République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieure et de la Recherche Scientifique
Université Abderrahmane Mira
Faculté de Technologie

Département d'Automatique, Télécommunication et
D'Electronique

Projet de Fin d'études

CONVOLUTIONAL NEURAL NETWORK POUR LA DETECTION DU PORT DU MASQUE

Préparé par :
MAZIR MELISSA
AMRIOU HANANE

Filière : Télécommunications.
Spécialité : Réseaux et Télécoms.

Examiné par :

Mr. MEKHMOUKH ABDENOUR.	Encadrant
Mr ALLICHE ABDENOUR.	Encadrant
Mr. SADJI MUSTAPHA	President
Mr BERRAH SMAIL.	Examinateur

Année universitaire : 2021/2022

Remerciement

À la fin de ce travail je tiens à remercier ALLAH le tout puissant de m'avoir donné la foi et de m'avoir permis D'arriver là.

Le travail présenté a été effectué sous la direction de Mr.Alliche à qui je tiens à adresser mes plus vifs remerciements, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter ma réflexion et ses encouragements lors de la réalisation de ce mémoire.

Mes remerciements et ma gratitude vont aussi les membres de jury d'voir accepté de faire partie du jury de soutenance de ce projet de fin d'études.

Toute ma reconnaissance va aussi aux professeurs et enseignants de département d'informatique ainsi que ses étudiants et son personnel côtoyés tout au long de mon cursus universitaire.

Que toute personne ayant œuvré de près ou de loin à la réalisation de ce projet par une quelconque forme de contribution, trouve ici le témoignage de ma plus profonde connaissance

Résumé

Les méthodes d'apprentissage profond, en particulier les réseaux de neurones convolutifs, ont obtenu des succès significatifs dans le domaine de classification d'images. La formation de modèles profonds montre de performances exceptionnelles avec de grands ensembles de données.

Ce travail de projet de fin d'étude propose un réseau de neurones à apprentissage profond pour apprendre sur un ensemble de données de taille réduite. Pour cela, un ensemble de d'image selon des classes différentes avec masque, sans masque et masque incorrect a été généré sous le logiciel de programmation Matlab.

Nous démontrons expérimentalement que le jeu de données d'apprentissage améliore réellement la puissance de généralisation des CNN. L'évaluation du modèle étudié a montré de bonnes performances en termes de précision avec un taux de classification de **91.5%.**

Tables des matières :

Chapitre II : Deep Learning pour la classification des images

Chapitre III : Implémentation d'une application de classification

Liste des figures

Chapitre 1 :

Chapitre 2 :

Chapitre 3 :

Lister des tableaux :

Introduction générale

Il existe de nombreux traitement d'images pour lutter contre une pandémie comme celle du COVID-19. On peut citer par exemple le traitement d'images pour détecter le port du masque afin de lutter contre cette maladie.

La détection de visage est un traitement indispensable et crucial avant la phase de reconnaissance. En effet, le processus de reconnaissance de visage ne pourra jamais devenir intégralement automatique s'il n'a pas été procédé par une détection efficace.

Plusieurs méthodes de reconnaissance de visage ont été proposées suivant deux axes : la reconnaissance à partir d'images fixes et la reconnaissance à partir d'une séquence d'images (vidéo). De manière générale, les méthodes de reconnaissance de visage peuvent être divisées en trois groupes : des méthodes globales un utilisant des techniques linéaires et non linéaires, des méthodes locales basées sur des approches comme : KNN et SVM et des méthodes hybrides dont le fonctionnement est semblable à un système de perception humain.

Durant les dernières années, un algorithme d'apprentissage considéré parmi les algorithmes les plus sophistiqués au monde qui est le « Deep Learning » a fait l'objet de nombreuses études et a obtenu des résultats remarquables dans le domaine de détection du visage. Il a montré ses performances face aux différents problèmes dépassant les algorithmes classiques.

La technique d'apprentissage profond repose sur le traitement par les ordinateurs de grandes quantités de données dans le but de détecter des images numériques et extraire des données afin de produire des informations numériques sous forme de décisions d'une manière efficace.

Dans notre travail, nous allons nous focaliser sur l'un des algorithmes les plus performants du Deep Learning, les CNN (Convolutional Neural Network ou réseau de neurone convolutifs en français), un modèle de programmation puissant permettant notamment la reconnaissance d'images en attribuant automatiquement à chaque image fournie en entrée une étiquette correspondant à sa classe d'appartenance.

Notre travail comprend essentiellement deux parties : théoriques et pratiques, il est organisé comme suit

Dans le premier chapitre nous allons présenter les différentes techniques de détection des visages ainsi que la description de la méthode de Viola et Jones.

Le chapitre deux illustre la technique d'apprentissage profond, et une description plus détaillée sur les réseaux de neurones convolutifs CNN qui est la méthode choisie dans notre projet.

Le dernier chapitre sera consacré à la partie pratique, par la présentation de la méthode implémentée sous le langage de programmation Matlab.

Enfin, nous terminons par conclusion générale qui résumera notre contribution et les perspectives sur les travaux futurs.

Liste des abréviations

DL : Deep Learning

IA : Intelligence Artificielle

ML : Machine Learning

CNN : Convolutional Neural Network

ReLu : Rectifiedlinear unit

DBN : Deep Belief Networrk

SVM : Support Vector Machine

KNN : K-Nearest Neighbors

Chapitre I : Techniques de détection de visage

I.1 Introduction

Pour construire un système automatisé qui analyse l'information contenue dans les images de visage, des algorithmes efficaces et robustes de détection de visage sont exigés. En effet, vue l'importance de la détection de visage pour n'importe quel système d'analyse et dans le but d'identifier toutes les régions d'image qui contiennent un visage indépendamment de la position, de l'orientation ou d'éclairage, plusieurs recherches ont été faites. En particulier, de nombreuses techniques ont été développées pour détecter des visages dans des images fixes.

I.2 Détection de visage

La détection de visage est une étape très intéressante dans le domaine de reconnaissance de visage. Plusieurs travaux de recherches ont été effectués dans ce domaine. Ils ont donné lieu au développement d'une multitude de techniques allant de la simple détection du visage, à la localisation précise des régions caractéristiques du visage, tels que les yeux, le nez, les sourcils, la bouche, les oreilles, etc. **[1]**

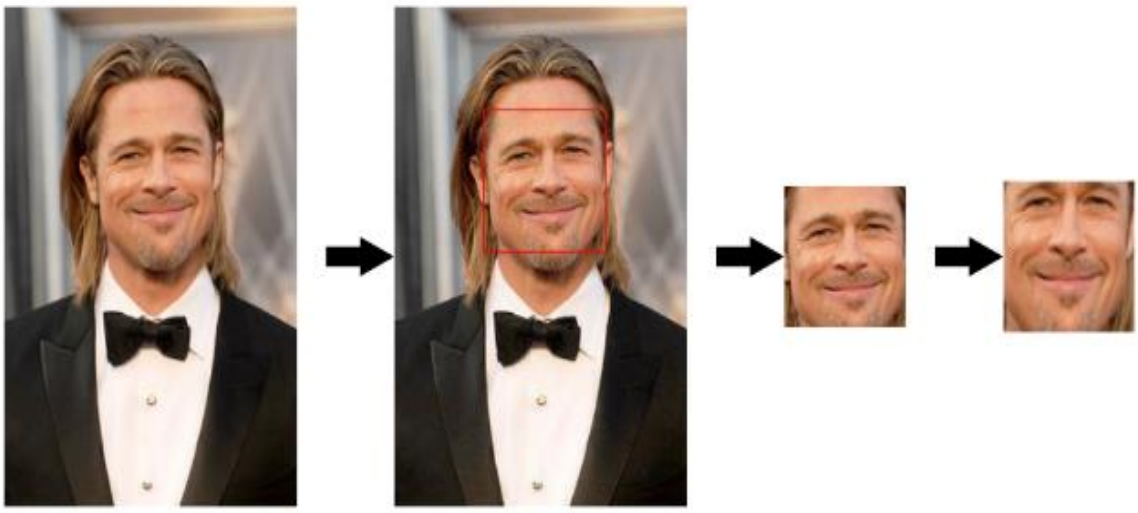


Figure 1: détection du visage [34]

I.3 Principales difficultés de la détection de visage

La détection de visage cherche à détecter la présence et la localisation précise d'un ou plusieurs visages dans une image numérique. C'est un sujet difficile, notamment dû à la grande variabilité d'apparence des visages dans des conditions non contraintes.

I.3.1 L'échelle

Dans une image donnée, un groupe de visages peut apparaitre dans différentes échelles (tailles).

Figure 2. Ensemble d'image sur diffèrent échelles [35]

I.3.2 Position

Les performances d'un système de détection de visages chutent d'une manière significative lorsque les variations de poses sont présentées.

Figure 3. Visage sur différents angles [36]

I.3.3 Occlusion

L'occlusion est une autre issue confrontée à la détection de visages. Les lunettes, les écharpes et les barbes tous changent l'aspect d'un visage

Figure 4. Visage occlus [37]

I.3.4 Illumination

En fait, les changements provoqués par des différences d'éclairage sont souvent plus grands que les différences qui existent entre les individus **[2]**

Figure 5. Visages sous différentes conditions de lumination [38]

I.4 Méthodes de détection de visage

Toutes ces difficultés font de la détection des visages un véritable challenge des chercheurs désirant relever les défis. Ainsi, durant ces vingt dernières années, de nombreuses approches ont été proposées, elles se répartissent en deux grandes catégories :

- La première catégorie englobe les approches dites **basées sur les traits du visage (features *based approches*) [3]** tels que : Règles généralisées de la connaissance, appariement de modèles et analyse bas niveau.

- La deuxième catégorie rassemble les approches **basées sur l'image (image based approches) [3]** dite approche basée sur l'apparence.

En résumé donc, on peut distinguer quatre principales techniques de détection de visage, tels qu'elles sont montrées dans la figure ci-dessous **:**

1. Règles généralisées de la connaissance (Approches basés sur les connaissances acquises)
2. Appariement de modèles (Approches basés sur Template matching)
3. Analyse bas niveau (approches basées sur des caractéristiques invariantes)
4. Approches basées sur l'image (sur l'apparence)

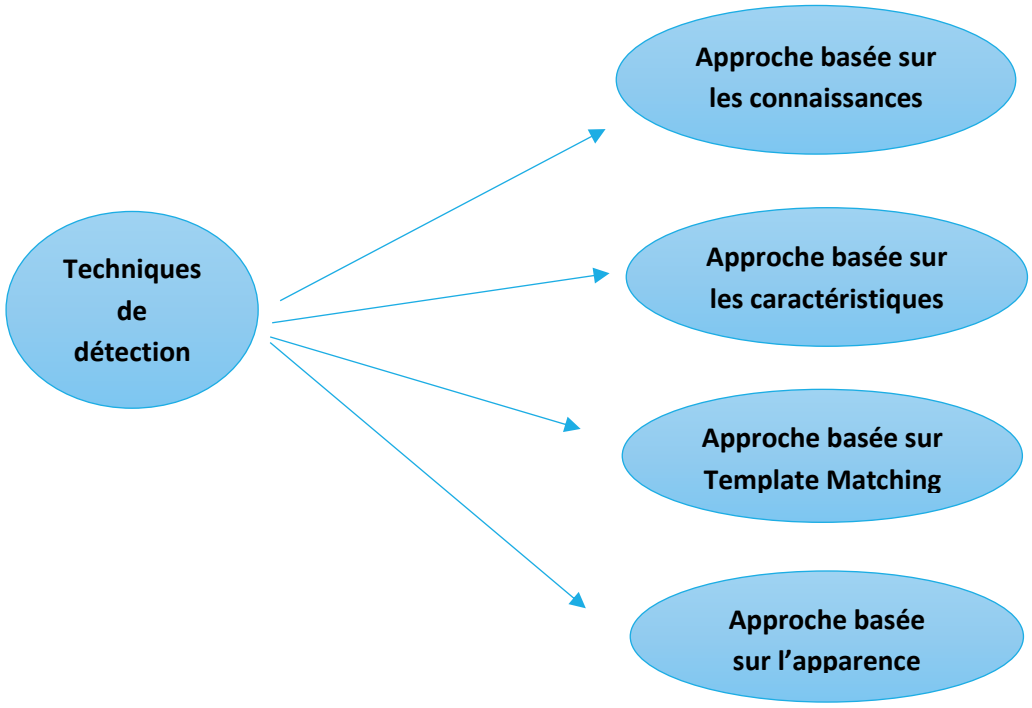


Figure 6. Les techniques de détection

I.4.1 Règles généralisés de la connaissance (Approches basés sur les Connaissances acquises)

Cette méthodologie s'intéresse aux parties caractéristiques du visage comme le nez, la bouche et les yeux, elle est basée sur la définition de règles strictes à partir des rapports entre les caractéristiques faciales, ces méthodes sont conçues principalement pour la localisation du

visage, mais malheureusement, cette technique occasionne de nombreuses fausses détections et un taux faible de détection**. [4]**

I.4.2 Appariement de modèles (Approches basés sur Template Matching)

La détection des visages se fait à travers un apprentissage d'exemples standards de visages ou d'images frontales contenant des visages. La procédure se fait en corrélant les images d'entrés et les exemples enregistrent (gabarits) et le résultat donne la décision finale soit de l'existence ou non d'un visage.

On trouve 2 types de corrélation suivant le type de gabarits :

- Faces de visages prédéfinies
- Modèles déformables.

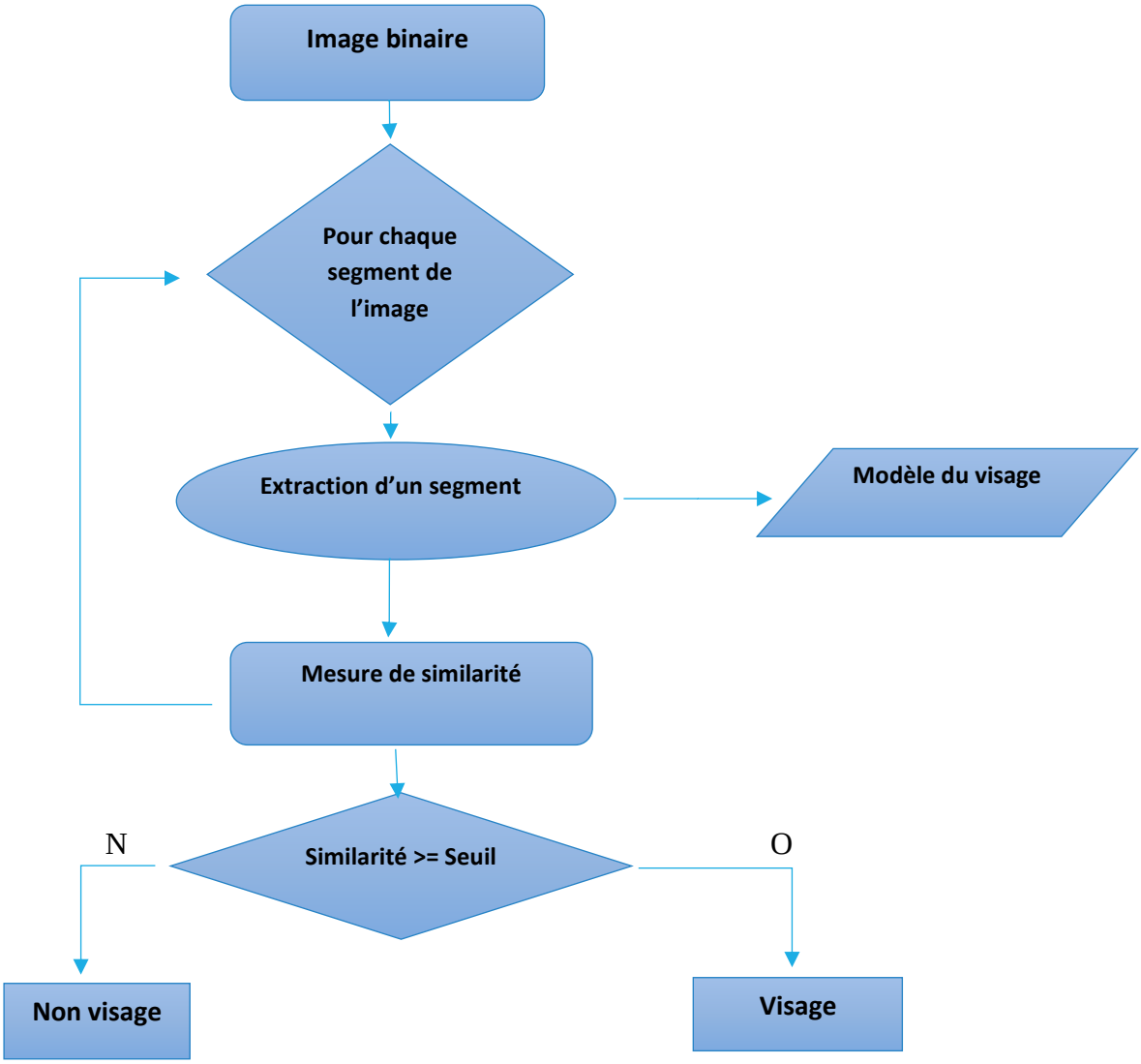


Figure 7. Schéma illustratif de Template Matching

Le modèle ci-dessus représente les différentes étapes qui permettent de détecter un visage en utilisant la méthode du « Template Matching » :

e. Extraction de l'image niveau de gris

D'abord, nous devons extraire l'image binaire du segment, nous multiplions l'image résultante par l'image originale en niveau de gris, le résultat sera une image en niveau de gris qui représente le segment.

