, J. Vijaya Menaka. " *Real time data analysis of face mask detection and social distance measurement using Matlab*". Department of ECE, M.Kumarasamy College of Engineering, India, Department of ICE, GRG Polytechnic College, India. Article, December 2020
- [2] M.Belahcen, "Identification et Authentification en Biométrie". Thèse de doctorat, Université Mohamed KHIDER, BISKRA, Algérie, 2013.
- [3] M. ANDRE, "Introduction aux : techniques de traitement d'images", 2e édition, Eyrolles, 1987.
- [4] Ahlem ADJIMI, "Classification des empreintes digitales par les réseaux de neurones flous". Master Académique en Informatique, Université de M'sila, Algérie, 2012.
- [5] M. BOUTERFAS, M. SOUIDI, "Développement d'un système de reconnaissance des Cellules sanguines". Thèse de doctorat, Université d'Aix Marseille, France, 2011.
- [6] Article vision par ordinateur et compréhension de l'image «détection du Visage : une enquête Tome 83 » Numéro 3 septembre 2001 pages 236_274
- [7] <https://www.google.com/url?sa=i&url=https%3A%2F%2Fmailient.com%2Fblog%2Fwh-does-edge-ai-mean-for-face-recognition-technology>. consulté le 01.04.2023
- [8] M.H. Yang, D.J. Kriegman, and N.Ahuja . « detecting faces in images : A survey ». IEEE Trans. On PAMI, pp 34-58, 2002.
- [9] E. Hjelm and B.K.Low . « Face détection : A survey ».
- [10] <https://www.google.com/url?sa=i&url=http%3A%2F%2Fthesis.univ-biskra.dz> consulté le 24.04.2023.
- [11] S.A. Sirohey, « Human Face Segmentation and Identification », Technical Report CS-TR3176, Univ. of Maryland, pp. 1-33, 1993.
- [12] J. Haddadnia , M. Ahmadi , and K .Faez , « an efficient feature extraction method with peude zernik moment in RBF neural network based human face recognition system » , eurasip, jasp , vol.9, pp.890891 , 1996.
- [13] <https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.researchgate>. consulté le 15.05.2023.
- [14] <https://www.google.com/url?>. consulté le 16.05.2023.
- [15] <https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.futura-sciences.com%2Ftech%2Factualites%2Ftechnologie-reconnaissance-faciale-algorithme-battu-humains>. consulté le 17.05.2023.
- [16] <https://www.google.com/url?sa=i&url=,> consulté le 17.4.2023
- [17] Jean-Paul HATON - LORIA – Institut Universitaire de France – Université de Lorraine – Nancy, France
- [18] Article de Céline Deluzarche, journaliste(1979-2021), le 7.02.2021

Références

- [19] Vinay Rao et al (2015). "Brain Tumor Segmentation with Deep Learning ", université de Southern California. Janvier 2015
- [20] <https://www.google.com/url?sa=i&url=https%3A%2F%2Fbdtechtalks.com%2F2019%2F02%2F15%2Fwhat-is-deep-learning-neural-networks>.consulté le 17.04.2023
- [21] CHAGAS, Edgar Thiago De Oliveira. "L'apprentissage profond et ses applications aujourd'hui", Revista Científica Multidisciplinar Núcleo do Conhecimento. 04 année, Ed. 05, vol. 04, pp. 05-26 mai 2019. ISSN: 2448
- [22] 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.
- [23] Donald J. Norris, "Machine Learning with the Raspberry Pi", page 213 Barrington, NH, USA ,2020.
- [24] O. GUENOUNOU, "Méthodologie de conception de contrôleurs intelligents par l'approche génétique : application à un bioprocédé", Thèse de Doctorat, Systèmes Automatiques, Université Toulouse III - Paul Sabatier, France,2009
- [25] K. Gurney, "An introduction to Neural Networks",2e édition, Dunod, Marseille, p-67, 1997.
- [26] Charu C. Aggarwal, "Neural Networks and Deep Learning ", IBM T. J. Watson Research Center International Business Machines, Yorktown Heights, NY, USA.
- [27] heba hakim et Ali Fadhil, "survey: Convolution Neural networks in Object Detection", Article dans Journal of Physics Conference Series,12 (7), pp.02, 2021.
- [28] Asifullah Khan et autre,"A Survey of the Recent Architectures of Deep Convolutional Neural Networks", Publié dans Artificial Intelligence Review DOI, 2020.
- [29] Renders, Jean-Michel,'Algorithmes génétiques et réseaux de neurone', Renders, Jean-Michel, Hermés Paris(1994).
- [30] https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2FMax-pooling-and-average-pooling-with-a-2-A-2-filter_4.consulté le 18.04.2023.
- [31] <https://www.google.com/url?sa=i&url=https%3A%2F%2Fpyimagesearch.com%2F2021%2F05%2F14%2Fconvolutional-neural-networks-cnns-and-layer-types>.consulté le 18.05.2023.
- [32] [https://www.google.com/imgres?.](https://www.google.com/imgres?) consulté le 18.05.2023.
- [33] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Advances in neural information processing systems Vol 25(2).]
- [34] A.Khelalef,Reconnaissances d'activités humaines en utilisant les descripteurs spatio-temporels 2D/3D,Thèse de doctorat,Faculté de technologie,Université Batna2-Mostafa Ben Bolaid,Batna,Algerie,2020.
- [35] Karen Simonyan & Andrew Zisserman, (2015). « Very Deep Convolutional Networks For Large-Scale Image Recognition », ICLR 2015.