f. Normalisation du segment

Cette procédure de normalisation porte sur l'image en niveau de gris du segment c'est-à-dire le résultat de l'étape précédente. La normalisation consiste à ajuster le segment selon son orientation et rendre ses dimensions égales à celles du modèle afin d'améliorer l'opération d'appariement.

g. Comparaison du segment avec le modèle

Cette comparaison s'effectue en balayant la surface du segment et en calculant la corrélation entre le modèle et la région sélectionnée du segment normalisé, le résultat est la valeur maximale calculée ainsi que la position de la région correspondante.

h. Décision

En fixant un seuil de similarité, nous pouvons juger un segment de peau comme étant un visage ou non. On distingue deux cas :

- Si Similarité >=Seuil donc le visage est détecté : visage

- Si la condition n'est pas réalisée donc le visage ne sera pas détecté : Non visage.

I.4.3 Analyse bas niveau (approches basées sur des caractéristiques invariantes)

Cette méthode a pour objectif de trouver les caractéristiques structurelles même si le visage en différentes positions, conditions lumineuses ou changement d'angle de vue**. [4]**

I.4.4 Approches basés image (sur l'apparence)

Ces approches appliquent généralement des techniques d'apprentissage automatique qui sont employés pour la détection. L'idée principale est de considérer que le problème de détection de visage est un problème de classification (visage, non visage). Ces méthodes présentent l'avantage de s'exécuter très rapidement mais demandent un long temps d'entrainement. **[4]**

L'inconvénient donc de cette approche réside dans le temps de calcul qui ne permet pas souvent de faire des traitements en temps réel**. [2]**

A partir de cette situation, nous présenterons le fonctionnement et le principe d'une nouvelle technique qui permet de détecter un visage efficacement et en temps réel, appelée méthode de **Viola et Jones**.

I.5 Méthode de Viola et Jones

Une avancée majeure dans le domaine a été réalisé par les chercheurs Paul Viola er Michael Jones dans leur article « Détection rapide d'objets utilisant une cascade boostée de fonctionnalités simples » en 2001 **[5].**

La méthode de Viola et Jones est une méthode de détection d'objet dans une image numérique, elle fait partie des toutes premières méthodes capables de détecter efficacement et en temps réel des objets dans une image. Inventée à l'origine pour détecter des visages et d'autres types d'objets.

Considérée comme étant l'une des méthodes les plus connues et les plus utilisées, en particulier pour la détection de visages et la détection de personnes.

I.5.1 Principe

Cette méthode est une approche basée sur l'apparence, qui consiste à parcourir l'ensemble de l'image en calculant un certain nombre de caractéristiques dans des zones rectangulaires. Elle a la particularité d'utiliser des caractéristiques très simples mais très nombreuses.

La méthode de Viola et Jones consiste à balayer une image à l'aide d'une fenêtre de détection de taille initiale 24px par 24px et de déterminer si un visage y est présent. Lorsque l'image a été parcourue entièrement, la taille de la fenêtre est augmentée et le balayage recommence, jusqu'à ce que la fenêtre fasse la taille de l'image. L'augmentation de la taille de

la fenêtre s'effectue par un facteur multiplicatif de 1.25. Le balayage, consiste simplement à décaler la fenêtre d'un pixel, ce décalage peut être changé afin d'accélérer le processus, mais un décalage d'un pixel assure une bonne précision.

Il existe d'autres méthodes mais celle de Viola et Jones est la plus performante à l'heure actuelle. Ce qui la différencie des autres est notamment :

- L'utilisation d'images intégrales qui permettent de calculer plus rapidement les caractéristiques.
- La sélection par boosting des caractéristiques.
- La combinaison en cascade de classifieurs boostés, qui apporte un gain meilleur de temps d'exécution.

I.5.2 Eléments de la méthode

I.5.2.1 Les caractéristiques pseudo Haar

Une caractéristique est une représentation synthétique et informative, calculée à partir des valeurs des pixels. Les caractéristiques utilisées ici sont les caractéristiques pseudo-Haar. Elles sont calculées par la différence des sommes de pixels de deux ou plusieurs zones rectangulaires adjacentes. Prenons un exemple : Voici deux zones rectangulaires adjacentes, la première en blanc, la deuxième en noire.

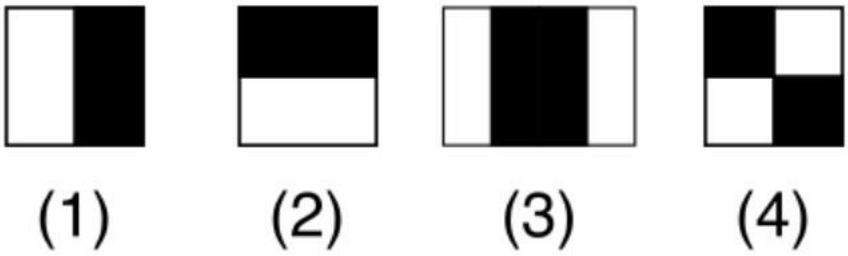


Figure 8.Caractéristiques pseudo-Haar à seulement deux caractéristiques [28]

Ces caractéristiques permettent de détecter des motifs. Par exemple, la reconnaissance des visages est rendue possible par :

- La variation de l'intensité de la lumière entre les yeux et le nez (caractéristique n°2).
- La variation de l'intensité de la lumière entre les yeux et les pommettes (caractéristique n°3) **[28].**

Valeur caractéristique = ∑ (pixels dans la zone blanche) – ∑ (pixels dans la zone noire).

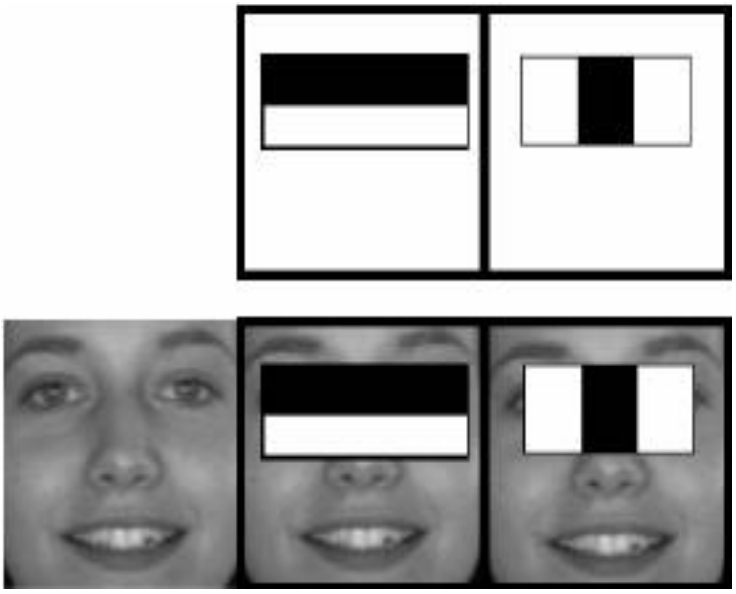


Figure 9. Exemple de classifieur Haar [39]

I.5.2.2 Le calcul

Les caractéristiques sont calculées en soustrayant la somme des pixels noirs à la somme des pixels blancs et, à toutes les positions et à toutes les échelles dans une fenêtre de détection de petite taille, typiquement de 24x24 pixels ou de 20x15 pixels. Un très grand nombre de caractéristiques par fenêtre est ainsi généré, Viola et Jones donnant l'exemple d'une fenêtre de taille 24 x 24 qui génère environ 160 000 caractéristiques.

L'image précédente présente des caractéristiques pseudo-haar à seulement deux caractéristiques mais il en existe d'autres, allant de 4 à 14, et avec différentes orientations. Malheureusement, le calcul de ces caractéristiques de manière classique induit un coût très important en termes de ressources processeur, c'est là qu'interviennent les images intégrales.

I.5.2.3 image intégrale

. Pour calculer rapidement et efficacement ces caractéristiques sur une image, les auteurs proposent également une nouvelle méthode, qu'ils appellent image intégrale.

C'est une représentation sous la forme d'une image, de même taille que l'image d'origine, elle contient en chacun de ses points la somme des pixels situés au-dessus et à gauche du pixel courant. Plus formellement, l'image intégrale ii au point (x, y) est définie à partir de l'image i par :

$$ii(x, y) = \sum_{x'<x, y'<y} i(x', y') \qquad \textbf{(1.1)}$$

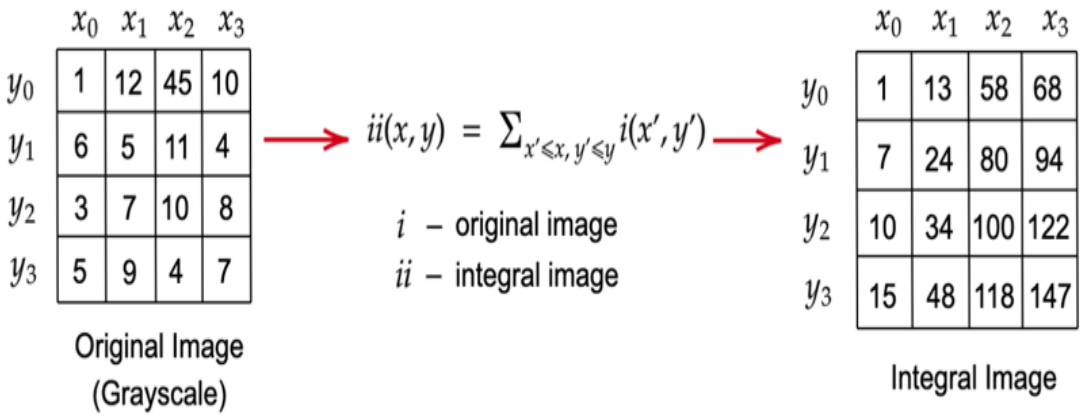


Figure 10.La représentation d'une image intégrale [40]

Tel que :

$\boldsymbol{ii}(\boldsymbol{x},\boldsymbol{y})$ **:** l'image intégrale ii au point $(\boldsymbol{x},\boldsymbol{y})$

$\boldsymbol{i}(\boldsymbol{x'},\boldsymbol{y'})$ **:** l'image originale i au point $(\boldsymbol{x'},\boldsymbol{y'})$

I.5.2.4 Algorithme d'apprentissage basé sur Adaboost

Le boosting est un principe qui consiste à construire un classifieur « fort » à partir d'une combinaison pondérée de classifieur « faibles », c'est-à-dire donnant en moyenne une réponse meilleure qu'un tirage aléatoire. Viola et Jones adaptent ce principe en assimilant une caractéristique à un classifier faible, en construisant un classifieur faible qui n'utilise qu'une seule caractéristique. L'apprentissage du classifieur faible consiste alors à trouver la valeur seuil de la caractéristique qui permet de mieux séparer les exemples positifs et des exemples négatifs.

I.5.2.5 Cascade de classifier

La méthode de Viola et Jones est basée sur une approche qui teste la présence de l'objet dans une fenêtre à toutes les positions et à plusieurs échelles. Cette approche est cependant extrêmement coûteuse en calcul. L'une des idées clés de la méthode pour réduire ce coût réside dans l'organisation de l'algorithme de détection en une cascade de classifier. Ces classifier prennent une décision d'acceptation si la fenêtre contient l'objet, l'exemple est alors passé au classifier suivant, ou de rejet si la fenêtre ne contient pas l'objet et dans ce cas l'exemple est définitivement écarté.

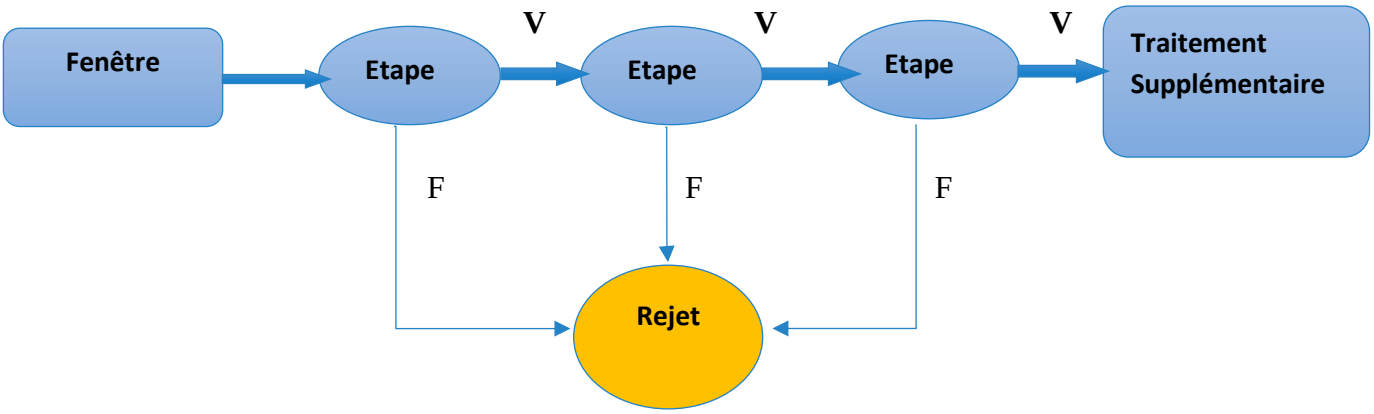


Figure 11. Architecture de la cascade

La figure ci-dessus présente l'architecture de la cascade, la première étape consiste à détecter l'image, une fenêtre de détection se décale donc pixel par pixel puis grandit jusqu'à faire toute l'image. Au sein de cette fenêtre de détection, il va y avoir une cascade de classifier qui réagira quand elle verra un visage.

Cette cascade de classifier s'exécute séquentiellement on se retrouve avec 2 situations possibles :

- « Vrai » marqué par **V** si l'étage 1 voit quelque chose, on passe notre fenêtre de détection à l'étage 2.
- « Faux » marqué par **F** dans le cas où le classifier ne détecte rien, on change de fenêtre de détection .la même opération sera répétée si l'étage 2 voit quelque chose on passe à l'étage 3 sinon on change de fenêtre de détection.

Le but est d'éliminer très vite les fenêtres de détection dans lesquelles il n'y a pas de visage pour économiser du temps de calcul, si tous les étages voient un visage alors, il y'a détection de visage.

On peut conclure que la méthode de Viola et Jones réduit le temps de calcul tout en obtenant une grande précision de détection. Elle a été utilisée pour construire un système de détection du visage qui est environ 15 fois plus rapide que toute approche précédente.

Conclusion

A travers ce chapitre, nous avons présenté les principales techniques de détection de visage en expliquant leurs principes de fonctionnement. Alors en se basant sur quelques recherches nous avons divisé les méthodes en quatre catégories : méthode basée sur les connaissances, les caractéristiques invariables, Template Matching et méthode basé sur l'apparence.

Ensuite, on a détaillé la méthode de Viola et Jones, cette technique basée sur l'apparence qui est capable de détecter efficacement et en temps réel des objets dans une image. La méthode de viola et Jones est l'une des méthodes les plus connues et les plus utilisées en particulier pour la détection de visages.

La liste des méthodes présentés dans ce chapitre est riche avec une multitude des méthodes et principes de fonctionnements, ce qui rend la détection de visage un domaine vaste et concurrent pour développer des méthodes efficaces et robustes.

Chapitre II : Deep Learning pour la classification des images

II.1 Introduction

Ces dernières années, le Deep Learning (apprentissage profond) attire beaucoup d'attention grâce au niveau de performance qui est extraordinaire

Il concerne les algorithmes (réseaux de neurones) inspirés par la structure et le fonctionnement du cerveau en ce qui concerne la classification des images.

Il existe deux principaux types d'apprentissage en Deep Learning : 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 notre travail on s'intéresse à l'approche supervisée basée sur les réseaux de neurones à convolution CNN, ses couches, ses avantages ainsi que les modèles les plus utilisés pour la classification d'images.

II.2 L'apprentissage profond (Deep Learning)

L'apprentissage profond « Deep Learning » est un type d'intelligence artificielle dérivé du machine Learning (apprentissage automatique) ou la machine est capable d'apprendre et de s'améliorer par elle-même selon le nombre de données. **[13]**

Les progrès de l'apprentissage profond ont été possibles notamment grâce à l'augmentation de la puissance des ordinateurs et au développement de grandes bases de données**. [14]**

L'apprentissage profond est basé sur des réseaux neurones artificiels, composés de milliers d'unités (les neurones) qui effectuent chacune de petites opérations simples. Les résultats d'une première couche de neurones servent d'entrer aux calculs d'une deuxième couche et ainsi de suite.

La figure suivante montre la relation entre les 3 concepts cités : Intelligence artificielle (AI), Machine Learning (ML) et Deep Learning (DL).

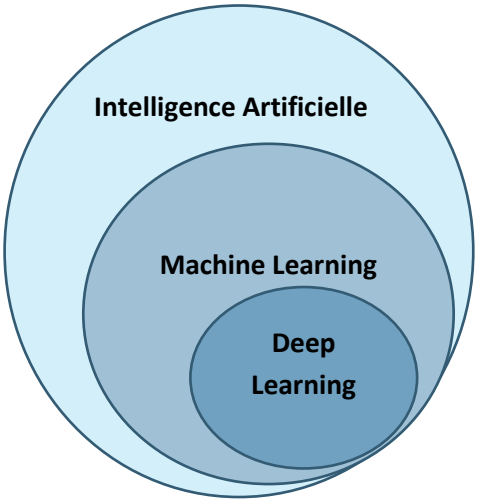


Figure 12. La relation entre l'IA, ML et le Deep Learning

II.3 Historique de Deep Learning

Le Deep Learning est un nouveau concept qui a été introduit pour la première fois par Dechtar (1986) au Machine Learning, et aux réseaux de neurones par Ainzenberg en 2000 **[29],** sa révolution est liée à la puissance des ordinateurs, également de la quantité de données qui ne cesse de s'accumuler.

On a les différentes phases par lesquelles le Deep Learning s'est développé :

- **1957 Perceptron** : Le perceptron est le réseau de neurone le plus simple, il est composé de neurones qui fonctionnent légèrement différemment que le neurone formel.
- **1986 MLP Les perceptrons multicouches** : les perceptrons multicouches ont pour objectif de classer différentes données selon leur étiquette. Pour cela le perceptron observe chacune des données qu'il possède et met à jour chaque poids de chaque neurone de son réseau afin de classifier au mieux cette base de données
- **1992 SVM Les machines à vecteurs de support** : sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression. Les SVM sont une généralisation des classifieurs linéaires. **[33]**
- **2010 Deep Neural Networks** : que l'on appelle un réseau de neurones profond est un perceptron avec au minimum deux couches cachées

II.4 Pourquoi le Deep Learning

Une des grandes différences entre le Deep Learning et les algorithmes de ML traditionnelles c'est qu'il s'adapte bien, plus la quantité de données fournie est grande plus les performances d'un algorithme de Deep Learning sont meilleures.

Autre différence entre les algorithmes de ML traditionnelles et les algorithmes de Deep Learning c'est l'étape de l'extraction de caractéristiques. Dans les algorithmes de ML traditionnelles l'extraction de caractéristiques est faite manuellement, c'est une étape difficile et coûteuse en temps et requiert un spécialiste en matière alors qu'en Deep Learning, cette étape est exécutée automatiquement par l'algorithme**. [32] [30]**

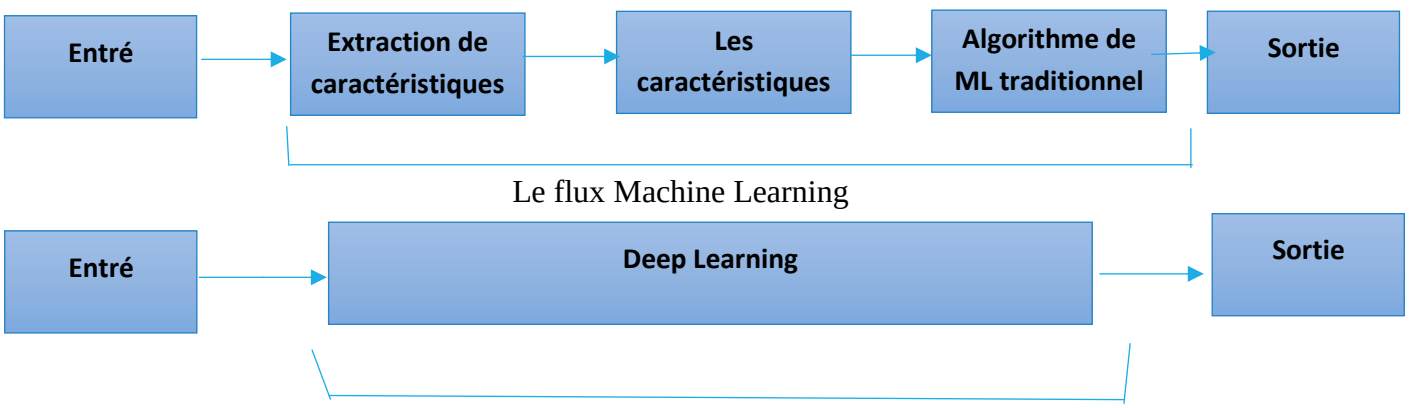


Figure 13. Comparaison entre Machine Learning et Deep Learning

II.5 Types du Deep Learning

Il existe différents types de Deep Learning, nous pouvons ainsi citer :

- **Les réseaux de neurones profonds (Deep Neural Networks) :** Ces réseaux sont similaires aux réseaux PMC mais avec plus de couches cachées. L'augmentation du nombre de couche, permet à un réseau de neurones de détecter de légères variations du modèle d'apprentissage favorisant le sur-apprentissage ou sur-ajustement.

- **Les réseaux de neurones convolutionnels (CNN ou Convolutional Neural Network) :** le problème est divisé en sous parties, la première partie réalise la convolution, et la seconde réalise la classification qui se base sur les sorties de convolution afin d'attribuer un label au niveau de la dernière couche, et qui fonctionne comme un MLP classique.

Ce type de réseau est au cœur de la plupart des systèmes de vision par ordinateur aujourd'hui, du marquage automatique de photos de Facebook, de voitures autonomes, et la classification d'image.

- **La machine de Boltzmann profonde (Deep Belief Network)** : Ce type basé sur un empilement de machines de Boltzmann restreintes**.** La **machine** de Boltzmann est un modèle de réseaux de neurones qui comporte de deux unités différentes : l'une des unités visibles et l'autre des unités cachées, complètement interconnectés. **[17]**

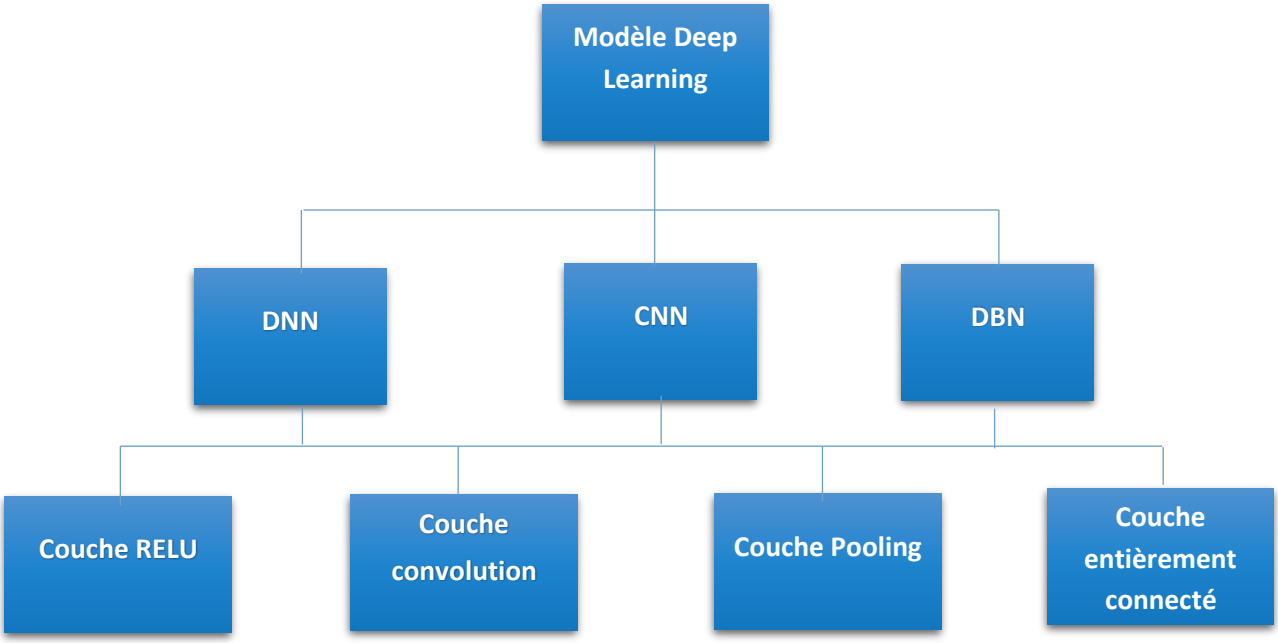


Figure 14. Différents modèles du Deep Learning

Dans notre travail, nous nous intéressons essentiellement aux réseaux de neurones convolutifs CNN qui est fondamental pour comprendre la reconnaissance de visage basée sur le Deep Learning**.**

II.6 Les réseaux de neurones convolutionnels

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 sont inspirés par des processus biologiques par les travaux Hubel et Wiesel en 1968 sur le cortex visuel des mammifères. Ces réseaux sont utilisés dans un grand nombre d'applications pour les systèmes de recommandation en traitement du langage naturel et la classification supervisée des images, cette exploitation qui a connu un grand succès grâce à leurs caractéristiques inspirées des systèmes visuels naturels**. [18]**

II.6.1 Les réseaux de neurones

Les réseaux de neurones ont été développés comme un modèle mathématique générique afin de modéliser les neurones biologiques. Ils comportent un certain nombre d'éléments de traitement appelé neurones.

Chaque neurone a son propre état interne interprété par la fonction d'activation. Il envoie son activation aux autres neurones sous forme de signaux. La connexion entre les neurones est réalisée via des liens orientés et pondérés**. [15]**

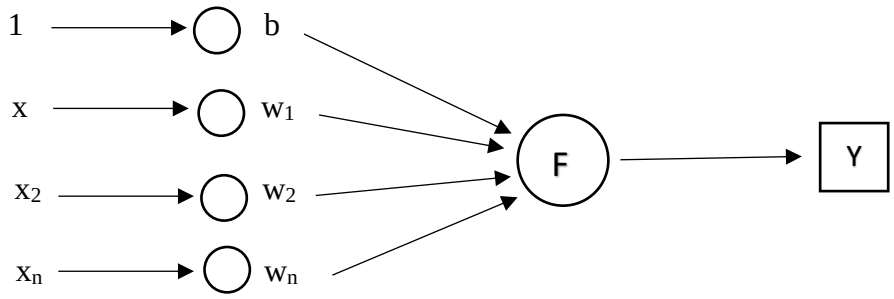


Figure 15. Modèle d'un réseau neurone

Le neurone y reçoit les valeurs d'entrés [x1 , x2….xn] , les poids des liens de connexion sont successivement [w1,w2….wn] , avec un biais =1. Son traitement consiste à effectuer à sa sortie y le résultat d'une fonction d'activation F de la somme pondérée.

La sortie du réseau est donnée par :

$$y = \sum_{i=0}^{n} w_i \ x_i + b \quad \textbf{(II.I) [16]}$$

Tel que :

x_i : les entrées du réseau

w_i : les poids synaptiques

b : le biais du réseau.

II.6.2 Architecture d'un réseau de neurone convolutif

L'architecture de base d'un réseau de neurone convolutif comporte essentiellement deux parties bien distinctes. En entrée, une image est fournie sous la forme d'une matrice de pixels. Elle a deux dimensions pour une image au niveau de gris, la couleur est représentée par une troisième dimension de profondeur 3 pour représenter les couleurs [Rouge, Vert, Bleu].

La première partie d'un réseau CNN est la partie convolutive. Elle fonctionne comme un extracteur de caractéristiques des images. Une image est passée à travers d'une succession de filtres, ou noyaux de convolution, créant de nouvelles images appelées cartes de convolution. Certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum local.

En fin, les cartes de convolution sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN.

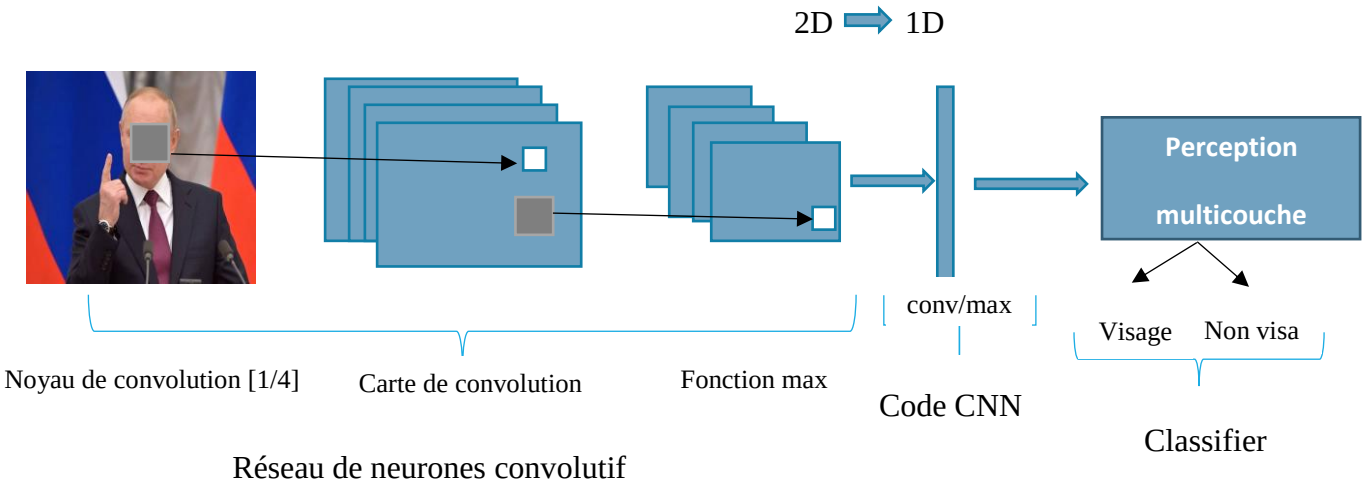


Figure 16. Architecture des réseaux de neurones convolutifs

Ce code CNN en sortie de la partie convolutive est ensuite branchée 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**. [19]**

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

II.7 Principe de fonctionnement d'un réseau CNN

Un réseau de neurone convolutif applique un ensemble d'opérations à une image afin d'en extraire les informations pertinentes et de les classifier.