Références

- [36] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, (2015). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9.
- [37] K. He, X. Zhang, S. Ren, and J. Sun, (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
- [38] <http://www.jobintree.com/dictionnaire/definition-matlab-915.html>. Consulté le 24/05/2023.
- [39] <https://www.lentiamo.fr/la-forme-de-votre-visage.html> Consulté le 13/05/2023.
- [40] <https://www.monvanityideal.com/conseils/cheveux/coiffures/visage-allonge-quelle-coupe-de-cheveux-adopter.html> consulté le 13/05/2023.

التعلم المعرق هو فرع من فروع التعلم الآلي الذي يركز على تدريب نماذج الذكاء الاصطناعي بكميات كبيرة من البيانات، علاوة على ذلك حققت الشبكات العصبية التلافيفية نجاحات ملحوظة في مجال تصنيف الصور باستخدام التعلم المعرق. عند تدريب هذه النماذج العميقة على مجموعات بيانات كبيرة، فإنها تعمل بشكل جيد للغاية. يقدم مشروع نهاية الدراسات هذا استخدام الشبكات العصبية التلافيفية التي قمنا بتنفيذها للكشف عن ارتداء القناع باستخدام مجموعة بيانات، وذلك باستخدام برنامج البرمجة Matlab. أظهرت التجارب المعدلة كفاءة استخدام هذه التقنية ومن خلال تجاربنا قدمنا عرضًا تجريبيًا على النموذج المقترح ببيانات مختلفة عن طريق تغيير حجم طبقة الالتفاف تظهر أن النموذج المكون من 11 طبقة بحجم المرشحات على التوالي (5 * 5.3 * 3) يصل إلى دقة 94.17%. الكلمات المفتاحية: التعلم المعرق، قناع، الشبكات العصبية التلافيفية، الكشف عن ارتداء القناع

Résumé

L'apprentissage en profondeur ou le Deep Learning est une branche de l'apprentissage automatique qui est basé sur l'entraînement à partir de grandes quantités de données. Les réseaux de neurones convolutionnels ont obtenu des succès remarquables dans le domaine de la classification d'images offrant des performances exceptionnelles.

Ce projet de fin d'études présente l'utilisation des réseaux de neurones convolutionnels CNN que nous avons implémenté pour la détection du port du masque en utilisant un ensemble de données, et cela en utilisant le logiciel de programmation Matlab.

Les expériences adaptées sur le modèle proposé avec différentes architectures en variant aussi la taille de la couche convolution révèle que le modèle avec 11 couches ayant la taille des filtres respectivement (5*5,3*3) atteint une précision de 94.17%.

Mots clés : Deep Learning, masque, CNN, détection du port de masque

Abstract

Deep learning is a branch of machine learning based on training with large amounts of data. Convolutional neural networks have achieved remarkable success in the field of image classification, delivering outstanding performance.

This graduation project presents the use of convolutional neural networks (CNNs), which we have implemented for mask-wearing detection using a dataset and Matlab programming software.

Adapted experiments on the proposed model with different architectures, also varying the size of the convolution layer, reveal that the model with 11 layers of respective filter sizes (5*5,3*3) achieves an accuracy of 94.17%.

Keywords : Deep Learning, mask, CNNs, mask-wearing detection