Ces types d'opérations sont les suivantes :

- La convolution
- Le Pooling
- La fonction d'activation de type ReLU
- L'opération Flattening
- Couche entièrement connectée (fully-connected)

II.7.1 Couche de convolution (CONV)

La couche de convolution est l'une des composantes les plus fondamentales et les plus importantes d'un réseau neuronal.

Son but est d'extraire les caractéristiques dans les images reçues en entrée et produit le résultat sous forme d'une carte de caractéristiques. Pour cela, on réalise un filtrage par convolution.

Le principe est de faire passer un filtre sur l'image, et de calculer le produit de convolution entre ce filtre et chaque portion de l'image balayée c'est-à-dire une multiplication matricielle par éléments et une somme sur la matrice résultante**. [13]**

La figure ci-dessous nous montre le processus de convolution :

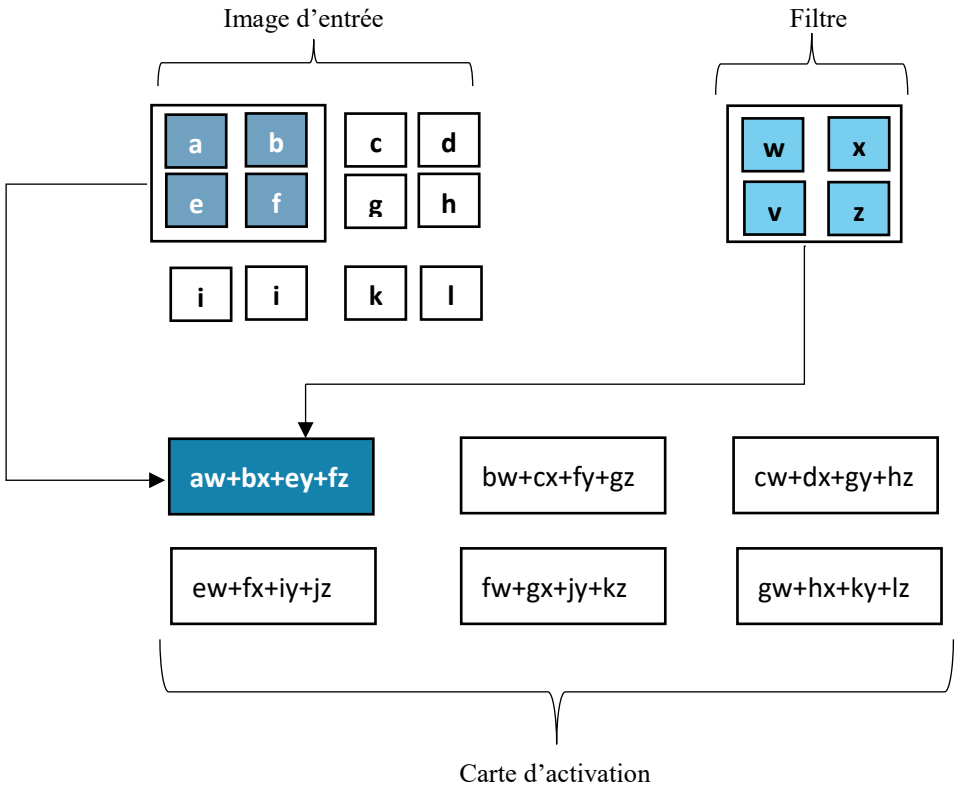


Figure 17. Exemple d'une convolution 2D avec un pas =1

II.7.2 Couche de Pooling

L'opération de Pooling est une technique qui permet de réduire la taille d'une image par extraire une valeur unique d'une région de valeurs tout en préservant ses caractéristiques importantes **[13].** La valeur extraite dépend du type de regroupement utilisé, les types de regroupement les plus courants sont au maximum (max) Pooling pour extraire la valeur la plus élevée, et Pooling moyen (moy) qui extrait la valeur moyenne d'une région.

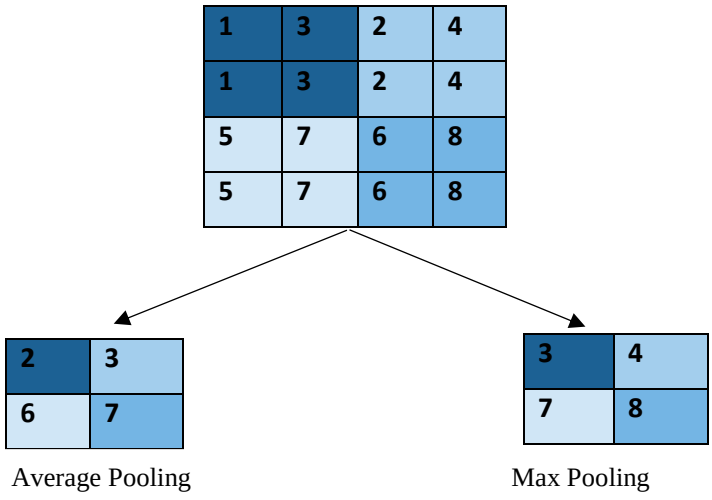


Figure 18. Les types de Pooling

II.7.3. La couche de correction ReLU

L'application de filtres de convolution génère souvent des intensités négatives, une fonction d'activation (correction) alors est utilisée afin d'améliorer l'efficacité du traitement du réseau neurone, cette fonction permet de remplacer toutes les valeurs négatives reçues en entrée de cette couche par des zéros.

La fonction ReLU (Rectified Linear Units) est donnée par l'équation suivante :

$$F(x) = \max(0,x) \qquad (2.1)\ [14]$$

II.7.4 L'opération Flattening

Les sorties que nous obtenons après les opérations de convolution et de Pooling, sont transformés en vecteur : cette opération est appelée Flattening. Elle consiste à regrouper toutes les caractéristiques de notre image extraite en un seul vecteur. **[22]**

II.7.5 Couche entièrement connectée (fully-connected)

Le vecteur appelé code CNN obtenu en sortie de la partie convolutive est ensuite branché en entrer d'une deuxième partie de classification, son objectif est d'attribuer à chaque échantillon de données une étiquette décrivant sa classe d'appartenance.

II.7.5.1 Classification des données

Diverses méthodes de classification existent. Elles peuvent etre classées en deux catégories : méthodes probabilistes qui permettent le calcul de l'appartenance d'un objet à une classe particulière et les méthodes qui séparatistes qui cherchent les frontières afin de séparer les classes. Dans notre cas, on s'intéresse à la deuxième catégorie basée sur le modèle réseau neurone perception multi couche PMC.

II.7.5.2 Description de la méthode utilisée

Le but est de réaliser une discrimination supervisée des données en deux classes. Cela signifie que les données sont étiquetées de façon que chaque objet est affecté à une classe spécifique. Pour cela, on a choisi d'utiliser un réseau de neurone à perception multi couche.

Ce réseau est composé d'une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie, chaque neurone est connecté avec les neurones de la couche suivante et le nombre de neurones de chaque couche est choisi par l'utilisateur. **[16]**

La figure ci-dessous donne l'exemple d'un réseau contenant 3 entrées, deux couches cachées et une couche de sortie.

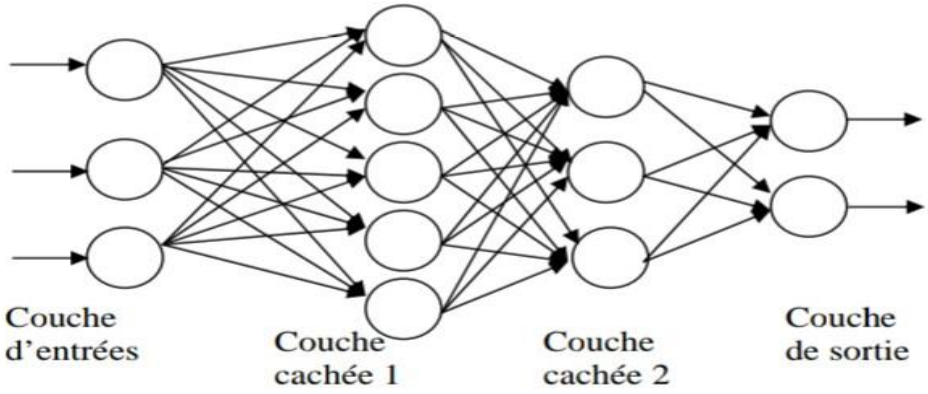


Figure 19. Exemple d'un réseau PMC [16]

II.8 Choix des paramètres

Il existe une longue liste de paramètres qui doivent être définis manuellement pour permettre au CNN d'avoir de meilleurs résultats**.**

II.8.1 Nombre de filtres

En pratique, un CNN apprend seul les valeurs de ses filtres pendant le processus d'entrainement. Les paramètres tel que le nombre de filtres, sont manuellement avant la tâche d'entrainement. Plus nous avons des filtres, plus les caractéristiques d'images sont extraites et plus le réseau devient performant au niveau de l'extraction des caractéristiques et de la classification des images.

II.8.2 Forme du filtre

Ils sont généralement choisis en fonction de l'ensemble des données. Les meilleurs résultats sur les images (28×28) sont habituellement dans la gamme de (5×5) sur la première

couche, tandis que les ensembles de données d'images naturelles ont tendance à utiliser de plus grands filtres de 12×12, voire 15×15. **[27]**

II.8.3 Forme de Max Pooling

Les 4 valeurs typiques sont 2×2. De très grands volumes d'entrée peuvent justifier un Pooling 4×4 dans les premières couches. Cependant, le choix de formes plus grandes va considérablement réduire la dimension du signal, et peut entrainer la perte de trop d'informations**.**

II.9 Les architectures neuronales convolutifs les plus utilisés

Il existe plusieurs modèles des réseaux CNN dont l'efficacité varie en fonction de la tâche, le nombre de convolution et la structure spécifiée.

Citons notamment :

II.9.1 LeNet

Réseau utilisé pour la reconnaissance de l'écriture manuscrite. Il se compose de 7 couches, sans compter la couche d'entrée. Les images d'entrée utilisées étaient de taille 32×32 pixels. La capacité de traiter des images à plus haute résolution nécessite des couches plus grandes et plus convolutive, de sorte que cette technique est limitée par la disponibilité des ressources informatiques**. [23]**

II.9.2 AlexNet

C'est l'un des premiers travaux à avoir popularisé les réseaux convolutifs. Le réseau avait une architecture très similaire à LeNet, mais étais plus profond et plus grand**. [26]** Cette architecture utilise 5 couches de convolution et trois couches de Pooling. La taille des noyaux de convolution est variable (11×11, 5×5, 3×3) en fonction de la couche considérée. **[24]**

II.9.3 VGGNet

VGGNet est un réseau qui se compose de 16 couches convolutives, très attrayant en raison de son architecture uniforme. Semblable à AlexNet mais avec une taille de filtres plus petite (3 × 3), considéré le choix le plus préféré pour extraire les caractéristiques d'une image.

Cependant, VGGNet se compose de 138 millions de paramètres, ce qui n'est pas facile à gérer**. [23]**

II.9.4 GoogLeNet

Cette architecture de CNN permet une réduction du temps de calcul par rapport à l'architecture VGG présentée précédemment. Pour cela, le GoogLeNet est composé de plusieurs couches appelées couches d'interception. Elles sont composées de plusieurs modules de convolution exécutés en parallèle sur la carte de caractéristiques résultant de la couche précédente. **[24]**

II.9.5 ResNet

Ce type de réseau CNN (réseau résiduel) permet l'apprentissage de réseaux très profonds : plus de 150 couches avec une complexité inférieure à VGGNet. L'idée développée dans ResNet est l'utilisation de connexions résiduelles permettant une meilleure optimisation des réseaux très profonds. Une connexion résiduelle permet de passer l'entrée dans deux filtres de convolution mais également de passer directement cette entrée aux couches suivantes. **[24]**

II.9.6 Unet

Ce modèle fameux de la forme U, issu du réseau de neurone convolutif traditionnel développé pour la segmentation des images biomédicales. Il permet la localisation et la segmentation en effectuant la classification sur chaque pixel, de sorte que l'entrée et la sortie partagent la même taille**. [20]**

II.9.7 SegNet

SegNet est un réseau neuronal à convolution profonde pour la segmentation d'images par pixels. Il est composé d'un certain nombre de couches représentant le réseau de codeurs et d'un réseau de décodeurs correspondants disposés les uns après les autres, suivis d'une couche finale de classification**. [20]**

II.10 Avantages des réseaux CNN

- Les réseaux de neurones convolutifs utilisent relativement peu de prétraitements, cela signifie que le réseau est responsable de faire évoluer tout seul ses propres filtres, ce qui n'est pas le cas avec d'autres algorithmes plus traditionnels.
- L'absence de paramétrage initial et d'intervention humaine est un atout majeur des CNN**. [14]**

Conclusion

Dans ce chapitre, nous avons présenté le Deep Learning qui utilise les réseaux de neurones convolutionnels les plus répondus.

Nous venons de voir que le fonctionnement d'un réseau CNN se décompose en deux parties : une partie de mise en évidence des caractéristiques d'un visage qui est gérée par la couche de convolution, et une partie de classification et de reconnaissance.

Nous pouvons conclure donc qu'un réseau de neurone convolutif présente un modèle très performant capable d'extraire des caractéristiques d'images présentés en entrée et de les classifier en sortie.

Chapitre III : Implémentation d’une application de classification

III.1 Introduction

Après avoir traité dans les chapitres précédents les principales techniques de détection du visage , l'apprentissage profond en basant sur les réseaux de neurones convolutifs, nous consacrons ce troisième chapitre à l'aspect expérimental du travail théorique lié principalement à un réseaux de neurone convolutifs **CNN,** capable de classifier des images et cela ne peut se faire qu'en réalisant des tests et des expériences à l'aide du logiciel de programmation MATLAB afin d'améliorer les performances du modèle en terme du temps et d'efficacité.

III.2 Présentation des outils de développement

III.2.1 Langage de programmation utilisé

Il existe plusieurs langages de programmation parmi ces langages on a choisi d'utiliser le langage de programmation MATLAB (Matrix Lonction MATLAB ABoratory) dans notre application car il représente un langage de haut niveau doublé d'un environnement de travail. Il est principalement utilisé dans les calculs scientifiques et les problèmes d'ingénierie parce qu'il permet de résoudre des problèmes numériques complexes en moins de temps requis par les langages de programmation courant, et ça grâce à une multitude de fonctions intégrées et à plusieurs programmes outils testés et regroupés selon usage dans des dossiers appelés boites à outils ou "toolbox". **[25]**

III.2.2 Exemple d'une fonction MATLAB intégrée

MatConvNet est une boite à Toolbox MATLAB implémentant des réseaux neuronaux convolutionnels (CNN) pour les applications de vision par ordinateur. C'est simple, efficace et peut fonctionner et apprendre des CNN à la fine pointe de la technologie. CNN préformés pour la classification d'image, la segmentation et la reconnaissance de visage.

III.3 Description du système

Notre intérêt par ce travail consiste essentiellement à la détection du visage, ensuite la classification des images selon trois cas : Avec masque, sans masque et masque incorrecte, pour cela nous répartissons notre travail en deux parties :

- Extraction de caractéristiques et détection du visage en utilisant l'algorithme de Viola-Jones
- Classification des visages avec l'approche CNN

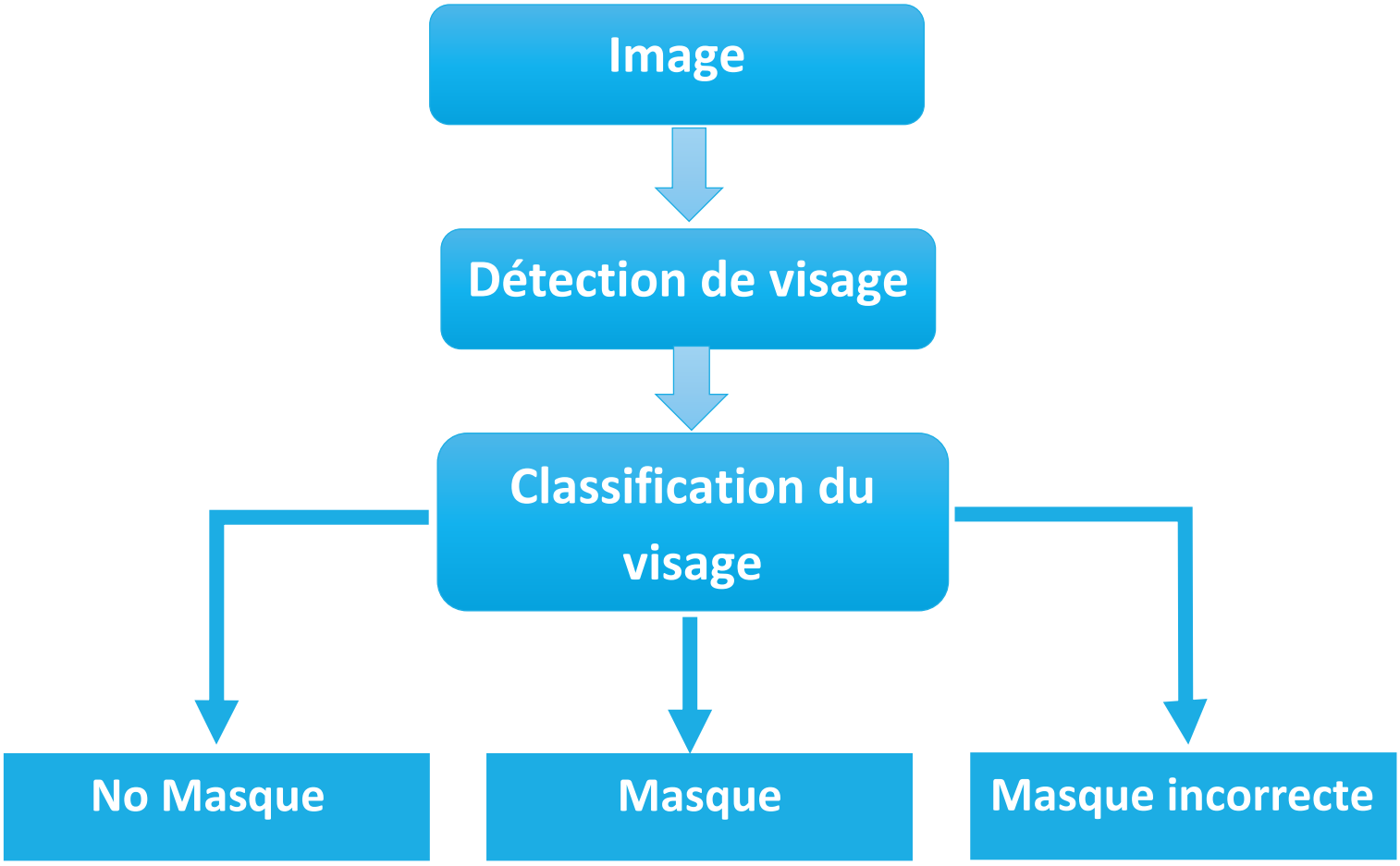


Figure 20. Description du système utilisé

L’efficacité d’un système dépend essentiellement de la méthode utilisée pour localiser le visage dans une image. C’est pour cela que nous optons pour la technique Viola-Jones afin de réaliser la phase de détection.

Si le visage est bien détecté, le système lance le processus d’extraction des caractéristiques qui va convertir les données des pixels à des représentations plus réduite pour que l’information extraite soit utilisée dans le processus de la classification.

L’objectif de l’étape de la classification est d’avoir une décision en fonction des caractéristiques extraites. Cette étape est basée sur l’utilisation de la technologie de réseau de neurones convolutionnels (CNN).

III.3.1 Détection du visage

Le bloc détection du visage, comme son nom l’indique a pour principal objectif de détecter un ou plusieurs visage(s) sur une image

Pour effectuer cette cruciale tache, de détection de visage, nous avons utilisé l'approche existante de l’algorithme Viola-Jones qui implémente le classifier **« vision.CascadeObjectDetector »** disponible dans MATLAB .Toutefois , elle ne détecte dans la majeure partie des cas que les visages frontaux ,c’est pour cette raison , que l’angle de prise de vue de l’image doit être le meilleur possible .

III.3.1.1 les images de test

Les images d’entrées choisies pour la phase de détection sont les suivantes :

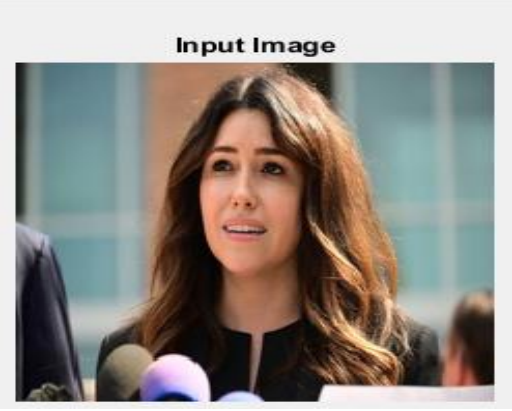


Figure 21. Image test contenant un seul visage [41]

Figure 22. Image test contenant plusieurs visages [42]

III.3.1.2 Détection du visage

La détection du visage par la méthode de Viola et Jones nous a permis d'obtenir les résultats présentés ci-dessous :

- **Cas d'image à un seul visage :**

Figure 23. Résultat d'application de la méthode Viola et Jones sur une image contenant

.

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

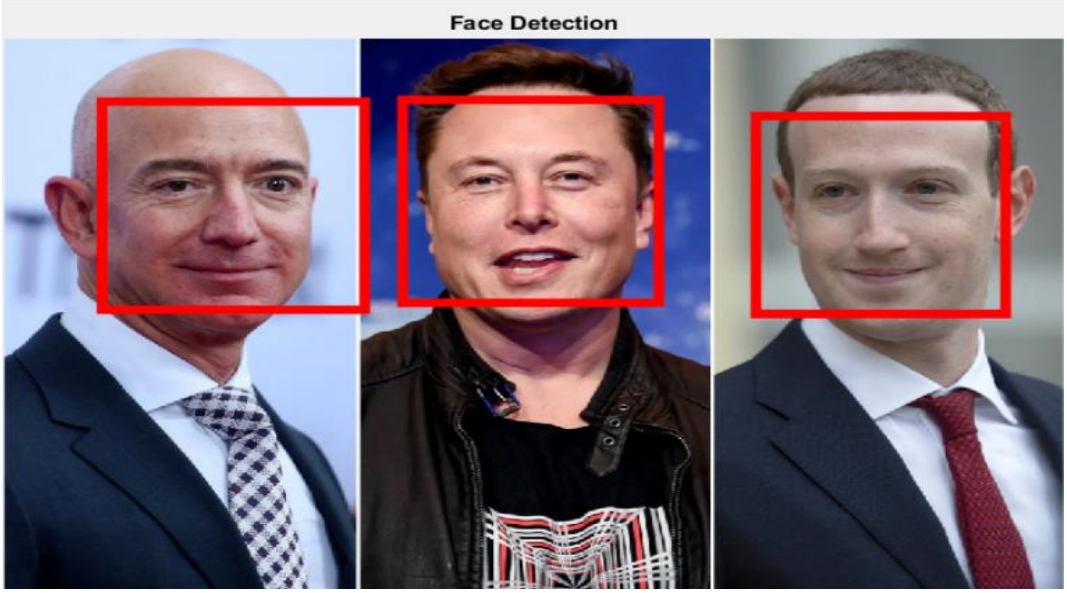


Figure 24. Résultat d'application de la méthode Viola et Jones sur une image contenant plusieurs visages

Les résultats obtenus montrent la capacité et l'efficacité de la méthode appliquée. Avec la méthode Viola et Jones on arrive à détecter le visage sur toutes les images que ce soit elles contiennent un seul visage ou plusieurs visages.

L'algorithme de Viola-Jones utilise une approche logique simple de détection des masques faciaux en détectant la présence ou l'absence des traits du visage.

- Il utilise l'idée de base que si le masque est porté correctement, les caractéristiques du nez et de la bouche ne seront absolument pas visibles.
- Par conséquent, si le modèle détecte le nez, la bouche ou le visage, la conclusion est que le masque n'est pas porté ou le maque est porter incorrectement.

Et afin de réaliser cette tache nous avons implémenté les fonctions propres à l'algorithme de Viola-Jones :

- « NoseDetect =vision.CascadeObjectDetector('Nose','MergeThreshold',16)» pour la détection du nez.
- « MouthDetect = vision.CascadeObjectDetector('Mouth','MergeThreshold',16)» pour la détection de la bouche.

• Ce modèle travaille directement sur les images en les important depuis la base de données « basedonnees ».

Les résultats obtenus sont montrés dans les figures ci-dessous :

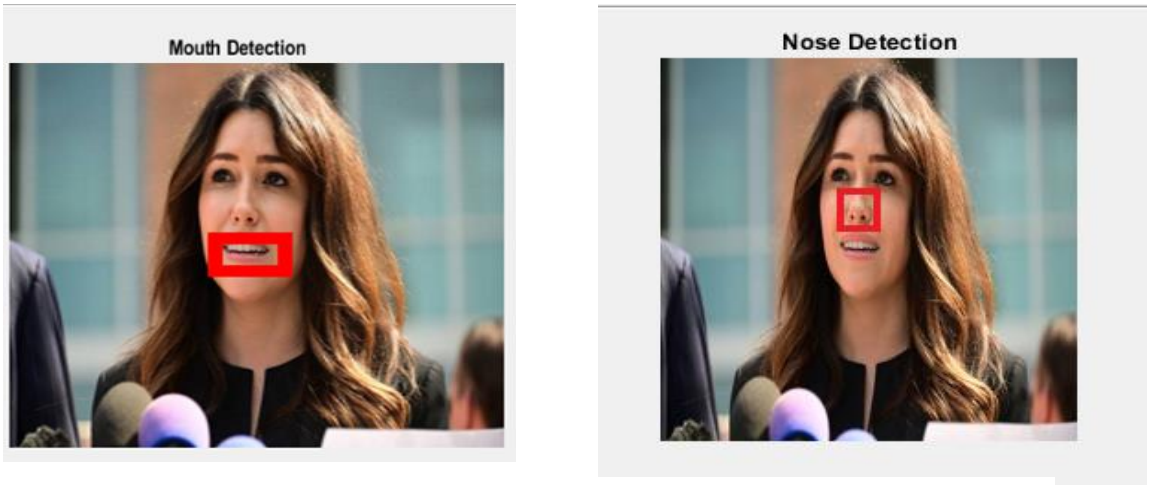


Figure 25. Détection du nez et la bouche dans le cas d'un seul visage

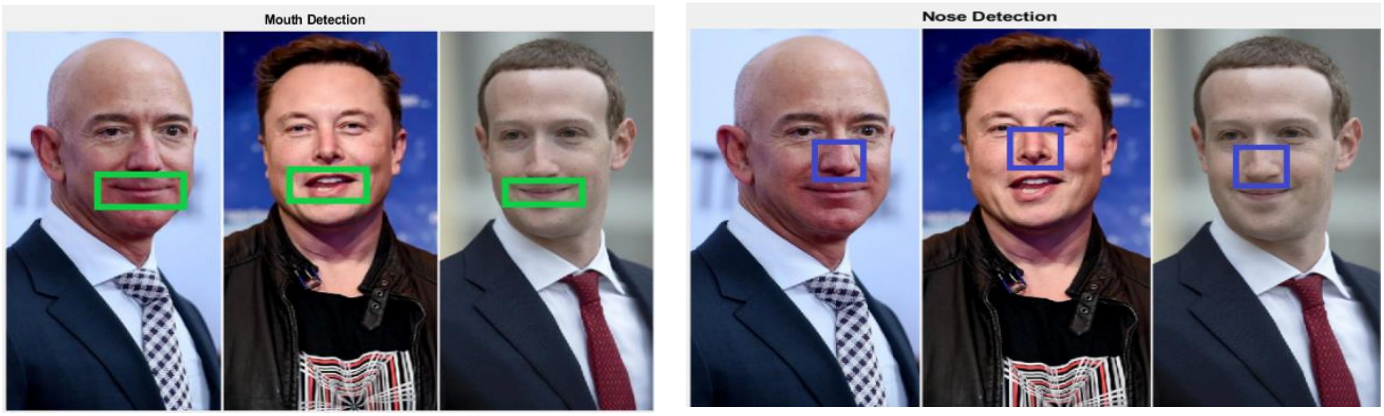


Figure 26. Détection du nez et la bouche dans le cas de plusieurs

D'une façon bien précise nous voyons bien que l'algorithme Viola-Jones arrive à détecter efficacement certaines caractéristiques comme le nez et la bouche. Lorsque ces derniers sont détectés, cela indique que le masque est absent (il n'est pas porté) et lorsque l'algorithme arrive à les détecter cela indique que le masque est présent.

III.3.2 Classification du visage

La réalisation de notre méthode de classification se fait en deux étapes comme le montre la figure ci-dessous :

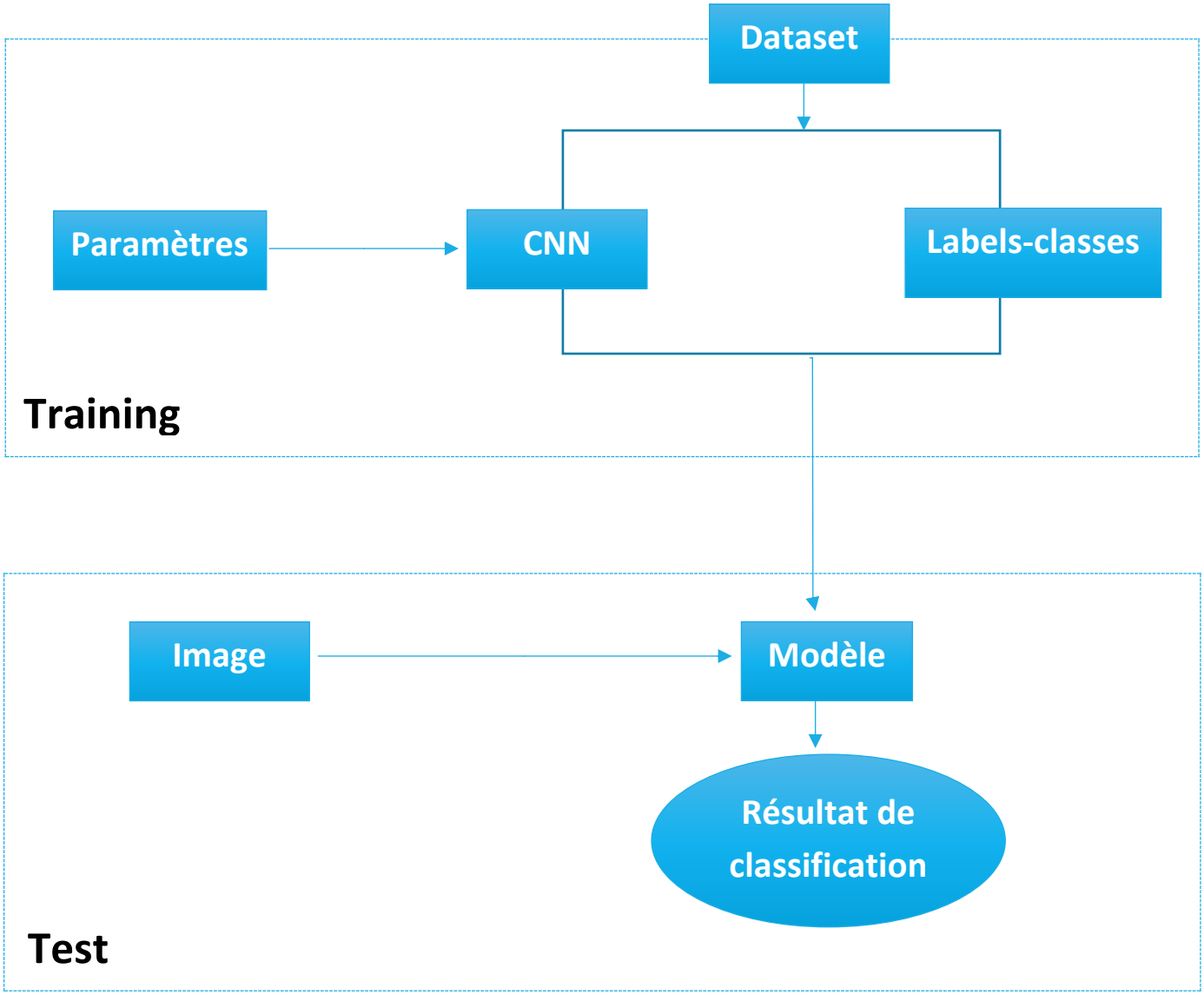


Figure 27. Conception du classifier

On remarque qu'il y'a deux grands processus qui permettent de réaliser notre expérience :

a. **Training :** c'est le processus le plus important parce qu'on va créer notre modèle grâce à des configurations précises.
    - **La Dataset :** c'est une base de données d'images répertoires en classes. Dans notre cas, on prend une Dataset de masque, on distingue les classes suivantes : avec masque, sans masque et masque incorrect
    - **Labels-classes :** c'est un fichier texte qui portera les noms des classes de notre Dataset.
    - **Réseau CNN :** on va exécuter le Dataset sur notre algorithme CNN pour générer un modèle puis on va utiliser ce modèle pour le Test

b. **Test :**

Dans le processus Test on retrouve :

- L'image : c'est l'entrée pour les tests. Ce sont plusieurs images ou une seule
- Le modèle : c'est un fichier généré dans notre training
- L'affichage de la classification : son nom résume son travail, on va afficher le résultat sorti du modèle qui est le nom d'une classe.

III.4 Présentation de l'application

Nous allons maintenant présenter les résultats obtenus grâce à des expériences réalisés par l'application de notre approche de classification des images.

La procédure de notre système est la suivante :

1. Création des classes.
2. La classification et l'apprentissage
3. Décision

III.4.1 Création des classes

Lorsque que l'on crée un modèle de réseaux de neurones convolutifs, nous devons l'entrainer avec des données afin qu'il puisse apprendre et être capable de reconnaitre un objet en particulier, pour cela on utilise une base de données contenant toutes nos images.

- Description de la base de données :

III.4.1.1 La base de données d'entrainement :

Utilisée pour entrainer le modèle CNN

Dans ce cas nous souhaitons reconnaitre 3 classes différentes : avec masque, sans masque et masque incorrecte. Chaque image de notre base de données doit appartenir à une de ces trois classes.

Afin de constituer une base de données pour le classifieur CNN nous avons pris des image prises de nos propres visages et d'autre sur google.

La figure si dessous représente quelques échantillons de notre base de données utilisé :

Figure 28. Échantillons de l'ensemble des trois classes avec masque, sans masque et masque

III.5 Architecture du réseau proposé

Nous avons effectué plusieurs tests afin de retenir la meilleure architecture CNN conduisant au meilleur taux de classification. Pour cela, nous avons varié le nombre de couches et nous avons opté pour un modèle composé de 15 couches comprenant des couches telles que : Imageinput, Convolutionlayer , Max Pooling layer , ReLu layer et SoftMax (sigmoid activation function) layer pour la classification et la minimisation des erreurs.

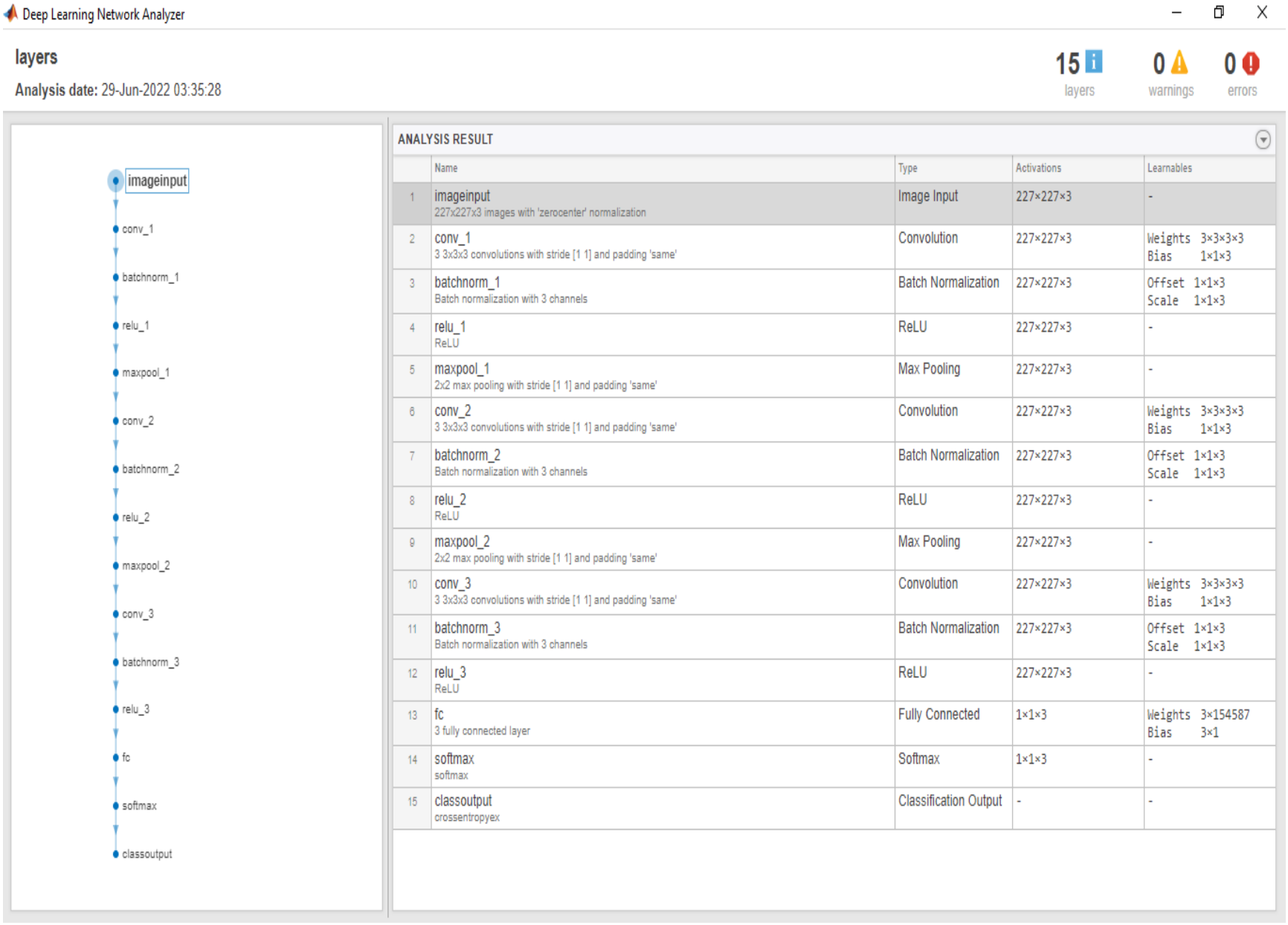
Le modèle implémenté que nous présentons dans la figure suivante est composé de trois couches de convolution, deux couches de max Pooling et d'une couche entièrement connectée (fully connected).

L’image en entrée est de taille 227*227, l’image passe d’abord à la première couche de convolution de taille 3*3, Chacune des couches de convolution est suivie d’une fonction d’activation ReLU cette fonction force les neurones à retourner des valeurs positives, après cette convolution, des cartes de caractéristiques (feature maps) seront créés.

Les feature maps qui sont obtenus auparavant ils sont donnés en entrée de la deuxième couche de convolution de la même taille aussi de 3*3, une fonction d’activation RELU est appliquée sur la couche de convolution, ensuite on applique Maxpooling pour réduire la taille de l’image ainsi la quantité de paramètres et de calcul. À la sortie de cette couche, nous aurons des feature maps de taille réduite.

La même opération sera répétée au niveau de la couche de convolutions trois, la fonction d’activation ReLU est appliquée toujours sur chaque convolution. À la sortie de cette couche, nous aurons feature maps de taille encore plus réduite que celle obtenue au niveau de la couche de convolution deux. Ces cartes de caractéristiques appelés aussi carte de convolution sont concaténées en un seul vecteur contenant ces caractéristiques.

Finalement, un softmax qui permet de calculer la distribution de probabilité des classes suivantes : avec masque, sans masque ou masque incorrect est appliqué.



	Name	Type	Activations	Learnables
1	imageinput 227x227x3 images with 'zerocenter' normalization	Image Input	227×227×3	-
2	conv_1 3 3x3x3 convolutions with stride [1 1] and padding 'same'	Convolution	227×227×3	Weights 3×3×3×3 Bias 1×1×3
3	batchnorm_1 Batch normalization with 3 channels	Batch Normalization	227×227×3	Offset 1×1×3 Scale 1×1×3
4	relu_1 ReLU	ReLU	227×227×3	-
5	maxpool_1 2x2 max pooling with stride [1 1] and padding 'same'	Max Pooling	227×227×3	-
6	conv_2 3 3x3x3 convolutions with stride [1 1] and padding 'same'	Convolution	227×227×3	Weights 3×3×3×3 Bias 1×1×3
7	batchnorm_2 Batch normalization with 3 channels	Batch Normalization	227×227×3	Offset 1×1×3 Scale 1×1×3
8	relu_2 ReLU	ReLU	227×227×3	-
9	maxpool_2 2x2 max pooling with stride [1 1] and padding 'same'	Max Pooling	227×227×3	-
10	conv_3 3 3x3x3 convolutions with stride [1 1] and padding 'same'	Convolution	227×227×3	Weights 3×3×3×3 Bias 1×1×3
11	batchnorm_3 Batch normalization with 3 channels	Batch Normalization	227×227×3	Offset 1×1×3 Scale 1×1×3
12	relu_3 ReLU	ReLU	227×227×3	-
13	fc 3 fully connected layer	Fully Connected	1×1×3	Weights 3×154587 Bias 3×1
14	softmax softmax	Softmax	1×1×3	-
15	classoutput crossentropyex	Classification Output	-	-

Figure 29. Architecture de gen_net pour 15 couche

III.6 Résultats obtenus et discussion

Afin de montrer les résultats obtenus pour notre modèle, on illustre dans ce qui suit les résultats en termes de précision et d'erreur par rapport au nombre d'itération ainsi que la matrice de confusion.

III.6.1 Graphe de précision et d'erreur

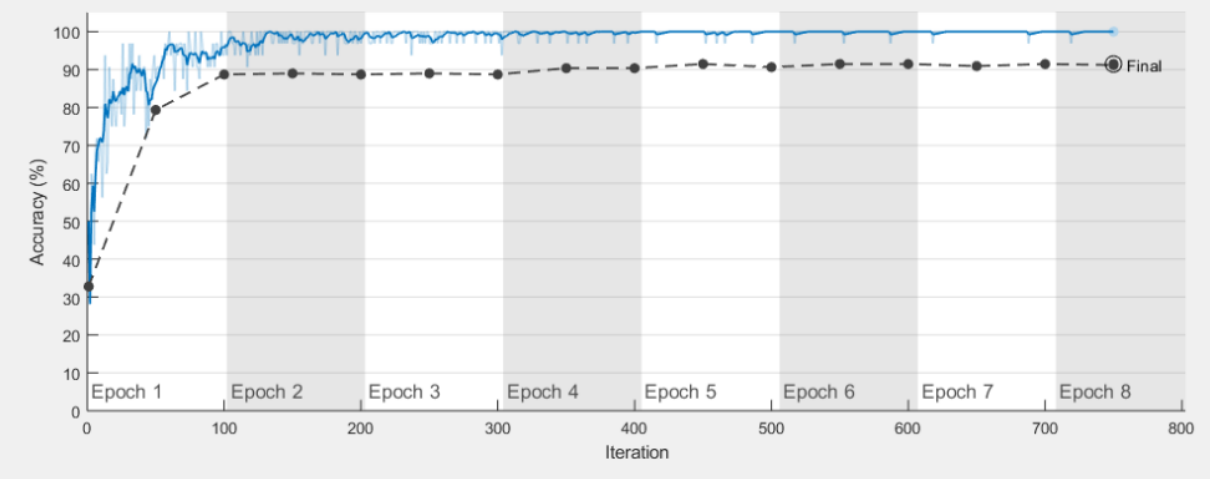


Figure 30. Graphe de précision

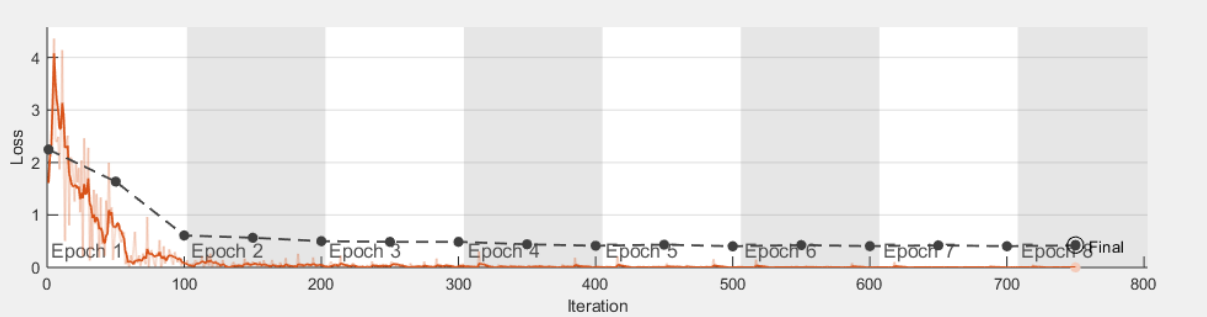


Figure 31. Graphe d'erreur

Les figures ci-dessus représentent la courbe de précision et de perte d'entrainement du modèle CNN avec un nombre d'Epochs égale à 10. Nous pouvons voir que la précision d'entrainement augmente à chaque itération jusqu'à atteindre 91.5% de précision, ceci explique qu'à chaque itération le modèle apprend plus d'information sur les images, pour mieux les classer et améliorer la précision ainsi que la performance du modèle.

D'autre part la ***Figure III.13***, montre que l'erreur de classification diminue au fur et à mesure à chaque itération.

III.6.2 Matrice de confusion

La matrice de confusion permet d'évaluer les performances de notre modèle, puisqu'elle reflète les métriques du Vrai positif, vrai négatif, faux positif et faux négatif. La figure illustre la position de ces métriques pour chaque classe.

L'avantage de ces métriques est qu'elles sont très simples à lire et à comprendre. Elles permettent de visualiser les données et les statiques d'une manière dynamique afin d'analyser les performances.

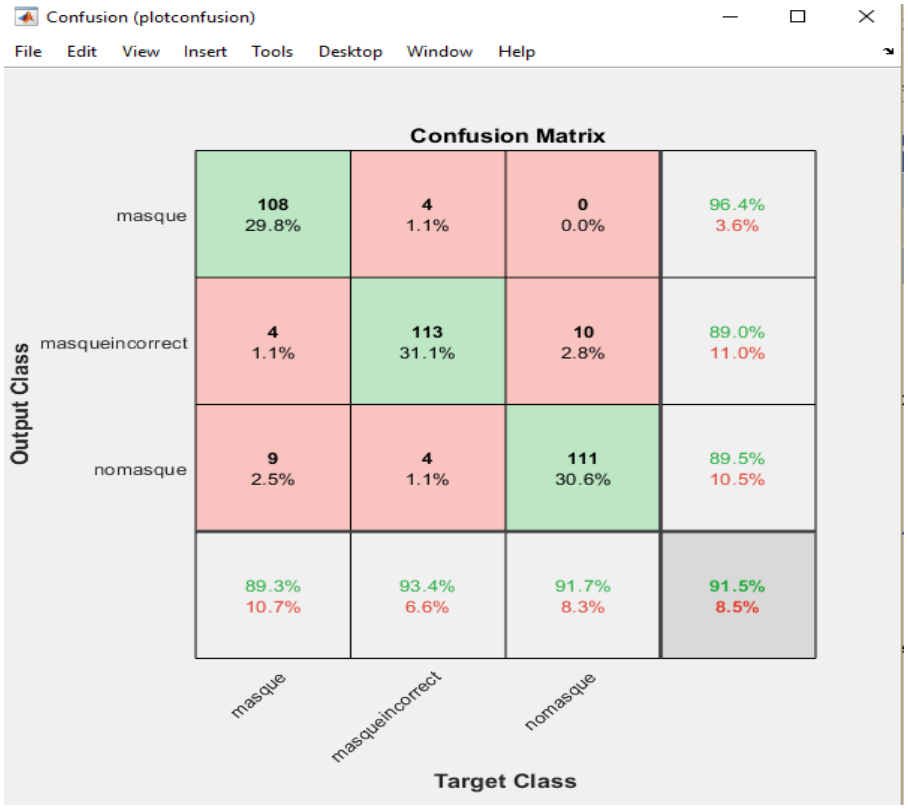


Figure 32. La matrice de confusion correspondante

La matrice de confusion de la figure résume la prédiction des images de test sur les trois classes :

Pour la classe masque : 108 images sont bien classées et 4 images mal classées ce qui donne un taux de précision de 96.8% et un taux d'erreur de 3.6%

Pour la classe masque incorrect : 113 images sont bien classées et 14 images mal classées ce qui donne un taux de précision de 89% et un taux d'erreur de 11%

Pour la classe no masque : 111 images sont bien classées et 13 images mal classées ce qui donne un taux de précision de 89.5% et un taux d'erreur de 10.5%

Les résultats montrent que le taux de la bonne classification est de **91.5%,** et le taux de classification erroné de **8.5%,** ce qui reflète la performance et l'efficacité de notre modèle.

III.7 Influence du nombre de couches de convolution

III.7.1 Modèle CNN à 10 couches

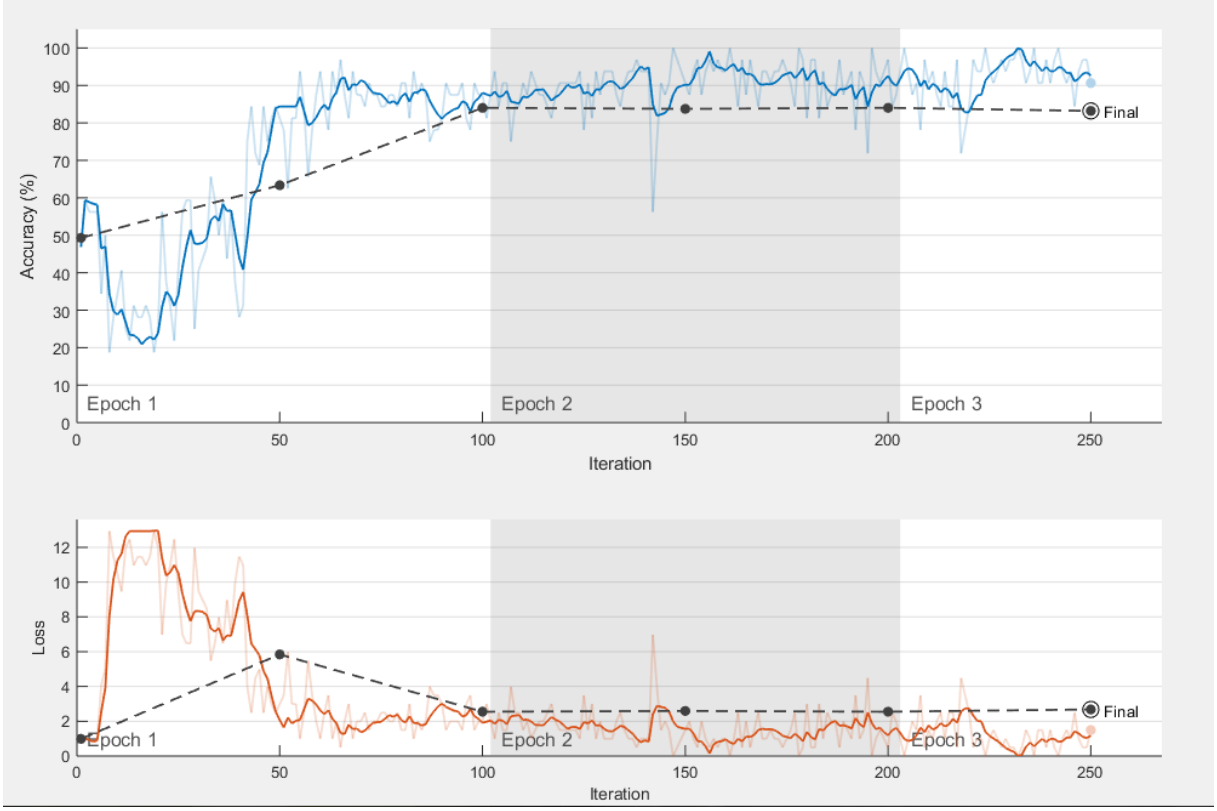


Figure 33. Graphe de précision et d'erreur du modèle 10 couches

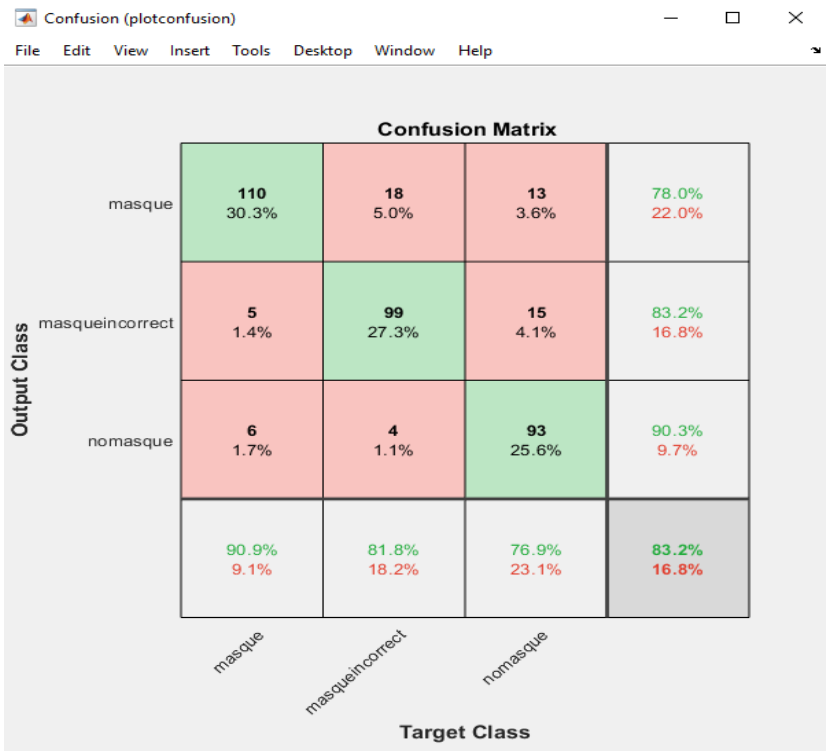


Figure 34. Matrice de confusion associée au modèle 10

La ***figure III.15*** représente les courbes de précision et de perte d'entrainement du modèle CNN à 10 couches avec un nombre d'Epochs égale à trois. D'après cette figure, nous constatons que la précision est en train d'augmenter jusqu'à atteindre 82.3% de précision, de même avec le graphe d'erreur qui diminue avec le nombre d'Epochs.

La ***figure III.16*** montre la matrice de confusion des trois classes dans le cas de la diminution du nombre de couches à 10. Nous trouvons donc que l'ensemble d'images bien classés situés sur la diagonale de la matrice pour ce modèle est de 302 images, et un totale de 61 d'images mal classés.

On remarque qu'il y'a une dégradation de la précision de la classification du modèle (83.2%) et du taux d'erreur (16.8%) par rapport au premier modèle ou la précision était de 91.5% et le taux d'erreur étais 8.5%.

III.8 Tableau de comparaison des résultats

Nombre de couches	Images bien classés	Images mal classés	Taux de précision	Taux d'erreur	Temps d'exécution
10	**302**	**61**	**82.3%**	**16.8%**	**31min,21sec**
15	**332**	**31**	**91.5%**	**8.5%**	**31 min**

Tableau 1 : Tableau de comparaison des résultats selon le nombre de couches

Nous étudions l'impact de l'augmentation du nombre de couches de convolution. Les résultats obtenus sont exprimés en termes de précision d'apprentissage, de validation, de tests et erreurs et enfin de temps d'exécution. Le temps varie selon la dimension de la base de données utilisée

D'après le modèle à **10 couches** le score de précision obtenu est **82.3%**. Par contre, avec le modèle à **15 couches**, nous arrivons à atteindre un score de précision de **91.5%**

Nous pouvons constater que le modèle à 15 couches présente les meilleurs résultats trouvés, le nombre de couches de convolution reflètent ces bons résultats.

Les résultats obtenus se sont améliorés à mesure que nous avons approfondie notre réseau et augmenté le jeu de données. La base d'apprentissage est également un élément déterminant dans les réseaux de neurones CNN, il faut avoir une base d'apprentissage de grande taille pour aboutir à des meilleurs résultats.

III.9 Influence du nombre d'image

Dans cette étape nous avons fais varier le nombre d'images de notre base de donner en se basant sur la model a 15 couche.

Pour cela, nous avons fait une comparaison entre un système de 200 images et un autre de 1207 image pour chaque classe afin d'arriver aux résultats qui nous permet d'avoir une bonne classification et de bonnes performances

- La section suivante nous montre les résultats obtenus pour 200 image :

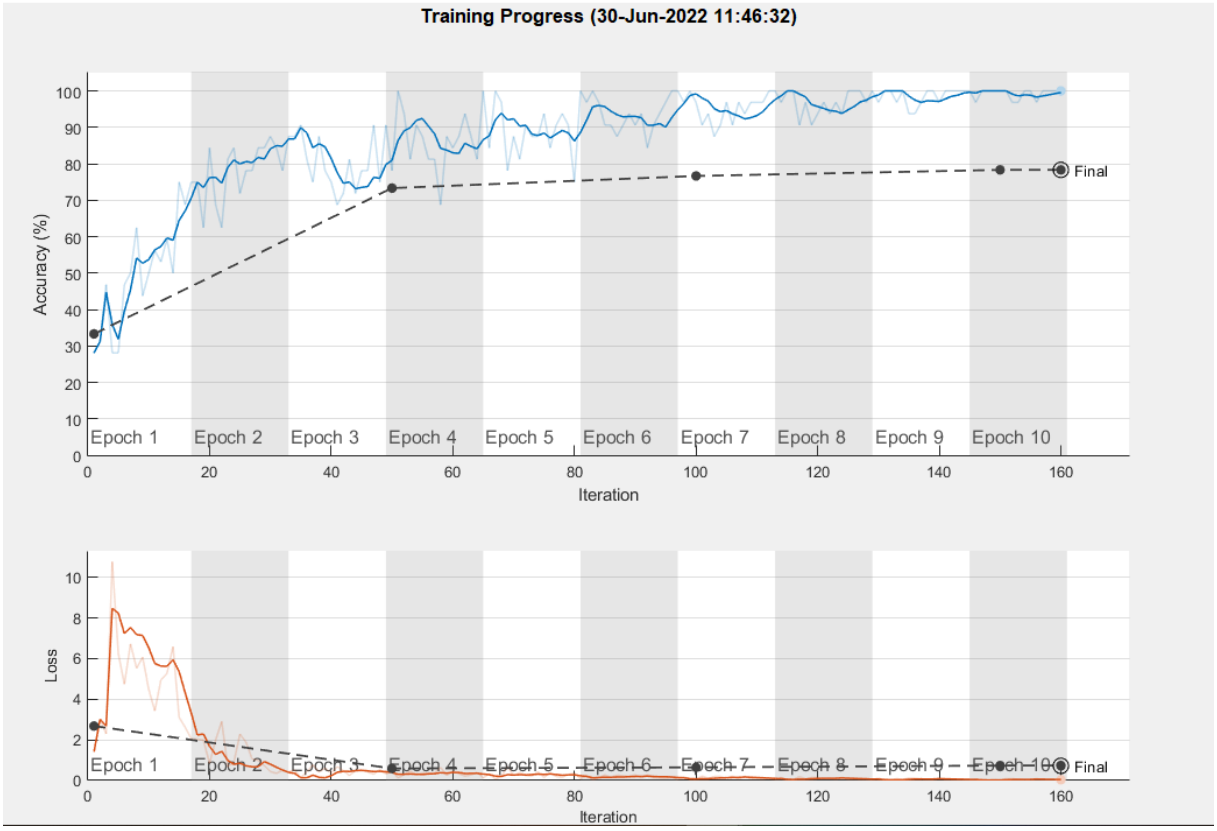


Figure 35. Le graphe de précisions et d'erreur pour 200 image

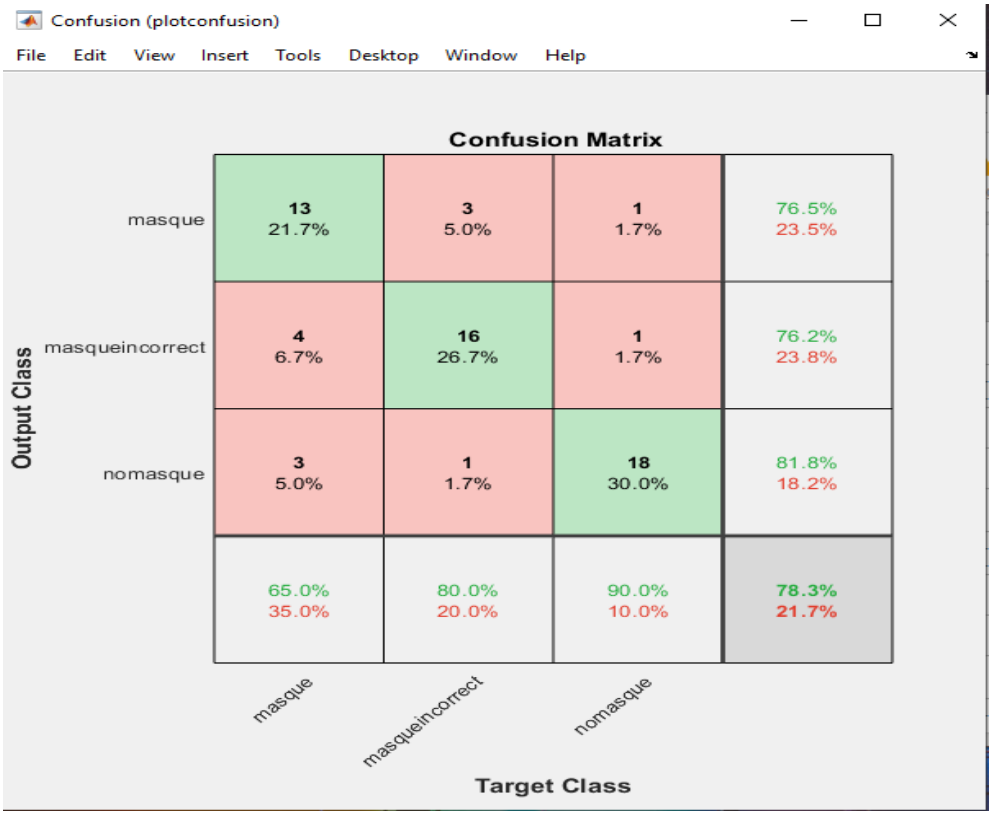


Figure 36. Matrice de confusion pour 200 images

La Figure 35 montre le graphe de précisions et de perte pour le model CNN a 200 images ; nous pouvons voir que la précision d'entrainement augmente à chaque itération jusqu'à atteindre 78.3% de précision, et de même pour le taux d'errer qui diminue pour chaque itération jusqu'à atteindre 21.7%

Les résultats obtenus sur la matrice de confusion représente l'emble d'image classer, 47 pour des images bien classés et 13 pour des images mal classer

- Le Tableau ci-dessous résume les résultats de classifications pour les deux bases de donner utiliser 200 et 1207 image pour chaque classe :

Nombre d'image	Images bien classés	Images mal classés	Taux de précision	Taux d'erreur	Temps d'exécution
200	**47**	**13**	**78.3%**	**21.7%**	**5 min et 24 sec**
1207	**332**	**31**	**91.5%**	**8.5%**	**31 min**

Tableau 2 : Tableau de comparaison des résultats selon le nombre d'image

D'après le tableau le model a 1207 images représente les meilleurs résultats par rapport au model a 200 image, de là nous pouvons déduire que la faite d'augmenter le nombre d'images permet d'améliorer les performances du système de classification

III.10 Choix des réseaux de convolution CNN

Il existe plusieurs techniques de classification comme, SVM et KNN, dans notre projet on a opté pour la méthode de classification CNN. Afin de justifier notre choix nous allons effectuer une étude comparative entre les résultats de classification du réseau étudié et les autres méthodes classiques.

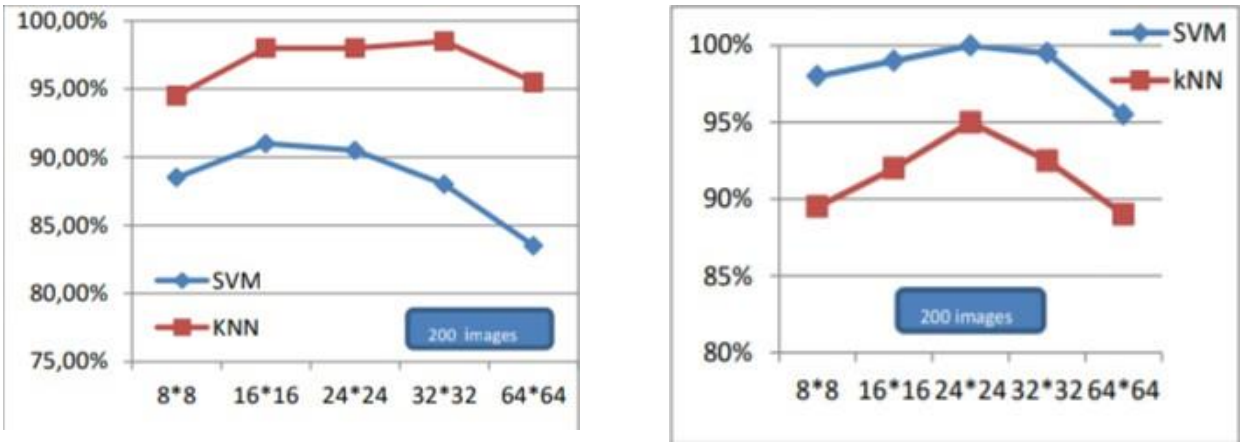


Figure 37. Représentation de taux de classification de visage sans et avec masque pour les classifier KNN et SVM

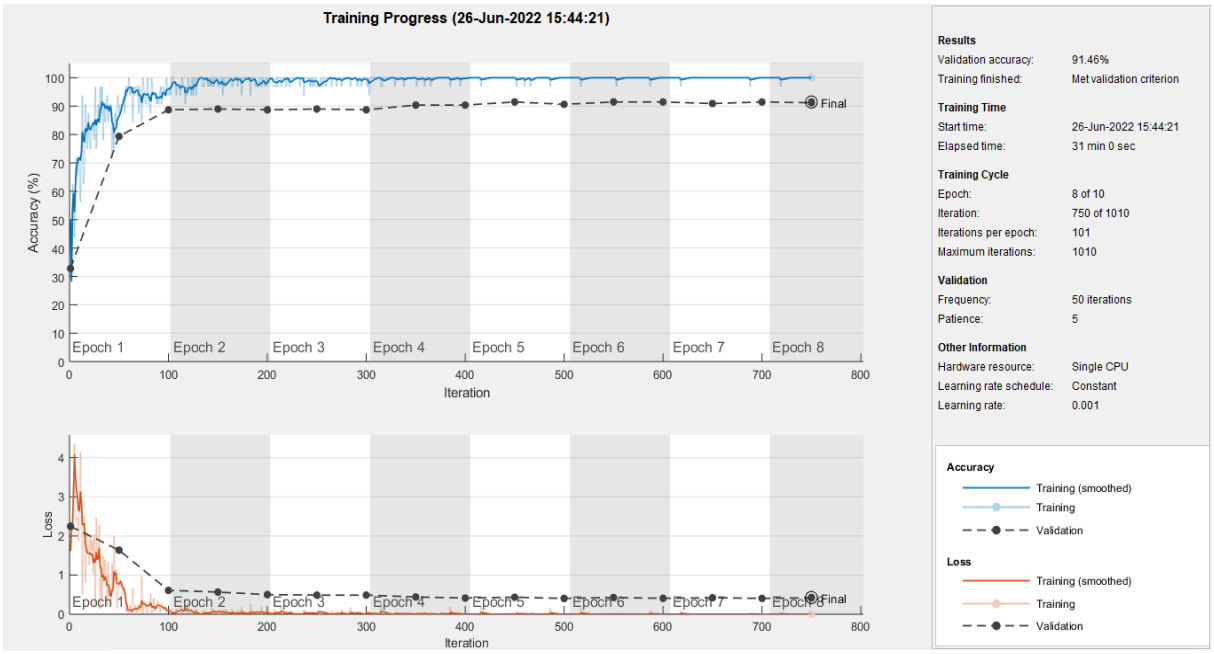


Figure 38. Représentation de taux de classification pour les classifier CNN

Le principe de fonctionnement de la méthode KNN est décrit par les paramètres suivants :

X : les entrées

D : une fonction de distance donnée par l'équation suivante :

- La distance de Manhattan : calcule la somme des valeurs absolus des différences entre les coordonnées de deux points :

$$D_m(x,y) = \sum_{i=1}^{n} |x_i - y_i|$$

K : un nombre entier

Y : prédiction de la valeur de sortie

D’après ces résultats, nous remarquons que les résultats des méthodes classiques pour la classe masque présentent un bon taux de classification sur l’analyse du bloc 24*24, mais avec moins précision pour les autres blocs (8*8, 16*16, 32*32, 64*34). De même pour les cas de la classe sans masque, les classifier arrivent à réaliser la classification que sur certains blocs avec un taux de classification plus moins par rapport à la première classe avec masque

D’autre part nous remarquons que le classifier CNN, arrive à détecter plus de deux classes, on constate que le taux de classification augmente à chaque itération et reste stable à un niveau de 100%, ce qui signifie que le classifier a appris à classifier toutes les images d’une manière plus performante par rapport aux approches statiques

Ces résultats justifient notre choix de classification avec les réseaux de neurones convolutifs, car ils permettent de gérer un grand nombre de données d’une façon efficace.

Conclusion

Nous avons présenté dans ce chapitre l’implémentation de l’approche de classification d’image basée sur des réseaux de neurones convolutifs CNN, pour cela nous avons utilisé deux modèles d’architectures, le premier à 15 couches et le second à 10 couches, qui nous permettent d’obtenir différents résultats. La comparaison des résultats trouvés a montré que la taille de la base de données et la profondeur du réseau, sont des facteurs importants pour l’obtention de meilleurs résultats.

Comparés à d’autres algorithmes de classification d’image, les réseaux de neurones convolutifs utilisent peu de prétraitement. Cela signifie que le réseau est responsable de faire évoluer ses paramètres selon la taille des donnés d’entrés, ce qui n’est pas le cas avec d’autres algorithmes traditionnels.

Conclusion générale

La classification d'images est une tâche importante dans le domaine de la vision par ordinateur.

Bien que les capacités des activités réalisés dans le domaine de classification des images soient nombreuses, aucune méthode n'est jugée faible à 100%, mais au fur et à mesure les nouveaux travaux essayent d'améliorer les scores pour de meilleurs résultats.

C'est dans ce cadre que s'inscrit notre travail, qui a pour objectif de proposer une application qui réalise une classification d'une base d'images un ensemble de classes (avec masque, sans masque et masque incorrecte).

Pour réaliser notre travail de classification on a utilisé le Deep Learning, la méthode d'apprentissage qui a montré ses performances ces dernières années et nous avons choisi la méthode CNN comme méthode de classification, ce choix est justifié par la simplicité et l'efficacité de la méthode.

Le résultat obtenu lors de la phase de test confirme l'efficacité de notre approche.

Notre travail n'est que dans sa version initiale, on peut dire que ce travail reste ouvert pour des travaux de comparaison avec d'autres méthodes de classification.

Bibliography

[1]: S. Djedi , « Etude comparative de PCA et KPCA associés au SVM en biométrie » , thèse de doctorat en informatique , université Mohamed Khider Biskra ,2012.

[2]: D.Afef , K.Delenda , « Développement d'un système de détection reconstruction et animation 3D de visage à partir d'une séquence vidéo » , mémoire , université de Sfax , juin 2011.

[3]: M.H . Yang , D.J. Kriegman , and N.Ahuja . « détecting faces in images : A survey ». IEEE Trans . on PAMI , pp 34-58, 2002.

[4]: E. Hjelmas and B.K.Low . « Face détection : A survey » . computer vision and image Understanding , pp 236-274 , 2001.

[5]: 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.

[6]: M.Zrelli , implémentation d'une méthode de détection et suivi de visage en temps réel, Ecole royale militaire bruxelles royaume de Belgique , 2006/2007.

[7]: Ch . Bencheriet , A/H . Boualleg & H. Tebbikh , B.guerziz & W.Belguidoum , « détection de visage par méthode hybride couleur de peau et template matching » , Laig ,université 8 mai 45 de Guelma BP 401 , algérie.

[8]: S.Guerfi . authentification d'individus par reconnaissance de caractéristique biométriques liées aux visages 2D/3D , thèse doctorat , université d'Evry-val d'Essonne , France , 2008.

[9]: Paul Viola and Michael Jones. Robust real-time object detection. In second international work shop on statistical and computation atheories of vision, van-couver, canada , July 13 2001.

[10]: face tracking implémentation de la méthode de viola & jones en C++ : site Web https://www.firediy.fr/article/face-tracking-implementation-de-la-methode -de-viola-jones-en-c.

[11]: A.L.C.Barazak , to xard and efficient implementation of a rotation invarient detection using Haar-like features . Proceeding of the 2009 IEEE /RSJ international conference on intelligent robotsand systems. Nouvelle Zélande, Dunedin.2005, P31-36.

[12]: M. Kolsch et M.Turk , Analysis of rotational Robustness of Hand detection with a Viola-Jones Detector, ICPR ,vol.3, 20004.

[13]: http://www.psychomedia.qc.ca/lexique/definition/apprentissage-profond.

[14]: Vinay Rao et al (2015) . Brain Tumor Segmentation with Deep Learning , université de Southern California . Janvier 2015 .

[15]: K. Gurney . an introduction to Neural Networks , 1997.

[16]: M. Demouche , A ; Ouakour , D.Aissani , B.Boudart . « Non Linear Classification with Neural Networks » Proceedings of the international conference méthodes et outils d’aide à la décision. Bejaia, pp.649-653,2007.

[17]: P. Beraud, « microsoft developer »: https://blogs.msdn.microsoft.com/mlfrance/2016/04/28/une-première-introduction-au-deep-learning/.

[18]: D. H. Hubel and T . N . Wiesel . Receptive fields , binocular interaction and functional architecture in the cat’s visual cortex. The journal of physiology.

[19]: I. Arel ,D.Rose , and T.Karnowski. « Deep machine learning-a new frontier in artificial intelligence research [research fronier] ». In : IEEE computational intelligence magazine 5.4(2010) , page 13-18

[20]: A. Kendall , V. Badrinarayanan, R. Cipolla , « Segnet : a deep convolutional encoder-decoder architecture for image segmentation » , IEEE transactions on pattern analysis and machine intelligence , 2017.

[21]: D. Rumelhart , G. Hinton , R. Wiliams , « Learning internal representations by error propagation » tech. Rep , DTIC Document , 1985.

[22]: « Convolution neural network : step 3 : Flattening » : https://www.superdatascience.com/convolutional-neural-networks-cnn-step-3-flattening/.

[23]: https://towardsdarascience.com/review-of-lenet-5-how-to-design-the-architecture-of-cnn-8ee92ff760ac.

[24]: F. Chabot . Analyse fine 2D/3D de véhicules par réseaux de neurones profonds, université Clermont auvergne, France, 2018.

[25] : http://www.jobintree.com/dictionnaire/definition-matlab-915.html

[26]: A. Krizhevsky , I . Sutskever , and G.E. Hinton . Imagenet classification with deep convolutional neural networks . In Advances in neural information processing systems , pages 1097-1105 , 2012.

[27]: https://idpoisson.fr/louchet/teaching/timo/Echegut.pdf

[28]: saad narimen, reconnaissance tridimensionnelle du vidage, Thèse de Doctorat, université Mohamed khider Bisbra , juin 2014.

[29]: Igor Aizenberg , Naum Aizenberg , and Joos Vandewall , Multi-Vlud and universal binary neurons :theory , learning and applications , Springer Science & Business Media, 2013.

[30]: http://deeplearning.stanford.edu/tutoriel/supervised/convolutioanalNeuralNetwork/

[31]: https://www.futura-sciences.com/tech/definitions/intelligence-artificielle-deep-learning-17262/.

[32]: 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.

[33]: https://www.studiosport.fr/guides/drones/tout-savoir-sur-les-drones.html.

[34]: https://lemnet.fr/blog/post/la-reconnaissance-faciale-accessible-a-tous-avec-openface-en-5-etapes-simples

[35]: https://www.genxys.com/large-group-of-people-of-divers-2/?lang=fr

[36]: https://www.shutterstock.com/fr/image-photo/set-portraits-collage-close-portrait-sexy-1927689461

[37]: https://www.researchgate.net/figure/Exemples-docclusion-du-visage-Image-recueillie-a-partir-dInternet_fig5_291345615

[38]: https://www.researchgate.net/figure/Exemple-dun-visage-dune-meme-personne-subissant-un-changement-de-luminosite-Limage_fig2_291345615

[39]: https://medium.datadriveninvestor.com/haar-cascade-classifiers-237c9193746b

[40]: https://towardsdatascience.com/understanding-face-detection-with-the-viola-jones-object-detection-framework-c55cc2a9da14

[41]: https://nouvellefr.com/camille-vasquez-revele-la-cle-de-la-victoire-judiciaire-de-johnny-depp/

[42]: https://nypost.com/2020/12/31/these-billionaires-wealth-ballooned-by-a-record-total-in-2020/